UNIVERSITATEA ECOLOGICĂ DIN BUCUREŞTI

FACULTATEA DE ECOLOGIE ŞI PROTECŢIA MEDIULUI

Program de studii universitare de master

EVALUAREA IMPACTULUI ASUPRA MEDIULUI

Forma de învăţământ: învăţământ cu frecvenţă

Material bibliografic pentru disciplina:

Utilizarea GIS pentru evaluarea impactului asupra mediului

Titular de disciplină:
Conf. dr. Giuliano Tevi

Bucureşti
2019

Introducere

Cursul este destinat studenţilor din cadrul programului de studii universitare de master Evaluarea Impactului asupra Mediului al Facultăţii de Ecologie şi Protecţia Mediului.

a. Date privind titularul de disciplină

Nume şi prenume:	**Conf. dr. Giuliano Tevi**
E-mail:	**giuliano.tevi@ueb.ro**

b. Date despre disciplină

Anul de studiu:	**II**
Semestrul:	**4**

c. Obiectivele disciplinei

Obiectivul general al disciplinei	• Însuşirea de către masteranzi a principalelor caracteristici pe care le au datele de mediu, modului lor de obţinere şi de prelucrare specifică
Obiectivele specifice	• Însuşirea noţiunilor fundamentale legate de metodele de asociere a datelor spaţiale şi a datelor atribut în cadrul bazei de date specifică aplicaţiilor GIS; • Iniţierea în utilizarea unor soft-uri de specialitate – pentru realizarea unor aplicaţii, utilizând GIS în scopul protecţiei mediului.

d. Competenţe acumulate după parcurgerea cursului

Competenţe profesionale	• Limbaj specific: o Cunoaşterea, înţelegerea şi utilizarea limbajului de specialitate. • Explicare şi interpretare: o Identificarea şi utilizarea surselor de informaţii necesare pentru întocmirea aplicaţiilor la noţiunile teoretice prezentate la curs. • Aplicare, transfer şi rezolvare de probleme: o Aplicarea noţiunilor teorectice în analiza studiilor de caz privind aspectele utilizării GIS şi produselor de teledetecţie în evaluarea impactului de mediu. • Reflecţie critică şi constructivă o Exprimarea clară a unui punct de vedere în privinţa locului GIS, ca instrument pentru studiile specifice ecologiei şi protecţiei mediului. • Conduită creativ-inovatovă: o Capacitatea de a elabora aplicaţii şi proiecte privind utilizarea GIS în protecţia mediului; • Capacitatea de a prezenta simplu şi clar, pentru un public neavizat, rezultatele aplicării GIS şi a informaţiilor obţinute prin utilizarea acestui instrument.

Competenţe transversale	• Asumarea funcției de conducere într-o echipă pluridisciplinară și aplicarea de diferite tehnici în vederea eficientizării activității echipei și / sau a instituției. • Aplicarea normelor și valorilor de etică profesională în contextul asumării responsabilității de coordonator al unor sarcini și obiective complexe. • Selectarea și utilizarea oportunităților de învățare și de formare profesională continuă, a surselor informaționale și a resurselor de comunicare.

e. Resurse şi mijloace de lucru

Pentru o pregătire temeinică vă sugerăm să consultaţi:

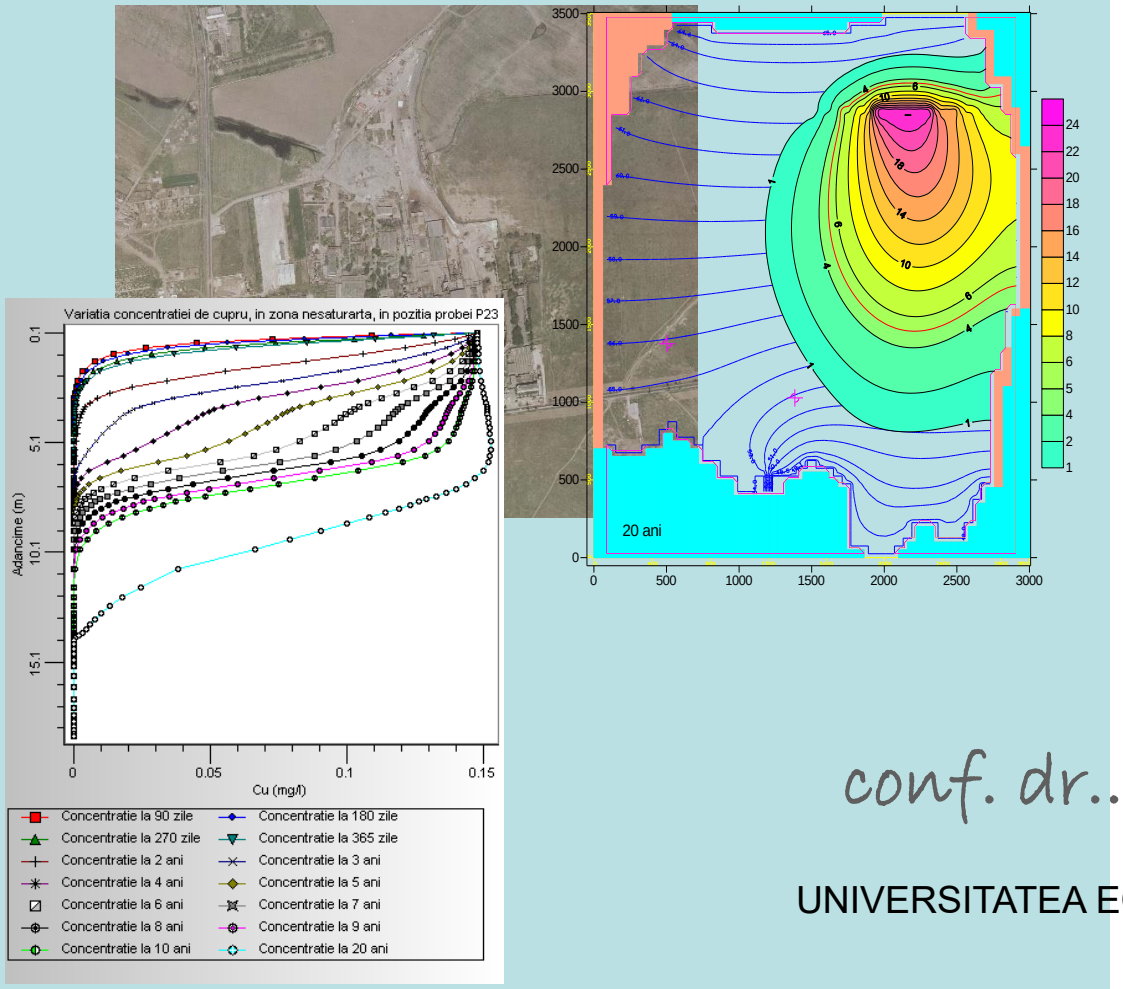
- Fazal, Shahab., GIS Basics, New Age International Publishers, 2008, New Delhi.
- Reddy, M. Anji., Remote Sensing and Geographical Information Systems, third edition, BS Publications, 2008.
- Shamsi, U.M., GIS Applications for Water, Wastewater and Stormwater Systems, CRC Press Book, 2005.

f. Evaluarea

Înainte de examen este indicat să parcurgeţi din nou toată materia, cu atenţie, durata estimată pentru această activitate fiind de aproximativ o săptămână.

Tip activitate	Criterii de evaluare	Metode de evaluare	Pondere din nota finală
Curs	Prezenţa la 50% din cursuri	Verificarea prezenţei	10%
Seminar	Acurateţea ideilor prezentate în cadrul referatelor; acurateţea şi detalierea observaţiilor, corelarea cu noţiunile teoretice, calitatea suportului ilustrativ	Punctaj întocmit pentru realizarea referatelor	30%
Evaluare finală	Gradul de cunoaştere a problematicii auditului și evaluării de mediu a organizațiilor, încadrarea în subiect, corectitudinea exprimării, capacitatea de a utiliza materialul bibliografic indicat	Examen scris	60%
Modalitatea de notare (calificativ sau notă): Notă			
Standard minim de performanţă: - cunoașterea instrumentelor gestiunea datelor de mediu în scopul evaluării impactului de mediu. - elaborarea unui proiect GIS de complexitate redusă.			

GIS & evaluare impact de mediu

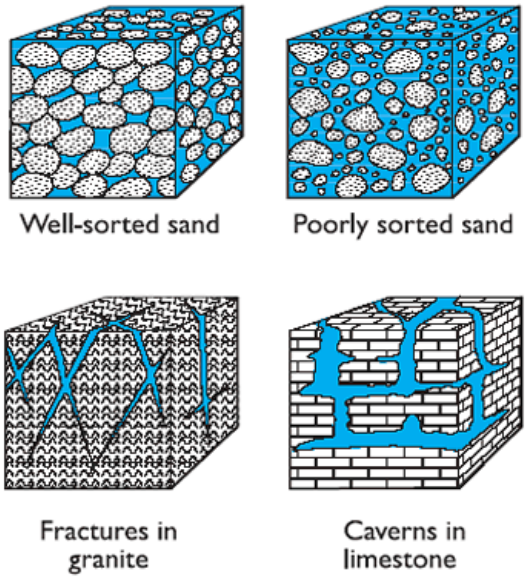


conf. dr.. giuliano tevi

UNIVERSITATEA ECOLOGICĂ DIN BUCUREŞTI

Mediul geologic este un sistem complex, situat la partea superioară a litosferei, limitat la zona în care omul îşi desfăşoară activitatea, format din roci ce pot conţine apă subterană, gaze şi resurse naturale neregenerabile.

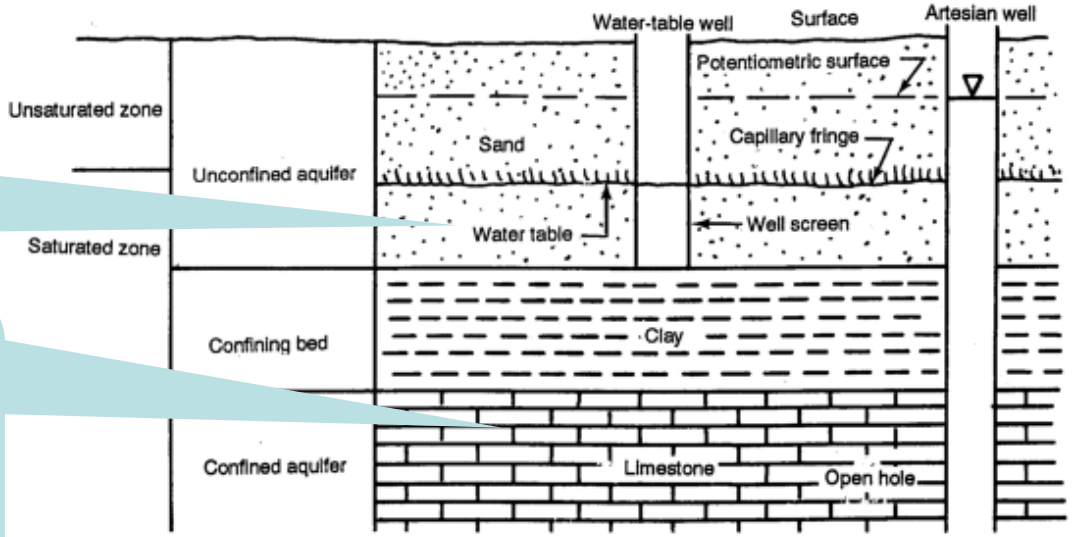
- teren permeabil (roci şi depozite sedimentare,
- roci cristaline (fisurate);
- ape subterane (+ fluide asociate, temperatură, viscozitate, greutate specifică, compresibilitate).



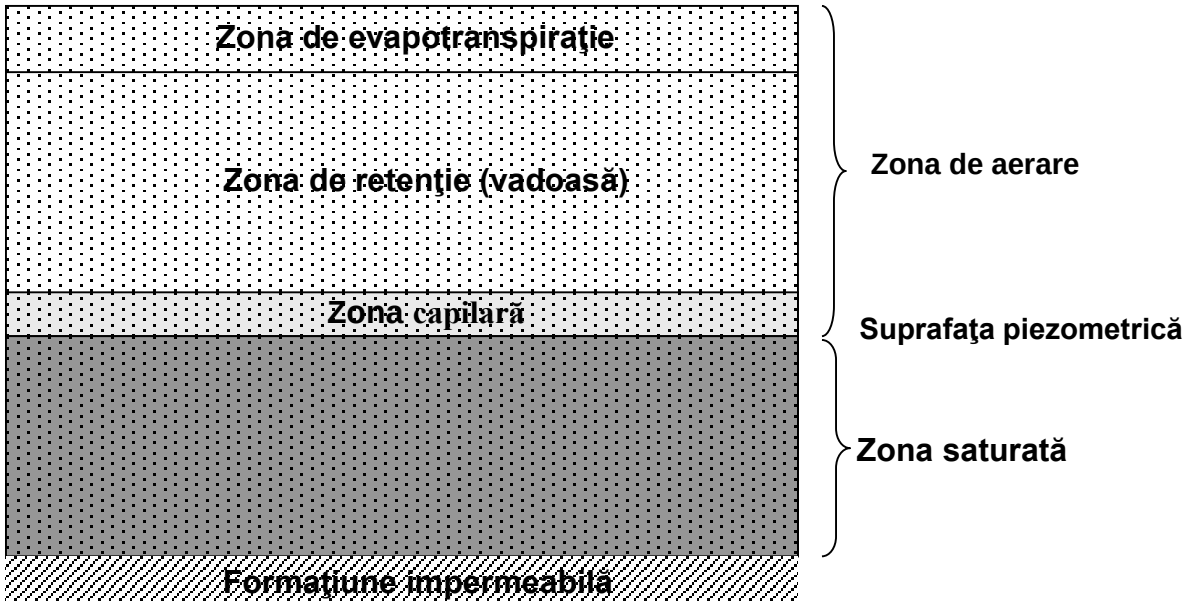
Acvifer – partea saturată cu apă a unei formaţiuni permeabile, suficient de conducătoare pentru a permite o curgere semnificativă a unui curent acvifer şi captarea profitabilă a apei subterane (conductivitate >0,1 m/zi, gradient hidraulic > 0,001 şi viteză minimă de 0,0001 m/zi).

acvifere cu nivel liber – dezvoltate în terenuri permeabile ce se dezvoltă uniform până la suprafaţa topografică

acvifere sub presiune – acumulat în terenuri permeabile care sunt delimitate în acoperiş de terenuri impermeabile şi nivelul piezometric se află la cote superioare acoperişului



Zonarea umidităţii pe veritcală



În cazul unui teren poros, permeabil, umiditatea în plan vertical este distribuită după următorul model tipic:

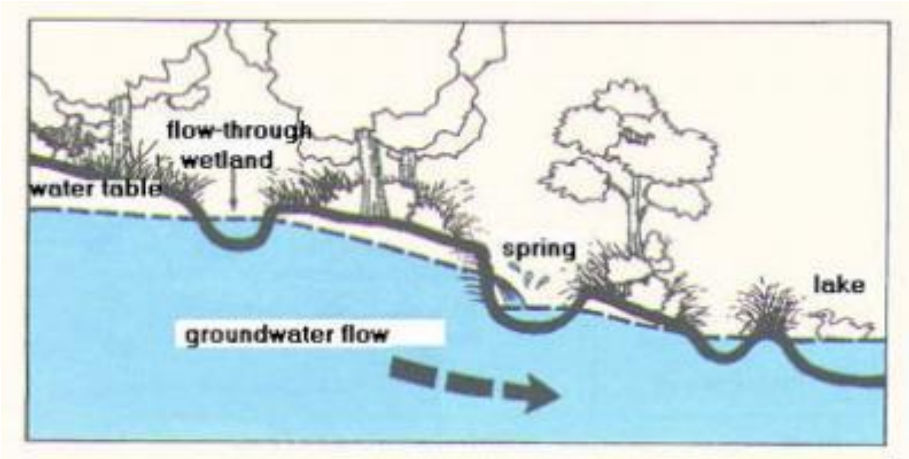
• ***Zona de aerare*** – aflată în partea superioară a profilului, între suprafaţa topografică şi suprafaţa acviferului. Ea este împărţită în mai multe zone:

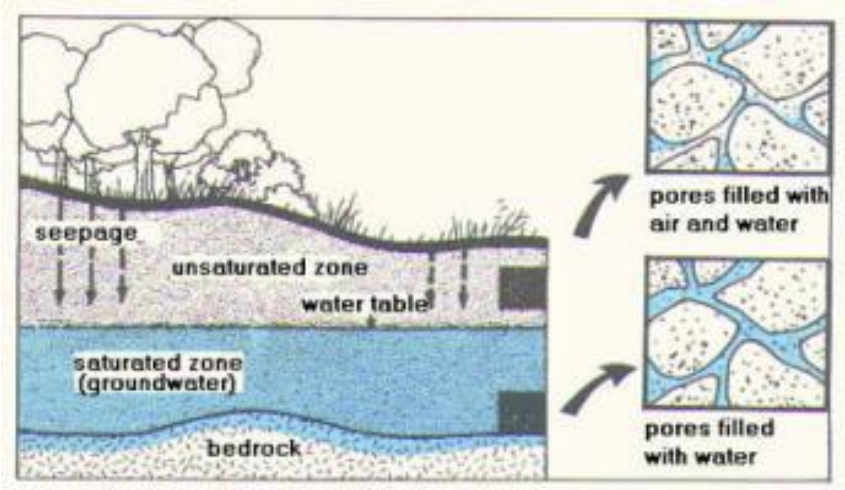
  • ***Zona de evapotranspiraţie*** – corespunzătoare practic profilului de sol. Umiditatea în această zonă este controlată de pierderea apei prin evapotranspiraţie, aici este prezentă o cantitate mare de apă sub formă de vapori.

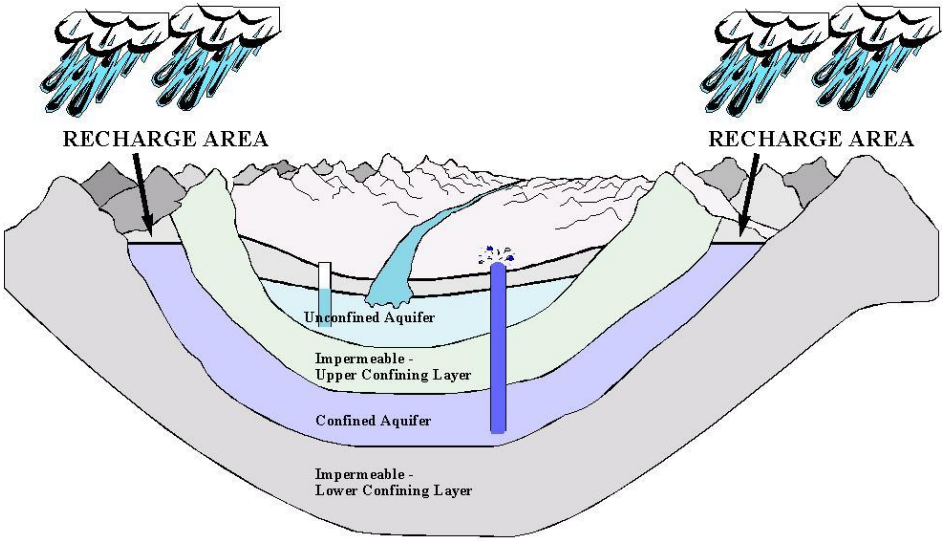
  • ***Zona de retenţie (zona vadoasă)*** – evapotranspiraţia este neglijabilă, apa este reţinută de către forţele higroscopice şi capilare, iar surplusul de apă din zona de evapotranspiraţie se deplasează descendent sub influenţa gravitaţiei. Grosimea acestei zone este variabilă funcţie de poziţia nivelului piezometric (de la 0 la sute de metrii în zonele aride).

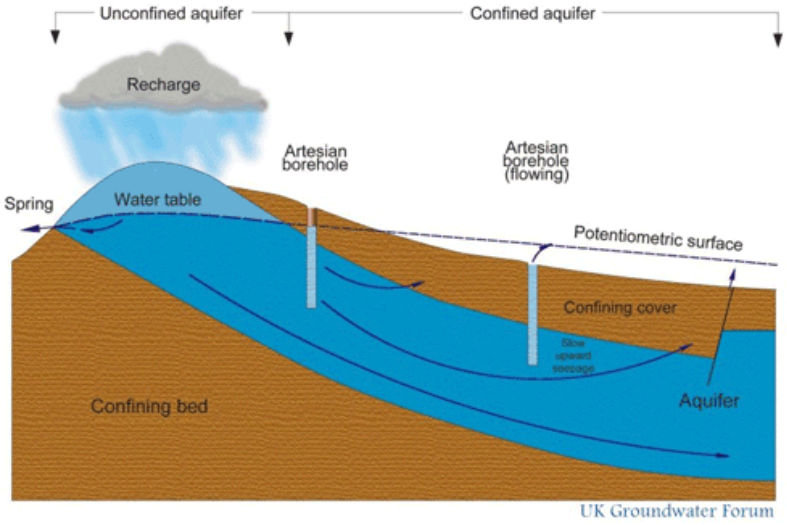
  • ***Zona capilară*** – este generată de suprafaţa piezometrică, grosimea acestei zone este invers proporţională cu granulometria (cca. 2,5 cm pentru pietrişuri până la 700 cm pentru silturi).

• ***Zona saturată*** – se dezvoltă sub suprafaţa piezometrică şi caracterizată de faptul că întregul volum al porilor este saturat de apă lichidă în diverse stări. La nivel general, limita inferioară a acestei zone este determinată de apariţia temperaturilor şi presiunilor critice (12 – 20 km. în domeniul continental).


flow-through
wetland
water table
spring
lake
groundwater flow


seepage
unsaturated zone
water table
saturated zone
(groundwater)
bedrock
pores filled with
air and water
pores filled
with water


RECHARGE AREA
RECHARGE AREA
Unconfined Aquifer
Impermeable -
Upper Confining Layer
Confined Aquifer
Impermeable -
Lower Confining Layer


Unconfined aquifer
Confined aquifer
Recharge
Artesian
borehole
Artesian
borehole
(flowing)
Spring
Water table
Potentiometric surface
Confining cover
Confining bed
Aquifer
UK Groundwater Forum

❖ **Caracteristici ale interacţiunii terenului cu apa subterană** → mobilitatea apelor in mediul poros

• *umiditatea terenurilor permeabile* (umiditate masică, volumică, grad de saturaţie);

• *tensiunea interfacială* la suprafaţa de separaţie fluid/aer sau fluid/rocă, tensiune responsabilă de fenomenele de capilaritate.

❖ **Caracteristicile interacţiunii apei subterane cu fluidele asociate** → *fluide miscibile, fluide imiscibile* ce influenţează procesul de migrare al apei subterane.

• variaţia în timp a concentraţiei *fluidelor miscibile* este influenţată de 5 procese:

- difuzia moleculară;
- dispersia;
- advecţia;
- sorbţia;
- reacţiile biochimice (influenţate în principal de *viteza de filtrare* a apelor subterane şi de *factorul de retardare* al contaminanţilor).
    - Izoterme de sorbţie (Freundlich, Langmuir)

• curgerea polifazică a *fluidelor imiscibile* (mai uşoare sau mai grele dacât apa) este controlată de permeabilitatea relativă (raport între permeabilitatea unui fluid aflat la anumit grad de saturaţie şi permeabilitatea intrinsecă a formaţiunii.)

Poluarea

O stare indezirabilă a mediului natural datorată contaminării cu substanţe dăunătoare ca o consecinţă a activităţii umane (Dicţionarul Webster)

Introducerea directă sau indirectă, ca rezultat al activităţii umane, de substanţe, vibraţii, încălzire sau zgomot în aer, apă sau sol, care pot fi periculoase pentru sănătatea umană, calitatea mediului sau pot crea pierderi materiale (Directiva Cadru a Resurselor de Apă).

Clasificarea surselor de poluare ale mediului geologic:

• *modul de dezvoltare în spaţiu* (concentrate şi difuze);
• *modul de evoluţie în timp* (instantanee şi continue);
• *gradul de mobilitate* (fixe şi mobile);
• *natura fenomenului de poluare* (poluarea de natură *fizică*, *chimică*, *biologică* şi *radioactivă*);
• *originea sursei* (industrială, agricolă şi/sau menajeră).

Clasificarea poluanţilor mediului geologic:

• *stare de agregare* (lichizi, solizi, gazoşi);
• *natura lor* (anorganici, organici, biologici, fizici);
• *solubilitatea în apă şi greutate specifică* (miscibili, imiscibili mai grei sau mai uşori decât apa)

Schematizarea hidrostructurilor

Simplificarea complexităţii hidrostructurilor în condiţiile menţinerii caracteristicilor reprezentative pentru:

• spaţiul în care are loc scurgerea subterană;

• caracteristicile hidrofizice ale terenurilor a căror distribuţie spaţială condiţionează caracteristicile curgerii apei subterane;

• condiţii hidrodinamice ale curgerii la limitele hidrostructurii şi în interiorul acesteia.

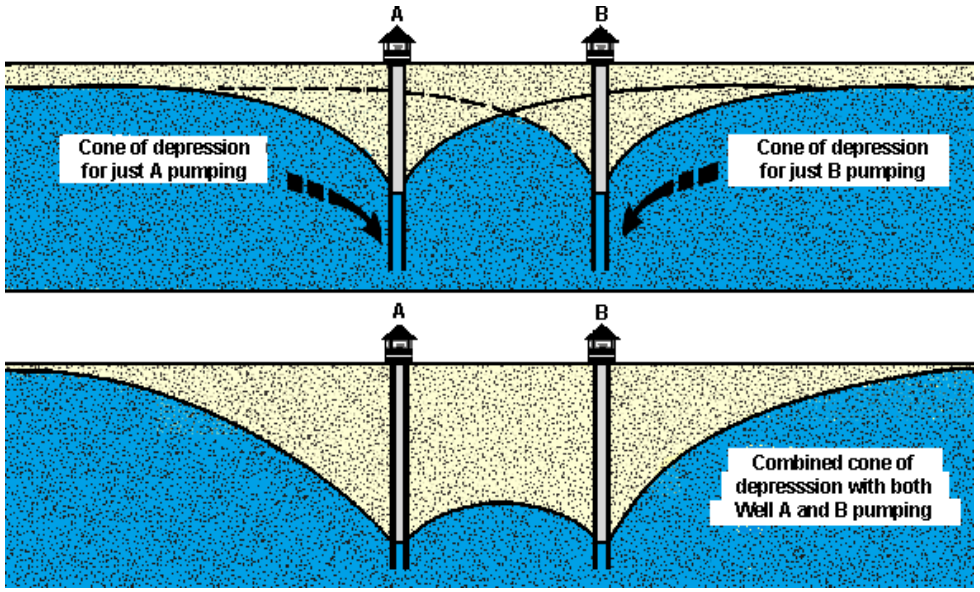
Etapele schematizării:

• *Schematizarea spaţială* – precizează geometria spaţiului în care are loc curgerea apelor subterane;

• *Schematizarea parametrică* – evaluarea distribuţiei spaţiale a parametrilor care descriu caracteristicile hidrofizice ale terenului;

• *Schematizarea hidrodinamică* – precizarea frontierelor hidrodinamice ale sistemului acvifer şi a condiţiilor hidrodinamice iniţiale în interiorul acestuia.

Reprezentativitatea modelului matematic utilizat pentru evaluarea curgerii apelor subterane este condiţionat de reprezentativitatea modelului conceptual al hidrostructurii şi asigură similaritatea reacţiei modelului cu cea a hidrostructurii reale, la o perturbare hidrodinamică sau fizico-chimică.



MODELUL CONCEPTUAL

Schematizarea spaţială

❖ Urmăreşte simplificarea formei suprafeţelor care definesc spaţial hidrostructura.

Morfologia este schematizată sub formă grafică prin secţiuni geologice, bloc diagrame, harţi structurale. Datele provin în general din cartarea de suprafaţă, foraje geologice şi hidrogeologice şi prospecţiuni geofizice.

Alături de elementele morfologice caracteristice ale hidrostructurii o atenţie deosebită se acordă precizării continuităţii hidrodinamice a depozitelor permeabile (determinări de niveluri piezometrice)

- Extinderea cercetării:
  - *extinderea regională* (evaluarea resurselor naturale ale hidrostructurii) – modelul conceptual se extinde până la limitele fizice ale structurii (zone de aflorare, falii impermeabile, zone de drenaj…);
  - *extinderea locală* (studii hidrogeologice de interes local) – modelul se extinde în general până la limita zonei de influenţă (estimată prin metode analitice simplificate sau prin măsurători de nivel)

MODELUL CONCEPTUAL

Schematizarea parametrică

❖ metodologia este aplicată tuturor parametrilor hidrostructurii, în special:

- porozităţii active;
- conductivităţii hidraulice;
- coeficientului de difuzivitate hidraulică şi înmagazinare;
- coeficientului de realimentare;
- coeficientului dispersiei hidrodinamice, etc.

Rezultatul acestei schematizări este imaginea distribuţiei spaţiale a parametrilor studiaţi (estimare prin kriging, simulare condiţionată pe baza legii de distribuţie a parametrului)

Variabilitatea parametrului este tradusă prin:

- variabilitatea parametrului de la un punct la altul (omogenitate sau neomogenitate);
- variaţia valorii pe diverse direcţii (izotropie sau anizotropie)

Mediul poate fi omogen şi izotrop sau neomogen şi anizotrop

- Extinderea cercetării:
  - *extindere regională* – presupune în general estomparea variabilităţii parametrilor , se tinde spre adoptarea unei distribuţii omogene şi izotrope;
  - *extinderea locală* – presupune o rigurozitate specifică legată de descrierea distribuţiei parametrilor.

MODELUL CONCEPTUAL

Schematizarea hidrodinamică

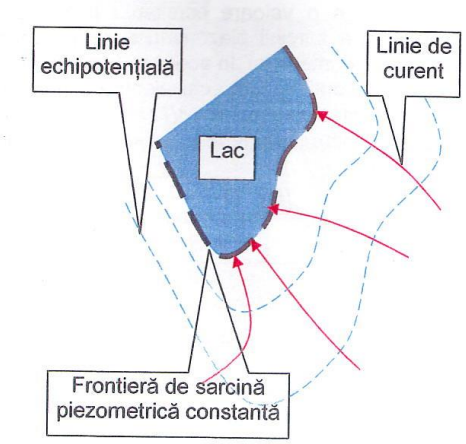
❖ se materializează prin precizarea:

• condiţiilor hidrodinamice pe frontierele modelului;

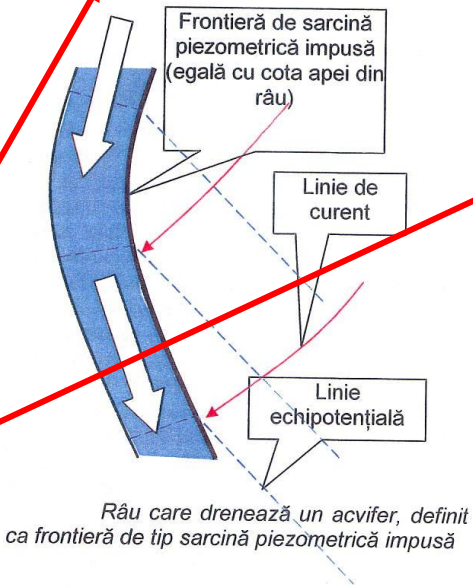
• condiţiilor hidrodinamice iniţiale în interiorul frontierelor.

Condiţiile hidrodinamice pe frontierele modelului

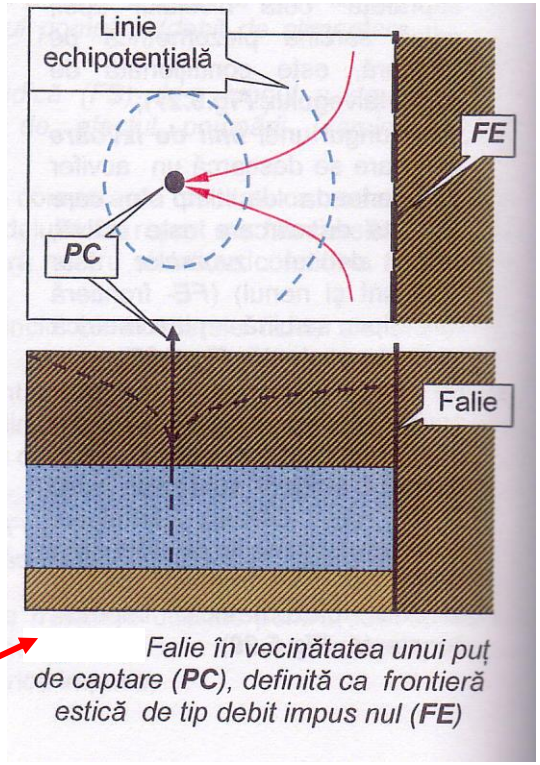
- frontiere de tip sarcină piezometrică impusă (Dirichlet);
- frontieră de tip debit impus (Newmann);



Malul unui lac definit ca frontieră/margine de tip sarcină piezometrică constantă pentru acviferul drenat



Râu care drenează un acvifer, definit ca frontieră de tip sarcină piezometrică impusă

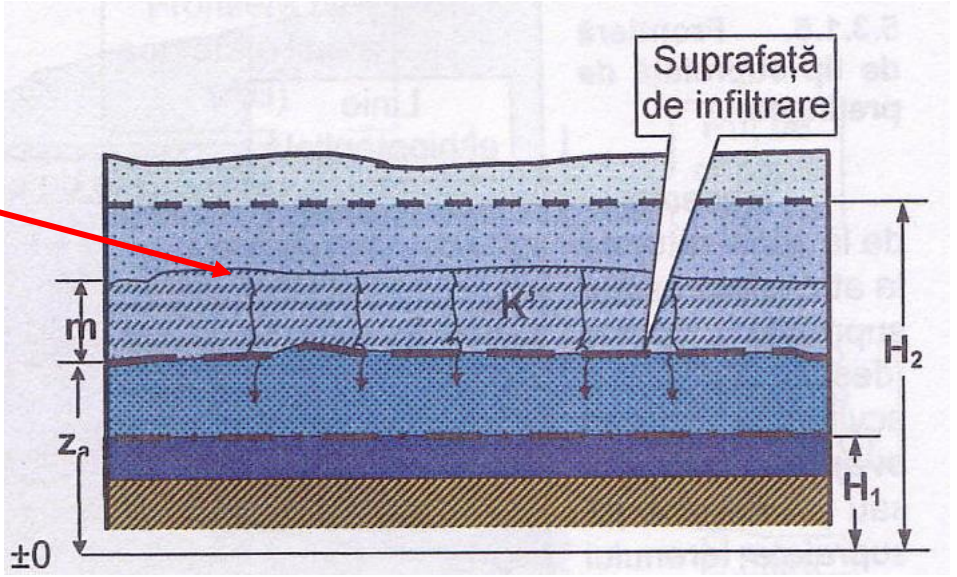


*Falie în vecinătatea unui puț de captare (**PC**), definită ca frontieră estică de tip debit impus nul (**FE**)*

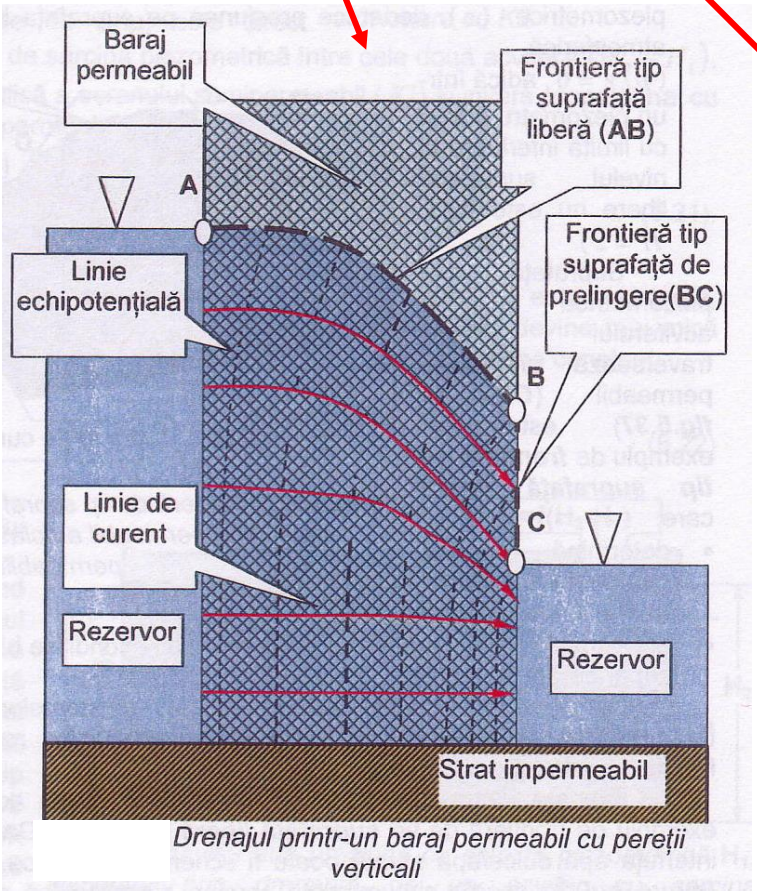
după Scrădeanu D., Gheorghe Al.

MODELUL CONCEPTUAL

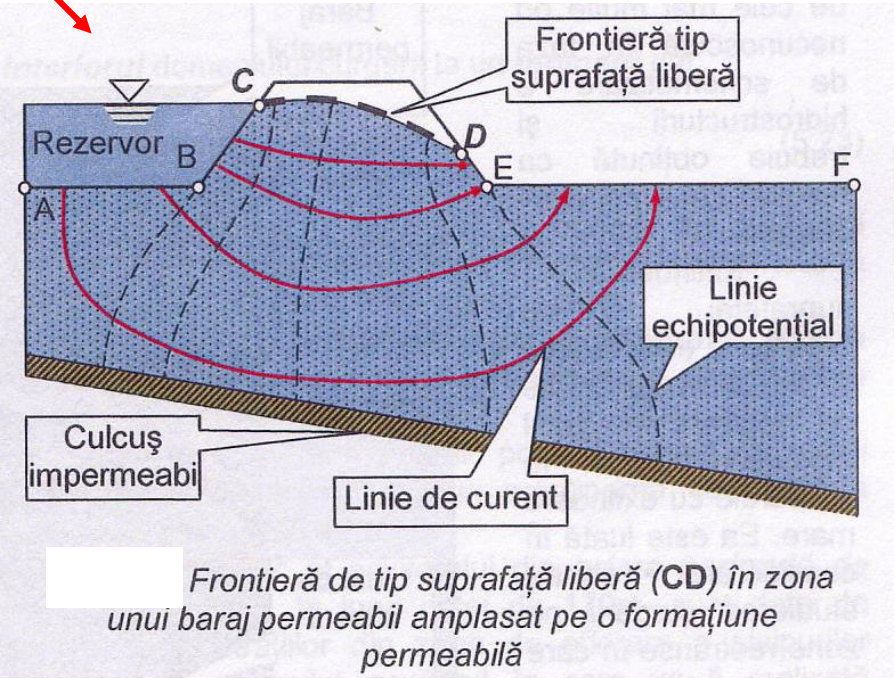
- frontieră de tip dependent de sarcina piezometrică;
- frontieră de tip suprafaţă liberă;
- frontieră de tip suprafaţă de prelingere.



după Scrădeanu D., Gheorghe Al.



Drenajul printr-un baraj permeabil cu pereții verticali



Frontieră de tip suprafață liberă (CD) în zona unui baraj permeabil amplasat pe o formațiune permeabilă

Alegerea variantei de modelare

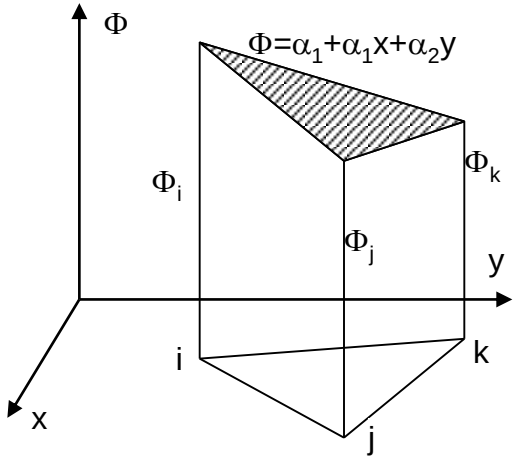
Ecuaţiile ce descriu procesele hidrodinamice şi de transport pot fi rezolvate utilizând modele *analitice* sau *numerice.*

Modelele analitice operează cu soluţii funcţionale fixe, au avantajul de a fi exacte, dar sunt aplicabile pentru soluţionarea unui număr limitat de probleme. Soluţionarea pe cale analitică a ecuaţiilor impune simplificarea formei şi extinderii structurii.

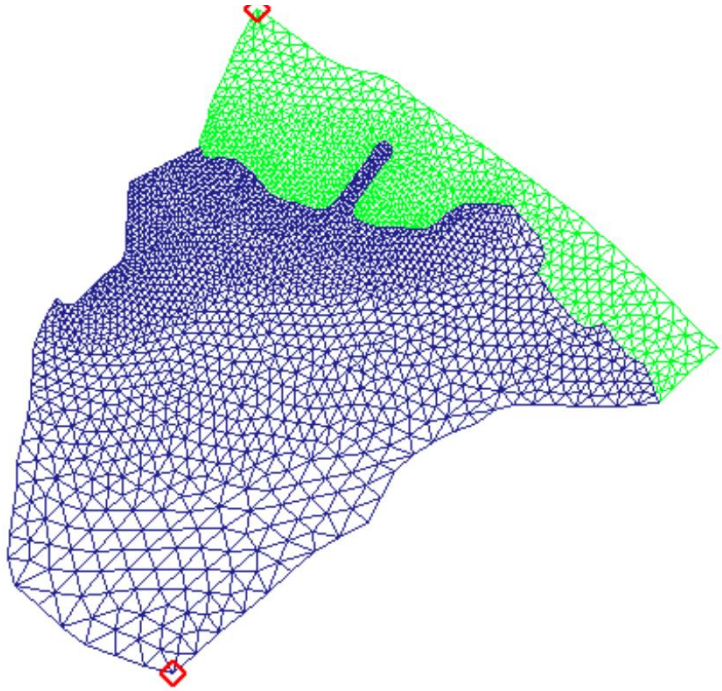
Modelele numerice sunt capabile să rezolve ecuaţii complexe ce caracterizează spectrul hidrodinamic al curgerii şi transportul poluanţilor. Modelele numerice folosesc aproximări pentru a rezolva ecuaţiile diferenţiale ce descriu curgerea apei subterane şi transportul în soluţie.

• Aproximarea prin metoda **elementului finit** – structura spaţială studiată este împărţită în elemente finite (de ex.: triunghiuri). Elementul finit este caracterizat prin:

- dezvoltarea dimensională (1D, 2D, 3D);
- număr de noduri;
- funcţia de interpolare asociată



Element finit triunghiular cu trei noduri



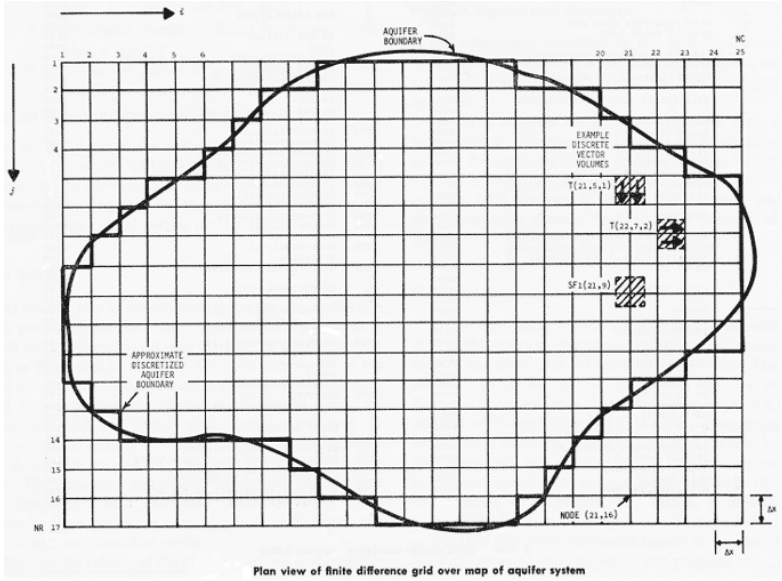
Aproximarea continuumului material cu ajutorul elementelor finite de formă triunghiulară

Alegerea variantei de modelare

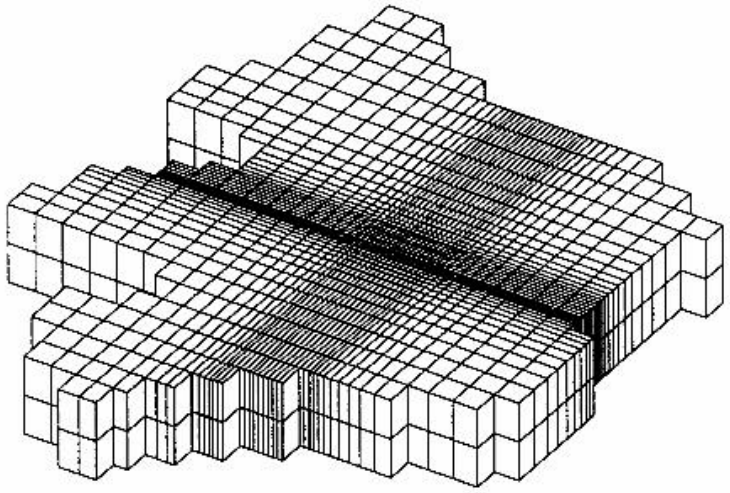
• Aproximarea prin metoda **diferenţelor finite** – presupune discretizarea structurii cu o reţea rectangulară (2D sau 3D), fiecare nod al reţelei fiind caracterizat de o valoare a parametrului.

În cazul metodei ***diferenţelor finite***, domeniul de integrare coincide cu reţeaua de discretizare, în fiecare nod al reţelei avem o valoare a parametrilor, aceasta fiind necunoscută în spaţiul dintre noduri. De al un element la altul al reţelei, variaţia este discontinuă.

În cazul metodei ***elementului finit*** domeniul de integrare este diferit de reţeaua de discretizare. Valoarea parametrilor este variabilă în interiorul domeniului de integrare iar variaţia este continuă. Spre deosebire de metoda diferenţelor finite, în acest caz valoarea unui parametru poate fi calculată în orice punct al domeniului de integrare al ecuaţiei.



Discretizarea 2D şi 3D a structurii cu ajutorul reţelei rectangulare



Alegerea variantei de modelare

Repere în alegerea tipului de model utilizat:

❖ *Modelele analitice* în cazul în care:

- curgerea apei subterane sau procesele de transport sunt relativ simple;
- este necesară evaluarea iniţială a condiţiilor hidrogeologice sau evaluarea unor condiţii de remediere.

❖ *Modelele numerice* în cazul în care:

- curgerea apei subterane sau procesele de transport sunt relativ complexe;
- variaţia debitului, condiţiilor hidrogeologice sau geochimice precum a altor caracteristici este foarte accentuată.

✓ *Modelele* hidrodinamice sau de transport *unidimensionale* în cazul în care:

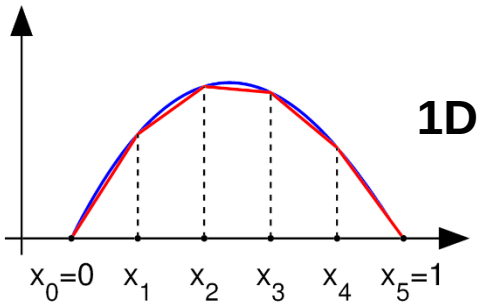
- se realizează evaluări iniţiale în care nu este cunoscut gradul de neomogenitate sau de anizotropie;
- situaţiile în care un potenţial receptor este în imediata vecinătate a unei surse de poluare.

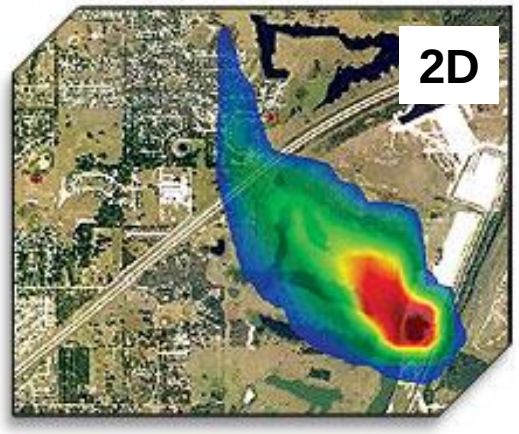
✓ *Modelele bidimensionale* în cazul în care:

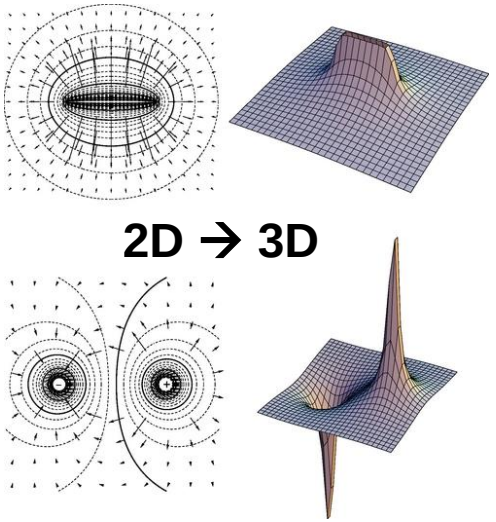
- zona de studiu cuprinde una sau mai multe zone de drenare sau de aport de apă (puţuri de pompare sau de injecţie, canalizări, râuri, etc.);
- zona este caracterizată prin direcţii de curgere ale apei subterane care au o evidentă distribuţie bidimensională;
- acviferul are variaţii distincte ale proprietăţilor hidraulice;
- se modelează migrarea poluanţilor, efectul dispersiei transversale este semnificativ şi este necesară evaluarea distribuţiei spaţiale a frontului de poluare.

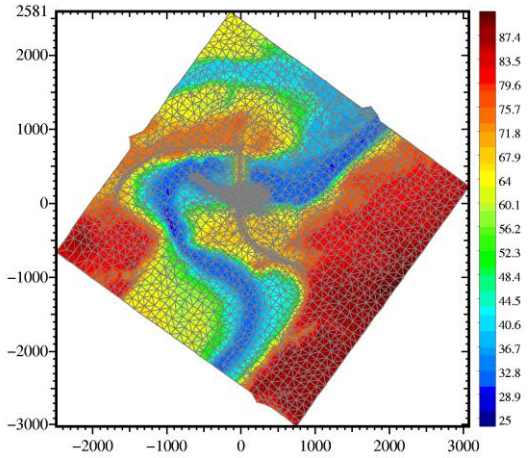
✓ *Modelele tridimensionale* în cazul în care:

- condiţiile hidrogeologice sunt bine cunoscute;
- sunt prezente acvifere multiple;
- mişcarea apei şi contaminanţilor în plan vertical este semnificativă.

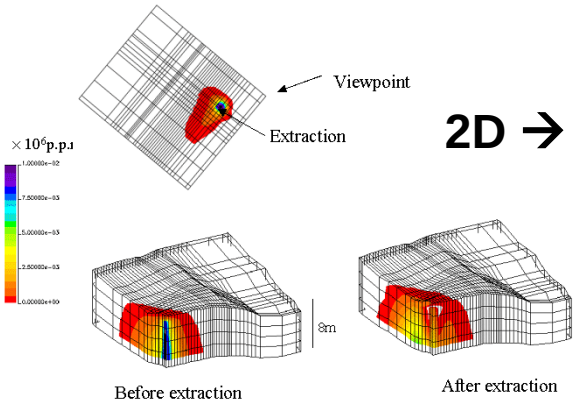


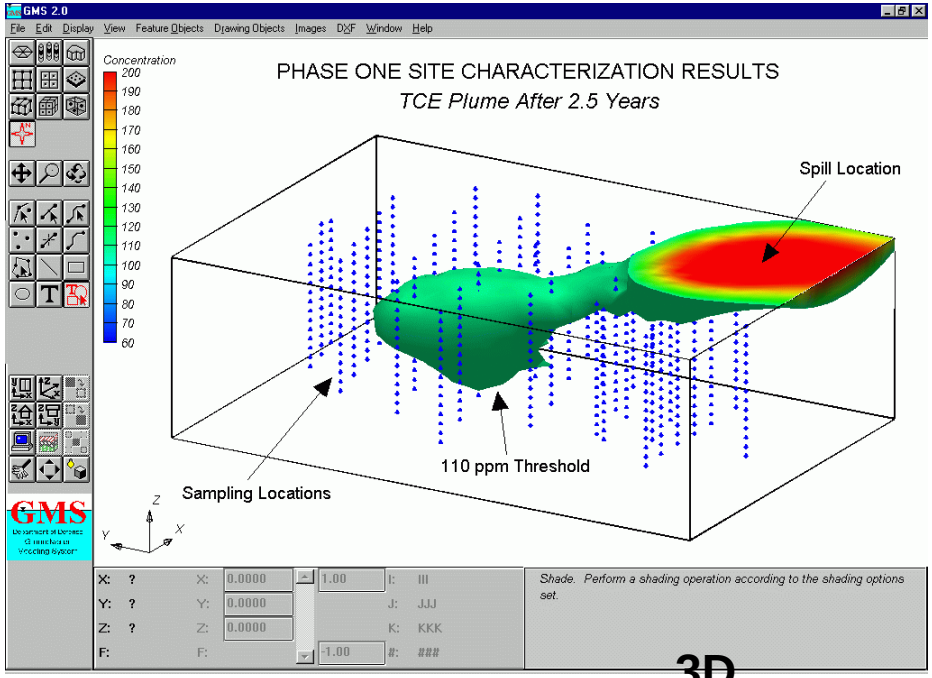






Elimination of the volatile contaminant by the vacuuming method





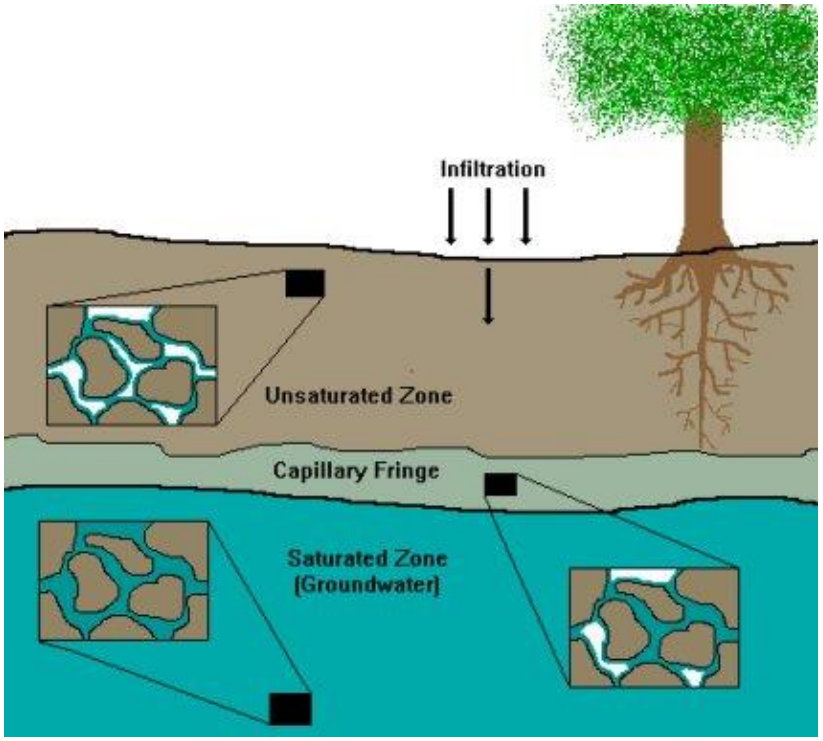
Zona nesaturată – zonă tampon între acvifer şi suprafaţa terenului

Condiţii caracteristice

• condiţionare climatică, care impune adoptarea anumitor condiţii specifice la limita superioară a modelului;

• variaţie a conţinutului de umiditate al solului şi al gradului de saturaţie, cu efect direct asupra conductivităţii hidraulice;

• procese foarte active se sorbţie, schimb ionic, biodegradare, etc.



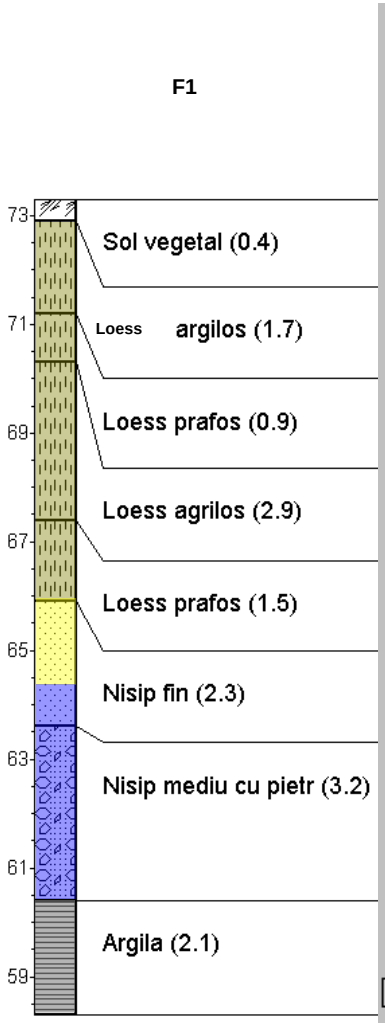
Capacitatea de autoepurare a zonei nesaturate este de 4 – 5 ori mai mare decât cea a acviferului. *Pentru o granulometrie de prafuri, nisipuri fine şi mijlocii, o particulă de apă poluată biologic are nevoie de un parcurs de 80 de metri în acvifer si de 16 metri in zona nesaturată.* *(Gheorghe, Al., Scrădeanu, D.)*

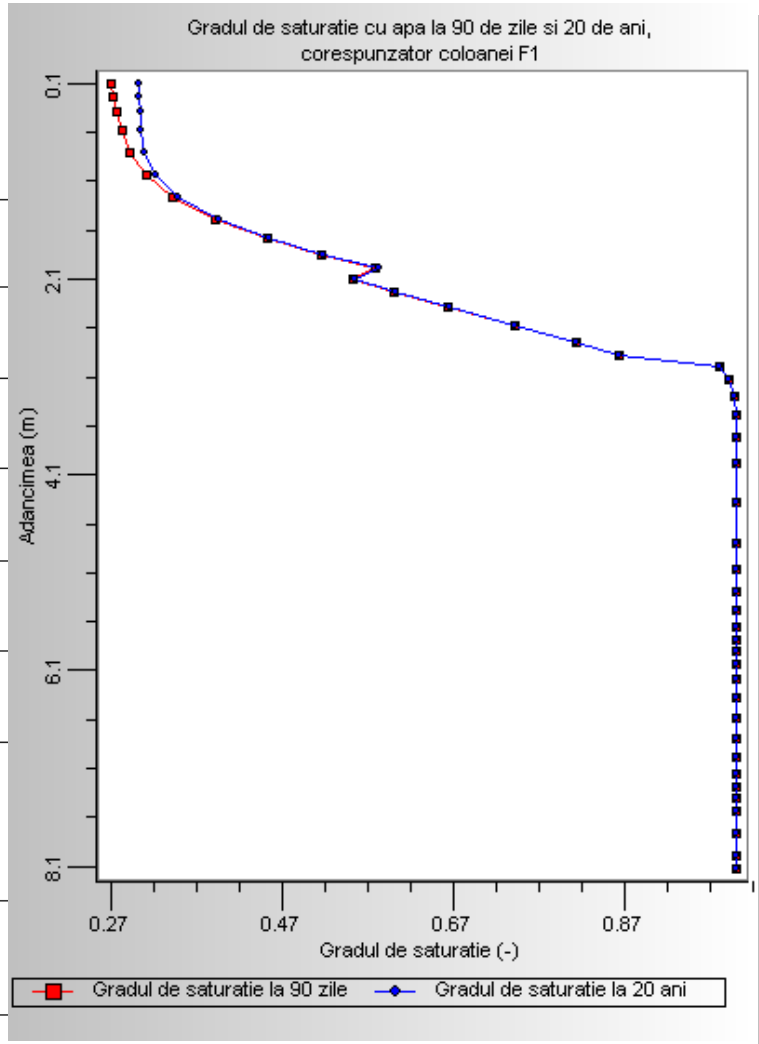
Importanţa cunoaşterii proceselor de transfer a poluanţilor şi de autoepurare a zonei nesaturare, este justificată de următoarele aspecte:

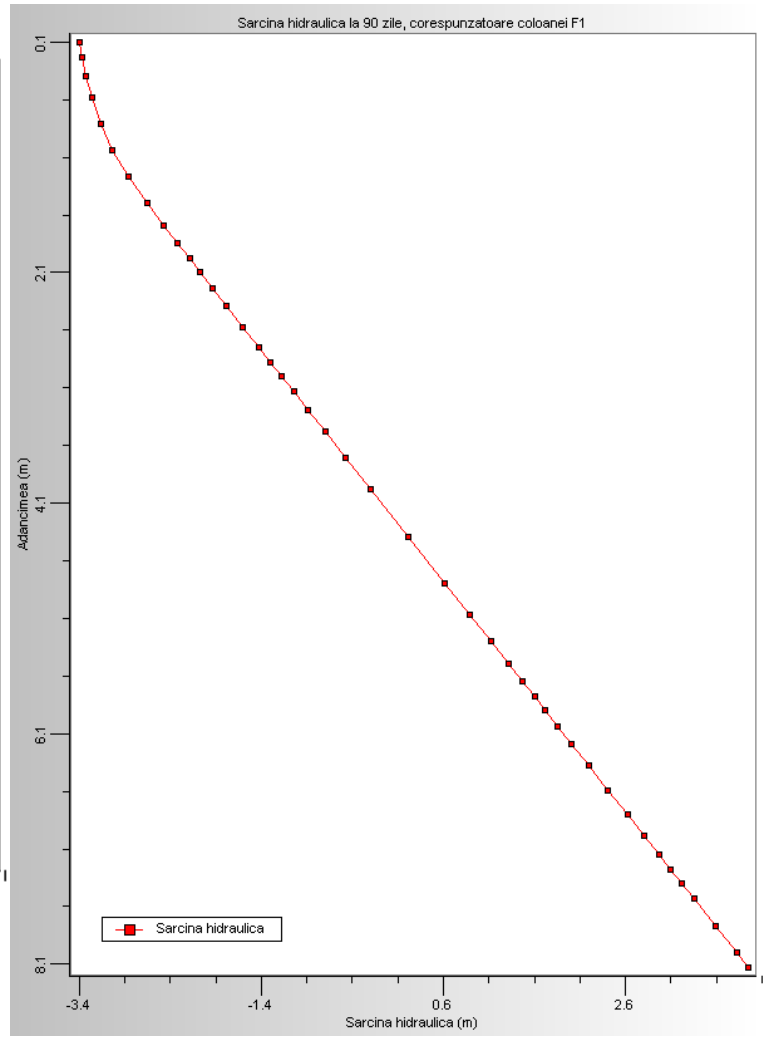
• detectarea poluanţilor în zona nesaturată se face în timp real (zile, săptămâni) în timp de monitorizarea calităţii acviferelor, prin foraje de monitorizare se face la intervale mai mari de timp (săptămâni, luni);

• vulnerabilitatea la poluare a acviferelor poate fi evaluată pe baza caracteristicilor geologice şi hidrogeologice ale zonei nesaturate.

Model 1D, distanţe finite, curgere şi transport metale grele - *UnSat Suite 2201*



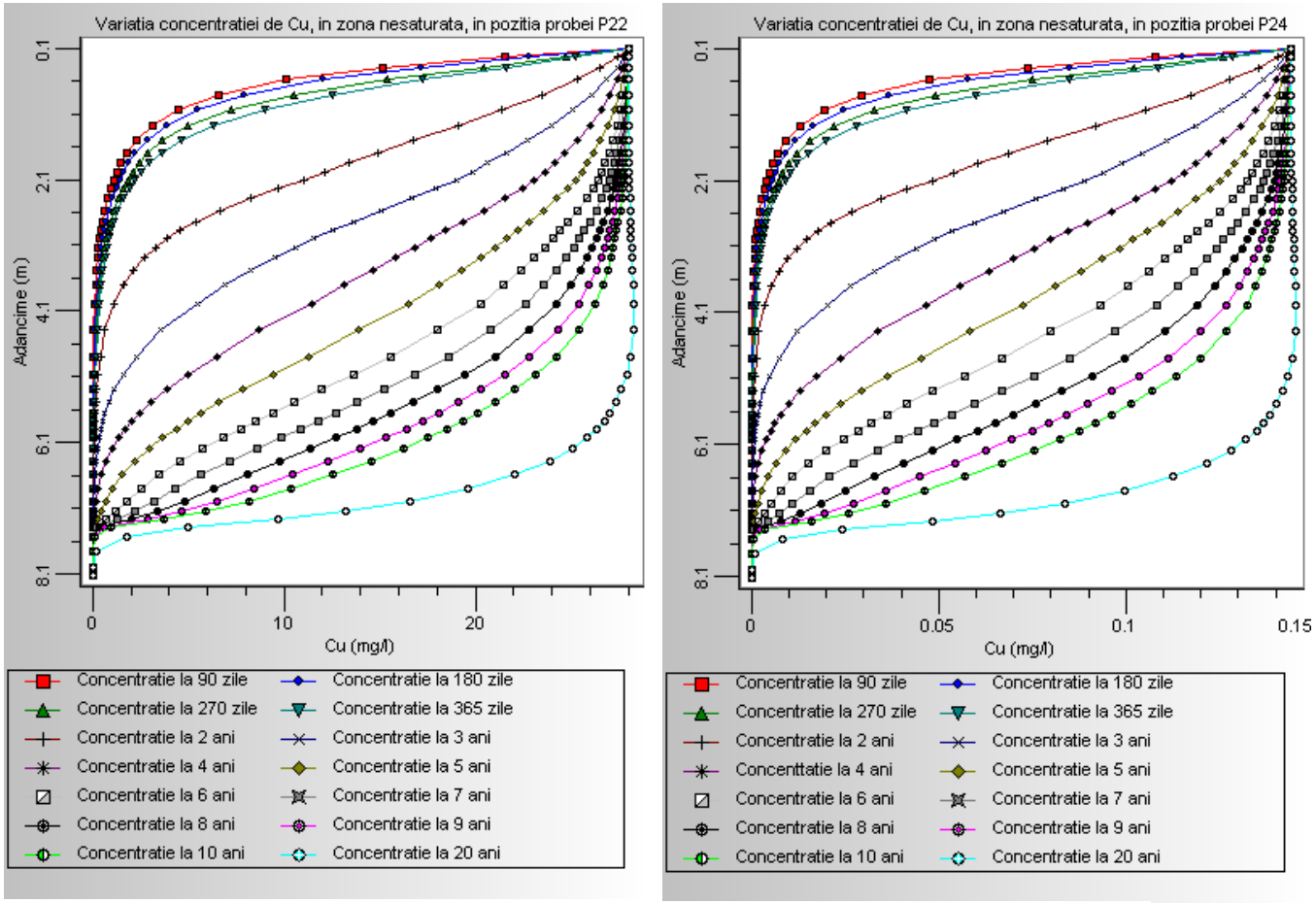




Coloana litologică corespunzătoare structurii zonei vadoase pentru care s-a realizat primul scenariu de simulare;

a. Variaţia gradului de saturaţie cu apă la începutul si la sfârşitul simulării

b. Variaţia sarcinii hidraulice la începutul simulării



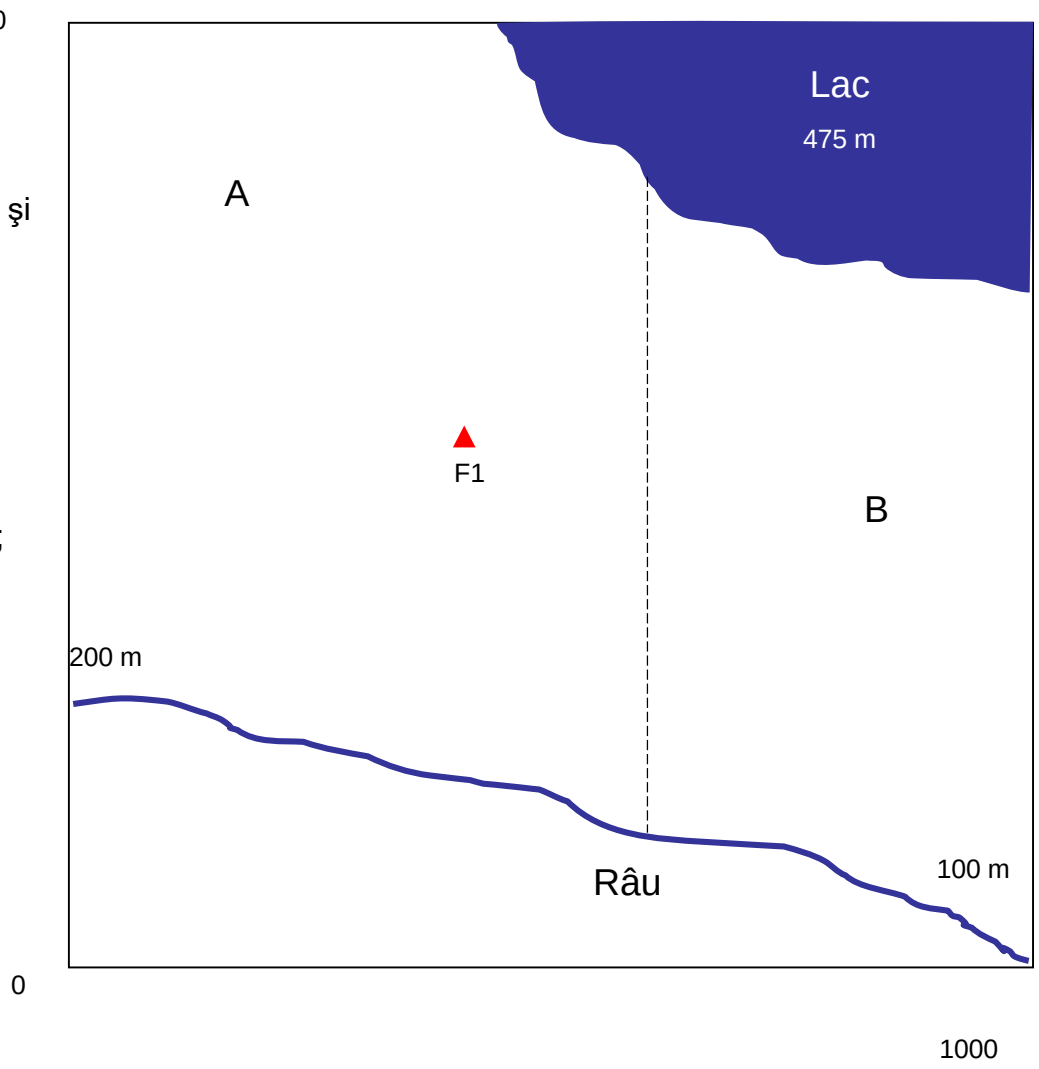
Variaţia concentraţiei de cupru, de-a lungul profilului, pentru concentraţii iniţiale (infiltrate) de 28 mg/l şi 0,144 mg/l

Modelele de curgere şi de transport pot fi folosite pentru a prezice evoluţia diverselor scenarii de poluare, adaptate potenţialei variaţii a condiţiilor climatice sau a concentraţiei sursei de poluare. De asemenea modelele pot fi folosite pentru a evalua diferite alternative de remediere sau pentru a asista evaluarea riscului.

Datorită faptului că până şi cel mai bine realizat şi calat model poate fi bazat pe date insuficiente sau pe simplificări numeroase, apar o serie de erori şi incertitudini în analiza curgerii sau transportului. Aceste posibile erori vor fi luate în considerare în evaluarea proceselor naturale sau episoadelor de contaminare studiate

Aplicaţie

- Acvifer freatic, în legătură hidraulică cu lacul în zona NE şi cu râul în zona S;
- cota culcuş – 70 m;
- alimentare acvifer 0,0009 m/zi
- formaţiuni litologice diferite (A şi B), caracterizate astfel:
  - conductivitate hidraulica teren A - K_A = 8,33 m/zi;
  - conductivitate hidraulică teren B – K_B = 5,63 m/zi;
  - porozitate activă teren A – na_A = 0.11
  - porozitate activă teren B – na_B = 0.15;
- F1 foraj de pompare cu un debit pompat Q = 100.000 m^3/zi;
- se va realiza modelul de curgere în regim natural şi în regim influenţat (metoda diferenţe finite, 2D, software FLOWPATH)

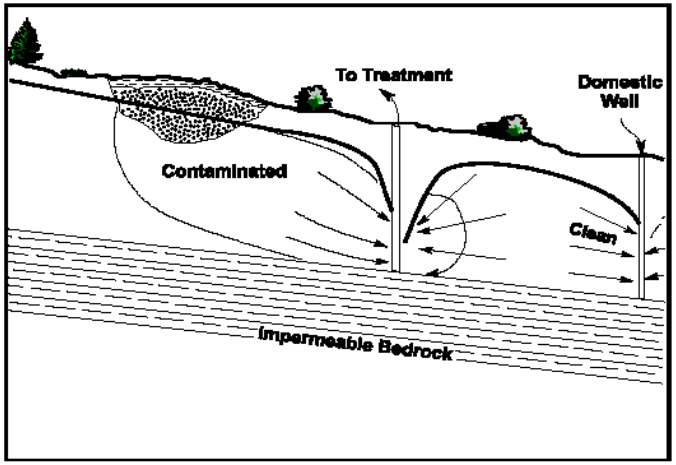


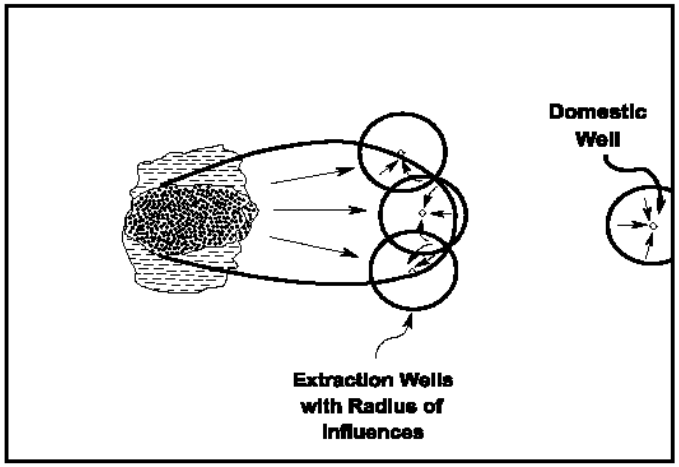
Combaterea poluării şi remedierea calităţii mediului geologic

Măsuri specifice în vederea combaterii şi reducerii poluării apelor subterane:

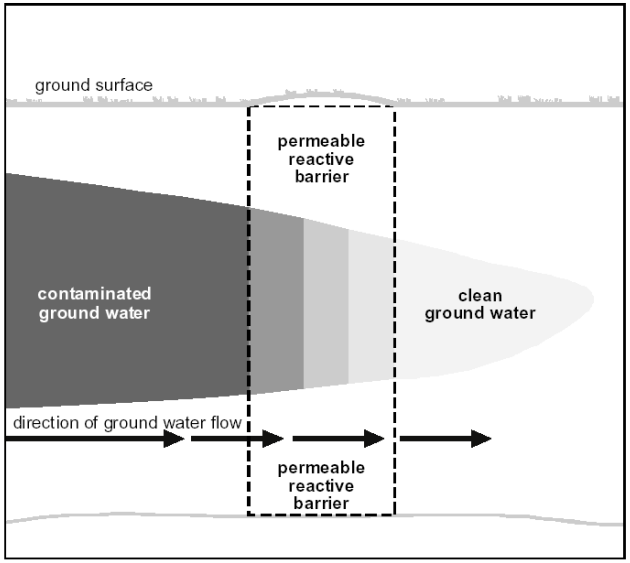
- Ecranarea prin pereţi etanşi:
- Bariere permeabile reactive;
- Izolarea hidraulică:
- Sisteme de drenaj
  - *Puţuri;*
  - *Drenuri.*
- Metode chimice şi biologice

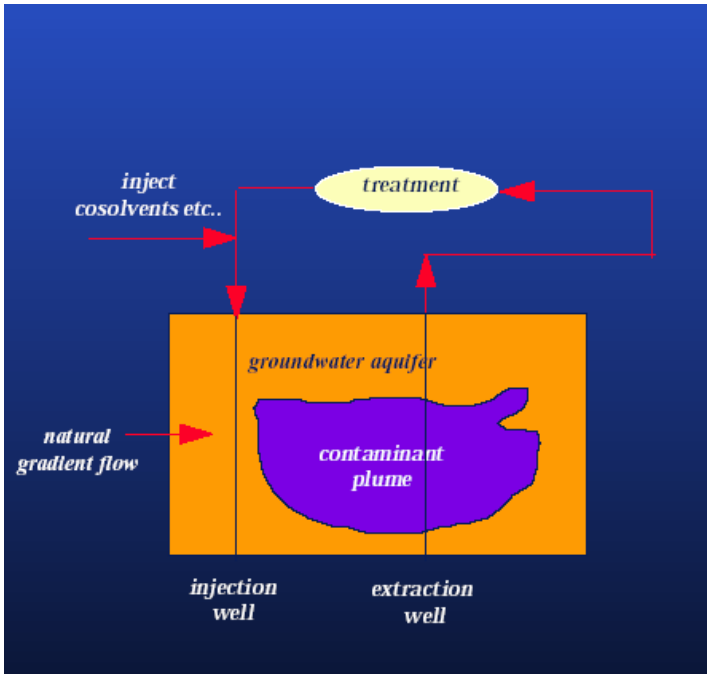
de tratare “in situ”.





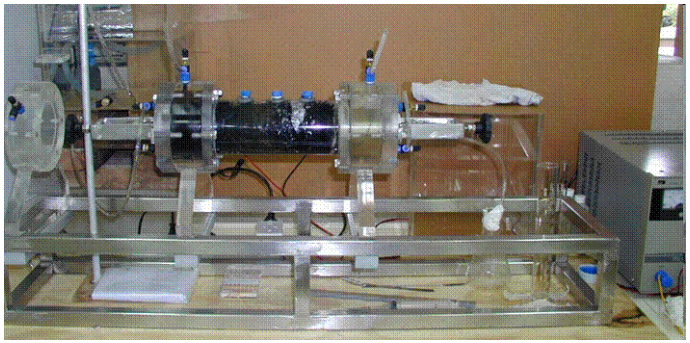
Specific example of hydraulic containment. Top - cross-sectional view. Bottom - plan view. (from Noyes Data Corp., 1985).

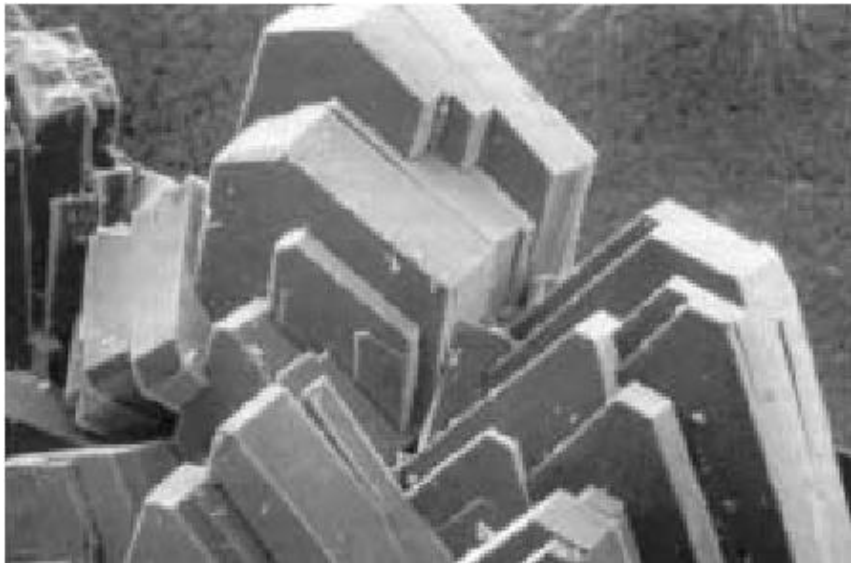


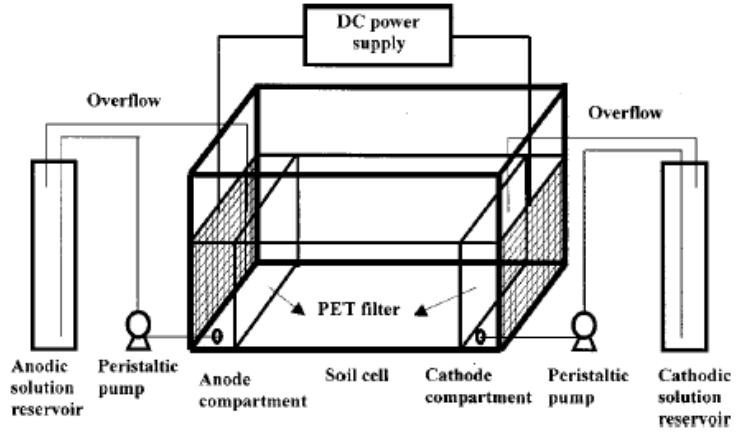


Măsuri specifice pentru combaterea şi reducerea poluării solului

- Bioremedierea (frecvent **fitoremedierea**);
- Remedierea electrocinetică;
- Remedierea prin spălare cu soluţii acide;
- Inactivarea metalelor prin utilizarea zeoliţilor naturali.







Atenuarea naturală monitorizată (ANM)

Presupune o serie de procese ce se desfăşoară "in situ" şi au drept rezultat reducerea masei, toxicităţii, mobilităţii, volumului sau concentraţiei poluantului. Cel mai adesea aceste procese sunt reprezentate de: biodegradare, dispersie, diluţie, sorbţie, volatilizare, stabilizare chimică sau biologică, transformarea sau distrugerea poluantului.

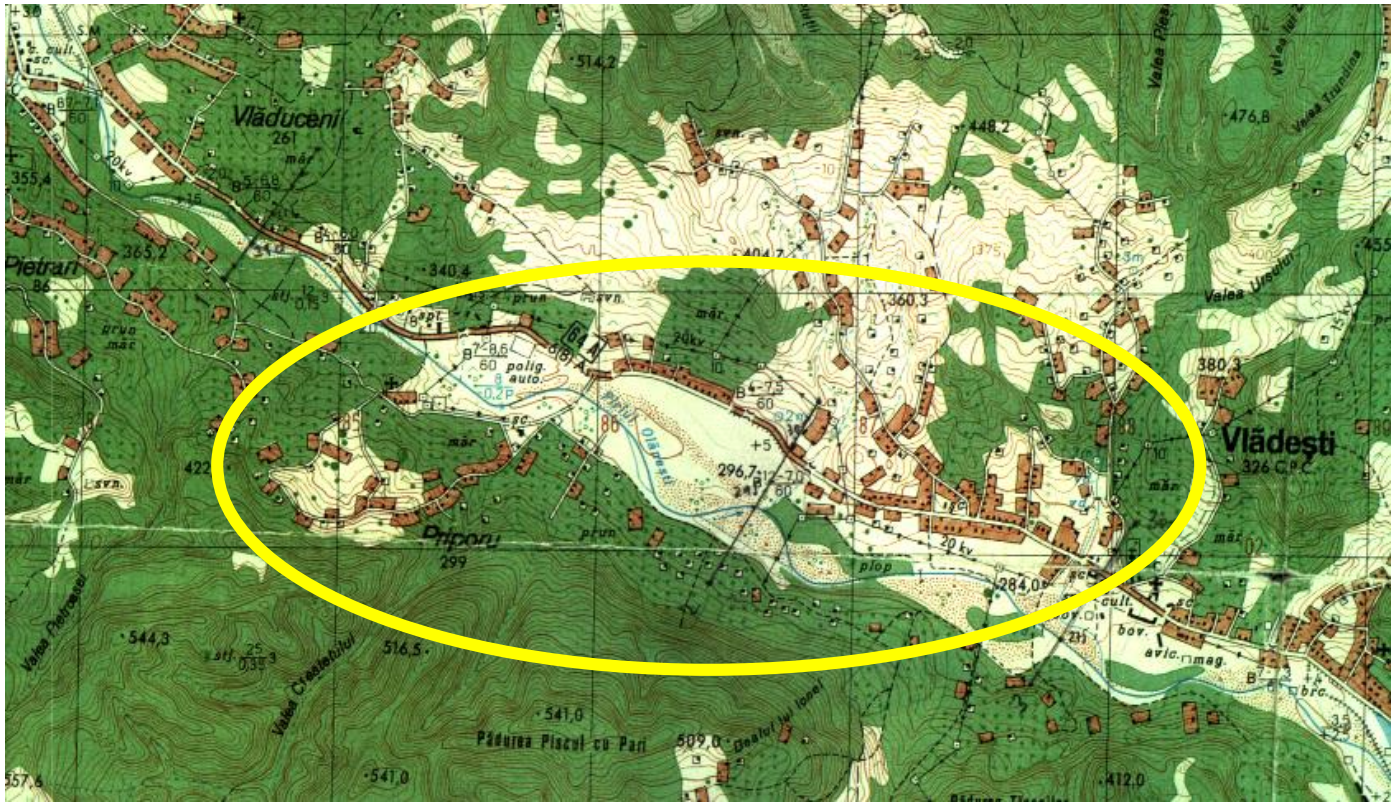
✓ Iniţial (1995) ANM a fost folosită ca soluţie de remediere pentru poluarea cu benzen, toluen, etilbenzen, xilen.
✓ Mai recent (2000) ANM a fost propusă pentru solvenţi cloruraţi, *metale grele*, radionuclizi şi alţi contaminanţi pentru care experienţa şi gradul de cunoaştere este mult mai redus.
✓ Succesul ANM depinde atât de natura poluantului cât şi de condiţiile zonei studiate. Se recomandă ca nivelul de investigare în vederea adoptării ANM să fie adaptat complexităţii pedologice, hidrogeologice şi naturii contaminantului. Mediile mai complexe necesită un grad mai ridicat de cunoaştere, necesar şi în situaţia în care atenuarea naturală se desfăşoară doar în condiţii specifice.
✓ *În principiu, procesul de atenuare naturală poate fi aplicat şi **poluanţilor anorganici** (ex. **metale grele**) cu toate că ei, spre deosebire de poluanţii organici, nu pot fi descompuşi chimic sau biologic. În anumite condiţii metalele grele pot fi imobilizate şi în consecinţă, disponibilitatea lor pentru procesele de transport în subteran este foarte redusă. Procesele geochimice majore ce contribuie la fixarea şi localizarea poluanţilor sunt: sorbţia şi precipitarea. Procesele de imobilizare a metalelor grele în sol/sedimente pot fi mediate de activitatea microbiană.*

Reacţiile de sorbţie ale unor metale dizolvate (Cu, Co, Ni, Zn, etc.) depind de condiţiile de pH. Aceste reacţii sunt reversive, rezultând o remobilizare a metalelor în momentul scăderii pH. Creşterea pH determină precipitarea Al la suprafaţa fragmentelor mediului ce formează acviferul.

Adoptarea unei decizii privind alegerea unei anume metode de remediere trebuie să ţină seama de următoarele etape:
• identificarea tuturor poluanţilor prezenţi în zonă;
• evaluarea riscului pentru mediu şi populaţie luând în considerare toţi poluanţii;
• investigarea spectrului hidrodinamic al curgerii subterane;
• estimarea volumului poluanţilor din zonă;
• analiza proceselor fizice, chimice sau biochimice de degradare sau atenuare a poluanţilor;
• calcularea duratei proceselor de remediere pentru toate metodele propuse;
• analiza cost-eficienţă pentru toate metodele propuse;
• investigarea reacţiei sociale pentru metodele propuse;
• *alegerea şi aplicarea metodei de remediere potrivite.*

STUDIU DE CAZ 1 - Delimitarea zonei de protecţie sanitară pentru captarea Vlădeşti (realizate conform **Hotărârea nr. 101 din 3 aprilie 1997**, în prezent înlocuită de Hotărârea 932/2005, modificată de Ordinul 1245/2005)

Captarea este alcătuită din şapte puţuri verticale cu adâncimi de 8-10 m şi un dren orizontal amplasate pe malul stâng al pârâului Olăneşti. Debitul de apă furnizat de captare este de 100 m^3/zi. Apa pompată provine din *acviferul freatic* din lunca pârâului Olăneşti. Acviferul freatic exploatat este în *comunicare hidraulică directă cu pârâul Olăneşti*. Apa furnizată de captare, după o simplă clorinare la puţul central al captării este utilizată de o parte a locuitorilor comunei Vlădeşti.



Datele climatice şi hidrologice implicate în: evaluarea *realimentării acviferelor* prin infiltraţii şi în cuantificarea *vulnerabilităţii la poluare a acviferelor*:

- ***Temperatura*** medie anuală este de **8-9^{o}C** în zona **Vlădeşti** şi de **10,2^{o}C** în zona staţiei de tratare **Cetăţuia**.
- ***Precipitaţiile*** medii anuale sunt de **850 mm**. Cantităţile medii lunare cele mai mari cad în luna iunie şi au totalizat **96,2 mm (la Govora),** iar cantitatea maximă căzută în 24 de ore a fost de **128,0 mm (17 iunie 1920).** Cantităţile medii lunare cele mai mici cad în **februarie (45,1 mm).**
- ***Evapotranspiraţia*** potenţială este de **680 mm/an** valoare care reprezintă **80%** din precipitaţia medie anuală În aceste condiţii cantitatea de apă rămasă pentru scurgerea de suprafaţă şi pentru infiltraţii este de **170mm/an.**

✓ ***Pârâul Olăneşti*** este principala apă de suprafaţă din zona captării Vlădeşti, fiind al doilea afluent principal al Oltului de la intrarea în judeţul Vâlcea. *Debitul* mediu multianual al pârâului Olăneşti este estimat între **3,8 şi 4,2 m3/s**. Lăţimea şesului aluvial variază între **200 şi 500 m**, uneori chiar mai mult.
✓ ***Râul Olt*** este principalul curs de apă din Judeţul Vâlcea, având un *debit* mediu multianual, în dreptul localităţii Râmnicul Vâlcea de **140 m^3/s**. În lunile foarte ploioase debitul poate ajunge la **1400 m^3/s**, la viitura din 1975 Oltul având debitul maxim de **2000 m^3/s**.
✓ În dreptul localităţii Râmnicul Vâlcea, respectiv în zona staţiei de tratare a apei, râul Olt este regularizat.

Acviferele freatice, în zona cercetată, sunt acumulate în depozite grosiere (**pietrişuri şi nisipuri**) cu grosimi cu prinse între **8 si 15 m**. ***Conductivitatea hidraulică*** a depozitelor permeabile este cuprinsă între **60 şi 600 m/zi** (valori determinate in forajele de captare realizate în zona Râmnicu Vâlcea).
Nivelele hidrostatice sunt la adâncimi cuprinse între **0,5 si 2,5m**, acviferele freatice fiind în legătură hidraulică directă cu reţeaua hidrografică. Legătura hidraulică directă cu reţeaua hidrografica are drept consecinţe:

- Creşterea potenţialului de debitare al acviferului prin alimentare de mal;
- Creşterea vulnerabilităţii la poluare a acviferului;
- *Calitatea apei* din acviferele freatice este dependentă de calitatea apei din reţeaua hidrografică.

Modelarea matematică a acviferului captat la Vlădeşti are ca obiective:

- Calculul direcţiilor şi vitezelor de curgere ale apei subterane în zona captării;
- Simularea unor scenarii de poluare a acviferului, cu efect imediat asupra calităţii apei captate.

Delimitarea spaţială a modelului hidrodinamic s-a realizat prin linia de cumpăna a apelor de suprafaţă care se închide:

- în amonte, la confluenţa *pârâului Cornetului* cu *pârâul Olăneşti*(secţiunea S1);
- în aval, la 250m de confluenţa *pârâului lui Ionel* cu *pârâul Olăneşti* (secţiunea S3).

Pentru precizarea *grosimii* (9-10m) şi *lăţimii* acviferului freatic în acest areal s-au realizat trei secţiuni hidrogeologice schematice (S1, S2, S3)

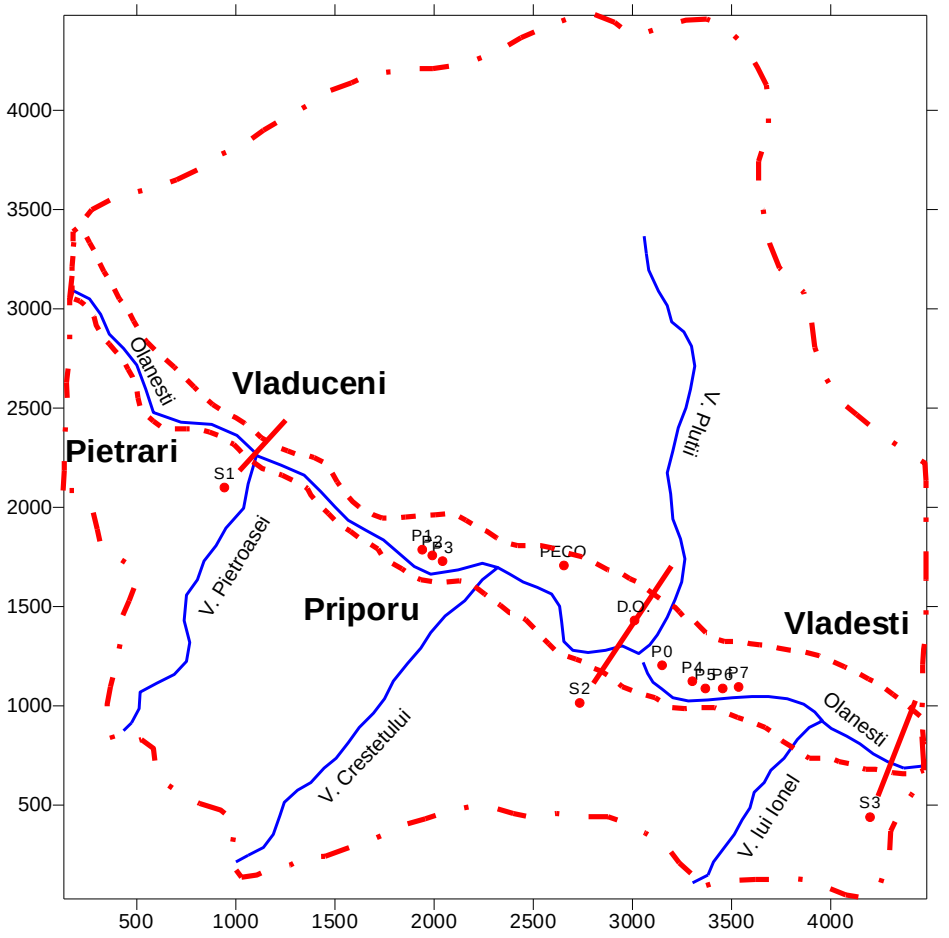
- *Condiţiile pe conturul modelului hidrodinamic* au fost alese de tip *sarcină piezometrică impusă (Dirichlet)* din lipsa unor informaţii privind alimentarea din infiltraţii a acviferului freatic şi a conductanţei talvegului pârâului Olăneneşti. Acest tip de condiţii pe conturul modelului va introduce o supraestimare a potenţialului de debitare a acviferului.
- *Conductivitatea hidraulică* medie utilizată este K= 60 m/zi, iar *porozitate activă* na=25%.

Curgerea este considerată *staţionară* şi *plan orizontală*, realimentarea prin infiltraţii fiind considerată neglijabilă, datorită extinderii reduse a şesului aluvionar şi evapotranspiraţiei potenţiale mari din zonă.

Modelul a fost încadrat într-un pătrat cu latura de 4550 m. Pentru rezolvarea numerică a fost utilizată metoda diferenţelor finite într-o reţea pătratică de:

10000 celule pentru curgerea în regimul natural

17000 celule pentru curgerea în regim influenţat (pentru îndesirea reţelei în zona puţurilor de captare şi a drenului orizontal.



Elemente de schematizare a conditiilor hidrogeologice pentru captarea Vladesti

Legenda

Cumpana apelor de suprafata

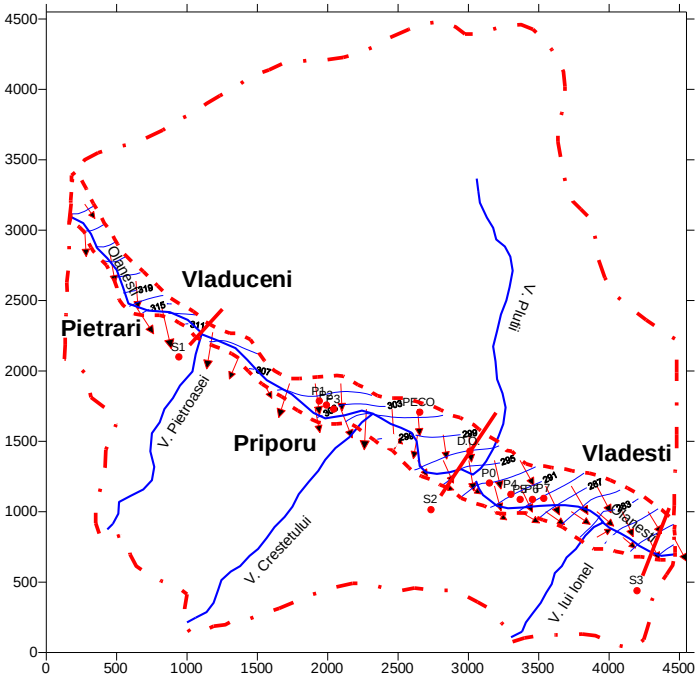
Limita zonei colinare

P1 Putul de captare P1

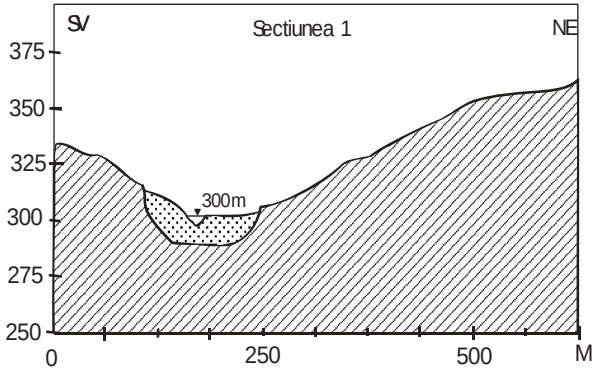
S1 Sectiunea hidrogeologica S1

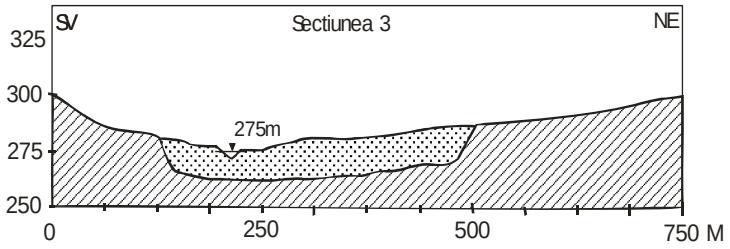
Modelul hidrodinamic în regim natural

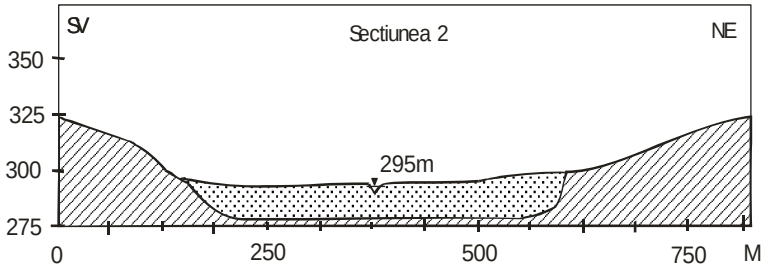
Regimul natural de curgere al acviferului freatic indică o drenare a acestuia de pârâul Olăneşti (*Fig.4*). Pentru zona acviferului freatic sau calculat 813 valori ale vitezei de curgere cuprinse între *0,8 şi 8,73 m/zi*.



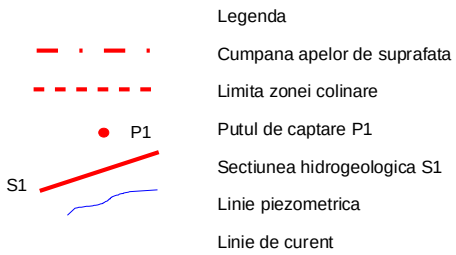
Spectru hidrodinamic al acviferului freatic din zona captarii Vladesti, în regim natural

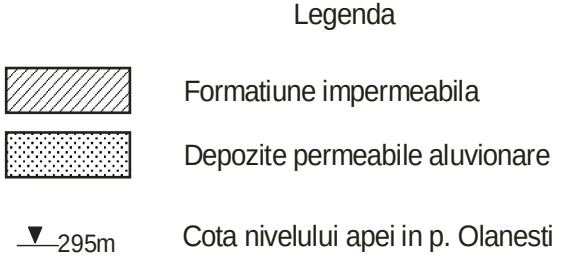






Sectiuni hidrogeologice in zona captarii Vladesti





Modelul hidrodinamic în regim influenţat

Regimul natural de curgere al acviferului freatic este influenţat de captarea Vlădeşti care din cele şapte puţuri verticale şi un dren orizontal pompează în reţeaua de alimentare a comunei Vlădeşti între 100 şi 130 m^3/zi.

Pe baza acestor debite şi a nivelelor dinamice măsurate în puţurile captării s-a calculat spectrul hidrodinamic în regim influenţat.

Se remarcă o *drenare accentuată a râului de către acvifer* în zona de influenţă a puţurilor şi drenului.

Vitezele de curgere au valori cuprinse între *0,004 şi 60,2 m/zi*.

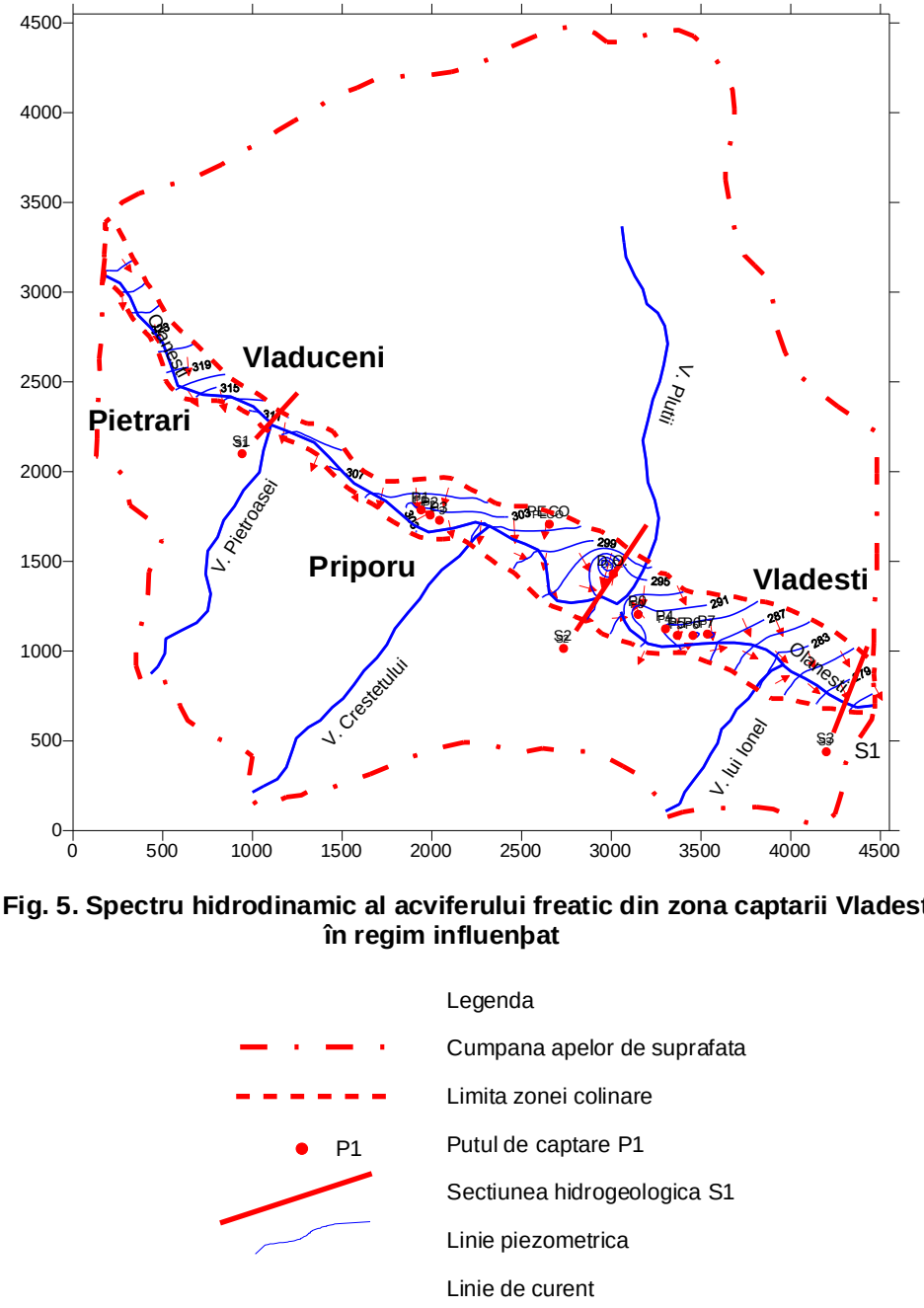
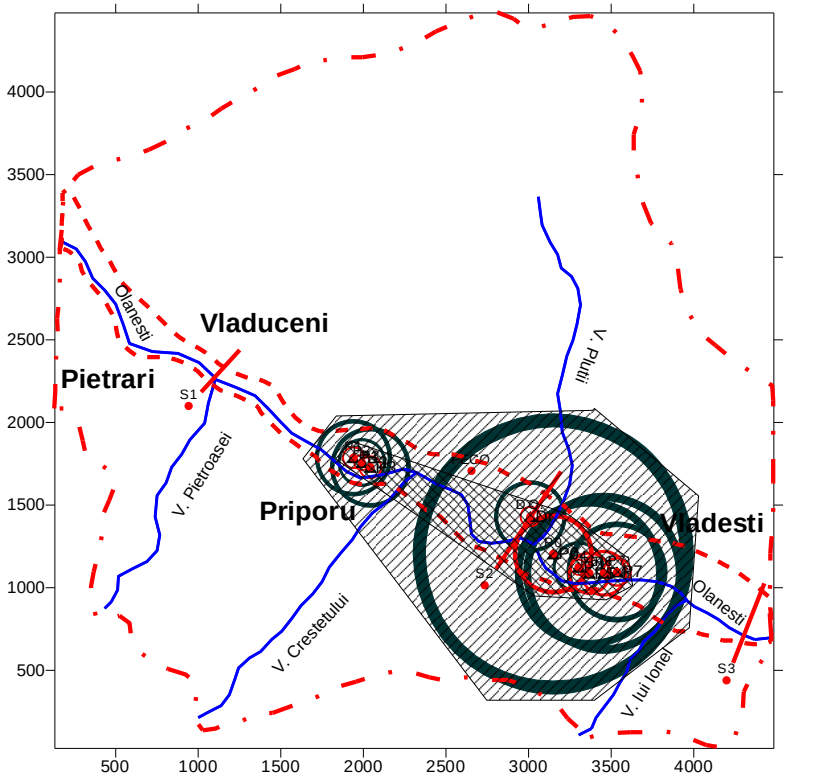


Fig. 5. Spectru hidrodinamic al acviferului freatic din zona captarii Vladesti, în regim influenþat

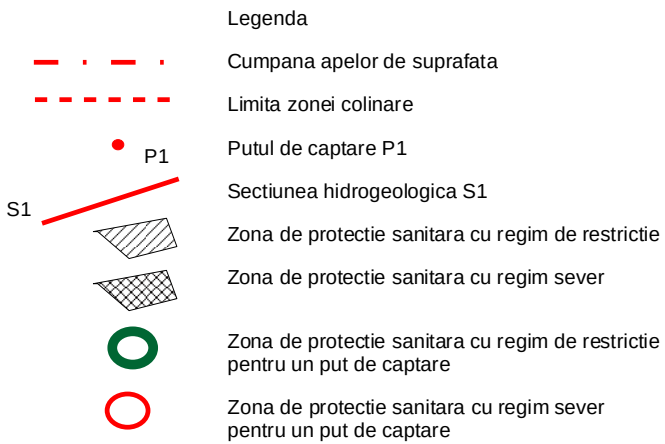
✓ Dimensionarea zonelor de protecţie sanitară cu regim sever şi cu regim de restricţie - criteriul *timpului mediu de tranzit al unei particule de apă hidrodinamic active,* folosindu-se în calcule parametrii acviferului şi rezultatele obţinute cu ajutorul modelului hidrodinamic.

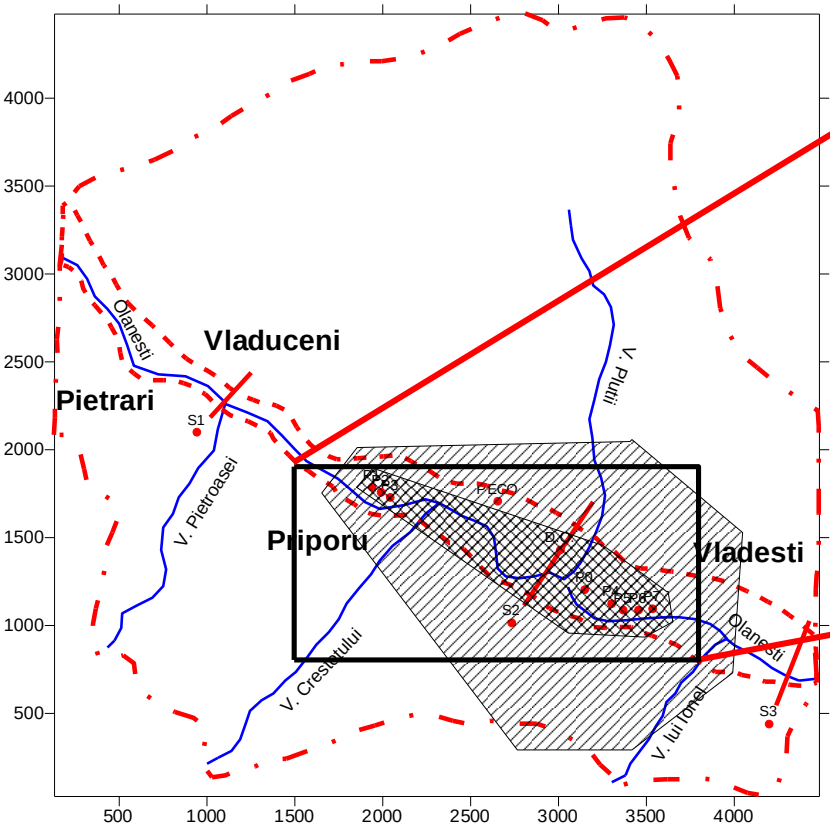
✓ Mărimea *zonei de protecţie sanitară cu regim sever* - să fie asigurată o durată de parcurgere de minim 20 zile pentru orice picătură de apă presupusă contaminată care s-ar infiltra la limita acestei zone şi ar ajunge la locul de captare al apei.

✓ Mărimea *zonei de protecţie sanitară cu regim de restricţii* - asigură protecţie faţă de contaminarea bacteriană şi impurificarea chimică, luând în considerare o durată de 50 zile pentru parcurgerea distanţei de la punctul de infiltrare până la limita zonei de protecţie sanitară cu regim sever.

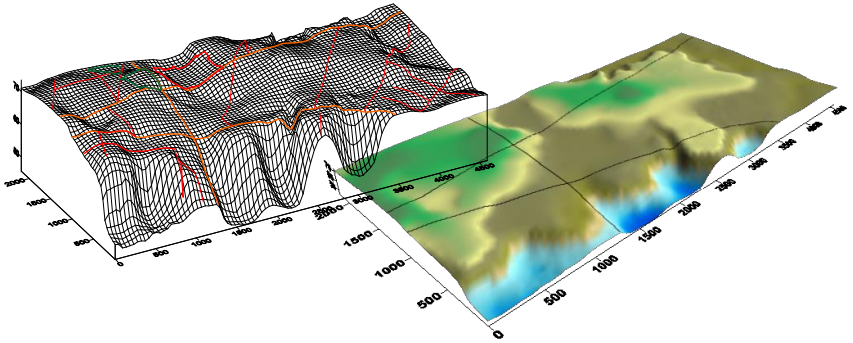


Zone de protectie sanitara propuse cu elementele fiecarui put al captarii

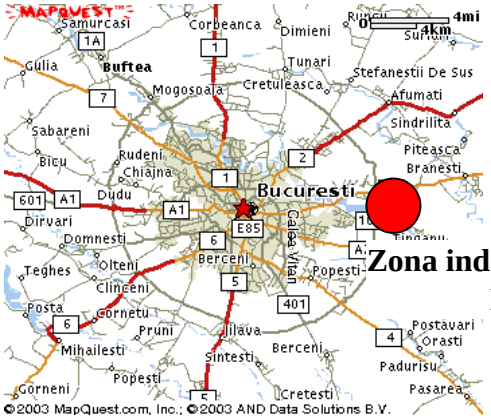


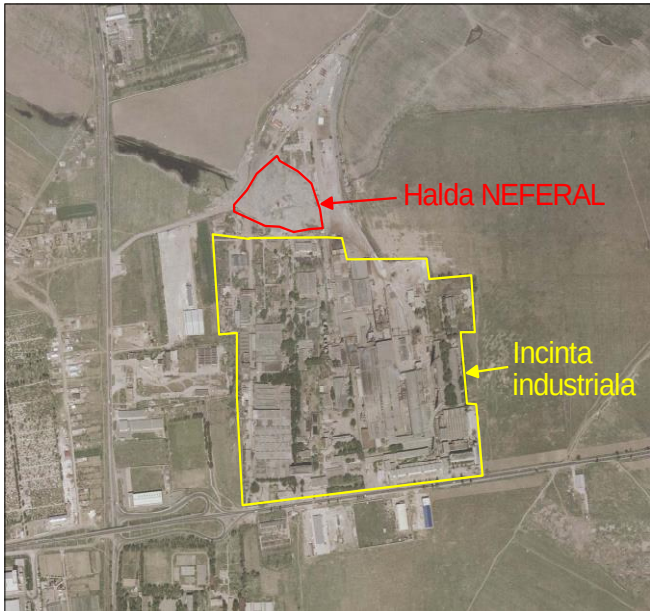

Vladuceni
Pietrari
Olanesti
V. Pietroasei
V. Crestatului
V. Pluti
V. lui Ionel
Priporu
Vladesti
S1
S2
S3
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13
P14
P15
P16
P17
P18
P19
P20
P21
P22
P23
P24
P25
P26
P27
P28
P29
P30
P31
P32
P33
P34
P35
P36
P37
P38
P39
P40

STUDIU DE CAZ 2 – Studiu hidrogeologic realizat pentru evaluarea poluării apelor subterane cu metale grele, în zona industrială Pantelimon

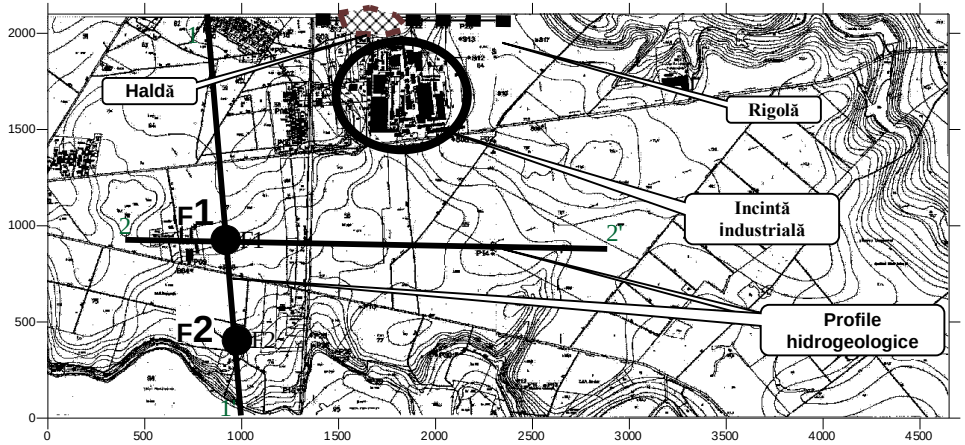


- Zonă interfluvială de câmpie altitudinea medie 80 m;
- depozite loessoide cu grosimi variind între 2 şi 6 m;
- acvifer freatic format în nisipuri fine cu intercalaţii de nisipuri cu pietrişuri, situat în baza depozitelor loessoide la adâncimi de 6 – 9 m, drenat de lacul Cernica;
- tip de sol – cernoziom argiloiluvial;
- temperatura medie multianuală 11,2^0. C cu medii lunare variind între -1,8^0. C şi 22.4^0. C;
- precipitaţii de 550 – 600 mm/an;
- valori ale evaportanspiraţiei reale 450 – 500 mm/an.





Zona ind. Acumulatorul - Neferal



Schematizarea principalelor elemente ale perimetrului studiat

S.C. NEFERAL S.A.

Secţia Cupru:

- bronzuri – 2.531 t/an;
- alame – 644 t/an;
- CuP9 – 16 t/an.

Secţia Plumb:

- aliaje de plumb pentru acumulatori – 3287 t/an;
- aliaje antifricţiune – 56 t/an;
- aliaje de lipit – 168 t/an;
- zamac – 162 t/an.

Secţia Aluminiu - aluminiu secundar şi aliaje de aluminiu – 1554 t/an.

Secţia Wolfram – amestecuri – 33.977 kg/an.

Secţia Al – etape tehnologice:

- **materia primă – materiale recuperabile sunt supuse operaţiilor de separare şi sfărâmare;**
- **separarea fierului în cuptoare de deferare;**
- **topire sub strat de săruri pentru evitarea oxidării;**
- **materialul topit se toarnă în lingouri → aliajul Al 380;**
- **zgura rezultă (466 t/an cu Al între 1 – 7 %) este depozitată în halda cu o suprafaţă de 16.000 mp.**

Prelevarea probelor

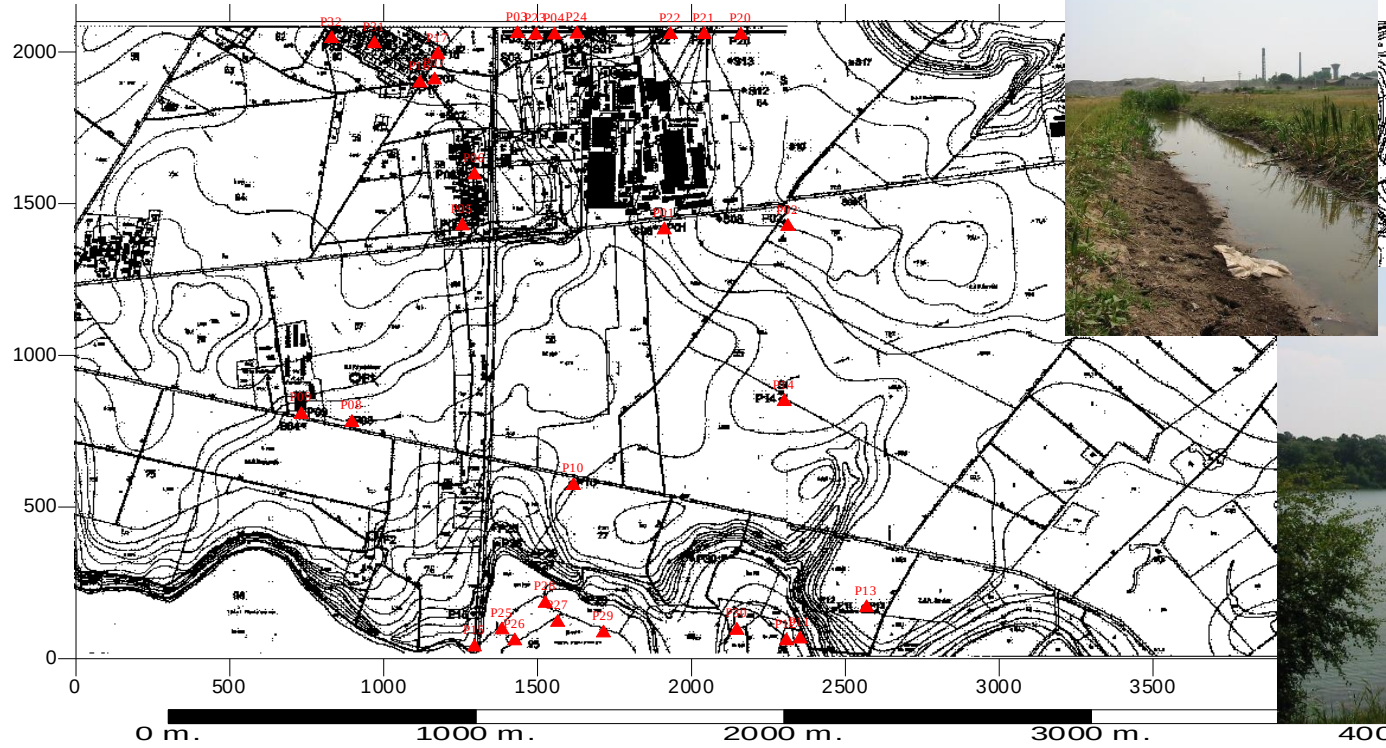
Probarea sursei de poluare:

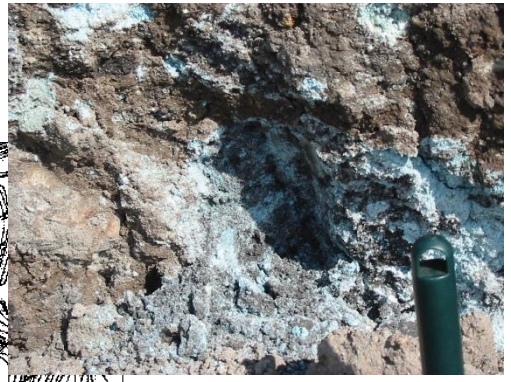
- **Halda** –10 – 20 cm adâncime din diverse puncte;
- **Rigola –** pe toată lungime sa, la intervale de 50 - 100 m.

Probarea acviferelor – au fost prelevate probe din forajele existente şi din puţurile de alimentare cu apă, atât din acviferul freatic cât şi din cel de medie adâncime.

Probarea lacului – Lacul Cernica a fost probat relativ echidistant pe toată latura sa nordică.

Locaţiile punctelor de prelevare a probelor





Etapa analitică

S-a determinat temperatura şi pH-ul în momentul probării, iar ulterior au fost pregătite după metodologia standard pentru analiza conţinutului de metale dizolvate (filtrate prin filtru de 0.45 µm şi acidifiate cu HNO3 până la pH = 1). Au fost conservate şi transportate la temperatură scăzută.

Metodele de analiză folosite:

Adsorbţie atomică în cuptor cu grafit pentru metale (Cu, Zn, Pb, Al);

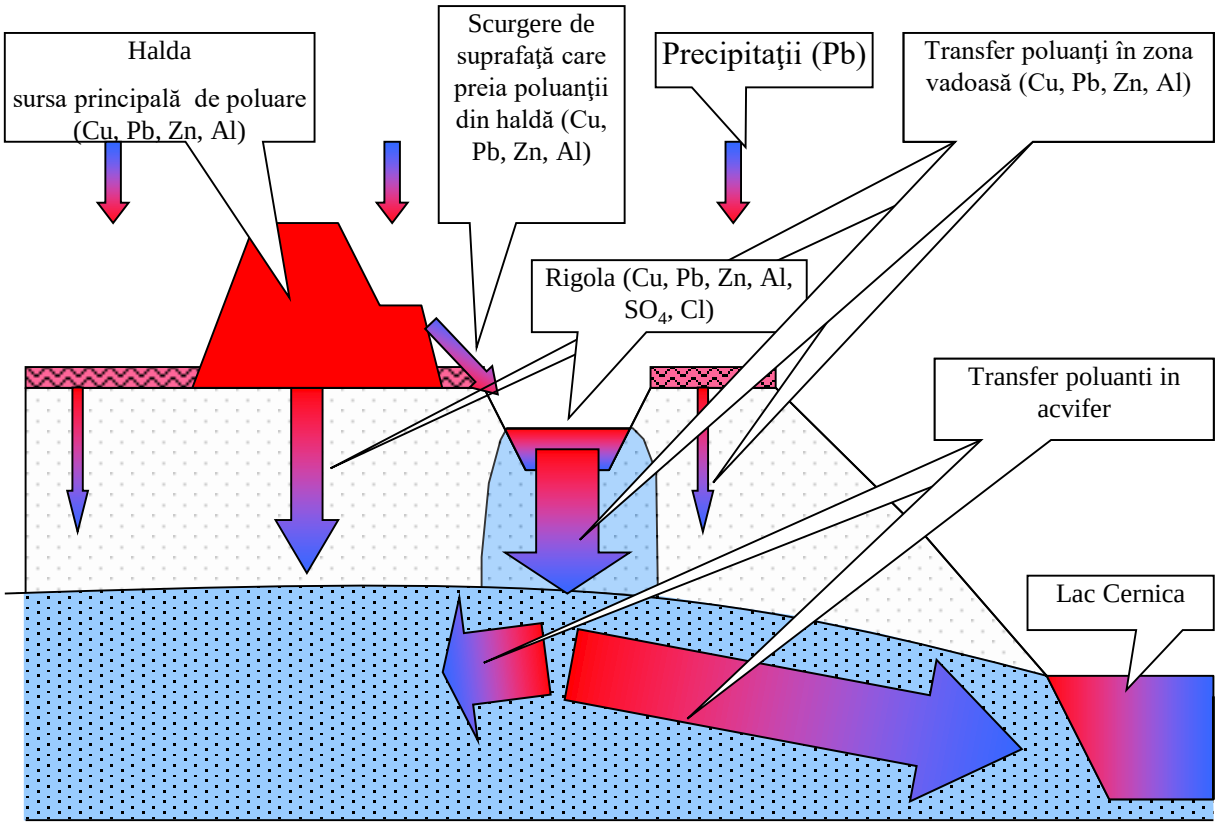
Cromatografie ionică pentru anioni (Cl, SO4)

Au fost analizate 34 de probe de apă după următoarea metodologie de selecţie a elementelor:

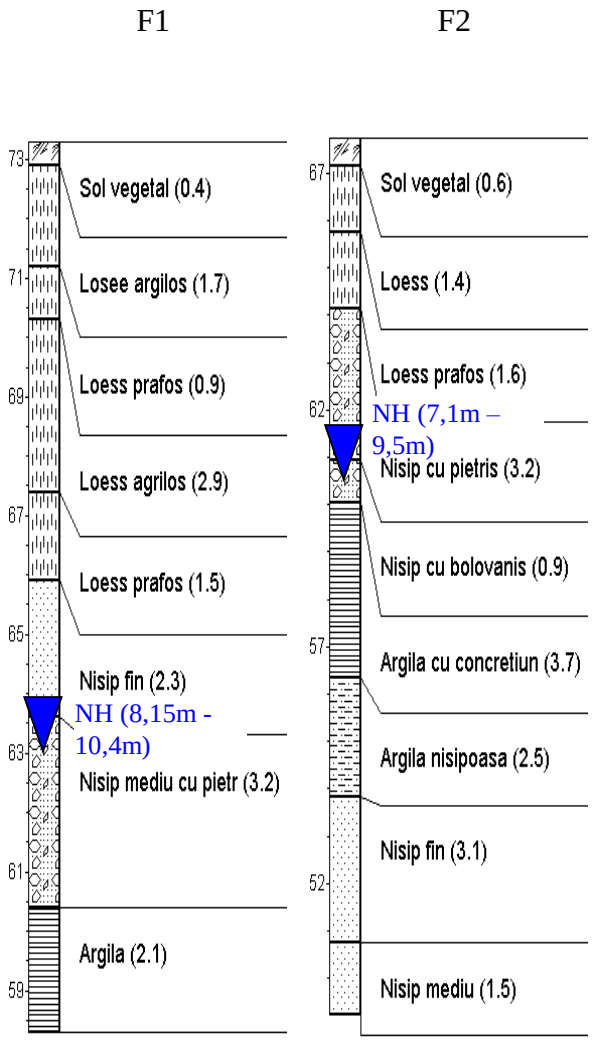
- determinarea pH şi temperatură la colectarea probei;
- pentru probele din rigolă s-a făcut o analiză chimică completă;
- s-au identificat metalele care prezintă conţinuturi semnificative – Al, Pb, Cu, Zn;
- pentru toate probele s-a determinat numai conţinutul de Al, Pb, Cu şi Zn;
- s-a determinat pentru toate probele conţinutul de Cl şi SO4.

Modelarea poluării cu metale grele

- Model pentru zona vadoasă;
- Model pentru zona saturată (acvifer freatic)



Modelul conceptual al sistemul în care de produce poluarea cu metale grele în zona Pantelimon



Modelul conceptual al sistemului în care se produce poluarea – spaţiul

• plan ***orizontal*** a spaţiului în care se modelează transportul metalelor grele este un ***dreptunghi*** orientat cu latura lungă (3500m) paralel cu direcţia nord-sud, direcţia principala de curgere în acviferul freatic şi latura mai scurtă (3000m) paralelă cu direcţia vest-est.

• plan ***vertical*** spaţiul în care se este modelat transportul metalelor grele are două componente distincte:

• **Zona vadoasă;**

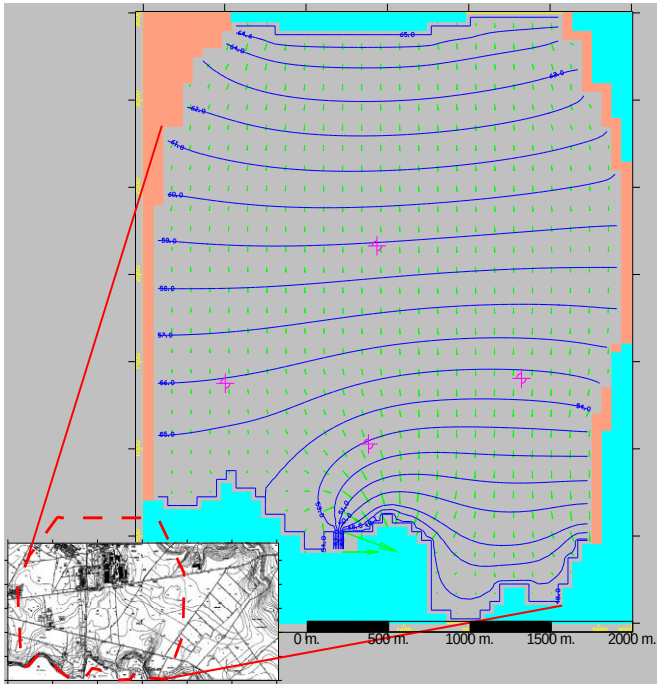
• **Acviferul freatic.**

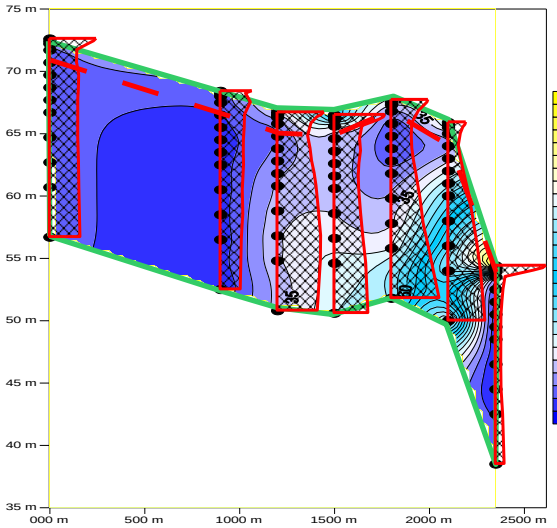
Grosimea zonei vadoase determinată atât direct în forajele cât şi prin SEV-uri – 2 – 10 m, compusă în principal din loess, loess argilos, loess prăfos, în alternanţă, în structuri cu grosimi variabile (0,5 – 3 m).

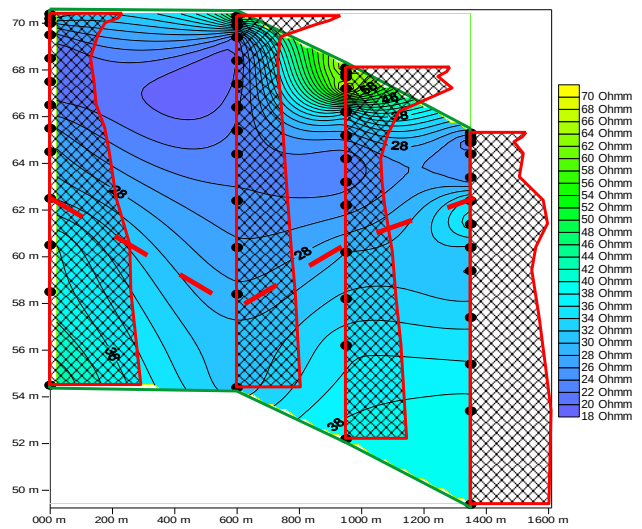
Acviferul freatic - depozite Pleistocene,

• culcuşul - adâncimi cuprinse între 12.9 m şi 14,1 m

• grosimea acviferului variază între 4 şi 5 m.





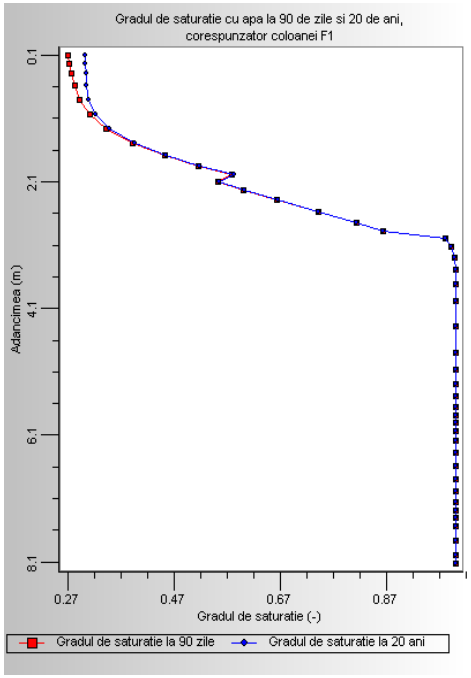


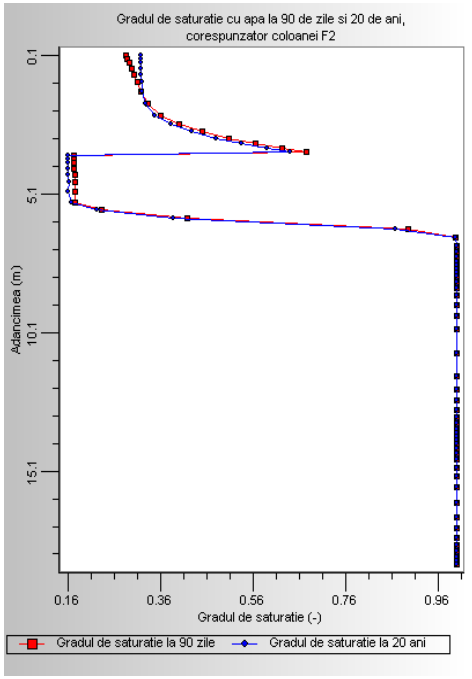
Secţiuni electrometrice N-S şi E-V realizate pe baza datelor obţinute prin SEV

Nivelul hidrostatic rezultat din variatia rezistivitatii

Modelul conceptual al sistemului în care se produce poluarea – **parametrii modelului**

- **parametrii zonei vadoase;**
- **parametrii acviferului;**
- **parametrii ai transferului poluanţilor.**





Parametrii zonei vadoase:

• infiltrarea eficace pe sezoane (**an mediu – 150 mm**):

• **vara – 29 mm;**

• **primăvară – 39 mm;**

• **vara – 50 mm;**

• **toamna - 32 mm.**

Conductivitatea hidraulică în regim ***saturat*** a depozitelor - cca. 0,3 – 0,35 m/zi (pentru depozitele loessoide) cu rare valori de 10 m/zi (pentru nisipurile fine), valorile cele mai mici corespund argilelor nisipoase – 0,1 m/zi.

Pentru ***umiditatea reală*** a acestor depozite conductivităţile sunt mult mai mici şi se evaluează pe baze profilelor de umiditate.

Parametrii acviferului:

• **porozitatea** – porozitate totală medie – 35%; porozitate efectivă – 25 %

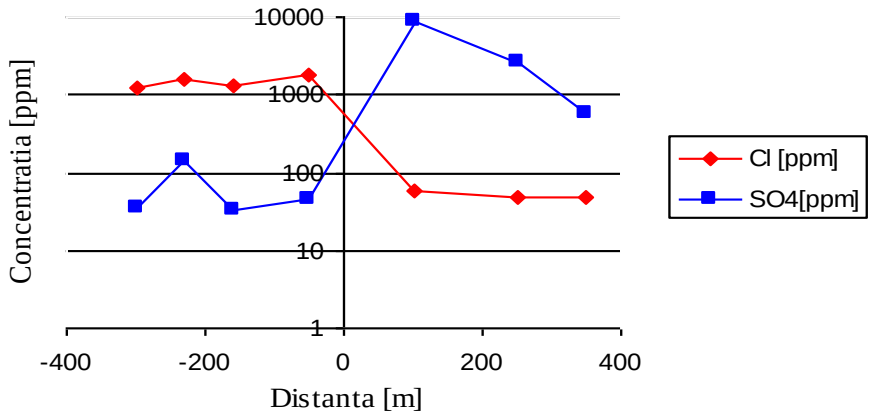
• **conductivitatea hidraulică** – conductivitatea medie 100 m/zi.

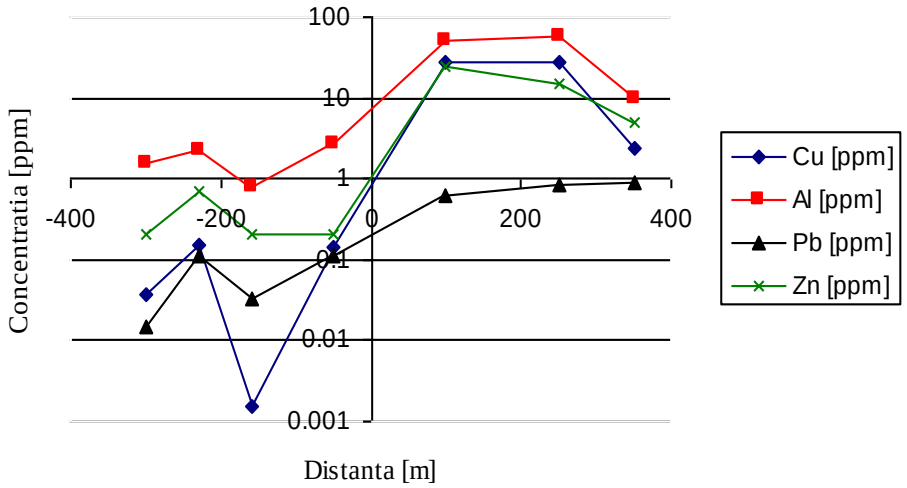
Modelul conceptual al sistemului în care se produce poluarea – parametrii modelului

Parametrii de transfer ai poluanţilor:

Parametrii specifici proceselor de migrare a contaminanţilor (***coeficientul de difuzie, coeficientul de dispersie hidrodinamică, coeficientul de anizotropie a dispersie, izotermele de sorbţie*** **etc**.) sunt evaluate în etapa de calare a modelului realizată cu ajutorul concentraţiilor determinate în apa recoltată din piezometre realizate în acviferul freatic, între haldă şi lacul Cernica.

Concentraţiile metalelor grele în apa recoltată din rigolă utilizate la modelarea poluării cu metale grele.		
Elementul	Concentraţii în apa rigolă[mg/l]	
	Amonte haldă	Aval haldă
Aluminiu	2,250	59,000
Cupru	0,148	28,000
Plumb	0,032	0,908
Zinc	0,713	25,000





Modelul conceptual al sistemului în care se produce poluarea – condiţiile hidrodinamice

• ***zona vadoasă*** delimitată în ***plan vertical***:

  • la partea superioară de o suprafaţă de alimentare continuă prin infiltrare;

  • la partea inferioară de suprafaţa piezometrică, staţionară a acviferului freatic;

• ***acviferul freatic*** delimitat în ***plan pe vertical***:

  • la partea superioară de suprafaţă piezometrica staţionară cu o alimentare nesemnificativa din infiltraţii;

  • la partea inferioară de formaţiuni argiloase impermeabile.

• În ***plan orizontal*** ambele componente sunt încadrate într-un dreptunghi 3500 m / 3000 m:

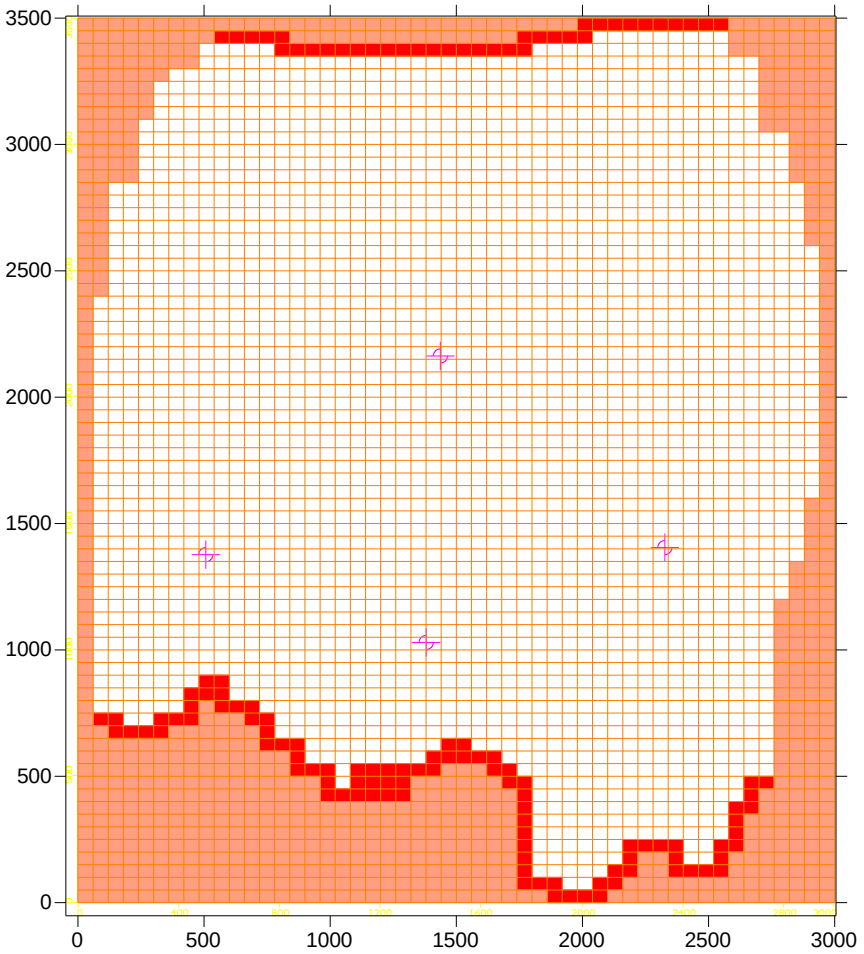
• latura nordică a modelului este o frontieră de tip **sarcină piezometrică constantă** (Dirichlet), plasată la nord de halda Acumulatorul - Neferal (Cota +65m);

• latura sudica este o frontieră de tip **sarcina piezometrică constanta** plasată de-a lungul malului lacului Cernica care drenează acviferul freatic (Cota variabilă de la vest:+54m la este:+48m);

• latura vestică este paralelă cu direcţia de curgere şi este o **frontieră de tip debit nul** (Neumann);

• latura estică este paralelă cu direcţia de curgere şi este o frontieră **de tip debit nul** (Neumann).

Frontierele modelului



Modelarea transferului metalelor grele – zona vadoasă

Evaluarea transferului s-a realizat cu programul WHI UnSat

- Rigola – sursă de poluare de concentraţie constantă;
- Doua scenarii diferenţiate din punct de vedere al litologiei, corespunzătoare forajelor F1 şi F2;

Condiţii iniţiale pentru realizarea modelelor:

• perioada maximă de simulare a fost de 20 ani, incluzând intervale specifice de: 90 zile, 180 zile, 270 zile, 1 an, 2 ani, 3 ani, 4 ani, 5 ani, 6 ani, 7 ani, 8 ani, 9 ani, 10 ani şi 20 ani.

Condiţii de curgere:

• la limita superioară – în primul an de simulare au fost considerate fluxuri variabile sezonier (29 mm/an, 39 mm/an, 50 mm/an respectiv 32 mm/an) iar pentru restul perioadei o valoare medie de 150 mm/an. la limita

• la limita inferioară – sarcină piezometrică constantă.

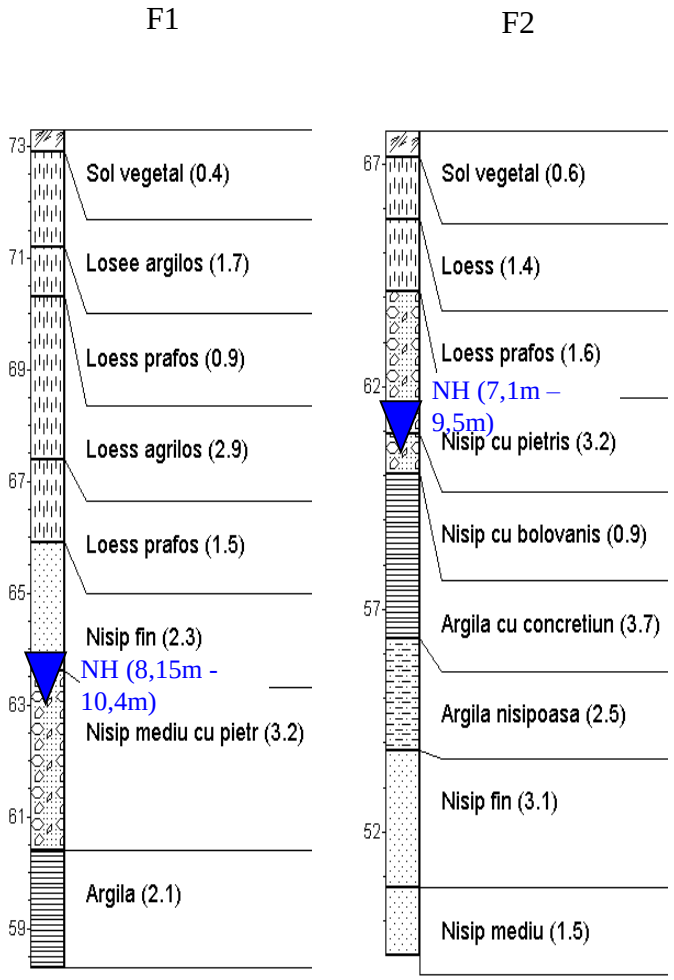
Condiţii de transfer:

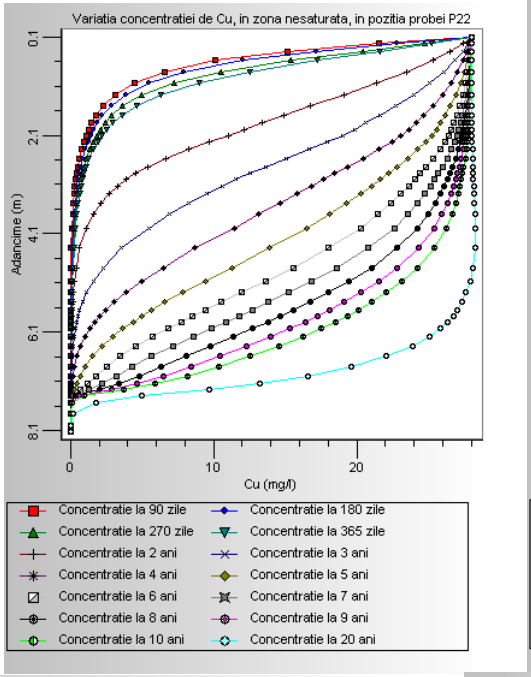
• la limita superioară – tip concentraţie constantă (*cupru* – 2,36 mg/l; 28 mg/l; 0,144 mg/l; *plumb* – 0,908 mg/l; 0,619 mg/l; *zinc* – 5 mg/l; 25 mg/l; 0,713 mg/l; *aluminiu* – 59 mg/l; 2,25 mg/l);

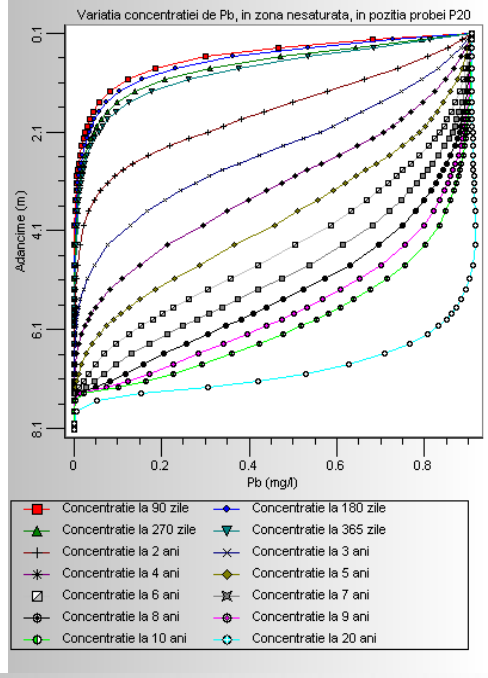
• la limita inferioară – limită caracteristică suprafeţei piezometrice.

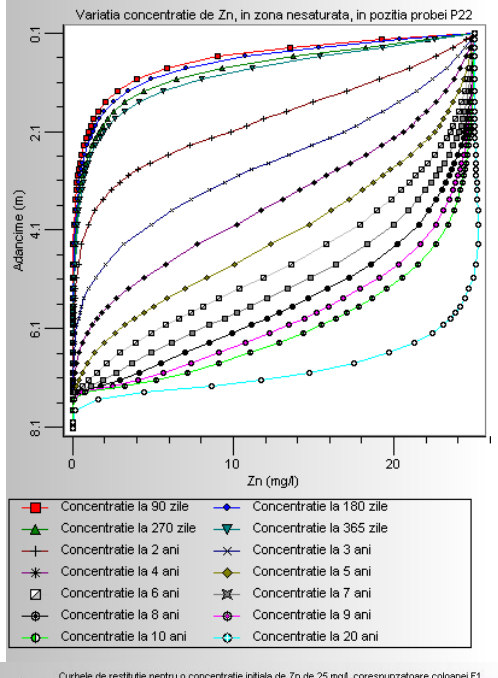
• Conţinutul de umiditate iniţial al solului a fost considerat variabil, funcţie de litologie (între 0,25 – 0,5 vol/vol).

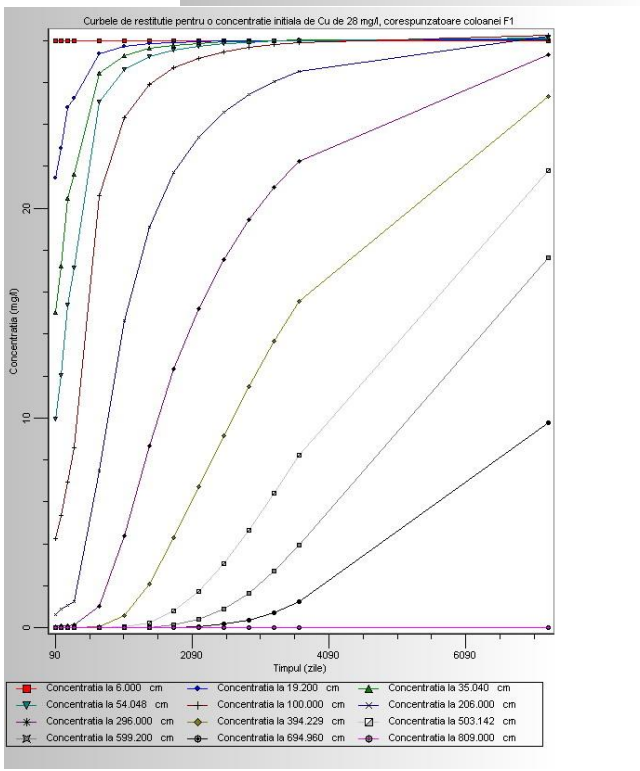
Simularea transferului poluanţilor s-a făcut ţinând cont de adsorbţie, pe baza izotermei lui Langmuir, iar funcţia hidraulică a solului este Van Genuchten.

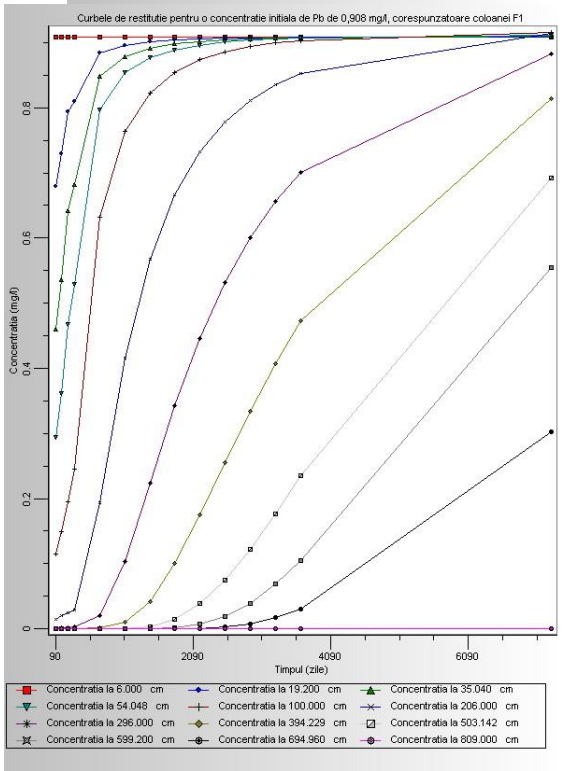


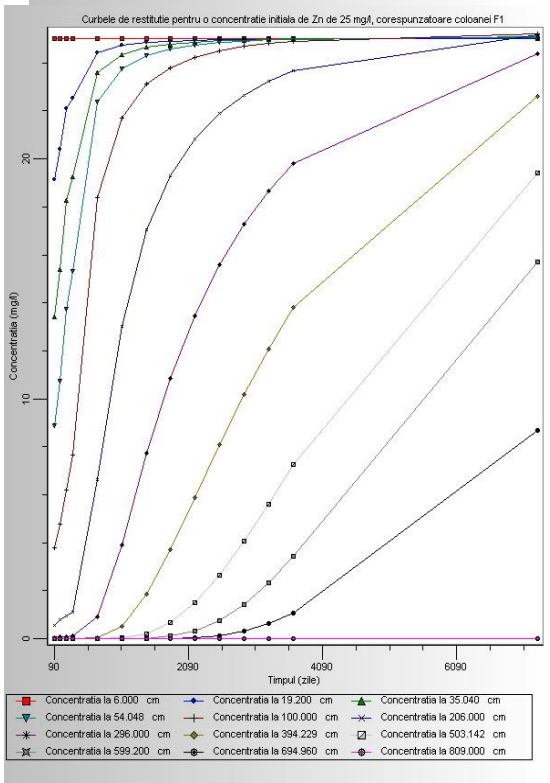

Variatia concentratiei de Cu, in zona nesaturata, in pozitia probei P22
Adancime (m)
Cu (mg/l)
Concentratie la 90 zile
Concentratie la 180 zile
Concentratie la 270 zile
Concentratie la 365 zile
Concentratie la 2 ani
Concentratie la 3 ani
Concentratie la 4 ani
Concentratie la 5 ani
Concentratie la 6 ani
Concentratie la 7 ani
Concentratie la 8 ani
Concentratie la 9 ani
Concentratie la 10 ani
Concentratie la 20 ani


Variatia concentratiei de Pb, in zona nesaturata, in pozitia probei P20
Adancime (m)
Pb (mg/l)
Concentratie la 90 zile
Concentratie la 180 zile
Concentratie la 270 zile
Concentratie la 365 zile
Concentratie la 2 ani
Concentratie la 3 ani
Concentratie la 4 ani
Concentratie la 5 ani
Concentratie la 6 ani
Concentratie la 7 ani
Concentratie la 8 ani
Concentratie la 9 ani
Concentratie la 10 ani
Concentratie la 20 ani


Variatia concentratie de Zn, in zona nesaturata, in pozitia probei P22
Adancime (m)
Zn (mg/l)
Concentratie la 90 zile
Concentratie la 180 zile
Concentratie la 270 zile
Concentratie la 365 zile
Concentratie la 2 ani
Concentratie la 3 ani
Concentratie la 4 ani
Concentratie la 5 ani
Concentratie la 6 ani
Concentratie la 7 ani
Concentratie la 8 ani
Concentratie la 9 ani
Concentratie la 10 ani
Concentratie la 20 ani


Curbele de restitutie pentru o concentratie initiala de Cu de 28 mg/l, corespunzatoare coloanei F1
Concentratia (mg/l)
Timpul (zile)
Concentratia la 6.000 cm
Concentratia la 19.200 cm
Concentratia la 35.040 cm
Concentratia la 54.048 cm
Concentratia la 100.000 cm
Concentratia la 206.000 cm
Concentratia la 296.000 cm
Concentratia la 394.229 cm
Concentratia la 503.142 cm
Concentratia la 599.200 cm
Concentratia la 694.960 cm
Concentratia la 809.000 cm


Curbele de restitutie pentru o concentratie initiala de Pb de 0,908 mg/l, corespunzatoare coloanei F1
Concentratia (mg/l)
Timpul (zile)
Concentratia la 6.000 cm
Concentratia la 19.200 cm
Concentratia la 35.040 cm
Concentratia la 54.048 cm
Concentratia la 100.000 cm
Concentratia la 206.000 cm
Concentratia la 296.000 cm
Concentratia la 394.229 cm
Concentratia la 503.142 cm
Concentratia la 599.200 cm
Concentratia la 694.960 cm
Concentratia la 809.000 cm


Curbele de restitutie pentru o concentratie initiala de Zn de 25 mg/l, corespunzatoare coloanei F1
Concentratia (mg/l)
Timpul (zile)
Concentratia la 6.000 cm
Concentratia la 19.200 cm
Concentratia la 35.040 cm
Concentratia la 54.048 cm
Concentratia la 100.000 cm
Concentratia la 206.000 cm
Concentratia la 296.000 cm
Concentratia la 394.229 cm
Concentratia la 503.142 cm
Concentratia la 599.200 cm
Concentratia la 694.960 cm
Concentratia la 809.000 cm

Modelarea transferului metalelor grele – acvifer freatic

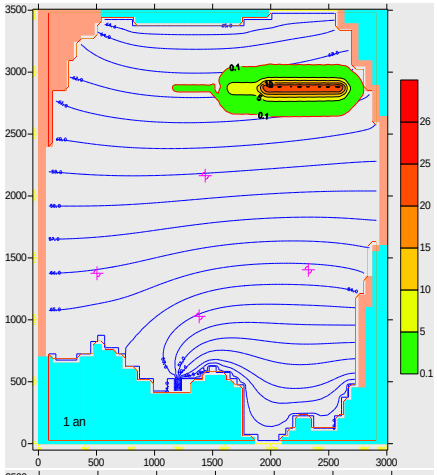
Evaluarea transferului s-a realizat cu programul Visual ModFlow

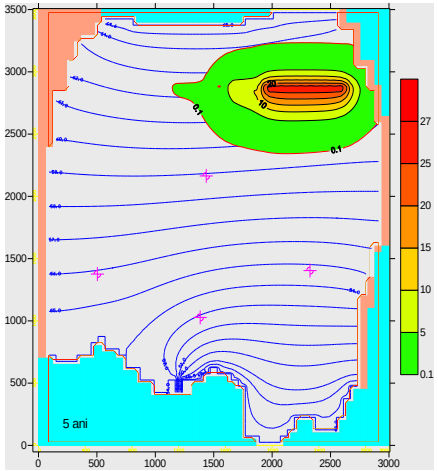
Parametrii utilizaţi pentru modelare:

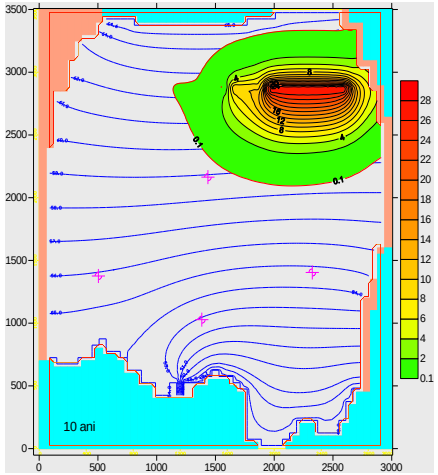
Concentraţiile metalelor grele în apa din acvifer, utilizate la modelarea poluării cu metale grele.		
Elementul	Concentraţii în apa rigolă[mg/l]	
	Amonte haldă	Aval haldă
Aluminiu	2,13	57,000
Cupru	0,140	26,910
Plumb	0,030	0,860
Zinc	0,675	23,98

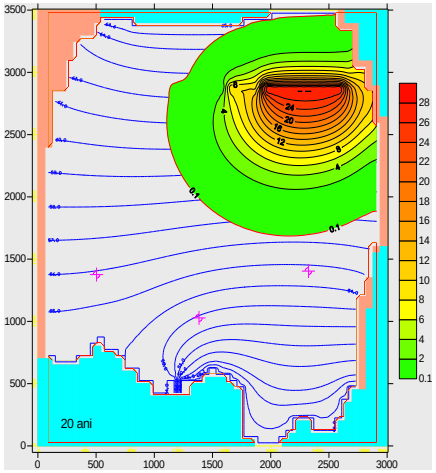
Pentru modelarea variaţiilor concentraţiilor metalelor grele selectate s-au utilizat următorii parametrii:

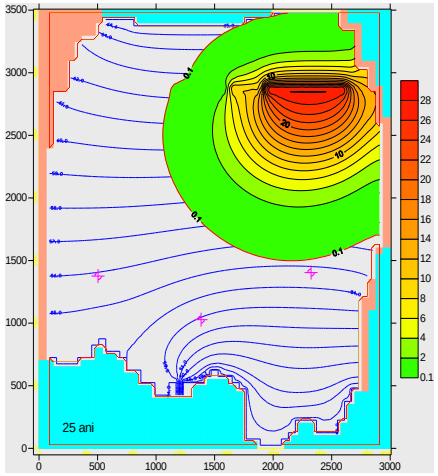
- **un coeficient de difuzie de 1 m2/zi;**
- **dispersivitatea de 100 m cu un raport de anizotropie de 0,1 între cea longitudinală şi cea transversală.**



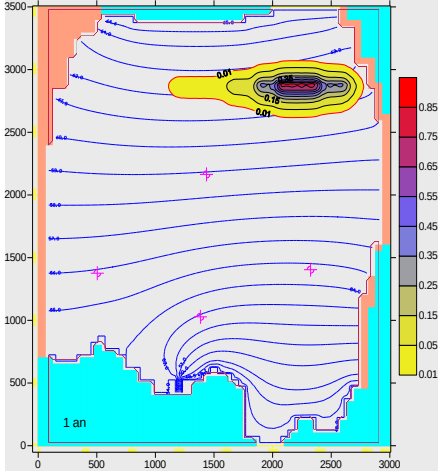


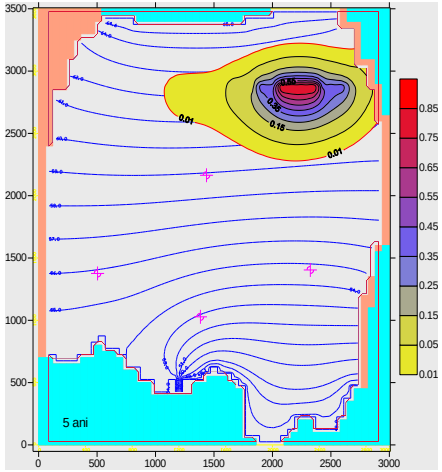


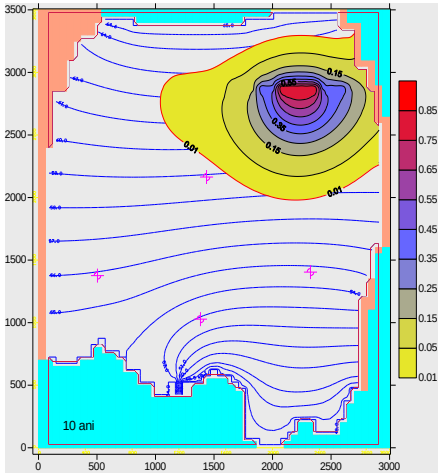


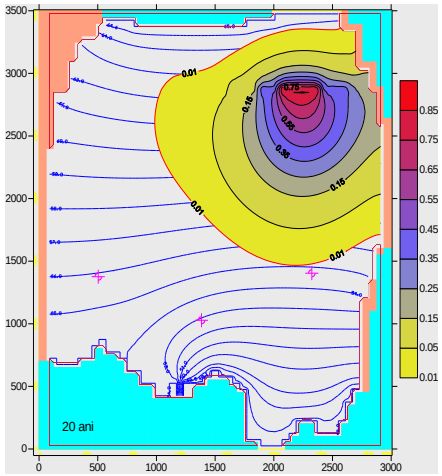


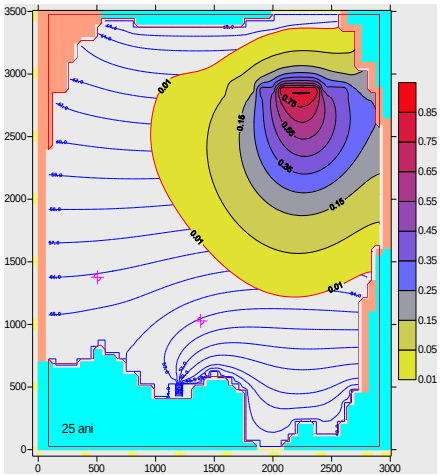
Distribuţia concentraţiei de cupru, la nivelul acviferului, după 1 an, 5 ani, 10 ani, 20 ani şi 25 ani.



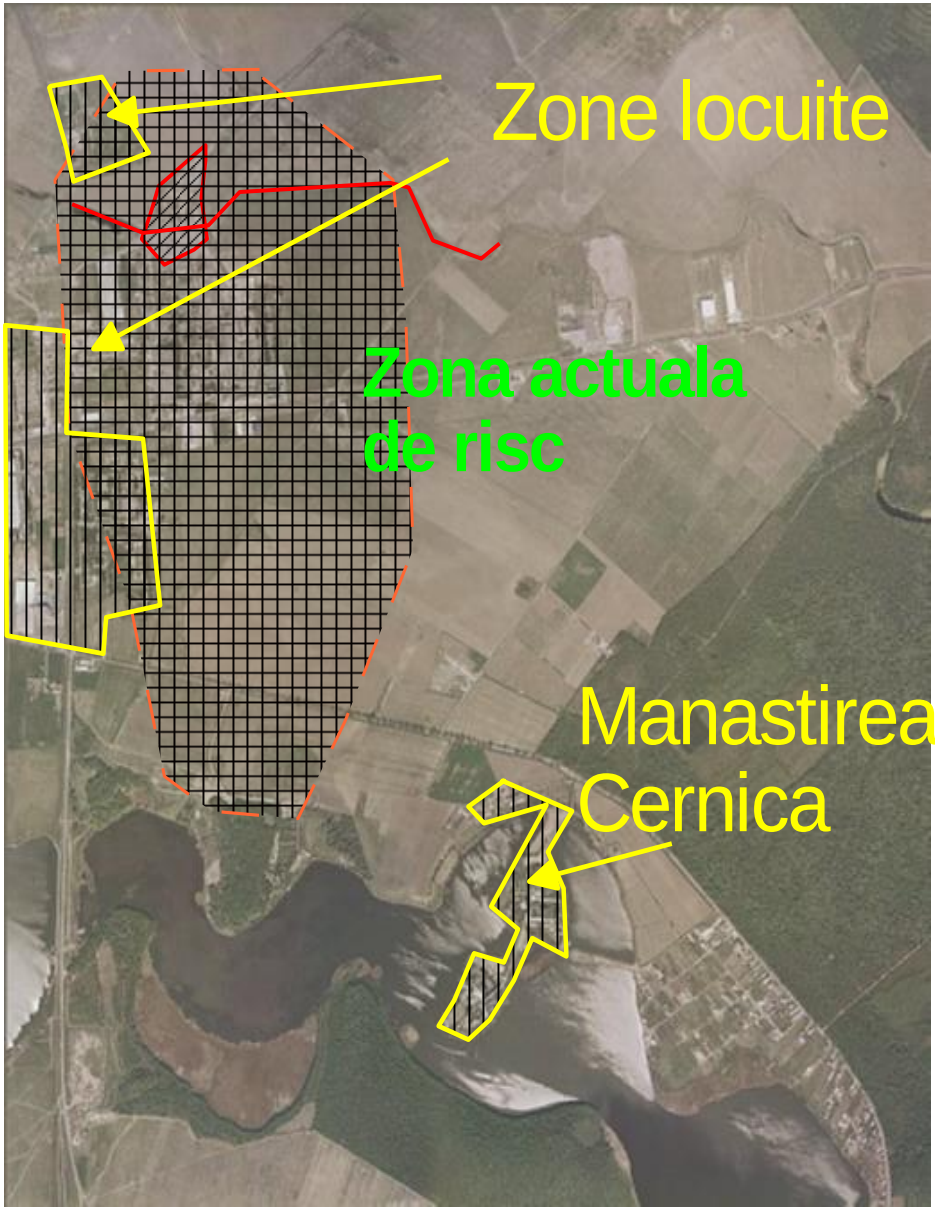


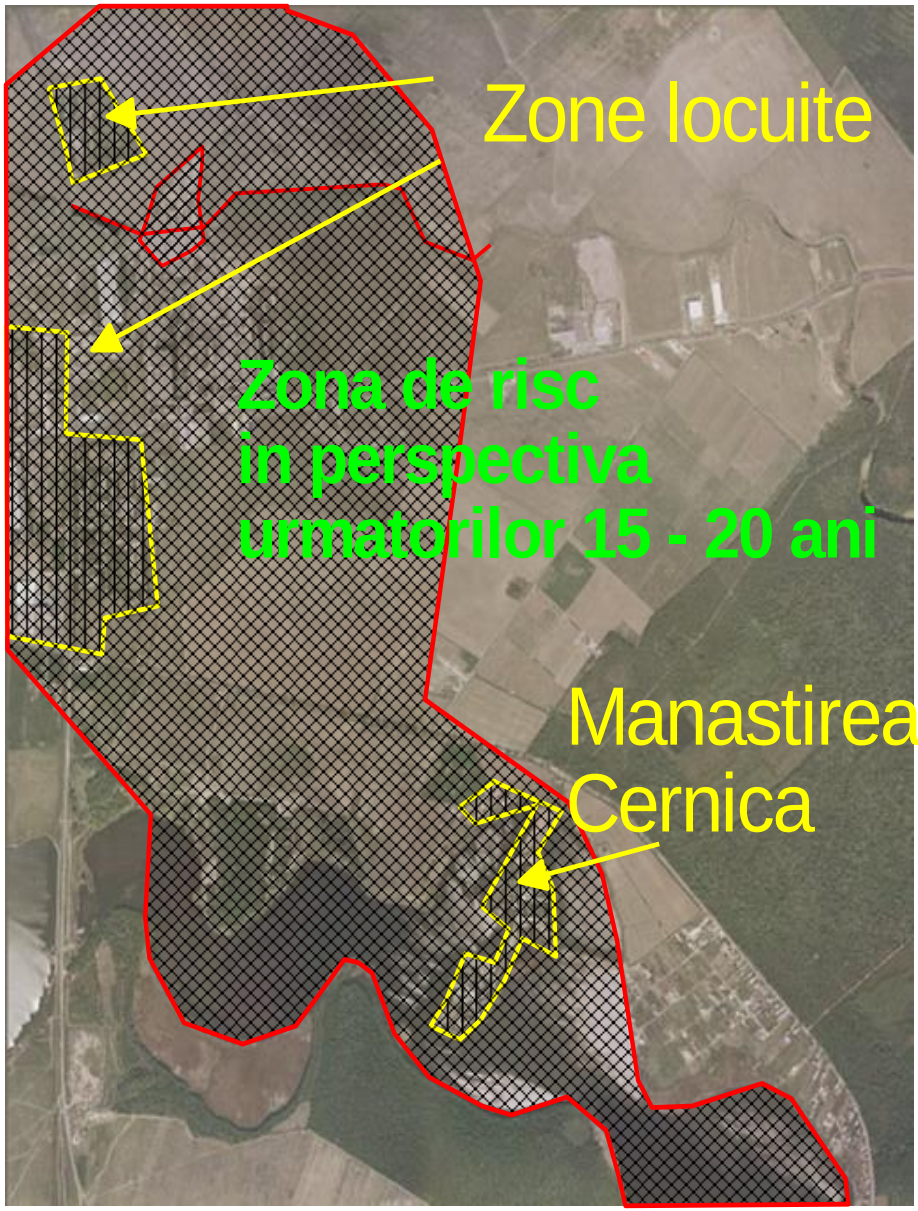






Distribuţia concentraţiei de plumb, la nivelul acviferului, după 1 an, 5 ani, 10 ani, 20 ani şi 25 ani.


Zone locuite
Zona actuala
de risc
Manastirea
Cernica


Zone locuite
Zona de risc
in perspectiva
urmatorilor 15 - 20 ani
Manastirea
Cernica

Utilizarea GIS în domeniul protecţiei mediului

GESTIUNEA DATELOR DE MEDIU

- **Categoriile de date**
  - **Numerice (spaţiale)**
  - **Alfanumeice**
- **Date de mediu**
  - **Apa**
  - **Aer**
  - **Sol**
  - **Biodiversitate**
  - **Deşeuri**
  - **Păduri**
  - **Radioactivitate**
- **Achiziţia datelor de mediu**
  - **Aparatura de măsura**
  - **Reţele de monitorizare**
  - **Data spaţiale**
- **Stocarea datelor**
  - **Tipuri de baze de date**
    - **Acces**
    - **Oracle**
    - **GIS**
- **Exploatarea bazelor de date**
  - **Actualizarea datelor de mediu**
  - **Selectarea datelor pentru prelucrări specifice**
  - **Reprezentarea distribuţiei spaţiale a datelor**

Un **GIS** este un sistem informatic care permite captarea (introducerea), stocarea, integrarea, manipularea, analiza şi vizualizarea datelor care au referinţă spaţială.

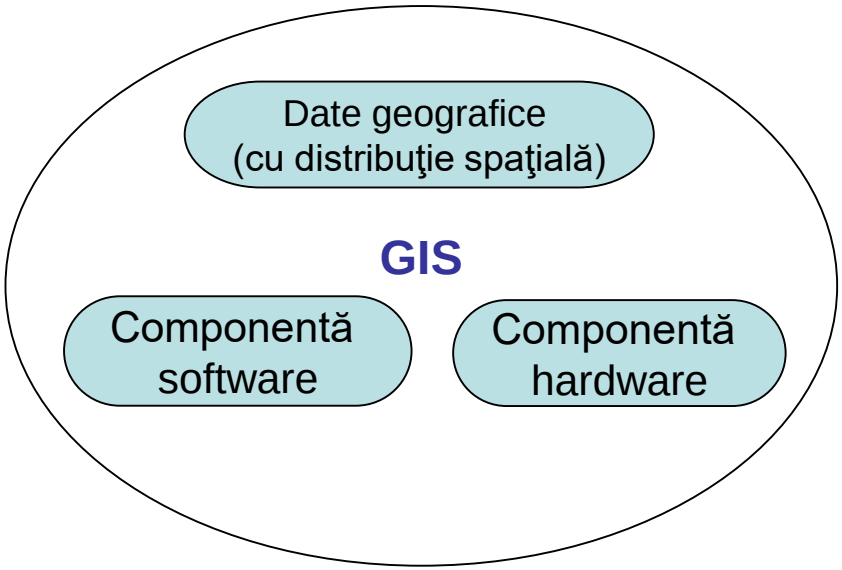
Un GIS, poate răspunde la următoarele întrebări:

• ***Ce este la……,*** adică localizarea unei anumite caracteristici. O locaţie poate fi descrisă în mai multe feluri. De ex…ce reprezintă un anumit areal, care sunt coordonatele geografice ale unui anumit punct…etc.;

• ***Unde se găseşte……,*** adică exprimarea unei condiţii. Mai precis, în loc să identificăm ce este la o anumită locaţie, dorim să ştim în ce locaţii sunt satisfăcute anumite condiţii;

• ***Ce s-a schimbat la……,*** adică evoluţia. Se determină variaţiile în timp a unui areal, de ex.... variaţiile precipitaţiilor zilnice ce cad pe o anumita suprafaţă în decursul unui an;

• ***Ce se întâmplă dacă……,*** adică modelarea. De exemplu ce impact asupra mediului este determinat de adăugarea unei şosele la reţeaua de drumuri, etc.



Domenii de aplicabilitate GIS

Utilităţi – Gestiunea reţelelor de apă, gaz, electricitate, telecomunicaţii, etc. Acest gen de aplicaţii necesită harţi foarte precise, iar modelele vectoriale domină acest domeniu. Tot aici putem include amplasarea staţiilor de emisie/recepţie din sistemul de telefonie celulară;

Mediu – Într-o primă variantă, produsele GIS sunt folosite pentru inventarierea teritoriilor afectate de poluare (ape, sol, aşezări). La un nivel mai ridicat se pot face studii privitoare la procesele de eroziune, alunecări de teren, studii de impact, studiul calităţii apei, etc.;

Amenajarea teritoriului – GIS pot fi utilizate în monitorizarea terenului, planuri de amenajare urbanistice, comunale, judeţene, regionale. Ca exemplu putem…studiul amplasării unor blocuri de locuinţe (coroborat cu date provenite de la utilităţi; harţi ale conductelor de gaz, apă, informaţii privitoare la dimensionarea acestora…etc.);

Agricultură şi silvicultură – inventarierea solurilor însoţită de datele atribut referitoare la tipul de sol, calitate, utilizare. Monitorizarea terenurilor în vederea obţinerii de producţii maxime. Inventarierea pădurilor, a zonelor protejate, etc. Proiectele GIS din acest domeniu sunt dublate de prelucrarea imaginilor luate prin teledetecţie;

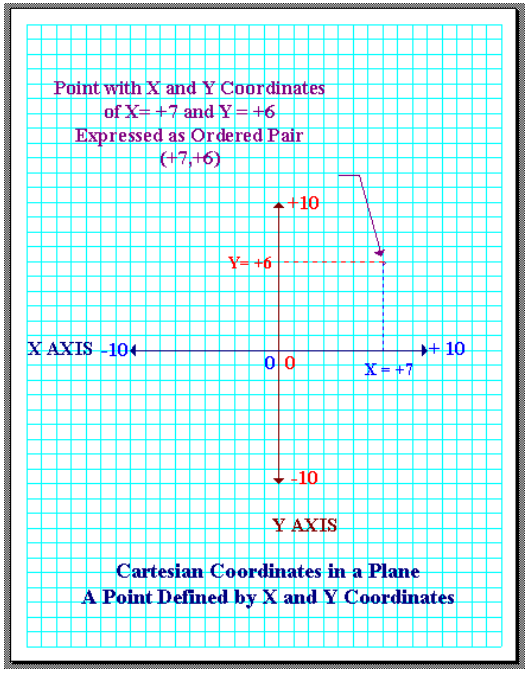
Resurse naturale – în ultimii ani se investeşte tot mai mult în proiecte care conduc la depistarea resurselor naturale (minereuri, petrol, gaz, apă) utilizând produse GIS. Şi această activitate este dublată de prelucrarea imaginilor digitale sau aeriene. Acest domeniu a beneficiat din plin de programele de teledetecţie Skylab din anii 70, când s-au descoperit multe resurse naturale exploatate în momentul de faţă (petrol şi gaz în Marea Nordului, petrol în Marea Neagră, etc.);

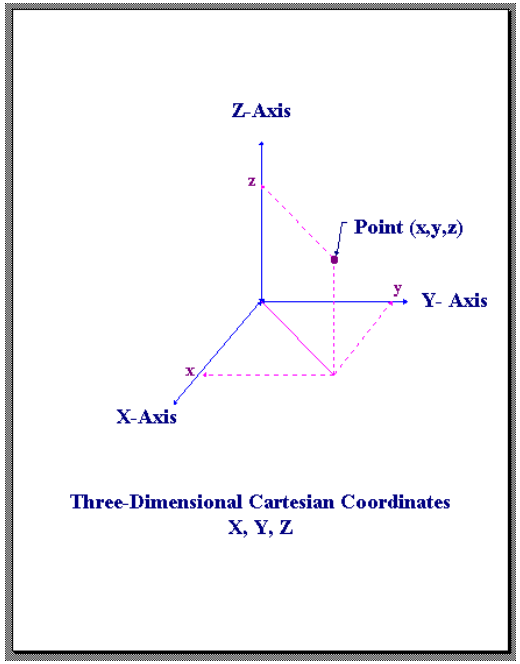
Transport – GIS are un potenţial considerabil în gestiunea şi optimizarea transportului urban sau regional (trasee optime pentru autobuze, tramvaie, trenuri, la care se adaugă determinarea numărului optim de mijloace de transport în diferite intervale de timp). Alegerea traseelor optime pentru maşinile de intervenţie (salvare, pompieri, poliţie). În transportul maritim hărţile electronice le înlocuiesc tot mai des pe cele clasice, iar poziţionarea se face cu instrumente specializate (GPS – sisteme de poziţionare globală), acestea fiind direct legate de hărţile digitale;

Demografie – baze de date referitoare la populaţie (structură pe vârste, religii, profesii, învăţământ, sănătate) asociate cu hărţi administrative la nivel de comună, etc.;

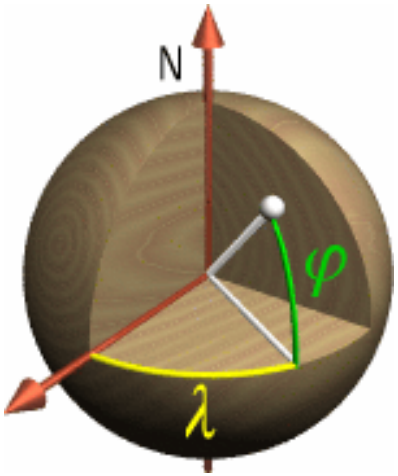
Marketing – Cu o hartă a unui oraş, asociată cu o bază de date ce conţine recensăminte, plus localizările firmelor se pot face corelaţii între clienţi şi ofertanţii de servicii. Se poate merge până la simularea amplasării unui magazin într-o anumită zonă. Rezultă astfel o hartă care prezintă modificarea clientelei magazinelor învecinate…etc.;

Cadastru – inventarierea şi întreţinerea datelor spaţiale şi atribut a tuturor terenurilor...





Coordonate carteziene - în plan sunt proiecţiile razei vectoare pe cele două axe, iar în spaţiu proiecţiile raze vectoare pe trei axe. Cele bidimensionale comportă o origine arbitrară, faţă de care se face referire şi care de obicei este astfel aleasă (în partea stângă, jos), astfel încât să avem numai valori pozitive pt *ox* şi *oy*. În cazul tridimensional originea este considerată în centrul Pământului, axa *oz* fiind confundată cu axa de rotaţie. Coordonatele unui punct sunt exprimate prin proiecţiile sale pe cele trei axe (*x, y, z)*



Coordonatele geografice - se exprimă într-un sistem de referinţă cu originea în centrul Pământului, axa fundamentală fiind axa de rotaţie, iar planul fundamental – planul ecuatorial. Se consideră astfel că meridianul Geenwich este conţinut în planul *xoz* şi ecuatorul este conţinut în planul *xoy.* În cazul coordonatelor geografice, altitudinea nu este exprimată prin lungimea razei vectoare (adică distanţa de la centrul Pământului până la punct)**,** ci este exprimată faţă de o suprafaţă de referinţă – nivelul mării, care implică faptul că Pământul este considerat un geoid. Suprafaţa de referinţă nu este cunoscută exact decât la nivelul mărilor. Sub continente, geoidul este o formă teoretică care este reconstituită prin măsurători.

Coordonatele geografice sunt: *latitudinea* unghiul format de planul ecuatorului terestru cu verticala locului şi *longitudinea* unghiul format de meridianul iniţial cu meridianul locului.

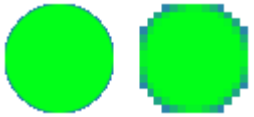
Datele spaţiale

Baza unui GIS o constituie datele spaţiale – reprezentate de harţi sub formă digitale.

Sisteme de reprezentare a datelor spaţiale

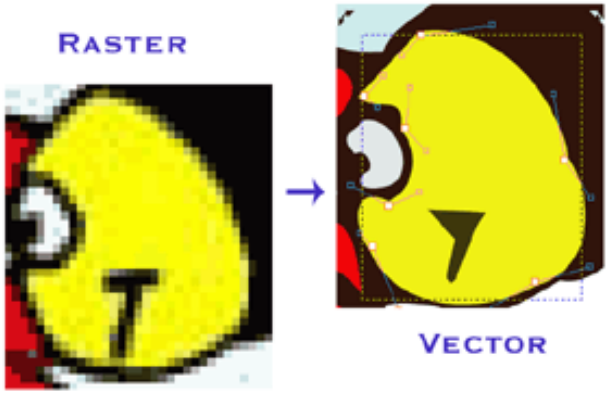
....cum să introducem o hartă în calculator???

Astfel reprezentarea unei harţi se poate face în două feluri: sistemul vector şi sistemul raster.



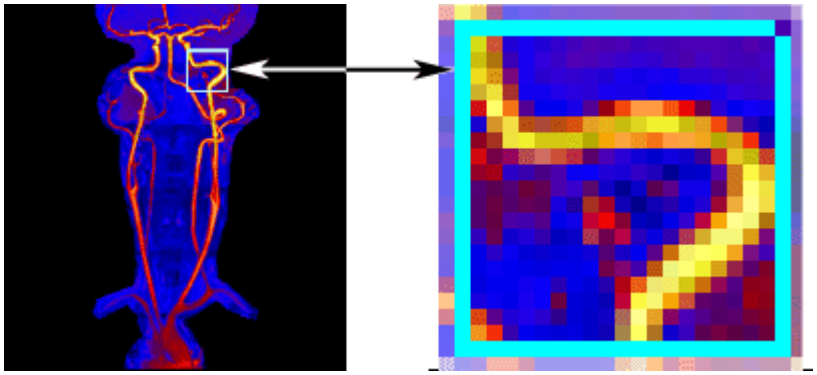
În ***sistemul vector*** harta este constituită, în mare, din puncte şi linii, fiecare punct si extremităţile liniilor fiind definite prin perechi de coordonate (x,z). acestea pot forma *arce*, ***suprafeţe*** sau *volume* (în cazul în care se mai atasază încă o coordonată). Elementele constitutive ale harţii vor fi reprezentate prin astfel de entităţi: o fântână va fi un punct, un râu va fi un arc, un lac va fi un poligon, etc.

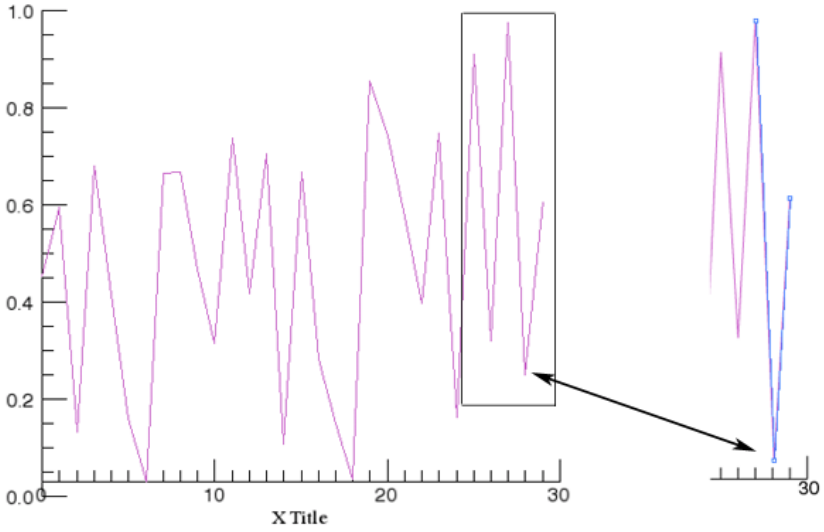
În ***sistemul raster*** imaginile sunt constituite din pixeli. Pixelul sau unitatea de imagine, este cel mai mic element de pe o suprafaţă de afişare căruia i se poate atribui în mod independent o intensitate sau o culoare. Fiecărui pixel i se va atribui un număr care va fi asociat cu o culoare. Entităţile grafice constituie multimi de pixeli. Un drum va fi desemnat de o mulţime de pixeli de aceeaşi valoare.



Între cele două sisteme există diferenţe privitoare la modul de stocare, manipulare şi afişare a datelor.

Ambele sisteme au avantaje şi dezavantaje. Principalul avantaj al sistemului vector constă în memorarea mai eficientă a datelor. În acest sistem toate coordonatele care descriu trăsăturile caracteristice ale imaginii trebuiesc codificate. Acest sistem se utilizează în general la realizarea hărţilor la scară mare. În sistemul raster fiecare pixel din imagine trebuie codificat. Grafica raster este utilizată în general când este nevoie să integram harţi tematice cu imagini luate prin teledetecţie.





Sistemul raster

Acest sistem generează un singur model numit *model raster* (sau model matricial). Acest sistem este compus din mici entităţi (pixeli) având o suprafaţă de regulă determinată de rezoluţia sistemului. Este important de precizat că este vorba de o reprezentare internă a harţii, care poate să coincidă sau nu cu rezoluţia monitorului sau a altor echipamente.

În general sistemul raster este mare consumator de resurse (ex: o imagine format A4, reprezentată cu o rezoluţie a unei imprimante laser este formată din aproximativ 9 milioane de celule.

Modelul raster este simplu, el conţinând două entităţi – celula şi imaginea. Poziţia unei celule este definită printr-un număr de linie şi un număr de coloană într-o imagine şi numai una.

Calitatea imaginilor raster este pusă în valoare atunci când se reprezintă fenomene de mare variabilitate. De ex. altimetria şi batimetria se pretează mai bine la o astfel de reprezentare. Analiza la nivel de celule permite evidenţierea unor proprietăţi importante ale terenului (funcţie şi de scara la care se lucrează).

Sistemul vector

Sistemul vector se bazează pe ***primitive grafice***. Primitiva grafică este cel mai mic element reprezentabil grafic utilizat la crearea şi stocarea unei imagini vectoriale şi recunoscută ca atare de sistem. Sistemul vectorial se bazează pe cinci primitive grafice:

Punctul

Arcul (sau linia care uneşte puncte);

Nodul (punct care marchează capetele unui arc sau care se află la contactul dintre arce);

Poligonul (arie determinată de arce);

Corpul (volum determinat de suprafeţe).

Obiectele cartografice simple sunt alcătuite din primitive iar obiectele geografice mai complexe sunt formate din combinarea obiectelor simple.

Punctul este unitatea elementară în geometrie sau în captarea fotogrametrică. El nu trebuie confundat cu celula din reprezentarea raster, deoarece el nu are suprafaţă şi nici dimensiune. El reprezintă doar o poziţionare într-un spaţiu cu 2 sau 3 dimensiuni. Înregistrarea punctelor în sistemul de calcul se face sub formă numerică, mai exact, fiecare punct va fi înregistrat într-un fişier, sub formă de tabel, ce conţine două coloane (prima coloană conţine un număr de identificare, care este unic, şi a doua coordonatele punctului respectiv în sistemul de referinţă ales).

Arcul este o succesiune de joncţiuni între o succesiune de puncte. Un arc este în general o linie frântă ce uneşte direct două puncte ale parcursului. Astfel, o linie frântă poate aproxima cu suficientă precizie orice curbă, prin micşorarea segmentelor.

Nodul este definit ca o extremitate de arc şi nu trebuie confundat cu un punct. Un arc este mărginit în mod obligatoriu de un punct de origine şi unul de destinaţie.

Poligonul este delimitat de un parcurs de arce, ele însele fiind conectate de noduri. Unui poligon îi este ataşat în mod obligatoriu un nod izolat numit centroid. Acest nod permite construirea suprafeţelor în jurul lui, până la limitele formate de arcele întâlnite.

Volumele există pachete grafice ce au posibilitatea de a lua în considerare, de a calcula şi de a reprezenta volume simple. Ele aproximează cu o precizie suficientă volumele de pe hărţile reprezentate în 3D.

Georeferenţierea

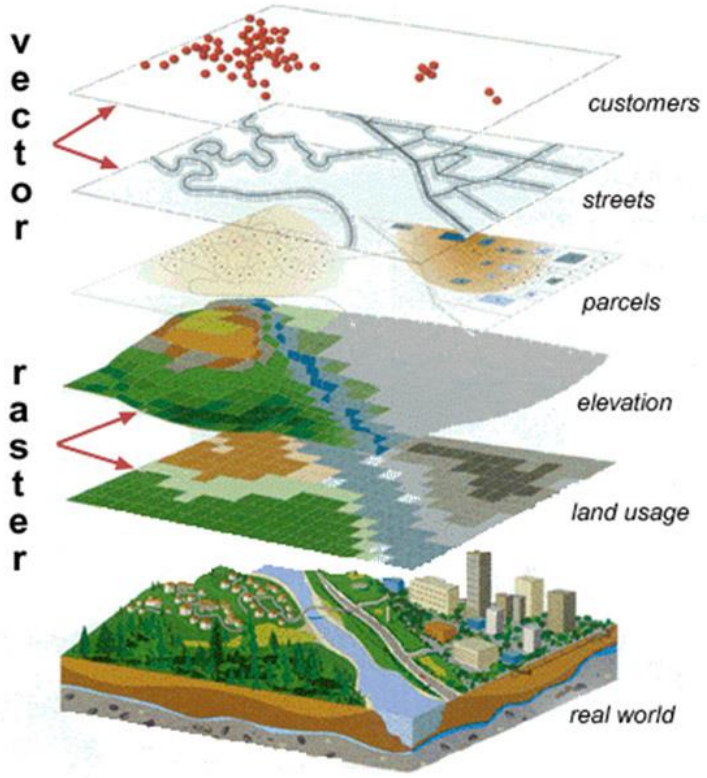
Este procesul prin care harta digitală este asociată cu coordonatele geografice reale. În unele aplicaţii nu este necesară raportarea la coordonatele geografice, fiind suficient un sistem de coordonate carteziene. În cazul harţilor vectoriale, care deja conţin un sistem de coordonate local (cartezian), trecerea la coordonate geografice se face prin transformări de coordonate. Practic georeferenţierea constă în determinarea coordonatelor geografice ale unor puncte cu mare precizie şi localizarea lor pe harta digitală, urmând ca restul punctelor sa fie calculate automat pe baza formulelor de transformare. Acest gen de georeferenţiere se numeşte georeferenţiere continuă.

În cazul raster nu avem nici un sistem de coordonate definit în imagine. Georeferenţierea constă în localizarea cu precizie maximă a unor pixeli dispersaţi pe imagine cărora li se asociază coordonate geografice cunoscute. Deoarece pixelul are dimensiune, lui îi va corespunde o suprafaţă de teren. În consecinţă rezoluţia imaginii are o mare diferenţă în determinarea coordonatelor. Georeferenţierea în sistemul raster se numeşte georeferenţiere discretă.

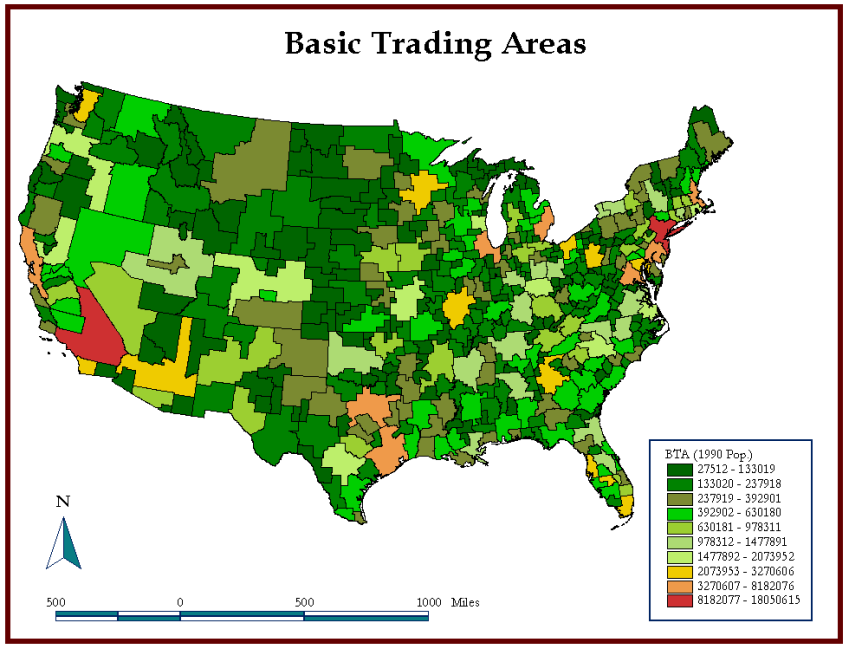
Georeferenţierea ridică probleme în momentul în care apar harţi digitale diseminate, adică provenite din diverse surse şi care trebuie folosite în comun.

Baza de date spaţială – constituie o colecţie de date, referenţiate spaţial, ce reprezintă un model al realităţii (reprezintă diverse variabile selectate sau o aproximare a unui fenomen)

Din punct de vedere al unui proiect GIS – hărţile digitale, implicate în prelucrarea datelor sub GIS reprezintă *baza de date spaţială*.

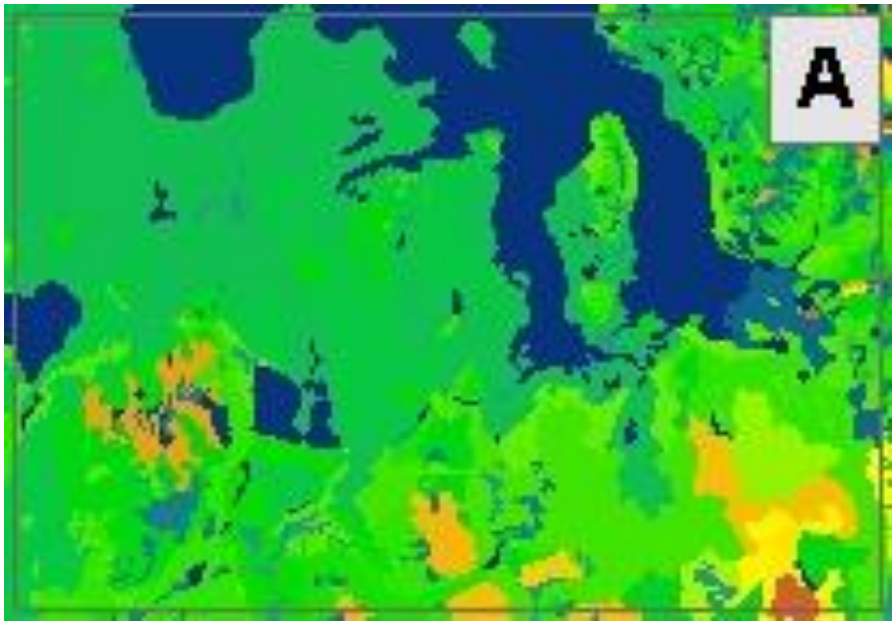


O hartă poate fi descompusă în mai multe straturi (layers) de informaţie şi invers, mai multe straturi pot compune o hartă. Aceste element stă la baza organizării *bazei de date spaţiale*. Straturile pot fi combinate pentru a forma harţi care nu există în mod tradiţional. Este important de ştiut ca entităţile dintr-un strat nu pot fi separate.



Un strat în **sistemul vector** se comportă ca un ansamblu de primitive grafice. Unele softuri nu permit utilizarea într-un strat a anumitor tipuri de primitive (ArcInfo). Se recomandă ca fiecare tip de primitivă (punct, poligon, etc.) să fie pe un strat separat. În funcţie de tema pe care o reprezintă harta, pot exista mai multe straturi care să conţină aceleaşi primitive grafice. Fiecare strat este însoţit de tabela de atribut proprie. Un strat constituie o harta tematică.

În **sistemul raster**, un strat (layer) reprezintă o hartă tematică. Acestea pot fi utilizate împreună cu straturile vector sau separat, funcţie de scopul urmărit.



Exemple se **straturi la scară mică**:

- limitele administrative,
- geologia,
- modul de utilizare al terenului,
- hidrografia,
- reţeaua de drumuri,
- căi ferate.

Exemple se **straturi la scară mare**:

- planul străzilor,
- reţeaua de apă,
- gaz,
- canalizare,
- electricitate.

Unele teme pot constitui o bază de lucru pentru celelalte (planul stradal, harta topo, etc.).

Specificaţiile pentru organizarea BDG se fac în faza de analiză a proiectului GIS, în funcţie de scopul urmărit. Aceasta este o etapă foarte importantă de care depinde practic exploatarea proiectului. Aici se stabileşte conţinutul fiecărui strat şi automat şi tabelele de atribut. Una din cele mai uzuale operaţii este suprapunerea straturilor. Pentru această operaţie este important să se tină seama de scară şi proiecţie.

Date atribut

Datele tabelare care se asociază harţilor digitale pot avea diferite formate (ASCII, dbf sau formate proprii). Tipul de format intern este propriu fiecărui produs GIS, ArcView lucrează cu *dbf*, acest format poate fi creat cu dBase şi Fox iar Excel-ul poate exporta fişierele proprii în format dbf, în anumite condiţii.

În ceea ce priveşte ArcView, este de menţionat că există două categorii de tabele (Atribute Table şi fişierele oarecare în format dbf care pot fi asociate cu acestea cu condiţia să aibă un câmp comun. Atribute Table se creează odată cu fişierul shape şi este legat de acesta. El conţine informaţii minime referitoare la tema respectivă. Pentru o mai uşoară asociere cu celelalte tabele se preferă o structură cat mai simplă pentru Atribut Table. Totalitatea datelor atribut formează ceea ce se numeşte Bază de Date Atribut (BDA).

Geocodificarea. Baza de date geografică.

Asocierea datelor tabelare cu cele spaţiale constituie partea care deosebeşte un produs GIS de produsele soft pentru cartografie digitală, de produsele CAD sau de programele de grafică (Corel Draw, etc). Procesul de asociere a celor două tipuri de date poartă numele de ***geocodificare***. Această operaţiune este diferă funcţie de sistem, la sistemul vector fiecare primitivă grafică este asociată cu un tabel (tabel atribut) ce conţine date alfanumerice referitoare la caracteristicile respectivei primitive grafice (ex. dacă o polilinie sau un arc reprezintă o porţiune de şosea, atunci tabela atribut va trebui să conţină un cod de identificare unic pentru polilinia respectivă, urmat de un minim de caracteristici – denumirea, calitatea, lungimea).

Daca harta este digitizată cu ajutorul altui produs (Surfer, AutoCad), tabela atribut nu există iar aceasta trebuie creată în momentul conversiei în produsul GIS. În acest caz, codurile de identificare ale atributelor trebuie introduse de la tastatură.

În sistemul raster, tabela de atribut va conţine drept cod, numărul asociat pixelului. O particularitate a sistemului raster este aceea că, atributul poate fi conţinut în imagine (de exemplu o harta raster poate să conţină tipul de sol, prin ea însăşi conţine şi atributul.

Este important de reţinut că organizarea internă a tabelei de atribut este de tip *baza de date relaţională*, adică fiecare coloană are un nume şi reprezintă câmpul, iar linia articolele.

Noţiunea de geocodificare (geocoding) poate avea şi alt înţeles. În produsele elaborate de ESRI, prin geocodificare se înţelege identificarea unei locaţii căreia i se atribuie o caracteristică (ex. codul poştal). În schimb ESRI foloseşte pentru geocodificare ceea ce ei numesc *modelul georelaţional* (Georelational Model). Un termen ce poate fi similar geocodificării *adress matching*.

Ansamblul celor două baze de date (spaţială – BDS şi atribut – BDA) formează ***Baza de Date Geografică*** (BDG). Acest termen este consacrat ca atare şi deci nu trebuie folosit cu alte înţelesuri. O dată geografică este un element al BDG şi în consecinţă cuprinde două componente: spaţial (poziţional) şi atribut.

Tehnici de introducere a datelor spaţiale.

În situaţia în care datele sunt disponibile în formă analogică (hărţi pe suport de hârtie sau fotograme) ele trebuie convertite în formă digitală. Această conversie se poate realiza pe două cai: *digitizarea* şi *scanarea*.

Digitizarea

Digitizarea constă în transpunerea datelor grafice din format analogic în format digital. Această operaţie se execută cu ajutorul unui digitizor. Fişierul rezultat în urma digitizării va conţine coordonate carteziene sau coordonate geografice. Aceste doua tipuri de coordonate por fi convertite.

Procesul de digitizare

Înainte de a începe, trebuie să stabilim scopul, să alegem hărţile care deja există pe suport de hârtie şi să definim straturile. Scopul va decide şi alegerea caracteristicilor hărţilor – temele, gradul de detaliere, scara, sistemul de coordonate. După fixarea temei vom decide câte straturi vor fi necesare.

Procesul de digitizare propriu-zis presupune următoarele etape:

Fixarea punctelor de control şi apoi digitizarea lor. După această operaţiune se va afişa o eroare calculată prin metoda celor mai mici pătrate (RMSE – Root Mean Square Error). Dacă aceasta este acceptată, se trece la pasul următor;

- Fixarea dimensiunilor hărţii;
- Digitizarea punctelor;
- Digitizarea arcelor;
- Digitizarea poligoanelor (daca este permis în acel strat);
- Salvarea fişierului.

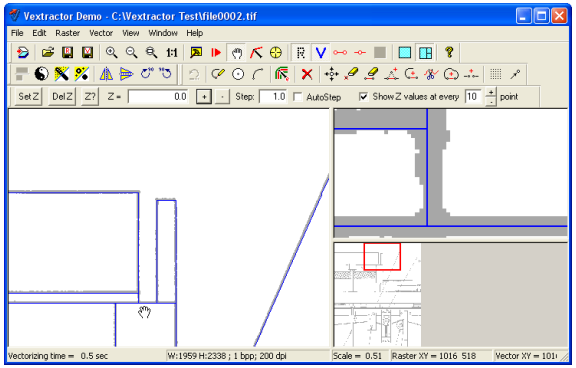
Scanarea

Procesul de scanare constă în conversia datelor din format analogic în format digital. Modul în care se realizează scanarea este următorul: imaginea este împărţită în puncte (matrice de puncte), fiecărui punct i se atribuie un număr în conformitate cu nuanţa de gri sau cu culoarea originalului. Procesul este analog cu fotocopierea. Un *scaner* copiază imaginea şi o stochează într-un fişier raster, care ulterior poate fi prelucrat utilizând un produs pentru procesarea imaginilor, cele mai uzual format este TIFF (Tag Image File Format).

După scanare, urmează faza de editare, în care, după determinarea celor trei categorii de date: *date utile* (puncte, linii, poligoane), *simboluri* (adnotaţii sau semne convenţionale) şi *zgomot*, se procedează la următoarele operaţiuni:

- Îndepărtarea zgomotului;
- Îndepărtarea simbolurilor (daca este necesar);
- Vectorizarea;
- Adăugarea de date suplimentare (dacă este necesar);
- Corecţia erorilor;
- Geocodificarea;
- Crearea topologiei;
- Georeferenţierea.

Dacă fişierul raster compilat va fi o simplă imagine compilată, doar zgomotul trebuie îndepărtat. În cazul în care acesta se doreşte a fi un strat tematic (coverage) trebuie îndepărtate şi simbolurile.



Procesul de vectorizare poate fi rezumat astfel:

Se înregistrează doar pixelii care formează o structură (o linie sau un poligon);

Dintre pixelii răspândiţi de-a lungul unei linii vor fi înregistraţi doar cei din mijloc;

Coordonatele se determina pentru punctul de start şi cel de sfârşit al fiecărui segment.

Surse de date spaţiale

Date preluate prin GPS *(Global Positioning System)*

Poziţionarea cu ajutorul sateliţilor - acest sistem a apărut în 1990 şi permite poziţionarea oricărui punct de pe suprafaţa pământului cu o precizie de câţiva cm, fără utilizarea reperelor geodezice existente.

Datele preluare prin GPS pot fi integrate în GIS. Staţiile performante permit cuplarea la un laptop care înregistrează coordonatele punctelor. Există programe specializare care permit conversia acestor date direct în produsele GIS mai cunoscute.

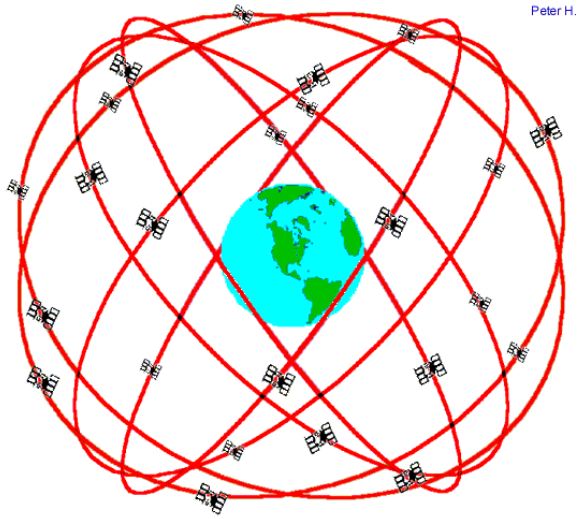
Datele CAD

Cel mai cunoscut produs din această categorie este AutoCAD. Cel mai răspandit format este *.dxf (Drawnig eXchange Format) el fiind creat şi citit de cele mai multe produse CAD si GIS.

Surse de date

Achiziţia de date reprezintă primul pas în implementarea unui GIS. Costurile legate de achiziţia datelor se pot ridica şi la 80% din costurile asociate unui proiect GIS.

Există o mare varietate de surse de date colectate de-a lungul timpului, privitoare al Europa. Unul din cele mai cunoscute programe ale Comunităţii Europene este programul CORINE care a avut drept rezultat înregistrarea întregii suprafeţe a Europei prin teledetecţie apelând la serviciile programelor spaţiale Landsat si SPOT. De asemenea compania ESRI (producătoarea ArcView GIS) pune la dispoziţie unele harţi digitale în mod gratuit (www.esri.com).



GPS Nominal Constellation
24 Satellites in 6 Orbital Planes
4 Satellites in each Plane
20,200 km Altitudes, 55 Degree Inclination

Analiza spaţială

În general posibilităţile de *analiză spaţială* sunt limitate în comparaţie cu cele de *stocare*, *manipulare* şi *vizualizare*. Este totuşi puţin probabil ca un produs GIS să satisfacă cerinţele unui utilizator în ceea ce priveşte analiza şi modelarea. Marea dificultate a firmelor producătoare este lipsa unui consens atât în ceea ce priveşte algoritmii implementaţi în analiza spaţială cât şi în definirea noţiunilor.

Esenţa analizei spaţiale este de a extrage datele cu semnificaţie din datele distribuite spaţial, care eventual au fost supuse unor prelucrări. Analiza spaţială este procesul de obţinere de asociaţii şi de tipare pe hartă în vederea caracterizării, previziunii sau înţelegerii fenomenelor spaţiale.

Modelarea prin GIS se realizează cu ajutorul procedurilor conţinute în cadrul analizei spaţiale, care conduce în final la o previziune în vederea luării deciziilor. Operaţiile implicate în modelare pot fi efectuate direct pe straturile de harţi şi în combinaţie cu atributele aferente. Datorită complexităţii proceselor de modelare, precum şi a particularităţilor domeniului modelat, nu există până în prezent un limbaj universal pentru modelarea în GIS.

Din punct de vedere al produselor GIS, analiza spaţială reprezintă mai mult un concept decât o componentă program a unui GIS. Operaţiile analizei spaţiale sunt conţinute în diverse module ale GIS (ex. ArcView deţine o extensie – care se livrează separat – *Spatial Analyst*.

În linii mari analiza spaţială trebuie să permită: **interogarea spaţială**, **interogarea atribut** şi generarea unui set de date noi din cele existente.

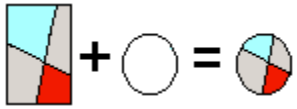
Această ultimă cerinţă este cea mai reprezentativă pentru analiza spaţială şi constituie esenţa prelucrării datelor geografice.

Analiza spaţială cuprinde o mare varietate de operaţii care se constituie în grupuri, subgrupuri şi operaţii elementare. O problemă dată se compune dintr-un şir de operaţii elementare, într-o ordine prestabilită, în care sunt implicate datele spaţiale, datele atribut şi proceduri (programe), care le prelucrează.

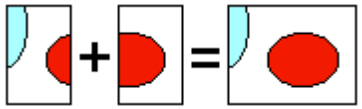
Operaţiile analizei spaţiale sunt numeroase, în principal au fost definite şapte clase de operaţii:

- Operaţii pe un singur strat;
- Operaţii pe straturi multiple;
- Analiză statistică;
- Analiza reţelelor;
- Analiza suprafeţelor;
- Analiză grid.

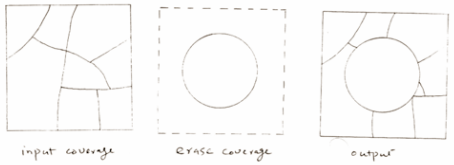
Modelarea spaţială complexă înglobează proceduri din toate grupurile de operaţii amintite.



CLIP



MAPJOIN



ERASE

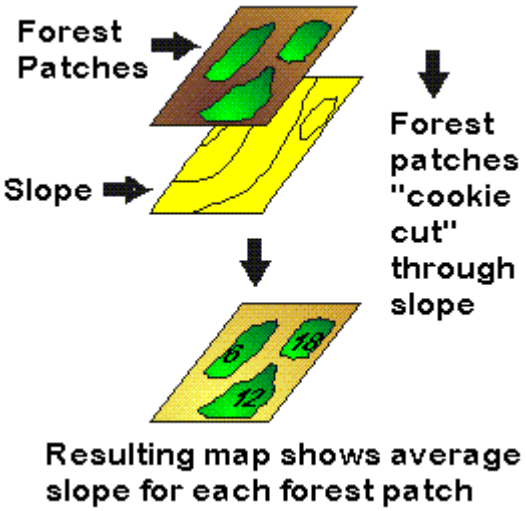
SPLIT

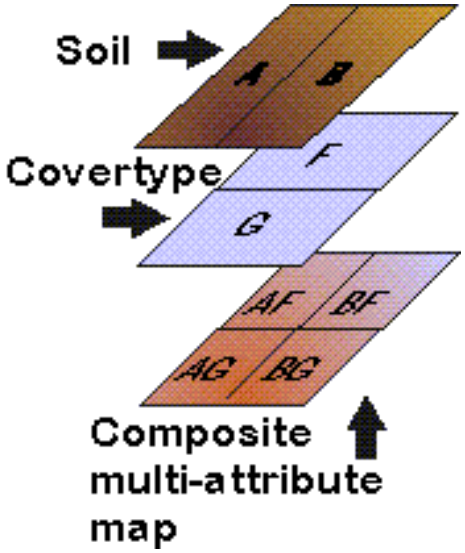
Operaţii pe un singur strat

Se mai numesc operaţii pe orizontală, ele constituie instrumentele de bază ale analizei spaţiale. În cazul harţilor vectoriale este necesar ca hărţile să conţină primitive grafice de acelaşi fel. În cazul în care un strat conţine mai multe tipuri de primitive grafice (puncte, arce, poligoane) se apelează la grupul de operaţii pe mai multe straturi. Există trei categorii mai importante de operaţii: *manipularea primitivelor grafice, selecţia lor* şi *clasificarea*.

Prima categorie conţine operaţii asupra contururilor şi analiza de proximitate. A doua categorie de operaţii se referă la identificarea de primitive grafice pe baza unor expresii logice. A treia categorie are în componenţă operaţii care permit gruparea primitivelor grafice în clase, în vederea unei analize statistice. Dintre operaţiile elementare care se fac asupra contururilor amintim: selectarea unui porţiuni dintr-un strat (**CLIP**), îndepărtarea unor primitive grafice (**ERASE**), crearea unor subdiviziuni (**SPLIT**), asamblarea a două sau patru hărţi adiacente (**MAPJOIN**), îndepărtarea limitelor care separă poligoanele de acelaşi tip (**DISOLVE**) şi eliminarea unor linii care au fost introduse în mod eronat (**ELIMINATE**). Analiza de proximitate implică determinarea unor contururi de distanţă egală la o anumită primitivă grafică. În mod uzual operaţia poartă numele de BUFFER. Există situaţii când un buffer are dimensiune variabilă (ex. un studiu de poluare a solului – extinderea gradului de contaminare este dependentă de concentraţia poluantului a cărei valoare este conţinută în tabela atribut).

Clasificarea este operaţia prin care datele spaţiale sunt simbolizate în conformitate cu atributele asociate lor. Operaţia poate fi făcută asupra oricărei primitive grafice. Clasele pot fi realizate automat sau fixate de către utilizator.





Operaţii pe straturi multiple

Operaţiile pe straturi multiple impun ca toate hărţile implicate în prelucrări să aibă acelaşi sistem de coordonate şi aceeaşi scară. Acest gen de operaţii sunt cunoscute sub numele de operaţii pe verticală şi se bazează pe relaţiile existente între datele aflate pe straturi diferite. Cu ajutorul acestor operaţii un strat complex poate fi descompus în straturi tematice şi invers, mai multe straturi pot fi combinate pentru a rezulta un strat complex.

Operaţiile pe straturi multiple pot fi împărţite în trei categorii: operaţii de tip ***overlay***, ***analiza de proximitate*** şi ***analiza corelaţiilor spaţiale.***

Analiza overlay creează combinaţii între primitivele grafice aflate pe straturi diferite. Obiectivul principal al acestui tip de analiză este acela de a construi conexiuni între date aparţinând unor straturi diferite pentru a stabili relaţii între entităţi geografice.

Analiza corelaţiilor spaţiale are în componenţă operaţii statistice. Obiectul principal al acestui modul este de a găsi relaţii între diferitele tipuri de date spaţiale distribuite pe mai multe straturi.

Eficienţa aplicării metodei corelaţiei depinde de enunţarea corectă a problemei studiate precum şi de aplicarea corectă a instrumentelor statistice. Caracterul complex al dependenţei statistice pune pe primul plan problema identificării existenţei legăturilor. Calculul indicatorilor de corelaţie este admis cu condiţia stabilirii unei legături cauzale reale între fenomenele cercetate.

O problemă importantă pentru cercetarea corelaţiilor este stabilirea *funcţiei de regresie*, care să exprime relaţia cantitativă dintre fenomenul efect şi fenomenul cauză. Forma legăturii dintre fenomene şi descrierea printr-o ecuaţie analitică, pe baza cărora se stabilesc valorile fenomenului efect, în funcţie numai de factorul (sau factorii) luaţi în considerare. De buna alegere a funcţiei de regresie depind rezultatele analizei de corelaţie.

Analiza statistică

Acest grup de operaţii se execută pe teme punct. Metode statistice se aplică în mod selectiv şi în celelalte tipuri de operaţii însă aici ele sunt implicate în mod exclusiv.

Distribuţia punctelor în plan este caracterizată de frecvenţă, densitate, centru de greutate, dispersie şi aranjare spaţială. În general, datele statistice, în forma lor brută, reprezintă o mulţime dezorganizată de puncte. Prima problemă este aceea de a realiza o grupare a datelor astfel încât să se înlesnească combinarea cu alte date de acelaşi gen şi să se poată face ulterior o analiză.

Analiza reţelelor

Analiza spaţială pe structuri liniare se numeşte analiza reţelelor. Domenii de utilizare: drumuri, căi ferate, telefonie, reţele apă, canalizare, etc.

O reţea constă dintr-un număr de arce interconectate. Fiecare arc are un nod de start şi unul final, iar parcursul său fiind direcţionat de puncte. Relaţiile topologice definite de noduri determină conectivitatea reţelei. Noţiunile de bază în analiza de reţele sunt legate de *teoria grafurilor*.

Analiza suprafeţelor

Analiza suprafeţelor reprezintă distribuţia unei variabile care reprezintă a treia dimensiune a datelor spaţiale. Atunci când vorbim de reprezentarea tridimensională subînţelegem implicarea elevaţiei. La modul general însă, a 3-a dimensiune poate fi reprezentată de orice parametru (funcţie de datele specifice). În mod convenţional, mărimea care reprezintă a treia dimensiune este notată cu z, ea fiind în general o funcţie de două variabile – z = f(x,y). Prin reprezentarea tridimensionala este necesară evaluarea parametrului z în orice punct. Tipuri de reprezentări 2D ale datelor – puncte de elevaţie dispuse neregulat, curbe de nivel, structura latice (asemănătoare *grid*). Reprezentări 3D – bloc diagram, structură TIN (Triangulated Iregular Network) – reţea de triunghiuri dispuse neregulat, bazată pe puncte de elevaţie cunoscute. Panta terenului este considerată constantă în interiorul fiecărui triunghi. Dimensiunile triunghiurilor sunt variabile în funcţie de particularităţile morfologie ale terenului. Acest sistem de reprezentare, chiar daca presupune o oarecare aproximare prezintă o serie de avantaje în analiza ulterioară.

Analiza grid

Modelul raster se bazează pe celule de arie egală (pixeli) cărora le sunt atribuite numere, acestea la rândul lor sunt asociate cu culori, în vederea reprezentării pe un terminal (monitor sau imprimantă). Operaţiile suportate de acest model sunt foarte diferite de cele corespunzătoare modelelor vectoriale, deoarece structura este total diferită. În acest caz nu putem distinge puncte, linii sau poligoane, elementele ieşind în evidenţă prin valori atribuite pixelilor.

Un strat raster poate fi asociat cu o tabelă atribut sau nu. Daca nu are asociat un atribut, acesta se recunoaşte după tema hărţii.

O hartă raster este de fapt o matrice de numere. Operaţiile analizei grid se referă la operaţii pe aceste numere.

În cazul analizei grid putem include următorul grup de operaţii: operaţii pe un singur strat, operaţii pe straturi multiple, analiza suprafeţelor şi analiză statistică.

Operaţii pe un singur strat – acest gen de operaţii se referă la modificarea valorilor celulelor prin înmulţirea sau împărţirea lor cu un număr (acest lucru corespunde înmulţirii unei matrice cu un scalar).

Operaţii pe straturi multiple – în general acest tip de operaţii implică două imagini în intrare şi una în ieşire. Prin repetarea acestei operaţii pot fi implicate mai multe straturi. Straturile trebuie să aibă acelaşi număr de linii şi de coloane (putem menţiona – adunarea, scăderea înmulţirea şi respectiv împărţirea a două imagini).

Operaţii statistice – structurile raster se pretează la o largă varietate de operaţii statistice.

Procedurile statistice au în componenţă una sau mai multe operaţii elementare şi sunt apelabile prin comenzi, care au o anumită sintaxă. Acestea au un nume care sugerează tipul de operaţie pe care-l conţine.

Studiu de caz – Evaluarea vulnerabilităţii la poluare a acviferului freatic din zona industrială Pantelimon, prin metoda DRASTIC

Vulnerabilitate la poluare a apelor subterane - include totalitatea factorilor naturali de risc ce pot determina fenomene de alterare a calităţii apelor subterane.

Vulnerabilitatea la poluare este condiţionată de următorii factori:

✓ ***Grosimea zonei vadoase a acviferului***

În această zonă mişcarea apei se face pe verticală până la atingerea suprafeţei piezometrice. Cu cât grosimea este mai mare, cu atât riscul ajungerii poluantului în apa subterană este mai mic.

✓ ***Puterea epuratoare a rocii care formează acviferul***

Este o caracteristică a acviferelor cu porozitate interstiţială şi variază în funcţie de granulometrie şi alcătuirea petrografică, fiind foarte mică sau nulă în acviferele fisurale

✓ ***Viteza de curgere a apei subterane***

Este factorul cel mai important care influenţează fenomenele de diluţie, dispersie şi absorbţie. Viteza este mică în acviferele omogene şi mare în cele fisurale şi carstice. În aceste condiţii, evoluţia extinderii unei zone poluate este mult mai rapidă în mediile fisurale comparativ cu cele granulare.

✓ ***Protecţia naturală a acviferului***

Prezenţa în acoperişul acviferului a unui strat impermeabil continuu asigură o protecţie naturală eficientă împotriva pătrunderii poluanţilor. Este evident că acviferele cele mai vulnerabile sunt cele freatice, situate în adâncimi mici, sub influenţa directă a factorilor atmosferici şi cele carstice.

DRASTIC *este un model empiric, dezvoltat de EPA în anii `80 în scopul evaluării potenţialului de poluare al acviferelor (Aller et. al. 1987). Acest model include o serie de parametrii hidrogeologici variaţi, ai căror caracteristici fizice pot afecta calitatea apelor subterane.*

Aceşti parametrii sunt:
- Adâncimea nivelului hidrostatic (**D**epth to water table);
- Alimentarea acviferului (**R**echarge);
- Caracteristicile mediului acvifer (**A**quifer Media);
- Caracteristicile solului (**S**oil Media);
- Topografia zonei – panta (**T**opography);
- Caracteristicile zonei vadoase (**I**mpact of the Vadose Zone);
- Conductivitatea hidraulică (Hydraulic **C**onductivity).

Fiecărui parametru i se atribuie o valoare între 1 şi 10, funcţie de condiţiile locale. Valorile mari corespund unei vulnerabilităţi ridicate. După stabilirea acestor valori pentru fiecare parametru, ele se multiplică cu un factor (între 1 şi 5) dependent de importanţa acelui parametru faţă de vulnerabilitatea acviferului. Prin însumarea valorilor obţinute se obţine un *indice* a cărei expresie este:

Indice DRASTIC = DrDw + RrRw + ArAw + SrSw + TrTw + IrIw + CrCw, unde:

Dr = Punctaj acordat funcţie de adâncimea nivelului hidrostatic;
Dw = Factor de multiplicare acordat adâncimii nivelului hidrostatic;
Rr = Punctaj acordat funcţie de alimentarea acviferului;
Rw = Factor de multiplicare acordat alimentării acviferului;
Ar = Punctaj acordat funcţie de caracteristicile mediului acvifer;
Aw = Factor de multiplicare acordat caracteristicilor mediului acvifer;
Sr = Punctaj acordat funcţie de caracteristicile solului;
Sw = Factor de multiplicare acordat caracteristicilor solului;
Tr = Punctaj acordat funcţie de topografie (pantă);
Tw = Factor de multiplicare acordat topografiei;
Ir = Punctaj acordat funcţie de caracteristicile zonei vadoase;
Iw = Factor de multiplicare acordat caracteristicilor zonei vadoase;
Cr = Punctaj acordat funcţie de conductivitatea hidraulică
Cw = Factor de multiplicare acordat conductivităţii hidraulice.

Grade de vulnerabilitate funcţie de valorile indicelui DRASTIC

	Grade de vulnerabilitate a acviferului			
	Scăzută	Moderată	Ridicată	Foarte ridicată
Indice DRASTIC	1 – 100	101 – 140	141 – 200	> 200

Particularizând aplicarea acestei metode pentru cazul acviferului studiat (acviferul freatic din zona Pantelimon – Acumulatorul) rezultă următoarea evaluare:

Layere GIS distincte, din integrarea cărora, condiţionată de principiile metodei DRASTIC, rezultă *harta vulnerabilităţii la poluare a zonei*

Parametrul evaluat	Coeficientul obţinut	Factor de multiplicare specific parametrului	Valoare finală
Adâncimea NH (5 – 15 m)	9	5	45
Alimentarea (50 – 100 mm/an)	3	4	12
Mediul acvifer (nisip, pietriş)	8	3	24
Sol (Praf argilos)	4	2	8
Pantă (0 – 2%)	10	1	10
Conductivitate hidraulică (50 – 100 m/zi)	8	3	24
Zona vadoasă (prafuri, argile)	3	5	15
Valoare indice DRASTIC			**138**

Layere GIS – Hărţi tematice - hidrogeologică, geologică, pedologică, geomorfologică şi climatică – ce constituie baza întocmirii hărţii vulnerabilitate la poluare a acviferului freatic.

Valoarea obţinută pentru acest indice, încadrează acviferul în categoria de acvifere freatice cu vulnerabilitate moderată (101 - 140)

Geographical Information Systems and Science

2nd Edition

Paul A. Longley *University College London, UK*

Michael F. Goodchild *University of California, Santa Barbara, USA*

David J. Maguire *ESRI Inc., Redlands, USA*

David W. Rhind *City University, London, UK*

John Wiley & Sons, Ltd

Geographical Information Systems and Science

Geographical Information Systems and Science

2nd Edition

Paul A. Longley *University College London, UK*

Michael F. Goodchild *University of California, Santa Barbara, USA*

David J. Maguire *ESRI Inc., Redlands, USA*

David W. Rhind *City University, London, UK*

John Wiley & Sons, Ltd

Copyright © 2005

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

ESRI Press logo is the trademark of ESRI and is used herein under licence.

Main cover image and first box from bottom, courtesy of NASA.
Second box, reproduced from Ordnance Survey.
Third box, reproduced by permission of National Geographic Maps.
Fourth box, reproduced from Ordnance Survey, courtesy @Last.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-87000-1 (HB)
ISBN 0-470-87001-X (PB)

Typeset in 9/10.5pt Times by Laserwords Private Limited, Chennai, India
Printed and bound in Spain by Grafos S.A., Barcelona, Spain
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

Foreword ix
Addendum x
Preface xi
List of Acronyms and Abbreviations xv

I *Introduction* 1

1 *Systems, science, and study* 3

1.1 Introduction: why does GIS matter? 4
1.2 Data, information, evidence, knowledge, wisdom 11
1.3 The science of problem solving 13
1.4 The technology of problem solving 16
1.5 The business of GIS 24
1.6 GISystems, GIScience, and GIStudies 28
1.7 GIS and geography 31
Questions for further study 33
Further reading 33

2 *A gallery of applications* 35

2.1 Introduction 36
2.2 Science, geography, and applications 39
2.3 Representative application areas and their foundations 41
2.4 Concluding comments 60
Questions for further study 60
Further reading 60

II *Principles* 61

3 *Representing geography* 63

3.1 Introduction 64
3.2 Digital representation 65
3.3 Representation for what and for whom? 67
3.4 The fundamental problem 68
3.5 Discrete objects and continuous fields 70
3.6 Rasters and vectors 74
3.7 The paper map 76
3.8 Generalization 80
3.9 Conclusion 82
Questions for further study 83
Further reading 83

4 *The nature of geographic data* 85

4.1 Introduction 86
4.2 The fundamental problem revisited 86
4.3 Spatial autocorrelation and scale 87
4.4 Spatial sampling 90
4.5 Distance decay 93
4.6 Measuring distance effects as spatial autocorrelation 95
4.7 Establishing dependence in space 101
4.8 Taming geographic monsters 104
4.9 Induction and deduction and how it all comes together 106
Questions for further study 107
Further reading 107

5 *Georeferencing* 109

5.1 Introduction 110
5.2 Placenames 112
5.3 Postal addresses and postal codes 113
5.4 Linear referencing systems 114
5.5 Cadasters and the US Public Land Survey System 114
5.6 Measuring the Earth: latitude and longitude 115
5.7 Projections and coordinates 117
5.8 Measuring latitude, longitude, and elevation: GPS 122
5.9 Converting georeferences 123

5.10 Summary 125
Questions for further study 126
Further reading 126

6 Uncertainty 127

6.1 Introduction 128
6.2 U1: Uncertainty in the conception of geographic phenomena 129
6.3 U2: Further uncertainty in the measurement and representation of geographic phenomena 136
6.4 U3: Further uncertainty in the analysis of geographic phenomena 144
6.5 Consolidation 152
Questions for further study 153
Further reading 153

III Techniques 155

7 GIS Software 157

7.1 Introduction 158
7.2 The evolution of GIS software 158
7.3 Architecture of GIS software 159
7.4 Building GIS software systems 165
7.5 GIS software vendors 165
7.6 Types of GIS software systems 167
7.7 GIS software usage 174
7.8 Conclusion 174
Questions for further study 174
Further reading 175

8 Geographic data modeling 177

8.1 Introduction 178
8.2 GIS data models 179
8.3 Example of a water-facility object data model 192
8.4 Geographic data modeling in practice 195
Questions for further study 196
Further reading 197

9 GIS data collection 199

9.1 Introduction 200
9.2 Primary geographic data capture 201
9.3 Secondary geographic data capture 205
9.4 Obtaining data from external sources (data transfer) 211
9.5 Capturing attribute data 215
9.6 Managing a data collection project 215
Questions for further study 216
Further reading 216

10 Creating and maintaining geographic databases 217

10.1 Introduction 218
10.2 Database management systems 218
10.3 Storing data in DBMS tables 222
10.4 SQL 225
10.5 Geographic database types and functions 226
10.6 Geographic database design 227
10.7 Structuring geographic information 229
10.8 Editing and data maintenance 235
10.9 Multi-user editing of continuous databases 235
10.10 Conclusion 237
Questions for further study 238
Further reading 239

11 Distributed GIS 241

11.1 Introduction 242
11.2 Distributing the data 244
11.3 The mobile user 250
11.4 Distributing the software: GIServices 257
11.5 Prospects 259
Questions for further study 259
Further reading 259

IV Analysis 261

12 Cartography and map production 263

12.1 Introduction 264
12.2 Maps and cartography 267

12.3 Principles of map design 270
12.4 Map series 281
12.5 Applications 284
12.6 Conclusions 287
Questions for further study 287
Further reading 287

13 *Geovisualization* 289

13.1 Introduction: uses, users, messages, and media 290
13.2 Geovisualization and spatial query 293
13.3 Geovisualization and transformation 297
13.4 Immersive interaction and PPGIS 302
13.5 Consolidation 309
Questions for further study 312
Further reading 313

14 *Query, measurement, and transformation* 315

14.1 Introduction: what is spatial analysis? 316
14.2 Queries 320
14.3 Measurements 323
14.4 Transformations 329
14.5 Conclusion 339
Questions for further study 339
Further reading 339

15 *Descriptive summary, design, and inference* 341

15.1 More spatial analysis 342
15.2 Descriptive summaries 343
15.3 Optimization 352
15.4 Hypothesis testing 359
15.5 Conclusion 361
Questions for further study 362
Further reading 362

16 *Spatial modeling with GIS* 363

16.1 Introduction 364
16.2 Types of model 369
16.3 Technology for modeling 376
16.4 Multicriteria methods 378
16.5 Accuracy and validity: testing the model 379
16.6 Conclusion 381
Questions for further study 381
Further reading 382

V *Management and Policy* 383

17 *Managing GIS* 385

17.1 The big picture 386
17.2 The process of developing a sustainable GIS 390
17.3 Sustaining a GIS – the people and their competences 399
17.4 Conclusions 401
Questions for further study 402
Further reading 403

18 *GIS and management, the Knowledge Economy, and information* 405

18.1 Are we all in 'managed businesses' now? 406
18.2 Management is central to the successful use of GIS 408
18.3 The Knowledge Economy, knowledge management, and GIS 413
18.4 Information, the currency of the Knowledge Economy 415
18.5 GIS as a business and as a business stimulant 422
18.6 Discussion 424
Questions for further study 424
Further reading 424

19 *Exploiting GIS assets and navigating constraints* 425

19.1 GIS and the law 426
19.2 GIS people and their skills 431
19.3 Availability of 'core' geographic information 434
19.4 Navigating the constraints 440
19.5 Conclusions 444

Questions for further study 445
Further reading 445

20 *GIS partnerships* 447

20.1 Introduction 448
20.2 Collaborations at the local level 448
20.3 Working together at the national level 450
20.4 Multi-national collaborations 458
20.5 Nationalism, globalization, politics, and GIS 459
20.6 Extreme events can change everything 464
20.7 Conclusions 470
Questions for further study 470
Further reading 470

21 *Epilog* 471

21.1 Introduction 472
21.2 A consolidation of some recurring themes 472
21.3 Ten 'grand challenges' for GIS 478
21.4 Conclusions 485
Questions for further study 485
Further reading 486

Index 487

Foreword

At the time of writing, the first edition of *Geographic Information Systems and Science* (GIS&S) has sold well over 25 000 copies – the most, it seems, of any GIS textbook. Its novel structure, content, and 'look and feel' expanded the very idea of what a GIS *is*, what it involves, and its pervasive importance. In so doing, the book introduced thousands of readers to the field in which we have spent much of our working lifetimes. Being human, we take pleasure in that achievement – but it is not enough. Convinced as we are of the benefits of thinking and acting geographically, we are determined to enthuse and involve many more people. This and the high rate of change in GIS&S (Geographic Information Systems *and* Science) demands a new edition that benefits from the feedback we have received on the first one.

Setting aside the (important) updates, the major changes reflect our changing world. The use of GIS was pioneered in the USA, Canada, various countries in Europe, and Australia. But it is expanding rapidly – and in innovative ways – in South East Asia, Latin America and Eastern Europe, for example. We have recognized this by broadening our geography of examples. The world of 2005 is not the same as that prior to 11 September 2001. Almost all countries are now engaged in seeking to protect their citizens against the threat of terrorism. Whilst we do not seek to exaggerate the contribution of GIS, there are many ways in which these systems and our geographic knowledge can help in this, the first duty of a national government. Finally, the sheen has come off much information technology and information systems: they have become consumer goods, ubiquitous in the market place. Increasingly they are recognized as a necessary underpinning of government and commerce – but one where real advantage is conferred by their ease of use and low price, rather than the introduction of exotic new functions. As we demonstrate in this book, GIS&S was never simply hardware and software. It has also always been about people and, in preparing this second edition, we have taken the decision to present an entirely new set of current GIS protagonists. This has inevitably meant that all boxes from the first edition pertaining to living individuals have been removed in order to create space: we hope that the individuals concerned will understand, and we congratulate them on their longevity! This second edition, then, remains about hardware, software, people – and also about geographic information, some real science, a clutch of partnerships, and much judgment. Yet we recognize the progressive 'consumerization' of our basic tool set and welcome it, for it means more can be done for greater numbers of beneficiaries for less money. Our new book reflects the continuing shift from tools to understanding and coping with the fact that, in the real world, 'everything is connected to everything else'!

We asked Joe Lobley, an individual unfamiliar with political correctness and with a healthy scepticism about the utterances of GIS gurus, to write the foreword for the first edition. To our delight, he is now cited in various academic papers and reviews as a stimulating, fresh, and lateral thinker. Sadly, at the time of going to press, Joe had not responded to our invitation to repeat his feat. He was last heard of on location as a GIS consultant in Afghanistan. So this Foreword is somewhat less explosive than last time. We hope the book is no less valuable.

Paul A. Longley
Michael F. Goodchild
David J. Maguire
David W. Rhind
October 2004

Addendum

Hi again! Greetings from Afghanistan, where I am temporarily resident in the sort of hotel that offers direct access to GPS satellite signals through the less continuous parts of its roof structure. Global communications mean I can stay in touch with the GIS world from almost anywhere. Did you know that when Abraham Lincoln was assassinated it took 16 days for the news to reach Britain? But when William McKinley was assassinated only 36 years later the telegraph (the first Internet) ensured it took only seconds for the news to reach Old and New Europe. Now I can pull down maps and images of almost anything I want, almost anywhere. Of course, I get lots of crap as well – the curse of the age – and some of the information is rubbish. What does Kabul's premier location prospector want with botox? But technology makes good (and bad) information available, often without payment (which I like), to all those with telecoms and access to a computer. Sure, I know that's still a small fraction of mankind but boy is that fraction growing daily. It's helped of course by the drop in price of hardware and even software: GIS tools are increasingly becoming like washing machines – manufactured in bulk and sold on price though there is a lot more to getting success than buying the cheapest.

I've spent lots of time in Asia since last we communicated and believe me there are some smart things going on there with IS and GIS. Fuelled by opportunism (and possibly a little beer) the guys writing this book have seen the way the wind is blowing and made a good stab at representing the *whole* world of GIS. So what else is new in this revised edition of what they keep telling me is the world's best-selling GIS textbook? I like the way homeland security issues are built in. All of us have to live with terrorist threats these days and GIS can help as a data and intelligence integrator. I like the revised structure, the continuing emphasis on business benefit and institutions and the new set of role models they have chosen (though 'new' is scarcely the word I would have used for Roger Tomlinson. . .). I like the same old unstuffy ways these guys write in proper American English, mostly avoiding jargon.

On the down-side, I still think they live in a rose-tinted world where they believe government and academia actually do useful things. If you share their strange views, tell me what the great National Spatial Data Infrastructure movement has *really* achieved worldwide except hype and numerous meetings in nice places? Wise up guys! You don't have to pretend. Now I do like the way that the guys recognize that places are unique (boy, my hotel is. . .), but don't swallow the line that digital representations of space are any less valid, ethical or usable than digital measures of time or sound. Boast a little more, and, while you're at it, say less about 'the' digital divide and more about digital differentiation. And keep well clear of patronizing, social theory stroking, box-ticking, self-congratulatory claptrap. The future of that is just people with spectacles who write books in garden sheds. Trade up from the caves of the pre-digital era and educate the wannabes that progress can be a good thing. And wise up that the real benefits of GIS do not depend on talking shops or gravy trains. What makes GIS unstoppable is what we can do with the tools, with decent data and with our native wit and training to make the world a better and more efficient place. Business and markets (mostly) will do that for you!

Joe Lobley

Preface

The field of geographic information systems (GIS) is concerned with the description, explanation, and prediction of patterns and processes at geographic scales. GIS is a science, a technology, a discipline, and an applied problem solving methodology. There are perhaps 50 other books on GIS now on the world market. We believe that this one has become one of the fastest selling and most used because we see GIS as providing a gateway to *science* and problem solving (geographic information systems 'and science' in general), and because we relate available software for handling *geographic information* to the scientific principles that should govern its use (geographic information: 'systems and science'). GIS is of enduring importance because of its central co-ordinating *principles*, the specialist *techniques* that have been developed to handle spatial data, the special *analysis* methods that are key to spatial data, and because of the particular *management* issues presented by geographic information (GI) handling. Each section of this book investigates the unique, complex, and difficult problems that are posed by geographic information, and together they build into a holistic understanding of all that is important about GIS.

Our approach

GIS is a proven technology and the basic operations of GIS today provide secure and established foundations for measurement, mapping, and analysis of the real world. GIScience provides us with the ability to devise GIS-based analysis that is robust and defensible. GI technology facilitates analysis, and continues to evolve rapidly, especially in relation to the Internet, and its likely successors and its spin-offs. Better technology, better systems, and better science make better management and exploitation of GI possible.

Fundamentally, GIS is an applications-led technology, yet successful applications need appropriate scientific foundations. Effective use of GIS is impossible if they are simply seen as black boxes producing magic. GIS is applied rarely in controlled, laboratory-like conditions. Our messy, inconvenient, and apparently haphazard real world is the laboratory for GIS, and the science of real-world application is the difficult kind – it can rarely control for, or assume away, things that we would prefer were not there and that get in the way of almost any given application. Scientific understanding of the inherent uncertainties and imperfections in representing the world makes us able to judge whether the conclusions of our analysis are sustainable, and is essential for everything except the most trivial use of GIS. GIScience is also founded on a search for understanding and predictive power in a world where human factors interact with those relating to the physical environment. Good science is also ethical and clearly communicated science, and thus the ways in which we analyze and depict geography also play an important role.

Digital geographic information is central to the practicality of GIS. If it does not exist, it is expensive to collect, edit, or update. If it does exist, it cuts costs and time – assuming it is fit for the purpose, or good enough for the particular task in hand. It underpins the rapid growth of trading in geographic information (g-commerce). It provides possibilities not only for local business but also for entering new markets or for forging new relationships with other organizations. It is a foolish individual who sees it only as a commodity like baked beans or shaving foam. Its value relies upon its coverage, on the strengths of its representation of diversity, on its truth within a constrained definition of that word, and on its availability.

Few of us are hermits. The way in which geographic information is created and exploited through GIS affects us as citizens, as owners of enterprises, and as employees. It has increasingly been argued that GIS is only a part – albeit a part growing in importance and size – of the Information, Communications, and Technology (ICT) industry. This is a limited perception, typical of the ICT supply-side industry which tends to see itself as the sole progenitor of change in the world (wrongly). Actually, it is much more sensible to take a balanced demand- and supply-side perspective: GIS and geographic information can and do underpin many operations of many organizations, but how GIS works in detail differs between different cultures, and can often also partly depend on whether an organization is in the private or public sector. Seen from this perspective, management of GIS facilities is crucial to the success of organizations – businesses as we term them later. The management of the organizations using our tools, information, knowledge, skills, and commitment is therefore what will ensure the ultimate local and global success of GIS. For this reason we devote an entire section of this book to management issues. We go far beyond how to choose, install, and run a GIS; that is only one part of the enterprise. We try to show how to use GIS and geographic information to contribute to the business success of your organization (whatever it is), and have it recognized as doing just that. To achieve that, you need to know what drives organizations and how they operate in the reality of their business environments. You need to know something about assets, risks, and constraints on actions – and how to avoid the last two and

nurture the first. And you need to be exposed – for that is reality – to the inter-dependencies in any organization and the tradeoffs in decision making in which GIS can play a major role.

Our audience

Originally, we conceived this book as a 'student companion' to a very different book that we also produced as a team – the second edition of the 'Big Book' of GIS (Longley et al 1999). This reference work on GIS provided a defining statement of GIS at the end of the last millennium: many of the chapters that are of enduring relevance are now available as an advanced reader in GIS (Longley et al 2005). These books, along with the first 'Big Book' of GIS (Maguire et al 1991) were designed for those who were already very familiar with GIS, and desired an advanced understanding of enduring GIS principles, techniques, and management practices. They were not designed as books for those being introduced to the subject.

This book is the companion for everyone who desires a rich understanding of how GIS is used in the real world. GIS today is both an increasingly mature technology and a strategically important interdisciplinary meeting place. It is taught as a component of a huge range of undergraduate courses throughout the world, to students that already have different skills, that seek different disciplinary perspectives on the world, and that assign different priorities to practical problem solving and the intellectual curiosities of science. This companion can be thought of as a textbook, though not in a conventionally linear way. We have not attempted to set down any kind of rigid GIS curriculum beyond the core organizing principles, techniques, analysis methods, and management practices that we believe to be important. We have structured the material in each of the sections of the book in a cumulative way, yet we envisage that very few students will start at Chapter 1 and systematically work through to Chapter 21 – much of learning is not like that any more (if ever it was), and most instructors will navigate a course between sections and chapters of the book that serves their particular disciplinary, curricular, and practical priorities. The ways in which three of us use the book in our own undergraduate and postgraduate settings are posted on the book's website (**www.wiley.com/go/longley**), and we hope that other instructors will share their best practices with us as time goes on (please see the website for instructions on how to upload instructor lists and offer feedback on those that are already there!). Our Instructor Manual (see **www.wiley.com/go/longley**) provides suggestions as to the use of this book in a range of disciplines and educational settings. The linkage of the book to reference material (specifically Longley et al (2005) and Maguire et al (1991) at **www.wiley.com/go/longley**) is a particular strength for GIS postgraduates and professionals. Such users might desire an up-to-date overview of GIS to locate their own particular endeavors, or (particularly if their previous experience lies outside the mainstream geographic sciences) a fast track to get up-to-speed with the range of principles, techniques, and practice issues that govern real-world application.

The format of the book is intended to make learning about GIS fun. GIS is an important transferable skill because people successfully use it to solve real-world problems. We thus convey this success through use of real (not contrived, conventional text-book like) applications, in clearly identifiable boxes throughout the text. But even this does not convey the excitement of learning about GIS that only comes from doing. With this in mind, an on-line series of laboratory classes have been created to accompany the book. These are available, free of charge, to any individual working in an institution that has an ESRI site license (see **www.esri.com**). They are cross-linked in detail to individual chapters and sections in the book, and provide learners with the opportunity to refresh the concepts and techniques that they have acquired through classes and reading, and the opportunity to work through extended examples using ESRI ArcGIS. This is by no means the only available software for learning GIS: we have chosen it for our own lab exercises because it is widely used, because one of us works for ESRI Inc. (Redlands, CA., USA) and because ESRI's cooperation enabled us to tailor the lab exercises to our own material. There are, however, many other options for lab teaching and distance learning from private and publicly funded bodies such as the UNIGIS consortium, the Worldwide Universities Network, and Pennsylvania State University in its World Campus (**www.worldcampus.psu.edu/pub/index.shtml**).

GIS is not just about machines, but also about people. It is very easy to lose touch with what is new in GIS, such is the scale and pace of development. Many of these developments have been, and continue to be, the outcome of work by motivated and committed individuals – many an idea or implementation of GIS would not have taken place without an individual to champion it. In the first edition of this book, we used boxes highlighting the contributions of a number of its champions to convey that GIS is a living, breathing subject. In this second edition, we have removed all of the living champions of GIS and replaced them with a completely new set – not as any intended slight upon the remarkable contributions that these individuals have made, but as a necessary way of freeing up space to present vignettes of an entirely new set of committed, motivated individuals whose contributions have also made a difference to GIS.

As we say elsewhere in this book, human attention is valued increasingly by business, while students are also seemingly required to digest ever-increasing volumes of material. We have tried to summarize some of the most important points in this book using short 'factoids', such as that below, which we think assist students in recalling core points.

Short, pithy, statements can be memorable.

We hope that instructors will be happy to use this book as a core teaching resource. We have tried to provide a number of ways in which they can encourage their

students to learn more about GIS through a range of assessments. At the end of each chapter we provide four questions in the following sequence that entail:

- **Student-centred learning** by doing.
- A **review** of material contained in the chapter.
- A **review and research** task – involving integration of issues discussed in the chapter with those discussed in additional external sources.
- A **compare and research** task – similar to the review and research task above, but additionally entailing linkage with material from one or more other chapters in the book.

The on-line lab classes have also been designed to allow learning in a self-paced way, and there are self-test exercises at the end of each section for use by learners working alone or by course evaluators at the conclusion of each lab class.

As the title implies, this is a book about geographic information systems, the practice of science in general, and the principles of geographic information science (GIScience) in particular. We remain convinced of the need for high-level understanding and our book deals with ideas and concepts – as well as with actions. Just as scientists need to be aware of the complexities of interactions between people and the environment, so managers must be well-informed by a wide range of knowledge about issues that might impact upon their actions. Success in GIS often comes from dealing as much with people as with machines.

The new learning paradigm

This is not a traditional textbook because:

- It recognizes that GISystems and GIScience do not lend themselves to traditional classroom teaching alone. Only by a combination of approaches can such crucial matters as principles, technical issues, practice, management, ethics, and accountability be learned. Thus the book is complemented by a website (**www.wiley.com/go/longley**) and by exercises that can be undertaken in laboratory or self-paced settings.
- It brings the principles and techniques of GIScience to those learning about GIS for the first time – and as such represents part of the continuing evolution of GIS.
- The very nature of GIS as an underpinning technology in huge numbers of applications, spanning different fields of human endeavor, ensures that learning has to be tailored to individual or small-group needs. These are addressed in the Instructor Manual to the book (**www.wiley.com/go/longley**).
- We have recognized that GIS is driven by real-world applications and real people, that respond to real-world needs. Hence, information on a range of applications and GIS champions is threaded throughout the text.
- We have linked our book to online learning resources throughout, notably the ESRI Virtual Campus.
- The book that you have in your hands has been completely restructured and revised, while retaining the best features of the (highly successful) first edition published in 2001.

Summary

This is a book that recognizes the growing commonality between the concerns of science, government, and business. The examples of GIS people and problems that are scattered through this book have been chosen deliberately to illuminate this commonality, as well as the interplay between organizations and people from different sectors. To differing extents, the five sections of the book develop common concerns with effectiveness and efficiency, by bringing together information from disparate sources, acting within regulatory and ethical frameworks, adhering to scientific principles, and preserving good reputations. This, then, is a book that combines the basics of GIS with the solving of problems which often have no single, ideal solution – the world of business, government, and interdisciplinary, mission-orientated holistic science.

In short, we have tried to create a book that remains attuned to the way the world works now, that understands the ways in which most of us increasingly operate as knowledge workers, and that grasps the need to face complicated issues that do not have ideal solutions. As with the first edition of the book, this is an unusual enterprise and product. It has been written by a multinational partnership, drawing upon material from around the world. One of the authors is an employee of a leading software vendor and two of the other three have had business dealings with ESRI over many years. Moreover, some of the illustrations and examples come from the customers of that vendor. We wish to point out, however, that neither ESRI (nor Wiley) has ever sought to influence our content or the way in which we made our judgments, and we have included references to other software and vendors throughout the book. Whilst our lab classes are part of ESRI's Virtual Campus, we also make reference to similar sources of information in both paper and digital form. We hope that we have again created something novel but valuable by our lateral thinking in all these respects, and would very much welcome feedback through our website (**www.wiley.com/go/longley**).

Conventions and organization

We use the acronym *GIS* in many ways in the book, partly to emphasize one of our goals, the interplay between geographic information *systems* and geographic information *science*; and at times we use two other possible interpretations of the three-letter acronym: geographic information

studies and geographic information *services*. We distinguish between the various meanings where appropriate, or where the context fails to make the meaning clear, especially in Section 1.6 and in the Epilog. We also use the acronym in both singular and plural senses, following what is now standard practice in the field, to refer as appropriate to a single geographic information system or to geographic information systems in general. To complicate matters still further, we have noted the increasing use of 'geospatial' rather than 'geographic'. We use 'geospatial' where other people use it as a proper noun/title, but elsewhere use the more elegant and readily intelligible 'geographic'.

We have organized the book in five major but interlocking sections: after two chapters that establish the foundations to GI Systems and Science and the real world of applications, the sections appear as Principles (Chapters 3 through 6), Techniques (Chapters 7 through 11), Analysis (12 through 16) and Management and Policy (Chapters 17 through 20). We cap the book off with an Epilog that summarizes the main topics and looks to the future. The boundaries between these sections are in practice permeable, but remain in large part predicated upon providing a systematic treatment of enduring principles – ideas that will be around long after today's technology has been relegated to the museum – and the knowledge that is necessary for an understanding of today's technology, and likely near-term developments. In a similar way, we illustrate how many of the analytic methods have had reincarnations through different manual and computer technologies in the past, and will doubtless metamorphose further in the future.

We hope you find the book stimulating and helpful. Please tell us – either way!

Acknowledgments

We take complete responsibility for all the material contained herein. But much of it draws upon contributions made by friends and colleagues from across the world, many of them outside the academic GIS community. We thank them all for those contributions and the discussions we have had over the years. We cannot mention all of them but would particularly like to mention the following.

We thanked the following for their direct and indirect inputs to the first edition of this book: Mike Batty, Clint Brown, Nick Chrisman, Keith Clarke, Andy Coote, Martin Dodge, Danny Dorling, Jason Dykes, Max Egenhofer, Pip Forer, Andrew Frank, Rob Garber, Gayle Gaynor, Peter Haggett, Jim Harper, Rich Harris, Les Hepple, Sophie Hobbs, Andy Hudson-Smith, Karen Kemp, Chuck Killpack, Robert Laurini, Vanessa Lawrence, John Leonard, Bob Maher, Nick Mann, David Mark, David Martin, Elanor McBay, Ian McHarg, Scott Morehouse, Lou Page, Peter Paisley, Cath Pyke, Jonathan Raper, Helen Ridgway, Jan Rigby, Christopher Roper, Garry Scanlan, Sarah Sheppard, Karen Siderelis, David Simonett, Roger Tomlinson, Carol Tullo, Dave Unwin, Sally Wilkinson, David Willey, Jo Wood, Mike Worboys.

Many of those listed above also helped us in our work on the second edition. But this time around we additionally acknowledge the support of: Tessa Anderson, David Ashby, Richard Bailey, Brad Baker, Bob Barr, Elena Besussi, Dick Birnie, John Calkins, Christian Castle, David Chapman, Nancy Chin, Greg Cho, Randy Clast, Rita Colwell, Sonja Curtis, Jack Dangermond, Mike de Smith, Steve Evans, Andy Finch, Amy Garcia, Hank Gerie, Muki Haklay, Francis Harvey, Denise Lievesley, Daryl Lloyd, Joe Lobley, Ian Masser, David Miller, Russell Morris, Doug Nebert, Hugh Neffendorf, Justin Norry, Geof Offen, Larry Orman, Henk Ottens, Jonathan Rhind, Doug Richardson, Dawn Robbins, Peter Schaub, Sorin Scortan, Duncan Shiell, Alex Singleton, Aidan Slingsby, Sarah Smith, Kevin Schürer, Josef Strobl, Larry Sugarbaker, Fraser Taylor, Bethan Thomas, Carolina Tobón, Paul Torrens, Nancy Tosta, Tom Veldkamp, Peter Verburg, and Richard Webber. Special thanks are also due to Lyn Roberts and Keily Larkins at John Wiley and Sons for successfully guiding the project to fruition. Paul Longley's contribution to the book was carried out under ESRC AIM Fellowship RES-331-25-0001, and he also acknowledges the guiding contribution of the CETL Center for Spatial Literacy in Teaching (Splint).

Each of us remains indebted in different ways to Stan Openshaw, for his insight, his energy, his commitment to GIS, and his compassion for geography.

Finally, thanks go to our families, especially Amanda, Fiona, Heather, and Christine.

Paul Longley, University College London
Michael Goodchild, University of California Santa Barbara
David Maguire, ESRI Inc., Redlands CA
David Rhind, City University, London

October 2004

Further reading

Maguire D.J., Goodchild M.F., and Rhind D.W. (eds) 1991 *Geographical Information Systems*. Harlow: Longman.

Longley P.A., Goodchild M.F., Maguire D.W., and Rhind D.W. (eds) 1999 *Geographical Information Systems: Principles, Techniques, Management and Applications (two volumes)*. New York, NJ: Wiley.

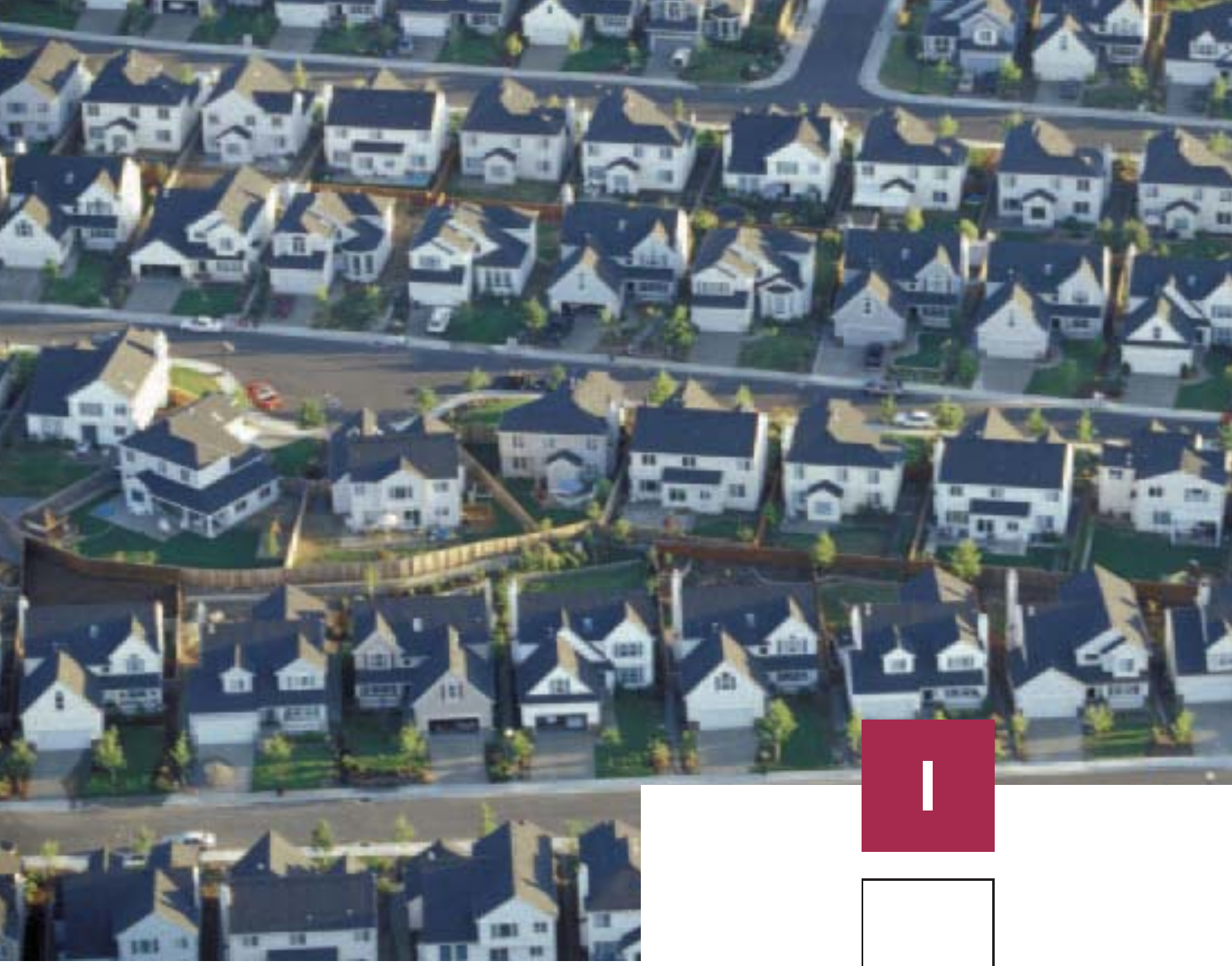
Longley P.A., Goodchild M.F., Maguire D.W., and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley.

List of Acronyms and Abbreviations

AA Automobile Association
ABM agent-based model
AGI Association for Geographic Information
AGILE Association of Geographic Information Laboratories in Europe
AHP Analytical Hierarchy Process
AM automated mapping
AML Arc Macro Language
API application programming interface
ARPANET Advanced Research Projects Agency Network
ASCII American Standard Code for Information Interchange
ASP Active Server Pages
AVIRIS Airborne Visible InfraRed Imaging Spectrometer
BBC British Broadcasting Corporation
BLM Bureau of Land Management
BLOB binary large object
CAD Computer-Aided Design
CAMA Computer Assisted Mass Appraisal
CAP Common Agricultural Policy
CASA Centre for Advanced Spatial Analysis
CASE computer-aided software engineering
CBD central business district
CD compact disc
CEN Comité Européen de Normalisation
CERN Conseil Européen pour la Recherche Nucléaire
CGIS Canada Geographic Information System
CGS Czech Geological Survey
CIA Central Intelligence Agency
CLI Canada Land Inventory
CLM collection-level metadata
COGO coordinate geometry
COM component object model
COTS commercial off-the-shelf
CPD continuing professional development
CSDGM Content Standards for Digital Geospatial Metadata
CSDMS Centre for Spatial Database Management and Solutions
CSO color separation overlay
CTA Chicago Transit Authority
DARPA Defense Advanced Research Projects Agency
DBA database administrator
DBMS database management system
DCL data control language
DCM digital cartographic model
DCW Digital Chart of the World
DDL data definition language
DEM digital elevation model
DGPS Differential Global Positioning System
DHS Department of Homeland Security
DIME Dual Independent Map Encoding
DLG digital line graph
DLM digital landscape model
DML data manipulation language
DRG digital raster graphic
DST Department of Science and Technology
DXF drawing exchange format
EBIS ESRI Business Information Solutions
EC European Commission
ECU Experimental Cartography Unit
EDA exploratory data analysis
EOSDIS Earth Observing System Data and Information System
EPA Environmental Protection Agency
EPS encapsulated postscript
ERDAS Earth Resource Data Analysis System
ERP Enterprise Resource Planning
ERTS Earth Resources Technology Satellite
ESDA exploratory spatial data analysis
ESRI Environmental Systems Research Institute
EU European Union
EUROGI European Umbrella Organisation for Geographic Information
FAO Food and Agriculture Organization
FEMA Federal Emergency Management Agency
FGDC Federal Geographic Data Committee
FIPS Federal Information Processing Standard
FM facility management
FOIA Freedom of Information Act
FSA Forward Sortation Area
GAO General Accounting Office
GBF-DIME Geographic Base Files – Dual Independent Map Encoding
GDI GIS data industry
GIO Geographic Information Officer
GIS geographic(al) information system
GIScience geographic(al) information science
GML Geography Markup Language
GNIS Geographic Names Information System
GOS geospatial one-stop
GPS Global Positioning System
GRASS Geographic Resources Analysis Support System
GSDI global spatial data infrastructure
GUI graphical user interface
GWR geographically weighted regression
HLS hue, lightness, and saturation
HTML hypertext markup language
HTTP hypertext transmission protocol
ICMA International City/County Management Association
ICT Information and Communication Technology
ID identifier
IDE Integrated Development Environment
IDW inverse-distance weighting
IGN Institut Géographique National

IMW International Map of the World
INSPIRE Infrastructure for Spatial Information in Europe
IP Internet protocol
IPR intellectual property rights
IS information system
ISCGM International Steering Committee for Global Mapping
ISO International Standards Organization
IT information technology
ITC International Training Centre for Aerial Survey
ITS intelligent transportation systems
JSP Java Server Pages
KE knowledge economy
KRIHS Korea Research Institute for Human Settlements
KSUCTA Kyrgyz State University of Construction, Transportation and Architecture
LAN local area network
LBS location-based services
LiDAR light detection and ranging
LISA local indicators of spatial association
LMIS Land Management Information System
MAT point of minimum aggregate travel
MAUP Modifiable Areal Unit Problem
MBR minimum bounding rectangle
MCDM multicriteria decision making
MGI Masters in Geographic Information
MIT Massachusetts Institute of Technology
MOCT Ministry of Construction and Transportation
MrSID Multiresolution Seamless Image Database
MSC Mapping Science Committee
NASA National Aeronautics and Space Administration
NATO North Atlantic Treaty Organization
NAVTEQ Navigation Technologies
NCGIA National Center for Geographic Information and Analysis
NGA National Geospatial-Intelligence Agency
NGIS National GIS
NILS National Integrated Land System
NIMA National Imagery and Mapping Agency
NIMBY not in my back yard
NMO national mapping organization
NMP National Mapping Program
NOAA National Oceanic and Atmospheric Administration
NPR National Performance Review
NRC National Research Council
NSDI National Spatial Data Infrastructure
NSF National Science Foundation
OCR optical character recognition
ODBMS object database management system
OEM Office of Emergency Management
OGC Open Geospatial Consortium
OLM object-level metadata
OLS ordinary least squares
OMB Office of Management and Budget
ONC Operational Navigation Chart
ORDBMS object-relational database management system
PAF postcode address file
PASS Planning Assistant for Superintendent Scheduling
PCC percent correctly classified
PCGIAP Permanent Committee on GIS Infrastructure for Asia and the Pacific
PDA personal digital assistant
PE photogrammetric engineering
PERT Program, Evaluation, and Review Techniques
PLSS Public Land Survey System
PPGIS public participation in GIS
RDBMS relational database management system
RFI Request for Information
RFP Request for Proposals
RGB red-green-blue
RMSE root mean square error
ROMANSE Road Management System for Europe
RRL Regional Research Laboratory
RS remote sensing
SAP spatially aware professional
SARS severe acute respiratory syndrome
SDE Spatial Database Engine
SDI spatial data infrastructure
SDSS spatial decision support systems
SETI Search for Extraterrestrial Intelligence
SIG Special Interest Group
SOHO small office/home office
SPC State Plane Coordinates
SPOT Système Probatoire d'Observation de la Terre
SQL Structured/Standard Query Language
SWMM Storm Water Management Model
SWOT strengths, weaknesses, opportunities, threats
TC technical committee
TIGER Topologically Integrated Geographic Encoding and Referencing
TIN triangulated irregular network
TINA there is no alternative
TNM The National Map
TOID Topographic Identifier
TSP traveling-salesman problem
TTIC Traffic and Travel Information Centre
UCAS Universities Central Admissions Service
UCGIS University Consortium for Geographic Information Science
UCSB University of California, Santa Barbara
UDDI Universal Description, Discovery, and Integration
UDP Urban Data Processing
UKDA United Kingdom Data Archive
UML Unified Modeling Language
UN United Nations
UNIGIS UNIversity GIS Consortium
UPS Universal Polar Stereographic
URISA Urban and Regional Information Systems Association
USGS United States Geological Survey
USLE Universal Soil Loss Equation
UTC urban traffic control
UTM Universal Transverse Mercator
VBA Visual Basic for Applications
VfM value for money
VGA video graphics array

ViSC visualization in scientific computing
VPF vector product format
WAN wide area network
WIMP windows, icons, menus, and pointers
WIPO World Intellectual Property Organization
WSDL Web Services Definition Language
WTC World Trade Center
WTO World Trade Organization
WWF World Wide Fund for Nature
WWW World Wide Web
WYSIWYG what you see is what you get
XML extensible markup language



I

Introduction

1 *Systems, science, and study*
2 *A gallery of applications*

1 *Systems, science, and study*

This chapter introduces the conceptual framework for the book, by addressing several major questions:

- What exactly is geographic information, and why is it important? What is special about it?
- What is information generally, and how does it relate to data, knowledge, evidence, wisdom, and understanding?
- What kinds of decisions make use of geographic information?
- What is a geographic information system, and how would I know one if I saw one?
- What is geographic information science, and how does it relate to the use of GIS for scientific purposes?
- How do scientists use GIS, and why do they find it helpful?
- How do companies make money from GIS?

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

At the end of this chapter you will:

- Know definitions of the terms used throughout the book, including GIS itself;
- Be familiar with a brief history of GIS;
- Recognize the sometimes invisible roles of GIS in everyday life, and the roles of GIS in business;
- Understand the significance of geographic information science, and how it relates to geographic information systems;
- Understand the many impacts GIS is having on society, and the need to study those impacts.

1.1 Introduction: why does GIS matter?

Almost everything that happens, happens somewhere. Largely, we humans are confined in our activities to the surface and near-surface of the Earth. We travel over it and in the lower levels of the atmosphere, and through tunnels dug just below the surface. We dig ditches and bury pipelines and cables, construct mines to get at mineral deposits, and drill wells to access oil and gas. Keeping track of all of this activity is important, and knowing where it occurs can be the most convenient basis for tracking. Knowing where something happens is of critical importance if we want to go there ourselves or send someone there, to find other information about the same place, or to inform people who live nearby. In addition, most (perhaps all) decisions have geographic consequences, e.g., adopting a particular funding formula creates geographic winners and losers, especially when the process entails zero sum gains. Therefore geographic location is an important attribute of activities, policies, strategies, and plans. Geographic information systems are a special class of information systems that keep track not only of events, activities, and things, but also of *where* these events, activities, and things happen or exist.

Almost everything that happens, happens somewhere. Knowing where something happens can be critically important.

Because location is so important, it is an issue in many of the problems society must solve. Some of these are so routine that we almost fail to notice them – the daily question of which route to take to and from work, for example. Others are quite extraordinary occurrences, and require rapid, concerted, and coordinated responses by a wide range of individuals and organizations – such as the events of September 11 2001 in New York (Box 1.1). Problems that involve an aspect of location, either in the information used to solve them, or in the solutions themselves, are termed *geographic problems*. Here are some more examples:

- Health care managers solve geographic problems (and may create others) when they decide where to locate new clinics and hospitals.
- Delivery companies solve geographic problems when they decide the routes and schedules of their vehicles, often on a daily basis.
- Transportation authorities solve geographic problems when they select routes for new highways.
- Geodemographics consultants solve geographic problems when they assess and recommend where best to site retail outlets.
- Forestry companies solve geographic problems when they determine how best to manage forests, where to cut, where to locate roads, and where to plant new trees.
- National Park authorities solve geographic problems when they schedule recreational path maintenance and improvement (Figure 1.3).
- Governments solve geographic problems when they decide how to allocate funds for building sea defenses.
- Travelers and tourists solve geographic problems when they give and receive driving directions, select hotels in unfamiliar cities, and find their way around theme parks (Figure 1.4).
- Farmers solve geographic problems when they employ new information technology to make better decisions about the amounts of fertilizer and pesticide to apply to different parts of their fields.

If so many problems are geographic, what distinguishes them from each other? Here are three bases for classifying geographic problems. First, there is the question of scale, or level of geographic detail. The architectural design of a building can present geographic problems, as in disaster management (Box 1.1), but only at a very detailed or local scale. The information needed to service the building is also local – the size and shape of the parcel, the vertical and subterranean extent of the building, the slope of the land, and its accessibility using normal and emergency infrastructure. The global diffusion of the 2003 severe acute respiratory syndrome (SARS) epidemic, or of bird flu in 2004 were problems at a much broader and coarser scale, involving information about entire national populations and global transport patterns.

Scale or level of geographic detail is an essential property of any GIS project.

Second, geographic problems can be distinguished on the basis of intent, or purpose. Some problems are strictly practical in nature – they must often be solved as quickly as possible and/or at minimum cost, in order to achieve such practical objectives as saving money, avoiding fines by regulators, or coping with an emergency. Others are better characterized as driven by human curiosity. When geographic data are used to verify the theory of continental drift, or to map distributions of glacial deposits, or to analyze the historic movements of people in anthropological or archaeological research (Box 1.2 and Figure 1.5), there is no sense of an immediate problem that needs to be solved – rather, the intent is the advancement of human understanding of the world, which we often recognize as the intent of science.

Although science and practical problem solving are often seen as distinct human activities, it is often argued that there is no longer any effective distinction between their methods. The tools and methods used by a scientist in a government agency to ensure the protection of an endangered species are essentially the same as the tools used by an academic ecologist to advance our scientific knowledge of biological systems. Both use the most accurate measurement devices, use terms whose meanings have been widely shared and agreed, insist that their results be replicable by others, and in general follow all of the principles of science that have evolved over the past centuries.

The use of GIS for both forms of activity certainly reinforces this idea that science and practical problem solving are no longer distinct in their methods, as does the fact that GIS is used widely in all kinds of organizations, from academic institutions to government agencies and corporations. The use of similar tools and methods right across science and problem solving is part of a shift from the pursuit of curiosity within traditional academic disciplines to solution centered, interdisciplinary team work.

Applications Box **1.1**

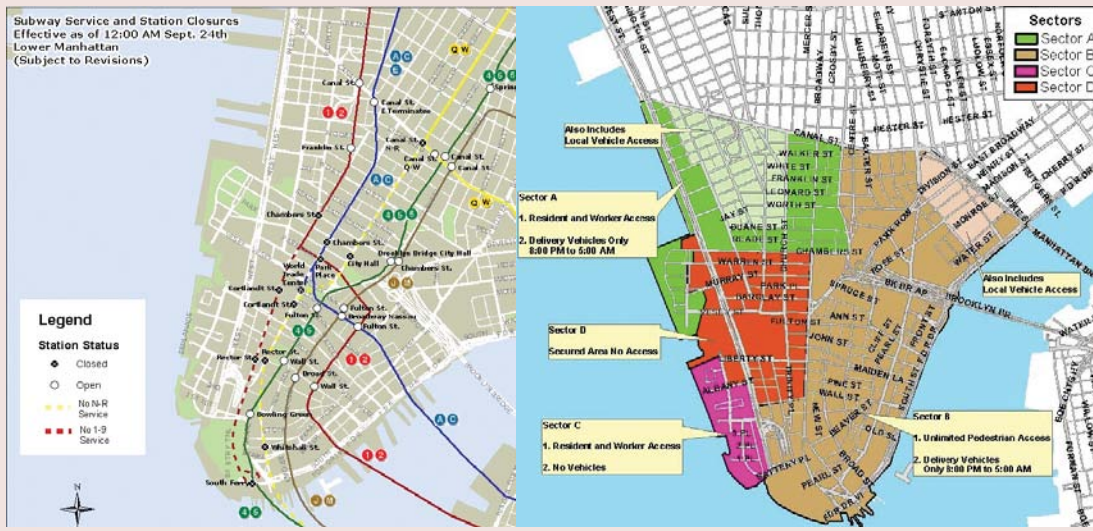
September 11 2001

Almost everyone remembers where they were when they learned of the terrorist atrocities in New York on September 11 2001. Location was crucial in the immediate aftermath and the emergency response, and the attacks had locational repercussions at a range of spatial

Figure 1.1 GIS in the Office of Emergency Management (OEM), first set up in the World Trade Center (WTC) complex immediately following the 2001 terrorist attacks on New York (Courtesy ESRI)

(geographic) and temporal (short, medium, and long time periods) *scales*. In the short term, the incidents triggered local emergency evacuation and disaster recovery procedures and global shocks to the financial system through the suspension of the New York Stock Exchange; in the medium term they blocked part of the New York subway system (that ran underneath the Twin Towers), profoundly changed regional work patterns (as affected workers became telecommuters) and had calamitous effects on the local retail economy; and in the

(A)



(B)

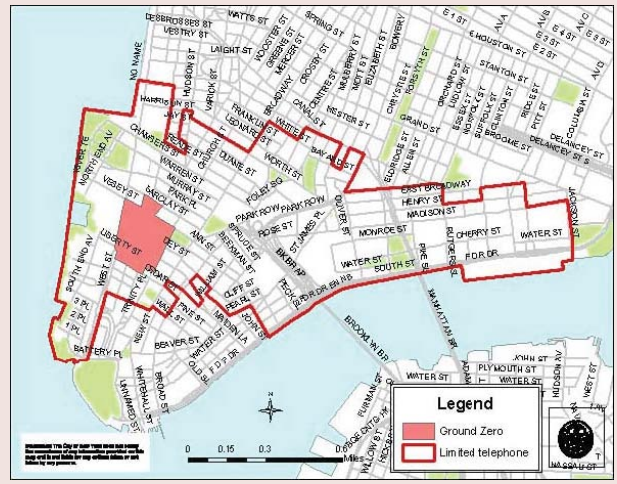


Figure 1.2 GIS usage in emergency management following the 2001 terrorist attacks on New York: (A) subway, pedestrian and vehicular traffic restrictions; (B) telephone outages; and (C) surface dust monitoring three days after the disaster. (Courtesy ESRI)

(C)

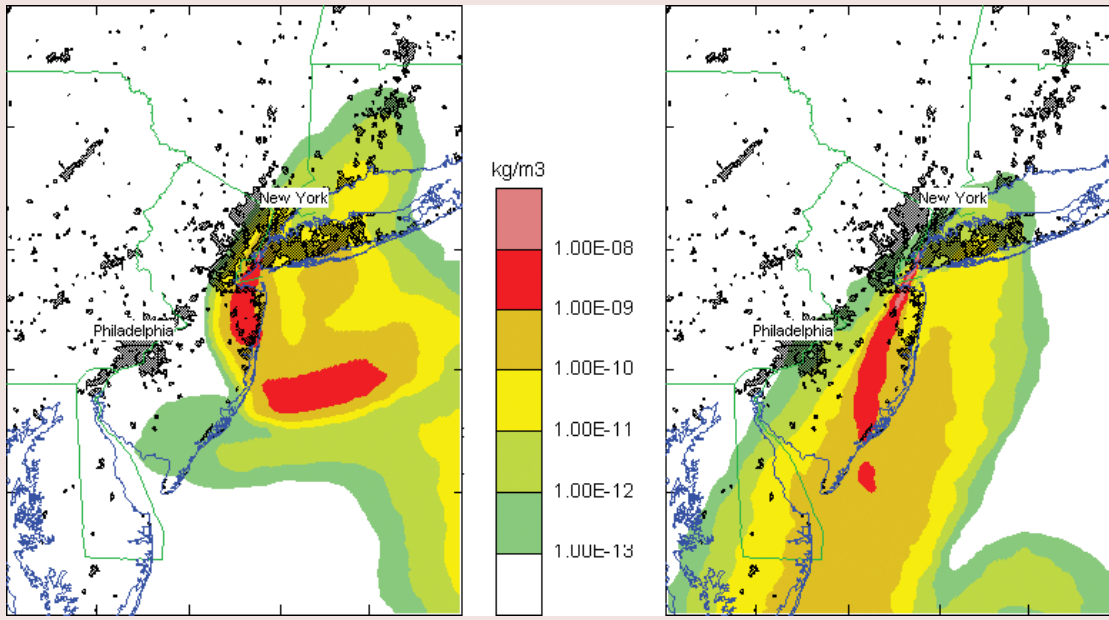


Figure 1.2 (*continued*)

long term, they have profoundly changed the way that we think of emergency response in our heavily networked society. Figures 1.1 and 1.2 depict some of the ways in which GIS was used for emergency management in New York in the immediate aftermath of the attacks. But the events also have much wider implications for the handling and management of geographic information, that we return to in Chapter 20.

At some points in this book it will be useful to distinguish between applications of GIS that focus on design, or so-called *normative* uses, and applications that advance science, or so-called *positive* uses (a rather confusing meaning of that term, unfortunately, but the one commonly used by philosophers of science – its use implies that science confirms theories by finding *positive* evidence in support of them, and rejects theories when negative evidence is found). Finding new locations for retailers is an example of a normative application of GIS, with its focus on design. But in order to predict how consumers will respond to new locations it is necessary for retailers to analyze and model the actual patterns of behavior they exhibit. Therefore, the models they use will be grounded in observations of messy reality that have been tested in a positive manner.

With a single collection of tools, GIS is able to bridge the gap between curiosity-driven science and practical problem-solving.

Third, geographic problems can be distinguished on the basis of their time scale. Some decisions are *operational*, and are required for the smooth functioning of an organization, such as how to control electricity inputs into grids that experience daily surges and troughs in usage (see Section 10.6). Others are *tactical*, and concerned with medium-term decisions, such as where to cut trees in next year's forest harvesting plan. Others are *strategic*, and are required to give an organization long-term direction, as when retailers decide to expand or rationalize their store networks (Figure 1.7). These terms are explored in the context of logistics applications of GIS in Section 2.3.4.6. The real world is somewhat more complex than this, of course, and these distinctions may blur – what is theoretically and statistically the 1000-year flood influences strategic and tactical considerations but may possibly arrive a year after the previous one! Other problems that interest geophysicists, geologists, or evolutionary biologists may occur on time scales that are much longer than a human lifetime, but are still geographic in nature, such as predictions about the future physical environment of Japan, or about the animal populations of Africa. Geographic databases are often *transactional* (see Sections 10.2.1 and 10.9.1), meaning

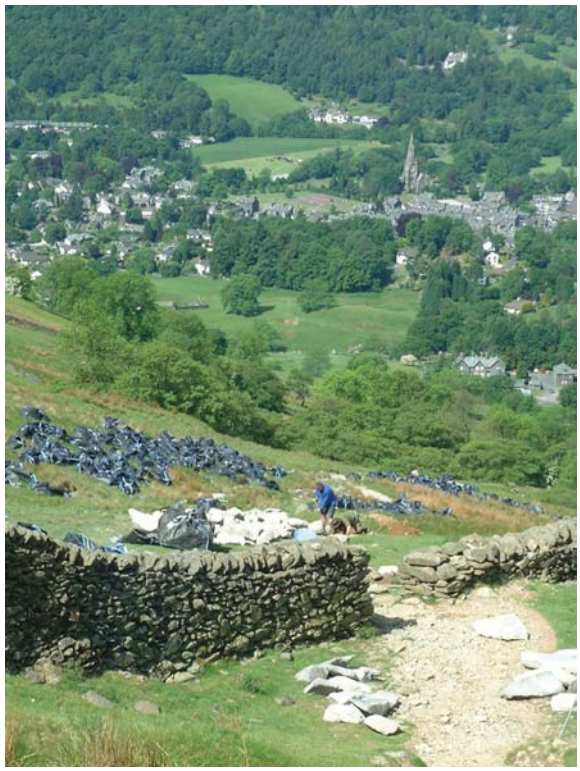


Figure 1.3 Maintaining and improving footpaths in National Parks is a geographic problem

that they are constantly being updated as new information arrives, unlike maps, which stay the same once printed.

Chapter 2 contains a more detailed discussion of the range and remits of GIS applications, and a view of how GIS pervades many aspects of our daily lives. Other applications are discussed to illustrate particular principles, techniques, analytic methods, and management practices as these arise throughout the book.

1.1.1 Spatial is special

The adjective *geographic* refers to the Earth's surface and near-surface, and defines the subject matter of this book, but other terms have similar meaning. *Spatial* refers to any space, not only the space of the Earth's surface, and it is used frequently in the book, almost always

Figure 1.4 Navigating tourist destinations is a geographic problem

with the same meaning as *geographic*. But many of the methods used in GIS are also applicable to other non-geographic spaces, including the surfaces of other planets, the space of the cosmos, and the space of the human body that is captured by medical images. GIS techniques have even been applied to the analysis of genome sequences on DNA. So the discussion of analysis in this book is of *spatial* analysis (Chapters 14 and 15), not geographic analysis, to emphasize this versatility.

Another term that has been growing in usage in recent years is *geospatial* – implying a subset of spatial applied specifically to the Earth's surface and near-surface. The former National Intelligence and Mapping Agency was renamed as the National Geospatial-Intelligence Agency in late 2003 by the US President and the Web portal for US Federal Government data is called Geospatial One-Stop. In this book we have tended to avoid geospatial, preferring geographic, and spatial where we need to emphasize generality (see Section 21.2.2).

People who encounter GIS for the first time are sometimes driven to ask why geography is so important – why *is* spatial special? After all, there is plenty of information around about geriatrics, for example, and in principle one could create a geriatric information system. So why has geographic information spawned an entire industry, if geriatric information hasn't to anything like the same extent? Why are there no courses in universities specifically in geriatric information systems? Part of the answer should be clear already – almost all human

Applications Box **1.2**

Where did your ancestors come from?

As individuals, many of us are interested in *where* we came from – socially and geographically. Some of the best clues to our ancestry come from our (family) surnames, and Western surnames have different types of origins – many of which are explicitly or implicitly geographic in origin (such clues are less important in some Eastern societies where family histories are generally much better documented). Research at

University College London is using GIS and historic censuses and records to investigate the changing local and regional geographies of surnames within the UK since the late 19th century (Figure 1.5). This tells us quite a lot about migration, changes in local and regional economies, and even about measures of local economic health and vitality. Similar GIS-based analysis can be used to generalize about

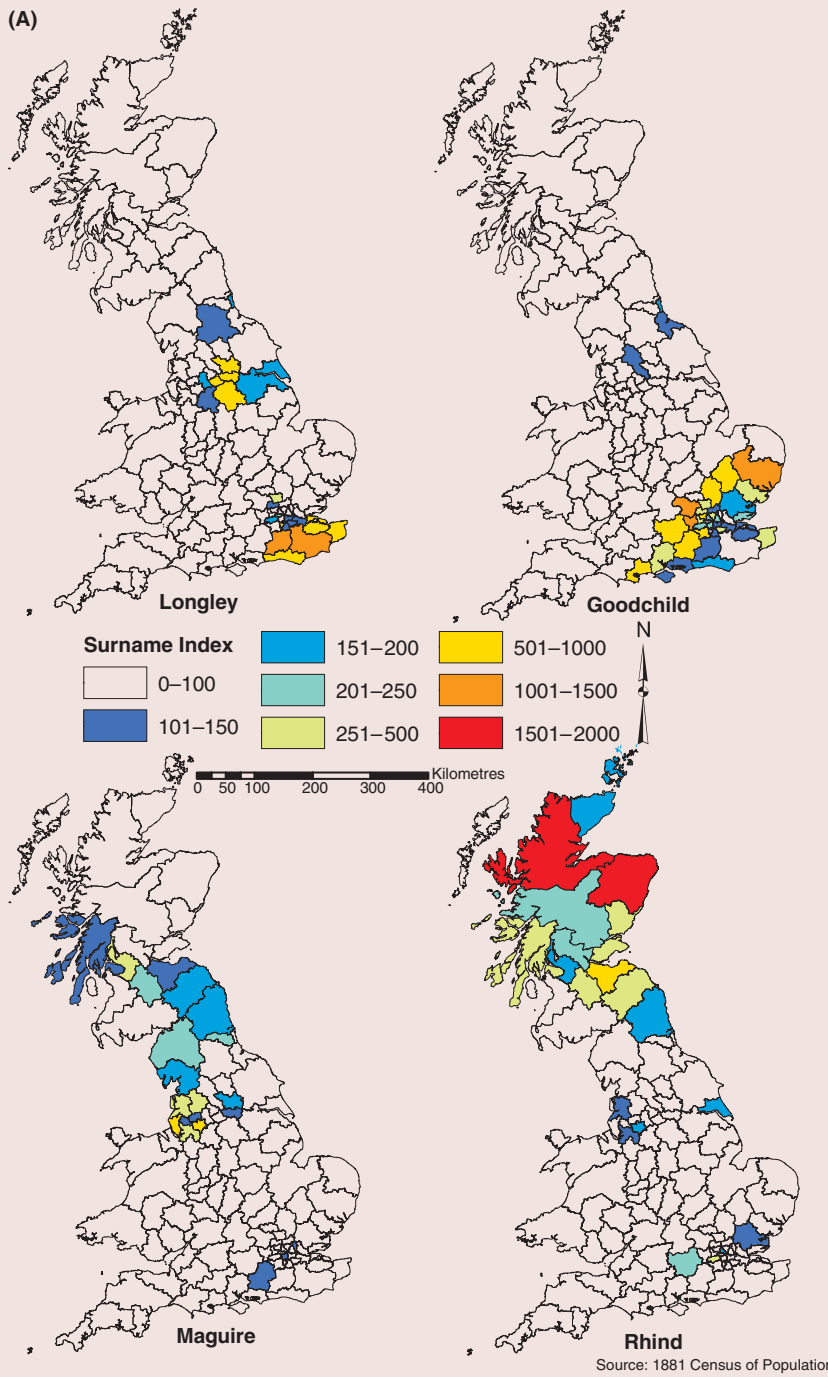


Figure 1.5 The UK geography of the Longleys, the Goodchilds, the Maguires, and the Rhinds in (A) 1881 and (B) 1998 (Reproduced with permission of Daryl Lloyd)

the characteristics of international emigrants (for example to North America, Australia, and New Zealand: Figure 1.6), or the regional naming patterns of immigrants to the US from the Indian sub-continent or China. In all kinds of senses, this helps us understand our place in the world. Fundamentally, this is curiosity-driven research: it is interesting to individuals to understand more about their origins, and it is interesting to everyone with planning or policy concerns with any particular place to understand the social and cultural mix of people that live there. But it is not central to resolving any specific problem within a specific timescale.

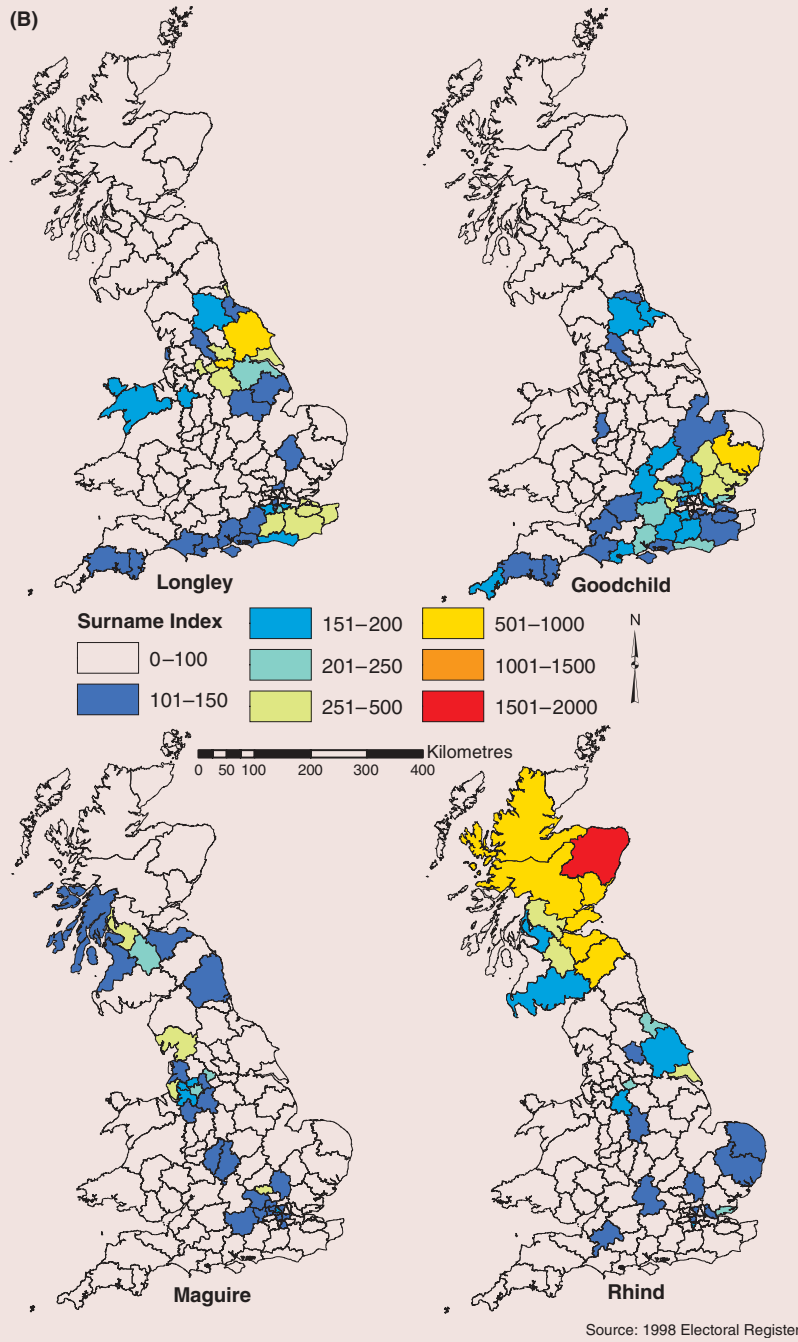


Figure 1.5 (*continued*)

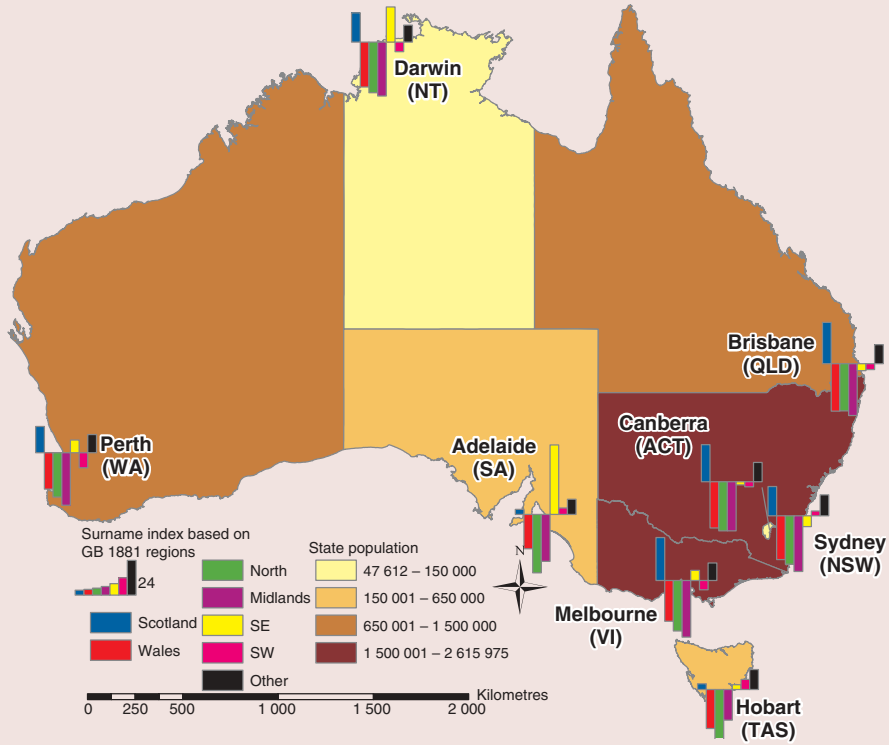


Figure 1.6 The geography of British emigrants to Australia (bars beneath the horizontal line indicate low numbers of migrants to the corresponding destination) (Reproduced with permission of Daryl Lloyd)

Figure 1.7 Store location principles are very important in the developing markets of Europe, as with Tesco's successful investment in Budapest, Hungary

activities and decisions involve a geographic component, and the geographic component is important. Another reason will become apparent in Chapter 3 – working with geographic information involves complex and difficult choices that are also largely unique. Other, more-technical reasons will become clear in later chapters, and are briefly summarized in Box 1.3.

1.2 Data, information, evidence, knowledge, wisdom

Information systems help us to manage *what we know*, by making it easy to organize and store, access and retrieve, manipulate and synthesize, and apply knowledge to the solution of problems. We use a variety of terms to describe what we know, including the five that head this section and that are shown in Table 1.2. There are no universally agreed definitions of these terms, the first two of which are used frequently in the GIS arena. Nevertheless it is worth trying to come to grips with their various meanings, because the differences between them can often be significant, and what follows draws upon many sources, and thus provides the basis for the use of these terms throughout the book. Data clearly refers to the most mundane kind of information, and wisdom to the most substantive.

Data consist of numbers, text, or symbols which are in some sense neutral and almost context-free. Raw geographic facts (see Box 18.7), such as the temperature at a specific time and location, are examples of data. When data are transmitted, they are treated as a stream of bits; a crucial requirement is to preserve the integrity of the dataset. The internal meaning of the data is irrelevant in

Technical Box **1.3**

Some technical reasons why geographic information is special

- It is multidimensional, because *two* coordinates must be specified to define a location, whether they be *x* and *y* or latitude and longitude.
- It is voluminous, since a geographic database can easily reach a terabyte in size (see Table 1.1).
- It may be represented at different levels of spatial resolution, e.g., using a representation equivalent to a 1:1 million scale map and a 1:24 000 scale one (see Box 4.2).
- It may be represented in different ways inside a computer (Chapter 3) and how this is done can strongly influence the ease of analysis and the end results.
- It must often be projected onto a flat surface, for reasons identified in Section 5.7.
- It requires many special methods for its analysis (see Chapters 14 and 15).
- It can be time-consuming to analyze.
- Although much geographic information is static, the process of updating is complex and expensive.
- Display of geographic information in the form of a map requires the retrieval of large amounts of data.

such considerations. Data (the noun is the plural of *datum*) are assembled together in a *database* (see Chapter 10), and the volumes of data that are required for some typical applications are shown in Table 1.1.

The term *information* can be used either narrowly or broadly. In a narrow sense, information can be treated as devoid of meaning, and therefore as essentially synonymous with data, as defined in the previous paragraph. Others define information as *anything* which can be digitized, that is, represented in digital form (Chapter 3), but also argue that information is differentiated from data by implying some degree of selection, organization, and preparation for particular purposes – information is data serving some *purpose*, or data that have been given some degree of *interpretation*. Information is often costly to produce, but once digitized it is cheap to reproduce and distribute. Geographic datasets, for example, may be very expensive to collect and assemble, but very cheap to copy and disseminate. One other characteristic of information is that it is easy to add value to it through processing, and through merger with other information. GIS provides an excellent example of the latter, because of the tools it provides for combining information from different sources (Section 18.3).

GIS does a better job of sharing data and information than knowledge, which is more difficult to detach from the knower.

Knowledge does not arise simply from having access to large amounts of information. It can be considered as information to which value has been added by interpretation based on a particular context, experience, and purpose. Put simply, the information available in a book or on the Internet or on a map becomes knowledge only when it has been read and understood. How the information is interpreted and used will be different for different readers depending on their previous experience, expertise, and needs. It is important to distinguish two types of knowledge: *codified* and *tacit*. Knowledge is codifiable if it can be written down and transferred relatively easily to others. Tacit knowledge is often slow to acquire and much more difficult to transfer. Examples include the knowledge built up during an apprenticeship, understanding of how a particular market works, or familiarity with using a particular technology or language. This difference in transferability means that codified and tacit knowledge need to be managed and rewarded quite differently. Because of its nature, tacit knowledge is often a source of competitive advantage.

Some have argued that knowledge and information are fundamentally different in at least three important respects:

- Knowledge entails a knower. Information exists independently, but knowledge is intimately related to people.

Table 1.1 Potential GIS database volumes for some typical applications (volumes estimated to the nearest order of magnitude). Strictly, bytes are counted in powers of 2 – 1 kilobyte is 1024 bytes, not 1000

1 megabyte	1 000 000	Single dataset in a small project database
1 gigabyte	1 000 000 000	Entire street network of a large city or small country
1 terabyte	1 000 000 000 000	Elevation of entire Earth surface recorded at 30 m intervals
1 petabyte	1 000 000 000 000 000	Satellite image of entire Earth surface at 1 m resolution
1 exabyte	1 000 000 000 000 000 000	A future 3-D representation of entire Earth at 10 m resolution?

Table 1.2 A ranking of the support infrastructure for decision making

Decision-making support infrastructure	Ease of sharing with everyone	GIS example
Wisdom ↑	*Impossible*	Policies developed and accepted by stakeholders
Knowledge ↑	*Difficult, especially tacit knowledge*	Personal knowledge about places and issues
Evidence ↑	*Often not easy*	Results of GIS analysis of many datasets or scenarios
Information ↑	*Easy*	Contents of a database assembled from raw facts
Data	*Easy*	Raw geographic facts

- Knowledge is harder to detach from the knower than information; shipping, receiving, transferring it between people, or quantifying it are all much more difficult than for information.
- Knowledge requires much more assimilation – we digest it rather than hold it. While we may hold conflicting information, we rarely hold conflicting knowledge.

Evidence is considered a half way house between information and knowledge. It seems best to regard it as a multiplicity of information from different sources, related to specific problems and with a consistency that has been validated. Major attempts have been made in medicine to extract evidence from a welter of sometimes contradictory sets of information, drawn from worldwide sources, in what is known as *meta-analysis*, or the comparative analysis of the results of many previous studies.

Wisdom is even more elusive to define than the other terms. Normally, it is used in the context of decisions made or advice given which is disinterested, based on all the evidence and knowledge available, but given with some understanding of the likely consequences. Almost invariably, it is highly individualized rather than being easy to create and share within a group. Wisdom is in a sense the top level of a hierarchy of decision-making infrastructure.

1.3 The science of problem solving

How are problems solved, and are geographic problems solved any differently from other kinds of problems? We humans have accumulated a vast storehouse about the world, including information both on how it *looks*, or its *forms*, and how it *works*, or its dynamic *processes*. Some of those processes are natural and built into the design of the planet, such as the processes of tectonic movement that lead to earthquakes, and the processes of atmospheric circulation that lead to hurricanes. Others are human in origin, reflecting the increasing influence that we have on our natural environment, through the burning of fossil fuels, the felling of forests, and the cultivation of crops (Figure 1.8). Others are imposed by us, in the form of laws, regulations, and practices. For example, zoning regulations affect the ways in which specific parcels of land can be used.

Knowledge about how the world works is more valuable than knowledge about how it looks, because such knowledge can be used to predict.

These two types of information differ markedly in their degree of generality. Form varies geographically, and the Earth's surface looks dramatically different in different places – compare the settled landscape of northern England with the deserts of the US Southwest (Figure 1.9). But processes can be very general. The ways in which the burning of fossil fuels affects the atmosphere are essentially the same in China as in Europe, although the two landscapes look very different. Science has always valued such general knowledge over knowledge of the specific, and hence has valued process knowledge over knowledge of form. Geographers in particular have witnessed a longstanding debate, lasting

Figure 1.8 Social processes, such as carbon dioxide emissions, modify the Earth's environment

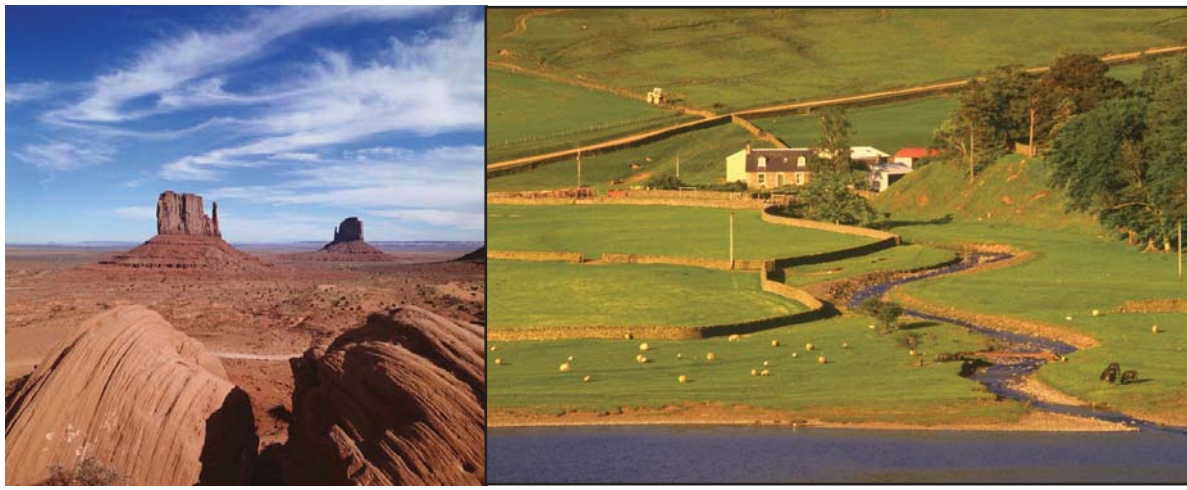


Figure 1.9 The form of the Earth's surface shows enormous variability, for example, between the deserts of the southwest USA and the settled landscape of northern England

centuries, between the competing needs of *idiographic* geography, which focuses on the description of form and emphasizes the unique characteristics of places, and *nomothetic* geography, which seeks to discover general processes. Both are essential, of course, since knowledge of general process is only useful in solving specific problems if it can be combined effectively with knowledge of form. For example, we can only assess the impact of soil erosion on agriculture in New South Wales if we know both how soil erosion is generally impacted by such factors as slope and specifically how much of New South Wales has steep slopes, and where they are located (Figure 1.10).

One of the most important merits of GIS as a tool for problem solving lies in its ability to combine the general with the specific, as in this example from New South Wales. A GIS designed to solve this problem would contain knowledge of New South Wales's slopes, in the form of computerized maps, and the programs executed by the GIS would reflect general knowledge of how slopes affect soil erosion. The *software* of a GIS captures and implements general knowledge, while the *database* of a GIS represents specific information. In that sense a GIS resolves the old debate between nomothetic and idiographic camps, by accommodating both.

GIS solves the ancient problem of combining general scientific knowledge with specific information, and gives practical value to both.

General knowledge comes in many forms. Classification is perhaps the simplest and most rudimentary, and is widely used in geographic problem solving. In many parts of the USA and other countries efforts have been made to limit development of wetlands, in the interests of preserving them as natural habitats and avoiding excessive impact on water resources. To support these efforts, resources have been invested in mapping wetlands, largely from aerial photography and satellite imagery. These maps simply classify land, using established rules that define what is and what is not a wetland (Figure 1.11).

Figure 1.10 Predicting landslides requires general knowledge of processes and specific knowledge of the area – both are available in a GIS (Reproduced with permission of PhotoDisc, Inc.)

More sophisticated forms of knowledge include *rule sets* – for example, rules that determine what use can be made of wetlands, or what areas in a forest can be legally logged. Rules are used by the US Forest Service

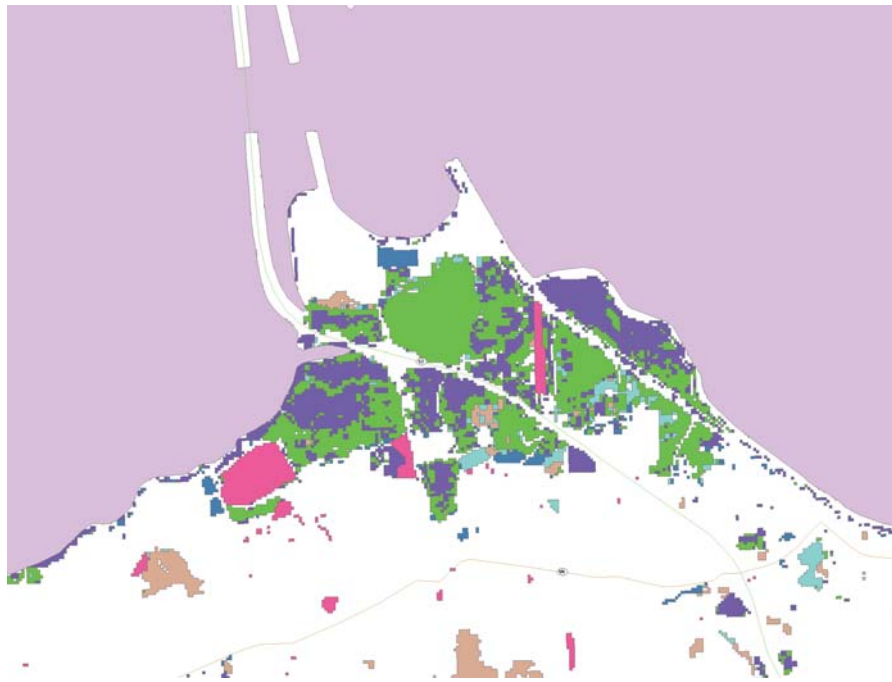


Figure 1.11 A wetland map of part of Erie County, Ohio, USA. The map has been made by classifying Landsat imagery at 30 m resolution. Brown = woods on hydric soil, dark blue = open water (excludes Lake Erie), green = shallow marsh, light blue = shrub/scrub wetland, blue-green = wet meadow, pink = farmed wetland. *Source*: Ohio Department of Natural Resources, **www.dnr.state.oh.us**

to define wilderness, and to impose associated regulations regarding the use of wilderness, including prohibition on logging and road construction.

Much of the knowledge gathered by the activities of scientists suggests the term *law*. The work of Sir Isaac Newton established the Laws of Motion, according to which all matter behaves in ways that can be perfectly predicted. From Newton's laws we are able to predict the motions of the planets almost perfectly, although Einstein later showed that certain observed deviations from the predictions of the laws could be explained with his Theory of Relativity. Laws of this level of predictive quality are few and far between in the geographic world of the Earth's surface. The real world is the only geographic-scale 'laboratory' that is available for most GIS applications, and considerable uncertainty is generated when we are unable to control for all conditions. These problems are compounded in the socioeconomic realm, where the role of human agency makes it almost inevitable that any attempt to develop rigid laws will be frustrated by isolated exceptions. Thus, while market researchers use spatial interaction models, in conjunction with GIS, to predict how many people will shop at each shopping center in a city, substantial errors will occur in the predictions. Nevertheless the results are of great value in developing location strategies for retailing. The Universal Soil Loss Equation, used by soil scientists in conjunction with GIS to predict soil erosion, is similar in its relatively low predictive power, but again the results are sufficiently accurate to be very useful in the right circumstances.

Solving problems involves several distinct components and stages. First, there must be an *objective*, or a goal that the problem solver wishes to achieve. Often this is a desire to maximize or minimize – find the solution of least cost, or shortest distance, or least time, or greatest profit; or to make the most accurate prediction possible. These objectives are all expressed in *tangible* form, that is, they can be measured on some well-defined scale. Others are said to be *intangible*, and involve objectives that are much harder, if not impossible to measure. They include maximizing *quality of life* and *satisfaction*, and minimizing *environmental impact*. Sometimes the only way to work with such intangible objectives is to involve human subjects, through surveys or focus groups, by asking them to express a preference among alternatives. A large body of knowledge has been acquired about such human-subjects research, and much of it has been employed in connection with GIS. For an example of the use of such mixed objectives see Section 16.4.

Often a problem will have *multiple objectives*. For example, a company providing a mobile snack service to construction sites will want to maximize the number of sites that can be visited during a daily operating schedule, and will also want to maximize the expected returns by visiting the most lucrative sites. An agency charged with locating a corridor for a new power transmission line may decide to minimize cost, while at the same time seeking to minimize environmental impact. Such problems employ methods known as *multicriteria decision making* (MCDM).

Many geographic problems involve multiple goals and objectives, which often cannot be expressed in commensurate terms.

1.4 The technology of problem solving

The previous sections have presented GIS as a technology to support both science and problem solving, using both specific and general knowledge about geographic reality. GIS has now been around for so long that it is, in many senses, a background technology, like word processing. This may well be so, but what exactly is this technology called GIS, and how does it achieve its objectives? In what ways is GIS more than a technology, and why does it continue to attract such attention as a topic for scientific journals and conferences?

Many definitions of GIS have been suggested over the years, and none of them is entirely satisfactory, though many suggest much more than a technology. Today, the label GIS is attached to many things: amongst them, a software product that one can buy from a vendor to carry out certain well-defined functions (*GIS software*); digital representations of various aspects of the geographic world, in the form of datasets (*GIS data*); a community of people who use and perhaps advocate the use of these tools for various purposes (the *GIS community*); and the activity of using a GIS to solve problems or advance science (*doing GIS*). The basic label works in all of these ways, and its meaning surely depends on the context in which it is used.

Nevertheless, certain definitions are particularly helpful (Table 1.3). As we describe in Chapter 3, GIS is much more than *a container of maps in digital form*. This can be a misleading description, but it is a helpful definition to give to someone looking for a simple explanation – a guest at a cocktail party, a relative, or a seat neighbor on an airline flight. We all know and appreciate the value of maps, and the notion that maps could be processed by a computer is clearly analogous to the use of word processing or spreadsheets to handle other types of information. A GIS is also *a computerized tool for solving geographic problems*, a definition that speaks to the purposes of GIS, rather than to its functions or physical form – an idea that is expressed in another definition, *a spatial decision support system*. A GIS is *a mechanized inventory of geographically distributed features and facilities*, the definition that explains the value of GIS to the utility industry, where it is used to keep track of such entities as underground pipes, transformers, transmission lines, poles, and customer accounts. A GIS is *a tool for revealing what is otherwise invisible in geographic information* (see Section 2.3.4.4), an interesting definition that emphasizes the power of a GIS as an analysis engine, to examine data and reveal its patterns, relationships, and anomalies – things that might not be apparent to someone looking at a map. A GIS is *a tool for performing operations on geographic data that are too tedious or expensive or inaccurate if performed by hand*, a definition that speaks to the problems associated with manual analysis of maps, particularly the extraction of simple measures, of area for example.

Table 1.3 Definitions of a GIS, and the groups who find them useful

A container of maps in digital form	The general public
A computerized tool for solving geographic problems	Decision makers, community groups, planners
A spatial decision support system	Management scientists, operations researchers
A mechanized inventory of geographically distributed features and facilities	Utility managers, transportation officials, resource managers
A tool for revealing what is otherwise invisible in geographic information	Scientists, investigators
A tool for performing operations on geographic data that are too tedious or expensive or inaccurate if performed by hand	Resource managers, planners

Everyone has their own favorite definition of a GIS, and there are many to choose from.

1.4.1 A brief history of GIS

As might be expected, there is some controversy about the history of GIS since parallel developments occurred in North America, Europe, and Australia (at least). Much of the published history focuses on the US contributions. We therefore do not yet have a well-rounded history of our subject. What is clear, though, is that the extraction of simple measures largely drove the development of the first real GIS, the Canada Geographic Information System or CGIS, in the mid-1960s (see Box 17.1). The Canada Land Inventory was a massive effort by the federal and provincial governments to identify the nation's land resources and their existing and potential uses. The most useful results of such an inventory are measures of area, yet area is notoriously difficult to measure accurately from a map (Section 14.3). CGIS was planned and developed as a measuring tool, a producer of tabular information, rather than as a mapping tool.

The first GIS was the Canada Geographic Information System, designed in the mid-1960s as a computerized map measuring system.

A second burst of innovation occurred in the late 1960s in the US Bureau of the Census, in planning the

tools needed to conduct the 1970 Census of Population. The DIME program (Dual Independent Map Encoding) created digital records of all US streets, to support automatic referencing and aggregation of census records. The similarity of this technology to that of CGIS was recognized immediately, and led to a major program at Harvard University's Laboratory for Computer Graphics and Spatial Analysis to develop a general-purpose GIS that could handle the needs of both applications – a project that led eventually to the ODYSSEY GIS of the late 1970s.

Early GIS developers recognized that the same basic needs were present in many different application areas, from resource management to the census.

In a largely separate development during the latter half of the 1960s, cartographers and mapping agencies had begun to ask whether computers might be adapted to their needs, and possibly to reducing the costs and shortening the time of map creation. The UK Experimental Cartography Unit (ECU) pioneered high-quality computer mapping in 1968; it published the world's first computer-made map in a regular series in 1973 with the British Geological Survey (Figure 1.12); the ECU also pioneered GIS work in education, post and zip codes as geographic references, visual perception of maps, and much else. National mapping agencies, such as Britain's Ordnance Survey, France's Institut Géographique National, and the US Geological Survey and the Defense Mapping Agency (now the National Geospatial-Intelligence Agency) began to investigate the use of computers to support the editing of maps, to avoid the expensive and slow process of hand correction and redrafting. The first automated cartography developments occurred in the 1960s, and by the late 1970s most major cartographic agencies were already computerized to some degree. But the magnitude of the task ensured that it was not until 1995 that the first country (Great Britain) achieved complete digital map coverage in a database.

Remote sensing also played a part in the development of GIS, as a source of technology as well as a source of data. The first military satellites of the 1950s were developed and deployed in great secrecy to gather intelligence, but the declassification of much of this material in recent years has provided interesting insights into the role played by the military and intelligence communities in the development of GIS. Although the early spy satellites used conventional film cameras to record images, digital remote sensing began to replace them in the 1960s, and by the early 1970s civilian remote sensing systems such as Landsat were beginning to provide vast new data resources on the appearance of the planet's surface from space, and to exploit the technologies of image classification and pattern recognition that had been developed earlier for military applications. The military was also responsible for the development in the 1950s of the world's first uniform system of measuring location, driven by the need for accurate targeting of intercontinental ballistic missiles, and this development led directly to the methods of

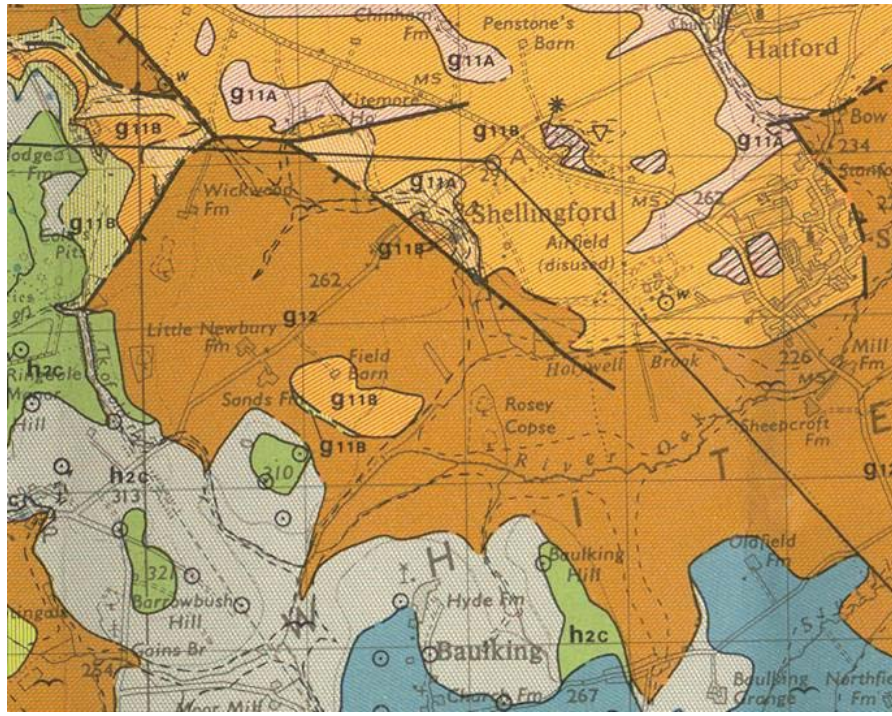


Figure 1.12 Section of the 1:63 360 scale geological map of Abingdon – the first known example of a map produced by automated means and published in a standard map series to established cartographic standards. (Reproduced by permission of the British Geological Survey and Ordnance Survey © NERC. All right reserved. IPR/59-13C)

positional control in use today (Section 5.6). Military needs were also responsible for the initial development of the Global Positioning System (GPS; Section 5.8).

Many technical developments in GIS originated in the Cold War.

GIS really began to take off in the early 1980s, when the price of computing hardware had fallen to a level that could sustain a significant software industry and cost-effective applications. Among the first customers were forestry companies and natural-resource agencies, driven by the need to keep track of vast timber resources, and to regulate their use effectively. At the time a modest computing system – far less powerful than today's personal computer – could be obtained for about $250 000, and the associated software for about $100 000. Even at these prices the benefits of consistent management using GIS, and the decisions that could be made with these new tools, substantially exceeded the costs. The market for GIS software continued to grow, computers continued to fall in price and increase in power, and the GIS software industry has been growing ever since.

The modern history of GIS dates from the early 1980s, when the price of sufficiently powerful computers fell below a critical threshold.

As indicated earlier, the history of GIS is a complex story, much more complex than can be described in this brief history, but Table 1.4 summarizes the major events of the past three decades.

1.4.2 Views of GIS

It should be clear from the previous discussion that GIS is a complex beast, with many distinct appearances. To some it is a way to automate the production of maps, while to others this application seems far too mundane compared to the complexities associated with solving geographic problems and supporting spatial decisions, and with the power of a GIS as an engine for analyzing data and revealing new insights. Others see a GIS as a tool for maintaining complex inventories, one that adds geographic perspectives to existing information systems, and allows the geographically distributed resources of a forestry or utility company to be tracked and managed. The sum of all of these perspectives is clearly too much for any one software package to handle, and GIS has grown from its initial commercial beginnings as a simple off-the-shelf package to a complex of software, hardware, people, institutions, networks, and activities that can be very confusing to the novice. A major software vendor such as ESRI today sells many distinct products, designed to serve very different needs: a major GIS workhorse (ArcInfo), a simpler system designed for viewing, analyzing, and mapping data (ArcView), an engine for supporting GIS-oriented websites (ArcIMS), an information system with spatial extensions (ArcSDE), and several others. Other vendors specialize in certain niche markets, such as the utility industry, or military and intelligence applications. GIS is a dynamic and evolving field, and its future is sure to be exciting, but speculations on where it might be headed are reserved for the final chapter.

Today a single GIS vendor offers many different products for distinct applications.

1.4.3 Anatomy of a GIS

1.4.3.1 The network

Despite the complexity noted in the previous section, a GIS does have its well-defined component parts. Today, the most fundamental of these is probably the *network*, without which no rapid communication or sharing of digital information could occur, except between a small group of people crowded around a computer monitor. GIS today relies heavily on the Internet, and on its limited-access cousins, the *intranets* of corporations, agencies, and the military. The Internet was originally designed as a network for connecting computers, but today it is rapidly becoming society's mechanism of information exchange, handling everything from personal messages to massive shipments of data, and increasing numbers of business transactions.

It is no secret that the Internet in its many forms has had a profound effect on technology, science, and society in the last few years. Who could have foreseen in 1990 the impact that the Web, e-commerce, digital government, mobile systems, and information and communication technologies would have on our everyday lives (see Section 18.4.4)? These technologies have radically changed forever the way we conduct business, how we communicate with our colleagues and friends, the nature of education, and the value and transitory nature of information.

The Internet began life as a US Department of Defense communications project called ARPANET (Advanced Research Projects Agency Network) in 1972. In 1980 Tim Berners-Lee, a researcher at CERN, the European organization for nuclear research, developed the hypertext capability that underlies today's World Wide Web – a key application that has brought the Internet into the realm of everyday use. Uptake and use of Web technologies have been remarkably quick, diffusion being considerably faster than almost all comparable innovations (for example, the radio, the telephone, and the television: see Figure 18.5). By 2004, 720 million people worldwide used the Internet (see Section 18.4.4 and Figure 18.8), and the fastest growth rates were to be found in the Middle East, Latin America, and Africa (**www.internetworldstats.com**). However, the global penetration of the medium remained very uneven – for example 62% of North Americans used the medium, but only 1% of Africans (Figure 1.13). Other Internet maps are available at the Atlas of Cybergeography maintained by Martin Dodge (**www.geog.ucl.ac.uk/casa/martin/atlas/atlas.html**).

Geographers were quick to see the value of the Internet. Users connected to the Internet could zoom in to parts of the map, or pan to other parts, using simple

Table 1.4 Major events that shaped GIS

Date	Type	Event	Notes
		The Era of Innovation	
1957	Application	First known automated mapping produced	Swedish meteorologists and British biologists
1963	Technology	CGIS development initiated	Canada Geographic Information System is developed by Roger Tomlinson and colleagues for Canadian Land Inventory. This project pioneers much technology and introduces the term GIS.
1963	General	URISA established	The Urban and Regional Information Systems Association founded in the US. Soon becomes point of interchange for GIS innovators.
1964	Academic	Harvard Lab established	The Harvard Laboratory for Computer Graphics and Spatial Analysis is established under the direction of Howard Fisher at Harvard University. In 1966 SYMAP, the first raster GIS, is created by Harvard researchers.
1967	Technology	DIME developed	The US Bureau of Census develops DIME-GBF (Dual Independent Map Encoding – Geographic Database Files), a data structure and street-address database for 1970 census.
1967	Academic and general	UK Experimental Cartography Unit (ECU) formed	Pioneered in a range of computer cartography and GIS areas.
1969	Commercial	ESRI Inc. formed	Jack Dangermond, a student from the Harvard Lab, and his wife Laura form ESRI to undertake projects in GIS.
1969	Commercial	Intergraph Corp. formed	Jim Meadlock and four others that worked on guidance systems for Saturn rockets form M&S Computing, later renamed Intergraph.
1969	Academic	'Design With Nature' published	Ian McHarg's book was the first to describe many of the concepts in modern GIS analysis, including the map overlay process (see Chapter 14).
1969	Academic	First technical GIS textbook	Nordbeck and Rystedt's book detailed algorithms and software they developed for spatial analysis.
1972	Technology	Landsat 1 launched	Originally named ERTS (Earth Resources Technology Satellite), this was the first of many major Earth remote sensing satellites to be launched.
1973	General	First digitizing production line	Set up by Ordnance Survey, Britain's national mapping agency.
1974	Academic	AutoCarto 1 Conference	Held in Reston, Virginia, this was the first in an important series of conferences that set the GIS research agenda.
1976	Academic	GIMMS now in worldwide use	Written by Tom Waugh (a Scottish academic), this vector-based mapping and analysis system was run at 300 sites worldwide.
1977	Academic	Topological Data Structures conference	Harvard Lab organizes a major conference and develops the ODYSSEY GIS.
		The Era of Commercialization	
1981	Commercial	ArcInfo launched	ArcInfo was the first major commercial GIS software system. Designed for minicomputers and based on the vector and relational database data model, it set a new standard for the industry.

(*continued overleaf*)

Table 1.4 (*continued*)

Date	Type	Event	Notes
1984	Academic	'Basic Readings in Geographic Information Systems' published	This collection of papers published in book form by Duane Marble, Hugh Calkins, and Donna Peuquet was the first accessible source of information about GIS.
1985	Technology	GPS operational	The Global Positioning System gradually becomes a major source of data for navigation, surveying, and mapping.
1986	Academic	'Principles of Geographical Information Systems for Land Resources Assessment' published	Peter Burrough's book was the first specifically on GIS principles. It quickly became a worldwide reference text for GIS students.
1986	Commercial	MapInfo Corp. formed	MapInfo software develops into first major desktop GIS product. It defined a new standard for GIS products, complementing earlier software systems.
1987	Academic	*International Journal of Geographical Information Systems*, now *IJGI Science*, introduced	Terry Coppock and others published the first journal on GIS. The first issue contained papers from the USA, Canada, Germany, and UK.
1987	General	Chorley Report	'Handling Geographical Information' was an influential report from the UK government that highlighted the value of GIS.
1988	General	*GISWorld* begins	*GISWorld*, now *GeoWorld*, the first worldwide magazine devoted to GIS, was published in the USA.
1988	Technology	TIGER announced	TIGER (Topologically Integrated Geographic Encoding and Referencing), a follow-on from DIME, is described by the US Census Bureau. Low-cost TIGER data stimulate rapid growth in US business GIS.
1988	Academic	US and UK Research Centers announced	Two separate initiatives, the US NCGIA (National Center for Geographic Information and Analysis) and the UK RRL (Regional Research Laboratory) Initiative show the rapidly growing interest in GIS in academia.
1991	Academic	*Big Book 1* published	Substantial two-volume compendium *Geographical Information Systems*; *principles and applications*, edited by David Maguire, Mike Goodchild, and David Rhind documents progress to date.
1992	Technical	DCW released	The 1.7 GB Digital Chart of the World, sponsored by the US Defense Mapping Agency, (now NGA), is the first integrated 1:1 million scale database offering global coverage.
1994	General	Executive Order signed by President Clinton	Executive Order 12906 leads to creation of US National Spatial Data Infrastructure (NSDI), clearinghouses, and Federal Geographic Data Committee (FGDC).
1994	General	OpenGIS® Consortium born	The OpenGIS® Consortium of GIS vendors, government agencies, and users is formed to improve interoperability.
1995	General	First complete national mapping coverage	Great Britain's Ordnance Survey completes creation of its initial database – all 230 000 maps covering country at largest scale (1:1250, 1:2500 and 1:10 000) encoded.
1996	Technology	Internet GIS products introduced	Several companies, notably Autodesk, ESRI, Intergraph, and MapInfo, release new generation of Internet-based products at about the same time.

Table 1.4 (*continued*)

Date	Type	Event	Notes
1996	Commercial	MapQuest	Internet mapping service launched, producing over 130 million maps in 1999. Subsequently purchased by AOL for $1.1 billion.
1999	General	GIS Day	First GIS Day attracts over 1.2 million global participants who share an interest in GIS.
		The Era of Exploitation	
1999	Commercial	IKONOS	Launch of new generation of satellite sensors: IKONOS claims 90 centimeter ground resolution; Quickbird (launched 2001) claims 62 cm resolution.
2000	Commercial	GIS passes $7 bn	Industry analyst Daratech reports GIS hardware, software, and services industry at $6.9 bn, growing at more than 10% per annum.
2000	General	GIS has 1 million users	GIS has more than 1 million core users, and there are perhaps 5 million casual users of GI.
2002	General	Launch of online National Atlas of the United States	Online summary of US national-scale geographic information with facilities for map making (**www.nationalatlas.gov**)
2003	General	Launch of online national statistics for the UK	Exemplar of new government websites describing economy, population, and society at local and regional scales (**www.statistics.gov.uk**)
2003	General	Launch of Geospatial One-Stop	A US Federal E-government initiative providing access to geospatial data and information (**www.geodata.gov/gos**)
2004	General	National Geospatial-Intelligence Agency (NGA) formed	Biggest GIS user in the world, National Imagery and Mapping Agency (NIMA), renamed NGA to signify emphasis on geo-intelligence

mouse clicks in their desktop WWW browser, without ever needing to install specialized software or download large amounts of data. This research project soon gave way to industrial-strength Internet GIS software products from mainstream software vendors (see Chapter 7).

The use of the WWW to give access to maps dates from 1993.

The recent histories of GIS and the Internet have been heavily intertwined; GIS has turned out to be a compelling application that has prompted many people to take advantage of the Web. At the same time, GIS has benefited greatly from adopting the Internet paradigm and the momentum that the Web has generated. Today there are many successful applications of GIS on the Internet, and we have used some of them as examples and illustrations at many points in this book. They range from using GIS on the Internet to disseminate information – a type of electronic yellow pages – (e.g., **www.yell.com**), to selling goods and services (e.g., **www.landseer.com.sg**, Figure 1.14), to direct revenue generation through subscription services (e.g., **www.mapquest.com/solutions/main.adp**), to helping members of the public to participate in important local, regional, and national debates.

The Internet has proven very popular as a vehicle for delivering GIS applications for several reasons. It is an established, widely used platform and accepted standard for interacting with information of many types. It also offers a relatively cost-effective way of linking together distributed users (for example, telecommuters and office workers, customers and suppliers, students and teachers). The interactive and exploratory nature of navigating linked information has also been a great hit with users. The availability of geographically enabled multi-content site gateways (geoportals) with powerful search engines has been a stimulus to further success.

Internet technology is also increasingly portable – this means not only that portable GIS-enabled devices can be used in conjunction with the wireless networks available in public places such as airports and railway stations, but also that such devices may be connected through broadband in order to deliver GIS-based representations on the move. This technology is being exploited in the burgeoning GIService (yet another use of the three-letter acronym GIS) sector, which offers distributed users access to centralized GIS capabilities. Later (Chapter 18 and onwards) we use the term *g-business* to cover all the myriad applications carried out in enterprises in different sectors that have a strong geographical component. The

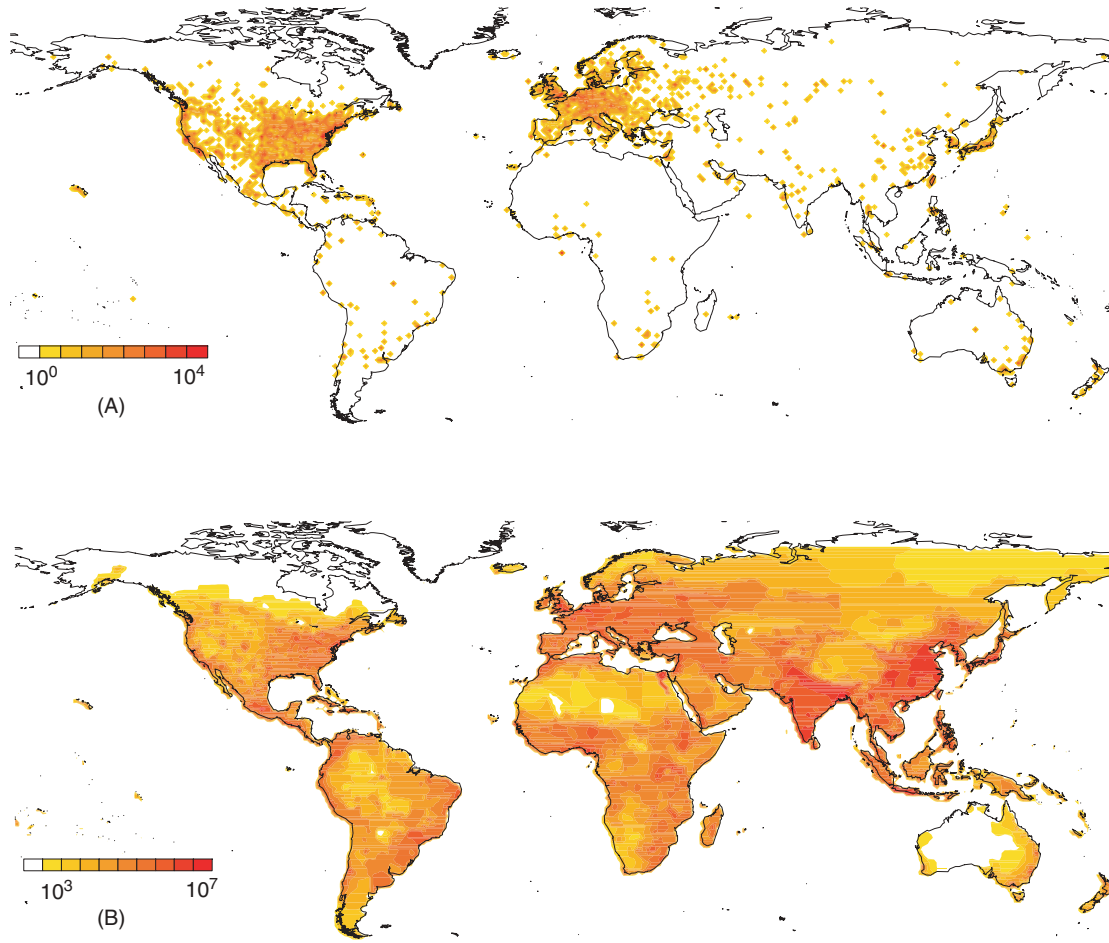


Figure 1.13 (A) The density of Internet hosts (routers) in 2002, a useful surrogate for Internet activity. The bar next to the map gives the range of values encoded by the color code per box (pixel) in the map. (B) This can be compared with the density of population, showing a strong correlation with Internet access in economically developed countries: elsewhere Internet access is sparse and is limited to urban areas. Both maps have a resolution of $1^\circ \times 1^\circ$. (Courtesy Yook S.-H., Jeong H. and Barabsi A.-L. 2002. 'Modeling the Internet's large-scale topology,' *Proceedings of the National Academy of Sciences* **99**, 13382–13386. See **www.nd.edu/~networks/PDF/Modeling%202002.pdf**) (Reproduced with permission of National Academy of Sciences, USA)

more restrictive term *g-commerce* is also used to describe types of electronic commerce (e-commerce) that include location as an essential element. Many GIServices are made available for personal use through mobile and handheld applications as *location-based services* (see Chapter 11). Personal devices, from pagers to mobile phones to Personal Digital Assistants, are now filling the briefcases and adorning the clothing of people in many walks of life (Figure 1.15). These devices are able to provide real-time geographic services such as mapping, routing, and geographic yellow pages. These services are often funded through advertisers, or can be purchased on a pay-as-you go or subscription basis, and are beginning to change the business GIS model for many types of applications.

A further interesting twist is the development of themed geographic networks, such as the US Geospatial One-Stop (**www.geo-one-stop.gov/**: see Box 11.4), which is one of 24 federal e-government initiatives to improve the coordination of government at local, state, and Federal levels. Its geoportal (**www.geodata.gov/gos**) identifies an integrated collection of geographic information providers and users that interact via the medium of the Internet. On-line content can be located using the interactive search capability of the portal and then content can be directly used over the Internet. This form of Internet application is explored further in Chapter 11.

The Internet is increasingly integrated into many aspects of GIS use, and the days of standalone GIS are mostly over.

1.4.3.2 The other five components of the GIS anatomy

The second piece of the GIS anatomy (Figure 1.16) is the user's hardware, the device that the user interacts with

Figure 1.14 Niche marketing of residential property in Singapore (**www.landseer.com.sg**) (Reproduced with permission of Landseer Property Services Pte Ltd.)

Figure 1.15 Wearable computing and personal data assistants are key to the diffusion and use of location-based services

directly in carrying out GIS operations, by typing, pointing, clicking, or speaking, and which returns information by displaying it on the device's screen or generating meaningful sounds. Traditionally this device sat on an office desktop, but today's user has much more freedom, because GIS functions can be delivered through laptops, personal data assistants (PDAs), in-vehicle devices, and even cellular telephones. Section 11.3 discusses the currently available technologies in greater detail. In the language of the network, the user's device is the *client*, connected through the network to a *server* that is probably handling many other user clients simultaneously. The client may be *thick*, if it performs a large part of the work locally, or *thin* if it does little more than link the user to the server. A PC or Macintosh is an instance of a thick client, with powerful local capabilities, while devices attached to TVs that offer little more than Web browser capabilities are instances of thin clients.

The third piece of the GIS anatomy is the software that runs locally in the user's machine. This can be as simple as a standard Web browser (Microsoft Explorer or Netscape) if all work is done remotely using assorted digital services offered on large servers. More likely it is a package bought from one of the GIS vendors, such as Intergraph Corp. (Huntsville, Alabama, USA; **www.ingr.com**), Environmental Systems Research Institute (ESRI; Redlands, California,

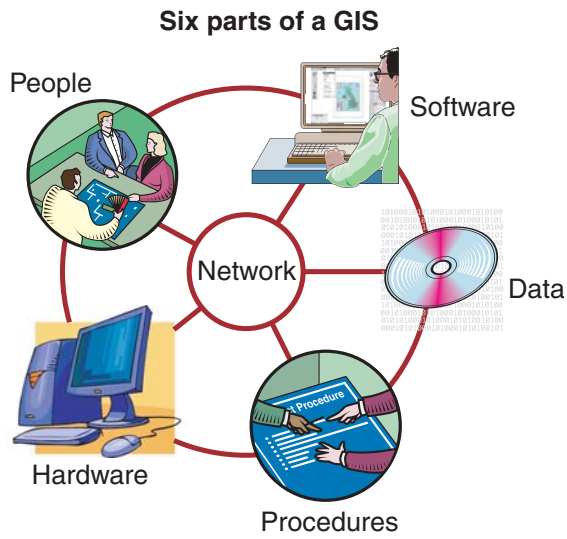


Figure 1.16 The six component parts of a GIS

USA; **www.esri.com**), Autodesk Inc. (San Rafael, California, USA; **www.autodesk.com**), or MapInfo Corp. (Troy, New York, USA; **www.mapinfo.com**). Each vendor offers a range of products, designed for different levels of sophistication, different volumes of data, and different application niches. IDRISI (Clark University, Worcester, Massachusetts, USA, **www.clarklabs.org**) is an example of a GIS produced and marketed by an academic institution rather than by a commercial vendor.

Many GIS tasks must be performed repeatedly, and GIS designers have created tools for capturing such repeated sequences into easily executed *scripts* or *macros* (Section 16.3.1) For example, the agency that needs to predict erosion of New South Wales's soils (Section 1.3) would likely establish a standard script written in the scripting language of its favorite GIS. The instructions in the script would tell the GIS how to model erosion given required data inputs and parameters, and how to output the results in suitable form. Scripts can be used repeatedly, for different areas or for the same area at different times. Support for scripts is an important aspect of GIS software.

GIS software can range from a simple package designed for a PC and costing a few hundred dollars, to a major industrial-strength workhorse designed to serve an entire enterprise of networked computers, and costing tens of thousands of dollars. New products are constantly emerging, and it is beyond the scope of this book to provide a complete inventory.

The fourth piece of the anatomy is the database, which consists of a digital representation of selected aspects of some specific area of the Earth's surface or near-surface, built to serve some problem solving or scientific purpose. A database might be built for one major project, such as the location of a new high-voltage power transmission corridor, or it might be continuously maintained, fed by the daily transactions that occur in a major utility company (installation of new underground pipes, creation of new customer accounts, daily service crew activities). It might be as small as a few megabytes (a few million bytes, easily stored on a few diskettes), or as large as a terabyte (roughly a trillion bytes, occupying a storage unit as big as a small office). Table 1.1 gives some sense of potential GIS database volumes.

GIS databases can range in size from a megabyte to a petabyte.

In addition to these four components – network, hardware, software, and database – a GIS also requires management. An organization must establish procedures, lines of reporting, control points, and other mechanisms for ensuring that its GIS activities stay within budgets, maintain high quality, and generally meet the needs of the organization. These issues are explored in Chapters 18, 19, and 20.

Finally, a GIS is useless without the people who design, program, and maintain it, supply it with data, and interpret its results. The people of GIS will have various skills, depending on the roles they perform. Almost all will have the basic knowledge needed to work with geographic data – knowledge of such topics as data sources, scale and accuracy, and software products – and will also have a network of acquaintances in the GIS community. We refer to such people in this book as *spatially aware professionals*, or SAPs, and the humor in this term is not intended in any way to diminish their importance, or our respect for what they know – after all, we would like to be recognized as SAPs ourselves! The next section outlines some of the roles played by the people of GIS, and the industries in which they work.

1.5 The business of GIS

Very many people play many roles in GIS, from software development to software sales, and from teaching about GIS to using its power in everyday activities. GIS is big business, and this section looks at the diverse roles that people play in the business of GIS, and is organized by the major areas of human activity associated with GIS.

1.5.1 The software industry

Perhaps the most conspicuous sector, although by no means the largest either in economic or human terms, is the GIS software industry. Some GIS vendors have their roots in other, larger computer applications: thus Intergraph and Autodesk, have roots in computer-assisted design software developed for engineering and architectural applications; and Leica Geosystems (ERDAS IMAGINE: **gis.leica-geosystems.com**) and PCI (**www.pcigeomatics.com**) have roots in remote sensing and image processing. Others began as specialists in GIS. Measured in economic terms, the GIS software industry currently accounts for over $1.8 billion in annual sales, although estimates vary, in part because of the

difficulty of defining GIS precisely. The software industry employs several thousand programmers, software designers, systems analysts, application specialists, and sales staff, with backgrounds that include computer science, geography, and many other disciplines.

The GIS software industry accounts for about $1.8 billion in annual sales.

1.5.2 The data industry

The acquisition, creation, maintenance, dissemination, and sale of GIS data also account for a large volume of economic activity. Traditionally, a large proportion of GIS data have been produced centrally, by national mapping agencies such as Great Britain's Ordnance Survey. In most countries the funds needed to support national mapping come from sales of products to customers, and sales now account for almost all of the Ordnance Survey's annual turnover of approximately $200 million. But federal government policy in the US requires that prices be set at no more than the cost of reproduction of data, and sales are therefore only a small part of the income of the US Geological Survey, the nation's premier civilian mapping agency.

In value of annual sales, the GIS data industry is much more significant than the software industry.

In recent years improvements in GIS and related technologies, and reductions in prices, along with various kinds of government stimulus, have led to the rapid growth of a private GIS data industry, and to increasing interest in data sales to customers in the public sector. In the socioeconomic realm, there is continuing investment in the creation and updating of general-purpose geodemographic indicators (Section 2.3.3), created using private sector datasets alongside traditional socioeconomic sources such as the Census. For example, UK data warehouse Experian's (Nottingham, UK) 2003 Mosaic product comprises 54% census data, with the balance of 46% coming from private sector sources and spatial indicators created using GIS. Data may also be packaged with software in order to offer integrated solutions, as with ESRI's Business Analyst product. Private companies are now also licensed to collect high-resolution data using satellites, and to sell it to customers – Space Imaging (**www.spaceimaging.com**) and its IKONOS satellite are a prominent instance (see Table 1.4). Other companies collect similar data from aircraft. Still other companies specialize in the production of high-quality data on street networks, a basic requirement of many delivery companies. Tele Atlas (**www.teleatlas.com** and its North American subsidiary, Geographic Data Technology **www.geographic.com**) is an example of this industry, employing some 1850 staff in producing, maintaining, and marketing high-quality street network data in Europe and North America.

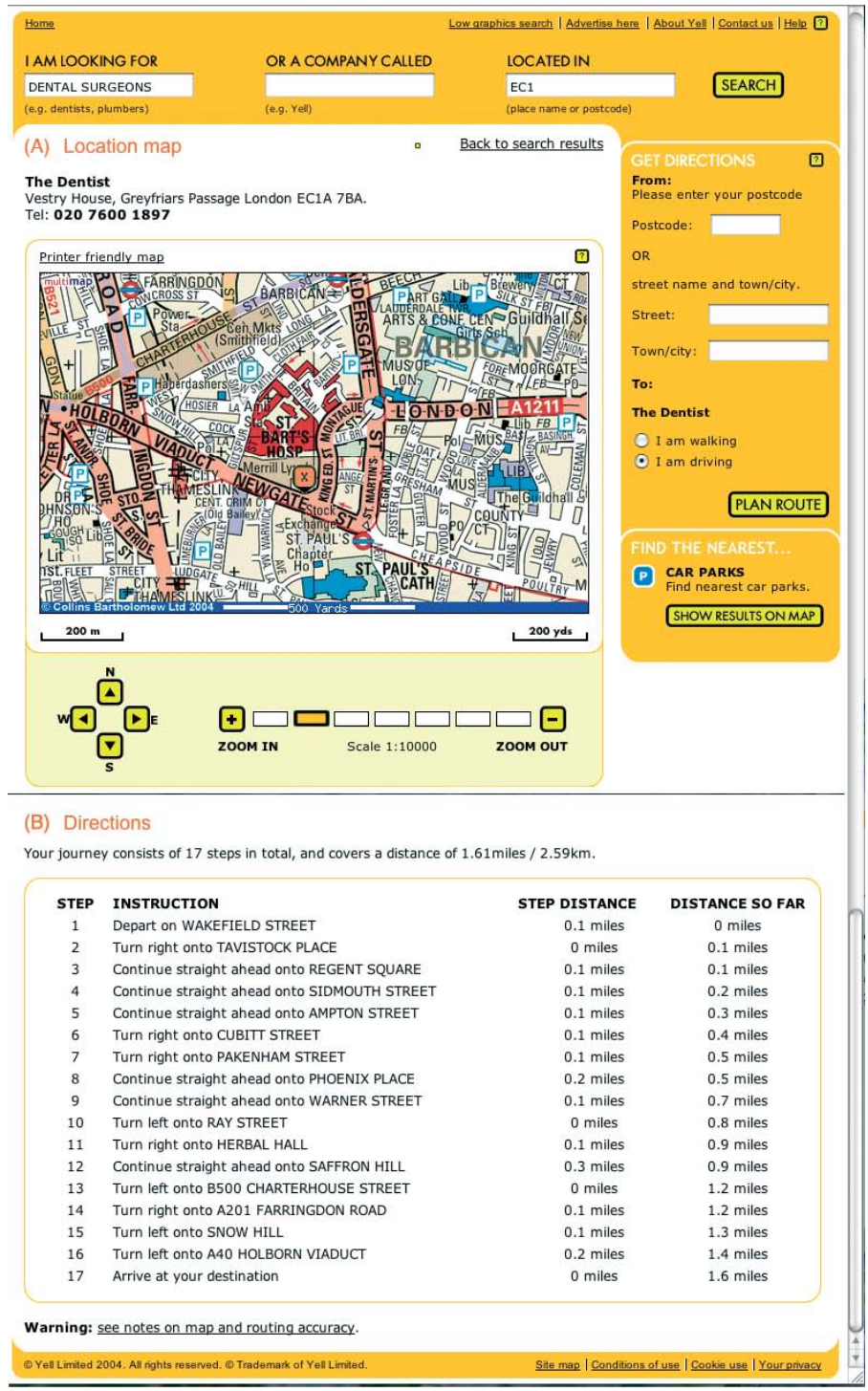
As developments in the information economy gather pace, many organizations are becoming focused upon delivering integrated business solutions rather than raw or value-added data. The Internet makes it possible for GIS users to access routinely collected data from sites that may be remote from locations where more specialized analysis and interpretation functions are performed. In these circumstances, it is no longer incumbent upon an organization to manage either its own data, or those that it buys in from value-added resellers. For example, ESRI offers a data management service, in which client data are managed and maintained for a range of clients that are at liberty to analyze them in quite separate locations. This may lead to greater vertical integration of the software and data industry – ESRI has developed an e-bis division and acquired its own geodemographic system (called Tapestry) to service a range of business needs. As GIS-based data handling becomes increasingly commonplace, so GIS is finding increasing application in new areas of public sector service provision, particularly where large amounts of public money are disbursed at the local level – as in policing, education provision, and public health. Many data warehouses and start-up organizations are beginning to develop public sector data infrastructures particularly where greater investment in public services is taking place.

1.5.3 The GIService industry

The Internet also allows GIS users to access specific functions that are provided by remote sites. For example, the US MapQuest site (**www.mapquest.com**) or the UK Yellow Pages site (**www.yell.com**) provide routing services that are used by millions of people every day to find the best driving route between two points. By typing a pair of street addresses, the user can execute a routing analysis (see Section 15.3.2) and receive the results in the form of a map and a set of written driving or walking directions (see Figure 1.17B). This has several advantages over performing the same analysis on one's own PC – there is no need to buy software to perform the analysis, there is no need to buy the necessary data, and the data are routinely updated by the GIService provider. There are clear synergies of interest between GIService providers and organizations providing location-based services (Section 1.4.3.1 and Chapter 11), and both activities are part of what we will describe as g-business in Chapter 19. Many sites that provide access to raw GIS data also provide GIServices.

GIServices are a rapidly growing form of electronic commerce.

GIServices continue to develop rapidly. In today's world one of the most important commodities is attention – the fraction of a second of attention given to a billboard, or the audience attention that a TV station sells to its advertisers. The value of attention also depends on the degree of fit between the message and the recipient – an advertiser will pay more for the attention of a small number of people if it knows that they include a large proportion of its typical customers. Advertising directed at the individual, based on an individual



STEP	INSTRUCTION	STEP DISTANCE	DISTANCE SO FAR
1	Depart on WAKEFIELD STREET	0.1 miles	0 miles
2	Turn right onto TAVISTOCK PLACE	0 miles	0.1 miles
3	Continue straight ahead onto REGENT SQUARE	0.1 miles	0.1 miles
4	Continue straight ahead onto SIDMOUTH STREET	0.1 miles	0.2 miles
5	Continue straight ahead onto AMPTON STREET	0.1 miles	0.3 miles
6	Turn right onto CUBITT STREET	0.1 miles	0.4 miles
7	Turn right onto PAKENHAM STREET	0.1 miles	0.5 miles
8	Continue straight ahead onto PHOENIX PLACE	0.2 miles	0.5 miles
9	Continue straight ahead onto WARNER STREET	0.1 miles	0.7 miles
10	Turn left onto RAY STREET	0 miles	0.8 miles
11	Turn right onto HERBAL HALL	0.1 miles	0.9 miles
12	Continue straight ahead onto SAFFRON HILL	0.3 miles	0.9 miles
13	Turn left onto B500 CHARTERHOUSE STREET	0 miles	1.2 miles
14	Turn right onto A201 FARRINGDON ROAD	0.1 miles	1.2 miles
15	Turn left onto SNOW HILL	0.1 miles	1.3 miles
16	Turn left onto A40 HOLBORN VIADUCT	0.2 miles	1.4 miles
17	Arrive at your destination	0 miles	1.6 miles

Figure 1.17 A GIS-enabled London electronic yellow pages: (A) location map of a dentist near St. Paul's Cathedral; and (B) written directions of how to get there from University College London Department of Geography

profile, is even more attractive to the advertiser. Direct-mail companies have exploited the power of geographic location to target specific audiences for many years, basing their strategies on neighborhood profiles constructed from census records. But new technologies offer to take this much further. For example, the technology already exists to identify the buying habits of a customer who stops at a gas pump and uses a credit card, and to direct targeted advertising through a TV screen at the pump.

1.5.4 The publishing industry

Much smaller, but nevertheless highly influential in the world of GIS, is the publishing industry, with its magazines, books, and journals. Several magazines are directed at the GIS community, as well as some increasingly significant news-oriented websites (see Box 1.4).

Several journals have appeared to serve the GIS community, by publishing new advances in GIS research. The oldest journal specifically targeted at the community is the *International Journal of Geographical Information Science*, established in 1987. Other older journals in areas such as cartography now regularly accept GIS articles, and several have changed their names and shifted focus significantly. Box 1.5 gives a list of the journals that emphasize GIS research.

1.5.5 GIS education

The first courses in GIS were offered in universities in the early 1970s, often as an outgrowth of courses

Technical Box **1.4**

Magazines and websites offering GIS news and related services

ArcNews and ArcUser Magazine (published by ESRI), see **www.esri.com**

Directions Magazine (Internet-centered and weekly newsletter publication by **directionsmag.com**), available online at **www.directionsmag.com**

GEO:connexion UK Magazine published quarterly by GEO:connexion Ltd., with website at **www.geoconnexion.com**

GEOInformatics published eight times a year by Cmedia Productions BV, with website at **www.geoinformatics.com**

GeoSpatial Solutions (published monthly by Advanstar Communications), and see their website **www.geospatial-online.com**. The company also publishes *GPSWorld GeoWorld* (published monthly by GEOTEC Media), available online at **www.geoplace.com**

GIS@development (published monthly for an Asian readership by GIS Development, India), with website at **www.GISDevelopment.net**

Spatial Business Online (published fortnightly in hard and electronic copy form by South Pacific Science Press), available online at **www.gisuser.com.au**

Some websites offering online resources for the GIS community:

www.gisdevelopment.net

www.geoconnexion.com

www.gis.com

www.giscafe.com

gis.about.com

www.geocomm.com and **www.spatialnews.com**

www.directionsmag.com

www.opengis.org/press/

Technical Box **1.5**

Some scholarly journals emphasizing GIS research

Annals of the Association of American Geographers

Cartography and Geographic Information Science

Cartography – The Journal

Computers and Geosciences

Computers, Environment and Urban Systems

Geographical Analysis

GeoInformatica

International Journal of Geographical Information Science (formerly *International Journal of Geographical Information Systems*)

ISPRS Journal of Photogrammetry and Remote Sensing

Journal of Geographical Systems

Photogrammetric Engineering and Remote Sensing (PE&RS)

Terra Forum

The Photogrammetric Record

Transactions in GIS

URISA Journal

Technical Box **1.6**

Sites offering Web-based education and training programs in GIS

- Birkbeck College (University of London) GIScOnline M.Sc. in Geographic Information Science at **www.bbk.ac.uk**
- City University (London) MGI – Masters in Geographic Information – a course with face-to-face or distance learning options at **www.city.ac.uk**
- Curtin University's distance learning programs in geographic information science at **www.cage.curtin.edu.au**
- ESRI's Virtual Campus at **campus.esri.com**
- Kingston Centre for GIS, Distance Learning Programme at **www.kingston.ac.uk**
- Pennsylvania State University Certificate Program in Geographic Information Systems at **www.worldcampus.psu.edu**
- UNIGIS International, Postgraduate Courses in GIS at **www.unigis.org**
- University of Southern California GIS distance learning certificate program at **www.usc.edu**

in cartography or remote sensing. Today, thousands of courses can be found in universities and colleges all over the world. Training courses are offered by the vendors of GIS software, and increasing use is made of the Web in various forms of remote GIS education and training (Box 1.6).

Often, a distinction is made between education and training in GIS – training in the use of a particular software product is contrasted with education in the fundamental principles of GIS. In many university courses, lectures are used to emphasize fundamental principles while computer-based laboratory exercises emphasize training. In our view, an education should be for life, and the material learned during an education should be applicable for as far into the future as possible. Fundamental principles tend to persist long after software has been replaced with new versions, and the skills learned in running one software package may be of very little value when a new technology arrives. On the other hand much of the fun and excitement of GIS comes from actually working with it, and fundamental principles can be very dry and dull without hands-on experience.

1.6 GISystems, GIScience, and GIStudies

Geographic information systems are useful tools, helping everyone from scientists to citizens to solve geographic problems. But like many other kinds of tools, such as computers themselves, their use raises questions that are sometimes frustrating, and sometimes profound. For example, how does a GIS user know that the results obtained are accurate? What principles might help a GIS user to design better maps? How can location-based services be used to help users to navigate and understand human and natural environments? Some of these are questions of GIS design, and others are about GIS data and methods. Taken together, we can think of them as questions that arise from the use of GIS – that are stimulated by exposure to GIS or to its products. Many of them are addressed in detail at many points in this book, and the book's title emphasizes the importance of both systems and science.

The term *geographic information science* was coined in a paper by Michael Goodchild published in 1992. In it, the author argued that these questions and others like them were important, and that their systematic study constituted a science in its own right. Information science studies the fundamental issues arising from the creation, handling, storage, and use of information – similarly, GIScience should study the fundamental issues arising from geographic information, as a well-defined class of information in general. Other terms have much the same meaning: *geomatics* and *geoinformatics*, *spatial information science*, *geoinformation engineering*. All suggest a scientific approach to the fundamental issues raised by the use of GIS and related technologies, though they all have different roots and emphasize different ways of thinking about problems (specifically geographic or more generally spatial, emphasizing engineering or science, etc.).

GIScience has evolved significantly in recent years. It is now part of the title of several renamed research journals (see Box 1.5), and the focus of the US University Consortium for Geographic Information Science (**www.ucgis.org**), an organization of roughly 60 research universities that engages in research agenda setting (Box 1.7), lobbying for research funding, and related activities. An international conference series on GIScience has been held in the USA biannually since 2000 (see **www.giscience.org**). The Varenius Project (**www.ncgia.org**) provides one disarmingly simple way to view developments in GIScience (Figure 1.18). Here, GIScience is viewed as anchored by three concepts – the individual, the computer, and society. These form the vertices of a triangle, and GIScience lies at its core. The various terms that are used to describe GIScience activity can be used to populate this triangle. Thus research about the individual is dominated by cognitive science, with its concern for understanding of spatial concepts,

Technical Box **1.7**

The 2002 research agenda of the US University Consortium for Geographic Information Science (www.ucgis.org), and related chapters in this book

1. **Long-term research challenges**
   a. Spatial ontologies (Chapters 3 and 6)
   b. Geographic representation (Chapter 3)
   c. Spatial data acquisition and integration (Chapters 9 and 10)
   d. Scale (Chapter 4)
   e. Spatial cognition (Chapter 3)
   f. Space and space/time analysis and modeling (Chapters 4, 14, 15, and 16)
   g. Uncertainty in geographic information (Chapter 6)
   h. Visualization (Chapters 12 and 13)
   i. GIS and society (Chapters 1 and 17)
   j. Geographic information engineering (Chapters 11 and 20)
2. **Short-term research priorities**
   a. GIS and decision making (Chapters 2, 17, and 18)
   b. Location-based services (Chapters 7 and 11)
   c. Social implications of LBS (Chapter 11)
   d. Identification of spatial clusters (Chapters 13 and 14)
   e. Geospatial semantic Web (Chapters 1 and 11)
   f. Incorporating remotely sensed data and information in GIS (Chapters 3 and 9)
   g. Geographic information resource management (Chapters 17 and 18)
   h. Emergency data acquisition and analysis (Chapter 9)
   i. Gradation and indeterminate boundaries (Chapter 6)
   j. Geographic information security (Chapter 17)
   k. Geospatial data fusion (Chapters 2 and 11)
   l. Institutional aspects of SDIs (Chapters 19 and 20)
   m. Geographic information partnering (Chapter 20)
   n. Geocomputation (Chapter 16)
   o. Global representation and modeling (Chapter 3)
   p. Spatialization (Chapters 3 and 13)
   q. Pervasive computing (Chapter 11)
   r. Geographic data mining and knowledge discovery (Chapter 14)
   s. Dynamic modeling (Chapter 16)

More detail on all of these topics, and additional topics added at more recent UCGIS assemblies, can be found at **www.ucgis.org/priorities/research/2002researchagenda.htm**

learning and reasoning about geographic data, and interaction with the computer. Research about the computer is dominated by issues of representation, the adaptation of new technologies, computation, and visualization. And finally, research about society addresses issues of impacts and societal context. Others have developed taxonomies of challenges facing the nascent discipline of GIScience, such as the US University Consortium for Geographic Information Science (Box 1.7). It is possible to imagine how the themes presented in Box 1.7 could be used to populate Figure 1.18 in relation to the three vertices of this triangle.

There are important respects in which GIScience is about using the software environment of GIS to redefine, reshape, and resolve pre-existing problems. Many of the research topics in GIScience are actually much older than GIS. The need for methods of spatial analysis, for example, dates from the first maps, and many methods were developed long before the first GIS appeared on the scene in the mid-1960s. Another way to look at GIScience is to see it as the body of knowledge

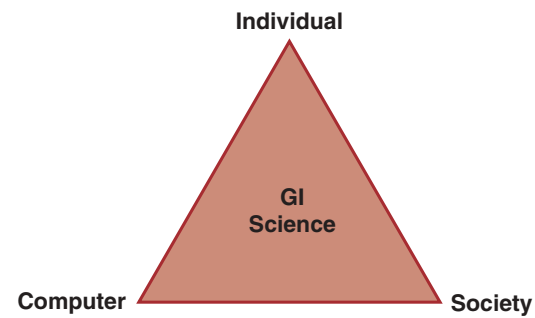


Figure 1.18 The remit of GIScience, according to Project Varenius (**www.ncgia.org**)

that GISystems implement and exploit. Map projections (Chapter 5), for example, are part of GIScience, and are used and transformed in GISystems. Another area of great importance to GIS is cognitive science, and

Biographical Box **1.8**

Reg Golledge, Behavioral Geographer

Figure 1.19 Reg Golledge, behavioral geographer

Reg Golledge was born in Australia but has worked in the US since completing his Ph.D. at the University of Iowa in 1966. He has worked at The Ohio State University (1967–1977), and since 1977 at the University of California, Santa Barbara (UCSB).

GIScience revisits many of the classic problems of spatial analysis, most of which assumed that people were rational and were optimizers in a very narrow sense. Over the last four decades, Reg's work has contributed much to our understanding of individual spatial behavior by relaxing these restrictive assumptions yet retaining the power of scientific generalization. Golledge's *analytical behavioral geography* has examined individual behavior using statistical and computational process models, particularly within the domain of transportation GIS (GIS-T: see Section 2.3.4), and has done much to make sense of the complexities and constraints that govern movement within urban systems. Related to this, analytical behavioral geography has also developed our understanding of individual cognitive awareness of urban networks and landmarks.

Reg's work is avowedly interdisciplinary. He has undertaken extensive work with cognitive psychologists at UCSB to develop personal guidance systems for use by visually-impaired travelers. This innovative work has linked GPS (for location and tracking) and GIS (for performing operations such as shortest path calculation, buffering, and orientation: see Chapters 14 and 15) with a novel auditory virtual system that presents users with the spatial relations between nearby environmental features. The device also allows users to personalize their representations of the environment.

Reg's enduring contribution to GIScience has been in modeling, explaining, and predicting disaggregate behaviors of individuals. This has been achieved through researching spatial cognition and cognitive science through GIS applications. He has established the importance of cognitive mapping to reasoning through GIScience, developed our understanding of the ways in which spatial concepts are embedded in GIS technology, and made vital contributions to the development of multimodal interfaces to GIS. These efforts have helped to develop new links to information science, information technology, and multimedia, and suggest ways of bridging the digital divides that threatens to further disadvantage disabled and elderly people. As a visually-impaired individual himself, Reg firmly believes that GIS technology and GIScience research are the most significant contributions that geography can make to truly integrated human and physical sciences, and sees a focus upon cognition as the natural bridge between these approaches to scientific inquiry.

particularly the scientific understanding of how people think about their geographic surroundings. If GISystems are to be easy to use they must fit with human ideas about such topics as driving directions, or how to construct useful and understandable maps. Box 1.8 introduces Reg Golledge, a quantitative and behavioral geographer who has brought diverse threads of cognitive science, transportation modeling, and analysis of geography and disability together under the umbrella of GIScience.

Many of the roots to GIS can be traced to the spatial analysis tradition in the discipline of geography.

In the 1970s it was easy to define or delimit a geographic information system – it was a single piece of software residing on a single computer. With time, and particularly with the development of the Internet and new approaches to software engineering, the old monolithic nature of GIS has been replaced by something much more fluid, and GIS is no longer an activity confined to the desktop (Chapter 11). The emphasis throughout this book is on this new vision of GIS, as the set of coordinated parts discussed earlier in Section 1.4. Perhaps the *system* part of GIS is no longer necessary – certainly the phrase *GIS data* suggests some redundancy, and many people have suggested that we could drop the 'S' altogether in favor of GI, for geographic information. GI*Systems* are only one part of the GI whole, which also includes the fundamental issues of GI*Science*. Much of this book is really about GI*Studies*, which can be defined as the systematic study of society's use of geographic information, including its institutions, standards, and procedures, and many of these

topics are addressed in the later chapters. Several of the UCGIS research topics suggest this kind of focus, including *GIS and society* and *geographic information partnering*. In recent years the role of GIS in society – its impacts and its deeper significance – has become the focus of extensive writing in the academic literature, particularly in the discipline of geography, and much of it has been critical of GIS. We explore these critiques in detail in the next section.

The importance of social context is nicely expressed by Nick Chrisman's definition of GIS which might also serve as an appropriate final comment on the earlier discussion of definitions:

> **The organized activity by which people: 1) measure aspects of geographic phenomena and processes; 2) represent these measurements, usually in the form of a computer database, to emphasize spatial themes, entities, and relationships; 3) operate upon these representations to produce more measurements and to discover new relationships by integrating disparate sources; and 4) transform these representations to conform to other frameworks of entities and relationships. These activities reflect the larger context (institutions and cultures) in which these people carry out their work. In turn, the GIS may influence these structures.**
>
> **(Chrisman 2003, p. 13)**

Chrisman's social structures are clearly part of the GIS whole, and as students of GIS we should be aware of the ethical issues raised by the technology we study. This is the arena of GIStudies.

1.7 GIS and geography

GIS has always had a special relationship to the academic discipline of geography, as it has to other disciplines that deal with the Earth's surface, including planning and landscape architecture. This section explores that special relationship and its sometimes tense characteristics. Non-geographers can conveniently skip this section, though much of its material might still be of interest.

Chapter 2 presents a gallery of successful GIS applications. This paints a picture of a field built around low-order concepts that actually stands in rather stark contrast to the scientific tradition in the academic discipline of geography. Here, the spatial analysis tradition has developed during the past 40 years around a range of more-sophisticated operations and techniques, which have a much more elaborate conceptual structure (see Chapters 14 through 16). One of the foremost proponents of the spatial analysis approach is Stewart Fotheringham, whose contribution is discussed in Box 1.9. As we will begin to see in Chapter 14, spatial analysis is the process by which we turn raw spatial data into useful spatial information. For the first half of its history, the principal focus of spatial analysis in most universities was upon development of theory, rather than working applications. Actual data were scarce, as were the means to process and analyze them.

In the 1980s GIS technology began to offer a solution to the problems of inadequate computation and limited data handling. However, the quite sensible priorities of vendors at the time might be described as solving the problems of 80% of their customers 80% of the time, and the integration of techniques based upon higher-order concepts was a low priority. Today's GIS vendors can probably be credited with solving the problems of at least 90% of their customers 90% of the time, and much of the remit of GIScience is to diffuse improved, curiosity-driven scientific understanding into the knowledge base of existing successful applications. But the drive towards improved applications has also been propelled to a significant extent by the advent of GPS and other digital data infrastructure initiatives by the late 1990s. New data handling technologies and new rich sources of digital data open up prospects for refocusing and reinvigorating academic interest in applied scientific problem solving.

Although repeat purchases of GIS technology leave the field with a buoyant future in the IT mainstream, there is enduring unease in some academic quarters about GIS applications and their social implications. Much of this unease has been expressed in the form of critiques, notably from geographers. John Pickles has probably contributed more to the debate than almost anyone else, notably through his 1993 edited volume *Ground Truth: The Social Implications of Geographic Information Systems*. Several types of arguments have surfaced:

- The ways in which GIS represents the Earth's surface, and particularly human society, favor certain phenomena and perspectives, at the expense of others. For example, GIS databases tend to emphasize homogeneity, partly because of the limited space available and partly because of the costs of more accurate data collection (see Chapters 3, 4, and 8). Minority views, and the views of individuals, can be submerged in this process, as can information that differs from the official or consensus view. For example, a soil map represents the geographic variation in soils by depicting areas of constant class, separated by sharp boundaries. This is clearly an approximation, and in Chapter 6 we explore the role of uncertainty in GIS. GIS often forces knowledge into forms more likely to reflect the view of the majority, or the official view of government, and as a result *marginalizes* the opinions of minorities or the less powerful.
- Although in principle it is possible to use GIS for any purpose, in practice it is often used for purposes that may be ethically questionable or may invade individual privacy, such as surveillance and the gathering of military and industrial intelligence. The technology may appear neutral, but it is always used in a social context. As with the debates over the

Biographical Box **1.9**

Stewart Fotheringham, geocomputation specialist

Figure 1.20 Stewart Fotheringham, quantitative geographer

There are many close synonyms for geographic information science (GIScience), one of which is *geocomputation* – a term first coined by the geographer Stan Openshaw to describe the scientific application of computationally-intensive techniques to problems with a spatial dimension. A. Stewart Fotheringham is Science Foundation Ireland Research Professor and Director of the National Centre for Geocomputation at the National University of Ireland in Maynooth. He is a spatial scientist who has considerable previous experience of the Anglo-American university systems – he has worked and studied at the Universities of Newcastle and Aberdeen in the UK, the State University of New York at Buffalo, the University of Florida, and Indiana University in the US, and McMaster University in Canada (Figure 1.20).

Like GIScience, geocomputation is fundamentally about satisfying human curiosity through systematic, scientific problem solving. Many of the roots to the scientific use of GIS in scientific problem solving can be traced to the 'Quantitative Revolution' in geography of the 1960s, which had the effect of popularizing systematic techniques of spatial analysis throughout the discipline – an approach that had its detractors then as well as now (Section 1.7). The Quantitative Revolution has not only bequeathed GIS a rich legacy of methods and techniques, but has also developed into a sustained concern for understanding the nature of spatial variations in relationships. The range of these methods and techniques is described in Stewart's 2000 book *Quantitative Geography: Perspectives on Spatial Data Analysis* (Sage, London: with co-researchers Chris Brunsdon and Martin Charlton), while spatial variations in relationships are considered in detail in his 2002 book *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (Wiley, Chichester: also with the same co-authors).

The methods and techniques that Stewart has developed and applied permeate the world of GIS applications that we consider in Chapter 2. Stewart remains evangelical about the importance of space and our need to use GIS to make spatial analysis sensitive to context. He says: 'We know that many spatially aggregated statements, such as the average temperature of the entire US on any given day, actually tell us very little. Yet when we seek to establish relationships between data, we all too often hypothesize that relationships are the same everywhere – that relationships are *spatially invariant*'. Stewart's work on geographically weighted regression (GWR) is part of a growing realization that relationships, or our measurements of them, can vary over space and that we need to investigate this potential non-stationarity further (see Chapter 4). Stewart's geocomputational approach is closely linked to GIS because it uses locational information as inputs and produces geocoded results as outputs that can be mapped and further analyzed. GWR exploits the property of spatial location to the full, and has led to geocomputational analysis of relationships by researchers working in many disciplines.

atomic bomb in the 1940s and 1950s, the scientists who develop and promote the use of GIS surely bear some responsibility for how it is eventually used. The idea that a tool can be inherently neutral, and its developers therefore immune from any ethical debates, is strongly questioned in this literature.

- The very success of GIS is a cause of concern. There are qualms about a field that appears to be led by technology and the marketplace, rather than by human need. There are fears that GIS has become *too* successful in modeling socioeconomic distributions, and that as a consequence GIS has become a tool of the 'surveillance society'.
- There are concerns that GIS remains a tool in the hands of the already powerful – notwithstanding the diffusion of technology that has accompanied the plummeting cost of computing and wide adoption of the Internet. As such, it is seen as maintaining the *status quo* in terms of power structures. By implication, any vision of GIS for all of society is seen as unattainable.
- There appears to be an absence of applications of GIS in *critical* research. This academic perspective is centrally concerned with the connections between human agency and particular social structures and contexts. Some of its protagonists are of the view that

such connections are not amenable to digital representation in whole or in part.

- Some view the association of GIS with the scientific and technical project as fundamentally flawed. More narrowly, there is a view that GIS applications are (like spatial analysis before it) inextricably bound to the philosophy and assumptions of the approach to science known as *logical positivism* (see also the reference to 'positive' in Section 1.1). As such, the argument goes, GIS can never be more than a logical positivist tool and a normative instrument, and cannot enrich other more critical perspectives in geography.

Many geographers remain suspicious of the use of GIS in geography.

We wonder where all this discussion will lead. For our own part, we have chosen a title that includes both systems and science, and certainly much more of this book is about the broader concept of geographic information than about isolated, monolithic software systems *per se*. We believe strongly that effective users of GIS require some awareness of *all* aspects of geographic information, from the basic principles and techniques to concepts of management and familiarity with applications. We hope this book provides that kind of awareness. On the other hand, we have chosen not to include GIStudies in the title. Although the later chapters of the book address many aspects of the social context of GIS, including issues of privacy, the context to GIStudies is rooted in social theory. GIStudies need the kind of focused attention that we cannot give, and we recommend that students interested in more depth in this area explore the specialized texts listed in the guide to further reading.

Questions for further study

1. Examine the geographic data available for the area within 50 miles (80 km) of either where you live or where you study. Use it to produce a short (2500 word) illustrated profile of either the socioeconomic or the physical environment. (See for example **www.geodata.gov/gos**; **www.geographynetwork.com**; **eu-geoportal.jrc.it**; or **www.magic.gov.uk**).
2. What are the distinguishing characteristics of the scientific method? Discuss the relevance of each to GIS.
3. We argued in Section 1.4.3.1 that the Internet has dramatically changed GIS. What are the arguments for and against this view?
4. Locate each of the issues identified in Box 1.7 in two triangular 'GIScience' diagrams like that shown in Figure 1.18 – one for long-term research challenges and one for short-term research priorities. Give short written reasons for your assignments. Compare the distribution of issues within each of your triangles in order to assess the relative importance of the individual, the computer, and society in the development of GIScience over both the short- and long-term.

Further reading

Chrisman N.R. 2003 *Exploring Geographical Information Systems* (2nd edn). Hoboken, NJ: Wiley.

Curry M.R. 1998 *Digital Places: Living with Geographic Information Technologies*. London: Routledge.

Foresman T.W. (ed) 1998 *The History of Geographic Information Systems: Perspectives from the Pioneers*. Upper Saddle River, NJ: Prentice Hall.

Goodchild M.F. 1992 'Geographical information science'. *International Journal of Geographical Information Systems* **6**: 31–45.

Longley P.A. and Batty M. (eds) 2003 *Advanced Spatial Analysis: The CASA Book of GIS*. Redlands, CA: ESRI Press.

Pickles J. 1993 *Ground Truth: The Social Implications of Geographic Information Systems*. New York: Guilford Press.

University Consortium for Geographic Information Science 1996 'Research priorities for geographic information science'. *Cartography and Geographic Information Systems* **23**(3): 115–127.

2 *A gallery of applications*

Fundamentally, GIS is about workable applications. This chapter gives a flavor of the breadth and depth of real-world GIS implementations. It considers:

- How GIS affects our everyday lives;
- How GIS applications have developed, and how the field compares with scientific practice;
- The goals of applied problem solving;
- How GIS can be used to study and solve problems in transportation, the environment, local government, and business.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After studying this chapter you will:

- Grasp the many ways in which we interact with GIS in everyday life;
- Appreciate the range and diversity of GIS applications in environmental and social science;
- Be able to identify many of the scientific assumptions that underpin real-world applications;
- Understand how GIS is applied in the representative application areas of transportation, the environment, local government, and business.

2.1 Introduction

2.1.1 One day of life with GIS

7:00 My alarm goes off... The energy to power the alarm comes from the local energy company, which uses a GIS to manage all its assets (e.g., electrical conductors, devices, and structures) so that it can deliver electricity continuously to domestic and commercial customers (Figure 2.1).

7:05 I jump in the shower... The water for the shower is provided by the local water company, which uses a hydraulic model linked to its GIS to predict water usage and ensure that water is always available to its valuable customers (Figure 2.2).

7:35 I open the mail... A property tax bill comes from a local government department that uses a GIS to store property data and automatically produce annual tax bills. This has helped the department to peg increases in property taxes to levels below retail price inflation.

There are also a small number of circulars addressed to me, sometimes called 'junk mail'. We spent our

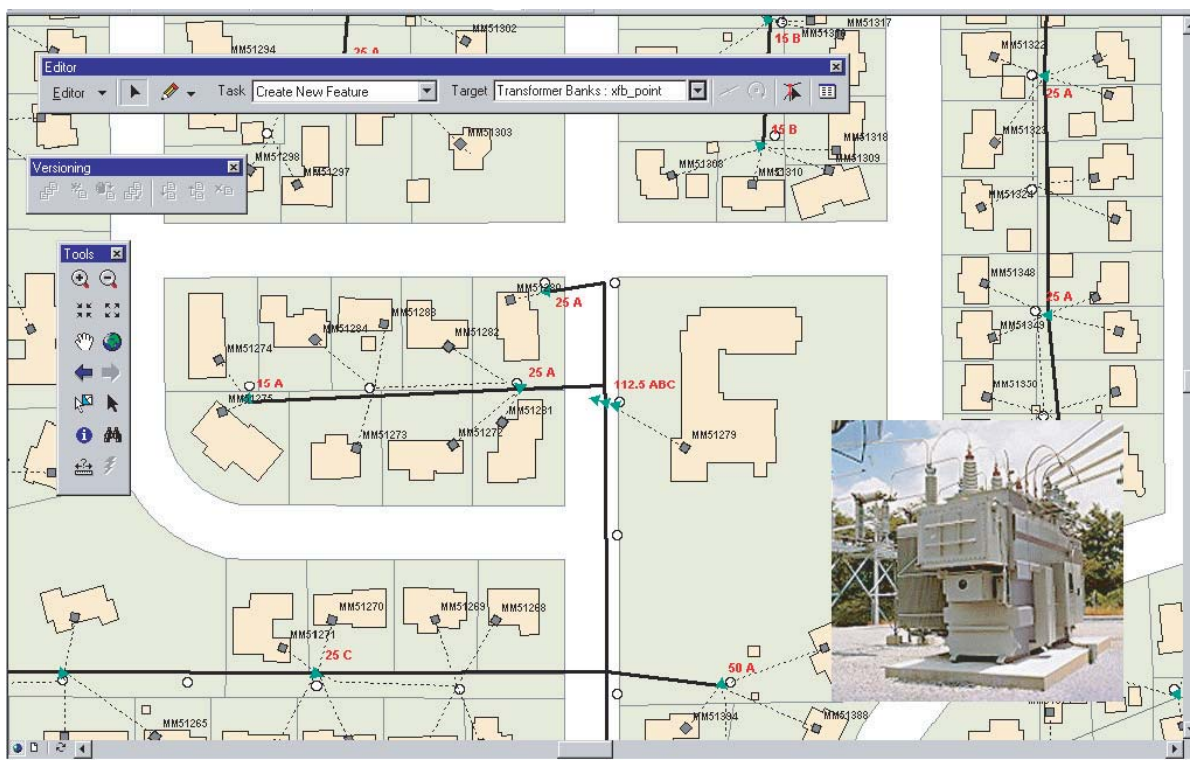


Figure 2.1 An electrical utility application of GIS

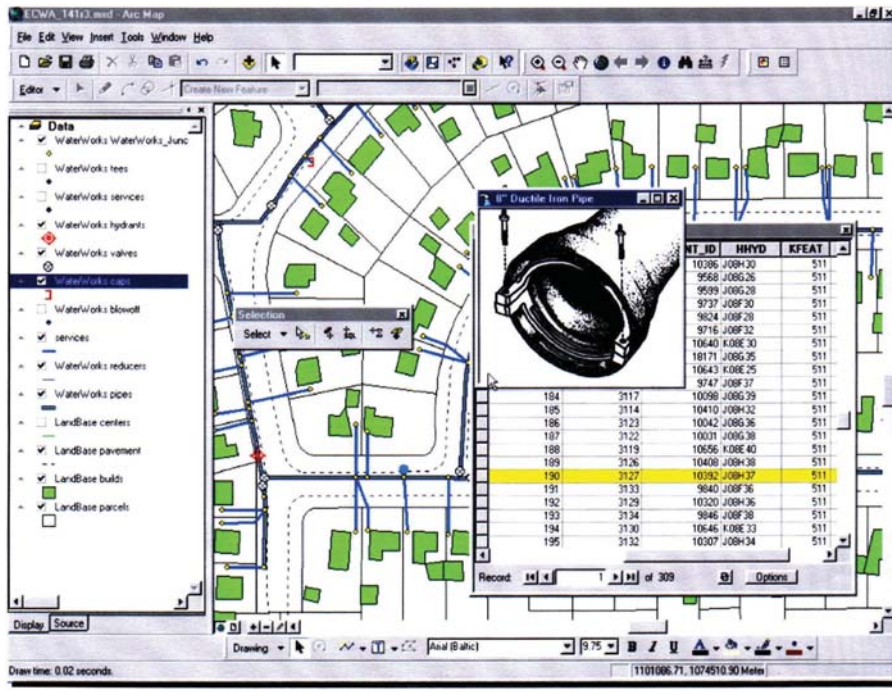


Figure 2.2 Application of a GIS for managing the assets of a water utility

vacation in Southlands and Santatol last year, and the holiday company uses its GIS to market similar destinations to its customer base – there are good deals for the Gower and Northampton this season. A second item is a special offer for property insurance, from a firm that uses its GIS to target neighborhoods with low past-claims histories. We get less junk mail than we used to (and we don't want to opt out of all programs), because geodemographic and lifestyles GIS is used to target mailings more precisely, thus reducing waste and saving time.

8:00 The other half leaves for work. . . He teaches GIS at one of the city community colleges. As a lecturer on one of the college's most popular classes he has a full workload and likes to get to work early.

8:15 I walk the kids to the bus stop. . . Our children attend the local middle school that is three miles away. The school district administrators use a GIS to optimize the routing of school buses (Figure 2.3). Introduction of this service enabled the district to cut their annual school busing costs by 16% and the time it takes the kids to get to school has also been reduced.

8:45 I catch a train to work. . . At the station the current location of trains is displayed on electronic maps on the platforms using a real-time feed from global positioning (GPS) receivers mounted on the trains. The same information is also available on the Internet so I was able to check the status of trains before I left the house.

9:15 I read the newspaper on the train. . . The paper for the newspaper comes from sustainable forests managed by a GIS. The forestry information system used by the forest products company indicates which areas are available for logging, the best access routes, and the likely yield (Figure 2.4).

9:30 I arrive at work. . . I am GIS Manager for the local City government. Today I have meetings to review annual budgets, plan for the next round of hardware and software acquisition, and deal with a nasty copyright infringement claim.

12:00 I grab a sandwich for lunch. . . The price of bread fell in real terms for much of the past decade. In some small part this is because of the increasing use of GIS in precision agriculture. This has allowed real-time mapping of soil nutrients and yield, and means that farmers can apply just the right amount of fertilizer in the right location and at the right time.

6:30 Shop till you drop. . . After work we go shopping and use some of the discount coupons that were in the morning mail. The promotion is to entice customers back to the renovated downtown Tesbury Center. We usually go to MorriMart on the far side of town, but thought we'd participate in the promotion. We actually bump into a few of our neighbors at Tesbury – I suspect the promotion was targeted by linking a marketing GIS to Tesbury's own store loyalty card data.

10:30 The kids are in bed. . . I'm on the Internet to try and find a new house. . . We live in a good neighborhood with many similarly articulate, well-educated folk, but it has become noisier since the new distributor road was routed close by. Our resident association mounted a vociferous campaign of protest, and its members filed numerous complaints to the website where the draft proposals were posted. But

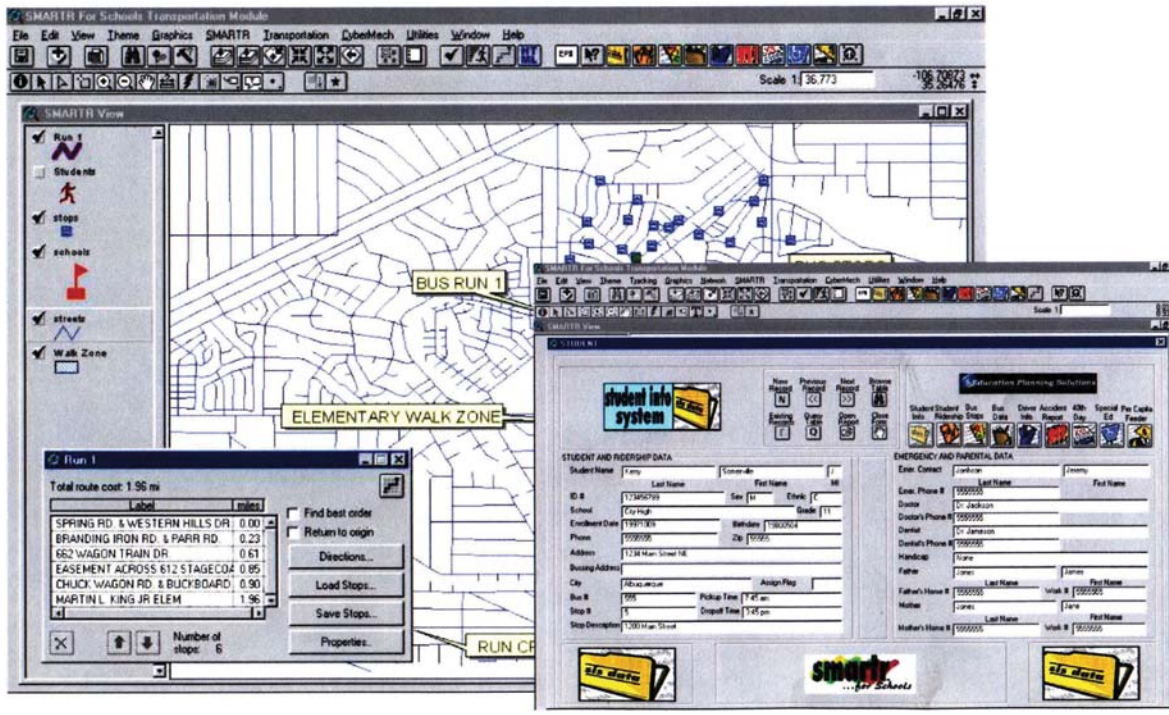


Figure 2.3 A GIS used for school bus routing

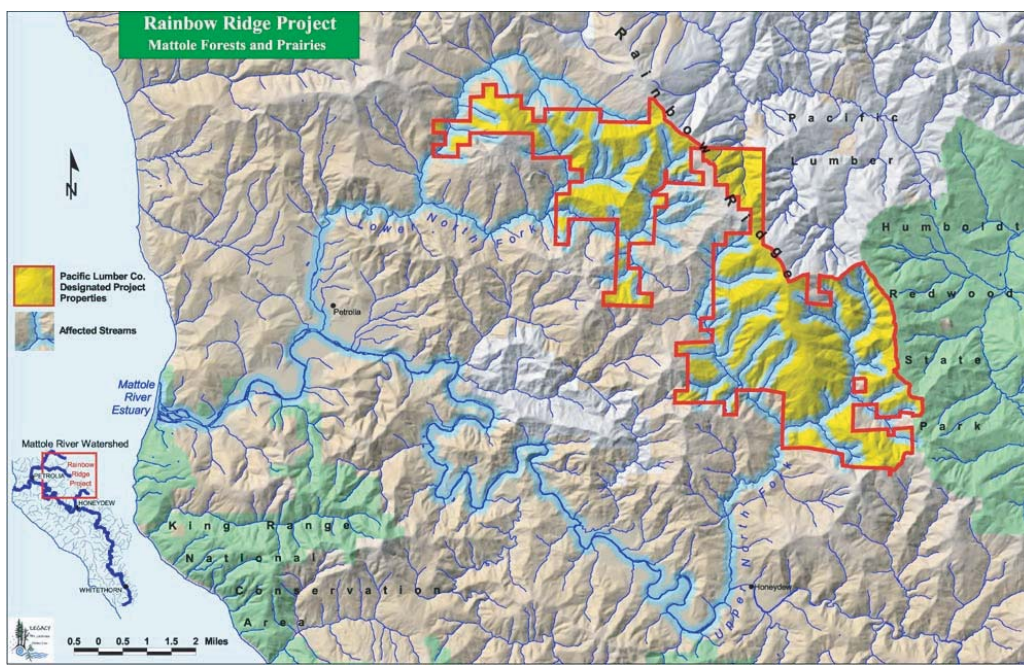


Figure 2.4 Forestry management GIS

the benefit-cost analysis carried out using the local authority's GIS clearly demonstrated that it was either a bit more noise for us, or the physical dissection of a vast swathe of housing elsewhere, and that we would have to grin and bear it. Post GIS, I guess that narrow interest NIMBY (Not In My Back Yard) protests don't get such a free run as they once did. So here I am using one of the on-line GIS-powered websites to find properties that match our criteria (similar to that in Figure 1.14). Once we have found a property, other mapping sites provide us with details about the local and regional facilities.

GIS is used to improve many of our day-to-day working and living arrangements.

This diary is fictitious of course, but most of the things described in it are everyday occurrences repeated hundreds and thousands of times around the world. It highlights a number of key things about GIS. GIS

- affects each of us, every day;
- can be used to foster effective short- and long-term decision making;
- has great practical importance;
- can be applied to many socio-economic and environmental problems;
- supports mapping, measurement, management, monitoring, and modeling operations;
- generates measurable economic benefits;
- requires key management skills for effective implementation;
- provides a challenging and stimulating educational experience for students;
- can be used as a source of direct income;
- can be combined with other technologies; and
- is a dynamic and stimulating area in which to work.

At the same time, the examples suggest some of the elements of the critique that has been leveled at GIS in recent years (see Section 1.7). Only a very small fraction of the world's population has access to information technologies of any kind, let alone high-speed access to the Internet. At the global scale, information technology can exacerbate the differences between developed and less-developed nations, across what has been called the *digital divide*, and there is also *digital differentiation* between rich and poor communities within nations. Uses of GIS for marketing often involve practices that border on invasion of privacy, since they allow massive databases to be constructed from what many would regard as personal information. It is important that we understand and reflect on issues like these while exploring GIS.

2.1.2 Why GIS?

Our day of life with GIS illustrates the unprecedented frequency with which, directly or indirectly, we interact with digital machines. Today, more and more individuals and organizations find themselves using GIS to answer the fundamental question, *where*? This is because of:

- Wider availability of GIS through the Internet, as well as through organization-wide local area networks.
- Reductions in the price of GIS hardware and software, because economies of scale are realized by a fast-growing market.
- Greater awareness that decision making has a geographic dimension.
- Greater ease of user interaction, using standard windowing environments.
- Better technology to support applications, specifically in terms of visualization, data management and analysis, and linkage to other software.
- The proliferation of geographically referenced digital data, such as those generated using Global Positioning System (GPS) technology or supplied by value-added resellers (VARs) of data.
- Availability of packaged applications, which are available commercially off-the-shelf (COTS) or 'ready to run out of the box'.
- The accumulated experience of applications that *work*.

2.2 Science, geography, and applications

2.2.1 Scientific questions and GIS operations

As we saw in Section 1.3, one objective of science is to solve problems that are of real-world concern. The range and complexity of scientific principles and techniques that are brought to bear upon problem solving will clearly vary between applications. Within the spatial domain, the goals of applied problem solving include, but are not restricted to:

- Rational, effective, and efficient allocation of resources, in accordance with clearly stated criteria – whether, for example, this entail physical construction of infrastructure in utilities applications, or scattering fertilizer in precision agriculture.
- Monitoring and understanding observed spatial distributions of attributes – such as variation in soil nutrient concentrations, or the geography of environmental health.
- Understanding the difference that *place* makes – identifying which characteristics are inherently similar between places, and what is distinctive and possibly unique about them. For example, there are regional and local differences in people's surnames (see Box 1.2), and regional variations in voting patterns are the norm in most democracies.
- Understanding of processes in the natural and human environments, such as processes of coastal erosion or river delta deposition in the natural environment, and understanding of changes in residential preferences or store patronage in the social.
- Prescription of strategies for environmental maintenance and conservation, as in national park management.

Understanding and resolving these diverse problems entails a number of general data handling operations – such

as inventory compilation and analysis, mapping, and spatial database management – that may be successfully undertaken using GIS.

GIS is fundamentally about solving real-world problems.

GIS has always been fundamentally an applications-led area of activity. The accumulated experience of applications has led to borrowing and creation of particular conventions for representing, visualizing, and to some extent analyzing data for particular classes of applications. Over time, some of these conventions have become useful in application areas quite different from those for which they were originally intended, and software vendors have developed general-purpose routines that may be customized in application-specific ways, as in the way that spatial data are visualized. The way that accumulated experience and borrowed practice becomes formalized into standard conventions makes GIS essentially an inductive field.

In terms of the definition and remit of GIScience (Section 1.6) the conventions used in applications are based on very straightforward concepts. Most data-handling operations are routine and are available as adjuncts to popular word-processing packages (e.g., Microsoft MapPoint: **www.microsoft.com/mappoint**). They work and are very widely used (e.g., see Figure 2.5), yet may not always be readily adaptable to scientific problem solving in the sense developed in Section 1.3.

2.2.2 GIScience applications

Early GIS was successful in depicting how the world looks, but shied away from most of the bigger questions concerning how the world works. Today GIScience is developing this extensive experience of applications into a bigger agenda – and is embracing a full range of conceptual underpinnings to successful problem solving.

GIS nevertheless remains fundamentally an applications-led technology, and many applications remain modest in both the technology that they utilize and the scientific tasks that they set out to accomplish. There is nothing fundamentally wrong with this, of course, as the most important test of geographic science and technology is whether or not it is useful for exploring and understanding the world around us. Indeed the broader relevance of geography as a discipline can only be sustained in relation to this simple goal, and no amount of scientific and technological ingenuity can salvage geographic representations of the world that are too inaccurate, expensive, cumbersome, or opaque to reveal anything new. In practice, this means that GIS applications must be grounded in sound concepts and theory if they are to resolve any but the most trivial of questions.

GIS applications need to be grounded in sound concepts and theory.

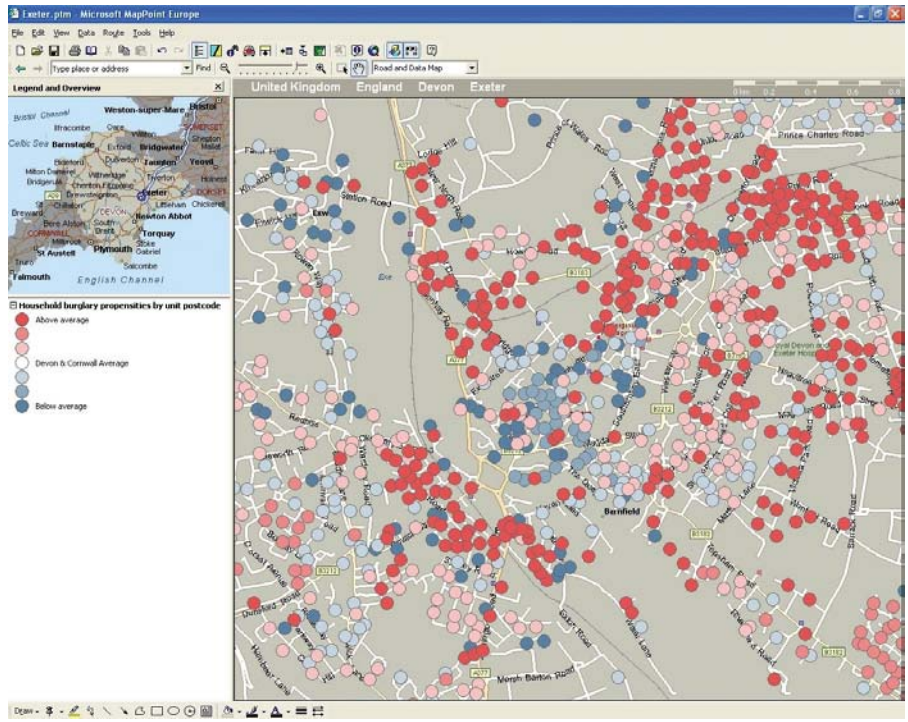


Figure 2.5 Microsoft MapPoint Europe mapping of spreadsheet data of burglary rates in Exeter, England using an adjunct to a standard office software package (courtesy D. Ashby. © 1988–2001 Microsoft Corp. and/or its suppliers. All rights reserved. © 2000 Navigation Technologies B.V. and its suppliers. All rights reserved. Selected Road Maps © 2000 by AND International Publishers N.V. All rights reserved. © Crown Copyright 2000. All rights reserved. License number 100025500. Additional demographic data courtesy of Experian Limited. © 2004 Experian Limited. All rights reserved.)

2.3 Representative application areas and their foundations

2.3.1 Introduction and overview

There is, quite simply, a huge range of applications of GIS, and indeed several pages of this book could be filled with a list of application areas. They include topographic base mapping, socio-economic and environmental modeling, global (and interplanetary!) modeling, and education. Applications generally set out to fulfill the five *Ms* of GIS: mapping, measurement, monitoring, modeling, management.

The five Ms of GIS application are mapping, measurement, monitoring, modeling, and management.

In very general terms, GIS applications may be classified as traditional, developing, and new. Traditional GIS application fields include military, government, education, and utilities. The mid-1990s saw the wide development of business uses, such as banking and financial services, transportation logistics, real estate, and market analysis. The early years of the 21st century are seeing new forward-looking application areas in small office/home office (SOHO) and personal or consumer applications, as well as applications concerned with security, intelligence, and counter-terrorism measures. This is a somewhat rough-and-ready classification, however, because the applications of some agencies (such as utilities) fall into more than one class.

A further way to examine trends in GIS applications is to examine the diffusion of GIS use. Figure 2.6 shows the classic model of GIS diffusion originally developed by Everett Rogers. Rogers' model divides the adopters of an innovation into five categories:

- Venturesome Innovators – willing to accept risks and sometimes regarded as oddballs.
- Respectable Early Adopters – regarded as opinion formers or 'role models'.
- Deliberate Early Majority – willing to consider adoption only after peers have adopted.
- Skeptical Late Majority – overwhelming pressure from peers is needed before adoption occurs.
- Traditional Laggards – people oriented to the past.

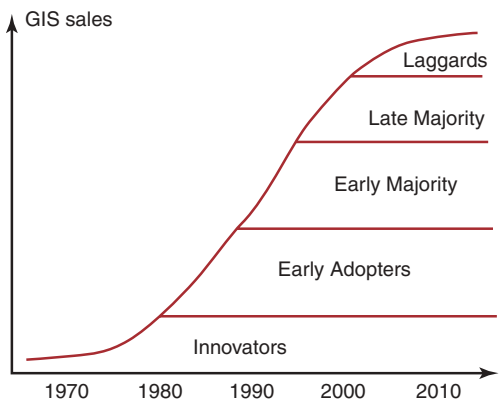


Figure 2.6 The classic Rogers model of innovation diffusion applied to GIS (After Rogers E.M. 2003 *Diffusion of Innovations* (5th edn). New York: Simon and Schuster.)

GIS is moving into the Late Majority stage, although some areas of application are more comprehensively developed than others. The Innovators who dominated the field in the 1970s were typically based in universities and research organizations. The Early Adopters were the users of the 1980s, many of whom were in government and military establishments. The Early Majority, typically in private businesses, came to the fore in the mid-1990s. The current question for potential users appears to be: do you want to gain competitive advantage by being part of the Majority user base or wait until the technology is completely accepted and contemplate joining the GIS community as a Laggard?

A wide range of motivations underpins the use of GIS, although it is possible to identify a number of common themes. Applications dealing with day-to-day issues typically focus on very practical concerns such as cost effectiveness, service provision, system performance, competitive advantage, and database creation, access, and use. Other, more strategic applications are more concerned with creating and evaluating scenarios under a range of circumstances.

Many applications involve use of GIS by large numbers of people. It is not uncommon for a large government agency, university, or utility to have more than 100 GIS seats, and a significant number have more than 1000. Once GIS applications become established within an organization, usage often spreads widely. Integration of GIS with corporate information system (IS) policy and with forward planning policy is an essential prerequisite for success in many organizations.

The scope of these applications is best illustrated with respect to representative application areas, and in the remainder of this chapter we consider:

1. Government and public service (Section 2.3.2)
2. Business and service planning (Section 2.3.3)
3. Logistics and transportation (Section 2.3.4)
4. Environment (Section 2.3.5)

We begin by identifying the range of applications within each of the four domains. Next, we go on to focus upon one application within each domain. Each application is chosen, first, for simplicity of exposition but also, second, for the scientific questions that it raises. In this book, we try to relate science and application in two ways. First, we flag the sections elsewhere in the book where the scientific issues raised by the applications are discussed. Second, the applications discussed here, and others like them, provide the illustrative material for our discussion of principles, techniques, analysis, and practices in the other chapters of the book.

A recurrent theme in each of the application classes is the importance of geographic location, and hence

what is *special* about the handling of georeferenced data (Section 1.1.1). The gallery of applications that we set out here intends to show how geographic data can provide crucial context to decision making.

2.3.2 Government and public service

2.3.2.1 Applications overview

Government users were among the first to discover the value of GIS. Indeed the first recognized GIS – the Canadian Geographic Information System (CGIS) – was developed for natural resource inventory and management by the Canadian government (see Section 1.4.1). CGIS was a national system and, unlike now, in the early days of GIS it was only national or federal organizations that could afford the technology. Today GIS is used at all levels of government from the national to the neighborhood, and government users still comprise the biggest single group of GIS professionals. It is helping to supplement traditional 'top down' government decision making with 'bottom up' representation of real communities in government decision making at all levels (Figure 2.7). We will see in later chapters how this deployment of GIS applications is consistent with greater supplementation of 'top down' deductivism with 'bottom up' inductivism in science. The importance of spatial variation to government and public service should not be underestimated – 70–80% of local government work should involve GIS in some way.

As GIS has become cheaper, so it has come to be used in government decision making at all levels from the nation to the neighborhood.

Today, local government organizations are acutely aware of the need to improve the quality of their products, processes, and services through ever-increasing efficiency of resource usage (see also Section 15.3). Thus GIS is used to inventory resources and infrastructure, plan transportation routing, improve public service delivery, manage land development, and generate revenue by increasing economic activity.

Local governments also use GIS in unique ways. Because governments are responsible for the long-term health, safety, and welfare of citizens, wider issues need to be considered, including incorporating public values in decision making, delivering services in a fair and equitable manner, and representing the views of citizens by working with elected officials. Typical GIS applications thus include monitoring public health risk, managing public housing stock, allocating welfare assistance funds, and tracking crime. Allied to analysis using geodemographics (see Section 2.3.3) they are also used for operational, tactical, and strategic decision making in law enforcement, health care planning, and managing education systems.

It is convenient to group local government GIS applications on the basis of their contribution to asset inventory, policy analysis, and strategic modeling/planning. Table 2.1 summarizes GIS applications in this way.

These applications can be implemented as centralized GIS or distributed desktop applications. Some will be designed for use by highly trained GIS professionals, while citizens will access others as 'front counter' or Internet systems. Chapter 8 discusses the different implementation models for GIS.

2.3.2.2 Case study application: GIS in tax assessment

Tax mapping and assessment is a classic example of the value of GIS in local government. In many countries local

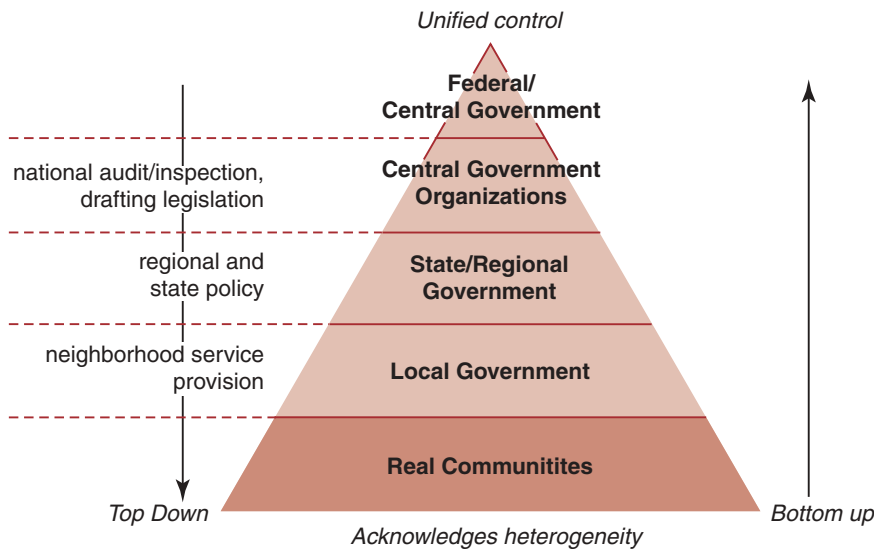


Figure 2.7 The use of GIS at different levels of government decision making

Table 2.1 GIS applications in local government (simplified from O'Looney 2000)

	Inventory Applications (locating property information such as ownership and tax assessments by clicking on a map)	**Policy Analysis Applications** (e.g., number of features per area, proximity to a feature or land use, correlation of demographic features with geological features)	**Management/Policy-Making Applications** (e.g., more efficient routing, modeling alternatives, forecasting future needs, work scheduling)
Economic Development	Location of major businesses and their primary resource demands	Analysis of resource demand by potential local supplier	Informing businesses of availability of local suppliers
Transportation and Services Routing	Identification of sanitation truck routes, capacities and staffing by area; identification of landfill and recycling sites	Analysis of potential capacity strain given development in certain areas; analysis of accident patterns by type of site	Identification of ideal high-density development areas based on criteria such as established transportation capacity
Housing	Inventory of housing stock age, condition, status (public, private, rental, etc.), durability, and demographics	Analysis of public support for housing by geographic area, drive time from low-income areas to needed service facilities, etc.	Analysis of funding for housing rehabilitation, location of related public facilities; planning for capital investment in housing based on population growth projections
Infrastructure	Inventory of roads, sidewalks, bridges, utilities (locations, names, conditions, foundations, and most recent maintenance)	Analysis of infrastructure conditions by demographic variables such as income and population change	Analysis to schedule maintenance and expansion
Health	Locations of persons with particular health problems	Spatial, time-series analysis of the spread of disease; effects of environmental conditions on disease	Analysis to pinpoint possible sources of disease
Tax Maps	Identification of ownership data by land plot	Analysis of tax revenues by land use within various distances from the city center	Projecting tax revenue change due to land-use changes
Human Services	Inventory of neighborhoods with multiple social risk indicators; location of existing facilities and services designated to address these risks	Analysis of match between service facilities and human services needs and capacities of nearby residents	Facility siting, public transportation routing, program planning, and place-based social intervention
Law Enforcement	Inventory of location of police stations, crimes, arrests, convicted perpetrators, and victims; plotting police beats and patrol car routing; alarm and security system locations	Analysis of police visibility and presence; officers in relation to density of criminal activity; victim profiles in relation to residential populations; police experience and beat duties	Reallocation of police resources and facilities to areas where they are likely to be most efficient and effective; creation of random routing maps to decrease predictability of police beats
Land-use Planning	Parcel inventory of zoning areas, floodplains, industrial parks, land uses, trees, green space, etc.	Analysis of percentage of land used in each category, density levels by neighborhoods, threats to residential amenities, proximity to locally unwanted land uses	Evaluation of land-use plan based on demographic characteristics of nearby population (e.g., will a smokestack industry be sited upwind of a respiratory disease hospital?)

(*continued overleaf*)

Table 2.1 (*continued*)

	Inventory Applications	Policy Analysis Applications	Management/Policy-Making Applications
Parks and Recreation	Inventory of park holdings/playscapes, trails by type, etc.	Analysis of neighborhood access to parks and recreation opportunities, age-related proximity to relevant playscapes	Modeling population growth projections and potential future recreational needs/playscape uses
Environmental Monitoring	Inventory of environmental hazards in relation to vital resources such as groundwater; layering of nonpoint pollution sources	Analysis of spread rates and cumulative pollution levels; analysis of potential years of life lost in a particular area due to environmental hazards	Modeling potential environmental harm to specific local areas; analysis of place-specific multilayered pollution abatement plans
Emergency Management	Location of key emergency exit routes, their traffic flow capacity and critical danger points (e.g., bridges likely to be destroyed by an earthquake)	Analysis of potential effects of emergencies of various magnitudes on exit routes, traffic flow, etc.	Modeling effect of placing emergency facilities and response capacities in particular locations
Citizen Information/ Geodemographics	Location of persons with specific demographic characteristics such as voting patterns, service usage and preferences, commuting routes, occupations	Analysis of voting characteristics of particular areas	Modeling effect of placing information kiosks at particular locations

government agencies have a mandate to raise revenue from property taxes. The amount of tax payable is partly or wholly determined by the value of taxable land and property. A key part of this process is evaluating the value of land and property fairly to ensure equitable distribution of a community's tax burden. In the United States the task of determining the taxable value of land and property is performed by the Tax Assessor's Office, which is usually a separate local government department. The Valuation Office Agency fulfills a similar role in the UK. The tax department can quickly get overwhelmed with requests for valuation of new properties and protests about existing valuations.

The Tax Assessor's Office is often the first home of GIS in local government.

Essentially, a Tax Assessor's role is to assign a value to properties using three basic methods: cost, income, and market. The cost method is based on the replacement cost of the property and the value of the land. The Tax Assessor must examine data on construction costs and vacant land values. The income method takes into consideration how much income a property would generate if it were rented. This requires details on current market rents, vacancy rates, operating expenses, taxes, insurance, maintenance, and other costs. The market method is the most popular. It compares the property to other recent sales that have a similar location, size, condition, and quality.

Collecting, storing, managing, analyzing, and displaying all this information is a very time-consuming activity and not surprisingly GIS has had a major impact on the way Tax Assessors go about their business.

2.3.2.3 Method

Tax Assessors, working in a Tax Assessor's Office, are responsible for accurately, uniformly, and fairly judging the value of all taxable properties in their jurisdiction. Details about properties are maintained on a tax assessment roll that includes information such as ownership, address, land and building value, and tax exemptions. The Assessor's Office is also responsible for processing applications for tax abatement, in cases of overvaluation, and exemptions for surviving spouses, veterans, and the elderly. Figure 2.8 shows some aspects of a tax assessment GIS in Ohio, USA.

A GIS is used to collect and manage the geographic boundaries and associated information about properties. Typically, data associated with properties is held in a Computer Assisted Mass Appraisal (CAMA) system that is responsible for sale analysis, evaluation, data management, and administration, and for generating notices to owners. CAMA systems are usually implemented on top of a database management system (DBMS) and can be linked to the parcel database using a common key (see Section 10.2 for further discussion of how this works).

The basic tax assessment task involves a geographic database query to locate all sales of similar properties within a predetermined distance of a given property. The property to be valued is first identified in the property database. Next, a geographic query is used to ascertain the values of all comparable properties within a predetermined search radius (typically one mile) of the property. These properties are then displayed on the assessor's screen. The assessor can then compare the

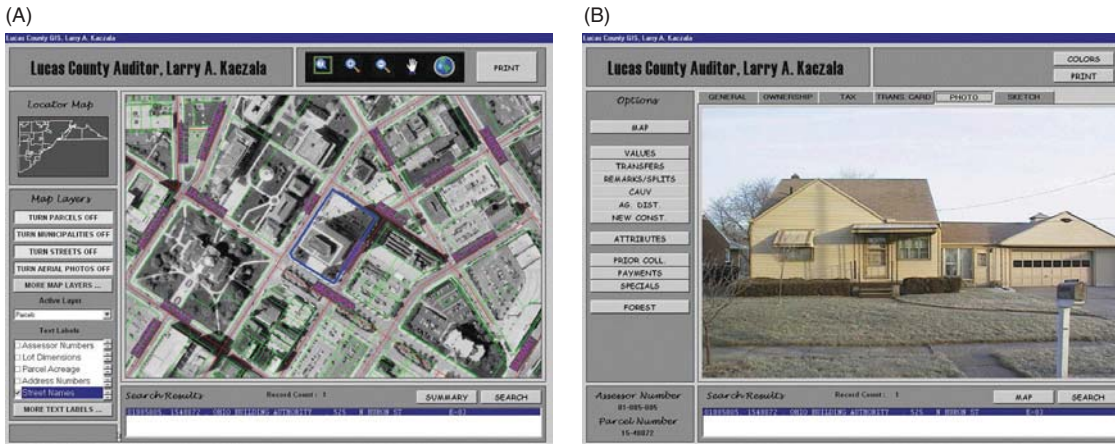


Figure 2.8 Lucas County, Ohio, USA tax assessment GIS: (A) tax map; (B) property attributes and photograph

characteristics of these properties (lot size, sales price and date of sale, neighborhood status, property improvements, etc.) and value the property.

2.3.2.4 Scientific foundations: principles, techniques, and analysis

Scientific foundations

Critical to the success of the tax assessment process is a high-quality, up-to-date geographic database that can be linked to a CAMA system. Considerable effort must expended to design, implement, and maintain the geographic database. Even for a small community of 50 000 properties it can take several months to assemble the geographic descriptions of property parcels with their associated attributes. Chapters 9 and 10 explain the processes involved in managing geographic databases such as this. Linking GIS and CAMA systems can be quite straightforward providing that both systems are based on DBMS technology and use a common identifier to effect linkage between a map feature and a property record. Typically, a unique parcel number (in the US) or unique property reference number (in the UK) is used.

A high-quality geographic database is essential to tax assessment.

Clearly, the system is dependent on an unambiguous definition of parcels, and common standards about how different characteristics (such as size, age, and value of improvements) are represented. The GIS can help enforce coding standards and can be used to derive some characteristics automatically in an objective fashion. For example, GIS makes it straightforward to calculate the area of properties using boundary coordinates.

Fundamentally, this application, like many others in GIS, depends upon an unambiguous and accurate inventory of geographic extent. To be effective it must link this with clear, intelligible, and stable attribute descriptors. These are all core characteristics of scientific investigation, and although the application is driven by results rather than scientific curiosity, it nevertheless follows scientific procedures of controlled comparison.

Principles

Tax assessment makes the assumption that, other things being equal, properties close together in space will have similar values. This is an application of Tobler's First Law of Geography, introduced in Section 3.1. However, it is left to the Assessor to identify comparator properties and to weight their relative importance. This seems rather straightforward, but in practice can prove very difficult – particularly where the exact extent of the effects of good and bad neighborhood attributes cannot be precisely delineated. In practice the value of location in a given neighborhood is often assumed to be uniform (see Section 4.7), and properties of a given construction type are also assumed to be identical. This assumption may be valid in areas where houses were constructed at the same time according to common standards; however, in older areas where infill has been common, properties of a given type vary radically in quality over short distances.

Techniques

Tax assessment requires a good database, a plan for system management and administration, and a workflow design. These procedures are set out in Chapters 9, 10 and 17. The alternative of manually sorting paper records, or even tabular data in a CAMA system, is very laborious and time-consuming, and thus the automated approach of GIS is very cost-effective.

Analysis

Tax assessment actually uses standard GIS techniques such as proximity analysis, and geographic and attribute query, mapping, and reporting. These must be robust and defensible when challenged by individuals seeking reductions in assessments. Chapter 14 sets out appropriate procedures, while Chapter 12 describes appropriate conventions for representing properties and neighborhoods cartographically.

2.3.2.5 Generic scientific questions arising from the application

This is not perhaps the most glamorous application of GIS, but its operational value in tax assessment cannot be

overestimated. It requires an up-to-date inventory of properties and information from several sources about sales and sale prices, improvements, and building programs.

To help tax assessors understand geographic variations in property characteristics it is also possible to use GIS for more strategic modeling activities. The many tools in GIS for charting, reporting, mapping, and exploratory data analysis help assessors to understand the variability of property value within their jurisdictions. Some assessors have also built models of property valuations and have clustered properties based on multivariate criteria (see Section 4.7). These help assessors to gain knowledge of the structure of communities and highlight unusually high or low valuations. Once a property database has been created, it becomes a very valuable asset, not just for the tax assessor's department, but also for many other departments in a local government agency. Public works departments may seek to use it to label access points for repairs and meter reading, housing departments may use it to maintain data on property condition, and many other departments may like shared access to a common address list for record keeping and mailings.

A property database is useful for many purposes besides tax assessment.

2.3.2.6 Management and policy

Tax assessment is a key local government application because it is a direct revenue generator. It is easy to develop a cost-benefit case for this application (Chapter 17) and it can pay for a complete department or corporate GIS implementation quickly (Chapter 18). Tax assessment is a service offered directly to members of the public. As such, the service must be reliable and achieve a quick turnaround (usually with one week). It is quite common for citizens to question the assessed value for their property, since this is the principal determinant of the amount of tax they will have to pay. A tax assessor must, therefore, be able to justify the method and data used to determine property values. A GIS is a great help and often convinces people of the objectivity involved (sometimes over-impressing people that it is totally scientific). As such, GIS is an important tool for efficiency and equitable local government.

2.3.3 Business and service planning

2.3.3.1 Applications overview

Business and service planning (sometimes called retailing) applications focus upon the use of geographic data to provide operational, tactical, and strategic context to decisions that involve the fundamental question, *where*? *Geodemographics* is a shorthand term for composite indicators of consumer behavior that are available at the small-area level (e.g., census output area, or postal zone). Figure 2.9 illustrates the profile of one geodemographic type from a UK classification called Mosaic, developed by market researcher and academic Richard Webber. The current version of Mosaic divides the UK population into 11

Figure 2.9 A geodemographic profile: Town Gown Transition (a Type within the Urban Intelligence Group of the 2001 MOSAIC classification). (Courtesy of Experian Limited. © 2004 Experian Limited. All rights reserved)

Groups (such as 'Happy Families', 'Urban Intelligence', and 'Blue Collar Enterprise'), which in turn are subdivided into a total of 61 Types. Geodemographic data are frequently used in business applications to identify geographic variations in the incidences of customer types. They are often supplemented by *lifestyles* data on the consumption choices and shopping habits of individuals who fill out questionnaires or participate in store loyalty programs. The term *market area analysis* describes the activity of assessing the distribution of retail outlets relative to the greatest concentrations of potential customers. The approach is increasingly being adapted to improving public service planning, in areas such as health, education, and law enforcement (see Box 2.1 and Section 2.3.2).

Geodemographic data are the basis for much market area analysis.

The tools of business applications typically range from simple desktop mapping to sophisticated decision support systems. Tools are used to analyze and inform the range of *operational*, *tactical*, and *strategic* functions of an organization. These tools may be part of standard GIS software, or they may be developed in-house by the organization, or they may be purchased (with or without accompanying data) as a 'business solution' product. We noted in Section 1.1 that operational functions concern the day-to-day processing of routine transactions and inventory analysis in an organization, such as stock management. Tactical functions require the allocation of resources to address specific (usually short term) problems, such as store sales promotions. Strategic functions contribute to the organization's longer-term goals and mission, and entail problems such as opening new stores or rationalizing existing store networks. Early business applications were simply concerned with mapping spatially referenced data, as a general descriptive indicator of the retail environment. This remains the first stage in most business applications, and in itself adds an important dimension to analysis of organizational function. More

Biographical Box **2.1**

Marc Farr, geodemographer

Figure 2.10 Marc Farr, geodemographer

'*City Adventurers* are young, well educated and open to new ideas and influences. They are cosmopolitan in their tastes and liberal in their social attitudes. Few have children. Many are in further education while others are moving into full-time employment. Most do not feel ready to make permanent commitments, whether to partners, professions or to specific employers. As higher education has become internationalized, the *City Adventurers* group has acquired many foreign-born residents, which further encourages ethnic and cultural variety.'

This is the geodemographic profile of the neighborhood in Hove, UK, where Marc Farr lives. Marc read economics and marketing at Lancaster University before going to work as a market researcher in London, first at the TMS Partnership and then at Experian. His work involved use of geodemographic data to analyze retail catchments, measure insurance risk, and analyze household expenditure patterns.

Over time, Marc gained increasing consultancy responsibilities for public sector clients in education, health, and law enforcement. As a consequence of his developing interests in the problems that they face, five years after graduating, he began to work on a Ph.D. that used geodemographics to analyze the ways in which prospective students in the UK choose the universities at which they want to study. He did this work in association with the UK Universities Central Admissions Service (UCAS). Speaking about his Ph.D., which was completed after five years' part time study, Marc says: 'I question the assumptions that the massive increases in numbers of people entering UK higher education during the late 1990s will reduce inequality between different socio-economic groups, or that they will necessarily improve economic and social mobility. Geodemographic analysis also suggests that we need to better understand the relationship between the geography of demand for higher education and its physical supply.'

Marc now works for the Dr. Foster consultancy firm, where he has responsibilities for the calculation of hospital and health authority performance statistics.

recently, decision support tools used by Spatially Aware Professionals (SAPs, Section 1.4.3.2) have created mainstream research and development roles for business GIS applications.

Some of the operational roles of GIS in business are discussed under the heading of logistics applications in Section 2.3.4. These include stock flow management systems and distribution network management, the specifics of which vary from industry sector to sector. Geodemographic analysis is an important operational tool in market area analysis, where it is used to plan marketing campaigns. Each of these applications can be described as assessing the circumstances of an organization.

The most obvious strategic application concerns the spatial *expansion* of a new entrant across a retail market. Expansion in a market poses fundamental spatial problems – such as whether to expand through *contagious diffusion* across space, or *hierarchical diffusion* down a settlement structure, or to pursue some combination of the two (Figure 2.11). Many organizations periodically experience spatial *consolidation* and branch *rationalization*. Consolidation and rationalization may occur: (a) when two organizations with overlapping networks merge; (b) in response to competitive threat; or (c) in response to changes in the retail environment. Changes in the retail environment may be short term and cyclic, as in the response to the recession phase of business cycles, or structural, as with the rationalization of clearing bank branches following the development of personal, telephone, and Internet-based banking (see Section 18.4.4). Still other organizations undergo spatial *restructuring*, as in the market repositioning of bank branches to supply a wider range of more profitable financial services. Spatial restructuring is often the consequence of technological change. For example, a 'clicks and mortar' strategy might be developed by a chain of conventional bookstores, whereby their retail outlets might be reconfigured to offer reliable pick-up points for Internet and telephone orders – perhaps in association with location-based services (Sections 1.4.3.1 and 11.3.2). This may confer advantage over new, purely Internet-based entrants. A final type of strategic operation involves *distribution* of goods and services, as in the case of so-called 'e-tailers', who use the Internet for merchandizing, but must create or buy into viable distribution networks. These various strategic operations require a range of spatial analytic

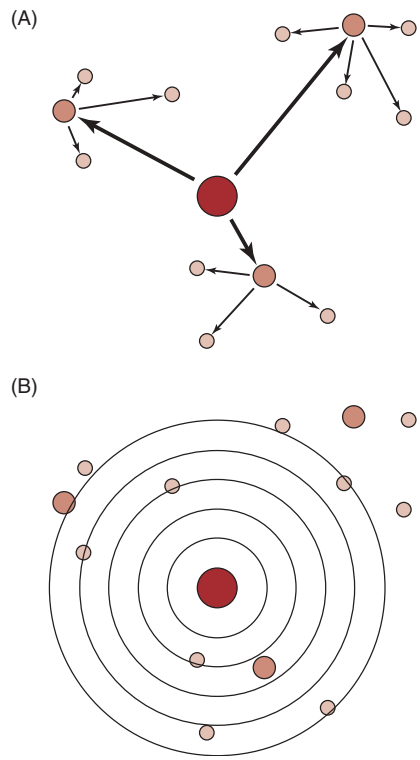


Figure 2.11 (A) Hierarchical and (B) contagious spatial diffusion

tools and data types, and entail a move from 'what-is' visualization to 'what-if' forecasts and predictions.

2.3.3.2 Case study application: hierarchical diffusion and convenience shopping

Tesco is, by some margin, the most successful grocery (food) retailer in the UK, and has used its knowledge of the home market to launch successful initiatives in Asia and the developing markets of Eastern Europe (see Figure 1.7). Achieving real sales growth in its core business of groceries is difficult, particularly in view of a strict national planning regime that prevents widespread development of new stores, and legislation to prevent the emergence, through acquisitions, of local spatial monopolies of supply. One way in which Tesco has succeeded in sustaining market growth in the domestic market in these circumstances is through strategic diversification into consumer durables and clothing in its largest (high order, colored dark red in Figure 2.11) stores. A second driver to growth has been the successful development of a store loyalty card program, which rewards members with money-off coupons or leisure experiences according to their weekly spend. This program generates *lifestyles* data as a very useful by-product, which enables Tesco to identify consumption profiles of its customers, not unlike the Mosaic geodemographic system. This enables the company to identify, for example, whether customers are 'value driven' and should be directed to budget food offerings, or whether they are principally motivated by quality and might be encouraged to purchase goods from the company's 'Finest' range. This is a very powerful marketing tool, although unlike geodemographic discriminators these data tell the company rather little about those households that are not their customers, or the products that their own customers buy elsewhere.

A third driver to growth entails the creation or acquisition of much smaller neighborhood stores (low order, in the lightest red in Figure 2.11). These provide a local community service and are not very intrusive on the retail landscape, and are thus much easier to create within the constraints of the planning system. The 'Express' format store shown in Figure 2.12A opened in 2003 and was planned by Tesco's in-house store location team using GE Smallworld GIS. Figure 2.12B shows its location (labeled T3) in Bournemouth, UK, in relation to the edge of the town and the locations of five competitor chains.

GIS can be used to predict the success of a retailer in penetrating a local market area.

2.3.3.3 Method

The location is in a suburban residential neighborhood, and as such it was anticipated that its customer base would be mainly local – that is, resident within a 1 km radius of the store. A budget was allocated for promoting the new establishment, in order to encourage repeat patronage. An established means of promoting new stores is through leaflet drops or enclosures with free local newspapers. However, such tactical interventions are limited by the coarseness of distribution networks – most organizations that deliver circulars will only undertake deliveries for complete postal sectors (typically 20 000 population size) and so this represents a rather crude and wasteful medium.

A second strategy would be to use the GIS to identify all of the households resident within a 1 km radius of the store. Each UK unit postcode (roughly equivalent to a US Zip+4 code, and typically comprising 18–22 addresses) is assigned the grid reference of the first mail delivery point on the 'mailman's walk'. Thus one of the quickest ways of identifying the relevant addresses entails plotting

Figure 2.12 (A) The site and (B) the location of a new Tesco Express store

(B)

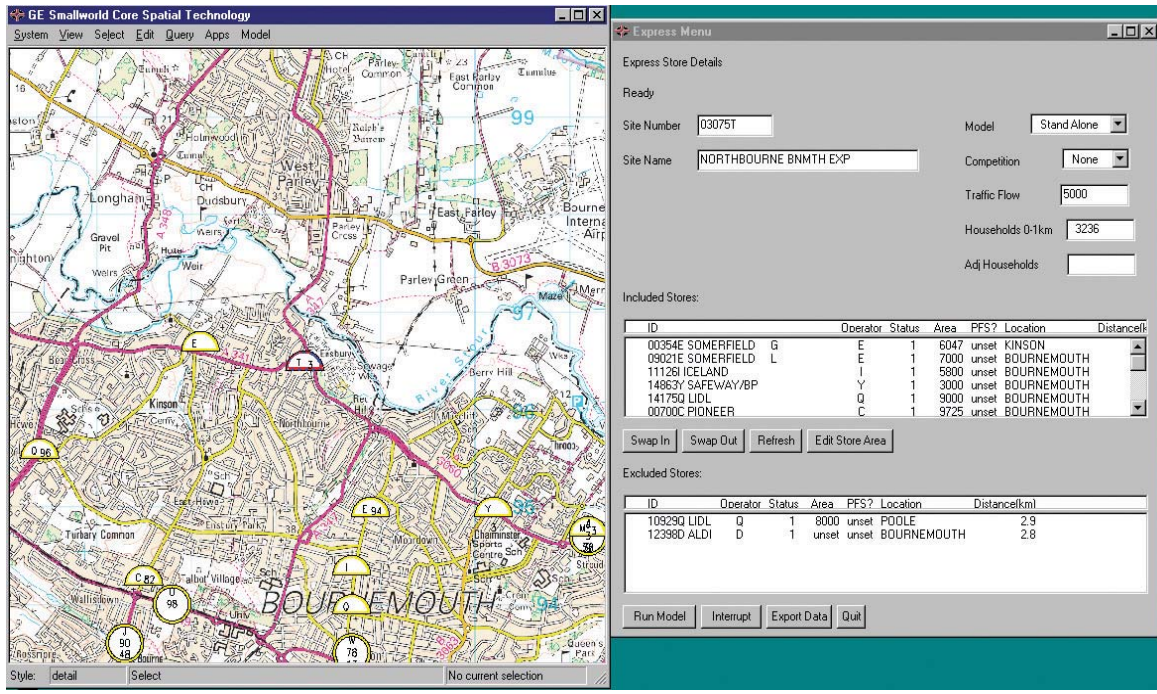


Figure 2.12 (*continued*)

the unit postcode addresses and selecting those that lie within the search radius. Matching the unit postcodes with the full postcode address file (PAF) then suggests that there are approximately 3236 households resident in the search area. Each of these addresses might then be mailed a circular, thus eliminating the largely wasteful activity of contacting around 16 800 households that were unlikely to use the store.

Yet even this tactic can be refined. Sending the same packet of money-off coupons to all 3236 households assumes that each has identical disposable incomes and consumption habits. There may be little point, for example, incentivizing a domestic-beer drinker to buy premium champagne, or vice-versa. Thus it makes sense to overlay the pattern of geodemographic profiles onto the target area, in order to tailor the coupon offerings to the differing consumption patterns of 'blue collar enterprise' neighborhoods versus those classified as belonging to 'suburban comfort', for example.

There is a final stage of refinement that can be developed for this analysis. Using its lifestyles (storecard) data, Tesco can identify those households who already prefer to use the chain, despite the previous nonavailability of a local convenience store. Some of these customers will use Tesco for their main weekly shop, but may 'top up' with convenience or perishable goods (such as bread, milk, or cut flowers) from a competitor. Such households might be offered stronger incentives to purchase particular ranges of goods from the new store, without the wasteful 'cannibalizing' activity of offering coupons towards purchases that are already made from Tesco in the weekly shop.

2.3.3.4 Scientific foundations: geographic principles, techniques, and analysis

The following assumptions and organizing principles are inherent to this case study.

Scientific foundations

Fundamental to the application is the assumption that the closer a customer lives to the store, the more likely he or she is to patronize it. This is formalized as Tobler's First Law of Geography in Section 3.1 and is accommodated into our *representations* as a distance decay effect. The *nature* (Chapter 4) of such distance decay effects does not have to be linear – in Section 4.5 we will introduce a range of *non-linear* effects.

The science of geodemographic profiling can be stated succinctly as 'birds of a feather flock together' – that is, the differences in the observed social and economic characteristics of residents *between* neighborhoods are greater than differences observed *within* them. The use of small-area geodemographic profiles to mix the coupon incentives that might be sent to prospective customers assumes that each potential customer is equally and utterly typical of the post code in which he or she resides. The individual resident *in* an area is thus assigned the characteristics *of* the area. In practice, of course, individuals within households have different characteristics, as do households within streets, zones, and any other aggregation. The practice of confounding characteristics of areas with individuals resident in them is known as committing the *ecological fallacy* – the term does not refer to any branch of biology, but shares with ecology a primary concern with describing the *linkage* of living organisms (individuals) to their

geographical surroundings. This is inevitable in most socio-economic GIS applications because data that enable sensitive characteristics of individuals must be kept confidential (see the point about GIS and the surveillance society in Section 1.7). Whilst few could take offence at error in mis-targeting money-off coupons as in this example, ecological analysis has the potential to cause distinctly unethical outcomes if individuals are penalized because of where they live – for example if individuals find it difficult to gain credit because of the credit histories of their neighborhoods. Such discriminatory activity is usually prevented by industry codes of conduct or even legislation. The use of lifestyles data culled from store loyalty card records enables individuals to be targeted precisely, but such individuals might well be geographically or socially *unrepresentative* of the population at large or the market as a whole.

More generally, geography is a science that has very few natural units of analysis – what, for example, is the natural unit for measuring a soil profile? In socio-economic applications, even if we have disaggregate data we might remain uncertain as to whether we should consider the individual or the household as the basic unit of analysis – sometimes one individual in a household always makes the important decisions, while in others this is a shared responsibility. We return to this issue in our discussion of *uncertainty* (Chapter 6).

The use of lifestyle data from store loyalty programs allows the retailer to enrich geodemographic profiling with information about its own customers. This is a cutting-edge marketing activity, but one where there is plenty of scope for relevant research that is able to take 'what is' information about existing customer characteristics and use it to conduct 'what if' analysis of behavior given a different constellation of retail outlets. We return to the issue of defining appropriate *predictor variables* and *measurement error* in our discussions of spatial dependence (Section 4.7) and uncertainty (Section 6.3). More fundamentally still, is it acceptable (predictively and ethically) to represent the behavior of consumers using any measurable socio-economic variables?

Principles

The definition of the primary market area that is to receive incentives assumes that a *linear* radial distance measure is intrinsically meaningful in terms of defining market area. In practice, there are a number of severe shortcomings in this. The simplest is that *spatial structure* will distort the radial measure of market area–the market is likely to extend further along the more important travel arteries, for example, and will be restricted by physical obstacles such as blocked-off streets and rivers, and by traffic management devices such as stop lights. Impediments to access may be perceived as well as real – it may be that residents of West Parley (Figure 2.12B) would never think of going into north Bournemouth to shop, for example, and that the store's customers will remain overwhelmingly drawn from the area south of the store. Such perceptions of *psychological distance* can be very important yet difficult to accommodate in representations. We return to the issue of appropriate distance metrics in Sections 4.6 and 14.3.1.

Techniques

The assignment of unit postcode coordinates to the catchment zone is performed through a procedure known as *point in polygon* analysis, which is considered in our discussion of *transformation* (Section 14.4.2).

The analysis as described here assumes that the principles that underpin consumer behavior in Bournemouth, UK, are essentially the same as those operating anywhere else on Planet Earth. There is no attempt to accommodate regional and local factors. These might include: adjusting the attenuating effect of distance (see above) to accommodate the different distances people are prepared to travel to find a convenience store (e.g., as between an urban and a rural area); and adjusting the likely attractiveness of the outlet to take account of ease of access, forecourt size, or a range of qualitative factors such as layout or branding. A range of spatial techniques is now available for making the general properties of spatial analysis more *sensitive to context*.

Analysis

Our stripped down account of the store location problem has not considered the *competition* from the other stores shown in Figure 2.12B – despite that fact that all residents almost certainly already purchase convenience goods somewhere! Our description does, however, address the phenomenon of *cannibalizing* – whereby new outlets of a chain poach customers from its existing sites. In practice, both of these issues may be addressed through analysis of the *spatial interactions* between stores. Although this is beyond the scope of this book, Mark Birkin and colleagues have described how the tradition of *spatial interaction modeling* is ideally suited to the problems of defining realistic catchment areas and estimating store revenues. A range of analytic solutions can be devised in order to accommodate the fact that store catchments often overlap.

2.3.3.5 Generic scientific questions arising from the application

A dynamic retail sector is fundamental to the functioning of all advanced economies, and many investments in location are so huge that they cannot possibly be left to chance. Doing nothing is simply not an option. Intuition tells us that the effects of distance to outlet, and the organization of existing outlets in the retail hierarchy *must* have some kind of impact upon patterns of store patronage. But, in intensely competitive consumer-led markets, the important question is *how much* impact?

Human decision making is complex, but predicting even a small part of it can be very important to a retailer.

Consumers are sophisticated beings and their shopping behavior is often complex. Understanding local patterns of convenience shopping is perhaps quite straightforward, when compared with other retail decisions that involve stores that have a wider range of attributes, in terms of floor space, range and quality of goods and services, price, and customer services offered. Different consumer groups

find different retailer attributes attractive, and hence it is the mix of *individuals* with particular characteristics that largely determines the likely store turnover of a particular location. Our example illustrates the kinds of simplifying assumptions that we may choose to make using the best available data in order to represent consumer characteristics and store attributes. However, it is important to remember that even blunt-edged tools can increase the effectiveness of operational and strategic R&D (research and development) activities many-fold. An untargeted leafleting campaign might typically achieve a 1% hit rate, while one informed by even quite rudimentary market area analysis might conceivably achieve a rate that is five times higher. The pessimist might dwell on the 95% failure rate that a supposedly scientific approach entails, yet the optimist should be more than happy with the fivefold increase in the efficiency of use of the marketing budget!

2.3.3.6 Management and policy

The geographic development of retail and business organizations has sometimes taken place in a haphazard way. However, the competitive pressures of today's markets require an understanding of branch location networks, as well as their abilities to anticipate and respond to threats from new entrants. The role of Internet technologies in the development of 'e-tailing' is important too, and these introduce further spatial problems to retailing – for example, in developing an understanding of the geographies of engagement with new information and communications technologies and in working out the logistics of delivering goods and services ordered in cyberspace to the geographic locations of customers (see Section 2.3.4).

Thus the role of the Spatially Aware Professional is increasingly as a mainstream manager alongside accountants, lawyers, and general business managers. SAPs complement understanding of corporate performance in national and international markets with performance at the regional and local levels. They have key roles in such areas of organizational activity as marketing, store revenue predictions, new product launch, improving retail networks, and the assimilation of pre-existing components into combined store networks following mergers and acquisitions.

Spatially Aware Professionals do much more than simple mapping of data.

Simple mapping packages alone provide insufficient scientific grounding to resolve retail location problems. Thus a range of *GIServices* have been developed – some in house, by large retail corporations (such as Tesco, above), some by software vendors that provide analytical and data services to retailers, and some by specialist consultancy services. There is ongoing debate as to which of these solutions is most appropriate to retail applications. The resolution of this debate lies in understanding the nature of particular organizations, their range of goods and services, and the priority that organizations assign to operational, tactical, and strategic concerns.

2.3.4 Logistics and transportation

2.3.4.1 Applications overview

Knowing where things are can be of enormous importance for the fields of logistics and transportation, which deal with the movement of goods and people from one place to another, and the infrastructure (highways, railroads, canals) that moves them. Highway authorities need to decide what new routes are needed and where to build them, and later need to keep track of highway condition. Logistics companies (e.g., parcel delivery companies, shipping companies) need to organize their operations, deciding where to place their central sorting warehouses and the facilities that transfer goods from one mode to another (e.g., from truck to ship), how to route parcels from origins to destinations, and how to route delivery trucks. Transit authorities need to plan routes and schedules, to keep track of vehicles and to deal with incidents that delay them, and to provide information on the system to the traveling public. All of these fields employ GIS, in a mixture of operational, tactical, and strategic applications.

The field of logistics addresses the shipping and transportation of goods.

Each of these applications has two parts: the static part that deals with the fixed infrastructure, and the dynamic part that deals with the vehicles, goods, and people that move on the static part. Of course, not even a highway network is truly static, since highways are often rebuilt, new highways are added, and highways are even sometimes moved. But the minute-to-minute timescale of vehicle movement is sharply different from the year-to-year changes in the infrastructure. Historically, GIS has been easier to apply to the static part, but recent developments in the technology are making it much more powerful as a tool to address the dynamic part as well. Today, it is possible to use GPS (Section 5.8) to track vehicles as they move around, and transit authorities increasingly use such systems to inform their users of the locations of buses and trains (Section 11.3.2 and Box 13.4).

GPS is also finding applications in dealing with emergency incidents that occur on the transportation network (Figure 2.13). The OnStar system (**www.onstar.com**) is one of several products that make use of the ability of GPS to determine location accurately virtually anywhere. When installed in a vehicle, the system is programmed to transmit location automatically to a central office whenever the vehicle is involved in an accident and its airbags deploy. This can be life-saving if the occupants of the vehicle do not know where they are, or are otherwise unable to call for help.

Many applications in transportation and logistics involve optimization, or the design of solutions to meet specified objectives. Section 15.3 discusses this type of analysis in detail, and includes several examples dealing with transportation and logistics. For example, a delivery company may need to deliver parcels to 200 locations in a given shift, dividing the work between 10 trucks. Different ways of dividing the work, and routing the

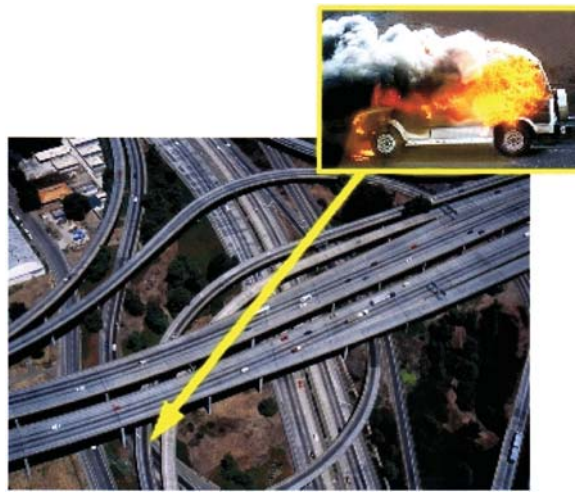


Figure 2.13 Systems such as OnStar allow information on the location of an accident, determined by a GPS unit in the vehicle, to be sent to a central office and compared to a GIS database of highways and streets, to determine the incident location so that emergency teams can respond

vehicles, can result in substantial differences in time and cost, so it is important for the company to use the most efficient solution (see Box 15.4 for an example of the daily workload of an elevator repair company). Logistics and related applications of GIS have been known to save substantially over traditional, manual ways of determining routes.

GIS has helped many service and delivery companies to substantially reduce their operating costs in the field.

2.3.4.2 Case study application: planning for emergency evacuation

Modern society is at risk from numerous types of disasters, including terrorist attacks, extreme weather events such as hurricanes, accidental spills of toxic chemicals resulting from truck collisions or train derailments, and earthquakes. In recent years several major events have required massive evacuation of civilian populations – for example, 800 000 people evacuated in Florida in advance of Hurricane Frances in 2004 (Figure 2.14).

In response to the threat of such events, most communities attempt to plan. But planning is made particularly difficult because the magnitude and location of the event can rarely be anticipated. Suppose, for example, that we attempt to develop a plan for dealing with a spill of a volatile toxic chemical resulting from a train derailment. It might make sense to plan for the worst case, for example the spillage of the entire contents of several cars loaded with chlorine gas. But the derailment might occur anywhere on the rail network, and the impact will depend on the strength and direction of the wind. Possible scenarios might involve people living within tens of kilometers of any point on the track network (see Section 14.4.1 for details of the buffer operation, which

Figure 2.14 Hurricane Frances approaching the coast of Florida, USA, September 3, 2004 (Courtesy US National Oceanic and Atmospheric Administration, NOAA)

would be used in such cases to determine areas lying within a specified distance). Locations can be anticipated for some disasters, such as those resulting from fire in buildings known to be storing toxic chemicals, but hurricanes and earthquakes can impact almost anywhere within large areas.

The magnitude and location of a disaster can rarely be anticipated.

To illustrate the value of GIS in evacuation planning, we have chosen the work of Tom Cova, an academic expert on GIS in emergency management. Tom's early work was strongly motivated by the problems that occurred in the Oakland Hills fire of October 1991 in Northern California, USA, which destroyed approximately 1580 acres and over 2700 structures in the East Bay Hills. This became the most expensive fire disaster in Californian history (Figure 2.15), taking 25 lives and causing over $1.68 billion in damages.

Cova has developed a planning tool that allows neighborhoods to rate the potential for problems associated with evacuation, and to develop plans accordingly. The tool

Figure 2.15 The Oakland Hills fire of October, 1991, which took 25 lives, in part because of the difficulty of evacuation

uses a GIS database containing information on the distribution of population in the neighborhood, and the street pattern. The result is an evacuation vulnerability map. Because the magnitude of a disaster cannot be known in advance, the method works by identifying the worst-case scenario that could affect a given location.

Suppose a specific household is threatened by an event that requires evacuation, such as a wildfire, and assume for the moment that one vehicle is needed to evacuate each household. If the house is in a cul-de-sac, the number of vehicles needing to exit the cul-de-sac will be equal to the number of households on the street. If the entire neighborhood of streets has only one exit, all vehicles carrying people from the neighborhood will need to use that one exit. Cova's method works by looking further and further from the household location, to find the most important bottleneck – the one that has to handle the largest amount of traffic. In an area with a dense network of streets traffic will disperse among several exits, reducing the bottleneck effect. But a densely packed neighborhood with only a single exit can be the source of massive evacuation problems, if a disaster requires the rapid evacuation of the entire neighborhood. In the Oakland Hills fire there were several critical bottlenecks – one-lane streets that normally carry traffic in both directions, but became hopelessly clogged in the emergency.

Figure 2.16 shows a map of Santa Barbara, California, USA, with streets colored according to Cova's measure of evacuation vulnerability. The color assigned to any location indicates the number of vehicles that would have to pass through the critical bottleneck in the worst-case evacuation, with red indicating that over 500 vehicles per lane would have to pass through the bottleneck. The red area near the shore in the lower left is a densely packed area of student housing, with very few routes out of the neighborhood. An evacuation of the entire neighborhood would produce a very heavy flow of vehicles on these exit routes. The red area in the upper left has a much lower population density, but has only one narrow exit.

2.3.4.3 Method

Two types of data are required for the analysis. Census data are used to determine population and household counts, and to estimate the number of vehicles involved in an evacuation. Census data are available as aggregate counts for areas of a few city blocks, but not for individual houses, so there will be some uncertainty regarding the exact numbers of vehicles needing to leave a specific street, though estimates for entire neighborhoods will be much more accurate. The locations of streets are obtained from so-called *street centerline* files, which give the geographic locations, names, and other details of individual streets (see Sections 9.4 and 10.8 for overviews of geographic data sources). The TIGER (Topologically Integrated Geographic Encoding and Referencing) files, produced by the US Bureau of the Census and the US Geological Survey and readily available from many sites on the Internet, are one free source of such data for the USA, and many private companies also offer such data, many adding new information such as traffic flow volumes or directions (for US sources, see, for example, GDT Inc., Lebanon, New Hampshire, now part of Tele Atlas, **www.geographic.com**; and NAVTEQ, formerly Navigation Technologies, Chicago, Illinois, **www.navteq.com**).

Street centerline files are essential for many applications in transportation and logistics.

The analysis proceeds by beginning at every street intersection, and working outwards following the street connections to reach new intersections. Every connection is tested to see if it presents a bottleneck, by dividing the total number of vehicles that would have to move out

Figure 2.16 Evacuation vulnerability map of the area of Santa Barbara, California, USA. Colors denote the difficulty of evacuating an area based on the area's worst-case scenario (Reproduced by permission of Tom Cova)

of the neighborhood by the number of exit lanes. After all streets have been searched out to a specified distance from the start, the worst-case value (vehicles per lane) is assigned to the starting intersection. Finally, the entire network is colored by the worst-case value.

2.3.4.4 Scientific foundations: geographic principles, techniques, and analysis

Scientific foundations

Cova's example is one of many applications that have been found for GIS in the general areas of logistics and transportation. As a planning tool, it provides a way of rating areas against a highly uncertain form of risk, a major evacuation. Although the worst-case scenario that might affect an area may never occur, the tool nevertheless provides very useful information to planners who design neighborhoods, giving them graphic evidence of the problems that can be caused by lack of foresight in street layout. Ironically, the approach points to a major problem with the modern style of street layout in subdivisions, which limits the number of entrances to subdivisions from major streets in the interests of creating a sense of community, and of limiting high-speed through traffic. Cova's analysis shows that such limited entrances can also be bottlenecks in major evacuations.

The analysis demonstrates the value of readily available sources of geographic data, since both major inputs – demographics and street layout – are available in digital form. At the same time we should note the limitations of using such sources. Census data are aggregated to areas that, while small, nevertheless provide only aggregated counts of population. The street layouts of TIGER and other sources can be out of date and inaccurate, particularly in new developments, although users willing to pay higher prices can often obtain current data from the private sector. And the essentially geometric approach cannot deal with many social issues: evacuation of the disabled and elderly, and issues of culture and language that may impede evacuation. In Chapter 16 we look at this problem using the tools of dynamic simulation modeling, which are much more powerful and provide ways of addressing such issues.

Principles

Central to Cova's analysis is the concept of *connectivity*. Very little would change in the analysis if the input maps were stretched or distorted, because what matters is how the network of streets is connected to the rest of the world. Connectivity is an instance of a *topological* property, a property that remains constant when the spatial framework is stretched or distorted. Other examples of topological properties are *adjacency* and *intersection*, both of which cannot be destroyed by stretching a map. We discuss the importance of topological properties and their representation in GIS in Section 10.7.1.

The analysis also relies on being able to find the *shortest path* from one point to another through a street network, and it assumes that people will follow such paths when they evacuate. Many forms of GIS analysis rely on being able to find shortest paths, and we discuss some of them in Section 15.3.3. Many WWW sites will find shortest paths between two street addresses (Figure 1.17). In practice, people will often not use the shortest path, preferring routes that may be quicker but longer, or routes that are more scenic.

Techniques

The techniques used in this example are widely available in GIS. They include *spatial interpolation* techniques, which are needed to assign worst-case values to the streets, since the analysis only produces values for the intersections. Spatial interpolation is widely applied in GIS to use information obtained at a limited number of sample points to guess values at other points, and is discussed in general in Box 4.3, and in detail in Section 14.4.4.

The shortest path methods used to route traffic are also widely available in GIS, along with other functions needed to create, manage, and visualize information about networks.

Analysis

Cova's technique is an excellent example of the use of GIS analysis to *make visible what is otherwise invisible*. By processing the data and mapping the results in ways that would be impossible by hand, he succeeds in exposing areas that are difficult to evacuate and draws attention to potential problems. This idea is so central to GIS that it has sometimes been claimed as the primary purpose of the technology, though that seems a little strong, and ignores many of the other applications discussed in this chapter.

2.3.4.5 Generic scientific questions arising from the application

Logistic and transportation applications of GIS rely heavily on representations of networks, and often must ignore off-network movement. Drivers who cut through parking lots, children who cross fields on their way to school, houses in developments that are not aligned along linear streets, and pedestrians in underground shopping malls all confound the network-based analysis that GIS makes possible. Humans are endlessly adaptable, and their behavior will often confound the simplifying assumptions that are inherent to a GIS model. For example, suppose a system is developed to warn drivers of congestion on freeways, and to recommend alternative routes on neighborhood streets. While many drivers might follow such recommendations, others will reason that the result could be severe congestion on neighborhood streets, and reduced congestion on the freeway, and ignore the recommendation. Residents of the neighborhood streets might also be tempted to try to block the use of such systems, arguing that they result in unwanted and inappropriate traffic, and risk to themselves. Arguments such as these are based on the notion that the transportation system can only be addressed as a whole, and that local modifications based on limited perspectives, such as the addition of a new freeway or bypass, may create more problems than they solve.

2.3.4.6 Management and policy

GIS is used in all three modes – operational, tactical, and strategic – in logistics and transportation. This section concludes with some examples in all three categories.

In *operational* systems, GIS is used:

- To monitor the movement of mass transit vehicles, in order to improve performance and to provide improved information to system users.
- To route and schedule delivery and service vehicles on a daily basis to improve efficiency and reduce costs.

In *tactical* systems:

- To design and evaluate routes and schedules for public bus systems, school bus systems, garbage collection, and mail collection and delivery.
- To monitor and inventory the condition of highway pavement, railroad track, and highway signage, and to analyze traffic accidents.

In *strategic* systems:

- To plan locations for new highways and pipelines, and associated facilities.
- To select locations for warehouses, intermodal transfer points, and airline hubs.

2.3.5 Environment

2.3.5.1 Applications overview

Although it is the last area to be discussed here, the environment drove some of the earliest applications of GIS, and was a strong motivating force in the development of the very first GIS in the mid-1960s (Section 1.4.1). Environmental applications are the subject of several GIS texts, so only a brief overview will be given here for the purposes of illustration.

The development of the Canada Geographic Information System in the 1960s was driven by the need for policies over the use of land. Every country's land base is strictly limited (although the Dutch have managed to expand theirs very substantially by damming and draining), and alternative uses must compete for space. Measures of area are critical to effective strategy – for example, how much land is being lost to agriculture through urban development and sprawl, and how will this impact upon the ability of future generations to feed themselves? Today, we have very effective ways of monitoring land use change through remote sensing from space, and are able to get frequent updates on the loss of tropical forest in the Amazon basin (see Figure 15.14). GIS is also allowing us to devise measures of urban sprawl in historically separate national settlement systems in Europe (Figure 2.17).

GIS allows us to compare the environmental conditions prevailing in different nations.

Generally, it is understood that the 21st century will see increasing proportions of the world's population resident in cities and towns, and so understanding of the environmental impacts of urban settlements is an increasingly important focus of attention in science and policy. Researchers have used GIS to investigate and understand how urban sprawl occurs, in order to understand the environmental consequences of sprawl and to predict its future consequences. Such predictions can be based on historic patterns of growth, together with information on the locations of roads, steeply sloping land unsuitable for development, land that is otherwise protected from urban use, and other factors that encourage or restrict urban development. Each of these factors may be represented in map form, as a layer in the GIS, while specialist software can be designed to simulate the processes that drive growth. These urban growth models are examples of *dynamic simulation models*, or computer programs designed to simulate the operation of some part of the human or environmental system. Figure 2.18, taken from the work of geographer Paul Torrens, presents a simple simulation of urban growth in the American Mid-West under four rather different growth scenarios: (A) uncontrolled suburban sprawl; (B) growth restricted to existing travel arteries; (C) 'leap-frog' development, occurring because of local zoning controls; and (D) development that is constrained to some extent.

Other applications are concerned with the simulation of processes principally in the natural environment. Many models have been coupled with GIS in the past decade, to simulate such processes as soil erosion, forest growth, groundwater movement, and runoff. Dynamic simulation modeling is discussed in detail in Chapter 16.

2.3.5.2 Case study application: deforestation on Sibuyan Island, the Philippines

If the increasing extent of urban areas, described above, is one side of the development coin, then the reduction in the extent of natural land cover is frequently the other. Deforestation is one important manifestation of land use change, and poses a threat to the habitat of many species in tropical and temperate forest areas alike. Ecologists, environmentalists, and urban geographers are therefore using GIS in interdisciplinary investigations to understand the local conditions that lead to deforestation, and to understand its consequences. Important evidence of the rate and patterning of deforestation has been provided through analysis of remote sensing images (again, see Figure 15.14 for the case of the Amazon), and these analyses of *pattern* need to be complemented by analysis at detailed levels of the causes and underlying driving factors of the *processes* that lead to deforestation. The negative environmental impacts of deforestation can be ameliorated by adequate spatial planning of natural parks and land development schemes. But the more strategic objective of sustainable development can only be achieved if a holistic approach is taken to ecological, social, and economic needs. GIS provides the medium of choice for integrating knowledge of natural and social processes in the interests of integrated environmental planning.

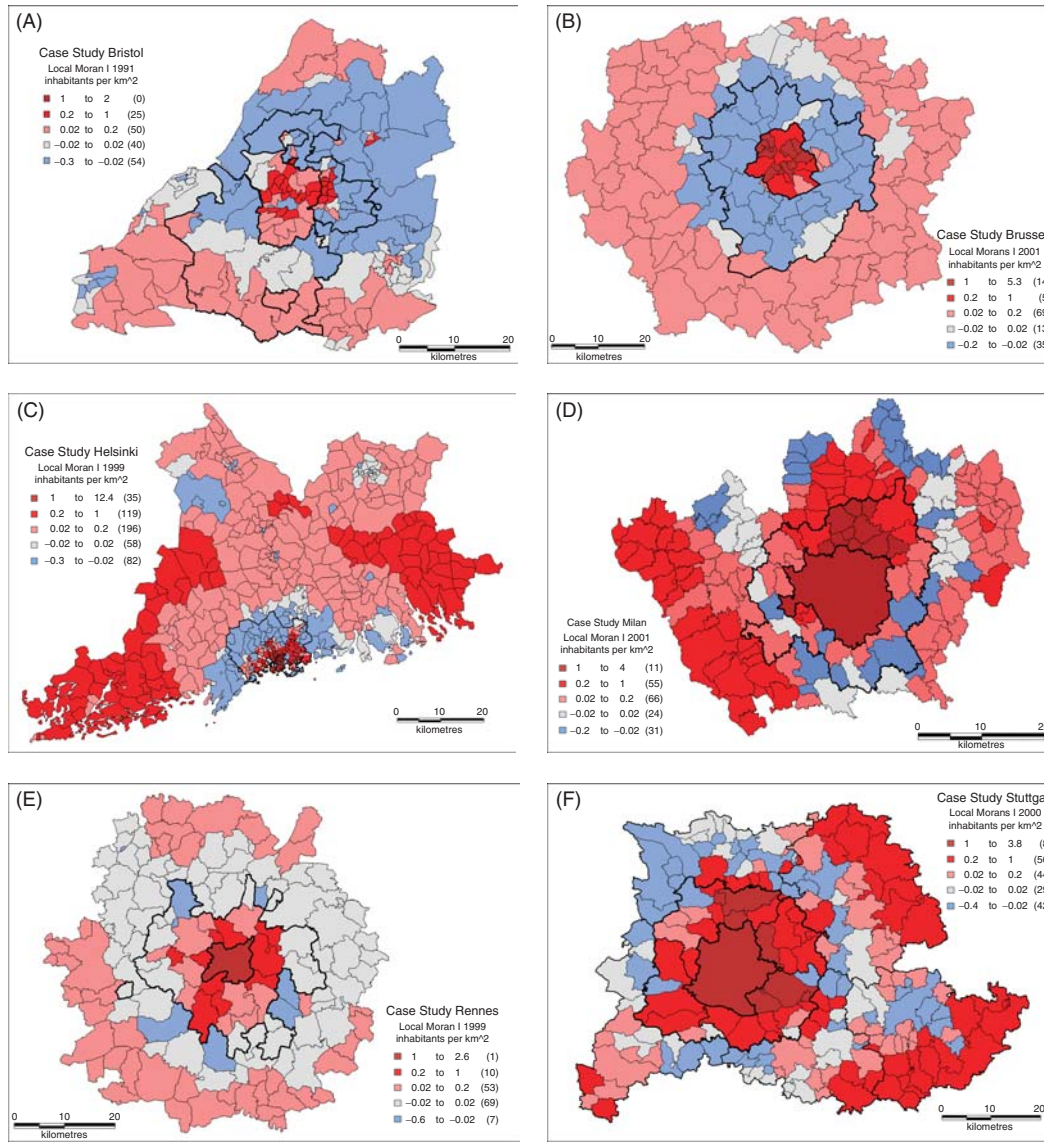


Figure 2.17 GIS enables standardized measures of sprawl for the different nation states of Europe (Reproduced by permission of Guenther Haag & Elena Besussi)

Working at the University of Wageningen in the Netherlands, Peter Verburg and Tom Veldkamp coordinate a research program that is using GIS to understand the sometimes complex interactions that exist between socio-economic and environmental systems, and to gauge their impact upon land use change in a range of different regions of the world (see **www.cluemodel.nl**). One of their case study areas is Sibuyan Island in the Philippines (Figure 2.19A), where deforestation poses a major threat to biodiversity. Sibuyan is a small island (area 456 km^2) of steep forested mountain slopes (Figure 2.19B) and gently sloping coastal land that is used mainly for agriculture, mining, and human settlement. The island has remarkable biodiversity – there are an estimated 700 plant species, of which 54 occur only on Sibuyan Island, and a unique local fauna. The objective of this case study was to identify a range of different development *scenarios* that make it possible to anticipate future land use and habitat change, and hence also anticipate changes in biodiversity.

2.3.5.3 Method

The initial stage of Verburg and Veldkamp's research was a *qualitative* investigation, involving interviews with different stakeholders on the island to identify a list of factors that are likely to influence land use patterns. Table 2.2 lists the data that provided direct or indirect indicators of pressure for land use change. For example, the suitability of the soil for agriculture or the accessibility of a location to local markets can increase the likelihood that a location will be stripped of forest and used for agriculture. They then used these data in a *quantitative* GIS-based analysis to calculate the probabilities of land use transition under three different

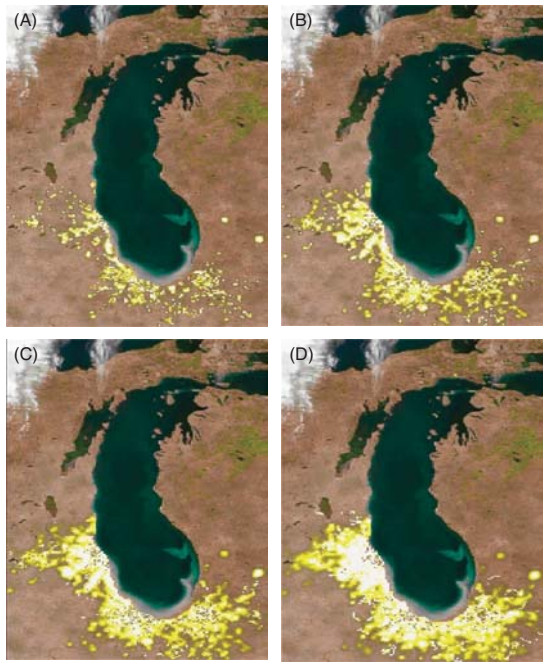


Figure 2.18 Growth in the American Mid-West under four different urban growth scenarios. Horizontal extent of image is 400 km (*Source*: P. Torrens 2005 'Simulating sprawl with geographic automata models', reproduced with permission of Paul Torrens)

scenarios of land use change – each of which was based on different spatial planning policies. Scenario 1 assumes no effective protection of the forests on the island (and a consequent piecemeal pattern of illegal logging), Scenario 2 assumes protection of the designated natural park area alone, and Scenario 3 assumes protection not only of the natural park but also a GIS-defined buffer zone. Figure 2.20 illustrates the forecasted remaining forest area under each of the scenarios at the end of a twenty-year simulation period (1999–2019). The three different scenarios not only resulted in different forest *areas* by 2019 but also different *spatial patterning* of the remaining forest. For example, gaps in the forest area under Scenario 1 were mainly caused by shifting cultivation and illegal logging within the area of primary rainforest, while most deforestation under Scenario 2 occurred in the lowland areas. Qualitative interpretations of the outcomes and aggregate statistics are supplemented by numerical spatial indices such as *fractal dimensions* in order to anticipate the effects of changes upon ecological processes – particularly the effects of disturbance at the edges of the remaining forest area. Such statistics make it possible to define the relative sizes of core and fragmented forest areas (for example, Scenario 1 in Figure 2.20 leads to the greatest fragmentation of the forest area), and this in turn makes it possible to measure the effects of development on biodiversity. Fragmentation statistics are discussed in Section 15.2.5.

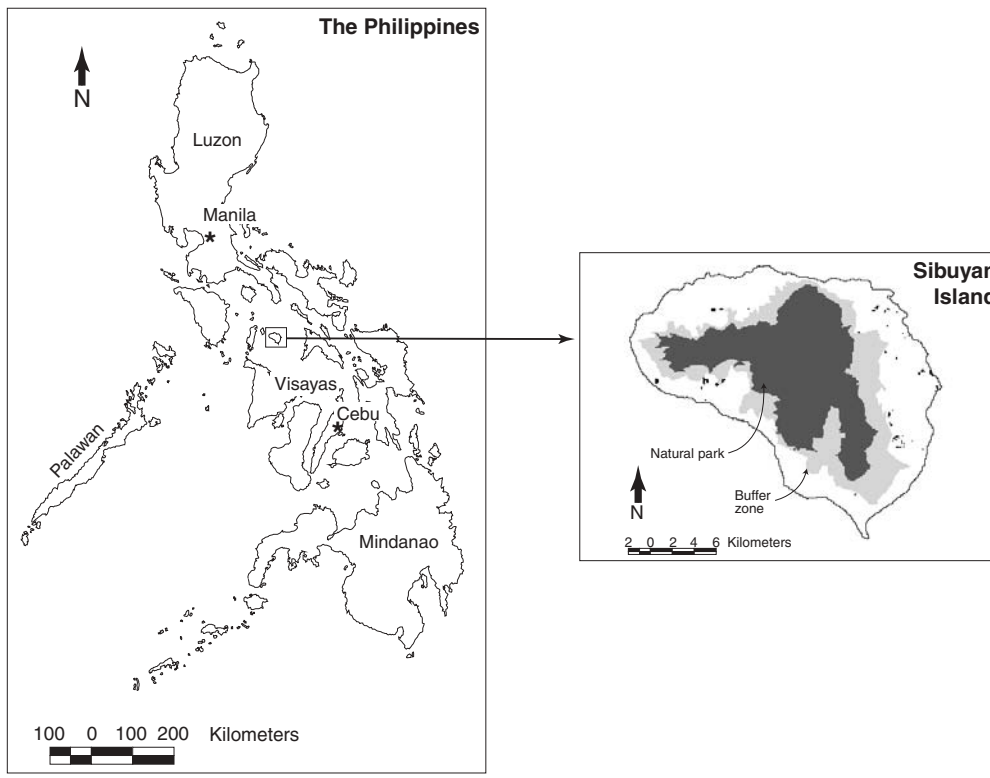


Figure 2.19 (A) Location of Sibuyan Island in the Philippines, showing location of the park and buffer zone, and (B) typical forested mountain landscape of Sibuyan Island

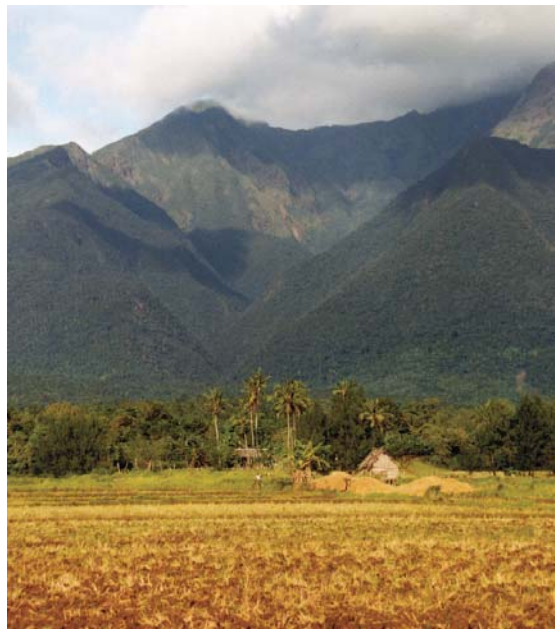


Figure 2.19 (*continued*)

Table 2.2 Data sources used in Verburg and Veldkamp's ecological analysis

Land use
Mangroves
Coconut plantations
Wetland rice cultivation
Grassland
Secondary forest
Swidden agriculture
Primary rainforest and mossy forest
Location factors
Accessibility of roads, rivers and populated places
Altitude
Slope
Aspect
Geology
Geomorphology
Population density
Population pressure
Land tenure
Spatial policies

2.3.5.4 Scientific foundations: geographic principles, techniques, and analysis

Scientific foundations

The goal of the research was to predict changes in the biodiversity of the island. Existing knowledge of a set of ecological *processes* led the researchers to the view that biodiversity would be compromised by changes in overall size of natural forest, and by changes in the *patterning* of the forest that remained. It was hypothesized that changes in patterning would be *caused* by the combined effects of a further set of physical, biological, and human *processes*. Thus the researchers take the observed existing land use pattern, and use understanding of the physical, biological, and human processes to predict future land use changes. The different forecasts of land use (based on different

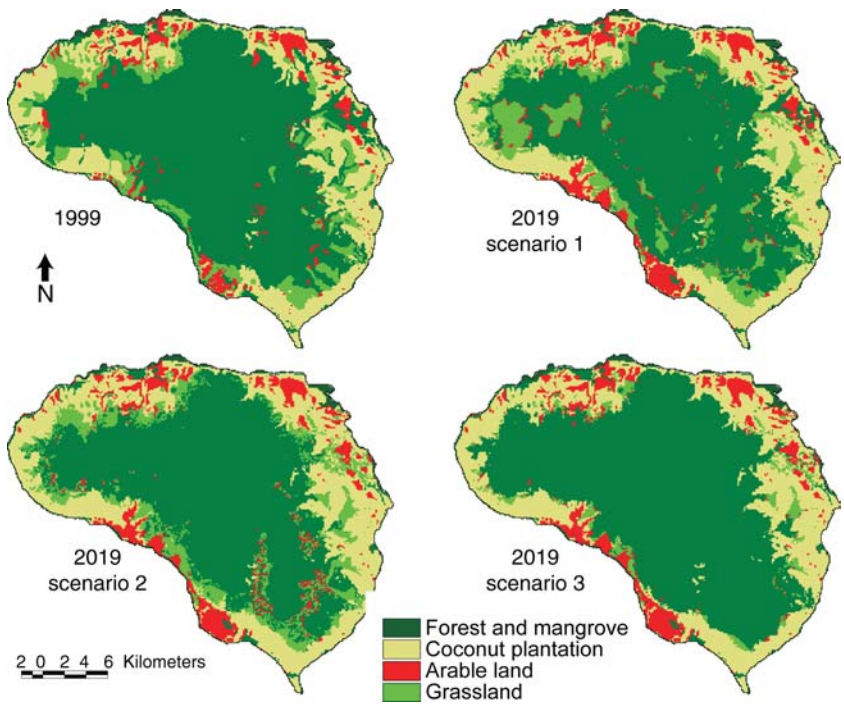


Figure 2.20 Forest area (dark green) in 1999 and at the end of the land use change simulations (2019) for three different scenarios

scenarios) can then be used see whether the functioning of ecological processes on the island will be changed in the future. This theme of inferring *process* from *pattern*, or *function* from *form*, is a common characteristic of GIScience applications.

Of course, it is not possible to identify a uniform set of physical, biological, and human processes that is valid in all regions of the world (the pure *nomothetic* approach of Section 1.3). But conversely it does not make sense to treat every location as unique (the *idiographic* approach of Section 1.3) in terms of the processes extant upon it. The art and science of ecological modeling requires us to make a good call not just on the range of relevant determining factors, but also their importance in the specific case study, with due consideration to the appropriate scale range at which each is relevant.

Geographic principles

GIS makes it possible to incorporate diverse physical, biological, and human elements, and to forecast the size, shape, scale, and dimension of land use parcels. Therefore, it is possible in this case study to predict habitat fragmentation and changes in biodiversity. Fundamental to the end use of this analysis is the assumption that the ecological consequences of future deforestation can be reliably predicted using a forecast land use map. The forecasting procedure also assumes that the various indicators of land development pressure are *robust, accurate*, and *reliable*. There are inevitable uncertainties in the ways in which these indicators are conceived and measured. Further uncertainties are generated by the scale of analysis that is carried out (Section 6.4), and, like our retailing example above, the qualitative importance of local context may be important.

Land use change is deemed to be a measurable response to a wide range of locationally variable factors. These factors have traditionally been the remit of different disciplines that have different intellectual traditions of measurement and analysis. As Peter Verburg says: 'The research assumes that GIS can provide a sort of "Geographic Esperanto" – that is, a common language to integrate diverse, geographically variable factors. It makes use of the core GIS idea that the world can be understood as a series of layers of different types of information, that can be added together meaningfully through overlay analysis to arrive at conclusions.'

Techniques

The *multicriteria* techniques used to harmonize the different location factors into a composite spatial indicator of development pressure are widely available in GIS, and are discussed in detail in Section 16.4. The individual component indicators are acquired using techniques such as *on-screen digitizing* and *classification of imagery* to obtain a land use map. All data are converted to a *raster* data structure with common resolution and extent. Relations between the location factors and land use are quantified using *correlation* and *regression* analysis based on the spatial dataset (Section 4.7).

Analysis

The GIS is used to simulate scenarios of future land use change based on different spatial policies. The application is predicated on the premise that changes in ecological *process* can be reliably inferred from predicted changes in land use *pattern*. Process is inferred not just through size measures, but also through spatial measures of connectivity and fragmentation – since these latter aspects affect the ability of a species to mix and breed without disturbance. The analysis of the extent and ways in which different land uses fill space is performed using specialized software (see Section 15.2.5). Although such spatial indices are useful tools, they may be more relevant to some aspects of biodiversity than others, since different species are vulnerable to different aspects of habitat change.

2.3.5.5 Generic scientific questions arising from the application

GIS applications need to be based on sound science. In environmental applications, this knowledge base is unlikely to be the preserve of any single academic discipline. Many environmental applications require recourse to use of GIS in the field, and field researchers often require multidisciplinary understanding of the full range of processes leading to land use change.

Irrespective of the quality of the measurement process, uncertainty will always creep into any prediction, for a number of reasons. Data are never perfect, being subject to measurement error (Chapter 6), and uncertainty arising out of the need selectively to generalize, abstract, and approximate (Chapter 3). Furthermore, simulations of land use change are subject to changes in exogenous forces such as the world economy. Any forecast can only be a selectively simplified representation of the real world and the processes operating within it. GIS users need to be aware of this, because the forecasts produced by a GIS will always appear to be precise in numerical terms, and spatial representations will usually be displayed using crisp lines and clear mapped colors.

GIS users should not think of systems as *black boxes*, and should be aware that explicit spatial forecasts may have been generated by invoking assumptions about process and data that are not as explicit. User awareness of these important issues can be improved through appropriate *metadata* and documentation of research procedures – particularly when interdisciplinary teams may be unaware of the disciplinary conventions that govern data creation and analysis in parts of the research. Interdisciplinary science and the cumulative development of algorithms and statistical procedures can lead GIS applications to conflict with an older principle of scientific reporting, that the results of analysis *should always be reported in sufficient detail to allow someone else to replicate them*. Today's science is complex, and all of us from time to time may find ourselves using tools developed by others that we do not fully understand. It is up to all of us to demand to know as many of the details of GIS analysis as is reasonably possible.

Users of GIS should always know exactly what the system is doing to their data.

2.3.5.6 Management and policy

GIS is now widely used in all areas of environmental science, from ecology to geology, and from oceanography to alpine geomorphology. GIS is also helping to reinvent environmental science as a discipline grounded in field observation, as data can be captured using battery-powered personal data assistants (PDAs) and notebooks, before being analyzed on a battery-powered laptop in a field tent, and then uploaded via a satellite link to a home institution. The art of scientific forecasting (by no means a contradiction in terms) is developing in a cumulative way, as interdisciplinary teams collaborate in the development and sharing of applications that range in sophistication from simple composite mapping projects to intensive numerical and statistical simulation experiments.

2.4 Concluding comments

This chapter has presented a selection of GIS application areas and specific instances within each of the selected areas. Throughout, the emphasis has been on the range of contexts, from day-to-day problem solving to curiosity-driven science. The principles of the scientific method have been stressed throughout – the need to maintain an enquiring mind, constantly asking questions about what is going on, and what it means; the need to use terms that are well-defined and understood by others, so that knowledge can be communicated; the need to describe procedures in sufficient detail so that they can be replicated by others; and the need for accuracy, in observations, measurements, and predictions. These principles are valid whether the context is a simple inventory of the assets of a utility company, or the simulation of complex biological systems.

Questions for further study

1. Devise a diary for your own activity patterns for a typical (or a special) day, like that described in Section 2.1.1, and speculate how GIS might affect your own daily activities. What activities are not influenced by GIS, and how might its use in some of these contexts improve your daily quality of life?
2. Compare and contrast the operational, tactical, and strategic priorities of the GIS specialists responsible for the specific applications described in Sections 2.3.2, 2.3.3, 2.3.4 and 2.3.5.
3. Look at one of the applications chapters in the CD in the Longley et al (2005) volume in the references below. To what extent do you believe that the author of your chapter has demonstrated that GIS has been 'successful' in application? Suggest some of the implicit and explicit assumptions that are made in order to achieve a 'successful' outcome.
4. Look at one of the applications areas in the CD in the Longley et al (2005) volume in the references below. Then re-examine the list of critiques of GIS at the end of Section 1.7. To what extent do you think that the critiques are relevant to the applications that you have studied?

Further reading

Birkin M., Clarke G.P., and Clarke M. 2002 *Retail Geography and Intelligent Network Planning*. Chichester, UK: Wiley.

Chainey S., and Ratcliffe J. 2005 *GIS and Crime Mapping*. Chichester, UK: Wiley.

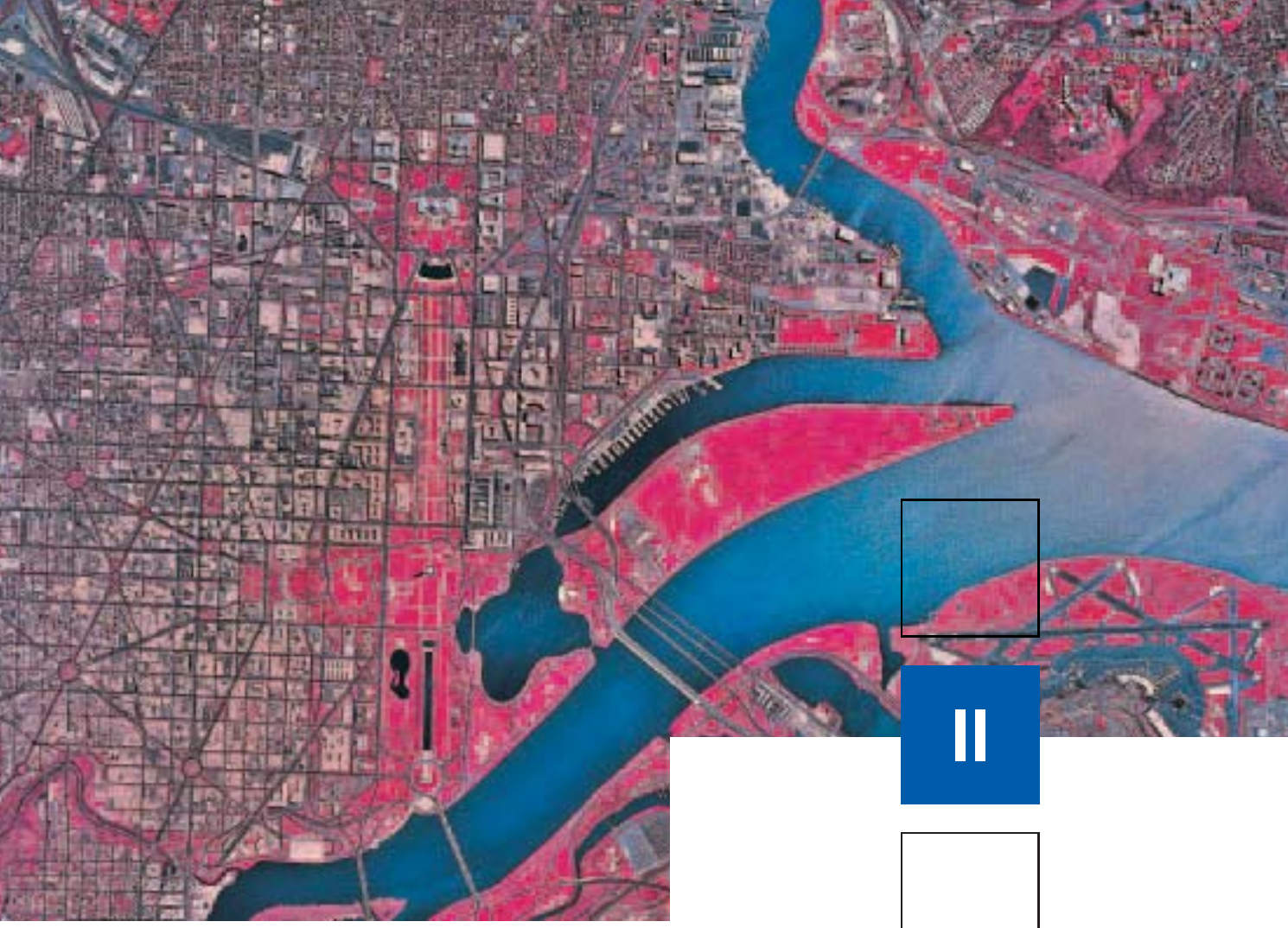
Greene R.W. 2000 *GIS in Public Policy*. Redlands, CA: ESRI Press.

Harris R., Sleight P., and Webber R. 2005 *Geodemographics, GIS and Neighbourhood Targeting*. Chichester, UK: Wiley.

Johnston C.A. 1998 *Geographic Information Systems in Ecology*. Oxford: Blackwell.

Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) 2005 *Geographical Information Systems: Principles; Techniques; Management and Applications (abridged edition)*. Hoboken, N.J.: Wiley.

O'Looney J. 2000 *Beyond Maps: GIS and Decision Making in Local Government*. Redlands, CA: ESRI Press.



II

Principles

3 *Representing geography*
4 *The nature of geographic data*
5 *Georeferencing*
6 *Uncertainty*

3 *Representing geography*

This chapter introduces the concept of representation, or the construction of a digital model of some aspect of the Earth's surface. Representations have many uses, because they allow us to learn, think, and reason about places and times that are outside our immediate experience. This is the basis of scientific research, planning, and many forms of day-to-day problem solving.

The geographic world is extremely complex, revealing more detail the closer one looks, almost *ad infinitum*. So in order to build a representation of any part of it, it is necessary to make choices, about what to represent, at what level of detail, and over what time period. The large number of possible choices creates many opportunities for designers of GIS software. Generalization methods are used to remove detail that is unnecessary for an application, in order to reduce data volume and speed up operations.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will know:

- The importance of understanding representation in GIS;
- The concepts of fields and objects and their fundamental significance;
- Raster and vector representation and how they affect many GIS principles, techniques, and applications;
- The paper map and its role as a GIS product and data source;
- The importance of generalization methods and the concept of representational scale;
- The art and science of representing real-world phenomena in GIS.

3.1 Introduction

We live on the surface of the Earth, and spend most of our lives in a relatively small fraction of that space. Of the approximately 500 million square kilometers of surface, only one third is land, and only a fraction of that is occupied by the cities and towns in which most of us live. The rest of the Earth, including the parts we never visit, the atmosphere, and the solid ground under our feet, remains unknown to us except through the information that is communicated to us through books, newspapers, television, the Web, or the spoken word. We live lives that are almost infinitesimal in comparison with the 4.5 billion years of Earth history, or the over 10 billion years since the universe began, and know about the Earth before we were born only through the evidence compiled by geologists, archaeologists, historians, etc. Similarly, we know nothing about the world that is to come, where we have only predictions to guide us.

Because we can observe so little of the Earth directly, we rely on a host of methods for learning about its other parts, for deciding where to go as tourists or shoppers, choosing where to live, running the operations of corporations, agencies, and governments, and many other activities. Almost all human activities at some time require knowledge (Section 1.2) about parts of the Earth that are outside our direct experience, because they occur either elsewhere in space, or elsewhere in time. Sometimes this knowledge is used as a *substitute* for directly sensed information, creating a *virtual* reality (see Section 11.3.1). Increasingly it is used to *augment* what we can see, touch, hear, feel, and smell, through the use of mobile information systems that can be carried around.

Our knowledge of the Earth is not created entirely freely, but must fit with the mental concepts we began to develop as young children – concepts such as containment (Paris is *in* France) or proximity (Dallas and Fort Worth are *close*). In digital representations, we formalize these concepts through *data models* (Chapter 8), the structures and rules that are programmed into a GIS to accommodate data. These concepts and data models together constitute our *ontologies*, the frameworks that we use for acquiring knowledge of the world.

Almost all human activities require knowledge about the Earth – past, present, or future.

One such ontology, a way to structure knowledge of movement through time, is a three-dimensional diagram, in which the two horizontal axes denote location on the Earth's surface, and the vertical axis denotes time. In Figure 3.1, the daily lives of a sample of residents of Lexington, Kentucky, USA are shown as they move by car through space and time, from one location to another, while going about their daily business of shopping, traveling to work, or dropping children at school. The diagram is crude, because each journey is represented by a series of straight lines between locations measured with GPS, and if we were able to examine each track or trajectory in more detail we would see the effects of having to follow streets, stopping at traffic lights, or slowing for congestion. If we looked even closer we might see details of each person's walk to and from the car. Each closer perspective would display more information, and a vast storehouse would be required to capture the precise trajectories of all humans throughout even a single day.

The real trajectories of the individuals shown in Figure 3.1 are complex, and the figure is only a representation of them – a model on a piece of paper, generated by a computer from a database. We use the terms *representation* and *model* because they imply a simplified relationship between the contents of the figure and the database, and the real-world trajectories of the individuals. Such representations or models serve many useful purposes, and occur in many different forms. For example, representations occur:

- in the human mind, when our senses capture information about our surroundings, such as the images captured by the eye, or the sounds captured by the ear, and memory preserves such representations for future use;
- in photographs, which are two-dimensional models of the light emitted or reflected by objects in the world into the lens of a camera;
- in spoken descriptions and written text, in which people describe some aspect of the world in language, in the form of travel accounts or diaries; or

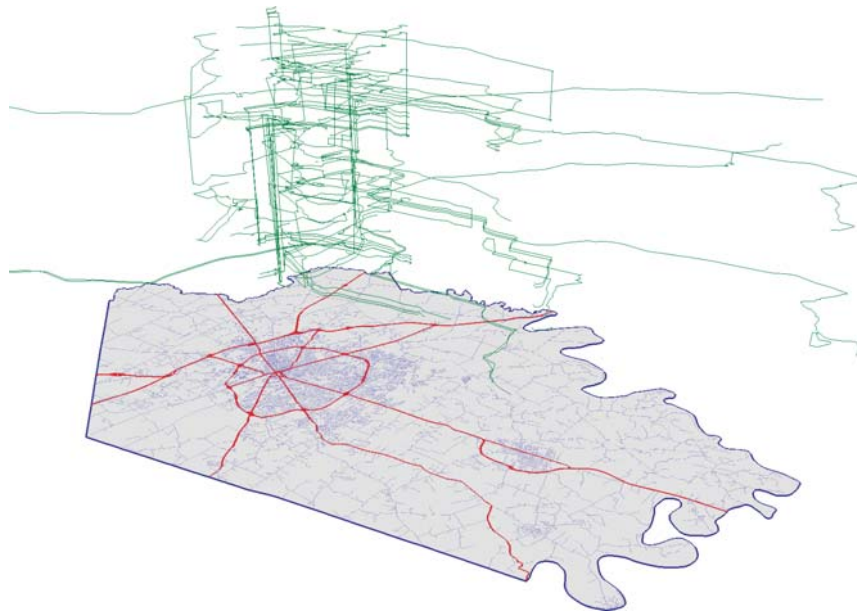


Figure 3.1 Schematic representation of the daily journeys of a sample of residents of Lexington, Kentucky, USA. The horizontal dimensions represent geographic space and the vertical dimension represents time of day. Each person's track plots as a three-dimensional line, beginning at the base in the morning and ending at the top in the evening. (Reproduced with permission of Mei-Po Kwan)

- in the numbers that result when aspects of the world are measured, using such devices as thermometers, rulers, or speedometers.

By building representations, we humans can assemble far more knowledge about our planet than we ever could as individuals. We can build representations that serve such purposes as planning, resource management and conservation, travel, or the day-to-day operations of a parcel delivery service.

Representations help us assemble far more knowledge about the Earth than is possible on our own.

Representations are reinforced by the rules and laws that we humans have learned to apply to the unobserved world around us. When we encounter a fallen log in a forest we are willing to assert that it once stood upright, and once grew from a small shoot, even though no one actually observed or reported either of these stages. We predict the future occurrence of eclipses based on the laws we have discovered about the motions of the Solar System. In GIS applications, we often rely on methods of spatial interpolation to guess the conditions that exist in places where no observations were made, based on the rule (often elevated to the status of a First Law of Geography and attributed to Waldo Tobler) that all places are similar, but nearby places are more similar than distant places.

Tobler's First Law of Geography: Everything is related to everything else, but near things are more related than distant things.

3.2 Digital representation

This book is about one particular form of representation that is becoming increasingly important in our society – representation in digital form. Today, almost all communication between people through such media as the telephone, FAX, music, television, newspapers and magazines, or email is at some time in its life in digital form. Information technology based on digital representation is moving into all aspects of our lives, from science to commerce to daily existence. Almost half of all households in some industrial societies now own at least one powerful digital information processing device (a computer); a large proportion of all work in offices now occurs using digital computing technology; and digital technology has invaded many devices that we use every day, from the microwave oven to the automobile.

One interesting characteristic of digital technology is that the representation itself is rarely if ever seen by the user, because only a few technical experts ever see the individual elements of a digital representation. What we see instead are *views*, designed to present the contents of the representation in a form that is meaningful to us.

The term *digital* derives from *digits*, or the fingers, and our system of counting based on the ten digits of the human hand. But while the counting system has ten symbols (0 through 9), the representation system in digital computers uses only two (0 and 1). In a sense, then, the term *digital* is a misnomer for a system that represents all information using some combination of the two symbols 0 and 1, and the more exact term *binary* is more appropriate. In this book we follow the convention

of using *digital* to refer to electronic technology based on binary representations.

Computers represent phenomena as binary digits. Every item of useful information about the Earth's surface is ultimately reduced by a GIS to some combination of 0s and 1s.

Over the years many standards have been developed for converting information into digital form. Box 3.1 shows the standards that are commonly used in GIS to store data, whether they consist of whole or decimal numbers or text. There are many competing coding standards for images and photographs (GIF, JPEG, TIFF, etc.) and for movies (e.g., MPEG) and sound (e.g., MIDI, MP3). Much of this book is about the coding systems used to represent geographic data, especially Chapter 8, and as you might guess that turns out to be comparatively complicated.

Digital technology is successful for many reasons, not the least of which is that all kinds of information share a common basic format (0s and 1s), and can be handled in ways that are largely independent of their actual meaning. The Internet, for example, operates on the basis of packets of information, consisting of strings of 0s and 1s, which are sent through the network based on the information contained in the packet's header. The network needs to know only what the header means, and how to read the instructions it contains regarding the packet's destination. The rest of the contents are no more than a collection of bits, representing anything from an email message to a short burst of music or highly secret information on its way from one military installation to another, and are almost never examined or interpreted during transmission. This allows one digital communications network to serve every need, from electronic commerce to chatrooms, and it allows manufacturers to build processing and storage technology for vast numbers of users who have very different applications in mind. Compare this to earlier ways of communicating, which required printing presses and delivery trucks for one application (newspapers) and networks of copper wires for another (telephone).

Digital representations of geography hold enormous advantages over previous types – paper maps, written reports from explorers, or spoken accounts. We can use

Technical Box **3.1**

The binary counting system

The binary counting system uses only two symbols, 0 and 1, to represent numerical information. A group of eight binary digits is known as a *byte*, and volume of storage is normally measured in bytes rather than bits (Table 1.1). There are only two options for a single digit, but there are four possible combinations for two digits (00, 01, 10, and 11), eight possible combinations for three digits (000, 001, 010, 011, 100, 101, 110, 111), and 256 combinations for a full byte. Digits in the binary system (known as binary digits, or *bits*) behave like digits in the decimal system but using powers of two. The rightmost digit denotes units, the next digit to the left denotes twos, the next to the left denotes fours, etc. For example, the binary number 11001 denotes one unit, no twos, no fours, one eight, and one sixteen, and is equivalent to 25 in the normal (decimal) counting system. We call this the *integer* digital representation of 25, because it represents 25 as a whole number, and is readily amenable to arithmetic operations. Whole numbers are commonly stored in GIS using either *short* (2-byte or 16-bit) or *long* (4-byte or 32-bit) options. Short integers can range from −65535 to +65535, and long integers from −4294967295 to +4294967295.

The 8-bit ASCII (American Standard Code for Information Interchange) system assigns codes to each symbol of text, including letters, numbers, and common symbols. The number 2 is assigned ASCII code 48 (00110000 in binary), and the number 5 is 53 (00110101), so if 25 were coded as two characters using 8-bit ASCII its digital representation would be 16 bits long (0011000000110101). The characters 2 = 2 would be coded as 48, 61, 48 (001100000011110100110000). ASCII is used for coding text, which consists of mixtures of letters, numbers, and punctuation symbols.

Numbers with decimal places are coded using *real* or *floating-point* representations. A number such as 123.456 (three decimal places and six significant digits) is first transformed by powers of ten so that the decimal point is in a standard position, such as the beginning (e.g., 0.123456×10^3). The fractional part (0.123456) and the power of 10 (3) are then stored in separate sections of a block of either 4 bytes (32 bits, *single* precision) or 8 bytes (64 bits, *double* precision). This gives enough precision to store roughly 7 significant digits in single precision, or 14 in double precision.

Integer, ASCII, and real conventions are adequate for most data, but in some cases it is desirable to associate images or sounds with places in GIS, rather than text or numbers. To allow for this GIS designers have included a BLOB option (standing for binary large object), which simply allocates a sufficient number of bits to store the image or sound, without specifying what those bits might mean.

the same cheap digital devices – the components of PCs, the Internet, or mass storage devices – to handle every type of information, independent of its meaning. Digital data are easy to copy, they can be transmitted at close to the speed of light, they can be stored at high density in very small spaces, and they are less subject to the physical deterioration that affects paper and other physical media. Perhaps more importantly, data in digital form are easy to transform, process, and analyze. Geographic information systems allow us to do things with digital representations that we were never able to do with paper maps: to measure accurately and quickly, to overlay and combine, and to change scale, zoom, and pan without respect to map sheet boundaries. The vast array of possibilities for processing that digital representation opens up is reviewed in Chapters 14 through 16, and is also covered in the applications that are distributed throughout the book.

Digital representation has many uses because of its simplicity and low cost.

3.3 Representation for what and for whom?

Thus far we have seen how humans are able to build representations of the world around them, but we have not yet discussed why representations are useful, and why humans have become so ingenious at creating and sharing them. The emphasis here and throughout the book is on one type of representation, termed *geographic*, and defined as a representation of some part of the Earth's surface or near-surface, at scales ranging from the architectural to the global.

Geographic representation is concerned with the Earth's surface or near-surface.

Geographic representations are among the most ancient, having their roots in the needs of very early societies. The tasks of hunting and gathering can be much more efficient if hunters are able to communicate the details of their successes to other members of their group – the locations of edible roots or game, for example. Maps must have originated in the sketches early people made in the dirt of campgrounds or on cave walls, long before language became sufficiently sophisticated to convey equivalent information through speech. We know that the peoples of the Pacific built representations of the locations of islands, winds, and currents out of simple materials to guide each other, and that very simple forms of representation are used by social insects such as bees to communicate the locations of food resources.

Hand-drawn maps and speech are effective media for communication between members of a small group, but much wider communication became possible with the invention of the printing press in the 15th century. Now large numbers of copies of a representation could be made and distributed, and for the first time it became possible to imagine that something could be known by every human being – that knowledge could be the common property of humanity. Only one major restriction affected what could be distributed using this new mechanism: the representation had to be flat. If one were willing to accept that constraint, however, paper proved to be enormously effective; it was cheap, light and thus easily transported, and durable. Only fire and water proved to be disastrous for paper, and human history is replete with instances of the loss of vital information through fire or flood, from the burning of the Alexandria Library in the 7th century that destroyed much of the accumulated knowledge of classical times to the major conflagrations of London in 1666, San Francisco in 1906, or Tokyo in 1945, and the flooding of the Arno that devastated Florence in 1966.

One of the most important periods for geographic representation began in the early 15th century in Portugal. Henry the Navigator (Box 3.2) is often credited with originating the Age of Discovery, the period of European history that led to the accumulation of large amounts of information about other parts of the world through sea voyages and land explorations. Maps became the medium for sharing information about new discoveries, and for administering vast colonial empires, and their value was quickly recognized. Although detailed representations now exist of all parts of the world, including Antarctica, in a sense the spirit of the Age of Discovery continues in the explorations of the oceans, caves, and outer space, and in the process of re-mapping that is needed to keep up with constant changes in the human and natural worlds.

It was the creation, dissemination, and sharing of accurate representations that distinguished the Age of Discovery from all previous periods in human history (and it would be unfair to ignore its distinctive negative consequences, notably the spread of European diseases and the growth of the slave trade). Information about other parts of the world was assembled in the form of maps and journals, reproduced in large numbers using the recently invented printing press, and distributed on paper. Even the modest costs associated with buying copies were eventually addressed through the development of free public lending libraries in the 19th century, which gave access to virtually everyone. Today, we benefit from what is now a longstanding tradition of free and open access to much of humanity's accumulated store of knowledge about the geographic world, in the form of paper-based representations, through the institution of libraries and the copyright doctrine that gives people rights to material for personal use (see Chapter 18 for a discussion of laws affecting ownership and access). The Internet has already become the delivery mechanism for providing distributed access to geographic information.

In the Age of Discovery maps became extremely valuable representations of the state of geographic knowledge.

It is not by accident that the list of important applications for geographic representations closely follows the list of applications of GIS (see Section 1.1 and Chapter 2),

Biographical Box **3.2**

Prince Henry the Navigator

Prince Henry of Portugal, who died in 1460, was known as Henry the Navigator because of his keen interest in exploration. In 1433 Prince Henry sent a ship from Portugal to explore the west coast of Africa in an attempt to find a sea route to the Spice Islands. This ship was the first to travel south of Cape Bojador (latitude 26 degrees 20 minutes N). To make this and other voyages Prince Henry assembled a team of map-makers, sea captains, geographers, ship builders, and many other skilled craftsmen. Prince Henry showed the way for Vasco da Gama and other famous 15th century explorers. His management skills could be applied in much the same way in today's GIS projects.

Figure 3.2 Prince Henry the Navigator, originator of the Age of Discovery in the 15th century, and promoter of a systematic approach to the acquisition, compilation, and dissemination of geographic knowledge

since representation is at the heart of our ability to solve problems using digital tools. Any application of GIS requires clear attention to questions of *what* should be represented, and *how*. There is a multitude of possible ways of representing the geographic world in digital form, none of which is perfect, and none of which is ideal for all applications.

The key GIS representation issues are what to represent and how to represent it.

One of the most important criteria for the usefulness of a representation is its *accuracy*. Because the geographic world is seemingly of infinite complexity, there are always choices to be made in building any representation – what to include, and what to leave out. When US President Thomas Jefferson dispatched Meriwether Lewis to explore and report on the nature of the lands from the upper Missouri to the Pacific, he said Lewis possessed 'a fidelity to the truth so scrupulous that whatever he should report would be as certain as if seen by ourselves'. But he clearly didn't expect Lewis to report everything he saw in complete detail: Lewis exercised a large amount of judgment about what to report, and what to omit. The question of accuracy is taken up at length in Chapter 6.

One more vital interest drives our need for representations of the geographic world, and also the need for representations in many other human activities. When a pilot must train to fly a new type of aircraft, it is much cheaper and less risky for him or her to work with a flight simulator than with the real aircraft. Flight simulators can represent a much wider range of conditions than a pilot will normally experience in flying. Similarly, when decisions have to be made about the geographic world, it is effective to experiment first on models or representations, exploring different scenarios. Of course this works only if the representation behaves as the real aircraft or world does, and a great deal of knowledge must be acquired about the world before an accurate representation can be built that permits such simulations. But the use of representations for training, exploring future scenarios, and recreating the past is now common in many fields, including surgery, chemistry, and engineering, and with technologies like GIS is becoming increasingly common in dealing with the geographic world.

Many plans for the real world can be tried out first on models or representations.

3.4 The fundamental problem

Geographic data are built up from atomic elements, or facts about the geographic world. At its most primitive, an atom of geographic data (strictly, a datum) links a place, often a time, and some descriptive property. The first of these, place, is specified in one of several ways that are discussed at length in Chapter 5, and there are also many ways of specifying the second, time. We often use the term *attribute* to refer to the last of these three. For example, consider the statement 'The temperature at local noon on December 2nd 2004 at latitude 34 degrees

45 minutes north, longitude 120 degrees 0 minutes west, was 18 degrees Celsius'. It ties location and time to the property or attribute of atmospheric temperature.

Geographic data link place, time, and attributes.

Other facts can be broken down into their primitive atoms. For example, the statement 'Mount Everest is 8848 m high' can be derived from two atomic geographic facts, one giving the location of Mt Everest in latitude and longitude, and the other giving the elevation at that latitude and longitude. Note, however, that the statement would not be a geographic fact to a community that had no way of knowing where Mt Everest is located.

Many aspects of the Earth's surface are comparatively static and slow to change. Height above sea level changes slowly because of erosion and movements of the Earth's crust, but these processes operate on scales of hundreds or thousands of years, and for most applications except geophysics we can safely omit time from the representation of elevation. On the other hand atmospheric temperature changes daily, and dramatic changes sometimes occur in minutes with the passage of a cold front or thunderstorm, so time is distinctly important, though such climatic variables as mean annual temperature can be represented as static.

The range of attributes in geographic information is vast. We have already seen that some vary slowly and some rapidly. Some attributes are physical or environmental in nature, while others are social or economic. Some attributes simply *identify* a place or an entity, distinguishing it from all other places or entities – examples include street addresses, social security numbers, or the parcel numbers used for recording land ownership. Other attributes measure something at a location and perhaps at a time (e.g., atmospheric temperature or elevation), while others classify into categories (e.g., the class of land use, differentiating between agriculture, industry, or residential land). Because attributes are important outside the domain of GIS there are standard terms for the different types (see Box 3.3).

Geographic attributes are classified as nominal, ordinal, interval, ratio, and cyclic.

But this idea of recording atoms of geographic information, combining location, time, and attribute, misses a fundamental problem, which is that the world is in effect infinitely complex, and the number of atoms required for a complete representation is similarly infinite. The closer we look at the world, the more detail it reveals – and it seems that this process extends *ad infinitum*. The shoreline of Maine appears complex on a map, but even more complex when examined in greater detail, and as more detail is revealed the shoreline appears to get longer and longer,

Technical Box **3.3**

Types of attributes

The simplest type of attribute, termed *nominal*, is one that serves only to identify or distinguish one entity from another. Placenames are a good example, as are names of houses, or the numbers on a driver's license – each serves only to identify the particular instance of a class of entities and to distinguish it from other members of the same class. Nominal attributes include numbers, letters, and even colors. Even though a nominal attribute can be numeric it makes no sense to apply arithmetic operations to it: adding two nominal attributes, such as two drivers' license numbers, creates nonsense.

Attributes are *ordinal* if their values have a natural order. For example, Canada rates its agricultural land by classes of soil quality, with Class 1 being the best, Class 2 not so good, etc. Adding or taking ratios of such numbers makes little sense, since 2 is not twice as much of anything as 1, but at least ordinal attributes have inherent order. Averaging makes no sense either, but the *median*, or the value such that half of the attributes are higher-ranked and half are lower-ranked, is an effective substitute for the average for ordinal data as it gives a useful central value.

Attributes are *interval* if the differences between values make sense. The scale of Celsius temperature is interval, because it makes sense to say that 30 and 20 are as different as 20 and 10. Attributes are *ratio* if the ratios between values make sense. Weight is ratio, because it makes sense to say that a person of 100 kg is twice as heavy as a person of 50 kg; but Celsius temperature is only interval, because 20 is not twice as hot as 10 (and this argument applies to all scales that are based on similarly arbitrary zero points, including longitude).

In GIS it is sometimes necessary to deal with data that fall into categories beyond these four. For example, data can be directional or *cyclic*, including flow direction on a map, or compass direction, or longitude, or month of the year. The special problem here is that the number following 359 degrees is 0. Averaging two directions such as 359 and 1 yields 180, so the average of two directions close to North can appear to be South. Because cyclic data occur sometimes in GIS, and few designers of GIS software have made special arrangements for them, it is important to be alert to the problems that may arise.

and more and more convoluted (see Figure 4.18). To characterize the world completely we would have to specify the location of every person, every blade of grass, and every grain of sand – in fact, every subatomic particle, clearly an impossible task, since the Heisenberg uncertainty principle places limits on the ability to measure precise positions of subatomic particles. So in practice any representation must be partial – it must limit the level of detail provided, or ignore change through time, or ignore certain attributes, or simplify in some other way.

The world is infinitely complex, but computer systems are finite. Representations must somehow limit the amount of detail captured.

One very common way of limiting detail is by throwing away or ignoring information that applies only to small areas, in other words not looking too closely. The image you see on a computer screen is composed of a million or so basic elements or *pixels*, and if the whole Earth were displayed at once each pixel would cover an area roughly 10 km on a side, or about 100 sq km. At this level of detail the island of Manhattan occupies roughly 10 pixels, and virtually everything on it is a blur. We would say that such an image has a *spatial resolution* of about 10 km, and know that anything much less than 10 km across is virtually invisible. Figure 3.3 shows Manhattan at a spatial resolution of 250 m, detailed enough to pick out the shape of the island and Central Park.

It is easy to see how this helps with the problem of too much information. The Earth's surface covers about 500 million sq km, so if this level of detail is sufficient for an application, a property of the surface such as elevation can be described with only 5 million pieces of information, instead of the 500 million it would take to describe elevation with a resolution of 1 km, and the 500 trillion (500 000 000 000 000) it would take to describe elevation with 1 m resolution.

Another strategy for limiting detail is to observe that many properties remain constant over large areas. For example, in describing the elevation of the Earth's surface we could take advantage of the fact that roughly two-thirds of the surface is covered by water, with its surface at sea level. Of the 5 million pieces of information needed to describe elevation at 10 km resolution, approximately 3.4 million will be recorded as zero, a colossal waste. If we could find an efficient way of identifying the area covered by water, then we would need only 1.6 million real pieces of information.

Humans have found many ingenious ways of describing the Earth's surface efficiently, because the problem we are addressing is as old as representation itself, and as important for paper-based representations as it is for binary representations in computers. But this ingenuity is itself the source of a substantial problem for GIS: there are many ways of representing the Earth's surface, and users of GIS thus face difficult and at times confusing choices. This chapter discusses some of those choices, and the issues are pursued further in subsequent chapters on uncertainty (Chapter 6) and data modeling (Chapter 8). Representation remains a major concern of GIScience, and researchers are constantly looking for ways to extend GIS representations to accommodate new types of information (Box 3.5).

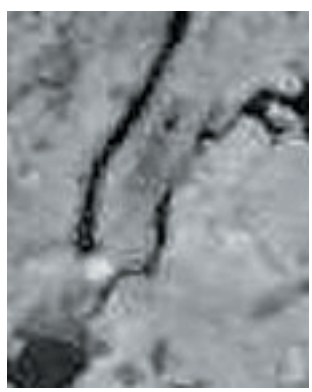


Figure 3.3 An image of Manhattan taken by the MODIS instrument on board the TERRA satellite on September 12, 2001. MODIS has a spatial resolution of about 250 m, detailed enough to reveal the coarse shape of Manhattan and to identify the Hudson and East Rivers, the burning World Trade Center (white spot), and Central Park (the gray blur with the Jacqueline Kennedy Onassis Reservoir visible as a black dot)

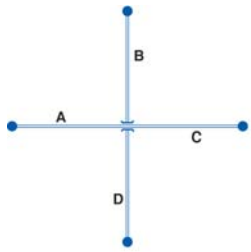
3.5 Discrete objects and continuous fields

3.5.1 Discrete objects

Mention has already been made of the level of detail as a fundamental choice in representation. Another, perhaps even more fundamental choice, is between two conceptual schemes. There is good evidence that we as humans like to simplify the world around us by naming things, and seeing individual things as instances of broader categories. We prefer a world of black and white, of good guys and bad guys, to the real world of shades of gray.

The two fundamental ways of representing geography are discrete objects and continuous fields.

This preference is reflected in one way of viewing the geographic world, known as the *discrete object* view. In this view, the world is empty, except where it is occupied by objects with well-defined boundaries that are instances of generally recognized categories. Just as the desktop is littered with books, pencils, or computers, the geographic world is littered with cars, houses, lampposts, and other discrete objects. Thus the landscape of Minnesota is littered with lakes, and the landscape of Scotland is littered with mountains. One characteristic of the discrete object view is that objects can be counted, so license plates issued by the State of Minnesota carry



From link	To link	Turn?
A	B	No
A	C	Yes
A	D	No
B	C	No
B	D	Yes
B	A	No
C	D	No
C	A	Yes
C	B	No
D	A	No
D	B	Yes
D	C	No

Figure 3.4 The problems of representing a three-dimensional world using a two-dimensional technology. The intersection of links A, B, C, and D is an overpass, so no turns are possible between such pairs as A and B

the legend '10 000 lakes', and climbers know that there are exactly 284 mountains in Scotland over 3000 ft (the so-called Munros, from Sir Hugh Munro who originally listed 277 of them in 1891 – the count was expanded to 284 in 1997).

The discrete object view represents the geographic world as objects with well-defined boundaries in otherwise empty space.

Biological organisms fit this model well, and this allows us to count the number of residents in an area of a city, or to describe the behavior of individual bears. Manufactured objects also fit the model, and we have little difficulty counting the number of cars produced in a year, or the number of airplanes owned by an airline. But other phenomena are messier. It is not at all clear what constitutes a mountain, for example, or exactly how a mountain differs from a hill, or when a mountain with two peaks should be counted as two mountains.

Geographic objects are identified by their dimensionality. Objects that occupy area are termed two-dimensional, and generally referred to as areas. The term *polygon* is also common for technical reasons explained later. Other objects are more like one-dimensional lines, including roads, railways, or rivers, and are often represented as one-dimensional objects and generally referred to as lines. Other objects are more like zero-dimensional points, such as individual animals or buildings, and are referred to as points.

Of course, in reality, all objects that are perceptible to humans are three dimensional, and their representation in fewer dimensions can be at best an approximation. But the ability of GIS to handle truly three-dimensional objects as volumes with associated surfaces is very limited. Some GIS allow for a third (vertical) coordinate to be specified for all point locations. Buildings are sometimes represented by assigning height as an attribute, though if this option is used it is impossible to distinguish flat roofs from any other kind. Various strategies have been used for representing overpasses and underpasses in transportation networks, because this information is vital for navigation but not normally represented in strictly two-dimensional network representations. One common strategy is to represent turning options at every intersection – so an overpass appears in the database as an intersection with no turns (Figure 3.4).

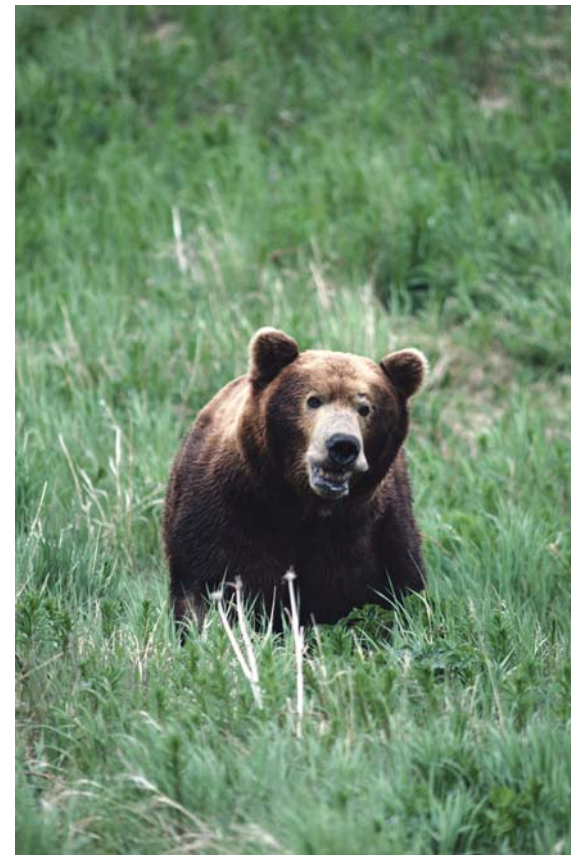


Figure 3.5 Bears are easily conceived as discrete objects, maintaining their identity as objects through time and surrounded by empty space

The discrete object view leads to a powerful way of representing geographic information about objects. Think of a class of objects of the same dimensionality – for example, all of the Brown bears (Figure 3.5) in the Kenai Peninsula of Alaska. We would naturally think of these objects as points. We might want to know the sex of each bear, and its date of birth, if our interests were in monitoring the bear population. We might also have a collar on each bear that transmitted the bear's location at regular intervals. All of this information could be expressed in a table, such as the one shown in Table 3.1, with each row corresponding to a different discrete object, and each column to an attribute of the object. To reinforce a point made earlier, this is a very efficient way of capturing raw geographic information on Brown bears.

But it is not perfect as a representation for all geographic phenomena. Imagine visiting the Earth from another planet, and asking the humans what they chose as a representation for the infinitely complex and beautiful environment around them. The visitor would hardly be impressed to learn that they chose tables, especially when the phenomena represented were natural phenomena such as rivers, landscapes, or oceans. Nothing on the natural Earth looks remotely like a table. It is not at all clear how the properties of a river should be represented as a table, or the properties of an ocean. So while the discrete object

Table 3.1 Example of representation of geographic information as a table: the locations and attributes of each of four Brown bears in the Kenai Peninsula of Alaska. Locations have been obtained from radio collars. Only one location is shown for each bear, at noon on July 31 2003 (imaginary data)

Bear ID	Sex	Estimated year of birth	Date of collar installation	Location, noon on 31 July 2003
001	M	1999	02242003	−150.6432, 60.0567
002	F	1997	03312003	−149.9979, 59.9665
003	F	1994	04212003	−150.4639, 60.1245
004	F	1995	04212003	−150.4692, 60.1152

view works well for some kinds of phenomena, it misses the mark badly for others.

3.5.2 Continuous fields

While we might think of terrain as composed of discrete mountain peaks, valleys, ridges, slopes, etc., and think of listing them in tables and counting them, there are unresolvable problems of definition for all of these objects. Instead, it is much more useful to think of terrain as a continuous surface, in which elevation can be defined rigorously at every point (see Box 3.4). Such continuous surfaces form the basis of the other common view of geographic phenomena, known as the *continuous field* view (and not to be confused with other meanings of the word field). In this view the geographic world can be described by a number of *variables*, each measurable at any point on the Earth's surface, and changing in value across the surface.

The continuous field view represents the real world as a finite number of variables, each one defined at every possible position.

Objects are distinguished by their dimensions, and naturally fall into categories of points, lines, or areas. Continuous fields, on the other hand, can be distinguished by what varies, and how smoothly. A continuous field of elevation, for example, varies much more smoothly in a landscape that has been worn down by glaciation or flattened by blowing sand than one recently created by cooling lava. Cliffs are places in continuous fields where elevation changes suddenly, rather than smoothly. Population density is a kind of continuous field, defined everywhere as the number of people per unit area, though the definition breaks down if the field is examined so closely that the individual people become visible. Continuous fields can also be created from classifications of land, into categories of land use, or soil type. Such fields change suddenly at the boundaries between different classes. Other types of fields can be defined by continuous variation along lines, rather than across space. Traffic density, for example, can be defined everywhere on a road network, and flow volume can be defined everywhere on a river. Figure 3.6 shows some examples of field-like phenomena.

Continuous fields can be distinguished by *what* is being measured at each point. Like the attribute types discussed in Box 3.3, the variable may be nominal, ordinal, interval, ratio, or cyclic. A *vector* field assigns two variables, magnitude and direction, at every point in space, and is used to represent flow phenomena such as winds or currents; fields of only one variable are termed *scalar* fields.

Here is a simple example illustrating the difference between the discrete object and field conceptualizations. Suppose you were hired for the summer to count the number of lakes in Minnesota, and promised that your answer would appear on every license plate issued by the state. The task sounds simple, and you were happy to get the job. But on the first day you started to run into difficulty (Figure 3.7). What about small ponds, do they count as lakes? What about wide stretches of rivers? What about swamps that dry up in the summer? What about a lake with a narrow section connecting two wider parts, is it one lake or two? Your biggest dilemma concerns the scale of mapping, since the number of lakes shown on a map clearly depends on the map's level of detail – a more detailed map almost certainly will show more lakes.

Your task clearly reflects a discrete object view of the phenomenon. The action of counting implies that lakes are discrete, two-dimensional objects littering an otherwise empty geographic landscape. In a continuous field view, on the other hand, all points are either lake or non-lake. Moreover, we could refine the scale a little to take account

Technical Box **3.4**

2.5 dimensions

Areas are two-dimensional objects, and volumes are three dimensional, but GIS users sometimes talk about '2.5-D'. Almost without exception the elevation of the Earth's surface has a single value at any location (exceptions include overhanging cliffs). So elevation is conveniently thought of as a continuous field, a variable with a value everywhere in two dimensions, and a full 3-D representation is only necessary in areas with an abundance of overhanging cliffs or caves, if these are important features. The idea of dealing with a three-dimensional phenomenon by treating it as a single-valued function of two horizontal variables gives rise to the term '2.5-D'. Figure 3.6B shows an example, in this case an elevation surface.

(A)

(B)

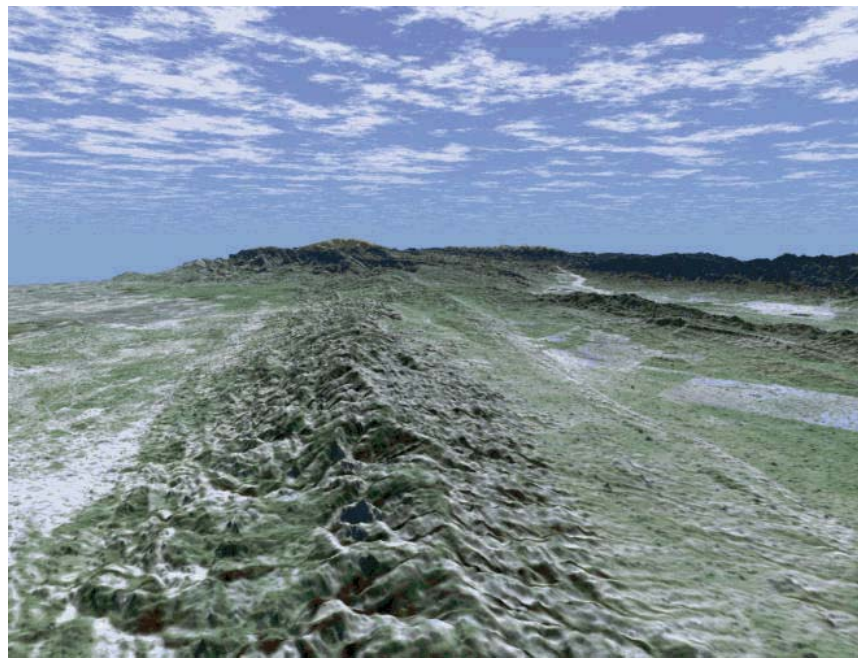


Figure 3.6 Examples of field-like phenomena. (A) Image of part of the Dead Sea in the Middle East. The lightness of the image at any point measures the amount of radiation captured by the satellite's imaging system. (B) A simulated image derived from the Shuttle Radar Topography Mission, a new source of high-quality elevation data. The image shows the Carrizo Plain area of Southern California, USA, with a simulated sky and with land cover obtained from other satellite sources (Courtesy NASA/JPL–Caltech)

of marginal cases; for example, we might define the scale shown in Table 3.2, which has five degrees of lakeness. The complexity of the view would depend on how closely we looked, of course, and so the scale of mapping would still be important. But all of the problems of defining a lake as a discrete object would disappear (though there would still be problems in defining the levels of the scale). Instead of counting, our strategy would be to lay a grid over the map, and assign each grid cell a score on the lakeness scale. The size of the grid cell would determine how accurately the result approximated the value we could theoretically obtain by visiting every one of the infinite

Figure 3.7 Lakes are difficult to conceptualize as discrete objects because it is often difficult to tell where a lake begins and ends, or to distinguish a wide river from a lake

Table 3.2 A scale of lakeness suitable for defining lakes as a continuous field

Lakeness	Definition
1	Location is always dry under all circumstances
2	Location is sometimes flooded in Spring
3	Location supports marshy vegetation
4	Water is always present to a depth of less than 1 m
5	Water is always present to a depth of more than 1 m

number of points in the state. At the end, we would tabulate the resulting scores, counting the number of cells having each value of lakeness, or averaging the lakeness score. We could even design a new and scientifically more reasonable license plate – 'Minnesota, 12% lake' or 'Minnesota, average lakeness 2.02'.

The difference between objects and fields is also illustrated well by photographs (e.g., Figure 3.6A). The image in a photograph is created by variation in the chemical state of the material in the photographic film – in early photography, minute particles of silver were released from molecules of silver nitrate when the unstable molecules were exposed to light, thus darkening the image in proportion to the amount of incident light. We think of the image as a field of continuous variation in color or darkness. But when we look at the image, the eye and brain begin to infer the presence of discrete objects, such as people, rivers, fields, cars, or houses, as they interpret the content of the image.

3.6 Rasters and vectors

Continuous fields and discrete objects define two conceptual views of geographic phenomena, but they do not solve the problem of digital representation. A continuous field view still potentially contains an infinite amount of information if it defines the value of the variable at every point, since there is an infinite number of points in any defined geographic area. Discrete objects can also require an infinite amount of information for full description – for example, a coastline contains an infinite amount of information if it is mapped in infinite detail. Thus continuous fields and discrete objects are no more than conceptualizations, or ways in which we think about geographic phenomena; they are not designed to deal with the limitations of computers.

Two methods are used to reduce geographic phenomena to forms that can be coded in computer databases, and we call these raster and vector. In principle, both can be used to code both fields and discrete objects, but in practice there is a strong association between raster and fields, and between vector and discrete objects.

Raster and vector are two methods of representing geographic data in digital computers.

3.6.1 Raster data

In a raster representation space is divided into an array of rectangular (usually square) cells (Figure 3.8). All geographic variation is then expressed by assigning properties or attributes to these cells. The cells are sometimes called pixels (short for *picture elements*).

Raster representations divide the world into arrays of cells and assign attributes to the cells.

One of the commonest forms of raster data comes from remote-sensing satellites, which capture information in this form and send it to ground to be distributed and analyzed. Data from the Landsat Thematic Mapper, for example, which are commonly used in GIS applications, come in cells that are 30 m a side on the ground, or approximately 0.1 hectare in area. Other similar data can be obtained from sensors mounted on aircraft. Imagery varies according to the spatial resolution (expressed as the length of a cell side as measured on the ground), and also according to the timetable of image capture by the

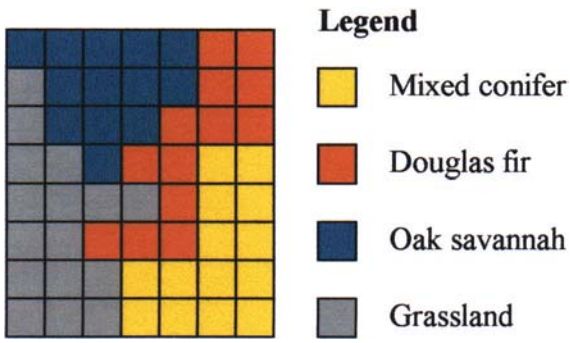


Figure 3.8 Raster representation. Each color represents a different value of a nominal-scale variable denoting land cover class

sensor. Some satellites are in *geostationary* orbit over a fixed point on the Earth, and capture images constantly. Others pass over a fixed point at regular intervals (e.g., every 12 days). Finally, sensors vary according to the part or parts of the spectrum that they sense. The visible parts of the spectrum are most important for remote sensing, but some invisible parts of the spectrum are particularly useful in detecting heat, and the phenomena that produce heat, such as volcanic activities. Many sensors capture images in several areas of the spectrum, or *bands*, simultaneously, because the relative amounts of radiation in different parts of the spectrum are often useful indicators of certain phenomena, such as green leaves, or water, on the Earth's surface. The AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) captures no fewer than 224 different parts of the spectrum, and is being used to detect particular minerals in the soil, among other applications. Remote sensing is a complex topic, and further details are available in Chapter 9.

Square cells fit together nicely on a flat table or a sheet of paper, but they will not fit together neatly on the curved surface of the Earth. So just as representations on paper require that the Earth be flattened, or projected, so too do rasters (because of the distortions associated with flattening, the cells in a raster can never be perfectly equal in shape or area on the Earth's surface). Projections, or ways of flattening the Earth, are described in Section 5.7. Many of the terms that describe rasters suggest the laying of a tile floor on a flat surface – we talk of raster cells *tiling* an area, and a raster is said to be an instance of a *tesselation*, derived from the word for a mosaic. The mirrored ball hanging above a dance floor recalls the impossibility of covering a spherical object like the Earth perfectly with flat, square pieces.

When information is represented in raster form all detail about variation within cells is lost, and instead the cell is given a single value. Suppose we wanted to represent the map of the counties of Texas as a raster. Each cell would be given a single value to identify a county, and we would have to decide the rule to apply when a cell falls in more than one county. Often the rule is that the county with the *largest share* of the cell's area gets the cell. Sometimes the rule is based on the *central point* of the cell, and the county at that point is

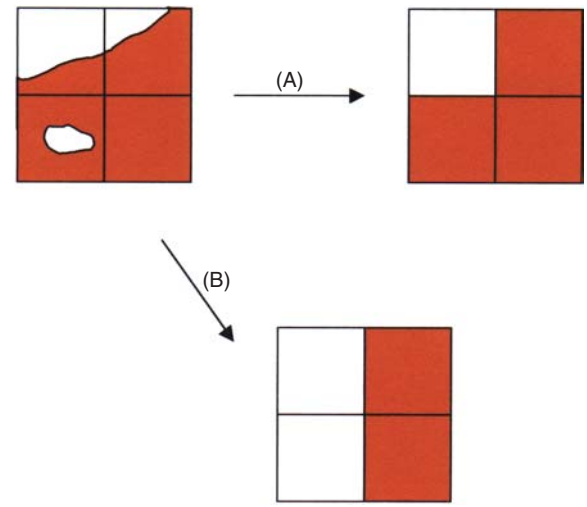


Figure 3.9 Effect of a raster representation using (A) the largest share rule and (B) the central point rule

assigned to the whole cell. Figure 3.9 shows these two rules in operation. The largest share rule is almost always preferred, but the central point rule is sometimes used in the interests of faster computing, and is often used in creating raster datasets of elevation.

3.6.2 Vector data

In a vector representation, all lines are captured as points connected by precisely straight lines (some GIS software allows points to be connected by curves rather than straight lines, but in most cases curves have to be approximated by increasing the density of points). An area is captured as a series of points or *vertices* connected by straight lines as shown in Figure 3.10. The straight edges between vertices explain why areas in vector representation are often called *polygons*, and in GIS-speak the terms polygon and area are often used

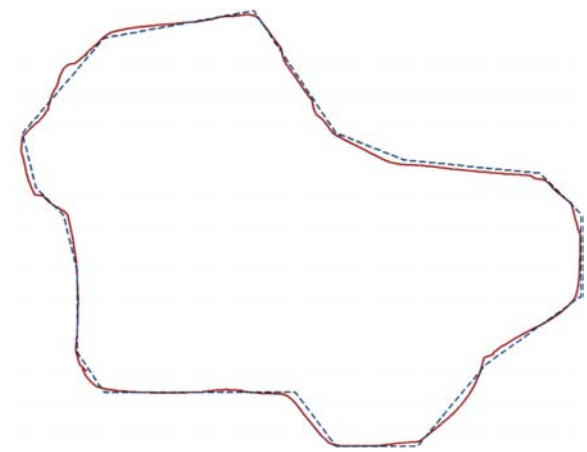


Figure 3.10 An area (red line) and its approximation by a polygon (blue line)

interchangeably. Lines are captured in the same way, and the term *polyline* has been coined to describe a curved line represented by a series of straight segments connecting vertices.

To capture an area object in vector form, we need only specify the locations of the points that form the vertices of a polygon. This seems simple, and also much more efficient than a raster representation, which would require us to list all of the cells that form the area. These ideas are captured succinctly in the comment 'Raster is vaster, and vector is correcter'. To create a precise approximation to an area in raster, it would be necessary to resort to using very small cells, and the number of cells would rise proportionately (in fact, every halving of the width and height of each cell would result in a quadrupling of the number of cells). But things are not quite as simple as they seem. The apparent precision of vector is often unreasonable, since many geographic phenomena simply cannot be located with high accuracy. So although raster data may look less attractive, they may be more honest to the inherent quality of the data. Also, various methods exist for compressing raster data that can greatly reduce the capacity needed to store a given dataset (see Chapter 8). So the choice between raster and vector is often complex, as summarized in Table 3.3.

3.6.3 Representing continuous fields

While discrete objects lend themselves naturally to representation as points, lines, or areas using vector methods, it is less obvious how the continuous variation of a field can be expressed in a digital representation. In GIS six alternatives are commonly implemented (Figure 3.11):

A. capturing the value of the variable at each of a grid of regularly spaced sample points (for example, elevations at 30 m spacing in a DEM);

B. capturing the value of the field variable at each of a set of irregularly spaced sample points (for example, variation in surface temperature captured at weather stations);

Table 3.3 Relative advantages of raster and vector representation

Issue	Raster	Vector
Volume of data	Depends on cell size	Depends on density of vertices
Sources of data	Remote sensing, imagery	Social and environmental data
Applications	Resources, environmental	Social, economic, administrative
Software	Raster GIS, image processing	Vector GIS, automated cartography
Resolution	Fixed	Variable

C. capturing a single value of the variable for a regularly shaped cell (for example, values of reflected radiation in a remotely sensed scene);

D. capturing a single value of the variable over an irregularly shaped area (for example, vegetation cover class or the name of a parcel's owner);

E. capturing the linear variation of the field variable over an irregularly shaped triangle (for example, elevation captured in a triangulated irregular network or TIN, Section 9.2.3.4);

F. capturing the isolines of a surface, as digitized lines (for example, digitized contour lines representing surface elevation).

Each of these methods succeeds in compressing the potentially infinite amount of data in a continuous field to a finite amount, using one of the six options, two of which (A and C) are raster, and four (B, D, E, and F) are vector. Of the vector methods one (B) uses points, two (D and E) use polygons, and one (F) uses lines to express the continuous spatial variation of the field in terms of a finite set of vector objects. But unlike the discrete object conceptualization, the objects used to represent a field are not real, but simply artifacts of the representation of something that is actually conceived as spatially continuous. The triangles of a TIN representation (E), for example, exist only in the digital representation, and cannot be found on the ground, and neither can the lines of a contour representation (F).

3.7 The paper map

The paper map has long been a powerful and effective means of communicating geographic information. In contrast to digital data, which use coding schemes such as ASCII, it is an instance of an *analog* representation, or a physical model in which the real world is scaled – in the case of the paper map, part of the world is scaled to fit the size of the paper. A key property of a paper map is its *scale* or *representative fraction*, defined as the ratio of distance on the map to distance on the Earth's surface. For example, a map with a scale of 1:24 000 reduces everything on the Earth to one 24 000th of its real size. This is a bit misleading, because the Earth's surface is curved and a paper map is flat, so scale cannot be exactly constant.

A paper map is: a source of data for geographic databases; an analog product from a GIS; and an effective communication tool.

Maps have been so important, particularly prior to the development of digital technology, that many of the ideas associated with GIS are actually inherited directly from paper maps. For example, scale is often cited as a property of a digital database, even though the definition of scale makes no sense for digital data – ratio of distance *in the*

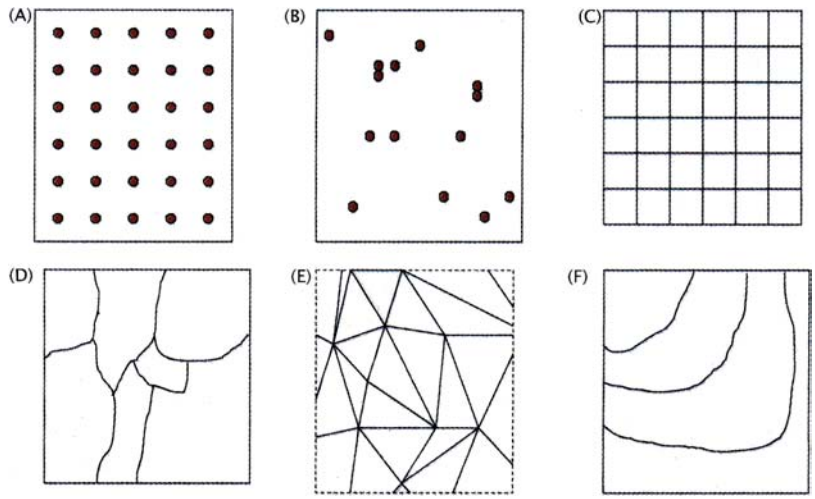


Figure 3.11 The six approximate representations of a field used in GIS. (A) Regularly spaced sample points. (B) Irregularly spaced sample points. (C) Rectangular cells. (D) Irregularly shaped polygons. (E) Irregular network of triangles, with linear variation over each triangle (the Triangulated Irregular Network or TIN model; the bounding box is shown dashed in this case because the unshown portions of complete triangles extend outside it). (F) Polylines representing contours (see the discussion of isopleth maps in Box 4.3) (Courtesy US Geological Survey)

Biographical Box **3.5**

May Yuan and new forms of representation

Figure 3.13 May Yuan, developer of new forms of representation

May Yuan received her Bachelor of Science degree in Geography from the National Taiwan University, where she was attracted to the fields of geomorphology and climatology. Continuing her fundamental interest in evolution of processes, she studied geographic representation and temporal GIS and earned both her Masters and PhD degrees in Geography from the State University of New York at Buffalo.

Currently, May is an Associate Professor of Geography at the University of Oklahoma. Severe weather in the Southern Plains of the United States (Figure 3.12) has inspired her to re-evaluate GIS representation of geographic dynamics, the complexity of events and processes at spatial and temporal scales, and GIS applications in meteorology (i.e., weather and climate). She investigates meteorological cases (e.g., convective storms and flash floods) to develop new ideas of using events and processes as the basis to integrate spatial and temporal data in GIS. Her publications address theoretical issues on representation of geographic dynamics and offer conceptual models and a prototype GIS to support spatiotemporal queries and analysis of dynamic geographic phenomena. Her temporal GIS research goes beyond merely considering time as an attribute or annotation of spatial objects to incorporate much richer spatiotemporal meaning. In her case study on convective storms, she has demonstrated that, by modeling storms as data objects, GIS is able to support information query about storm evolution, storm behaviors, and interactions with environments.

May developed a strong interest in physics in early childhood. Newton's theory of universal gravitation sparked her appreciation for simple principles that can explain how things work and for the use of graphical and symbolic representation to conceptualize complex processes. Planck's quantum theory and Heisenberg's uncertainty principle further stimulated her thinking on the nature of matter and its behavior at different scales of observations. Shaped by Einstein's theory of relativity, May developed her world view as a four-dimensional space-time continuum populated with events and phenomena. Before she pursued a career in GIScience, May studied fluvial processes and developed a model to classify waterfalls and explain

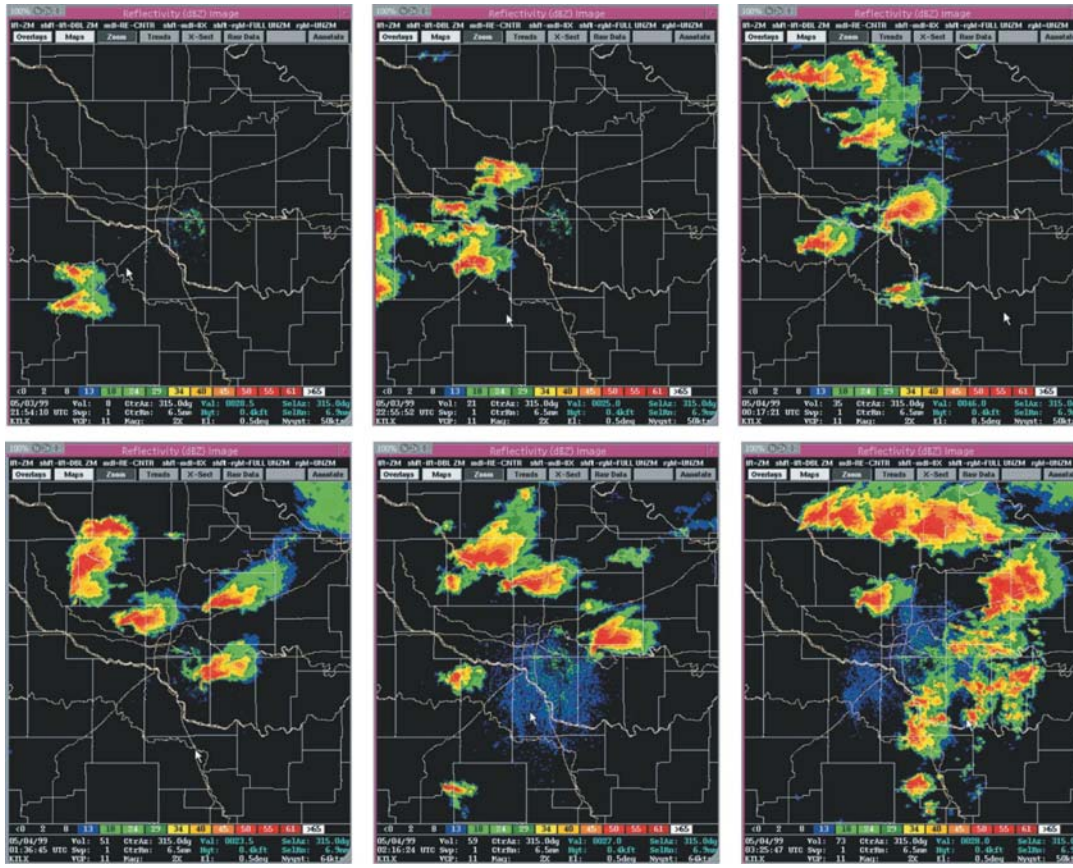


Figure 3.12 Representative radar images showing the evolution of supercell storms that produced F5 tornadoes in Oklahoma City, May 3, 1999. WSR-88D radar TKLX scanned the supercells every five minutes, but the images shown here were selected approximately every two hours

their formation. She went on to study paleoclimatology by analyzing soil and speleothem sediments. Both studies, as well as her dissertation research on wildfire representation, reinforced her interest in developing conceptual models of processes and examining the relationships between space and time. Since she moved to the University of Oklahoma, a suite of world-class meteorological research initiatives has offered her unique opportunities to extend her interest in physics to fundamental research in GIScience through meteorological applications. Weather and climate offer rich cases that emphasize movement, processes, and evolution and pose grand challenges to GIScience research regarding representation, object-field duality, and uncertainty. May enjoys the challenges that ultimately connect to her fundamental interest in how things work.

computer to distance on the ground; how can there be distances in a computer? What is meant is a little more complicated: when a scale is quoted for a digital database it is usually the scale of the map that formed the source of the data. So if a database is said to be at a scale of 1:24 000 one can safely assume that it was created from a paper map at that scale, and includes representations of the features that are found on maps at that scale. Further discussion of scale can be found in Box 4.2 and in Chapter 6, where it is important to the concept of uncertainty.

There is a close relationship between the contents of a map and the raster and vector representations discussed in the previous section. The US Geological Survey, for example, distributes two digital versions of its topographic maps, one in raster form and one in vector form, and both attempt to capture the contents of the map as closely as possible. In the raster form, or

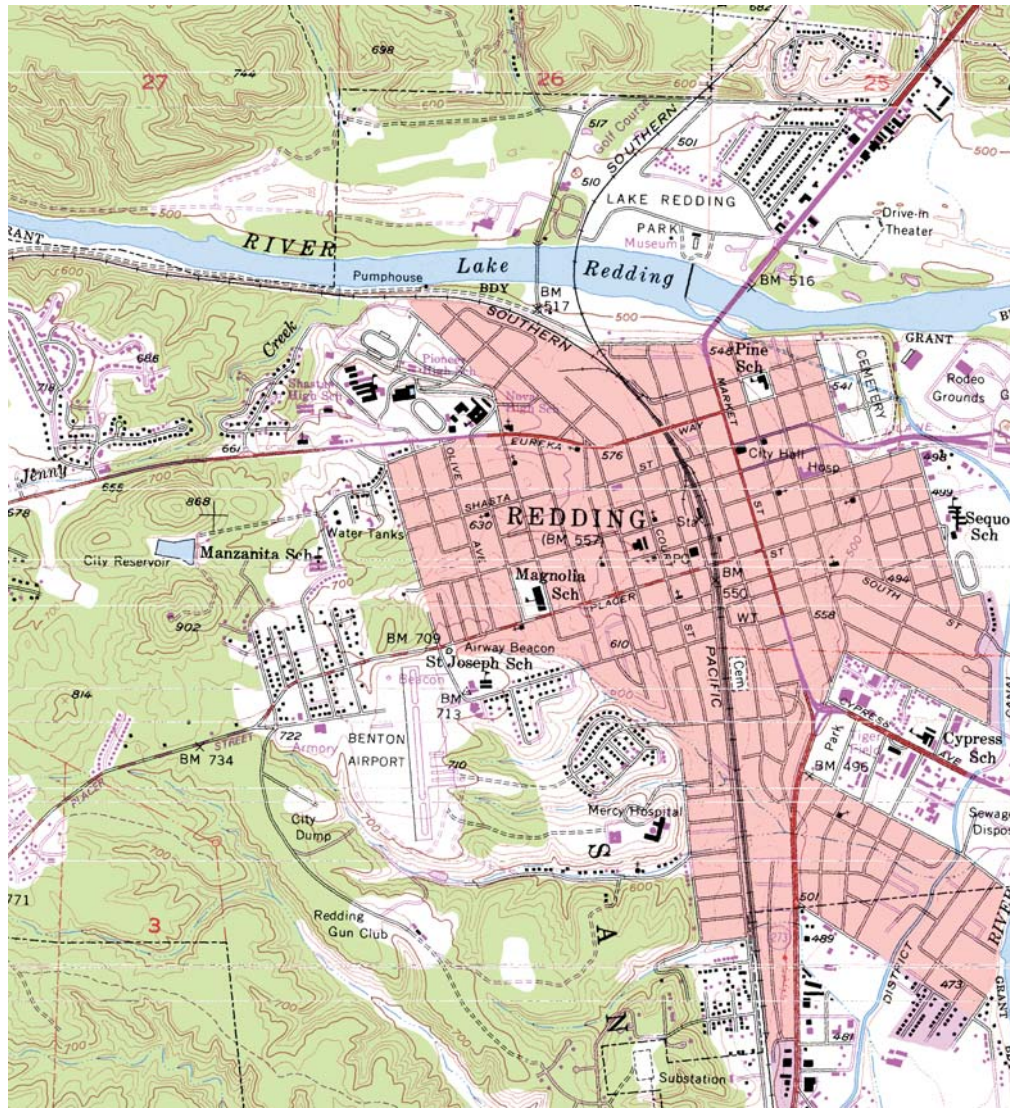


Figure 3.14 Part of a Digital Raster Graphic, a scan a US Geological Survey 1:24 000 topographic map

digital raster graphic (DRG), the map is scanned at a very high density, using very small pixels, so that the raster looks very much like the original (Figure 3.14). The coding of each pixel simply records the color of the map picked up by the scanner, and the dataset includes all of the textual information surrounding the actual map.

In the vector form, or *digital line graph* (DLG), every geographic feature shown on the map is represented as a point, polyline, or polygon. The symbols used to represent point features on the map, such as the symbol for a windmill, are replaced in the digital data by points with associated attributes, and must be regenerated when the data are displayed. Contours, which are shown on the map as lines of definite width, are replaced by polylines of no width, and given attributes that record their elevations.

In both cases, and especially in the vector case, there is a significant difference between the analog representation of the map and its digital equivalent. So it is quite misleading to think of the contents of a digital representation as a map, and to think of a GIS as a container of digital maps. Digital representations can include information that would be very difficult to show on maps. For example, they can represent the curved surface of the Earth, without the need for the distortions associated with flattening. They can represent changes, whereas maps must be static because it is very difficult to change their contents once they have been printed or drawn. Digital databases can represent all three spatial dimensions, including the vertical, whereas maps must always show two-dimensional views. So while the paper map is a useful metaphor for the contents of a geographic database, we must be careful not to let it limit our thinking about what is possible in the way of representation. This issue is pursued at greater length in Chapter 8, and map production is discussed in detail in Chapter 12.

3.8 Generalization

In Section 3.4 we saw how thinking about geographic information as a collection of atomic links – between a place, a time (not always, because many geographic facts are stated as if they were permanently true), and a property – led to an immediate problem, because the potential number of such atomic facts is infinite. If seen in enough detail, the Earth's surface is unimaginably complex, and its effective description impossible. So instead, humans have devised numerous ways of simplifying their view of the world. Instead of making statements about each and every point, we describe entire areas, attributing uniform characteristics to them, even when areas are not strictly uniform; we identify features on the ground and describe their characteristics, again assuming them to be uniform; or we limit our descriptions to what exists at a finite number of sample points, hoping that these samples will be adequately representative of the whole (Section 4.4).

A geographic database cannot contain a perfect description – instead, its contents must be carefully selected to fit within the limited capacity of computer storage devices.

From this perspective some degree of generalization is almost inevitable in all geographic data. But cartographers often take a somewhat different approach, for which this observation is not necessarily true. Suppose we are tasked to prepare a map at a specific scale, say 1:25 000, using the standards laid down by a national mapping agency, such as the Institut Géographique National (IGN) of France. Every scale used by IGN has its associated rules of representation. For example, at a scale of 1:25 000 the rules lay down that individual buildings will be shown only in specific circumstances, and similar rules apply to the 1:24 000 series of the US Geological Survey. These rules are known by various names, including *terrain nominal* in the case of IGN, which translates roughly but not very helpfully to 'nominal ground', and is perhaps better translated as 'specification'. From this perspective a map that represents the world by following the rules of a specification precisely can be perfectly accurate *with respect to the specification*, even though it is not a perfect representation of the full detail on the ground.

A map's specification defines how real features on the ground are selected for inclusion on the map.

Consider the representation of vegetation cover using the rules of a specification. For example, the rules might state that at a scale of 1:100 000, a vegetation cover map should not show areas of vegetation that cover less than 1 hectare. But small areas of vegetation almost certainly exist, so deleting them inevitably results in information loss. But under the principle discussed above, a map that adheres to this rule must be accurate, *even though it differs substantively from the truth as observed on the ground.*

3.8.1 Methods of generalization

A GIS dataset's level of detail is one of its most important properties, as it determines both the degree to which the dataset approximates the real world, and the dataset's complexity. It is often necessary to remove detail, in the interests of compressing data, fitting them into a storage device of limited capacity, processing them faster, or creating less confusing visualizations that emphasize general trends. Consequently many methods have been devised for generalization, and several of the more important are discussed in this section.

McMaster and Shea (1992) identify the following types of generalization rules:

- *simplification*, for example by weeding out points in the outline of a polygon to create a simpler shape;
- *smoothing*, or the replacement of sharp and complex forms by smoother ones;
- *aggregation*, or the replacement of a large number of distinct symbolized objects by a smaller number of new symbols;
- *amalgamation*, or the replacement of several area objects by a single area object;
- *merging*, or the replacement of several line objects by a smaller number of line objects;
- *collapse*, or the replacement of an area object by a combination of point and line objects;
- *refinement*, or the replacement of a complex pattern of objects by a selection that preserves the pattern's general form;
- *exaggeration*, or the relative enlargement of an object to preserve its characteristics when these would be lost if the object were shown to scale;
- *enhancement*, through the alteration of the physical sizes and shapes of symbols; and
- *displacement*, or the moving of objects from their true positions to preserve their visibility and distinctiveness.

The differences between these types of rules are much easier to understand visually and Figure 3.15 reproduces McMaster's and Shea's original example drawings. In addition, they describe two forms of generalization of attributes, as distinct from geometric forms of generalization. *Classification* generalization reclassifies the attributes of objects into a smaller number of classes, while *symbolization* generalization changes the assignment of symbols to objects. For example, it might replace an elaborate symbol including the words 'Mixed Forest' with a color identifying that class.

3.8.2 Weeding

One of the commonest forms of generalization in GIS is the process known as weeding, or the simplification of the representation of a line represented as a polyline. The process is an instance of McMaster and Shea's simplification. Standard methods exist in GIS for doing

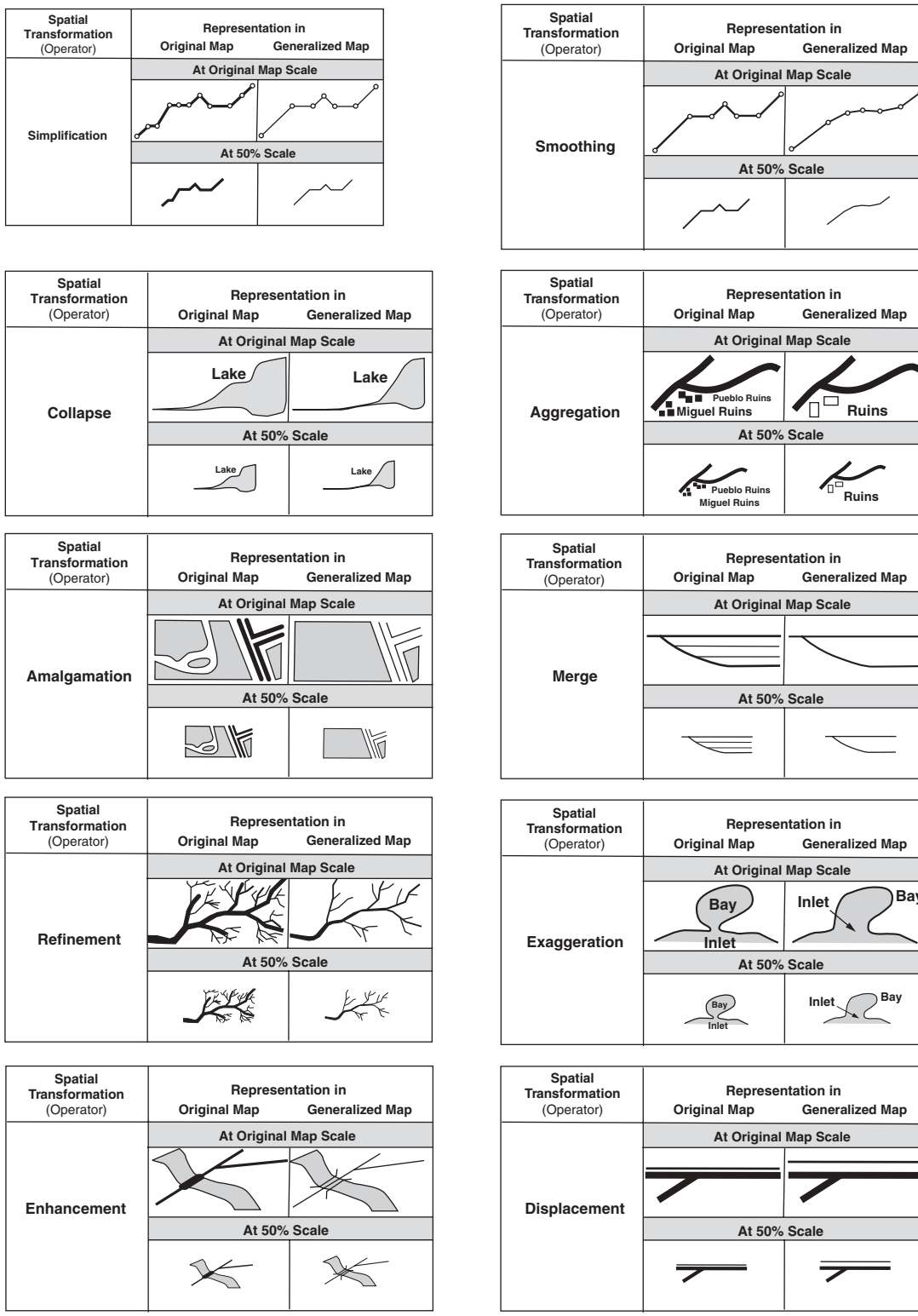


Figure 3.15 Illustrations from McMaster and Shea (1992) of their ten forms of generalization. The original feature is shown at its original level of detail, and below it at 50% coarser scale. Each generalization technique resolves a specific problem of display at coarser scale and results in the acceptable version shown in the lower right

Figure 3.16 The Douglas–Poiker algorithm is designed to simplify complex objects like this shoreline by reducing the number of points in its polyline representation

this, and the commonest by far is the method known as the Douglas–Poiker algorithm (Figure 3.16) after its inventors, David Douglas and Tom Poiker. The operation of the Douglas–Poiker weeding algorithm is shown in Figure 3.17.

Weeding is the process of simplifying a line or area by reducing the number of points in its representation.

Note that the algorithm relies entirely on the assumption that the line is represented as a polyline, in other words as a series of straight line segments. GIS increasingly support other representations, including arcs of circles, arcs of ellipses, and Bézier curves, but there is little consensus to date on appropriate methods for weeding or generalizing them, or on methods of analysis that can be applied to them.

3.9 Conclusion

Representation, or more broadly ontology, is a fundamental issue in GIS, since it underlies all of our efforts to express useful information about the surface of the Earth in a digital computer. The fact that there are so many ways of doing this makes GIS at once complex and interesting,

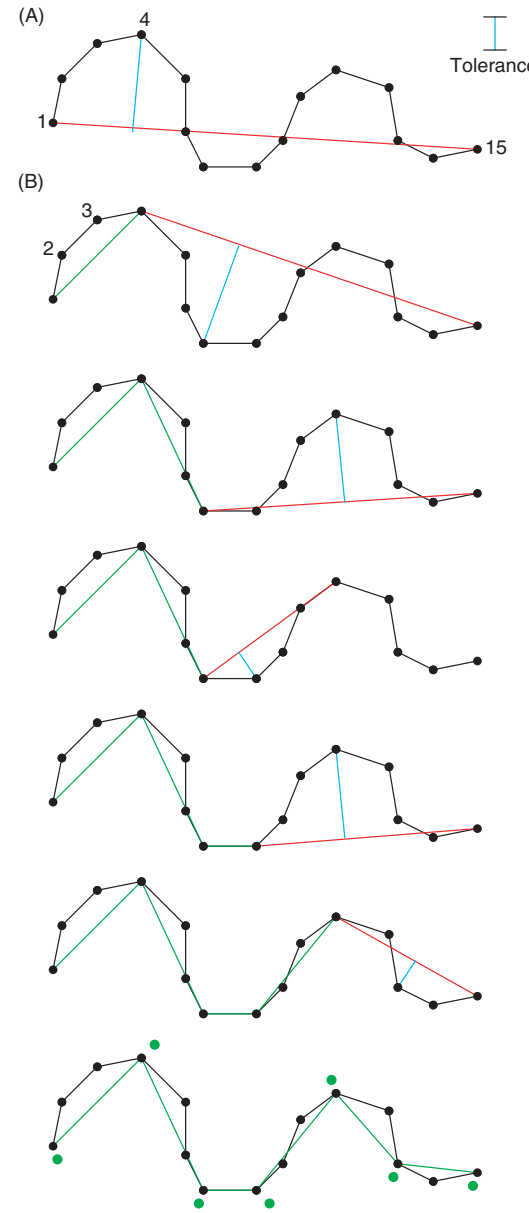


Figure 3.17 The Douglas–Poiker line simplification algorithm in action. The original polyline has 15 points. In (A) Points 1 and 15 are connected (red), and the furthest distance of any point from this connection is identified (blue). This distance to Point 4 exceeds the user-defined tolerance. In (B) Points 1 and 4 are connected (green). Points 2 and 3 are within the tolerance of this line. Points 4 and 15 are connected, and the process is repeated. In the final step 7 points remain (identified with green disks), including 1 and 15. No points are beyond the user-defined tolerance distance from the line

a point that will become much clearer on reading the technical chapter on data modeling, Chapter 8. But the broader issues of representation, including the distinction between field and object conceptualizations, underlie not only that chapter but many other issues as well, including uncertainty (Chapter 6), and Chapters 14 through 16 on analysis and modeling.

Questions for further study

1. What fraction of the Earth's surface have you experienced in your lifetime? Make diagrams like that shown in Figure 3.1, at appropriate levels of detail, to show a) where you have lived in your lifetime, b) how you spent last weekend. How would you describe what is missing from each of these diagrams?
2. Table 3.3 summarized some of the arguments between raster and vector representations. Expand on these arguments, providing examples, and add any others that would be relevant in a GIS application.
3. The early explorers had limited ways of communicating what they saw, but many were very effective at it. Examine the published diaries, notebooks, or dispatches of one or two early explorers and look at the methods they used to communicate with others. What words did they use to describe unfamiliar landscapes and how did they mix words with sketches?
4. Identify the limits of your own neighborhood, and start making a list of the discrete objects you are familiar with in the area. What features are hard to think of as discrete objects? For example, how will you divide up the various roadways in the neighborhood into discrete objects – where do they begin and end?

Further reading

Chrisman N.R. 2002 *Exploring Geographic Information Systems* (2nd edn). New York: Wiley.

McMaster R.B. and Shea K.S. 1992 *Generalization in Digital Cartography*. Washington, DC: Association of American Geographers.

National Research Council 1999 *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academy Press. Available: **www.nap.edu**.

4 *The nature of geographic data*

This chapter elaborates on the *spatial is special* theme by examining the nature of geographic data. It sets out the distinguishing characteristics of geographic data, and suggests a range of guiding principles for working with them. Many geographic data are correctly thought of as sample observations, selected from the larger universe of possible observations that could be made. This chapter describes the main principles that govern scientific sampling, and the principles that are invoked in order to infer information about the gaps between samples. When devising spatial sample designs, it is important to be aware of the nature of spatial variation, and here we learn how this is formalized and measured as spatial autocorrelation. Another key property of geographic information is the level of detail that is apparent at particular scales of analysis. The concept of fractals provides a solid theoretical foundation for understanding scale when building geographic representations.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will understand:

- How Tobler's First Law of Geography is formalized through the concept of spatial autocorrelation;
- The relationship between scale and the level of geographic detail in a representation;
- The principles of building representations around geographic samples;
- How the properties of smoothness and continuous variation can be used to characterize geographic variation;
- How fractals can be used to measure and simulate surface roughness.

4.1 Introduction

In Chapter 1 we identified the central motivation for scientific applications of GIS as the development of representations, not only of how the world *looks*, but also how it *works*. Chapter 3 established three governing principles that help us towards this goal, namely that:

1. the representations we build in GIS are of *unique places*;
2. our representations of them are necessarily *selective* of reality, and hence incomplete;
3. in building representations, it is useful to think of the world as either comprising continuously varying *fields* or as an empty space littered with *objects* that are crisp and well-defined.

In this chapter we build on these principles to develop a fuller understanding of the ways in which the *nature of spatial variation* is represented in GIS. We do this by asserting three further principles:

4. that proximity effects are key to understanding spatial variation, and to joining up incomplete representations of unique places;
5. that issues of geographic scale and level of detail are key to building appropriate representations of the world;
6. that different measures of the world *co-vary*, and understanding the nature of co-variation can help us to predict.

Implicit in all of this is one further principle, that we will develop in Chapter 6:

7. because almost all representations of the world are necessarily incomplete, they are *uncertain*.

GIS is about representing spatial and temporal phenomena in the real world and, because the real world is complicated, this task is difficult and error prone. The real world provides an intriguing laboratory in which to examine phenomena, but is one in which it can be impossible to control for variation in all characteristics – be they relevant to landscape evolution, consumer behavior, urban growth, or whatever. In the terminology of Section 1.3, generalized *laws* governing spatial distributions and temporal dynamics are therefore most unlikely to work perfectly. We choose to describe the seven points above as 'principles', rather than 'laws' (see Section 1.3) because, like our discussion in Chapter 2, this chapter is grounded in empirical generalization about the real world. A more elevated discussion of the way that these principles build into 'fundamental laws of GIScience' has been published by Goodchild.

4.2 The fundamental problem revisited

Consider for a moment a GIS-based representation of your own life history to date. It is infinitesimally small compared with the geographic extent and history of the world but, as we move to finer spatial and temporal scales than those shown in Figure 3.1, nevertheless very intricate in detail. Viewed in aggregate, human behavior where you live exhibits structure in geographic space, as the aggregated outcomes of day-to-day (often repetitive) decisions about where to go, what to do, how much time to spend doing it, and longer-term (one-off) decisions about where to live, how to achieve career objectives, and how to balance work, leisure, and family pursuits. It is helpful to distinguish between *controlled* and *uncontrolled* variation – the former oscillates around a steady state (daily, weekly) pattern, while the latter (career changes, residential moves) does not.

When relating our own daily regimes and life histories, or indeed any short or long term *time series* of events, we are usually mindful of the contexts in which our decisions (to go to work, to change jobs, to marry) are made – 'the past is the key to the present' aptly summarizes the effect of temporal context upon our actions. The day-to-day operational context to our activities is very much determined by where we live and work. The longer-term strategic context may well be provided by where we were born, grew up, and went to college.

Our behavior in geographic space often reflects past patterns of behavior.

The relationship between consecutive events in *time* can be formalized in the concept of *temporal*

autocorrelation. The analysis of time series data is in some senses straightforward, since the direction of causality is only one way – past events are sequentially related to the present and to the future. This chapter (and book) is principally concerned with spatial, rather than temporal, autocorrelation. Spatial autocorrelation shares some similarities with its temporal counterpart. Yet time moves in one direction only (forward), making temporal autocorrelation one-dimensional, while spatial events can potentially have consequences anywhere in two-dimensional or even three-dimensional space.

Explanation in time need only look to the past, but explanation in space must look in all directions simultaneously.

Assessment of spatial autocorrelation can be informed by knowledge of the degree and nature of *spatial heterogeneity* – the tendency of geographic places and regions to be different from each other. Everyone would recognize the extreme difference of landscapes between such regions as the Antarctic, the Nile delta, the Sahara desert, or the Amazon basin, and many would recognize the more subtle differences between the Central Valley of California, the Northern Plain of China, and the valley of the Ganges in India. Heterogeneity occurs both in the way the landscape looks, and in the way processes act on the landscape (the form/process distinction of Section 1.3). While the spatial variation in some processes simply oscillates about an average (controlled variation), other processes vary ever more the longer they are observed (uncontrolled variation). For example, controlled variation characterizes the operational environment of GIS applications in utility management (Section 2.1.1), or the tactical environment of retail promotions (Section 2.3.3), while longer-term processes such as global warming or deforestation may exhibit uncontrolled variation. As a general rule, spatial data exhibit an increasing range of values, hence increased heterogeneity, with increased distance. In this chapter we focus on the ways in which phenomena vary across space, and the general nature of geographic variation: later, in Chapter 14, we return to the techniques for measuring spatial heterogeneity.

Also, this requires us to move beyond thinking of GIS data as abstracted only from the continuous spatial distributions implied by the Tobler Law (Section 3.1) and from sequences of events over continuous time. Some events, such as the daily rhythm of the journey to work, are clearly incremental extensions of past practice, while others, such as residential relocation, constitute sudden breaks with the past. Similarly, landscapes of gently undulating terrain are best thought of as smooth and continuous, while others (such as the landscapes developed about fault systems, or mountain ranges) are best conceived as discretely bounded, jagged, and irregular. Smoothness and irregularity turn out to be among the most important distinguishing characteristics of geographic data.

Some geographic phenomena vary smoothly across space, while others can exhibit extreme irregularity, in violation of Tobler's Law.

Finally, it is highly likely that a representation of the real world that is suitable for predicting future change will need to incorporate information on how two or more factors *co-vary*. For example, planners seeking to justify improvements to a city's public transit system might wish to point out how house prices increase with proximity to existing rail stops. It is highly likely that patterns of spatial autocorrelation in one variable will, to a greater or lesser extent, be mirrored in another. Whilst this is helpful in building representations of the real world, the property of spatial autocorrelation can frustrate our attempts to build inferential statistical models of the co-variation of geographic phenomena.

Spatial autocorrelation helps us to build representations, but frustrates our efforts to predict.

The nature of geographic variation, the scale at which uncontrolled variation occurs, and the way in which different geographic phenomena co-vary are all key to building effective representations of the real world. These principles are of practical importance and guide us to answering questions such as: What is an appropriate scale or level of detail at which to build a representation for a particular application? How do I design my spatial sample? How do I generalize from my sample measurements? And what formal methods and techniques can I use to relate key spatial events and outcomes to one another?

Each of these questions is a facet of the fundamental problem of GIS, that is of selecting what to leave in and what to take out of our digital representations of the real world (Section 3.2). The Tobler Law (Section 3.1), that everything is related to everything else, but near things are more related than distant things, amounts to a succinct definition of spatial autocorrelation. An understanding of the nature of the spatial autocorrelation that characterizes a GIS application helps us to *deduce* how best to collect and assemble data for a representation, and also how best to develop inferences between events and occurrences.

The concept of geographic *scale* or level of detail will be fundamental to observed measures of the likely strength and nature of autocorrelation in any given application. Together, the scale and spatial structure of a particular application suggest ways in which we should *sample* geographic reality, and the ways in which we should *weight* sample observations in order to build our representation. We will return to the key concepts of scale, sampling, and weighting throughout much of this book.

4.3 Spatial autocorrelation and scale

In Chapter 3 (Box 3.3) we classified attribute data into the nominal, ordinal, interval, ratio, and cyclic scales of measurement. Objects existing in space are described by locational (spatial) descriptors, and are conventionally classified using the taxonomy shown in Box 4.1.

Technical Box **4.1**

Types of spatial objects

We saw in Section 3.4 that geographic objects are classified according to their *topological dimension*, which provides a measure of the way they fill space. For present purposes we assume that dimensions are restricted to *integer* (whole number) values, though in later sections (Sections 4.8 and 15.2.5) we relax this constraint and consider geographic objects of non-integer (fractional, or *fractal*) dimension. All geometric objects can be used to represent occurrences at absolute locations (*natural* objects), or they may be used to summarize spatial distributions (*artificial* objects).

A *point* has neither length nor breadth nor depth, and hence is said to be of dimension 0. Points may be used to indicate spatial occurrences or events, and their spatial patterning. *Point pattern analysis* is used to identify whether occurrences or events are interrelated – as in the analysis of the incidence of crime, or in identifying whether patterns of disease infection might be related to environmental or social factors (Section 15.2.3). The *centroid* of an area object is an artificial point reference, which is located so as to provide a summary measure of the location of the object (Section 15.2.1).

Lines have length, but not breadth or depth, and hence are of dimension 1. They are used to represent linear entities such as roads, pipelines, and cables, which frequently build together into networks. They can also be used to measure distances between spatial objects, as in the measurement of inter-centroid distance. In order to reduce the burden of data capture and storage, lines are often held in GIS in *generalized* form (see Section 3.8).

Area objects have the two dimensions of length and breadth, but not depth. They may be used to represent natural objects, such as agricultural fields, but are also commonly used to represent artificial aggregations, such as census tracts (see below). Areas may bound linear features and enclose points, and GIS functions can be used to identify whether a given area encloses a given point (Section 14.4.2).

Volume objects have length, breadth, and depth, and hence are of dimension 3. They are used to represent natural objects such as river basins, or artificial phenomena such as the population potential of shopping centers or the density of resident populations (Section 14.4.5).

Time is often considered to be the fourth dimension of spatial objects, although GIS remains poorly adapted to the modeling of temporal change.

The relationship between higher- and lower-dimension spatial objects is analogous to that between higher- and lower-order attribute data, in that lower-dimension objects can be derived from those of higher dimension but not vice versa. Certain phenomena, such as population, may be held as natural or artificially imposed spatial object types. The chosen way of representing phenomena in GIS not only defines the apparent nature of geographic variation, but also the way in which geographic variation may be analyzed. Some objects, such as agricultural fields or digital terrain models, are represented in their natural state. Others are transformed from one spatial object class to another, as in the transformation of population data from individual points to census tract areas, for reasons of confidentiality or convention. Some high-order representations are created by interpolation between lower-order objects, as in the creation of digital terrain models (DTMs) from spot height data (Chapter 8).

The classification of spatial phenomena into object types is dependent fundamentally upon scale. For example, on a less-detailed map of the world, New York is represented as a one-dimensional point. On a more-detailed map such as a road atlas it will be represented as a two-dimensional area. Yet if we visit the city, it is very much experienced as a three-dimensional entity, and virtual reality systems seek to represent it as such (see Section 13.4.2).

Spatial autocorrelation measures attempt to deal simultaneously with similarities in the location of spatial objects (Box 4.1) and their attributes (Box 3.3). If features that are similar in location are also similar in attributes, then the pattern as a whole is said to exhibit *positive spatial autocorrelation*. Conversely, *negative spatial autocorrelation* is said to exist when features which are close together in space tend to be more dissimilar in attributes than features which are further apart (in opposition to Tobler's Law). Zero autocorrelation occurs when attributes are independent of location. Figure 4.1 presents some simple field representations of a geographic variable in 64 cells that can each take one of two values, coded blue and white. Each of the five illustrations

contains the same set of attributes, 32 white cells and 32 blue cells, yet the spatial arrangements are very different. Figure 4.1A presents the familiar chess board, and illustrates extreme negative spatial autocorrelation between neighboring cells. Figure 4.1E presents the opposite extreme of positive autocorrelation, when blue and white cells cluster together in homogeneous regions. The other illustrations show arrangements which exhibit intermediate levels of autocorrelation. Figure 4.1C corresponds to spatial independence, or no autocorrelation, Figure 4.1B shows a relatively dispersed arrangement, and Figure 4.1D a relatively clustered one.

Spatial autocorrelation is determined both by similarities in position, and by similarities in attributes.

The patterns shown in Figure 4.1 are examples of a particular case of spatial autocorrelation. In terms of the classification developed in Chapter 3 (Box 3.3) the attribute data are *nominal* (blue and white simply identify two different possibilities, with no implied order and no possibility of difference, or ratio) and their spatial distribution is conceived as a field, with a single value everywhere. The figure gives no clue as to the true dimensions of the area being represented. Usually, similarities in attribute values may be more precisely measured on higher-order measurement scales, enabling continuous measures of spatial variation (See Section 6.3.2.2 for a discussion of precision). As we see below, the way in which we define what we mean by *neighboring* in investigating spatial arrangements may be more or less sophisticated. In considering the various arrangements shown in Figure 4.1, we have only considered the relationship between the attributes of a cell and those of its four *immediate* neighbors. But we could include a cell's four diagonal neighbors in the comparison, and more generally there is no reason why we should not interpret Tobler's Law in terms of a gradual incremental attenuating effect of distance as we traverse successive cells.

We began this chapter by considering a time series analysis of events that are highly, even perfectly, repetitive in the short term. Activity patterns often exhibit strong positive temporal autocorrelation (where you were at this time last week, or this time yesterday is likely to affect where you are now), but only if measures are made at the same time every day – that is, at the temporal scale of the daily interval. If, say, sample measurements were taken every 17 hours, measures of the temporal autocorrelation of your activity patterns would likely be much lower. Similarly, if the measures of the blue/white property were made at intervals that did not coincide with the dimensions of the squares of the chess boards in Figure 4.1, then the spatial autocorrelation measures would be different. Thus the issue of *sampling interval* is of direct importance in the measurement of spatial autocorrelation, because spatial events and occurrences may or may not accommodate spatial structure. In general, measures of spatial and temporal autocorrelation are *scale dependent* (see Box 4.2). Scale is often integral to the trade off between the level of spatial resolution and the

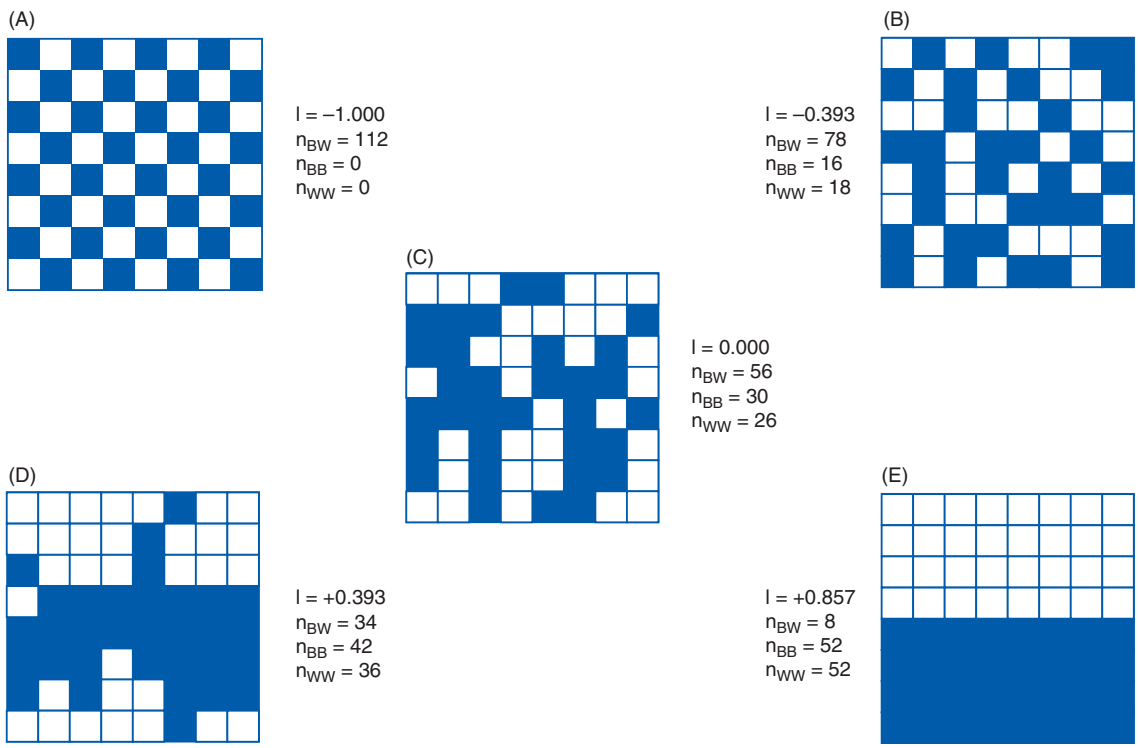


Figure 4.1 Field arrangements of blue and white cells exhibiting: (A) extreme negative spatial autocorrelation; (B) a dispersed arrangement; (C) spatial independence; (D) spatial clustering; and (E) extreme positive spatial autocorrelation. The values of the *I* statistic are calculated using the equation in Section 4.6 (*Source*: Goodchild 1986 CATMOG, GeoBooks, Norwich)

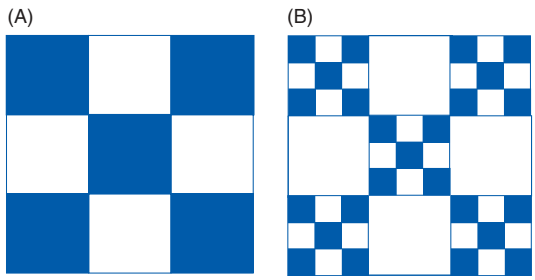


Figure 4.2 A Sierpinski carpet at two levels of resolution: (A) coarse scale and (B) finer scale

degree of attribute detail that can be stored in a given application – as in the trade off between spatial and spectral resolution in remote sensing.

Quattrochi and Goodchild have undertaken an extensive discussion of these and other meanings of scale (e.g., the degree of spectral or temporal coarseness), and their implications.

A further important property is that of *self-similarity*. This is illustrated using a mosaic of squares in Figure 4.2. Figure 4.2A presents a coarse-scale representation of attributes in nine squares, and a pattern of negative spatial autocorrelation. However, the pattern is self-replicating at finer scales, and in Figure 4.4B, a finer-scale representation reveals that the smallest blue cells replicate the pattern of the whole area in a recursive manner. The pattern of spatial autocorrelation at the coarser scale is replicated at the finer scale, and the overall pattern is said to exhibit the property of *self-similarity*. Self-similar structure is characteristic of natural as well as social systems: for example, a rock may resemble the physical form of the mountain from which it was broken (Figure 4.3), small coastal features may resemble larger

Figure 4.3 Individual rocks may resemble larger-scale structures, such as the mountains from which they are broken, in form

bays and inlets in structure and form, and neighborhoods may be of similar population size and each offer similar ranges of retail facilities right across a metropolitan area. Self-similarity is a core concept of fractals, a topic introduced in Section 4.8.

4.4 Spatial sampling

The quest to represent the myriad complexity of the real world requires us to abstract, or sample, events and occurrences from a *sample frame*, defined as the universe of eligible elements of interest. Thus the process of sampling elements from a sample frame can very much determine the apparent nature of geographic data. A spatial sampling frame might be bounded by the extent

Technical Box **4.2**

The many meanings of *scale*

Unfortunately the word *scale* has acquired too many meanings in the course of time. Because they are to some extent contradictory, it is best to use other terms that have clearer meaning where appropriate.

Scale is in the details. Many scientists use scale in the sense of spatial resolution, or the level of spatial detail in data. Data are fine-scaled if they include records of small objects, and coarse-scaled if they do not.

Scale is about extent. Scale is also used by scientists to talk about the geographic extent or scope of a project: a large-scale project covers a large area, and a small-scale project covers a small area. Scale can also refer to other aspects of the project's scope, including the cost, or the number of people involved.

The scale of a map. Geographic data are often obtained from maps, and often displayed in map form. Cartographers use the term scale to refer to a map's *representative fraction* (the ratio of distance on the map to distance on the ground – see Section 3.7). Unfortunately this leads to confusion (and often bemusement) over the meaning of *large* and *small* with respect to scale. To a cartographer a large scale corresponds to a large representative fraction, in other words to plenty of geographic detail. This is exactly the opposite of what an average scientist understands by a large-scale study. In this book we have tried to avoid this problem by using *coarse* and *fine* instead.

of a field of interest, or by the combined extent of a set of areal objects. We can think of sampling as the process of selecting points from a continuous field or, if the field has been represented as a mosaic of areal objects, of selecting some of these objects while discarding others. Scientific sampling requires that each element in the sample frame has a known and prespecified chance of selection.

In some important senses, we can think of any geographic representation as a kind of sample, in that the elements of reality that are retained are abstracted from the real world in accordance with some overall design. This is the case in remote sensing, for example (see Section 3.6.1), in which each pixel value is a spatially averaged reflectance value calculated at the spatial resolution characteristic of the sensor. In many situations, we will need consciously to select some observations, and not others, in order to create a generalizable abstraction. This is because, as a general rule, the resources available to any given project do not stretch to measuring every single one of the elements (soil profiles, migrating animals, shoppers) that we know to make up our population of interest. And even if resources were available, science tells us that this would be wasteful, since procedures of *statistical inference* allow us to infer from samples to the populations from which they were drawn. We will return to the process of statistical inference in Sections 4.7 and 15.4. Here, we will confine ourselves to the question, how do we ensure a good sample?

Geographic data are only as good as the sampling scheme used to create them.

Classical statistics often emphasizes the importance of randomness in sound sample design. The purest form, simple random sampling, is well known: each element in the sample frame is assigned a unique number, and a prespecified number of elements are selected using a random number generator. In the case of a spatial sample from continuous space, x, y coordinate pairs might be randomly sampled within the range of x and y values (see Section 5.7 for information on coordinate systems). Since each randomly selected element has a known and prespecified probability of selection, it is possible to make robust and defensible generalizations to the population from which the sample was drawn. A spatially random sample is shown in Figure 4.4A. Random sampling is integral to probability theory, and this enables us to use the distribution of values in our sample to tell us something about the likely distribution of values in the parent population from which the sample was drawn.

However, sheer bad luck can mean that randomly drawn elements are disproportionately concentrated amongst some parts of the population at the expense of others, particularly when the size of our sample is small relative to the population from which it was drawn. For example, a survey of household incomes might happen to select households with unusually low incomes. Spatially systematic sampling aims to circumvent this problem and ensure greater evenness of coverage across the sample frame. This is achieved by identifying a regular sampling interval k (equal to the reciprocal of the sampling fraction N/n, where n is the required sample size and N is the size of the population) and proceeding to select every kth element. In spatial terms, the sampling interval of spatially systematic samples maps into a regularly spaced grid, as shown in Figure 4.4B. This advantage over simple random sampling may be two-edged, however, if the sampling interval and the spatial structure of the study area coincide, that is, the sample frame exhibits *periodicity*. A sample survey of urban land use along streets originally surveyed under the US Public Land Survey System (PLSS: Section 5.5) would be ill-advised to take a sampling interval of one mile, for example, for this was the interval at which blocks within townships were originally laid out, and urban structure is still likely to be repetitive about this original design. In such instances, there may be a consequent failure to detect the true extent of heterogeneity of population attributes (Figure 4.4B) – for example, it is extremely unlikely that the attributes of street intersection

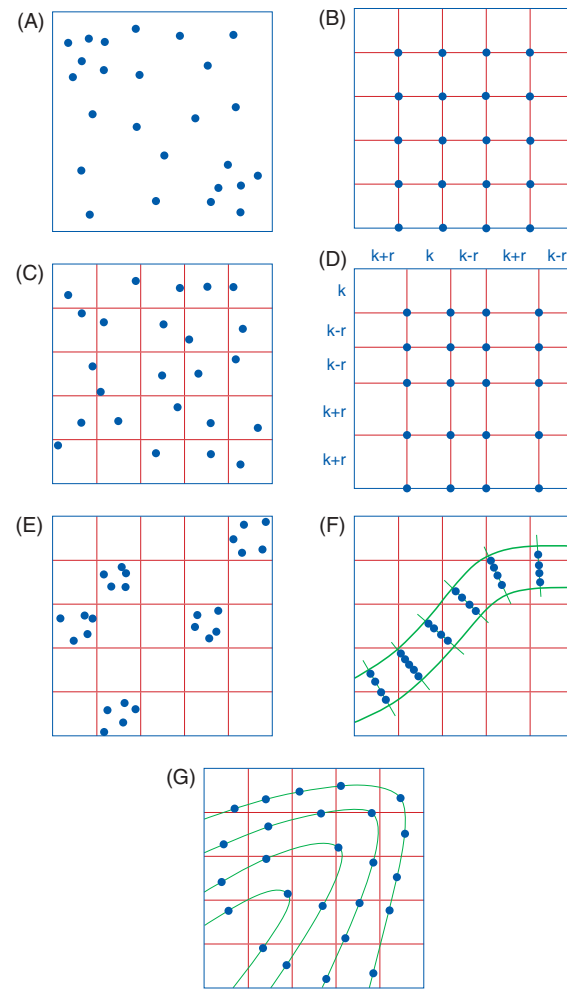


Figure 4.4 Spatial sample designs: (A) simple random sampling; (B) stratified sampling; (C) stratified random sampling; (D) stratified sampling with random variation in grid spacing; (E) clustered sampling; (F) transect sampling; and (G) contour sampling

locations would be representative of land uses elsewhere in the block structure. A number of systematic and quasi-systematic sample designs have been devised to get around the vulnerability of spatially systematic sample designs to periodicity, and the danger that simple random sampling may generate freak samples. These include stratified random sampling to ensure evenness of coverage (Figure 4.4C) and periodic random changes in the grid width of a spatially systematic sample (Figure 4.4D), perhaps subject to minimum spacing intervals.

In certain circumstances, it may be more efficient to restrict measurement to a specified range of sites – because of the prohibitive costs of transport over large areas, for example. Clustered sample designs, such as that shown in Figure 4.4E, may be used to generalize about attributes if the cluster presents a microcosm of surrounding conditions. In fact this provides a legitimate use of a comprehensive study of one area to say something about conditions beyond it – so long as the study area is known to be representative of the broader study region. For example, political opinion polls are often taken in shopping centers where shoppers can be deemed broadly representative of the population at large. However, instances where they provide a comprehensive detailed picture of spatial structure are likely to be the exception rather than the rule.

Use of either simple random or spatially systematic sampling presumes that each observation is of equal importance, and hence of equal weight, in building a representation. As such, these sample designs are suitable for circumstances in which spatial structure is weak or non-existent, or where (as in circumstances fully described by Tobler's Law) the attenuating effect of distance is constant in all directions. They are also suitable in circumstances where spatial structure is unknown. Yet in most practical applications, spatial structure is (to some extent at least) known, even if it cannot be wholly explained by Tobler's Law. These conditions make it both more efficient and necessary to devise application-specific sample designs. This makes for improved quality of representation, with minimum resource costs of collecting data. Relevant sample designs include sampling along a transect, such as a soil profile (Figure 4.4F), or along a contour line (Figure 4.4G).

Consider the area of Leicestershire, UK, illustrated in Figure 4.5. It depicts a landscape in which the hilly relief of an upland area falls away sharply towards a river's flood plain. In identifying the sample spot heights that we might measure and hold in a GIS to create a representation of this area, we would be advised to sample a disproportionate number of observations in the upland area of the study area where the local variability of heights is greatest.

In a socio-economic context, imagine that you are required to identify the total repair cost of bringing all housing in a city up to a specified standard. (Such applications are common, for example, in forming bids for Federal or Central Government funding.) A GIS that showed the time period in which different neighborhoods were developed (such as the Mid-West settlements simulated in Figure 2.18) would provide a useful guide to effective use of sampling resources. Newer houses are all likely to be in more or less the same condition, while the repair costs of the older houses are likely to be much more variable and dependent upon the attention that the occupants have lavished upon them. As a general rule, the older neighborhoods warrant a greater sampling frequency than the newer ones, but other considerations may also be accommodated into the sampling design as well – such as construction type (duplex versus apartment, etc.) and local geology (as an indicator of risk of subsidence).

In any application, where the events or phenomena that we are studying are spatially heterogeneous, we will require a large sample to capture the full variability of attribute values at all possible locations. Other parts of the study area may be much more homogeneous in attributes, and a sparser sampling interval may thus be more appropriate. Both simple random and systematic sample designs (and their variants) may be adapted in order to allow a differential sampling interval over a given study area (see Section 6.3.2 for more on this issue with respect to sampling vegetation cover). Thus it may be sensible to partition the sample frame into sub-areas, based on our knowledge of spatial structure – specifically our knowledge of the likely variability of the attributes that we are measuring.

Other application-specific special circumstances include:

- whether source data are ubiquitous or must be specially collected;
- the resources available for any survey undertaking; and
- the accessibility of all parts of the study area to field observation (still difficult even in the era of ubiquitous availability of Global Positioning System receivers: Section 5.8).

Stratified sampling designs attempt to allow for the unequal abundance of different phenomena on the Earth's surface.

It is very important to be aware that this discussion of sampling is appropriate to problems where there is a large hypothetical population of evenly distributed locations (elements, in the terminology of sampling theory, or atoms of information in the terminology of Section 3.4), that each have a known and prespecified probability of selection. Random selection of elements plays a part in each of the sample designs illustrated in Figure 4.4, albeit that the probability of selecting an element may be greater for clearly defined sub-populations – that lie along a contour line or across a soil transect, for example. In circumstances where spatial structure is either weak or is explicitly incorporated through clear definition of sub-populations, standard statistical theory provides a robust framework for inferring the attributes of the population from those of the sample. But the reality is somewhat messier. In most GIS applications, the population of elements (animals, glacial features, voters) may not be large, and its distribution across space may be far from random and independent. In these circumstances, conventional wisdom suggests a number of 'rules of

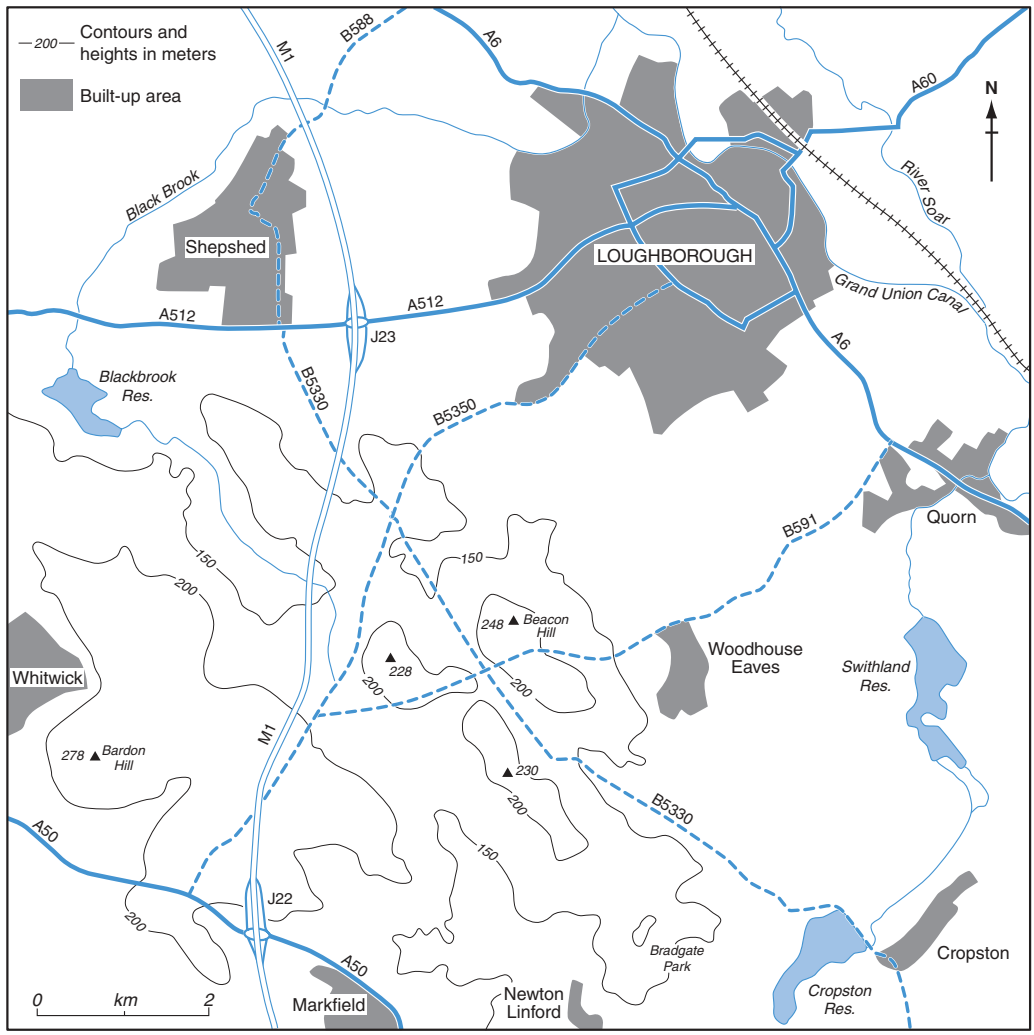


Figure 4.5 An example of physical terrain in which differential sampling would be advisable to construct a representation of elevation (Reproduced by permission of M. Langford, University of Glamorgan)

thumb' to compensate for the likely increase in error in estimating the true population value – as in clustered sampling, where slightly more than doubling the sample size is usually taken to accommodate the effects of spatial autocorrelation within a spatial cluster. However, it may be considered that the existence of spatial autocorrelation fundamentally undermines the inferential framework and invalidates the process of generalizing from samples to populations. We return to discuss this in more detail in our discussion of inference and hypothesis testing in Section 15.4.1.

Finally, it is also worth noting that this discussion assumes that we have the luxury of collecting our own data for our own particular purpose. The reality of analysis in our data-rich world is that more and more of the data that we use are collected by other parties for other purposes: in such cases the metadata of the dataset are crucially important in establishing their provenance for the particular investigation that we may wish to undertake (see Section 11.2.1).

4.5 Distance decay

In selectively abstracting, or sampling, part of reality to hold within a representation, judgment is required to fill in the gaps between the observations that make up a representation. This requires understanding of the likely attenuating effect of distance between the sample observations, and thus of the nature of geographic data (Figure 4.6). That is to say, we need to make an informed judgment about an appropriate *interpolation* function and how to *weight* adjacent observations. A literal interpretation of Tobler's Law implies a continuous, smooth, attenuating effect of distance upon the attribute values of adjacent or contiguous spatial objects, or incremental variation in attribute values as we traverse a field. The polluting effect of a chemical or oil spillage decreases in a predictable (and in still waters, uniform) fashion with distance from the point source; aircraft noise

Figure 4.6 We require different ways of interpolating between points, as well as different sample designs, for representing mountains and forested hillsides

decreases in a linear fashion with distance from the flight path; and the number of visits to a National Park decreases at a regular rate as we traverse the counties that adjoin it. This section focuses on principles, and introduces some of the functions that are used to describe effects over distance, or the nature of geographic variation, while Section 14.4.4 discusses ways in which the principles of distance decay are embodied in techniques of spatial interpolation.

The precise nature of the function used to represent the effects of distance is likely to vary between applications, and Figure 4.7 illustrates several hypothetical types. In mathematical terms, we take b as a parameter that affects the rate at which the weight w_{ij} declines with distance: a small b produces a slow decrease, and a large b a more rapid one. In most applications, the choice of distance attenuation function is the outcome of past experience, the fit of a particular application dataset, and convention. Figure 4.7A presents the simple case of linear distance decay, given by the expression:

$$w_{ij} = a - bd_{ij},$$

for $d_{ij} < a/b$, as might reflect the noise levels experienced across a transect perpendicular to an aircraft flight path. Figure 4.7B presents a negative power distance decay function, given by the expression:

$$w_{ij} = d_{ij}^{-b},$$

which has been used by some researchers to describe the decline in the density of resident population with distance from historic central business district (CBD) areas. Figure 4.7C illustrates a negative exponential statistical fit, given by the expression:

$$w_{ij} = e^{-bd_{ij}},$$

conventionally used in human geography to represent the decrease in retail store patronage with distance from it.

Each of the attenuation functions illustrated in Figure 4.7 is idealized, in that the effects of distance are presumed to be regular, continuous, and *isotropic* (uniform in every direction). This may be appropriate for many applications. The notion of smooth and continuous variation underpins many of the representational traditions in cartography, as in the creation of *isopleth* (or isoline) maps. This is described in Box 4.3. To some extent at least, high school math also conditions us to think of spatial variation as continuous, and as best represented by interpolating smooth curves between everything. Yet our understanding of spatial structure tells us that variation is often far from smooth and continuous. The Earth's surface and geology, for example, are discontinuous at cliffs and fault lines, while the socio-economic geography of cities can be similarly characterized by abrupt changes. Some illustrative physical and social issues pertaining to the catchment of a grocery store are presented in Figure 4.8. A naïve GIS analysis might assume that the maximum extent of the catchment of the store is bounded by an approximately circular area, and that within this area the likelihood (or probability) of shoppers using the store decreases the further away from it that they live. On this basis we might assume a negative exponential distance decay function (Figure 4.7C) and, for practical purposes, an absolute cut-off in patronage beyond a ten

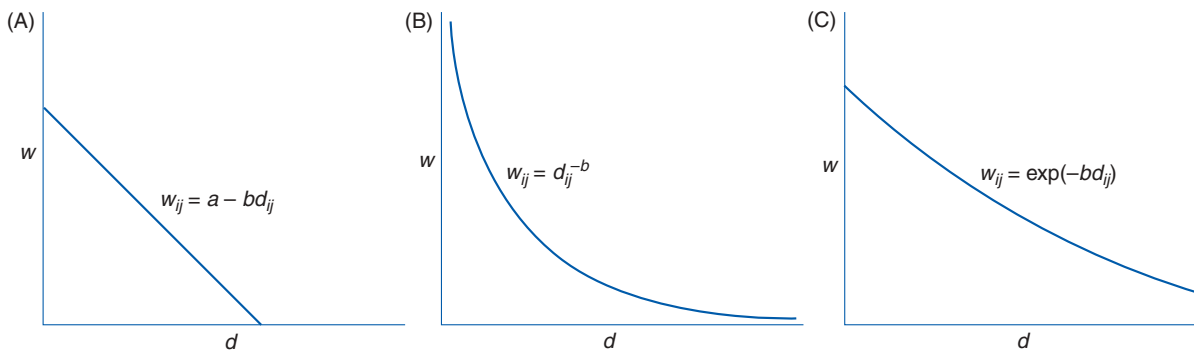


Figure 4.7 The attenuating effect of distance: (A) linear distance decay, $w_{ij} = a - bd_{ij}$; (B) negative power distance decay, $w_{ij} = {d_{ij}}^{-b}$; and (C) negative exponential distance decay, $w_{ij} = \exp(-bd_{ij})$

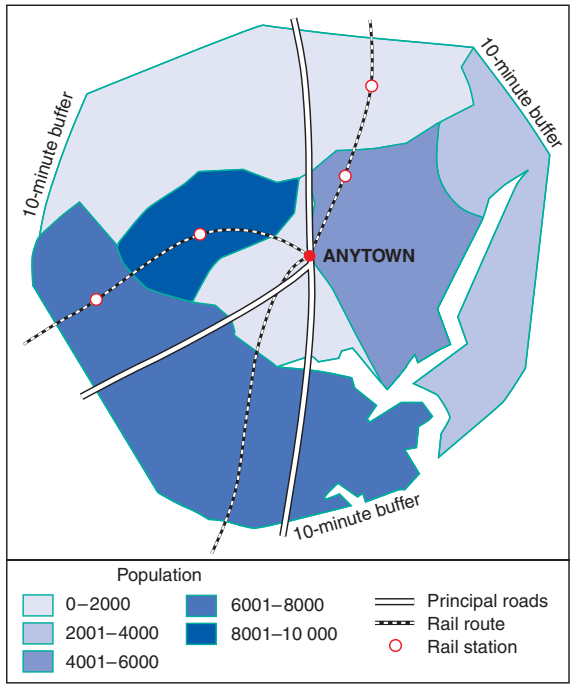


Figure 4.8 Discontinuities in a retail catchment (*Source*: Adapted from Birkin M., Clarke G. P., Clarke M. and Wilson A. G. (1996) *Intelligent GIS*. Cambridge, UK: GeoInformation International). Reproduced by permission of John Wiley & Sons Inc.

minute drive time at average speed. Yet in practice, the catchment also depends upon:

- physical factors, such as rivers and relief;
- road and rail infrastructure and associated capacity, congestion, and access (e.g., rail stations and road access ramps);
- socio-economic factors, that are manifest in differences in customer store preferences;
- administrative geographies, that modify the shape of the circle because census counts of population are only available for administrative zones;
- overlapping trade areas of competing stores, that are likely to truncate the trade area from particular directions;
- a demand constraint, that requires probabilities of patronizing all available stores at any point to sum to 1 (unless people opt out of shopping).

Additionally, the population base to Figure 4.8 raises an important issue of the representation of spatial structure. Remember that we said that the circular retail catchment had to be adapted to fit the administrative geography of population enumeration. The distribution of population is shown using choropleth mapping (Box 4.3), which implicitly assumes that the mapped property is uniformly distributed within zones and that the only important changes in distribution take place at zone boundaries. Such representations can obscure continuous variations and mask the true pattern of distance attenuation.

4.6 Measuring distance effects as spatial autocorrelation

An understanding of spatial structure helps us to deduce a good sampling strategy, to use an appropriate means of interpolating between sampled points, and hence to build a spatial representation that is fit for purpose. Knowledge of the actual or likely nature of spatial autocorrelation can thus be used *deductively* in order to help build a spatial representation of the world. However, in many applications we do not understand enough about geographic variability, distance effects, and spatial structure to invoke deductive reasoning. A further branch of spatial analysis thus emphasizes the *measurement* of spatial autocorrelation as an end in itself. This amounts to a more *inductive* approach to developing an understanding of the nature of a geographic dataset.

Induction reasons from data to build up understanding, while deduction begins with theory and principle as a basis for looking at data.

In Section 4.3 we saw that spatial autocorrelation measures the extent to which similarities in position match similarities in attributes. Methods of measuring spatial autocorrelation depend on the types of objects used as the basis of a representation, and as we saw in Section 4.2, the scale of attribute measurement is important too. Interpretation depends on how the objects relate to our conceptualization of the phenomena they represent. If the phenomenon of interest is conceived as a field, then spatial autocorrelation measures the smoothness of the field using data from the sample points, lines, or areas that represent the field. If the phenomena of interest are conceived as discrete objects, then spatial autocorrelation measures how the attribute values are distributed among the objects, distinguishing between arrangements that are clustered, random, and locally contrasting. Figure 4.11 shows examples of each of the four object types, with associated attributes, chosen to represent situations in which a scientist might wish to measure spatial autocorrelation. The point data in Figure 4.11A comprise data on well bores over an area of 30 km^2, and together provide information on the depth of an aquifer beneath the surface (the blue shading identifies those within a given threshold). We would expect values to exhibit strong spatial autocorrelation, with departures from this indicative of changes in bedrock structure or form. The line data in Figure 4.11B present numbers of accidents for links of road over a lengthy survey period in the Southwestern Ontario, Canada, provincial highway network. Low spatial autocorrelation in these statistics implies that local causative factors (such as badly laid out junctions) account for most accidents, whereas strong spatial autocorrelation would imply a more regional scale of variation, implying a link between accident rates and lifestyles, climate, or population density. The area data in Figure 4.11C illustrate the socio-economic patterning of the south east of England, and beg the question of

Technical Box **4.3**

Isopleth and choropleth maps

Isopleth maps are used to visualize phenomena that are conceptualized as fields, and measured on interval or ratio scales. An *isoline* connects points with equal attribute values, such as contour lines (equal height above sea level), *isohyets* (points of equal precipitation), *isochrones* (points of equal travel time), or *isodapanes* (points of equal transport cost). Figure 4.9 illustrates the procedures that are used to create a surface about a set of point measurements (Figure 4.9A), such as might be collected from rain gauges across a study region (and see Section 14.4.4 for more technical detail on the process of spatial interpolation). A parsimonious number of user-defined values is identified to define the contour intervals (Figure 4.9B). The GIS then interpolates a contour between point observations of greater and lesser value (Figure 4.9C) using standard procedures of inference, and the other contours are then interpolated using the same procedure (Figure 4.9D). Hue or shading can be added to improve user interpretability (Figure 4.9E).

Choropleth maps are constructed from values describing the properties of non-overlapping areas, such as counties or census tracts. Each area is colored, shaded, or cross-hatched to symbolize the value of a specific variable, as in Figure 4.10. Geographic rules define what happens to the properties of objects when they are split or merged (e.g., see Section 13.3.3). Figure 4.10 compares a map of total population (a spatially extensive variable) with a map of population density (a spatially intensive variable). Spatially extensive variables take values that are true only of entire areas, such as total population, or total number of children under 5 years of age. They are highly misleading – the same color is applied uniformly to each part of an area, yet we know that the mapped property cannot be true of each part of the area. The values taken by spatially intensive variables could potentially be true of every part of an area, if the area were homogeneous – examples include densities, rates, or proportions. Conceptually, a spatially intensive variable is a field, averaged over each area, whereas a spatially extensive variable is a field of density whose values are summed or integrated to obtain each area's value.

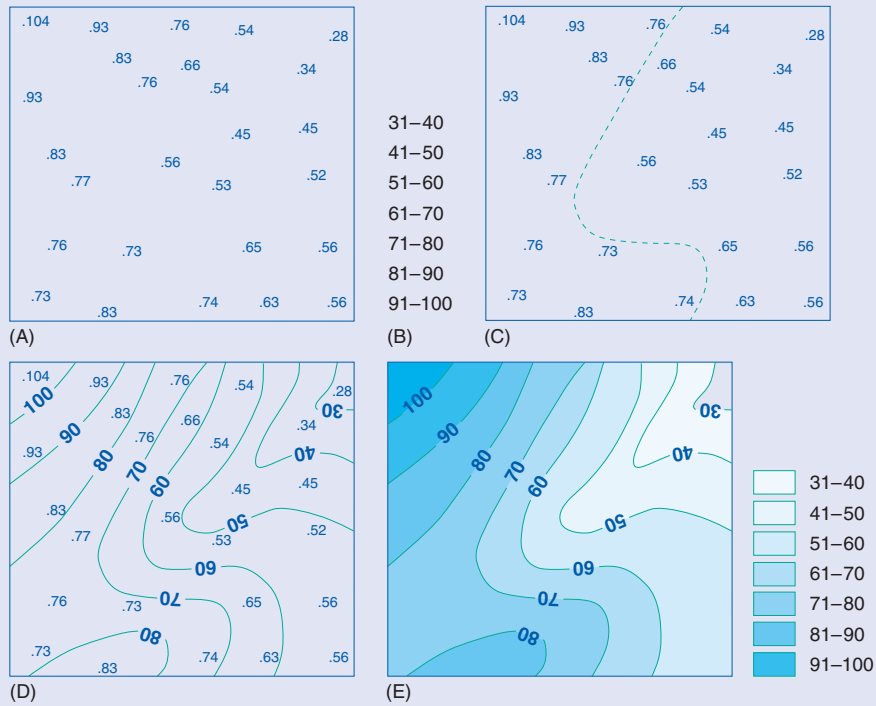


Figure 4.9 The creation of isopleth maps: (A) point attribute values; (B) user-defined classes; (C) interpolation of class boundary between points; (D) addition and labeling of other class boundaries; and (E) use of hue to enhance perception of trends (after Kraak and Ormeling 1996: 161)

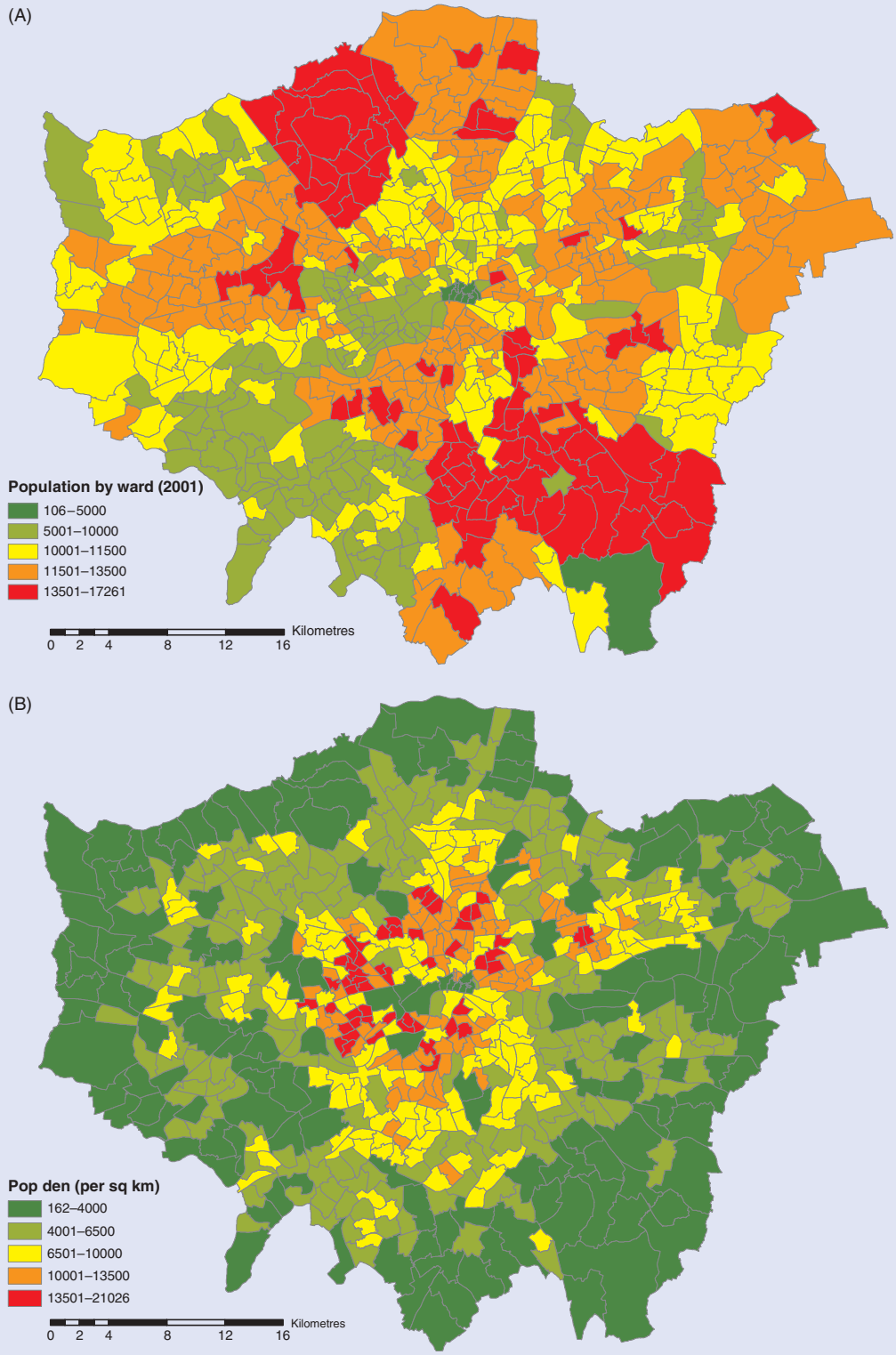
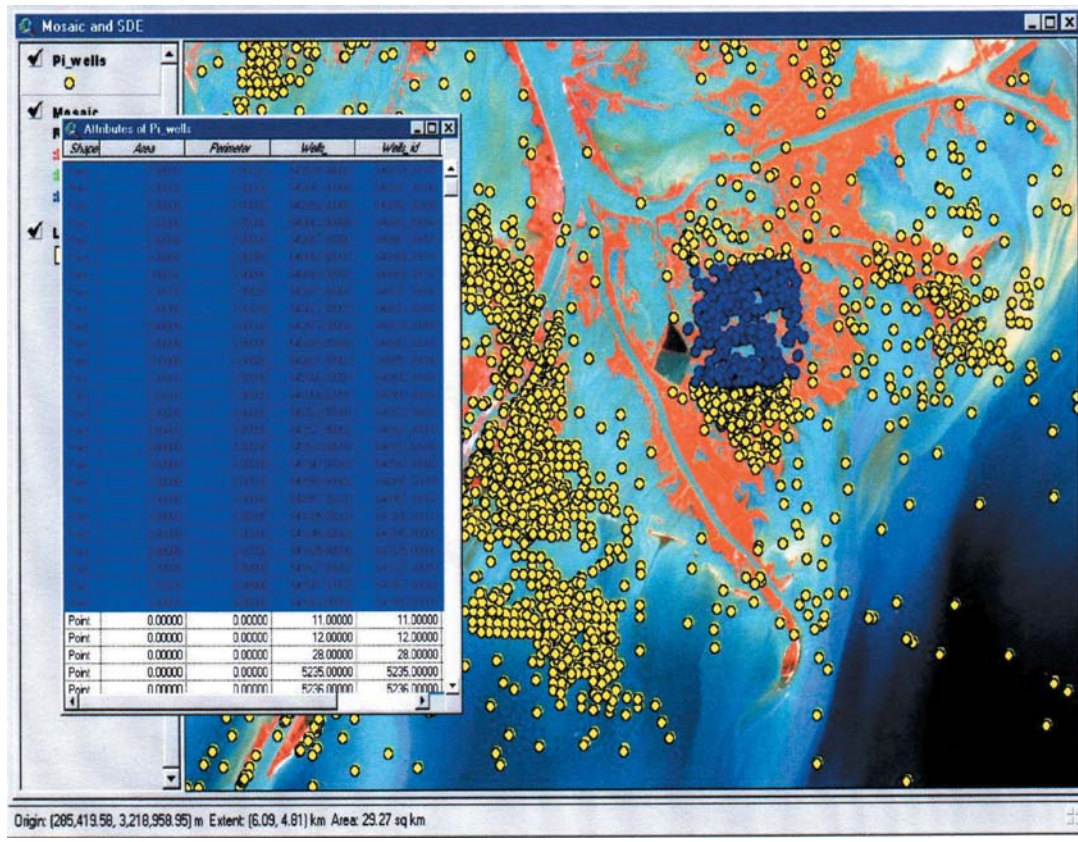


Figure 4.10 Choropleth maps of (A) a spatially extensive variable, total population, and (B) a related but spatially intensive variable, population density. Many cartographers would argue that (A) is misleading, and that spatially extensive variables should always be converted to spatially intensive form (as densities, ratios, or proportions) before being displayed as choropleth maps. (Reproduced with permission of Daryl Lloyd)

(A)



(B)

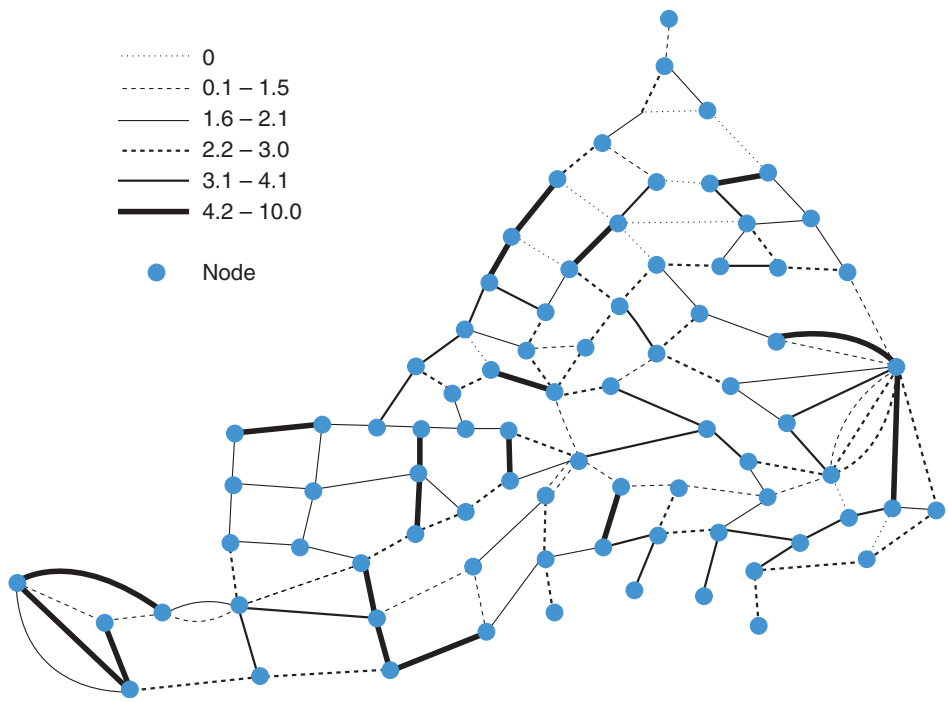


Figure 4.11 Situations in which a scientist might want to measure spatial autocorrelation: (A) point data (wells with attributes stored in a spreadsheet); (B) line data (accident rates in the Southwestern Ontario provincial highway network); (C) area data (percentage of population that are old age pensioners (OAPs) in South East England); and (D) volume data (elevation and volume of buildings in Seattle). (A) and (D) courtesy ESRI; (C) reproduced with permission of Daryl Lloyd

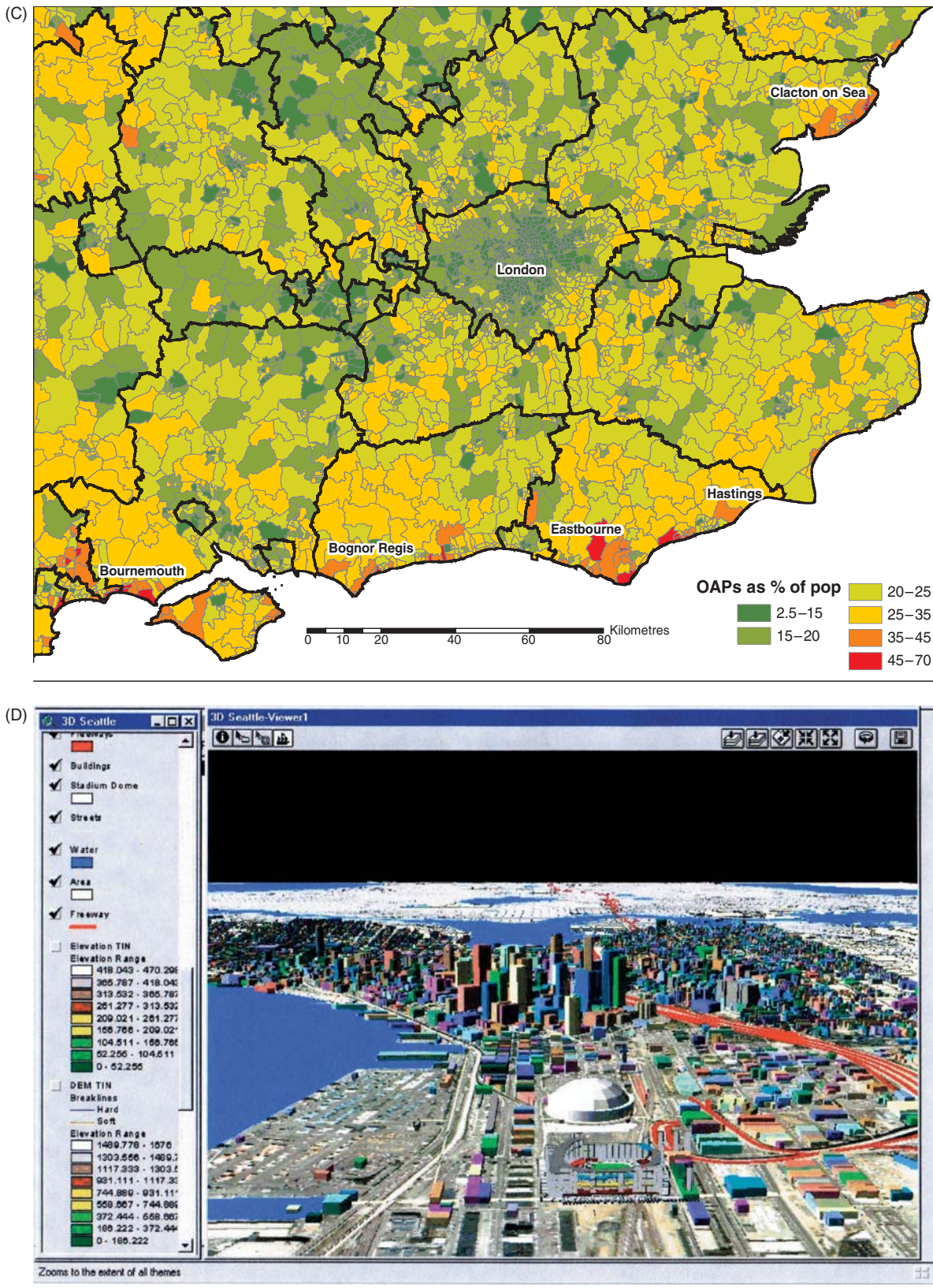


Figure 4.11 (*continued*)

Technical Box **4.4**

Measuring similarity between neighbors

In the simple example shown in Figure 4.12, we compare neighboring values of spatial attributes by defining a weights matrix **W** in which each element w_{ij} measures the locational similarity of i and j (i identifies the row and j the column of the matrix). We use a simple measure of contiguity, coding $w_{ij} = 1$ if regions i and j are contiguous and $w_{ij} = 0$ otherwise. w_{ii} is set equal to 0 for all i. This is shown in Table 4.1.

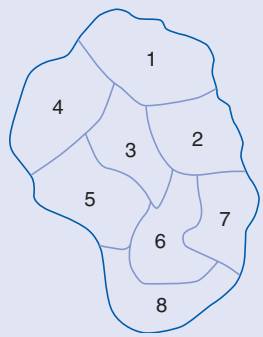


Figure 4.12 A simple mosaic of zones

Table 4.1 The weights matrix **W** derived from the zoning system shown in Figure 4.12

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	1	1	1	0	0	0	0
2	1	0	1	0	0	1	1	0
3	1	1	0	1	1	1	0	0
4	1	0	1	0	1	0	0	0
5	0	0	1	1	0	1	0	1
6	0	1	1	0	1	0	1	1
7	0	1	0	0	0	1	0	1
8	0	0	0	0	1	1	1	0

The weights matrix provides a simple way of representing similarities between location and attribute values, in a region of contiguous areal objects. Autocorrelation is identified by the presence of neighboring cells or zones that take the same (binary) attribute value. More sophisticated measures of w_{ij} include a decreasing function (such as one of those shown in Figure 4.7) of the straight line distance between points at the centers of zones, or the lengths of common boundaries. A range of different spatial metrics may also be used, such as existence of linkage by air, or a decreasing function of travel time by air, road, or rail, or the strength of linkages between individuals or firms on some (non-spatial) network.

The weights matrix makes it possible to develop measures of spatial autocorrelation using four of the attribute types (nominal, ordinal, interval, and ratio, but not, in practice, cyclic) in Box 3.3 and the dimensioned classes of spatial objects in Box 4.1. Any measure of spatial autocorrelation seeks to compare a set of locational similarities w_{ij} (contained in a weights matrix) with a corresponding set of attribute similarities c_{ij}, combining them into a single index in the form of a cross-product:

$$\sum_i \sum_j c_{ij} w_{ij},$$

This expression is the total obtained by multiplying every cell in the **W** matrix with its corresponding entry in the **C** matrix, and summing.

There are different ways of measuring the attribute similarities, c_{ij}, depending upon whether they are measured on the nominal, ordinal, interval, or ratio scale. For nominal data, the usual approach is to set c_{ij} to 1 if i and j take the same attribute value, and zero otherwise. For ordinal data, similarity is usually based on comparing the ranks of i and j. For interval and ratio data, the attribute of interest is denoted z_i, and the product $(z_i - \overline{z})(z_j - \overline{z})$ is calculated, where $\overline{z}$ denotes the average of the zs.

One of the most widely used spatial autocorrelation statistics for the case of area objects and interval-scale attributes is the Moran Index. This is positive when nearby areas tend to be similar in attributes, negative when they tend to be more dissimilar than one might expect, and approximately zero when attribute values are arranged randomly and independently in space. It is given by the expression:

$$I = \frac{n \sum_i \sum_j w_{ij}(z_i - \overline{z})(z_j - \overline{z})}{\sum_i \sum_j w_{ij} \sum_i (z_i - \overline{z})^2}$$

where n is the number of areal objects in the set. This brief exposition is provided at this point to emphasize the way in which spatial autocorrelation measures are able to accommodate attributes scaled as nominal, ordinal, interval, and ratio data, and to illustrate that there is flexibility in the nature of contiguity (or adjacency) relations that may be specified. Further techniques for measuring spatial autocorrelation are reviewed in connection with spatial interpolation in Section 14.4.4.

whether, at a regional scale, there are commonalties in household structure. The volume data in Figure 4.11D allow some measure of the spatial autocorrelation of high-rise structures to be made, perhaps as part of a study of the way that the urban core of Seattle functions. The way that spatial autocorrelation might actually be calculated for the data used to construct Figure 4.11C is described in Box 4.4.

4.7 Establishing dependence in space

Spatial autocorrelation measures tell us about the interrelatedness of phenomena across space, one attribute at a time. Another important facet to the nature of geographic data is the tendency for relationships to exist between different phenomena at the same place – between the values of two different fields, between two attributes of a set of discrete objects, or between the attributes of overlapping discrete objects. This section introduces one of the ways of describing such relationships (see also Box 4.5).

How the various properties of a location are related is an important aspect of the nature of geographic data.

In a formal statistical sense, regression analysis allows us to identify the *dependence* of one variable upon one or more *independent* variables. For example, we might hypothesize that the value of individual properties in a city is dependent upon a number of variables such as floor area, distance to local facilities such as parks and schools, standard of repair, local pollution levels, and so forth. Formally this may be written:

$$Y = f(X_1, X_2, X_3, \ldots, X_K)$$

where Y is the dependent variable and X_1 through X_K are all of the possible independent variables that might impact upon property value. It is important to note that it is the independent variables that together affect the dependent variable, and that the hypothesized causal relationship is one way – that is, that property value is *responsive* to floor area, distance to local facilities, standard of repair, and pollution, and not vice versa. For this reason the dependent variable is termed the *response* variable and the independent variables are termed *predictor* variables in some statistics textbooks.

In practice, of course, we will never successfully predict the exact values of any sample of properties. We can identify two broad classes of reasons why this might be the case. First, a property price outcome is the response to a huge range of factors, and it is likely that we will have evidence of and be able to measure only a small

Biographical Box **4.5**

Dawn Wright, marine geographer

Figure 4.13 Dawn Wright, marine geographer, and friend Lydia

Dawn Wright (a.k.a. 'Deepsea Dawn' by colleagues and friends: Figure 4.13) is a professor of Geography and Oceanography at Oregon State University (OrSt) in Corvallis, Oregon, USA, where she also directs Davey Jones Locker, a seafloor mapping and marine GIS research laboratory.

Shortly after the deepsea vehicle *Argo I* was used to discover the *HMS Titanic* in 1986, Dawn used some of the first GIS datasets that it collected to develop her Ph.D. at the University of California, Santa Barbara. It was then that she became acutely aware of the challenges of applying GIS to deep ocean environments. When we discuss the nature of the Earth's surface (this chapter) and the way in which it is georeferenced (Chapter 5), we implicitly assume that it is above sea level. Dawn has written widely on the nature of geographic data with regard to the entirety of the Earth's surface – especially the 70% covered by water. Research issues endemic to oceanographic applications of GIS include the handling of spatial data structures that can vary their relative positions and values over time, geostatistical interpolation (Box 4.3 and Section 14.4.4) of data that are sparser in one dimension as compared to the others, volumetric analysis, and the input and management of very large spatial databases. Dawn's research has described the range of these issues and applications, as well as recent advances in marine map-making, charting, and scientific visualization.

Dawn remains a strong advocate of the potential of these issues to not only advance the body of knowledge in GIS design and architecture, but also to inform many of the long-standing research challenges of geographic information science. She says, 'The ocean forces us to think about the nature of geographic data in different ways and to consider radically different ways of representing space and time – we have to go ''super-dimensional'' to get our minds and our maps around the natural processes at work. We cannot fully rely on the absolute coordinate systems that are so familiar to us in a GIS, or ignore the dissimilarity between the horizontal and the vertical dimension when measuring geographic features and objects. How deep is the ocean at any precise moment in time? How do we represent all of the relevant attributes of the habitat of marine mammals? How can we enforce marine protected area boundaries at depth? Much has been written about the importance of error and uncertainty in geographic analysis. The challenge of gathering data in dynamic marine environments using platforms that are constantly in motion in all directions (roll, pitch, yaw, heave), or of tracking fish, mammals, and birds at sea, creates critical challenges in managing uncertainty in marine position.' These issues of uncertainty (see Chapter 6) also have implications for the establishment of dependence in space (Section 4.7). Dawn and her students continue to develop methods, techniques, and tools for handling data in GIS, but with a unique oceanographic take on data modeling, geocomputation, and the incorporation of spatio-temporal data standards and protocols. Take a dive into Davey Jones Locker to learn more (**dusk.geo.orst.edu/djl**).

subset of these. Second, even if we were able to identify and measure every single relevant independent variable, we would in practice only be able to do so to a given level of measurement precision (for a more detailed discussion of what we mean by precision see Box 6.3 and Section 6.3.2.2). Such caveats do not undermine the wider rationale for trying to generalize, since any assessment of the effects of variables we know about is better than no assessment at all. But our conceptual solution to the problems of unknown and imprecisely measured variables is to subsume them all within a statistical error term, and to revise our regression model so that it looks like this:

$$Y = f(X_1, X_2, X_3, \ldots, X_K) + \varepsilon$$

where ε denotes the error term.

We assume that this relationship holds for each case (which we denote using the subscript i) in our population of interest, and thus:

$$Y_i = f(X_{i1}, X_{i2}, X_{i3}, \ldots, X_{iK}) + \varepsilon_i$$

The essential task of regression analysis is to identify the *direction* and *strength* of the association implied by this equation. This becomes apparent if we rewrite it as:

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + \cdots + b_KX_{iK} + \varepsilon_i$$

where b_1 through b_K are termed regression *parameters*, which measure the direction and strength of the influence of the independent variables X_1 through X_K on Y. b_0 is termed the *constant* or *intercept* term. This is illustrated in simplified form as a scatterplot in Figure 4.14. Here, for reasons of clarity, the values of the dependent (Y) variable (property value) are regressed and plotted against just one independent (X) variable (floorspace; for more on scatterplots see Section 14.2). The scatter of points

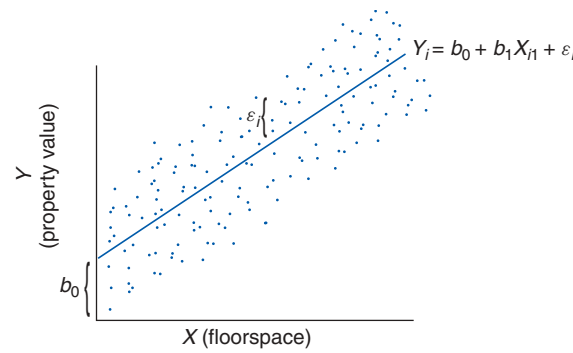


Figure 4.14 The fit of a regression line to a scatter of points, showing intercept, slope and error terms

exhibits an upward trend, suggesting that the response to increased floorspace is a higher property price. A *best fit* line has been drawn through this scatter of points. The gradient of this line is calculated as the b parameter of the regression, and the upward trend of the regression line means that the gradient is positive. The greater the magnitude of the b parameter, the stronger the (in this case positive) effect of marginal increases in the X variable. The value where the regression line intersects the Y axis identifies the property value when floorspace is zero (which can be thought of as the value of the land parcel when no property is built upon it), and gives us the intercept value b_0. The more general multiple regression case works by extension of this principle, and each of the b parameters gauges the marginal effects of its respective X variable.

This kind of effect of floorspace area upon property value is intuitively plausible, and a survey of any sample of individual properties is likely to yield the kind of well-behaved plot illustrated in Figure 4.14. In other cases the overall trend may not be as unambiguous. Figure 4.15A presents a hypothetical plot of the effect of

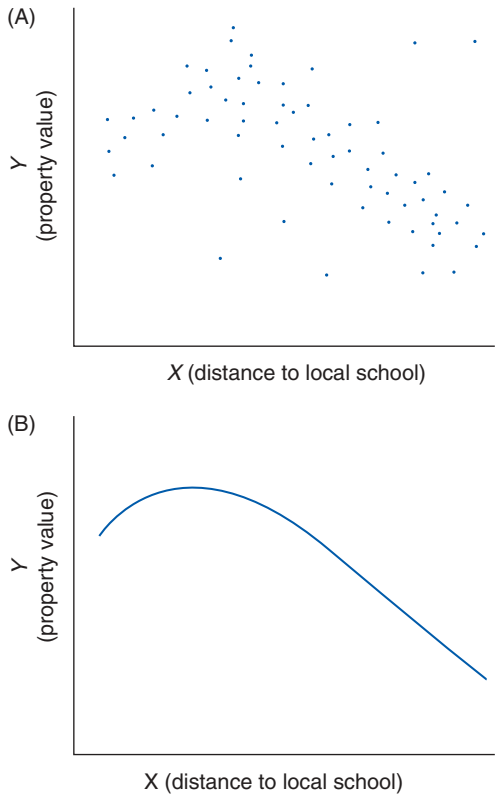


Figure 4.15 (A) A scatterplot and (B) hypothetical relationship between distance to local school and domestic property value

distance to a local school (measured perhaps as straight line distance; see Section 14.3.1 for more on measuring distance in a GIS) upon property value. Here the plot is less well behaved: the overall fit of the regression line is not as good as it might be, and a number of poorly fitting observations (termed high-residual and high-leverage points) present exceptions to a weak general trend. A number of formal statistical measures (notably t statistics and the R^2 measure) as well as less formal diagnostic procedures exist to gauge the statistical fit of the regression model to the data. Details of these can be found in any introductory statistics text, and will not be examined here.

It is easiest to assume that a relationship between two variables can be described by a straight line or linear equation, and that assumption has been followed in this discussion. But although a straight line may be a good first approximation to other functional forms (curves, for example), there is no reason to suppose that linear relationships represent the *truth*, in other words, how the world's social and physical variables are actually related. For example, it might be that very close proximity to the school has a negative effect upon property value (because of noise, car parking, and other localized nuisance) and it is properties at intermediate distances that gain the greatest positive neighborhood effect from this amenity. This is shown in Figure 4.15B: these and other effects might be accommodated by changing the intrinsic *functional form* of the model.

A straight line or linear distance relationship is the easiest assumption to make and analyze, but it may not be the correct one.

Figure 4.16 identifies the discrepancy between one observed property value and the value that is predicted by the regression line. This difference can be thought of as the error term for individual property i (strictly speaking, it is termed a *residual* when the scatterplot depicts a sample and not a population). The precise slope and intercept of the best fit line is usually identified using the principle of *ordinary least squares* (OLS). OLS regression fits the line through the scatter of points such that the sum of squared residuals across the entire sample is minimized. This procedure is robust and statistically efficient, and yields estimates of the b parameters. But in many situations it is common to try to go further, by *generalizing* results. Suppose the data being analyzed can be considered a representative sample of some larger group. In the field case, sample points might be representative of all of the infinite number of sample points one might select to represent the continuous variation of the field variable (for example, weather stations measuring temperature in new locations, or more soil pits dug to measure soil properties). In the discrete object case, the data analyzed might be only a selection of all of the objects. If this is the case, then statistics provides methods for making accurate and unbiased statements about these larger populations.

Generalization is the process of reasoning from the nature of a sample to the nature of a larger group.

For this to work, several conditions have to hold. First, the sample must be representative, which means for example that every case (element) in the larger group or *population* has a prespecified and independent chance of being selected. The sampling designs discussed in Section 4.4 are one way of ensuring this. But all too

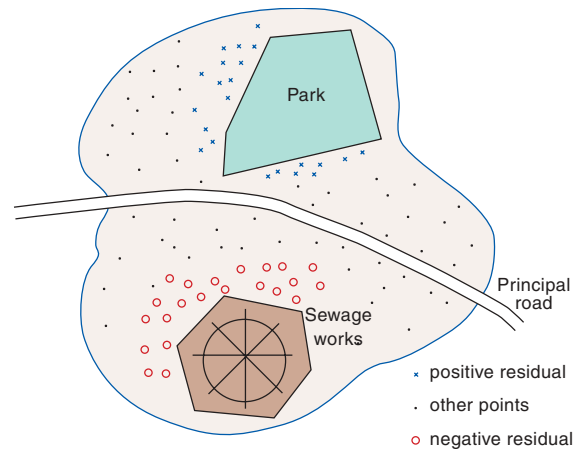


Figure 4.16 A hypothetical spatial pattern of residuals from a regression analysis

often in the analysis of geographic data it turns out to be difficult or impossible to imagine such a population. It is inappropriate, for example, to try to generalize from one study to statements about all of the Earth's surface, if the study was conducted in one area. Generalizations based on samples taken in Antarctica are clearly not representative of all of the Earth's surface. Often GIS provide complete coverage of an area, allowing us to analyze all of the census tracts in a city, or all of the provinces of China. In such cases the apparatus of generalization from samples to populations is unnecessary, and indeed, becomes meaningless.

In addition, the statistical apparatus that allows us to make inferences assumes that there is no *autocorrelation* between errors across space or time. This assumption clearly does not accord with Tobler's Law, where the greater relatedness of near things to one another than distant things is manifest in positive spatial autocorrelation. If strong (positive or negative) spatial autocorrelation is present, the inference apparatus of the ordinary least squares regression procedure rapidly breaks down. The consequence of this is that estimates of the population *b* parameters become imprecise and the statistical validity of the tests used to confirm the strength and direction of apparent relationships is seriously weakened.

The assumption of zero spatial autocorrelation that is made by many methods of statistical inference is in direct contradiction to Tobler's Law.

The spatial patterning of residuals can provide clues as to whether the structure of space has been correctly specified in the regression equation. Figure 4.16 illustrates the hypothetical spatial distribution of residuals in our property value example – the high clustering of negative residuals around the school suggests that some distance threshold should be added to the specification, or some function that negatively weights property values that are very close to the school. The spatial clustering of residuals can also help to suggest omitted variables that should have been included in the regression specification. Such variables might include the distance to a neighborhood facility that might have strong positive (e.g., a park) or negative (e.g., a sewage works) effect upon values in our property example.

A second assumption of the multiple regression model is that there is no intercorrelation between the independent variables, that is, that no two or more variables essentially measure the same construct. The statistical term for such intercorrelation is *multicollinearity*, and this is a particular problem in GIS applications. GIS is a powerful technology for combining information about a place, and for examining relationships between attributes, whether they be conceptualized as fields, or as attributes of discrete objects. The implication is that each attribute makes a distinct contribution to the total picture of geographic variability. In practice, however, geographic layers are almost always highly correlated. It is very difficult to imagine that two fields representing different variables over the same geographic area would not somehow reveal their common geographic location through similar patterns. For example, a map of rainfall and a map of population density would clearly have similarities, whether population was dependent on agricultural production and thus rainfall or tended to avoid steep slopes and high elevations where rainfall was also highest.

It is almost impossible to imagine that two maps of different phenomena over the same area would not reveal some similarities.

From this brief overview, it should be clear that there are many important questions about the applicability of such procedures to establishing statistical relationships using spatial data. We return to discuss these in more detail in Section 15.4.

4.8 Taming geographic monsters

Thus far in our discussion of the nature of geographic data we have assumed that spatial variation is smooth and continuous, apart from when we encounter abrupt truncations and discrete shifts at boundaries. However, much spatial variation is not smooth and continuous, but rather is jagged and apparently irregular. The processes which give rise to the form of a mountain range produce features that are spatially autocorrelated (for example, the highest peaks tend to be clustered), yet it would be wholly inappropriate to represent a mountainscape using smooth interpolation between peaks and valley troughs.

Jagged irregularity is a property which is also often observed across a range of scales, and detailed irregularity may resemble coarse irregularity in shape, structure, and form. We commented on this in Section 4.3 when we suggested that a rock broken off a mountain may, for reasons of lithology, represent the mountain in form, and this property is often termed *self-similarity*. Urban geographers also recognize that cities and city systems are also self-similar in organization across a range of scales, and the ways in which this echoes many of the earlier ideas of Christaller's Central Place Theory have been discussed in the academic literature. It is unlikely that idealized smooth curves and conventional mathematical functions will provide useful representations for self-similar, irregular spatial structures: at what scale, if any, does it become appropriate to approximate the San Andreas Fault system by a continuous curve? Urban geographers, for example, have long sought to represent the apparent decline in population density with distance from historic central business districts (CBDs), yet the three-dimensional profiles of cities are characterized by urban canyons between irregularly spaced high-rise buildings (Figure 4.11D). Each of these phenomena is characterized by spatial trends (the largest faults, the largest mountains, and the largest skyscrapers tend to be close to one another), but they are not contiguous and smoothly joined, and the kinds of surface functions shown in Figure 4.7 present inappropriate generalizations of their structure.

For many years, such features were considered geometrical monsters that defied intuition. More recently,

however, a more general geometry of the irregular, termed *fractal geometry* by Benoît Mandelbrot, has come to provide a more appropriate and general means of summarizing the structure and character of spatial objects. Fractals can be thought of as geometric objects that are, literally, between Euclidean dimensions, as described in Box 4.6.

In a self-similar object, each part has the same nature as the whole.

Fractal ideas are important, and for many phenomena a measurement of fractal dimension is as important as measures of spatial autocorrelation, or of medians and modes in standard statistics. An important application of

Technical Box **4.6**

The strange story of the lengths of geographic objects

How long is the coastline of Maine (Figure 4.17)? (Benoît Mandelbrot, a French mathematician, originally posed this question in 1967 with regard to the coastline of Great Britain.)

Figure 4.17 Part of the Maine coastline

We might begin to measure the stretch of coastline shown in Figure 4.18A. With dividers set to measure 100 km intervals, we would take approximately 3.4 swings and record a length of 340 km (Figure 4.18B).

If we then halved the divider span so as to measure 50 km swings, we would take approximately 7.1 swings and the measured length would increase to 355 km (Figure 4.18C).

If we halved the divider span once again to measure 25 km swings, we would take approximately 16.6 swings and the measured length would increase still further to 415 km (Figure 4.18D).

And so on until the divider span was so small that it picked up all of the detail on this particular representation of the coastline. But even that would not be the end of the story.

If we were to resort instead to field measurement, using a tape measure or the Distance Measuring Instruments (DMIs) used by highway departments, the length would increase still further, as we picked up detail that even the most detailed maps do not seek to represent.

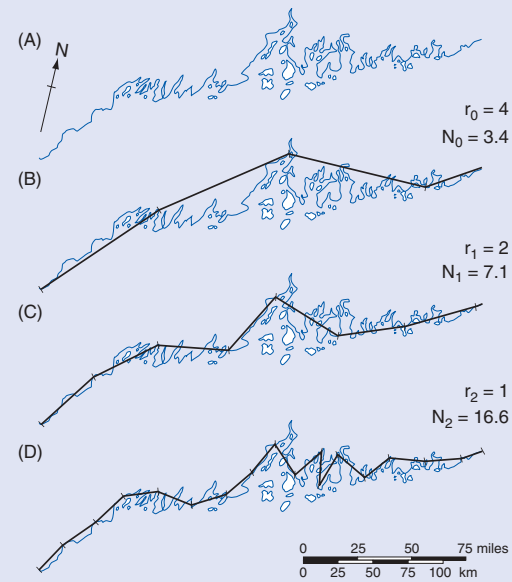


Figure 4.18 The coastline of Maine, at three levels of recursion: (A) the base curve of the coastline; (B) approximation using 100 km steps; (C) 50 km step approximation; and (D) 25 km step approximation

If we were to use dividers, or even microscopic measuring devices, to measure every last grain of sand or earth particle, our recorded length measurement would stretch towards infinity, seemingly without limit.

In short, the answer to our question is that the length of the Maine coastline is indeterminate. More helpfully, perhaps, any approximation is *scale-dependent* – and thus any measurement must also specify scale. The line representation of the coastline also possesses two other properties. First, where small deviations about the overall trend of the coastline resemble larger deviations in form, the coast is said to be *self-similar*. Second, as the path of the coast traverses space, its intricate structure comes to fill up more space than a one-dimensional straight line but less space than a two-dimensional area. As such, it is said to be of *fractional dimension* (and is termed a *fractal*) between 1 (a line) and 2 (an area).

fractal concepts is discussed in Section 15.2.5, and we return again to the issue of length estimation in GIS in Section 14.3.1. Ascertaining the fractal dimension of an object involves identifying the scaling relation between its length or extent and the yardstick (or level of detail) that is used to measure it. Regression analysis, as described in the previous section, provides one (of many) means of establishing this relationship. If we return to the Maine coastline example in Figures 4.17 and 4.18, we might obtain scale dependent coast length estimates (L) of 13.6 (4×3.4), 14.1 (2×7.1) and 16.6 (1×16.6) units for the step lengths (r) used in Figures 4.18B, 4.18C and 4.18D respectively. (It is arbitrary whether the steps are measured in miles or kilometers.) If we then plot the natural log of L (on the y-axis) against the natural log or r for these and other values, we will build up a scatterplot like that shown in Figure 4.19. If the points lie more or less on a straight line and we fit a regression line through it, the value of the slope (b) parameter is equal to $(1 - D)$, where D is the fractal dimension of the line. This method for analyzing the nature of geographic lines was originally developed by Lewis Fry Richardson (Box 4.7).

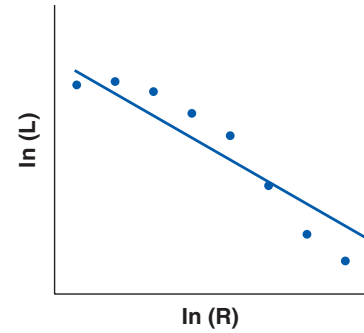


Figure 4.19 The relationship between recorded length (L) and step length (R)

4.9 Induction and deduction and how it all comes together

The abiding message of this chapter is that spatial is special – that geographic data have a unique nature. Tobler's Law presents an elementary general rule about spatial structure, and a starting point for the measurement and simulation of spatially autocorrelated structures. This in turn assists us in devising appropriate spatial sampling schemes and creating improved representations, which tell us still more about the real world and how we might represent it. A goal of GIS is often to establish causality between different geographically referenced data, and the multiple regression model potentially provides one means of relating spatial variables to one another, and of inferring from samples to the properties of the populations from which they were drawn. Yet statistical techniques often need to be recast in order to accommodate the special properties of spatial data, and regression analysis is no exception in this regard.

Spatial data provide the foundations to operational and strategic applications of GIS, foundations that must

Biographical Box **4.7**

Lewis Fry Richardson

Lewis Fry Richardson (1881–1953: Figure 4.20) was one of the founding fathers of the ideas of scaling and fractals. He was brought up a Quaker, and after earning a degree at Cambridge University went to work for the Meteorological Office, but his pacifist beliefs forced him to leave in 1920 when the Meteorological Office was militarized under the Air Ministry. His early work on how atmospheric turbulence is related at different scales established his scientific reputation. Later he became interested in the causes of war and human conflict, and in order to pursue one of his investigations found that he needed a rigorous way of defining the length of a boundary between two states. Unfortunately published lengths tended to vary dramatically, a specific instance being the difference between the lengths of the Spanish–Portuguese border as stated by Spain and by Portugal. He developed a method of walking a pair of dividers along a mapped line, and analyzed the relationship between the length estimate and the setting of the dividers, finding remarkable predictability. In the 1960s Benoît Mandelbrot's concept of fractals finally provided the theoretical framework needed to understand this result.

Figure 4.20 Lewis Fry Richardson: the formalization of scale effects

be used creatively yet rigorously if they are to support the spatial analysis superstructure that we wish to erect. This entails much more than technical competence with software. An understanding of the nature of spatial data allows us to use induction (reasoning from observations) and deduction (reasoning from principles and theory) alongside each other to develop effective spatial representations that are safe to use.

Questions for further study

1. Many jurisdictions tout the number of miles of shoreline in their community – for example, Ottawa County, Ohio, USA claims 107 miles of Lake Erie shoreline. What does this mean, and how could you make it more meaningful?
2. The apparatus of inference was developed by statisticians because they wanted to be able to reason from the results of experiments involving small samples to make conclusions about the results of much larger, hypothetical experiments – for example, in using samples to test the effects of drugs. Summarize the problems inherent in using this apparatus for geographic data in your own words.
3. How many definitions and uses of the word *scale* can you identify?
4. What important aspects of the nature of geographic data have not been covered in this chapter?

Further reading

Batty M. and Longley P.A. 1994 *Fractal Cities: A Geometry of Form and Function*. London: Academic Press.

Mandelbrot B.B. 1983 *The Fractal Geometry of Nature*. San Francisco: Freeman.

Quattrochi D.A. and Goodchild M.F. (eds) 1996 *Scale in Remote Sensing and GIS*. Boca Raton, Florida: Lewis Publishers.

Tate N.J. and Atkinson P.M. (eds) 2001 *Modelling Scale in Geographical Information Science*. Chichester: Wiley.

Wright D. and Bartlett D. (eds) 2000 *Marine and Coastal Geographical Information Systems*. London: Taylor and Francis.

Wright D. 2002 *Undersea with GIS*. Redlands, CA: ESRI Press.

5 *Georeferencing*

Geographic location is the element that distinguishes geographic information from all other types, so methods for specifying location on the Earth's surface are essential to the creation of useful geographic information. Humanity has developed many such techniques over the centuries, and this chapter provides a basic guide for GIS students – what you need to know about georeferencing to succeed in GIS. The first section lays out the principles of georeferencing, including the requirements that any effective system must satisfy. Subsequent sections discuss commonly used systems, starting with the ones closest to everyday human experience, including placenames and street addresses, and moving to the more accurate scientific methods that form the basis of geodesy and surveying. The final sections deal with issues that arise over conversions between georeferencing systems, with the Global Positioning System (GPS), with georeferencing of computers and cellphones, and with the concept of a gazetteer.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

By the end of this chapter you will:

- Know the requirements for an effective system of georeferencing;
- Be familiar with the problems associated with placenames, street addresses, and other systems used every day by humans;
- Know how the Earth is measured and modeled for the purposes of positioning;
- Know the basic principles of map projections, and the details of some commonly used projections;
- Understand the principles behind GPS, and some of its applications.

5.1 Introduction

Chapter 3 introduced the idea of an atomic element of geographic information: a triple of location, optionally time, and attribute. To make GIS work there must be techniques for assigning values to all three of these, in ways that are understood commonly by people who wish to communicate. Almost all the world agrees on a common calendar and time system, so there are only minor problems associated with communicating that element of the atom when it is needed (although different time zones, different names of the months in different languages, the annual switch to Summer or Daylight Saving Time, and systems such as the classical Japanese convention of dating by the year of the Emperor's reign all sometimes manage to confuse us).

Time is optional in a GIS, but location is not, so this chapter focuses on techniques for specifying location, and the problems and issues that arise. Locations are the basis for many of the benefits of GIS: the ability to map, to tie different kinds of information together because they refer to the same place, or to measure distances and areas. Without locations, data are said to be *non-spatial* or *aspatial* and would have no value at all within a geographic information system.

Time is an optional element in geographic information, but location is essential.

Commonly, several terms are used to describe the act of assigning locations to atoms of information. We use the verbs *to georeference*, *to geolocate*, and *to geocode*, and say that facts have been *georeferenced* or *geocoded*. We talk about *tagging* records with geographic locations, or about *locating* them. The term *georeference* will be used throughout this chapter.

The primary requirements of a georeference are that it must be *unique*, so that there is only one location associated with a given georeference, and therefore no confusion about the location that is referenced; and that its meaning be *shared* among all of the people who wish to work with the information, including their geographic information systems. For example, the georeference 909 West Campus Lane, Goleta, California, USA points to a single house – there is no other house anywhere on Earth with that address – and its meaning is shared sufficiently widely to allow mail to be delivered to the address from virtually anywhere on the planet. The address may not be meaningful to everyone living in China, but it will be meaningful to a sufficient number of people within China's postal service, so a letter mailed from China to that address will likely be delivered successfully. Uniqueness and shared meaning are sufficient also to allow people to link different kinds of information based on common location: for example, a driving record that is georeferenced by street address can be linked to a record of purchasing. The negative implications of this kind of record linking for human privacy are discussed further by Mark Monmonier (see Box 5.2).

To be as useful as possible a georeference must be *persistent through time*, because it would be very confusing if georeferences changed frequently, and very expensive to update all of the records that depend on them. This can be problematic when a georeferencing system serves more than one purpose, or is used by more than one agency with different priorities. For example, a municipality may expand by incorporating more land, creating problems for mapping agencies, and for researchers who wish to study the municipality through time. Street names sometimes change, and postal agencies sometimes revise postal codes. Changes even occur in the names of cities (Saigon to Ho Chi Minh City), or in their conventional transcriptions into the Roman alphabet (Peking to Beijing).

To be most useful, georeferences should stay constant through time.

Every georeference has an associated spatial resolution (Section 3.4), equal to the size of the area that is assigned that georeference. A mailing address could be said to have a spatial resolution equal to the size of the mailbox, or perhaps to the area of the parcel of land or structure assigned that address. A US state has a spatial resolution that varies from the size of Rhode Island to that of Alaska, and many other systems of georeferencing have similarly wide-ranging spatial resolutions.

Many systems of georeferencing are unique only within an area or *domain* of the Earth's surface. For example, there are many cities with the name Springfield in the USA (18 according to a recent edition of the Rand McNally Road Atlas; similarly there are nine places called Whitchurch in the 2003 AA Road Atlas of the United Kingdom). City name is unique within

the domain of a US state, however, a property that was engineered with the advent of the postal system in the 19th century. Today there is no danger of there being two Springfields in Massachusetts, and a driver can confidently ask for directions to 'Springfield, Massachusetts' in the knowledge that there is no danger of being sent to the wrong Springfield. But people living in London, Ontario, Canada are well aware of the dangers of talking about 'London' without specifying the appropriate domain. Even in Toronto, Ontario a reference to 'London' may be misinterpreted as a reference to the older (UK) London on a different continent, rather than to the one 200 km away in the same province (Figure 5.1). Street name is unique in the USA within municipal domains, but not within larger domains such as county or state. The six digits of a UK National Grid reference repeat every 100 km, so additional letters are needed to achieve uniqueness within the national domain (see Box 5.1). Similarly there are 120 places on the Earth's surface with the same Universal Transverse Mercator coordinates (see Section 5.7.2), and a zone number and hemisphere must be added to make a reference unique in the global domain.

While some georeferences are based on simple names, others are based on various kinds of *measurements*, and are called *metric* georeferences. They include latitude and longitude and various kinds of coordinate systems, all of which are discussed in more detail below, and are essential to the making of maps and the display of mapped information in GIS. One enormous advantage of such systems is that they provide the potential for infinitely fine spatial resolution: provided we have sufficiently accurate measuring devices, and use enough decimal places, it is possible with such systems to locate information to any level of accuracy. Another advantage is that from measurements of two or more locations it is possible to compute distances, a very important requirement of georeferencing in GIS.

Figure 5.1 Placenames are not necessarily unique at the global level – there are many Londons, for example, besides the largest and most prominent one in the UK. People living in other Londons must often add additional information (e.g., London, *Ontario, Canada*) to resolve ambiguity

Metric georeferences are much more useful, because they allow maps to be made and distances to be calculated.

Other systems simply *order* locations. In most countries mailing addresses are ordered along streets, often using the odd integers for addresses on one side and the even integers for addresses on the other. This means that it is possible to say that 3000 State Street and 100 State Street are further apart than 200 State Street and 100 State Street, and allows postal services to sort mail for easy

Technical Box **5.1**

A national system of georeferencing: the National Grid of Great Britain

The National Grid is administered by the Ordnance Survey of Great Britain, and provides a unique georeference for every point in England, Scotland, and Wales. The first designating letter defines a 500 km square, and the second defines a 100 km square (see Figure 5.2). Within each square, two measurements, called easting and northing, define a location with respect to the lower left corner of the square. The number of digits defines the precision – three digits for easting and three for northing (a total of six) define location to the nearest 100 m.

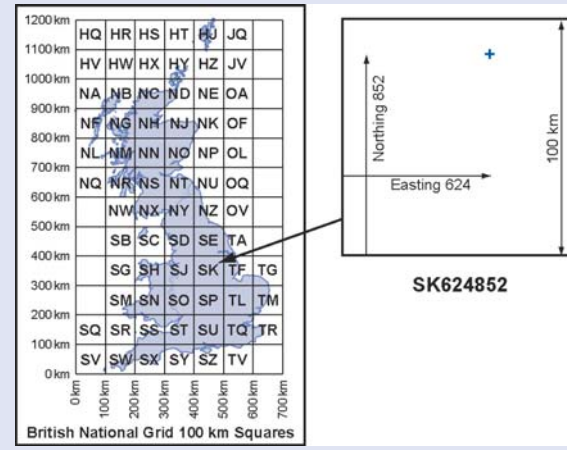


Figure 5.2 The National Grid of Great Britain, illustrating how a point is assigned a grid reference that locates it uniquely to the nearest 100 m (Reproduced by permission of Peter H. Dana)

Table 5.1 Some commonly used systems of georeferencing

System	Domain of uniqueness	Metric?	Example	Spatial resolution
Placename	varies	no	London, Ontario, Canada	varies by feature type
Postal address	global	no, but ordered along streets in most countries	909 West Campus Lane, Goleta, California, USA	size of one mailbox
Postal code	country	no	93117 (US ZIP code); WC1E 6BT (UK unit postcode)	area occupied by a defined number of mailboxes
Telephone calling area	country	no	805	varies
Cadastral system	local authority	no	Parcel 01452954, City of Springfield, Mass, USA	area occupied by a single parcel of land
Public Land Survey System	Western USA only, unique to Prime Meridian	yes	Sec 5, Township 4N, Range 6E	defined by level of subdivision
Latitude/longitude	global	yes	119 degrees 45 minutes West, 34 degrees 40 minutes North	infinitely fine
Universal Transverse Mercator	zones six degrees of longitude wide, and N or S hemisphere	yes	563146E, 4356732N	infinitely fine
State Plane Coordinates	USA only, unique to state and to zone within state	yes	55086.34E, 75210.76N	infinitely fine

delivery. In the western United States, it is often possible to infer estimates of the distance between two addresses on the same street by knowing that 100 addresses are assigned to each city block, and that blocks are typically between 120 m and 160 m long.

This section has reviewed some of the general properties of georeferencing systems, and Table 5.1 shows some commonly used systems. The following sections discuss the specific properties of the systems that are most important in GIS applications.

5.2 Placenames

Giving names to places is the simplest form of georeferencing, and was most likely the one first developed by early hunter-gatherer societies. Any distinctive feature on the landscape, such as a particularly old tree, can serve as a point of reference for two people who wish to share information, such as the existence of good game in the tree's vicinity. Human landscapes rapidly became littered with names, as people sought distinguishing labels to use in describing aspects of their surroundings, and other people adopted them. Today, of course, we have a complex system of naming oceans, continents, cities, mountains, rivers, and other prominent features. Each country maintains a system of authorized naming, often through national or state committees assigned with the task of standardizing geographic names. Nevertheless multiple names are often attached to the same feature, for example when cultures try to preserve the names given to features by the original or local inhabitants (for example, Mt Everest to many, but Chomolungma to many Tibetans), or when city names are different in different languages (Florence in English, Firenze in Italian)

Many commonly used placenames have meanings that vary between people, and with the context in which they are used.

Language extends the power of placenames through words such as 'between', which serve to refine references to location, or 'near', which serve to broaden them. 'Where State Street crosses Mission Creek' is an instance of combining two placenames to achieve greater refinement of location than either name could achieve individually. Even more powerful extensions come from combining placenames with directions and distances, as in '200 m north of the old tree' or '50 km west of Springfield'.

But placenames are of limited use as georeferences. First, they often have very coarse spatial resolution. 'Asia' covers over 43 million sq km, so the information that something is located 'in Asia' is not very helpful in pinning down its location. Even Rhode Island, the smallest state of the USA, has a land area of over 2700 sq km. Second, only certain placenames are officially authorized by national or subnational agencies. Many more are recognized only locally, so their use is limited to communication between people in the local community. Placenames may even be lost through time: although there are many contenders, we do not know with certainty where the 'Camelot' described in the English legends of King Arthur was located, if indeed it ever existed.

The meaning of certain placenames can become lost through time.

5.3 Postal addresses and postal codes

Postal addresses were introduced after the development of mail delivery in the 19th century. They rely on several assumptions:

- Every dwelling and office is a potential destination for mail;
- Dwellings and offices are arrayed along paths, roads, or streets, and numbered accordingly;
- Paths, roads, and streets have names that are unique within local areas;
- Local areas have names that are unique within larger regions; and
- Regions have names that are unique within countries.

If the assumptions are true, then mail address provides a unique identification for every dwelling on Earth.

Today, postal addresses are an almost universal means of locating many kinds of human activity: delivery of mail, place of residence, or place of business. They fail, of course, in locating anything that is not a potential destination for mail, including almost all kinds of natural features (Mt Everest does not have a postal address, and neither does Manzana Creek in Los Padres National Forest in California, USA). They are not as useful when dwellings are not numbered consecutively along streets, as happens in some cultures (notably in Japan, where street numbering can reflect date of construction, not sequence along the street – it is temporal, rather than spatial) and in large building complexes like condominiums. Many GIS applications rely on the ability to locate activities by postal address, and to convert addresses to some more universal system of georeferencing, such as latitude and longitude, for mapping and analysis.

Postal addresses work well to georeference dwellings and offices, but not natural features.

Postal codes were introduced in many countries in the late 20th century in order to simplify the sorting of mail. In the Canadian system, for example, the first three characters of the six-character code identify a Forward Sortation Area, and mail is initially sorted so that all mail directed to a single FSA is together. Each FSA's incoming mail is accumulated in a local sorting station, and sorted a second time by the last three characters of the code, to allow it to be delivered easily. Figure 5.3 shows a map of the FSAs for an area of the Toronto metropolitan region. The full six characters are unique to roughly ten houses, a single large business, or a single building. Much effort went into ensuring widespread adoption of the coding system by the general public and businesses, and computer programs were developed to assign codes automatically to addresses for large-volume mailers.

Postal codes have proven very useful for many purposes besides the sorting and delivery of mail.

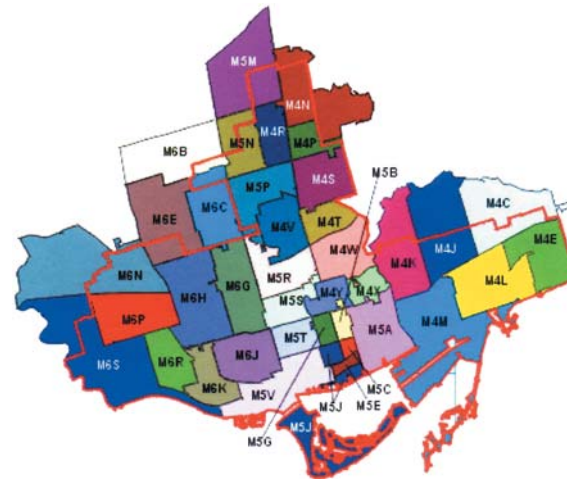


Figure 5.3 Forward Sortation Areas (FSAs) of the central part of the Toronto metropolitan region. FSAs form the first three characters of the six-character Canadian postal code

Although the area covered by a Canadian FSA or a US ZIP code varies, and can be changed whenever the postal authorities want, it is sufficiently constant to be useful for mapping purposes, and many businesses routinely make maps of their customers by counting the numbers present in each postal code area, and dividing by total population to get a picture of market penetration. Figure 5.4 shows an example of summarizing data by ZIP code. Most people know the postal code of their home, and in some instances postal codes have developed popular images (the ZIP code for Beverly Hills, California, 90210, became the title of a successful television series).

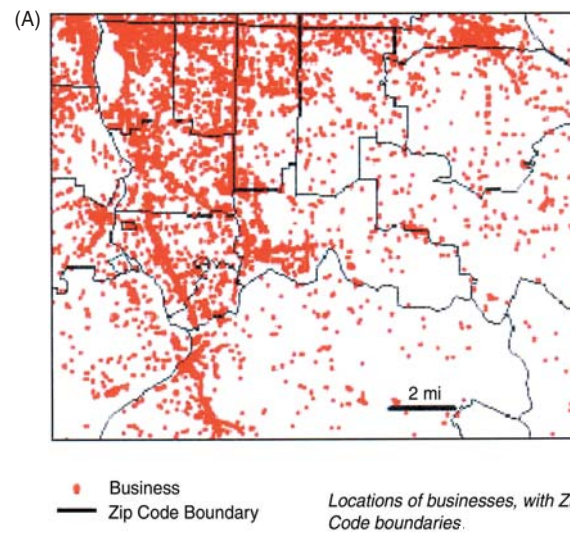


Figure 5.4 The use of ZIP codes boundaries as a convenient basis for summarizing data. (A) In this instance each business has been allocated to its ZIP code, and (B) the ZIP code areas have been shaded according to the density of businesses per square mile

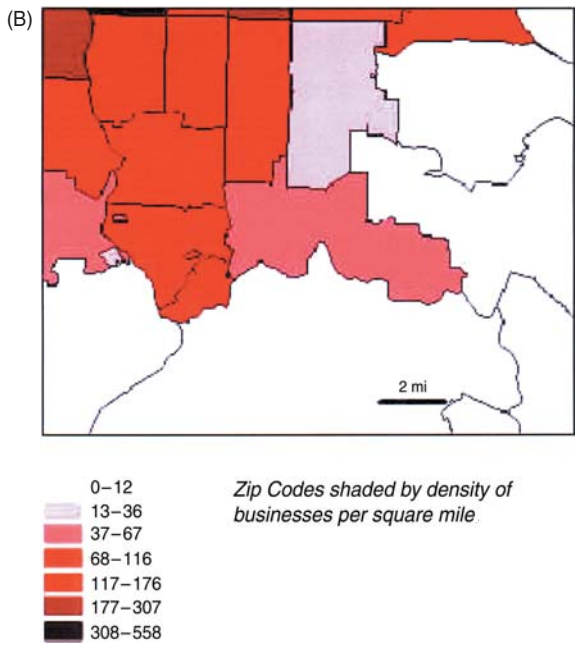


Figure 5.4 (*continued*)

5.4 Linear referencing systems

A linear referencing system identifies location on a network by measuring distance from a defined point of reference along a defined path in the network. Figure 5.5 shows an example, an accident whose location is reported as being a measured distance from a street intersection, along a named street. Linear referencing is closely related to street address, but uses an explicit measurement of distance rather than the much less reliable surrogate of street address number.

Linear referencing is widely used in applications that depend on a linear network. This includes highways (e.g., Mile 1240 of the Alaska Highway), railroads (e.g., 25.9 miles from Paddington Station in London on the main line to Bristol, England), electrical transmission, pipelines, and canals. Linear references are used by highway agencies to define the locations of bridges, signs, potholes, and accidents, and to record pavement condition.

Linear referencing systems are widely used in managing transportation infrastructure and in dealing with emergencies.

Linear referencing provides a sufficient basis for georeferencing for some applications. Highway departments often base their records of accident locations on linear references, as well as their inventories of signs and bridges (GIS has many applications in transportation that are known collectively as GIS-T, and in the developing field of intelligent transportation systems or ITS). But for other applications it is important to be able to convert between linear references and other forms, such as latitude and longitude. For example, the Onstar system that is installed in many Cadillacs sold in the USA is designed to radio the position of a vehicle automatically as soon as it is involved in an accident. When the airbags deploy, a GPS receiver determines position, which is then relayed to a central dispatch office. Emergency response centers often use street addresses and linear referencing to define the locations of accidents, so the latitude and longitude received from the vehicle must be converted before an emergency team can be sent to the accident.

Linear referencing systems are often difficult to implement in practice in ways that are robust in all situations. In an urban area with frequent intersections it is relatively easy to measure distance from the nearest one (e.g., on Birch St 87 m west of the intersection with Main St). But in rural areas it may be a long way from the nearest intersection. Even in urban areas it is not uncommon for two streets to intersect more than once (e.g., Birch may have two intersections with Columbia Crescent). There may also be difficulties in defining distance accurately, especially if roads include steep sections where the distance driven is significantly longer than the distance evaluated on a two-dimensional digital representation (Section 14.3.1).

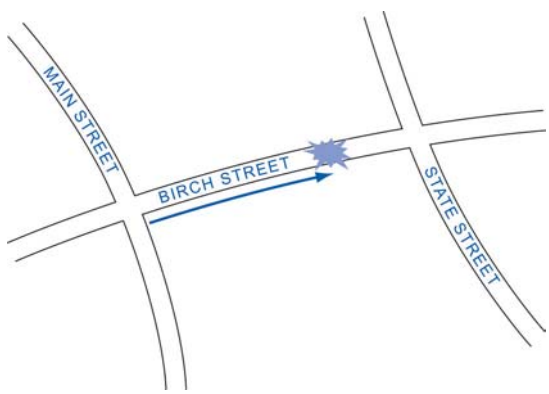


Figure 5.5 Linear referencing – an incident's position is determined by measuring its distance (87 m) along one road (Birch St) from a well-defined point (its intersection with Main St)

5.5 Cadasters and the US Public Land Survey System

The *cadaster* is defined as the map of land ownership in an area, maintained for the purposes of taxing land, or of creating a public record of ownership. The process of *subdivision* creates new parcels by legally subdividing existing ones.

Parcels of land in a cadaster are often uniquely identified, by number or by code, and are also reasonably persistent through time, and thus satisfy the requirements of a georeferencing system. But very few people know

the identification code of their home parcel, and use of the cadaster as a georeferencing system is thus limited largely to local officials, with one major exception.

The US Public Land Survey System (PLSS) evolved out of the need to survey and distribute the vast land resource of the Western USA, starting in the early 19th century, and expanded to become the dominant system of cadaster for all of the USA west of Ohio, and all of Western Canada. Its essential simplicity and regularity make it useful for many purposes, and understandable by the general public. Its geometric regularity also allows it to satisfy the requirement of a metric system of georeferencing, because each georeference is defined by measured distances.

The Public Land Survey System defines land ownership over much of western North America, and is a useful system of georeferencing.

To implement the PLSS in an area, a surveyor first lays out an accurate north–south line or *prime meridian*. Rows are then laid out six miles apart and perpendicular to this line, to become the *townships* of the system. Then blocks or *ranges* are laid out in six mile by six mile squares on either side of the prime meridian (see Figure 5.6). Each square is referenced by township number, range number, whether it is to the east or to the west, and the name of the prime meridian. Thirty-six *sections* of one mile by one mile are laid out inside each township, and numbered using a standard system (note how the numbers reverse in every other row). Each section is divided into four quarter-sections of a quarter of a square mile, or 160 acres, the size of the nominal family farm or homestead in the original conception of the PLSS. The process can be continued by subdividing into four to obtain any level of spatial resolution.

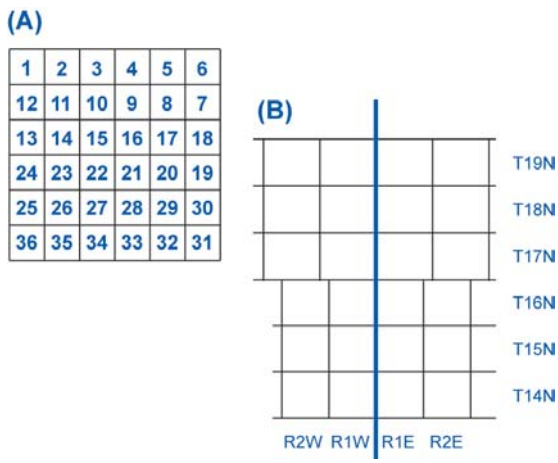


Figure 5.6 Portion of the Township and Range system (Public Lands Survey System) widely used in the western USA as the basis of land ownership (shown on the right). Townships are laid out in six-mile squares on either side of an accurately surveyed Prime Meridian. The offset shown between ranges 16 N and 17 N is needed to accommodate the Earth's curvature (shown much exaggerated). The square mile sections within each township are numbered as shown in the upper left

The PLSS would be a wonderful system if the Earth were flat. To account for its curvature the squares are not perfectly six miles by six miles, and the rows must be offset frequently; and errors in the original surveying complicate matters still further, particularly in rugged landscapes. Figure 5.6 shows the offsetting exaggerated for a small area. Nevertheless, the PLSS remains an efficient system, and one with which many people in the Western USA and Western Canada are familiar. It is often used to specify location, particularly in managing natural resources in the oil and gas industry and in mining, and in agriculture. Systems have been built to convert PLSS locations automatically to latitude and longitude.

5.6 Measuring the Earth: latitude and longitude

The most powerful systems of georeferencing are those that provide the potential for very fine spatial resolution, that allow distance to be computed between pairs of locations, and that support other forms of spatial analysis. The system of latitude and longitude is in many ways the most comprehensive, and is often called the *geographic* system of coordinates, based on the Earth's rotation about its center of mass.

To define latitude and longitude we first identify the *axis* of the Earth's rotation. The Earth's center of mass lies on the axis, and the plane through the center of mass perpendicular to the axis defines the *Equator*. Slices through the Earth parallel to the axis, and perpendicular to the plane of the Equator, define lines of constant longitude (Figure 5.7), rather like the segments of an

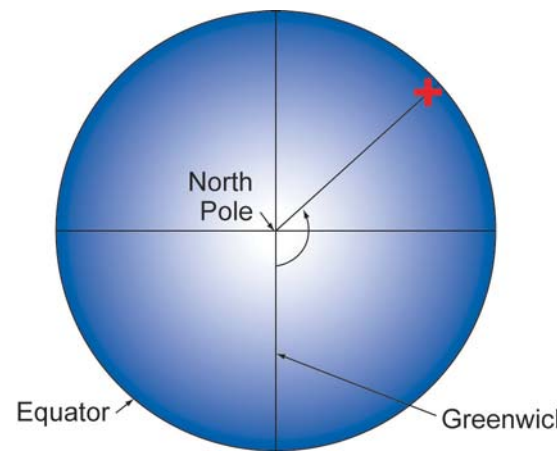


Figure 5.7 Definition of longitude. The Earth is seen here from above the North Pole, looking along the Axis, with the Equator forming the outer circle. The location of Greenwich defines the Prime Meridian. The longitude of the point at the center of the red cross is determined by drawing a plane through it and the axis, and measuring the angle between this plane and the Prime Meridian

orange. A slice through a line marked on the ground at the Royal Observatory in Greenwich, England defines zero longitude, and the angle between this slice and any other slice defines the latter's measure of longitude. Each of the 360 degrees of longitude is divided into 60 minutes and each minute into 60 seconds. But it is more conventional to refer to longitude by degrees East or West, so longitude ranges from 180 degrees West to 180 degrees East. Finally, because computers are designed to handle numbers ranging from very large and negative to very large and positive, we normally store longitude in computers as if West were negative and East were positive; and we store parts of degrees using decimals rather than minutes and seconds. A line of constant longitude is termed a *meridian*.

Longitude can be defined in this way for any rotating solid, no matter what its shape, because the axis of rotation and the center of mass are always defined. But the definition of latitude requires that we know something about the shape. The Earth is a complex shape that is only approximately spherical. A much better approximation or *figure of the Earth* is the *ellipsoid of rotation*, the figure formed by taking a mathematical ellipse and rotating it about its shorter axis (Figure 5.8). The term *spheroid* is also commonly used.

The difference between the ellipsoid and the sphere is measured by its *flattening*, or the reduction in the minor axis relative to the major axis. Flattening is defined as:

$$f = (a - b)/a$$

where a and b are the lengths of the major and minor axes respectively (we usually refer to the *semi*-axes, or half the lengths of the axes, because these are comparable to radii). The actual flattening is about 1 part in 300.

The Earth is slightly flattened, such that the distance between the Poles is about 1 part in 300 less than the diameter at the Equator.

Much effort was expended over the past 200 years in finding ellipsoids that best approximated the shape of the Earth in particular countries, so that national mapping agencies could measure position and produce accurate maps. Early ellipsoids varied significantly in their basic parameters, and were generally not centered on the Earth's center of mass. But the development of intercontinental ballistic missiles in the 1950s and the need to target them accurately, as well as new data available from satellites, drove the push to a single international standard. Without a single standard, the maps produced by different countries using different ellipsoids could never be made to fit together along their edges, and artificial steps and offsets were often necessary in moving from one country to another (navigation systems in aircraft would have to be corrected, for example).

The ellipsoid known as WGS84 (the World Geodetic System of 1984) is now widely accepted, and North American mapping is being brought into conformity with it through the adoption of the virtually identical North American Datum of 1983 (NAD83). It specifies a semi-major axis (distance from the center to the Equator) of 6378137 m, and a flattening of 1 part in 298.257. But many other ellipsoids remain in use in other parts of the world, and many older data still adhere to earlier standards, such as the North American Datum of 1927 (NAD27). Thus GIS users sometimes need to convert between datums, and functions to do that are commonly available.

We can now define latitude. Figure 5.9 shows a line drawn through a point of interest perpendicular to the ellipsoid at that location. The angle made by this line with the plane of the Equator is defined as the point's latitude, and varies from 90 South to 90 North. Again, south latitudes are usually stored as negative numbers and north latitudes as positive. Latitude is often symbolized by the Greek letter phi (ϕ) and longitude by the Greek letter lambda (λ), so the respective ranges can be expressed in mathematical shorthand as: $-180 \leq \lambda \leq 180$; $-90 \leq \phi \leq 90$. A line of constant latitude is termed a *parallel*.

It is important to have a sense of what latitude and longitude mean in terms of distances on the surface. Ignoring the flattening, two points on the same north–south line of longitude and separated by one degree

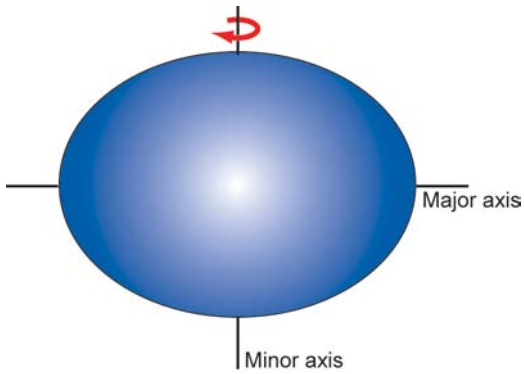


Figure 5.8 Definition of the ellipsoid, formed by rotating an ellipse about its minor axis (corresponding to the axis of the Earth's rotation)

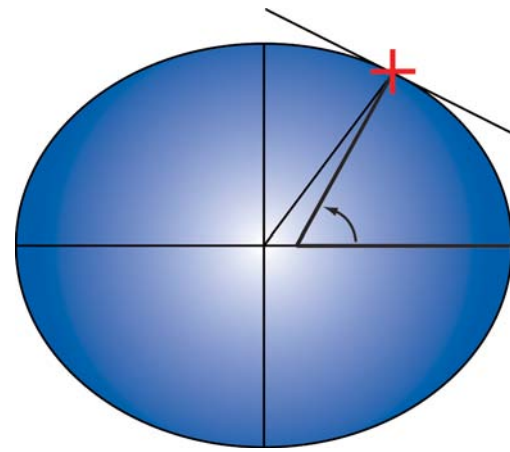


Figure 5.9 Definition of the latitude of the point marked with the red cross, as the angle between the Equator and a line drawn perpendicular to the ellipsoid

of latitude are 1/360 of the circumference of the Earth, or about 111 km, apart. One minute of latitude corresponds to 1.86 km, and also defines one nautical mile, a unit of distance that is still commonly used in navigation. One second of latitude corresponds to about 30 m. But things are more complicated in the east–west direction, and these figures only apply to east–west distances along the Equator, where lines of longitude are furthest apart. Away from the Equator the length of a line of latitude gets shorter and shorter, until it vanishes altogether at the poles. The degree of shortening is approximately equal to the cosine of latitude, or $\cos\phi$, which is 0.866 at 30 degrees North or South, 0.707 at 45 degrees, and 0.500 at 60 degrees. So a degree of longitude is only 55 km along the northern boundary of the Canadian province of Alberta (exactly 60 degrees North).

Lines of latitude and longitude are equally far apart only at the Equator; towards the Poles lines of longitude converge.

Given latitude and longitude it is possible to determine distance between any pair of points, not just pairs along lines of longitude or latitude. It is easiest to pretend for a moment that the Earth is spherical, because the flattening of the ellipsoid makes the equations much more complex. On a spherical Earth the shortest path between two points is a *great circle*, or the arc formed if the Earth is sliced through the two points and through its center (Figure 5.10; an off-center slice creates a *small circle*). The length of this arc on a spherical Earth of radius R is given by:

$$R \arccos[\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2)]$$

where the subscripts denote the two points (and see the discussion of Measurement in Section 14.3). For example, the distance from a point on the Equator at longitude 90 East (in the Indian Ocean between Sri Lanka and the Indonesian island of Sumatra) and the North Pole is found by evaluating the equation for $\phi_1 = 0$, $\lambda_1 = 90$, $\phi_2 = 90$, $\lambda_2 = 90$. It is best to work in radians (1 radian is 57.30 degrees, and 90 degrees is $\pi/2$ radians). The equation evaluates to R arccos 0, or $R\ \pi/2$, or one quarter of the circumference of the Earth. Using a radius of 6378 km this comes to 10 018 km, or close to 10 000 km (not surprising, since the French originally defined the meter in the late 18th century as one ten millionth of the distance from the Equator to the Pole).

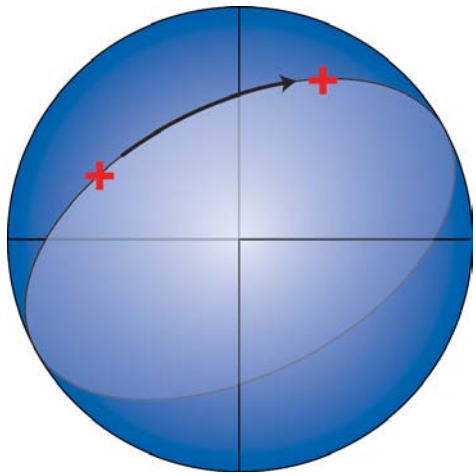


Figure 5.10 The shortest distance between two points on the sphere is an arc of a great circle, defined by slicing the sphere through the two points and the center (all lines of longitude, and the Equator, are great circles). The circle formed by a slice that does not pass through the center is a small circle (all lines of latitude except the Equator are small circles)

5.7 Projections and coordinates

Latitude and longitude define location on the Earth's surface in terms of angles with respect to well-defined references: the Royal Observatory at Greenwich, the center of mass, and the axis of rotation. As such, they constitute the most comprehensive system of georeferencing, and support a range of forms of analysis, including the calculation of distance between points, on the curved surface of the Earth. But many technologies for working with geographic data are inherently flat, including paper and printing, which evolved over many centuries long before the advent of digital geographic data and GIS. For various reasons, therefore, much work in GIS deals with a flattened or *projected* Earth, despite the price we pay in the distortions that are an inevitable consequence of flattening. Specifically, the Earth is often flattened because:

- paper is flat, and paper is still used as a medium for inputting data to GIS by scanning or digitizing (see Chapter 9), and for outputting data in map or image form;
- rasters are inherently flat, since it is impossible to cover a curved surface with equal squares without gaps or overlaps;
- photographic film is flat, and film cameras are still used widely to take images of the Earth from aircraft to use in GIS;
- when the Earth is seen from space, the part in the center of the image has the most detail, and detail drops off rapidly, the back of the Earth being invisible; in order to see the whole Earth with approximately equal detail it must be distorted in some way, and it is most convenient to make it flat.

The Cartesian coordinate system (Figure 5.11) assigns two coordinates to every point on a flat surface, by measuring distances from an origin parallel to two axes drawn at right angles. We often talk of the two axes as x and y, and of the associated coordinates as the x and y coordinate, respectively. Because it is common to align the y axis with North in geographic applications, the coordinates of a projection on a flat sheet are often termed *easting* and *northing*.

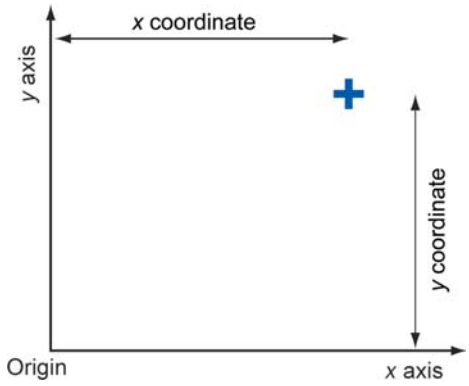


Figure 5.11 A Cartesian coordinate system, defining the location of the blue cross in terms of two measured distances from the Origin, parallel to the two axes

Although projections are not absolutely required, there are several good reasons for using them in GIS to flatten the Earth.

One way to think of a map projection, therefore, is that it transforms a position on the Earth's surface identified by latitude and longitude (ϕ, λ) into a position in Cartesian coordinates (x, y). Every recognized map projection, of which there are many, can be represented as a pair of mathematical functions:

$$x = f(\phi, \lambda)$$
$$y = g(\phi, \lambda)$$

For example, the famous Mercator projection uses the functions:

$$x = \lambda$$
$$y = \ln \tan[\phi/2 + \pi/4]$$

where ln is the natural log function. The inverse transformations that map Cartesian coordinates back to latitude and longitude are also expressible as mathematical functions: in the Mercator case they are:

$$\lambda = x$$
$$\phi = 2 \arctan e^{y} - \pi/2$$

where e denotes the constant 2.71828. Many of these functions have been implemented in GIS, allowing users to work with virtually any recognized projection and datum, and to convert easily between them.

Two datasets can differ in both the projection and the datum, so it is important to know both for every data set.

Projections necessarily distort the Earth, so it is impossible in principle for the scale (distance on the map compared to distance on the Earth, for a discussion of scale see Box 4.2) of any flat map to be perfectly uniform, or for the pixel size of any raster to be perfectly constant. But projections can preserve certain properties, and two such properties are particularly important, although any projection can achieve at most one of them, not both:

- the *conformal* property, which ensures that the shapes of small features on the Earth's surface are preserved on the projection: in other words, that the scales of the projection in the x and y directions are always equal;
- the *equal area* property, which ensures that areas measured on the map are always in the same proportion to areas measured on the Earth's surface.

The conformal property is useful for navigation, because a straight line drawn on the map has a constant bearing (the technical term for such a line is a *loxodrome*). The equal area property is useful for various kinds of analysis involving areas, such as the computation of the area of someone's property.

Besides their distortion properties, another common way to classify map projections is by analogy to a physical model of how positions on the map's flat surface are related to positions on the curved Earth. There are three major classes (Figure 5.12):

- *cylindrical* projections, which are analogous to wrapping a cylinder of paper around the Earth, projecting the Earth's features onto it, and then unwrapping the cylinder;
- *azimuthal* or *planar* projections, which are analogous to touching the Earth with a sheet of flat paper; and
- *conic* projections, which are analogous to wrapping a sheet of paper around the Earth in a cone.

In each case, the projection's *aspect* defines the specific relationship, e.g., whether the paper is wrapped around the Equator, or touches at a pole. Where the paper coincides with the surface the scale of the projection is 1, and where the paper is some distance outside the surface the projected feature will be larger than it is on the Earth. *Secant* projections attempt to minimize distortion by allowing the paper to cut through the surface, so that scale can be both greater and less than 1 (Figure 5.12; projections for which the paper touches the Earth and in which scale is always 1 or greater are called *tangent*).

All three types can have either conformal or equal area properties, but of course not both. Figure 5.13 shows examples of several common projections, and shows how the lines of latitude and longitude map onto the projection, in a (distorted) grid known as a *graticule*.

The next sections describe several particularly important projections in detail, and the coordinate systems that they produce. Each is important to GIS, and users are likely to come across them frequently. The map projection (and datum) used to make a dataset is sometimes not known to the user of the dataset, so it is helpful to know enough about map projections and coordinate systems to make intelligent guesses when trying to combine such a dataset with other data. Several excellent books on map projections are listed in the References.

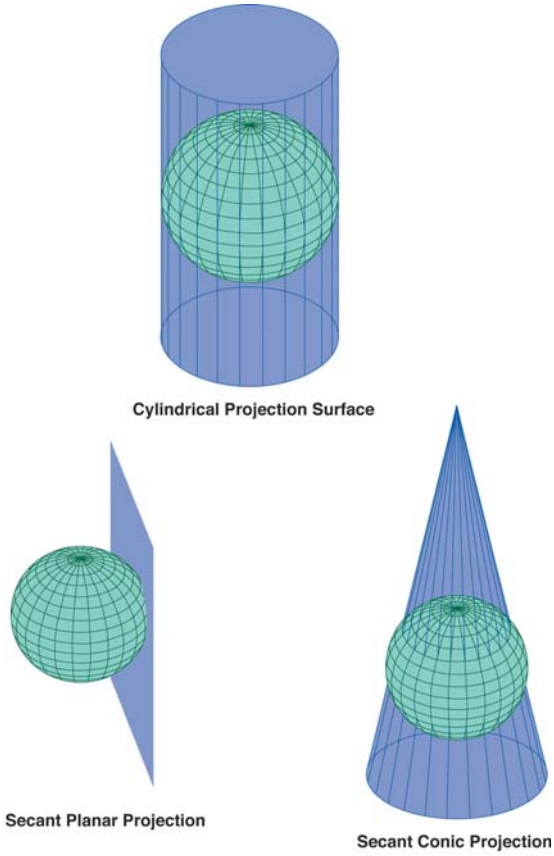


Figure 5.12 The basis for three types of map projections – cylindrical, planar, and conic. In each case a sheet of paper is wrapped around the Earth, and positions of objects on the Earth's surface are projected onto the paper. The cylindrical projection is shown in the *tangent* case, with the paper touching the surface, but the planar and conic projections are shown in the *secant* case, where the paper cuts into the surface (Reproduced by permission of Peter H. Dana)

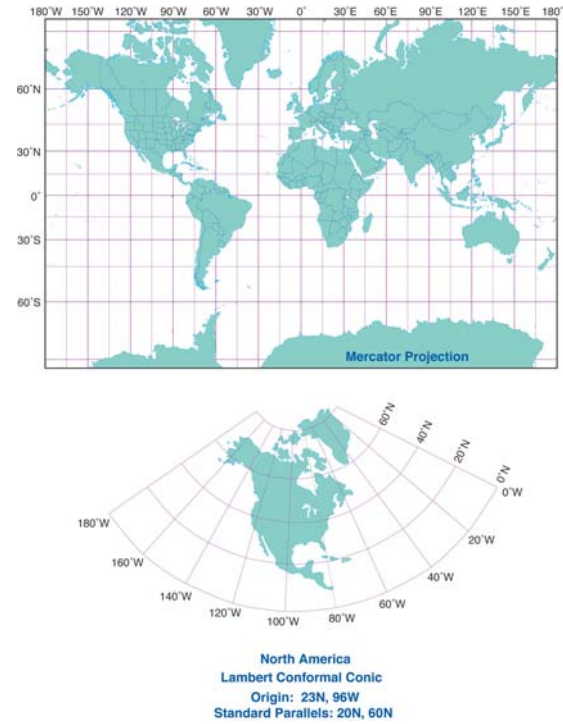


Figure 5.13 Examples of some common map projections. The Mercator projection is a tangent cylindrical type, shown here in its familiar Equatorial aspect (cylinder wrapped around the Equator). The Lambert Conformal Conic projection is a secant conic type. In this instance the cone onto which the surface was projected intersected the Earth along two lines of latitude: 20 North and 60 North (Reproduced by permission of Peter H. Dana)

5.7.1 The Plate Carrée or Cylindrical Equidistant projection

The simplest of all projections simply maps longitude as x and latitude as y, and for that reason is also known informally as the *unprojected* projection. The result is a heavily distorted image of the Earth, with the poles smeared along the entire top and bottom edges of the map, and a very strangely shaped Antarctica. Nevertheless, it is the view that we most often see when images are created of the entire Earth from satellite data (for example in illustrations of sea surface temperature that show the El Niño or La Niña effects). The projection is not conformal (small shapes are distorted) and not equal area, though it does maintain the correct distance between every point and the Equator. It is normally used only for the whole Earth, and maps of large parts of the Earth, such as the USA or Canada, look distinctly odd in this projection. Figure 5.14 shows the projection applied to the world, and also shows a comparison of three familiar projections of the United States: the Plate Carrée, Mercator, and Lambert Conformal Conic.

When longitude is assigned to *x* and latitude to *y* a very odd-looking Earth results.

Serious problems can occur when doing analysis using this projection. Moreover, since most methods of analysis in GIS are designed to work with Cartesian coordinates rather than latitude and longitude, the same problems can arise in analysis when a dataset uses latitude and longitude, or so-called geographic coordinates. For example, a command to generate a circle of radius one unit in this projection will create a figure that is two degrees of latitude across in the north–south direction, and two degrees of longitude across in the east–west direction. On the Earth's surface this figure is not a circle at all, and at high latitudes it is a very squashed ellipse. What happens if you ask your favorite GIS to generate a circle and add it to a dataset that is in geographic coordinates? Does it recognize that you are using geographic coordinates and automatically compensate for the differences in distances east–west and north–south away from the Equator, or does it in effect operate on a Plate Carrée projection and create a figure that is an ellipse on the Earth's surface? If you ask it to compute distance between two points defined by

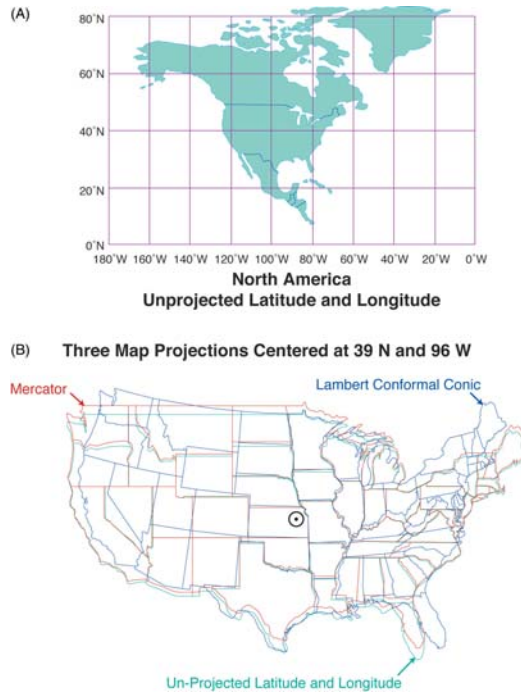


Figure 5.14 (A) The so-called *unprojected* or Plate Carrée projection, a tangent cylindrical projection formed by using longitude as x and latitude as y. (B) A comparison of three familiar projections of the USA. The Lambert Conformal Conic is the one most often encountered when the USA is projected alone, and is the only one of the three to curve the parallels of latitude, including the northern border on the 49th Parallel (Reproduced by permission of Peter H. Dana)

latitude and longitude, does it use the true shortest (great circle) distance based on the equation in Section 5.6, or the formula for distance in a Cartesian coordinate system on a distorted plane?

It is wise to be careful when using a GIS to analyze data in latitude and longitude rather than in projected coordinates, because serious distortions of distance, area, and other properties may result.

5.7.2 The Universal Transverse Mercator projection

The UTM system is often found in military applications, and in datasets with global or national coverage. It is based on the Mercator projection, but in *transverse* rather than Equatorial aspect, meaning that the projection is analogous to wrapping a cylinder around the Poles, rather than around the Equator. There are 60 zones in the system, and each zone corresponds to a half cylinder wrapped along a particular line of longitude, each zone being 6 degrees wide. Thus Zone 1 applies to longitudes from 180 W to 174 W, with the half cylinder wrapped along 177 W; Zone 10 applies to longitudes from 126 W to 120 W centered on 123 W, etc. (Figure 5.15).

The UTM system is secant, with lines of scale 1 located some distance out on both sides of the central meridian. The projection is conformal, so small features appear with the correct shape and scale is the same in all directions. Scale is 0.9996 at the central meridian and at most 1.0004 at the edges of the zone. Both parallels

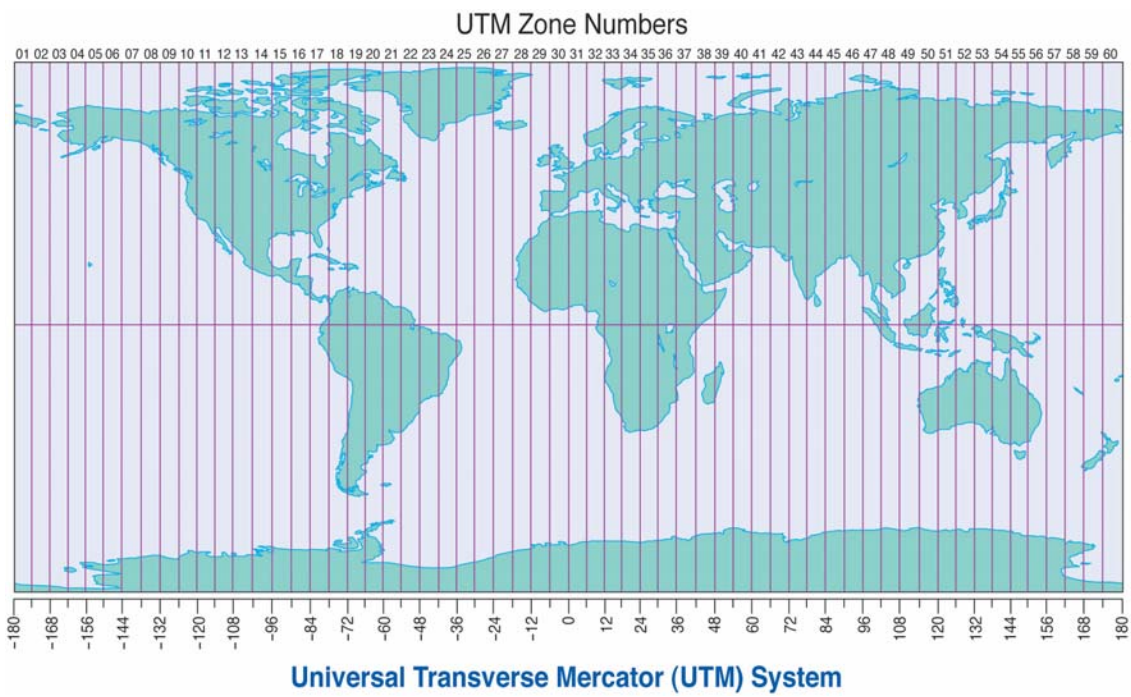


Figure 5.15 The system of zones of the Universal Transverse Mercator system. The zones are identified at the top. Each zone is six degrees of longitude in width (Reproduced by permission of Peter H. Dana)

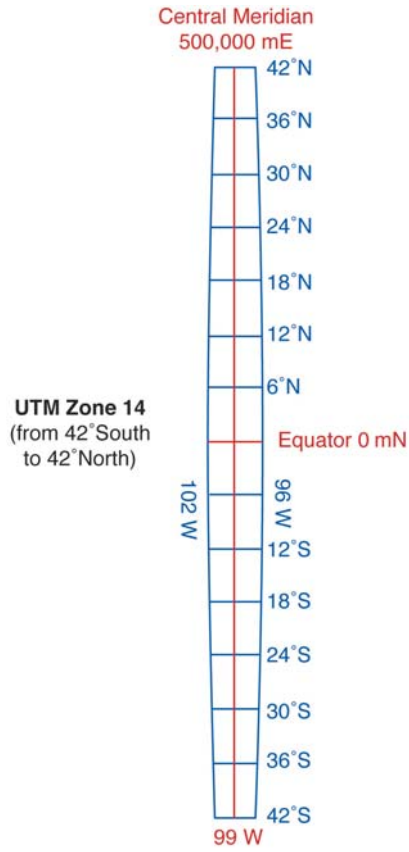


Figure 5.16 Major features of UTM Zone 14 (from 102 W to 96 W). The central meridian is at 99 W. Scale factors vary from 0.9996 at the central meridian to 1.0004 at the zone boundaries. See text for details of the coordinate system (Reproduced by permission of Peter H. Dana)

and meridians are curved on the projection, with the exception of the zone's central meridian and the Equator. Figure 5.16 shows the major features of one zone.

The coordinates of a UTM zone are defined in meters, and set up such that the central meridian's easting is always 500 000 m (a *false* easting), so easting varies from near zero to near 1 000 000 m. In the Northern Hemisphere the Equator is the origin of northing, so a point at northing 5 000 000 m is approximately 5000 km from the Equator. In the Southern Hemisphere the Equator is given a false northing of 10 000 000 m and all other northings are less than this.

UTM coordinates are in meters, making it easy to make accurate calculations of short distances between points.

Because there are effectively 60 different projections in the UTM system, maps will not fit together across a zone boundary. Zones become so much of a problem at high latitudes that the UTM system is normally replaced with azimuthal projections centered on each Pole (known as the UPS or Universal Polar Stereographic system) above 80 degrees latitude. The problem is especially critical for cities that cross zone boundaries, such as Calgary, Alberta, Canada (crosses the boundary at 114 W between Zone 11 and Zone 12). In such situations one zone can be extended to cover the entire city, but this results in distortions that are larger than normal. Another option is to define a special zone, with its own central meridian selected to pass directly through the city's center. Italy is split between Zones 32 and 33, and many Italian maps carry both sets of eastings and northings.

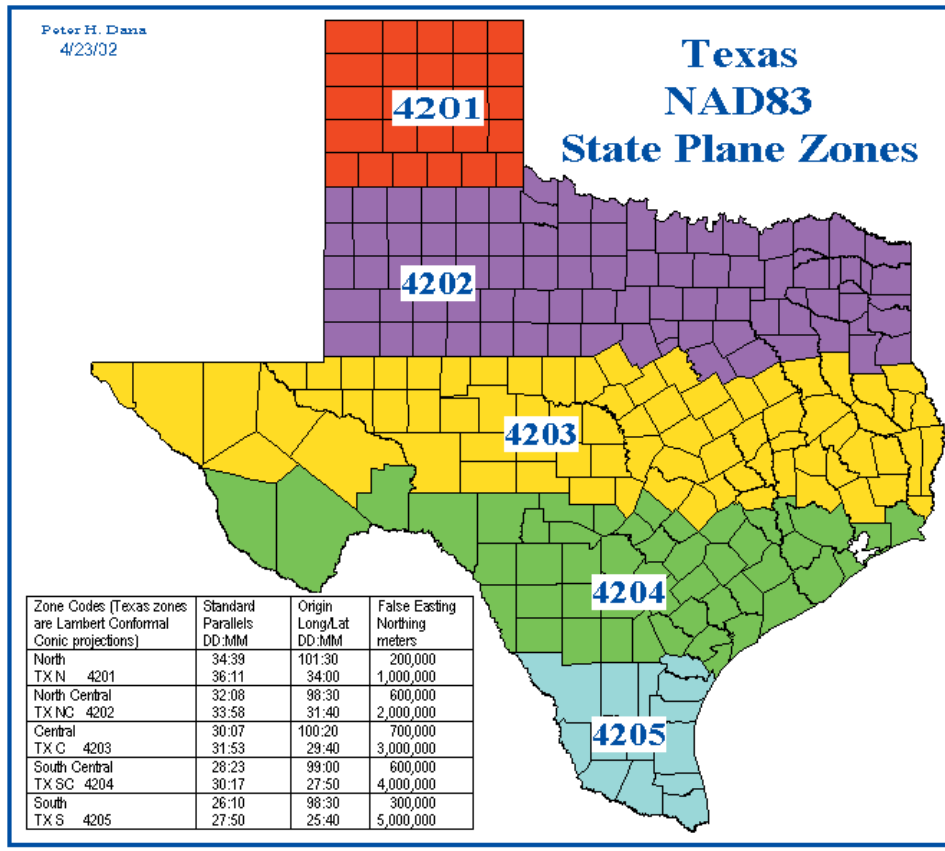
UTM coordinates are easy to recognize, because they commonly consist of a six-digit integer followed by a seven-digit integer (and decimal places if precision is greater than a meter), and sometimes include zone numbers and hemisphere codes. They are an excellent basis for analysis, because distances can be calculated from them for points within the same zone with no more than 0.04% error. But they are complicated enough that their use is effectively limited to professionals (the so-called 'spatially aware professionals' or SAPs defined in Section 1.4.3.2) except in applications where they can be hidden from the user. UTM grids are marked on many topographic maps, and many countries project their topographic maps using UTM, so it is easy to obtain UTM coordinates from maps for input to digital datasets, either by hand or automatically using scanning or digitizing (Chapter 9).

5.7.3 State Plane Coordinates and other local systems

Although the distortions of the UTM system are small, they are nevertheless too great for some purposes, particularly in accurate surveying. Zone boundaries also are a problem in many applications, because they follow arbitrary lines of longitude rather than boundaries between jurisdictions. In the 1930s each US state agreed to adopt its own projection and coordinate system, generally known as State Plane Coordinates (SPC), in order to support these high-accuracy applications. Projections were chosen to minimize distortion over the area of the state, so choices were often based on the state's shape. Some large states decided that distortions were still too great, and designed their SPCs with internal zones (for example, Texas has five zones based on the Lambert Conformal Conic projection, Figure 5.17, while Hawaii has five zones based on the Transverse Mercator projection). Many GIS have details of SPCs already stored, so it is easy to transform between them and UTM, or latitude and longitude. The system was revised in 1983 to accommodate the shift to the new North American Datum (NAD83).

All US states have adopted their own specialized coordinate systems for applications such as surveying that require very high accuracy.

Many other countries have adopted coordinate systems of their own. For example, the UK uses a single projection and coordinate system known as the National Grid that is based on the Oblique Mercator projection (see Box 5.1)



Zone Codes (Texas zones are Lambert Conformal Conic projections)	Standard Parallels DD:MM	Origin Long/Lat DD:MM	False Easting Northing meters
North TX N 4201	34:39 36:11	101:30 34:00	200,000 1,000,000
North Central TX NC 4202	32:08 33:58	98:30 31:40	600,000 2,000,000
Central TX C 4203	30:07 31:53	100:20 29:40	700,000 3,000,000
South Central TX SC 4204	28:23 30:17	99:00 27:50	600,000 4,000,000
South TX S 4205	26:10 27:50	98:30 25:40	300,000 5,000,000

Figure 5.17 The five State Plane Coordinate zones of Texas. Note that the zone boundaries are defined by counties, rather than parallels, for administrative simplicity (Reproduced by permission of Peter H. Dana)

and is marked on all topographic maps. Canada uses a uniform coordinate system based on the Lambert Conformal Conic projection, which has properties that are useful at mid to high latitudes, for applications where the multiple zones of the UTM system would be problematic.

5.8 Measuring latitude, longitude, and elevation: GPS

The Global Positioning System and its analogs (GLONASS in Russia, and the proposed Galileo system in Europe) have revolutionized the measurement of position, for the first time making it possible for people to know almost exactly where they are anywhere on the surface of the Earth. Previously, positions had to be established by a complex system of relative and absolute measurements. If one was near a point whose position was accurately known (a survey *monument*, for example), then position could be established through a series of accurate measurements of distances and directions starting from the monument. But if no monuments existed, then position had to be established through absolute measurements. Latitude is comparatively easy to measure, based on the elevation of the sun at its highest point (local noon), or on the locations of the sun, moon, or fixed stars at precisely known times. But longitude requires an accurate method of measuring time, and the lack of accurate clocks led to massively incorrect beliefs about positions during early navigation. For example, Columbus and his contemporary explorers had no means of measuring longitude, and believed that the Earth was much smaller than it is, and that Asia was roughly as far west of Europe as the width of the Atlantic. The strength of this conviction is still reflected in the term we use for the islands of the Caribbean (the West Indies) and the first rapids on the St Lawrence in Canada (Lachine, or China). The fascinating story of the measurement of longitude is recounted by Dava Sobel.

The GPS consists of a system of 24 satellites (plus some spares), each orbiting the Earth every 12 hours on distinct orbits at a height of 20 200 km and transmitting radio pulses at very precisely timed intervals. To determine position, a receiver must make precise calculations from the signals, the known positions of the satellites, and the velocity of light. Positioning in three dimensions (latitude, longitude, and elevation) requires that at least four satellites are above the horizon, and accuracy depends on the number of such satellites and their positions (if elevation is not needed then only three

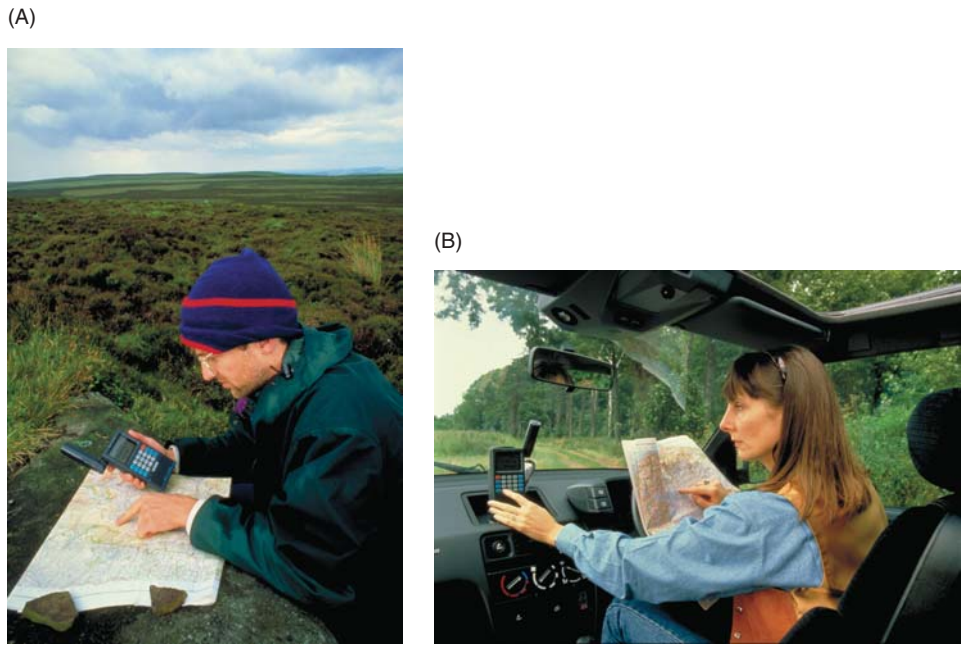


Figure 5.18 A simple GPS can provide an essential aid to wayfinding when (A) hiking or (B) driving (Reproduced by permission of David Parker, SPL, Photo Researchers)

satellites need be above the horizon). Several different versions of GPS exist, with distinct accuracies.

A simple GPS, such as one might buy in an electronics store for $100, or install as an optional addition to a laptop, cellphone, PDA (personal digital assistant, such as a Palm Pilot or iPAQ), or vehicle (Figure 5.18), has an accuracy within 10 m. This accuracy will degrade in cities with tall buildings, or under trees, and GPS signals will be lost entirely under bridges or indoors. Differential GPS (DGPS) combines GPS signals from satellites with correction signals received via radio or telephone from base stations. Networks of such stations now exist, at precisely known locations, constantly broadcasting corrections; corrections are computed by comparing each known location to its apparent location determined from GPS. With DGPS correction, accuracies improve to 1 m or better. Even greater accuracies are possible using various sophisticated techniques, or by remaining fixed and averaging measured locations over several hours.

GPS is very useful for recording ground control points when building GIS databases, for locating objects that move (for example, combine harvesters, tanks, cars, and shipping containers), and for direct capture of the locations of many types of fixed objects, such as utility assets, buildings, geological deposits, and sample points. Other applications of GPS are discussed in Chapter 11 on Distributed GIS, and by Mark Monmonier (see Box 5.2).

Some care is needed in using GPS to measure elevation. First, accuracies are typically lower, and a position determined to 10 m in the horizontal may be no better than plus or minus 50 m in the vertical. Second, a variety of reference elevations or vertical datums are in common use in different parts of the world and by different agencies – for example, in the USA the topographic and hydrographic definitions of the vertical datum are significantly different.

5.9 Converting georeferences

GIS are particularly powerful tools for converting between projections and coordinate systems, because these transformations can be expressed as numerical operations. In fact this ability was one of the most attractive features of early systems for handling digital geographic data, and drove many early applications. But other conversions, e.g., between placenames and geographic coordinates, are much more problematic. Yet they are essential operations. Almost everyone knows their mailing address, and can identify travel destinations by name, but few are able to specify these locations in coordinates, or to interact with geographic information systems on that basis. GPS technology is attractive precisely because it allows its user to determine his or her latitude and longitude, or UTM coordinates, directly at the touch of a button.

Methods of converting between georeferences are important for:

- converting lists of customer addresses to coordinates for mapping or analysis (the task known as *geocoding*; see Box 5.3);
- combining datasets that use different systems of georeferencing;

Biographical Box **5.2**

Mark Monmonier, Cartographer

Figure 5.19 Mark Monmonier, cartographer

Mark Monmonier (Figure 5.19) is Distinguished Professor of Geography in the Maxwell School of Citizenship and Public Affairs at Syracuse University. He has published numerous papers on map design, automated map analysis, cartographic generalization, the history of cartography, statistical graphics, geographic demography, and mass communications. But he is best known as author of a series of widely read books on major issues in cartography, including *How to Lie with Maps* (University of Chicago Press, 1991; 2nd edition, revised and expanded, 1996) and *Rhumb Lines and Map Wars: A Social History of the Mercator Projection* (University of Chicago Press, 2004).

Commenting on the power of GPS, he writes:

One of the more revolutionary aspects of geospatial technology is the ability to analyze maps without actually looking at one. Even more radical is the ability to track humans or animals around the clock by integrating a GPS fix with spatial data describing political boundaries or the street network. Social scientists recognize this kind of constant surveillance as a *panoptic gaze*, named for the Panopticon, a hypothetical prison devised by Jeremy Bentham, an eighteenth-century social reformer intrigued with knowledge and power. Bentham argued that inmates aware they could be watched secretly at any time by an unseen warden were easily controlled. Location tracking can achieve similar results without walls or shutters.

GPS-based tracking can be beneficial or harmful depending on your point of view. An accident victim wants the Emergency-911 dispatcher to know where to send help. A car rental firm wants to know which client violated the rental agreement by driving out of state. Parents want to know where their children are. School principals want to know when a paroled pedophile is circling the playground. And the Orwellian thought police want to know where dissidents are gathering. Few geospatial technologies are as ambiguous and potentially threatening as location tracking.

Merge GPS, GIS, and wireless telephony, and you have the location-based services (LBS) industry (Chapter 11), useful for dispatching tow trucks, helping us find restaurants or gas stations, and letting a retailer, police detective, or stalker know where we've been. Our locational history is not only marketable but potentially invasive. How lawmakers respond to growing concern about location privacy (Figure 5.20) will determine whether we control our locational history or society lets our locational history control us.

Figure 5.20 An early edition of Mark Monmonier's government identification card. Today the ability to link such records to other information about an individual's whereabouts raises significant concerns about privacy

Technical Box **5.3**

Geocoding: conversion of street addresses to coordinates

Geocoding is the name commonly given to the process of converting street addresses to latitude and longitude, or some similarly universal coordinate system. It is very widely used as it allows any database containing addresses, such as a company mailing list or a set of medical records, to be input to a GIS and mapped. Geocoding requires a database containing records representing the geometry of street segments between consecutive intersections, and the address ranges on each side of each segment (a *street centerline database*, see Chapter 9). Addresses are geocoded by finding the appropriate street segment record, and estimating a location based on linear interpolation within the address range. For example, 950 West Broadway in Columbia, Missouri, USA, lies on the side of the segment whose address range runs from 900 to 998, or $50/98 = 51.02\%$ of the distance from the start of the segment to the end. The segment starts at 92.3503 West longitude, 38.9519 North latitude, and ends at 92.3527 West, 38.9522 North. Simple arithmetic gives the address location as 92.3515 West, 38.9521 North. Four decimal places suggests an accuracy of about 10 m, but the estimate depends also on the accuracy of the assumption that addresses are uniformly spaced, and on the accuracy of the street centerline database.

- converting to projections that have desirable properties for analysis, e.g., no distortion of area;
- searching the Internet or other distributed data resources for data about specific locations;
- positioning GIS map displays by recentering them on places of interest that are known by name (these last two are sometimes called *locator* services).

The oldest method of converting georeferences is the *gazetteer*, the name commonly given to the index in an atlas that relates placenames to latitude and longitude, and to relevant pages in the atlas where information about that place can be found. In this form the gazetteer is a useful locator service, but it works only in one direction as a conversion between georeferences (from placename to latitude and longitude). Gazetteers have evolved substantially in the digital era, and it is now possible to obtain large databases of placenames and associated coordinates and to access services that allow such databases to be queried over the Internet (e.g., the Alexandria Digital Library gazetteer, **www.alexandria.ucsb.edu**; the US Geographic Names Information System, **geonames.usgs.gov**).

5.10 Summary

This chapter has looked in detail at the complex ways in which humans refer to specific locations on the planet, and how they measure locations. Any form of geographic information must involve some kind of georeference, and so it is important to understand the common methods, and their advantages and disadvantages. Many of the benefits of GIS rely on accurate georeferencing – the ability to link different items of information together through common geographic location; the ability to measure distances and areas on the Earth's surface, and to perform more complex forms of analysis; and the ability to communicate geographic information in forms that can be understood by others.

Georeferencing began in early societies, to deal with the need to describe locations. As humanity has progressed, we have found it more and more necessary to describe locations accurately, and over wider and wider domains, so that today our methods of georeferencing are able to locate phenomena unambiguously and to high accuracy anywhere on the Earth's surface. Today, with modern methods of measurement, it is possible to direct another person to a point on the other side of the Earth to an accuracy of a few centimeters, and this level of accuracy and referencing is achieved regularly in such areas as geophysics and civil engineering.

But georeferences can never be perfectly accurate, and it is always important to know something about spatial resolution. Questions of measurement accuracy are discussed at length in Chapter 6, together with techniques for representation of phenomena that are inherently fuzzy, such that it is impossible to say with certainty whether a given point is inside or outside the georeference.

Questions for further study

1. Visit your local map library, and determine: (1) the projections and datums used by selected maps; (2) the coordinates of your house in several common georeferencing systems.
2. Summarize the arguments for and against a single global figure of the Earth, such as WGS84.
3. How would you go about identifying the projection used by a common map source, such as the weather maps shown by a TV station or in a newspaper?
4. Chapter 14 discusses various forms of measurement in GIS. Review each of those methods, and the issues involved in performing analysis on databases that use different map projections. Identify the map projections that would be best for measurement of (1) area, (2) length, (3) shape.

Further reading

Bugayevskiy L.M. and Snyder J.P. 1995 *Map Projections: A Reference Manual*. London: Taylor and Francis.

Kennedy M. 1996 *The Global Positioning System and GIS: An Introduction*. Chelsea, Michigan: Ann Arbor Press.

Maling D.H. 1992 *Coordinate Systems and Map Projections* (2nd edn). Oxford: Pergamon.

Sobel D. 1995 *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. New York: Walker.

Snyder J.P. 1997 *Flattening the Earth: Two Thousand Years of Map Projections*. Chicago: University of Chicago Press.

Steede-Terry K. 2000 *Integrating GIS and the Global Positioning System*. Redlands, CA: ESRI Press.

6 *Uncertainty*

Uncertainty in geographic representation arises because, of necessity, almost all representations of the world are incomplete. As a result, data in a GIS can be subject to measurement error, out of date, excessively generalized, or just plain wrong. This chapter identifies many of the sources of geographic uncertainty and the ways in which they operate in GIS-based representations. Uncertainty arises from the way that GIS users conceive of the world, how they measure and represent it, and how they analyze their representations of it. This chapter investigates a number of conceptual issues in the creation and management of uncertainty, before reviewing the ways in which it may be measured using statistical and other methods. The propagation of uncertainty through geographical analysis is then considered. Uncertainty is an inevitable characteristic of GIS usage, and one that users must learn to live with. In these circumstances, it becomes clear that all decisions based on GIS are also subject to uncertainty.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

By the end of this chapter you will:

- Understand the concept of uncertainty, and the ways in which it arises from imperfect representation of geographic phenomena;
- Be aware of the uncertainties introduced in the three stages (conception, measurement and representation, and analysis) of database creation and use;
- Understand the concepts of vagueness and ambiguity, and the uncertainties arising from the definition of key GIS attributes;
- Understand how and why scale of geographic measurement and analysis can both create and propagate uncertainty.

6.1 Introduction

GIS-based representations of the real world are used to reconcile science with practice, concepts with applications, and analytical methods with social context. Yet, almost always, such reconciliation is imperfect, because, necessarily, representations of the world are incomplete (Section 3.4). In this chapter we will use *uncertainty* as an umbrella term to describe the problems that arise out of these imperfections. Occasionally, representations may approach perfect accuracy and precision (terms that we will define in Section 6.3.2.2) – as might be the case, for example, in the detailed site layout layer of a utility management system, in which strenuous efforts are made to reconcile fine-scale multiple measurements of built environments. Yet perfect, or nearly perfect, representations of reality are the exception rather than the rule. More usually, the inherent complexity and detail of our world makes it virtually impossible to capture every single facet, at every possible scale, in a digital representation. (Neither is this usually desirable: see the discussion of sampling in Section 4.4.) Furthermore, different individuals see the world in different ways, and in practice no single view is likely to be accepted universally as the best or to enjoy uncontested status. In this chapter we discuss how the processes and procedures of abstraction create differences between the contents of our (geographic and attribute) database and real-world phenomena. Such differences are almost inevitable and understanding of them can help us to manage uncertainty, and to live with it.

It is impossible to make a perfect representation of the world, so uncertainty about it is inevitable.

Various terms are used to describe differences between the real world and how it appears in a GIS, depending upon the context. The established scientific notion of measurement *error* focuses on differences between observers or between measuring instruments. As we saw in a previous chapter (Section 4.7), the concept of error in multivariate statistics arises in part from omission of some relevant aspects of a phenomenon – as in the failure to fully specify all of the predictor variables in a multiple regression model, for example. Similar problems arise when one or more variables are omitted from the calculation of a composite indicator – as, for example, in omitting road accessibility in an index of land value, or omitting employment status from a measure of social deprivation (see Section 16.2.1 for a discussion of indicators). More generally, the Dutch geostatistician Gerard Heuvelink (who we will introduce in Box 6.1) has defined *accuracy* as the difference between reality and *our* representation of reality. Although such differences might principally be addressed in formal mathematical terms, the use of the word *our* acknowledges the varying views that are generated by a complex, multi-scale, and inherently uncertain world.

Yet even this established framework is too simple for understanding quality or the defining standards of geographic data. The terms *ambiguity* and *vagueness* identify further considerations which need to be taken into account in assessing the *quality* of a GIS representation. Quality is an important topic in GIS, and there have been many attempts to identify its basic dimensions. The US Federal Geographic Data Committee's various standards list five components of quality: attribute accuracy, positional accuracy, logical consistency, completeness, and lineage. Definitions and other details on each of these and several more can be found on the FGDC's Web pages (**www.fgdc.gov**). Error, inaccuracy, ambiguity, and vagueness all contribute to the notion of uncertainty in the broadest sense, and uncertainty may thus be defined as a measure of the user's understanding of the difference between the contents of a dataset, and the real phenomena that the data are believed to represent. This definition implies that phenomena are real, but includes the possibility that we are unable to describe them exactly. In GIS, the term uncertainty has come to be used as the catch-all term to describe situations in which the digital representation is simply incomplete, and as a measure of the general quality of the representation.

Many geographic representations depend upon inherently vague definitions and concepts

The views outlined in the previous paragraph are themselves controversial, and a rich ground for endless philosophical discussions. Some would argue that uncertainty can be inherent in phenomena themselves, rather than just in their description. Others would argue for distinctions between *vagueness, uncertainty, fuzziness, imprecision, inaccuracy*, and many other terms that most people use as if they were essentially synonymous. Information scientist

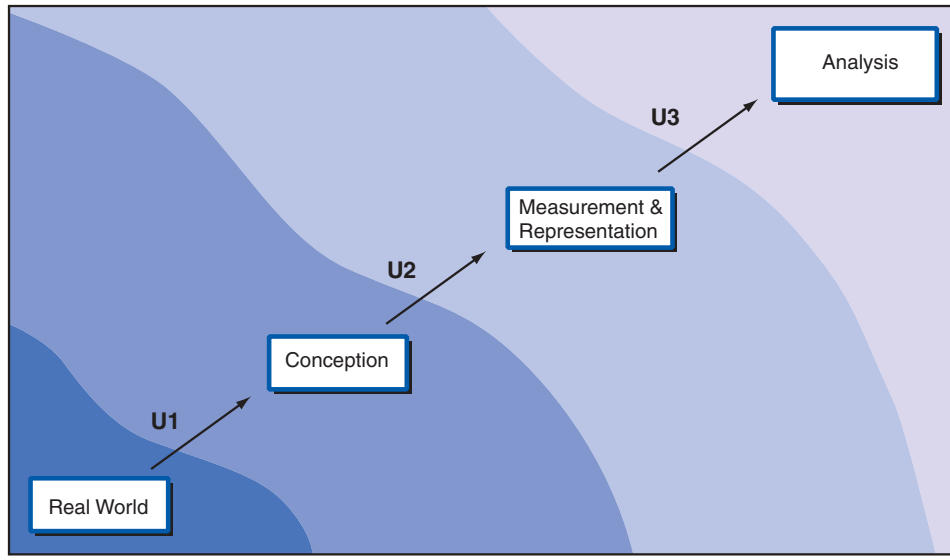


Figure 6.1 A conceptual view of uncertainty. The three filters, U1, U2, and U3 can distort the way in which the complexity of the real world is conceived, measured and represented, and analyzed in a cumulative way

Peter Fisher has provided a useful and wide-ranging discussion of these terms. We take the catch-all view here, and leave the detailed arguments to further study.

In this chapter, we will discuss some of the principal sources of uncertainty and some of the ways in which uncertainty degrades the quality of a spatial representation. The way in which we conceive of a geographic phenomenon very much prescribes the way in which we are likely to set about measuring and representing it. The measurement procedure, in turn, heavily conditions the ways in which it may be analyzed within a GIS. This chain sequence of events, in which *conception* prescribes *measurement and representation*, which in turn prescribes *analysis* is a succinct way of summarizing much of the content of this chapter, and is summarized in Figure 6.1. In this diagram, U1, U2, and U3 each denote *filters* that selectively distort or transform the representation of the real world that is stored and analyzed in GIS: a later chapter (Section 13.2.1) introduces a fourth filter that mediates interpretation of analysis, and the ways in which feedback may be accommodated through improvements in representation.

6.2 U1: Uncertainty in the conception of geographic phenomena

6.2.1 Units of analysis

Our discussion of Tobler's Law (Section 3.1) and of spatial autocorrelation (Section 4.6) established that geographic data handling is different from all other classes of non-spatial applications. A further characteristic that sets geographic information science apart from most every other science is that it is only rarely founded upon *natural* units of analysis. What is the natural unit of measurement for a soil profile? What is the spatial extent of a *pocket* of high unemployment, or a *cluster* of cancer cases? How might we delimit an environmental impact study of spillage from an oil tanker (Figure 6.2)? The questions become still more difficult in bivariate (two variable) and multivariate (more than two variable) studies. At what scale is it appropriate to investigate any relationship between background radiation and the incidence of leukemia? Or to assess any relationship between labor-force qualifications and unemployment rates?

In many cases there are no natural units of geographic analysis.

Figure 6.2 How might the spatial impact of an oil tanker spillage be delineated? We can measure the dispersion of the pollutants, but their impacts extend far beyond these narrowly defined boundaries (Reproduced by permission of Sam C. Pierson, Jr., Photo Researchers)

The discrete object view of geographic phenomena is much more reliant upon the idea of natural units of analysis than the field view. Biological organisms are almost always natural units of analysis, as are groupings such as households or families – though even here there are certainly difficult cases, such as the massive networks of fungal strands that are often claimed to be the largest living organisms on Earth, or extended families of human individuals. Things we manipulate, such as pencils, books, or screwdrivers, are also obvious natural units. The examples listed in the previous paragraph fall almost entirely into one of two categories – they are either instances of fields, where variation can be thought of as inherently continuous in space, or they are instances of poorly defined aggregations of discrete objects. In both of these cases it is up to the investigator to make the decisions about units of analysis, making the identification of the objects of analysis inherently subjective.

6.2.2 Vagueness and ambiguity

6.2.2.1 Vagueness

The frequent absence of objective geographic individual units means that, in practice, the labels that we assign to zones are often vague best guesses. What absolute or relative incidence of oak trees in a forested zone qualifies it for the label *oak woodland* (Figure 6.3)? Or, in a developing-country context in which aerial photography rather than ground enumeration is used to estimate population size, what rate of incidence of dwellings identifies a zone of *dense* population? In each of these instances, it is expedient to transform point-like events (individual trees or individual dwellings) into area objects, and pragmatic decisions must be taken in order to create a working definition of a spatial distribution. These decisions have no absolute validity, and raise two important questions:

- Is the defining boundary of a zone crisp and well-defined?
- Is our assignment of a particular label to a given zone robust and defensible?

Figure 6.3 Seeing the wood for the trees: what absolute or relative incidence rate makes it meaningful to assign the label 'oak woodland'? (Reproduced by permission of Ellan Young, Photo Researchers)

Uncertainty can exist both in the positions of the boundaries of a zone and in its attributes.

The questions have statistical implications (can we put numbers on the confidence associated with boundaries or labels?), cartographic implications (how can we convey the meaning of vague boundaries and labels through appropriate symbols on maps and GIS displays?), and cognitive implications (do people subconsciously attempt to force things into categories and boundaries to satisfy a deep need to simplify the world?).

6.2.2.2 Ambiguity

Many objects are assigned different labels by different national or cultural groups, and such groups perceive space differently. Geographic prepositions like *across, over*, and *in* (used in the Yellow Pages query in Figure 1.17) do not have simple correspondences with terms in other languages. Object names and the topological relations between them may thus be inherently *ambiguous*. Perception, behavior, language, and cognition all play a part in the conception of real-world entities and the relationships between them. GIS cannot present a value-neutral view of the world, yet it can provide a formal framework for the reconciliation of different worldviews. The geographic nature of this ambiguity may even be exploited to identify regions with shared characteristics and worldviews. To this end, Box 6.1 describes how different surnames used to describe essentially the same historic occupations provide an enduring measure in region building.

Many linguistic terms used to convey geographic information are inherently ambiguous.

Ambiguity also arises in the conception and construction of *indicators* (see also Section 16.2.1). *Direct* indicators are deemed to bear a clear correspondence with a mapped phenomenon. Detailed household income figures, for example, provide a direct indicator of the likely geography of expenditure and demand for goods and services; tree diameter at breast height can be used to estimate stand value; and field nutrient measures can be used to estimate agronomic yield. *Indirect* indicators are used when the best available measure is a perceived surrogate link with the phenomenon of interest. Thus the incidence of central heating amongst households, or rates of multiple car ownership, might provide a surrogate for (unavailable) household income data, while local atmospheric measurements of nitrogen dioxide might provide an indirect indicator of environmental health. Conception of the (direct or indirect) linkage between any indicator and the phenomenon of interest is subjective, hence ambiguous. Such measures will create (possibly systematic) errors of measurement if the correspondence between the two is imperfect. So, for example, differences in the conception of what hardship and deprivation entail can lead to specification of different composite indicators, and different geodemographic systems include different cocktails of census variables (Section 2.3.3). With regard to the natural environment, conception of critical defining properties

Applications Box **6.1**

Historians need maps of our uncertain past

In the study of history, there are many ways in which 'spatial is special' (Section 1.1.1). For example, it is widely recognized that although what our ancestors did (their occupations) and the social groups (classes) to which they belonged were clearly important in terms of demographic behavior, location and place were of equal if not greater importance. Although population changes occur in particular socio-economic circumstances, they are also strongly influenced by the unique characteristics, or 'cultural identities', of particular places. In Great Britain today, as almost everywhere else in the world, most people still think of their nation as made up of 'regions', and their stability and defining characteristics are much debated by cultural geographers and historians.

Yet analyzing and measuring human activity by place creates particular problems for historians. Most obviously, the past was very much less data rich than the present, and few systematic data sources survive. Moreover, the geographical administrative units by which the events of the past were recorded are both complex and changing. In an ideal world, perhaps, physical and cultural boundaries would always coincide, but physical features alone rarely provide appropriate *indicators* of the limits of socio-economic conditions and cultural circumstance.

Unfortunately many mapped historical data are still presented using high-level aggregations, such as counties or regions. This achieves a measure of standardization but may depict demography in only the most arbitrary of ways. If data are forced into geographic administrative units that were delineated for other purposes, regional maps may present nothing more than misleading, or even meaningless, spatial means (see Box 1.9).

In England and in many other countries, the daily activities of most individuals historically revolved around small numbers of contiguous civil *parishes*, of which there were more than 16 000 in the 19th century. These are the smallest administrative units for which data are systematically available. They provide the best available building blocks for meaningful region building. But how can we group parishes in order to identify non-overlapping geographic territories to which people felt that they belonged? And what indicators of regional identity are likely to have survived for all individuals in the population?

Historian Kevin Schürer (Box 13.2) has investigated these questions using a historical GIS to map digital surname data from the 1881 Census of England and Wales. The motivation for the GIS arises from the observation that many surnames contain statements of regional identity, and the suggestion that distinct zones of similar surnames might be described as homogeneous regions. The digitized records of the 1881 Census for England and Wales cover some 26 million people: although some 41 000 different surnames are recorded, a fifth of the population shared just under 60 surnames, and half of the population were accounted for by some 600 surnames. Schürer suggests that these aggregate statistics conceal much that we might learn about regional identity and diversity.

Many surnames of European origin are formed from occupational titles. Occupations often have uneven regional distributions and sometimes similar occupations are described using different names in different places (at the global scale, today's 'realtors' in the US perform much the same functions as their 'estate agent' counterparts in the UK, for example). Schürer has investigated the 1881 geographical distribution of three occupational surnames – Fuller, Tucker, and Walker. These essentially refer to the same occupation; namely someone who, from around the 14th century onwards, worked in the preparation of textiles by scouring or beating cloth as a means of finishing or cleansing it. Using GIS, Schürer confirms that the geographies of these 14th century surnames remained of enduring importance in defining the regional geography of England in 1881. Figure 6.4 illustrates that in 1881 Tuckers remained concentrated in the West Country, while Fullers occurred principally in the east and Walkers resided in the Midlands and north. This map also shows that there was not much mixing of the surnames in the transition zones between names, suggesting that the maps provide a useful basis to region building.

The enduring importance of surnames as evidence of the strength and durability of regional cultures has been confirmed in an update to the work by Daryl Lloyd at University College London: Lloyd used the 2003 UK Electoral Register to map the distribution of the same three surnames (Figure 6.5) and identified persistent regional concentrations.

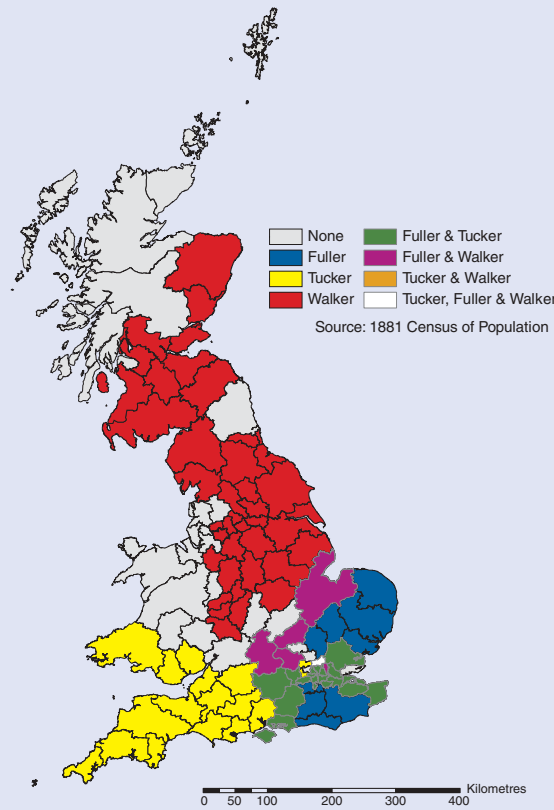


Figure 6.4 The 1881 geography of the Fullers, Tuckers, and Walkers (Reproduced with permission of K. Schürer)

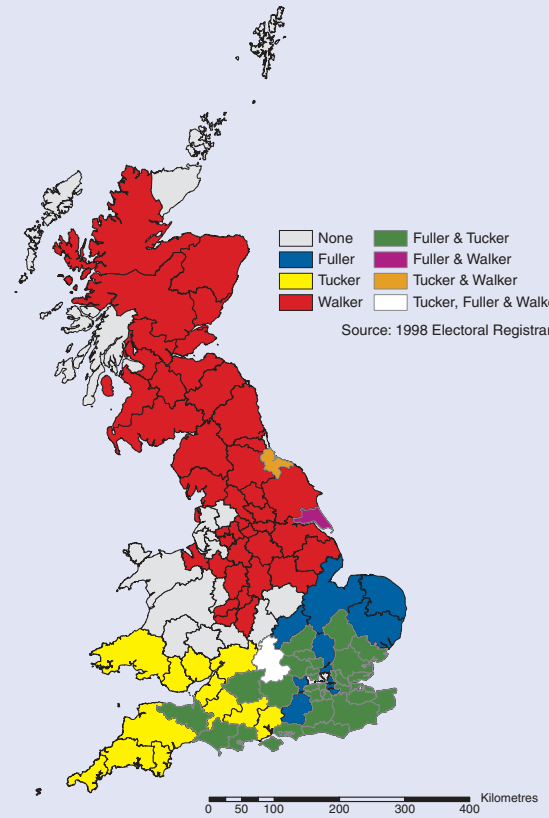


Figure 6.5 The 2003 geography of the Fullers, Tuckers, and Walkers (Reproduced with permission of Daryl Lloyd)

of soils can lead to inherent ambiguity in their classification (see Section 6.2.4).

Ambiguity is introduced when imperfect indicators of phenomena are used instead of the phenomena themselves.

Fundamentally, GIS has upgraded our abilities to generalize about spatial distributions. Yet our abilities to do so may be constrained by the different taxonomies that are conceived and used by data-collecting organizations within our overall study area. A study of wetland classification in the US found no fewer than six agencies engaged in mapping the same phenomena over the same geographic areas, and each with their own definitions of wetland types (see Section 1.2). If wetland maps are to be used in regulating the use of land, as they are in many areas, then uncertainty in mapping clearly exposes regulatory agencies to potentially damaging and costly lawsuits. How might soils data classified according to the UK national classification be assimilated within a pan-European soils map, which uses a classification honed to the full range and diversity of soils found across the European continent rather than those just on an offshore island? How might different national geodemographic classifications be combined into a form suitable for a pan-European marketing exercise? These are all variants of the question:

- How may mismatches between the categories of different classification schema be reconciled?

Differences in definitions are a major impediment to integration of geographic data over wide areas.

Like the process of pinning down the different nomenclatures developed in different cultural settings, the process of reconciling the semantics of different classification schema is an inherently *ambiguous* procedure. Ambiguity arises in data concatenation when we are unsure regarding the *meta-category* to which a particular class should be assigned.

6.2.3 Fuzzy approaches

One way of resolving the assignment process is to adopt a probabilistic interpretation. If we take a statement like

'the database indicates that this field contains wheat, but there is a 0.17 probability (or 17% chance) that it actually contains barley', there are at least two possible interpretations:

(a) If 100 randomly chosen people were asked to make independent assessments of the field on the ground, 17 would determine that it contains barley, and 83 would decide it contains wheat.

(b) Of 100 similar fields in the database, 17 actually contained barley when checked on the ground, and 83 contained wheat.

Of the two we probably find the second more acceptable because the first implies that people cannot correctly determine the crop in the field. But the important point is that, in conceptual terms, both of these interpretations are *frequentist*, because they are based on the notion that the probability of a given outcome can be defined as the proportion of times the outcome occurs in some real or imagined experiment, when the number of tests is very large. Yet while this is reasonable for classic statistical experiments, like tossing coins or drawing balls from an urn, the geographic situation is different – there is only one field with precisely these characteristics, and one observer, and in order to imagine a number of tests we have to invent more than one observer, or more than one field (the problems of imagining larger populations for some geographic samples are discussed further in Section 15.4).

In part because of this problem, many people prefer the *subjectivist* conception of probability – that it represents a judgment about relative likelihood that is not the result of any frequentist experiment, real or imagined. Subjective probability is similar in many ways to the concept of fuzzy sets, and the latter framework will be used here to emphasize the contrast with frequentist probability.

Suppose we are asked to examine an aerial photograph to determine whether a field contains wheat, and we decide that we are not sure. However, we are able to put a number on our degree of uncertainty, by putting it on a scale from 0 to 1. The more certain we are, the higher the number. Thus we might say we are 0.90 sure it is wheat, and this would reflect a greater degree of certainty than 0.80. This degree of belonging to the class *wheat* is termed the *fuzzy membership*, and it is common though not necessary to limit memberships to the range 0 to 1. In effect, we have changed our view of membership in classes, and abandoned the notion that things must either belong to classes or not belong to them – in this new world, the boundaries of classes are no longer clean and crisp, and the set of things assigned to a set can be fuzzy.

In fuzzy logic, an object's degree of belonging to a class can be partial.

One of the major attractions of fuzzy sets is that they appear to let us deal with sets that are not precisely defined, and for which it is impossible to establish membership cleanly. Many such sets or classes are found in GIS applications, including land use categories, soil types, land cover classes, and vegetation types. Classes used for maps are often fuzzy, such that two people asked to classify the same location might disagree, not because of measurement error, but because the classes themselves are not perfectly defined and because opinions vary. As such, mapping is often forced to stretch the rules of scientific repeatability, which require that two observers will always agree. Box 6.2 shows a typical extract from the legend of a soil map, and it is easy to see how two people might disagree, even though both are experts with years of experience in soil classification.

Figure 6.6 shows an example of mapping classes using the fuzzy methods developed by A-Xing Zhu of the

Technical Box **6.2**

Fuzziness in classification: description of a soil class

Following is the description of the Limerick series of soils from New England, USA (the type location is in Chittenden County, Vermont), as defined by the National Cooperative Soil Survey. Note the frequent use of vague terms such as 'very', 'moderate', 'about', 'typically', and 'some'. Because the definition is so loose it is possible for many distinct soils to be lumped together in this one class – and two observers may easily disagree over whether a given soil belongs to the class, even though both are experts. The definition illustrates the extreme problems of defining soil classes with sufficient rigor to satisfy the criterion of scientific repeatability.

The Limerick series consists of very deep, poorly drained soils on flood plains. They formed in loamy alluvium. Permeability is moderate. Slope ranges from 0 to 3 percent. Mean annual precipitation is about 34 inches and mean annual temperature is about 45 degrees F.

Depth to bedrock is more than 60 inches. Reaction ranges from strongly acid to neutral in the surface layer and moderately acid to neutral in the substratum. Textures are typically silt loam or very fine sandy loam, but lenses of loamy very fine sand or very fine sand are present in some pedons. The weighted average of fine and coarser sands, in the particle-size control section, is less than 15 percent.

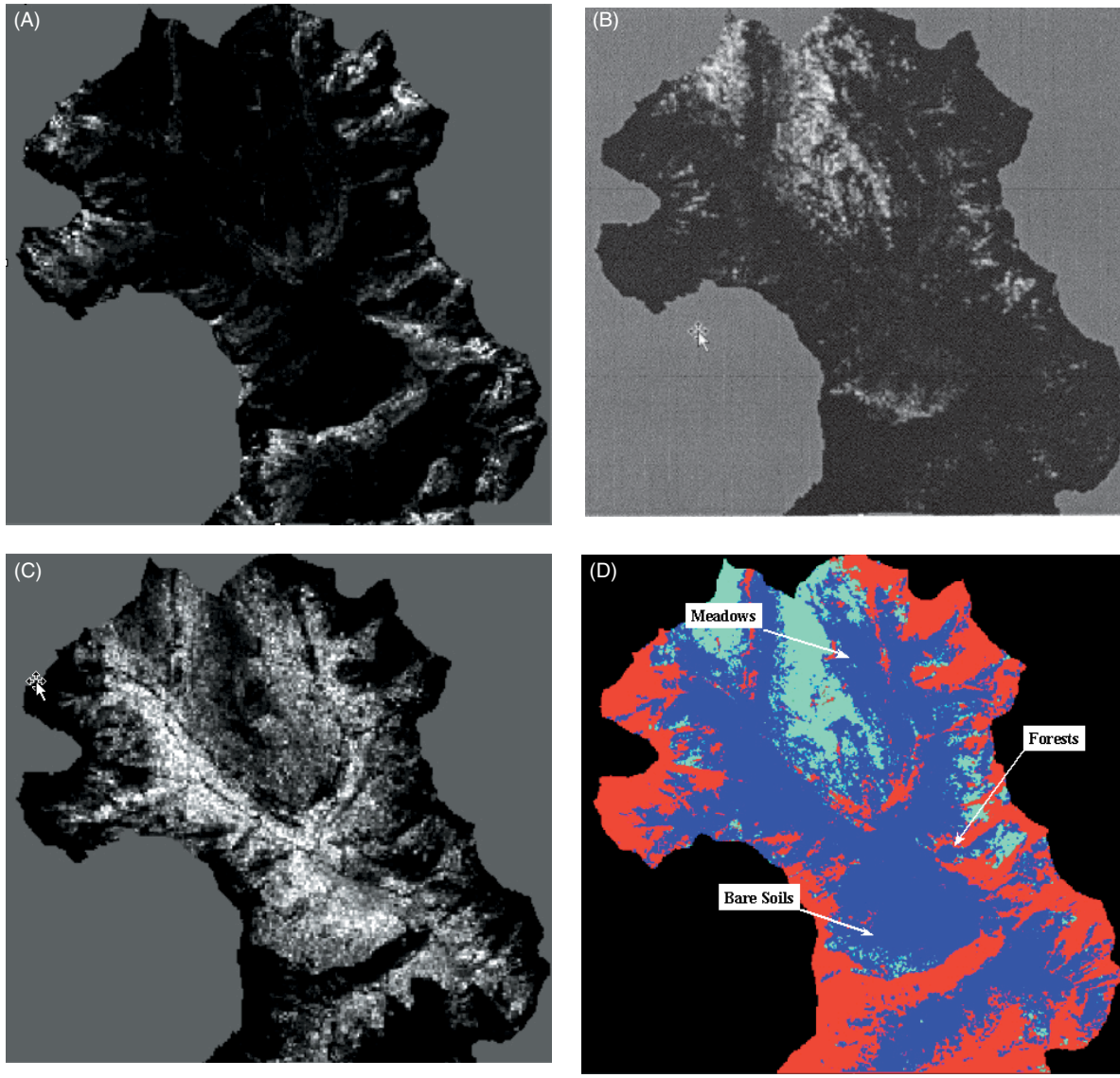


Figure 6.6 (A) Membership map for bare soils in the Upper Lake McDonald basin, Glacier National Park. High membership values are in the ridge areas where active colluvial and glacier activities prevent the establishment of vegetation. (B) Membership map for forest. High membership values are in the middle to lower slope areas where the soils are both stable and better drained. (C) Membership map for alpine meadows. High membership values are on gentle slopes at high elevation where excessive soil water and low temperature prevent the growth of trees. (D) Spatial distribution of the three cover types from hardening the membership maps. (Reproduced by permission of A-Xing Zhu)

University of Wisconsin-Madison, USA, which take both remote sensing images and the opinions of experts as inputs. There are three classes, and each map shows the fuzzy membership values in one class, ranging from 0 (darkest) to 1 (lightest). This figure also shows the result of converting to *crisp* categories, or *hardening* – to obtain Figure 6.6D, each pixel is colored according to the class with the highest membership value.

Fuzzy approaches are attractive because they capture the uncertainty that many of us feel about the assignment of places on the ground to specific categories. But researchers have struggled with the question of whether they are more *accurate*. In a sense, if we are uncertain about which class to choose then it is more accurate to say so, in the form of a fuzzy membership, than to be forced into assigning a class without qualification. But that does not address the question of whether the fuzzy membership value is accurate. If Class A is not well defined, it is hard to see how one person's assignment of a fuzzy membership of 0.83 in Class A can be meaningful to another person, since there is no reason to believe that the two people share the same notions of what Class A means, or of what 0.83 means, as distinct from 0.91, or 0.74. So while fuzzy approaches make sense at an intuitive level, it is more difficult to see how they could be helpful in the

process of communication of geographic knowledge from one person to another.

6.2.4 The scale of geographic individuals

There is a sense in which vagueness and ambiguity in the conception of *usable* (rather than *natural*) units of analysis undermines the very foundations of GIS. How, in practice, may we create a sufficiently secure base to support geographic analysis? Geographers have long grappled with the problems of defining systems of zones and have marshaled a range of deductive and inductive approaches to this end (see Section 4.9 for a discussion of what deduction and induction entail). The long-established regional geography tradition is fundamentally concerned with the delineation of zones characterized by internal homogeneity (with respect to climate, economic development, or agricultural land use, for example), within a zonal scheme which maximizes between-zone heterogeneity, such as the map illustrated in Figure 6.7. Regional geography is fundamentally about delineating *uniform* zones, and many employ multivariate statistical techniques such as cluster analysis to supplement, or post-rationalize, intuition.

Identification of homogeneous zones and spheres of influence lies at the heart of traditional regional geography as well as contemporary data analysis.

Other geographers have tried to develop *functional* zonal schemes, in which zone boundaries delineate the breakpoints between the spheres of influence of adjacent facilities or features – as in the definition of travel-to-work areas (Figure 6.8) or the definition of a river catchment. Zones may be defined such that there is maximal interaction within zones, and minimal between zones. The scale at which uniformity or functional integrity is conceived clearly conditions the ways it is measured – in terms of the magnitude of within-zone heterogeneity that must be accommodated in the case of uniform zones, and the degree of leakage between the units of functional zones.

Scale has an effect, through the concept of spatial autocorrelation outlined in Section 4.3, upon the outcome of geographic analysis. This was demonstrated more than half a century ago in a classic paper by Yule and Kendall, where the correlation between wheat and potato yields was shown systematically to increase as English county units were amalgamated through a succession of coarser scales (Table 6.1). A succession of research papers has subsequently reaffirmed the existence of similar scale effects in multivariate analysis. However, rather discouragingly, scale effects in multivariate cases do not follow any consistent or predictable trends. This theme of dependence of results on the geographic units of analysis is pursued further in Section 6.4.3.

Relationships typically grow stronger when based on larger geographic units.

GIS appears to trivialize the task of creating composite thematic maps. Yet inappropriate conception of the scale of geographic phenomena can mean that apparent spatial

Figure 6.7 The regional geography of Russia. (*Source*: de Blij H.J. and Muller P.O. 2000 *Geography: Realms, Regions and Concepts* (9th edn) New York: Wiley, p. 113)

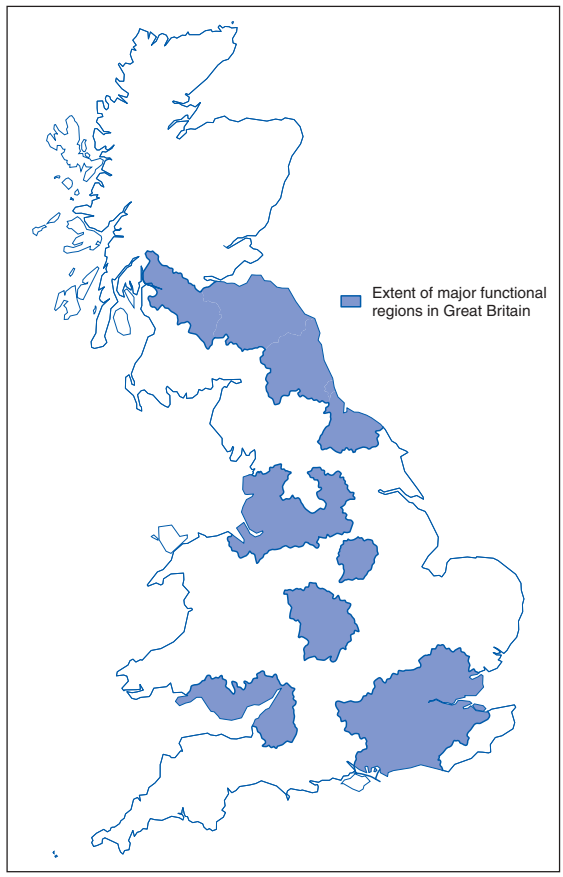


Figure 6.8 Dominant functional regions of Great Britain. (*Source*: Champion A.G., Green A.E., Owen D.W., Ellin D.J., Coombes M.G. 1987 *Changing Places: Britain's Demographic, Economic and Social Complexion*, London: Arnold, p. 9)

Table 6.1 In 1950 Yule and Kendall used data for wheat and potato yields from the (then) 48 counties of England to demonstrate that correlation coefficients tend to increase with scale. They aggregated the 48-county data into zones so that there were first 24, then 12, then 6, and finally just 3 zones. The range of their results, from near zero (no correlation) to over 0.99 (almost perfect positive correlation) demonstrates the range of results that can be obtained, although subsequent research has suggested that this range of values is atypical

No. of geographic areas	Correlation
48	0.2189
24	0.2963
12	0.5757
6	0.7649
3	0.9902

patterning (or the lack of it) in mapped data may be oversimplified, crude, or even illusory. It is also clearly inappropriate to conceive of boundaries as crisp and well-defined if significant leakage occurs across them (as happens, in practice, in the delineation of most functional regions), or if geographic phenomena are by nature fuzzy, vague, or ambiguous.

6.3 U2: Further uncertainty in the measurement and representation of geographic phenomena

6.3.1 Measurement and representation

The conceptual models (fields and objects) that were introduced in Chapter 3 impose very different filters upon reality, and their usual corresponding representational models (raster and vector) are characterized by different uncertainties as a consequence. The vector model enables a range of powerful analytical operations to be performed (see Chapters 14 through 16), yet it also requires *a priori* conceptualization of the nature and extent of geographic individuals and the ways in which they nest together into higher-order zones. The raster model defines individual elements as square cells, with boundaries that bear no relationship at all to natural features, but nevertheless provides a convenient and (usually) efficient structure for data handling within a GIS. However, in the absence of effective automated pattern recognition techniques, human interpretation is usually required to discriminate between real-world spatial entities as they appear in a rasterized image.

Although quite different representations of reality, vector and raster data structures are both attractive in their logical consistency, the ease with which they are able to handle spatial data, and (once the software is written) the ease with which they can be implemented in GIS. But neither abstraction provides easy measurement fixes and there is no substitute for robust conception of geographic units of analysis (Section 6.2). This said, however, the conceptual distinction between fields and discrete objects is often useful in dealing with uncertainty. Figure 6.9 shows a coastline, which is often conceptualized as a discrete line object. But suppose we recognize that its position is uncertain. For example, the coastline shown on a 1:2 000 000 map is a gross generalization, in which major liberties are taken, particularly in areas where the coast is highly indented and irregular. Consequently the 1:2 000 000 version leaves substantial uncertainty about the true location of the shoreline. We might approach this by changing from a line to an area, and mapping the area where the actual coastline lies, as shown in the figure. But another approach would be to reconceptualize the coastline as a field, by mapping a variable whose value represents the probability that a point is land. This is shown in the figure as a raster representation. This would have far more information content, and consequently much more value in many applications. But at the same time it would be difficult to find an appropriate data source for the representation – perhaps a fuzzy classification of an air photo, using one of an

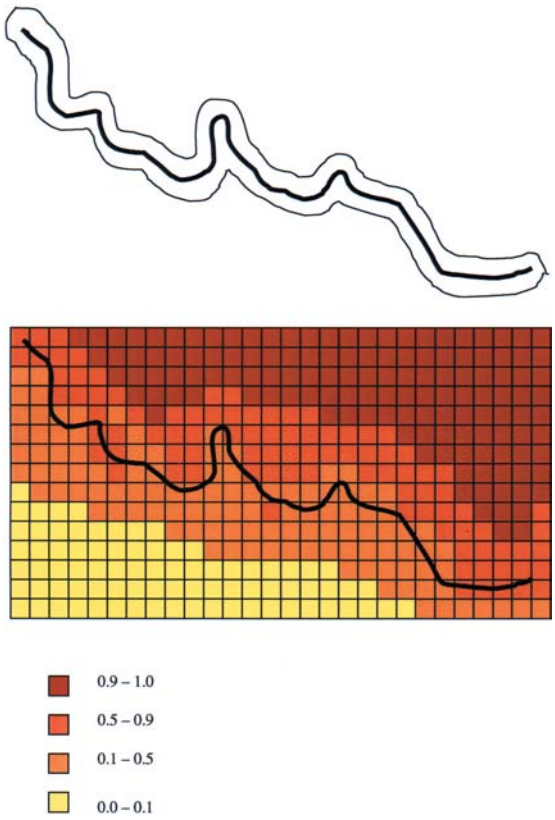


Figure 6.9 The contrast between discrete object (top) and field (bottom) conceptualizations of an uncertain coastline. In the discrete object view the line becomes an area delimiting where the true coastline might be. In the field view a continuous surface defines the probability that any point is land

increasing number of techniques designed to produce representations of the uncertainty associated with objects discovered in images.

Uncertainty can be measured differently under field and discrete object views.

Indeed, far from offering quick fixes for eliminating or reducing uncertainty, the measurement process can actually increase it. Given that the vector and raster data models impose quite different filters on reality, it is unsurprising that they can each generate additional uncertainty in rather different ways. In field-based conceptualizations, such as those that underlie remotely sensed images expressed as rasters, spatial objects are not defined *a priori*. Instead, the classification of each cell into one or other category builds together into a representation. In remote sensing, when resolution is insufficient to detect all of the detail in geographic phenomena, the term *mixel* is often used to describe raster cells that contain more than one class of land – in other words, elements in which the outcome of statistical classification suggests the occurrence of multiple land cover categories. The total area of cells classified as mixed should decrease as the resolution of the satellite sensor increases, assuming the number of categories remains constant, yet a completely mixel-free classification is very unlikely at any level of resolution. Even where the Earth's surface is covered with perfectly homogeneous areas, such as agricultural fields growing uniform crops, the failure of real-world crop boundaries to line up with pixel edges ensures the presence of at least some mixels. Neither does higher-resolution imagery solve all problems: medium-resolution data (defined as pixel size of between 30 m × 30 m and 1000 m × 1000 m) are typically classified using between 3 and 7 bands, while high-resolution data (pixel sizes 10 × 10 m or smaller) are typically classified using between 7 and 256 bands, and this can generate much greater heterogeneity of spectral values with attendant problems for classification algorithms.

A pixel whose area is divided among more than one class is termed a mixel.

The vector data structure, by contrast, defines spatial entities and specifies explicit topological relations (see Section 3.6) between them. Yet this often entails transformations of the inherent characteristics of spatial objects (Section 14.4). In conceptual terms, for example, while the true individual members of a population might each be defined as point-like objects, they will often appear in a GIS dataset only as aggregate counts for apparently *uniform* zones. Such aggregation can be driven by the need to preserve confidentiality of individual records, or simply by the need to limit data volume. Unlike the field conceptualization of spatial phenomena, this implies that there are good reasons for partitioning space in a particular way. In practice, partitioning of space is often made on grounds that are principally pragmatic, yet are rarely completely random (see Section 6.4). In much of socio-economic GIS, for example, zones which are designed to preserve the anonymity of survey respondents may be largely *ad hoc* containers. Larger aggregations are often used for the simple reason that they permit comparisons of measures over time (see Box 6.1). They may also reflect the way that a cartographer or GIS interpolates a boundary between sampled points, as in the creation of isopleth maps (Box 4.3).

6.3.2 Statistical models of uncertainty

Scientists have developed many widely used methods for describing errors in observations and measurements, and these methods may be applicable to GIS if we are willing to think of databases as collections of measurements. For example, a digital elevation model consists of a large number of measurements of the elevation of the Earth's surface. A map of land use is also in a sense a collection of measurements, because observations of the land surface have resulted in the assignment of classes to locations. Both of these are examples of observed or measured attributes, but we can also think of location as a property that is measured.

A geographic database is a collection of measurements of phenomena on or near the Earth's surface.

Here we consider errors in nominal class assignment, such as of types of land use, and errors in continuous (interval or ratio) scales, such as elevation (see Section 3.4).

6.3.2.1 Nominal case

The values of nominal data serve only to distinguish an instance of one class from an instance of another, or to identify an object uniquely. If classes have an inherent ranking they are described as ordinal data, but for purposes of simplicity the ordinal case will be treated here as if it were nominal.

Consider a single observation of nominal data – for example, the observation that a single parcel of land is being used for agriculture (this might be designated by giving the parcel Class A as its value of the 'Land Use Class' attribute). For some reason, perhaps related to the quality of the aerial photography being used to build the database, the class may have been recorded falsely as Class G, Grassland. A certain proportion of parcels that are truly Agriculture might be similarly recorded as Grassland, and we can think of this in terms of a probability, that parcels that are truly Agriculture are falsely recorded as Grassland.

Table 6.2 shows how this might work for all of the parcels in a database. Each parcel has a true class, defined by accurate observation in the field, and a recorded class as it appears in the database. The whole table is described as a *confusion matrix*, and instances of confusion matrices are commonly encountered in applications dominated by class data, such as classifications derived from remote sensing or aerial photography. The true class might be determined by ground check, which is inherently more accurate than classification of aerial photographs, but much more expensive and time-consuming.

Ideally all of the observations in the confusion matrix should be on the principal diagonal, in the cells that correspond to agreement between true class and database class. But in practice certain classes are more easily confused than others, so certain cells off the diagonal will have substantial numbers of entries.

A useful way to think of the confusion matrix is as a set of rows, each defining a vector of values. The vector for row i gives the proportions of cases in which what appears to be Class i is actually Class 1, 2, 3, etc. Symbolically, this can be represented as a vector $\{p_1, p_2, \ldots, p_i, \ldots, p_n\}$, where n is the number of classes, and p_i represents the proportion of cases for which what appears to be the class according to the database is actually Class i.

Table 6.2 Example of a misclassification or confusion matrix. A grand total of 304 parcels have been checked. The rows of the table correspond to the land use class of each parcel as recorded in the database, and the columns to the class as recorded in the field. The numbers appearing on the principal diagonal of the table (from top left to bottom right) reflect correct classification

	A	B	C	D	E	Total
A	80	4	0	15	7	106
B	2	17	0	9	2	30
C	12	5	9	4	8	38
D	7	8	0	65	0	80
E	3	2	1	6	38	50
Total	104	36	10	99	55	304

There are several ways of describing and summarizing the confusion matrix. If we focus on one row, then the table shows how a given class in the database falsely records what are actually different classes on the ground. For example, Row A shows that of 106 parcels recorded as Class A in the database, 80 were confirmed as Class A in the field, but 15 appeared to be truly Class D. The proportion of instances in the diagonal entries represents the proportion of correctly classified parcels, and the total of off-diagonal entries in the row is the proportion of entries in the database that appear to be of the row's class but are actually incorrectly classified. For example, there were only 9 instances of agreement between the database and the field in the case of Class D. If we look at the table's columns, the entries record the ways in which parcels that are truly of that class are actually recorded in the database. For example, of the 10 instances of Class C found in the field, 9 were recorded as such in the database and only 1 was misrecorded as Class E.

The columns have been called the *producer's* perspective, because the task of the producer of an accurate database is to minimize entries outside the diagonal cell in a given column, and the rows have been called the *consumer's* perspective, because they record what the contents of the database actually mean on the ground; in other words, the accuracy of the database's contents.

Users and producers of data look at misclassification in distinct ways.

For the table as a whole, the proportion of entries in diagonal cells is called the *percent correctly classified* (PCC), and is one possible way of summarizing the table. In this case 209/304 cases are on the diagonal, for a PCC of 68.8%. But this measure is misleading for at least two reasons. First, chance alone would produce some correct classifications, even in the worst circumstances, so it would be more meaningful if the scale were adjusted such that 0 represents chance. In this case, the number of chance hits on the diagonal in a random assignment is 76.2 (the sum of the row total times the column total divided by the grand total for each of the five diagonal cells). So the actual number of diagonal hits, 209, should be compared to this number, not 0. The more useful index of success is the *kappa index*, defined as:

$$\kappa = \frac{\sum_{i=1}^{n} c_{ii} - \sum_{i=1}^{n} c_{i.}c_{.i}/c_{..}}{c_{..} - \sum_{i=1}^{n} c_{i.}c_{.i}/c_{..}}$$

where c_{ij} denotes the entry in row i column j, the dots indicate summation (e.g., $c_{i.}$ is the summation over

all columns for row i, that is, the row i total, and $c_{..}$ is the grand total), and n is the number of classes. The first term in the numerator is the sum of all the diagonal entries (entries for which the row number and the column number are the same). To compute PCC we would simply divide this term by the grand total (the first term in the denominator). For kappa, both numerator and denominator are reduced by the same amount, an estimate of the number of hits (agreements between field and database) that would occur by chance. This involves taking each diagonal cell, multiplying the row total by the column total, and dividing by the grand total. The result is summed for each diagonal cell. In this case kappa evaluates to 58.3%, a much less optimistic assessment than PCC.

The second issue with both of these measures concerns the relative abundance of different classes. In the table, Class C is much less common than Class A. The confusion matrix is a useful way of summarizing the characteristics of nominal data, but to build it there must be some source of more accurate data. Commonly this is obtained by ground observation, and in practice the confusion matrix is created by taking samples of more accurate data, by sending observers into the field to conduct spot checks. Clearly it makes no sense to visit every parcel, and instead a sample is taken. Because some classes are commoner than others, a random sample that made every parcel equally likely to be chosen would be inefficient, because too many data would be gathered on common classes, and not enough on the relatively rare ones. So, instead, samples are usually chosen such that a roughly equal number of parcels are selected in each class. Of course these decisions must be based on the class as recorded in the database, rather than the true class. This is an instance of sampling that is systematically *stratified* by class (see Section 4.4).

Sampling for accuracy assessment should pay greater attention to the classes that are rarer on the ground.

Parcels represent a relatively easy case, if it is reasonable to assume that the land use class of a parcel is uniform over the parcel, and class is recorded as a single attribute of each parcel object. But as we noted in Section 6.2, more difficult cases arise in sampling natural areas (for example in the case of vegetation cover class), where parcel boundaries do not exist. Figure 6.10 shows a typical vegetation cover class map, and is obviously highly generalized. If we were to apply the previous strategy, then we would test each area to see if its assigned vegetation cover class checks out on the ground. But unlike the parcel case, in this example the boundaries between areas are not fixed, but are themselves part of the observation process, and we need to ask whether they are correctly located. Error in this case has two forms: misallocation of an area's class and mislocation of an area's boundaries. In some cases the boundary between two areas may be fixed, because it coincides with a clearly defined line on the ground; but in other cases, the boundary's location is as much a matter of judgment as the allocation of an area's class. Peter Burrough and Andrew Frank have discussed many of the implications of uncertain boundaries in GIS.

Figure 6.10 An example of a vegetation cover map. Two strategies for accuracy assessment are available: to check by area (polygon), or to check by point. In the former case a strategy would be devised for field checking each area, to determine the area's correct class. In the latter, points would be sampled across the state and the correct class determined at each point

Errors in land cover maps can occur in the locations of boundaries of areas, as well as in the classification of areas.

In such cases we need a different strategy, that captures the influence both of mislocated boundaries and of misallocated classes. One way to deal with this is to think of error not in terms of classes assigned to areas, but in terms of classes assigned to points. In a raster dataset, the cells of the raster are a reasonable substitute for individual points. Instead of asking whether area classes are confused, and estimating errors by sampling areas, we ask whether the classes assigned to raster cells are confused, and define the confusion matrix in terms of misclassified cells. This is often called *per-pixel* or *per-point* accuracy assessment, to distinguish it from the previous strategy of *per-polygon* accuracy assessment. As before, we would want to stratify by class, to make sure that relatively rare classes were sampled in the assessment.

6.3.2.2 Interval/ratio case

The second case addresses measurements that are made on interval or ratio scales. Here, error is best thought of not as a change of class, but as a change of value, such that the observed value x' is equal to the true value

x plus some distortion δx, where δx is hopefully small. δx might be either positive or negative, since errors are possible in both directions. For example, the measured and recorded elevation at some point might be equal to the true elevation, distorted by some small amount. If the average distortion is zero, so that positive and negative errors balance out, the observed values are said to be *unbiased*, and the average value will be true.

Error in measurement can produce a change of class, or a change of value, depending on the type of measurement.

Sometimes it is helpful to distinguish between *accuracy*, which has to do with the magnitude of δx, and *precision*. Unfortunately there are several ways of defining precision in this context, at least two of which are regularly encountered in GIS. Surveyors and others concerned with measuring instruments tend to define precision through the performance of an instrument in making repeated measurements of the same phenomenon. A measuring instrument is precise according to this definition if it repeatedly gives similar measurements, whether or not these are actually accurate. So a GPS receiver might make successive measurements of the same elevation, and if these are similar the instrument is said to be precise. Precision in this case can be measured by the variability among repeated measurements. But it is possible, for example, that all of the measurements are approximately 5 m too high, in which case the measurements are said to be biased, even though they are precise, and the instrument is said to be inaccurate. Figure 6.11 illustrates this meaning of precise, and its relationship to accuracy.

The other definition of precision is more common in science generally. It defines precision as the number of digits used to report a measurement, and again it is not necessarily related to accuracy. For example, a GPS receiver might measure elevation as 51.3456 m. But if the receiver is in reality only accurate to the nearest 10 cm, three of those digits are spurious, with no real meaning. So, although the precision is one ten thousandth of a meter, the accuracy is only one tenth of a meter. Box 6.3 summarizes the rules that are used to ensure that reported measurements do not mislead by appearing to have greater accuracy than they really do.

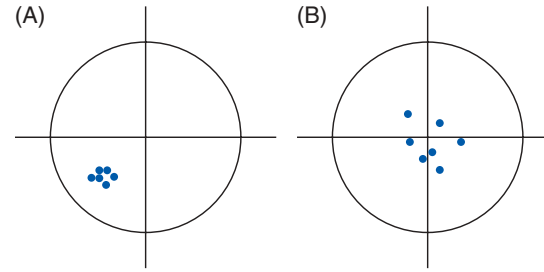


Figure 6.11 The term *precision* is often used to refer to the repeatability of measurements. In both diagrams six measurements have been taken of the same position, represented by the center of the circle. In (A) successive measurements have similar values (they are *precise*), but show a bias away from the correct value (they are *inaccurate*). In (B), precision is lower but accuracy is higher

To most scientists, precision refers to the number of significant digits used to report a measurement, but it can also refer to a measurement's repeatability.

In the interval/ratio case, the magnitude of errors is described by the *root mean square error* (RMSE), defined as the square root of the average squared error, or:

$$\left[\sum \delta x^2 / n\right]^{1/2}$$

Technical Box **6.3**

Good practice in reporting measurements

Here are some simple rules that help to ensure that people receiving measurements from others are not misled by their apparently high precision.

1. The number of digits used to report a measurement should reflect the measurement's accuracy. For example, if a measurement is accurate to 1 m then no decimal places should be reported. The measurement 14.4 m suggests accuracy to one tenth of a meter, as does 14.0, but 14 suggests accuracy to 1 m.
2. Excess digits should be removed by rounding. Fractions above one half should be rounded up, fractions below one half should be rounded down. The following examples reflect rounding to two decimal places:

   14.57803 rounds to 14.58
   14.57397 rounds to 14.57
   14.57999 rounds to 14.58
   14.57499 rounds to 14.57
3. These rules are not effective to the left of the decimal place – for example, they give no basis for knowing whether 1400 is accurate to the nearest unit, or to the nearest hundred units.
4. If a number is known to be exactly an integer or whole number, then it is shown with no decimal point.

where the summation is over the values of δx for all of the n observations. The RMSE is similar in a number of ways to the standard deviation of observations in a sample. Although RMSE involves taking the square root of the average squared error, it is convenient to think of it as approximately equal to the average error in each observation, whether the error is positive or negative. The US Geological Survey uses RMSE as its primary measure of the accuracy of elevations in digital elevation models, and published values range up to 7 m.

Although the RMSE can be thought of as capturing the magnitude of the average error, many errors will be greater than the RMSE, and many will be less. It is useful, therefore, to know how errors are *distributed* in magnitude – how many are large, how many are small. Statisticians have developed a series of models of error distributions, of which the commonest and most important is the Gaussian distribution, otherwise known as the error function, the 'bell curve', or the Normal distribution. Figure 6.12 shows the curve's shape. If observations are unbiased, then the mean error is zero (positive and negative errors cancel each other out), and the RMSE is also the distance from the center of the distribution (zero) to the points of inflection on either side, as shown in the figure. Let us take the example of a 7 m RMSE on elevations in a USGS digital elevation model; if error follows the Gaussian distribution, this means that some errors will be more than 7 m in magnitude, some will be less, and also that the relative abundance of errors of any given size is described by the curve shown. 68% of errors will be between −1.0 and +1.0 RMSEs, or −7 m and +7 m. In practice many distributions of error do follow the Gaussian distribution, and there are good theoretical reasons why this should be so.

The Gaussian distribution predicts the relative abundances of different magnitudes of error.

To emphasize the mathematical formality of the Gaussian distribution, its equation is shown below. The symbol

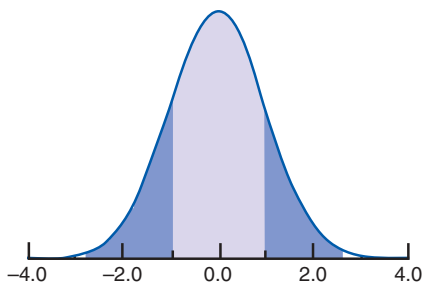


Figure 6.12 The Gaussian or Normal distribution. The height of the curve at any value of x gives the relative abundance of observations with that value of x. The area under the curve between any two values of x gives the probability that observations will fall in that range. The range between −1 standard deviation and +1 standard deviation is in light purple. It encloses 68% of the area under the curve, indicating that 68% of observations will fall between these limits

σ denotes the standard deviation, μ denotes the mean (in Figure 6.12 these values are 1 and 0 respectively), and exp is the exponential function, or '2.71828 to the power of'. Scientists believe that it applies very broadly, and that many instances of measurement error adhere closely to the distribution, because it is grounded in rigorous theory. It can be shown mathematically that the distribution arises whenever a large number of random factors contribute to error, and the effects of these factors combine additively – that is, a given effect makes the same additive contribution to error whatever the specific values of the other factors. For example, error might be introduced in the use of a steel tape measure over a large number of measurements because some observers consistently pull the tape very taught, or hold it very straight, or fastidiously keep it horizontal, or keep it cool, and others do not. If the combined effects of these considerations always contributes the same amount of error (e.g., +1 cm, or −2 cm), then this contribution to error is said to be additive.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

We can apply this idea to determine the inherent uncertainty in the locations of contours. The US Geological Survey routinely evaluates the accuracies of its digital elevation models (DEMs), by comparing the elevations recorded in the database with those at the same locations in more accurate sources, for a sample of points. The differences are summarized in a RMSE, and in this example we will assume that errors have a Gaussian distribution with zero mean and a 7 m RMSE. Consider a measurement of 350 m. According to the error model, the truth might be as high as 360 m, or as low as 340 m, and the relative frequencies of any particular error value are as predicted by the Gaussian distribution with a mean of zero and a standard deviation of 7. If we take error into account, using the Gaussian distribution with an RMSE of 7 m, it is no longer clear that a measurement of 350 m lies exactly on the 350 m contour. Instead, the truth might be 340 m, or 360 m, or 355 m. Figure 6.13 shows the implications of this in terms of the location of this contour in a real-world example. 95% of errors would put the contour within the colored zone. In areas colored red the observed value is less than 350 m, but the truth might be 350 m; in areas colored green the observed value is more than 350 m, but the truth might be 350 m. There is a 5% chance that the true location of the contour lies outside the colored zone entirely.

6.3.3 Positional error

In the case of measurements of position, it is possible for every coordinate to be subject to error. In the two-dimensional case, a measured position (x', y') would be subject to errors in both x and y; specifically, we might write $x' = x + \delta x$, $y' = y + \delta y$, and similarly in the three-dimensional case where all three coordinates are measured, $z' = z + \delta z$. The *bivariate Gaussian distribution* describes errors in the two horizontal dimensions,

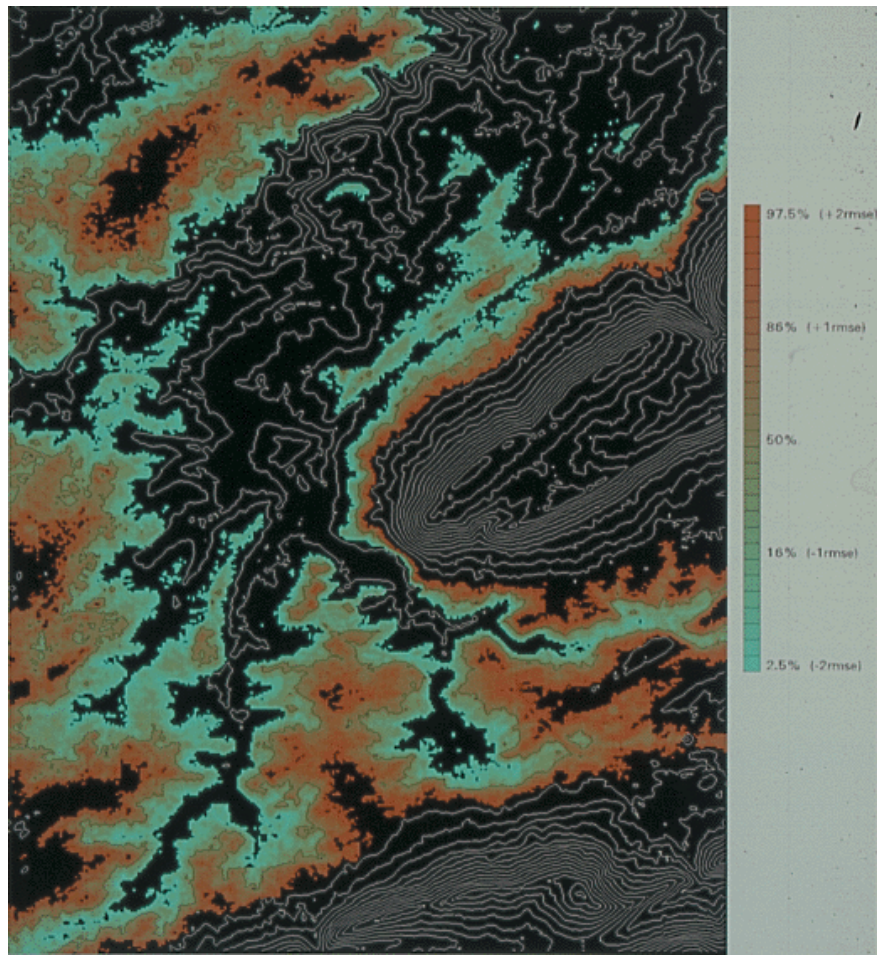


Figure 6.13 Uncertainty in the location of the 350 m contour in the area of State College, Pennsylvania, generated from a US Geological Survey DEM with an assumed RMSE of 7 m. According to the Gaussian distribution with a mean of 350 m and a standard deviation of 7 m, there is a 95% probability that the true location of the 350 m contour lies in the colored area, and a 5% probability that it lies outside (*Source*: Hunter G. J. and Goodchild M. F. 1995 'Dealing with error in spatial databases: a simple case study'. *Photogrammetric Engineering and Remote Sensing* **61**: 529–37)

and it can be generalized to the three-dimensional case. Normally, we would expect the RMSEs of x and y to be the same, but z is often subject to errors of quite different magnitude, for example in the case of determinations of position using GPS. The bivariate Gaussian distribution also allows for correlation between the errors in x and y, but normally there is little reason to expect correlations.

Because it involves two variables, the bivariate Gaussian distribution has somewhat different properties from the simple (univariate) Gaussian distribution. 68% of cases lie within one standard deviation for the univariate case (Figure 6.12). But in the bivariate case with equal standard errors in x and y, only 39% of cases lie within a circle of this radius. Similarly, 95% of cases lie within two standard deviations for the univariate distribution, but it is necessary to go to a circle of radius equal to 2.15 times the x or y standard deviations to enclose 90% of the bivariate distribution, and 2.45 times standard deviations for 95%.

National Map Accuracy Standards often prescribe the positional errors that are allowed in databases. For example, the 1947 US National Map Accuracy Standard specified that 95% of errors should fall below 1/30 inch (0.85 mm) for maps at scales of 1:20 000 and finer (more detailed), and 1/50 inch (0.51 mm) for other maps (coarser, less detailed, levels of granularity than 1:20 000). A convenient rule of thumb is that positions measured from maps are subject to errors of up to 0.5 mm at the scale of the map. Table 6.3 shows the distance on the ground corresponding to 0.5 mm for various common map scales.

A useful rule of thumb is that features on maps are positioned to an accuracy of about 0.5 mm.

6.3.4 The spatial structure of errors

The confusion matrix, or more specifically a single row of the matrix, along with the Gaussian distribution, provide convenient ways of describing the error present in a single

Table 6.3 A useful rule of thumb is that positions measured from maps are accurate to about 0.5 mm on the map. Multiplying this by the scale of the map gives the corresponding distance on the ground

Map scale	Ground distance corresponding to 0.5 mm map distance
1:1250	0.625 m
1:2500	1.25 m
1:5000	2.5 m
1:10 000	5 m
1:24 000	12 m
1:50 000	25 m
1:100 000	50 m
1:250 000	125 m
1:1 000 000	500 m
1:10 000 000	5 000 m

observation of a nominal or interval/ratio measurement respectively. When a GIS is used to respond to a simple query, such as 'tell me the class of soil at this point', or 'what is the elevation here?', then these methods are good ways of describing the uncertainty inherent in the response. For example, a GIS might respond to the first query with the information 'Class A, with a 30% probability of Class C', and to the second query with the information '350 m, with an RMSE of 7 m'. Notice how this makes it possible to describe nominal data as accurate to a percentage, but it makes no sense to describe a DEM, or any measurement on an interval/ratio scale, as accurate to a percentage. For example, we cannot meaningfully say that a DEM is '90% accurate'.

However, many GIS operations involve more than the properties of single points, and this makes the analysis of error much more complex. For example, consider the query 'how far is it from this point to that point?' Suppose the two points are both subject to error of position, because their positions have been measured using GPS units with mean distance errors of 50 m. If the two measurements were taken some time apart, with different combinations of satellites above the horizon, it is likely that the errors are independent of each other, such that one error might be 50 m in the direction of North, and the other 50 m in the direction of South. Depending on the locations of the two points, the error in distance might be as high as +100 m. On the other hand, if the two measurements were made close together in time, with the same satellites above the horizon, it is likely that the two errors would be similar, perhaps 50 m North and 40 m North, leading to an error of only 10 m in the determination of distance. The difference between these two situations can be measured in terms of the degree of *spatial autocorrelation*, or the interdependence of errors at different points in space (Section 4.6).

The spatial autocorrelation of errors can be as important as their magnitude in many GIS operations.

Spatial autocorrelation is also important in errors in nominal data. Consider a field that is known to contain a single crop, perhaps wheat. When seen from above, it is possible to confuse wheat with other crops, so there may be error in the crop type assigned to points in the field. But since the field has only one crop, we know that such errors are likely to be strongly correlated. Spatial autocorrelation is almost always present in errors to some degree, but very few efforts have been made to measure it systematically, and as a result it is difficult to make good estimates of the uncertainties associated with many GIS operations.

An easy way to visualize spatial autocorrelation and interdependence is through animation. Each frame in the animation is a single possible map, or *realization* of the error process. If a point is subject to uncertainty, each realization will show the point in a different possible location, and a sequence of images will show the point shaking around its mean position. If two points have perfectly correlated positional errors, then they will appear to shake in unison, as if they were at the ends of a stiff rod. If errors are only partially correlated, then the system behaves as if the connecting rod were somewhat elastic.

The spatial structure or autocorrelation of errors is important in many ways. DEM data are often used to estimate the slope of terrain, and this is done by comparing elevations at points a short distance apart. For example, if the elevations at two points 10 m apart are 30 m and 35 m respectively, the slope along the line between them is 5/10, or 0.5. (A somewhat more complex method is used in practice, to estimate slope at a point in the x and y directions in a DEM raster, by analyzing the elevations of nine points – the point itself and its eight neighbors. The equations in Section 14.4 detail the procedure.)

Now consider the effects of errors in these two elevation measurements on the estimate of slope. Suppose the first point (elevation 30 m) is subject to an RMSE of 2 m, and consider possible true elevations of 28 m and 32 m. Similarly the second point might have true elevations of 33 m and 37 m. We now have four possible combinations of values, and the corresponding estimates of slope range from $(33 - 32)/10 = 0.1$ to $(37 - 28)/10 = 0.9$. In other words, a relatively small amount of error in elevation can produce wildly varying slope estimates.

The spatial autocorrelation between errors in geographic databases helps to minimize their impacts on many GIS operations.

What saves us in this situation, and makes estimation of slope from DEMs a practical proposition at all, is spatial autocorrelation among the errors. In reality, although DEMs are subject to substantial errors in absolute elevation, neighboring points nevertheless tend to have similar errors, and errors tend to persist over quite large areas. Most of the sources of error in the DEM production process tend to produce this kind of persistence of error over space, including errors due to misregistration of aerial photographs. In other words, errors in DEMs exhibit strong positive spatial autocorrelation.

Another important corollary of positive spatial autocorrelation can also be illustrated using DEMs. Suppose an area of low-lying land is inundated by flooding, and our task is to estimate the area of land affected. We are asked to do this using a DEM, which is known to have an RMSE of 2 m (compare Figure 6.13). Suppose the data points in the DEM are 30 m apart, and preliminary analysis shows that 100 points have elevations below the flood line. We might conclude that the area flooded is the area represented by these 100 points, or 900 × 100 sq m, or 9 hectares. But because of errors, it is possible that some of this area is actually above the flood line (we will ignore the possibility that other areas outside this may also be below the flood line, also because of errors), and it is possible that *all* of the area is above. Suppose the recorded elevation for each of the 100 points is 2 m below the flood line. This is one RMSE (recall that the RMSE is equal to 2 m) below the flood line, and the Gaussian distribution tells us that the chance that the true elevation is actually above the flood line is approximately 16% (see Figure 6.12). But what is the chance that *all* 100 points are actually above the flood line?

Here again the answer depends on the degree of spatial autocorrelation among the errors. If there is none, in other words if the error at each of the 100 points is independent of the errors at its neighbors, then the answer is $(0.16)^{100}$, or 1 chance in 1 followed by roughly 70 zeroes. But if there is strong positive spatial autocorrelation, so strong that all 100 points are subject to exactly the same error, then the answer is 0.16. One way to think about this is in terms of *degrees of freedom*. If the errors are independent, they can vary in 100 independent ways, depending on the error at each point. But if they are strongly spatially autocorrelated, the effective number of degrees of freedom is much less, and may be as few as 1 if all errors behave in unison. Spatial autocorrelation has the effect of reducing the number of degrees of freedom in geographic data below what may be implied by the volume of information, in this case the number of points in the DEM.

Spatial autocorrelation acts to reduce the effective number of degrees of freedom in geographic data.

6.4 U3: Further uncertainty in the analysis of geographic phenomena

6.4.1 Internal and external validation through spatial analysis

In Chapter 1 we identified one remit of GIS as the resolution of scientific or decision-making problems through spatial analysis, which we defined in Section 1.7 as 'the process by which we turn raw spatial data into useful spatial information'. Good science needs secure foundations, yet Sections 6.2 and 6.3 have shown the conception and measurement of many geographic phenomena to be inherently uncertain. How can the outcome of spatial analysis be meaningful if it has such uncertain foundations?

Uncertainties in data lead to uncertainties in the results of analysis.

Once again, there are no easy answers to this question, although we can begin by examining the consequences of accommodating possible errors of positioning, or of aggregating clearly defined units of analysis into artificial geographic individuals (as when people are aggregated by census tracts, or disease incidences are aggregated by county). In so doing, we will illustrate how potential problems might arise, but will not present any definitive solutions – for the simple reason that the truth is inherently uncertain. The conception, measurement, and representation of geographic individuals may distort the outcome of spatial analysis by masking or accentuating apparent variation across space, or by restricting the nature and range of questions that can be asked of the GIS.

There are three ways of dealing with this risk. First, although we can only rarely tackle the *source* of distortion (we are rarely empowered to collect new, completely disaggregate data, for example), we can quantify the way in which it is likely to *operate* (or *propagates*) within the GIS, and can gauge the magnitude of its likely impacts. Second, although we may have to work with aggregated data, GIS allows us to model within-zone spatial distributions in order to ameliorate the worst effects of artificial zonation. Taken together, GIS allows us to gauge the effects of scale and aggregation through simulation of different possible outcomes. This is *internal validation* of the effects of scale, point placement, and spatial partitioning.

Because of the power of GIS to merge diverse data sources, it also provides a means of *external validation* of the effects of zonal averaging. In today's advanced GIService economy (Section 1.5.3), there may be other data sources that can be used to gauge the effects of aggregation upon our analysis. In Chapter 13 we will refine the basic model that was presented in Figure 6.1 to consider how GIS provides a medium for visualizing models of spatial distributions and patterns of homogeneity and heterogeneity.

GIS gives us maximum flexibility when working with aggregate data, and helps us to validate our data with reference to other available sources.

6.4.2 Internal validation: error propagation

The examples of Section 6.3.4 are cases of error propagation, where the objective is to measure the effects of known levels of data uncertainty on the outputs of

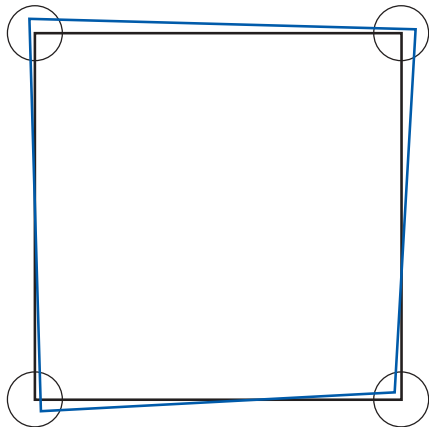


Figure 6.14 Error in the measurement of the area of a square 100 m on each side. Each of the four corner points has been surveyed; the errors are subject to bivariate Gaussian distributions with standard deviations in x and y of 1 m (dashed circles). The blue polygon shows one possible surveyed square (one *realization* of the error model)

GIS operations. We have seen how the spatial structure of errors plays a role, and how the existence of strong positive spatial autocorrelation reduces the effects of uncertainty upon estimates of properties such as slope or area. Yet the cumulative effects of error can also produce impacts that are surprisingly large, and some of the examples in this section have been chosen to illustrate the substantial uncertainties that can be produced by apparently innocuous data errors.

Error propagation measures the impacts of uncertainty in data on the results of GIS operations.

In general two strategies are available for evaluating error propagation. The examples in the previous section were instances in which it was possible to obtain a complete description of error effects based upon known measures of likely error. These enable a complete analysis of uncertainty in slope estimation, and can be applied in the DEM flooding example described in Section 6.3.4. Another example that is amenable to analysis is the calculation of the area of a polygon given knowledge of the positional uncertainties of its vertices.

For example, Figure 6.14 shows a square approximately 100 m on each side. Suppose the square has been surveyed by determining the locations of its four corner points using GPS, and suppose the circumstances of the measurements are such that there is an RMSE of 1 m in both coordinates of all four points, and that errors are independent. Suppose our task is to determine the area of the square. A GIS can do this easily, using a standard algorithm (see Figure 14.9). Computers are precise (in the sense of Section 6.3.2.2 and Box 6.3), and capable of working to many significant digits, so the calculation might be reported by printing out a number to eight digits, such as 10014.603 sq m, or even more. But the number of significant digits will have been determined by the precision of the machine, and not by the accuracy of the determination. Box 6.3 summarized some simple rules for ensuring that the precision used to report a measurement reflects as far as possible its accuracy, and clearly those rules will have been violated if the area is reported to eight digits. But what is the appropriate precision?

In this case we can determine exactly how positional accuracy affects the estimate of area. It turns out that area has an error distribution which is Gaussian, with a standard deviation (RMSE) in this case of 200 sq m – in other words, each attempt to measure the area will give a different result, the variation between them having a standard deviation of 200 sq m. This means that the five rightmost digits in the estimate are spurious, including two digits to the left of the decimal point. So if we were to follow the rules of Box 6.3, we would print 10 000 rather than 10014.603 (note the problem with standard notation here, which does not let us omit digits to the left of the decimal point even if they are spurious, and so leaves some uncertainty about whether the tens and units digits are certain or not – and note also the danger that if the number is printed as an integer it may be interpreted as exactly the whole number). We can also turn the question around and ask how accurately the points would have to be measured to justify eight digits, and the answer is approximately 0.01 mm, far beyond the capabilities of normal surveying practice.

Analysis can be applied to many other kinds of GIS analysis, and Gerard Heuvelink (Box 6.4) discusses several further examples in his excellent text on error propagation in GIS. But analysis is a difficult strategy when spatial autocorrelation of errors is present, and many problems of error propagation in GIS are not amenable to analysis. This has led many researchers to explore a more general strategy of simulation to evaluate the impacts of uncertainty on results.

In essence, simulation requires the generation of a series of realizations, as defined earlier, and it is often called Monte Carlo simulation in reference to the realizations that occur when dice are tossed or cards are dealt in various games of chance. For example, we could simulate error in a single measurement from a DEM by generating a series of numbers with a mean equal to the measured elevation, and a standard deviation equal to the known RMSE, and a Gaussian distribution. Simulation uses everything that is known about a situation, so if any additional information is available we would incorporate it in the simulation. For example, we might know that elevations must be whole numbers of meters, and would simulate this by rounding the numbers obtained from the Gaussian distribution. With a mean of 350 m and an RMSE of 7 m the results of the simulation might be 341, 352, 356, 339, 349, 348, 355, 350, . . .

Simulation is an intuitively simple way of getting the uncertainty message across.

Because of spatial autocorrelation, it is impossible in most circumstances to think of databases as decomposable into component parts, each of which can be independently disturbed to create alternative realizations, as in the previous example. Instead, we have to think of

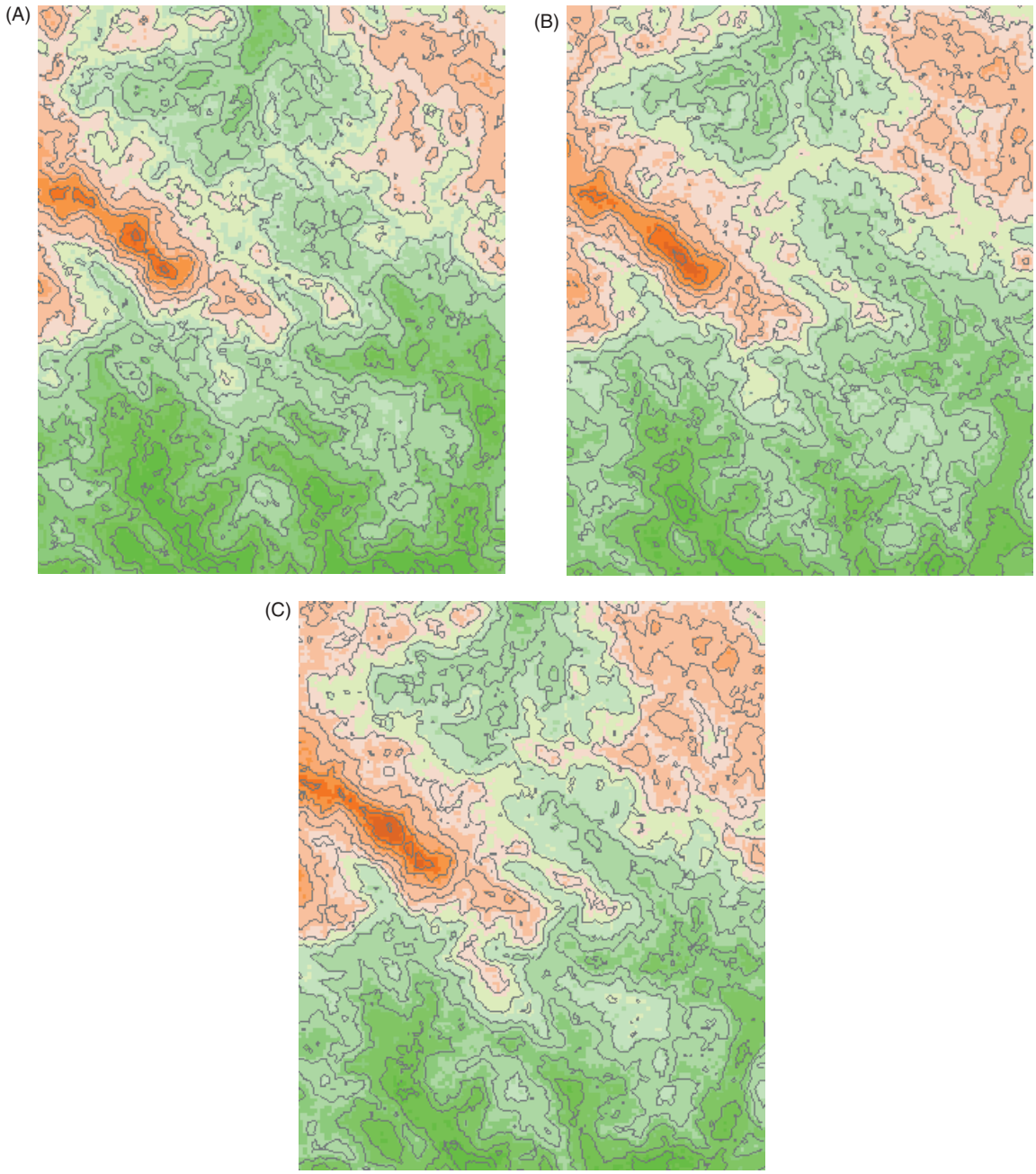


Figure 6.15 Three realizations of a model simulating the effects of error on a digital elevation model. The three datasets differ only to a degree consistent with known error. Error has been simulated using a model designed to replicate the known error properties of this dataset – the distribution of error magnitude, and the spatial autocorrelation between errors. (Reproduced by permission of Ashton Shortridge.)

the entire database as a realization, and create alternative realizations of the database's contents that preserve spatial autocorrelation. Figure 6.15 shows an example, simulating the effects of uncertainty on a digital elevation model. Each of the three realizations is a complete map, and the simulation process has faithfully replicated the strong correlations present in errors across the DEM.

6.4.3 Internal validation: aggregation and analysis

We have seen already that a fundamental difference between geography and other scientific disciplines is that the definition of its objects of study is only rarely unambiguous and, in practice, rarely precedes

Biographical Box **6.4**

Gerard Heuvelink, geostatistician

Figure 6.16 Gerard Heuvelink, geostatistician

Understanding the limitations of spatial data and spatial models is essential both for managing environmental systems effectively and for encouraging safe use of GIS. Gerard Heuvelink (Figure 6.16) of the Wageningen University and Research Centre, the Netherlands, has dedicated much of his scientific career to this end, through statistical modeling of the uncertainty in spatial data and analysis of the ways in which uncertainty is propagated through GIS.

Trained as a mathematician, Gerard undertook a Ph.D. in Physical Geography working with Professor Peter Burrough of Utrecht University. His 1998 research monograph *Error Propagation in Environmental Modelling with GIS* has subsequently become the key reference in spatial uncertainty analysis. Gerard is firmly of the view that GI scientists should pay more attention to statistical validation and exploration of data, and he is actively involved in a series of symposia on 'Spatial Accuracy Assessment in Natural Resources and Environmental Sciences' (**www.spatial-accuracy.org**).

Gerard's background in mathematics and statistics has left him with the view that spatial uncertainty analysis requires a sound statistical basis. In his view, understanding uncertainty in the position of spatial objects and in their attribute values entails use of probability distribution functions, and measuring spatial autocorrelation (Section 4.6) with uncertainties in other objects in spatial databases. He says: 'I remain disappointed with the amount of progress made in understanding the fundamental problems of uncertainty over the last fifteen years. We have moved forward in the sense that we now have a broader view of various aspects of spatial data quality. The 1980s and early 1990s were dedicated to technical topics such as uncertainty propagation in map overlay operations and the development of statistical models for representing positional uncertainty. More recently the research community has addressed a range of user-centric topics, such as visualization and communication of uncertainty, decision making under uncertainty and the development of error-aware GIS. But these developments do not hide the fact that we still do not have the statistical basics right. Until this is achieved, we run the risk of building elaborate representations on weak and uncertain foundations.'

Gerard and co-worker James Brown from the University of Amsterdam are working to contribute to filling this gap, by developing a general probabilistic framework for characterizing uncertainty in the positions and attribute values of spatial objects.

our attempts to measure their characteristics. In socio-economic GIS applications, these objects of study (geographic individuals) are usually aggregations, since the spaces that human individuals occupy are geographically unique, and confidentiality restrictions usually dictate that uniquely attributable information must be anonymized in some way. Even in natural-environment applications, the nature of sampling in the data collection process (Section 4.4) often makes it expedient to collect data pertaining to aggregations of one kind or another. Thus in socio-economic and environmental applications alike, the measurement of geographic individuals is unlikely to be determined with the end point of particular spatial-analysis applications in mind. As a consequence, we cannot be certain in ascribing even dominant characteristics *of* areas to true individuals or point locations *in* those areas. This source of uncertainty is known as the *ecological fallacy*, and has long bedevilled the analysis of spatial distributions (the opposite of ecological fallacy is atomistic fallacy, in which the individual is considered in isolation from his or her environment). This is illustrated in Figure 6.17.

Inappropriate inference from aggregate data about the characteristics of individuals is termed the ecological fallacy.

We have also seen that the *scale* at which geographic individuals are conceived conditions our measures of association between the mosaic of zones represented within a GIS. Yet even when scale is fixed, there is a multitude of ways in which basic areal units of analysis can be *aggregated* into zones, and the requirement of spatial contiguity represents only a weak constraint upon the huge combinatorial range. This gives rise to the related *aggregation* or *zonation* problem, in which different combinations of a given number of geographic individuals into coarser-scale areal units can yield widely different results. In a classic 1984 study, the geographer Stan

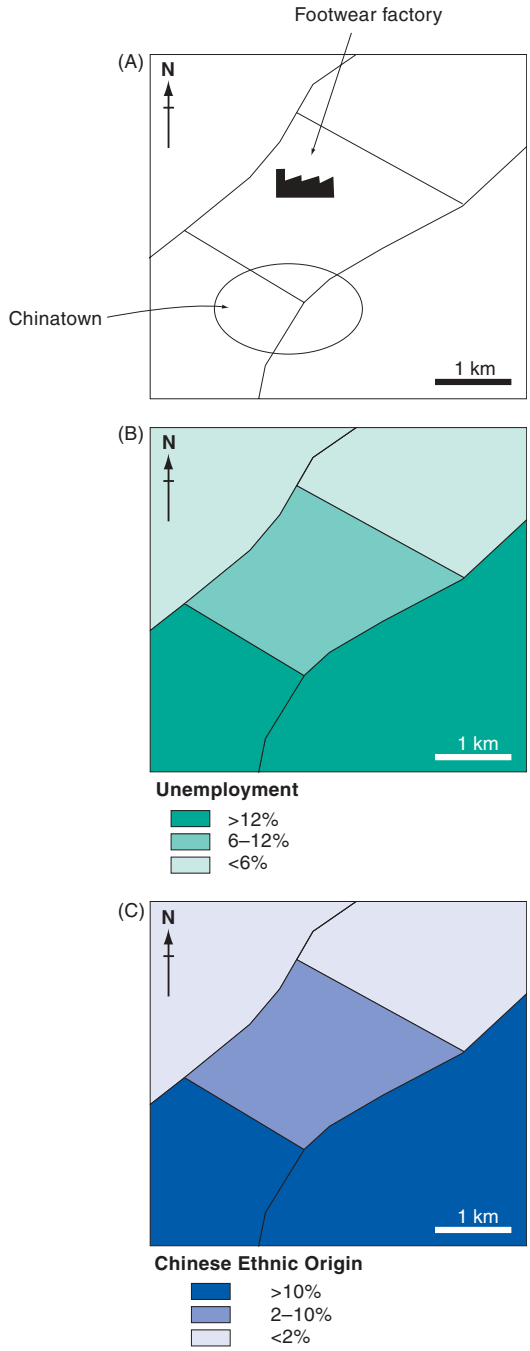


Figure 6.17 The problem of ecological fallacy. Before it closed down, the Anytown footwear factory drew its labor from blue-collar neighborhoods in its south and west sectors. Its closure led to high local unemployment, but not amongst the residents of Chinatown, who remain employed in service industries. Yet comparison of choropleth maps B and C suggests a spurious relationship between Chinese ethnicity and unemployment

Openshaw applied correlation and regression analysis to the attributes of a succession of zoning schemes. He demonstrated that the constellation of elemental zones within aggregated areal units could be used to manipulate the results of spatial analysis to a wide range of quite different prespecified outcomes. These numerical experiments have some sinister counterparts in the real world, the most notorious example of which is the political gerrymander of 1812 (see Section 14.3.2). Chance or design might therefore conspire to create apparent spatial distributions which are unrepresentative of the scale and configuration of real-world geographic phenomena. The outcome of multivariate spatial analysis is also similarly sensitive to the particular zonal scheme that is used. Taken together, the effects of scale and aggregation are generally known as the *Modifiable Areal Unit Problem* (MAUP).

The ecological fallacy and the MAUP have long been recognized as problems in applied spatial analysis and, through the concept of spatial autocorrelation (Section 4.3), they are also understood to be related problems. Increased technical capacity for numerical processing and innovations in scientific visualization have refined the quantification and mapping of these measurement effects, and have also focused interest on the effects of within-area spatial distributions upon analysis.

6.4.4 External validation: data integration and shared lineage

Goodchild and Longley (1999) use the term *concatenation* to describe the integration of two or more different data sources, such that the contents of each are accessible in the product. The polygon overlay operation that will be discussed in Section 14.4.3, and its field-view counterpart, is one simple form of concatenation. The term *conflation* is used to describe the range of functions that attempt to overcome differences between datasets, or to merge their contents (as with rubber-sheeting: see Section 9.3.2.3). Conflation thus attempts to replace two or more versions of the same information with a single version that reflects the pooling, or weighted averaging, of the sources.

The individual items of information in a single geographic dataset often share lineage, in the sense that more than one item is affected by the same error. This happens, for example, when a map or photograph is registered poorly, since all of the data derived from it will have the same error. One indicator of shared lineage is the persistence of error – because all points derived from the same misregistration will be displaced by the same, or a similar, amount. Because neighboring points are more likely to share lineage than distant points, errors tend to exhibit strong positive spatial autocorrelation.

Conflation combines the information from two data sources into a single source.

When two datasets that share no common lineage are concatenated (for example, they have not been subject to the same misregistration), then the relative positions of objects inherit the absolute positional errors of both, even over the shortest distances. While the shapes of objects in each dataset may be accurate, the relative locations

of pairs of neighboring objects may be wildly inaccurate when drawn from different datasets. The anecdotal history of GIS is full of examples of datasets which were perfectly adequate for one application, but which failed completely when an application required that they be merged with some new dataset that had no common lineage. For example, merging GPS measurements of point positions with streets derived from the US Bureau of the Census TIGER files may lead to surprises where points appear on the wrong sides of streets. If the absolute positional accuracy of a dataset is 50 m, as it is with parts of the TIGER database, points located less than 50 m from the nearest street will frequently appear to be misregistered.

Datasets with different lineages often reveal unsuspected errors when overlaid.

Figure 6.18 shows an example of the consequences of overlaying data with different lineages. In this case, two datasets of streets produced by different commercial vendors using their own process fail to match in position by amounts of up to 100 m, and also fail to match in the names of many streets, and even the existence of streets.

The integrative functionality of GIS makes it an attractive possibility to generate multivariate indicators from diverse sources. Yet such data are likely to have been collected at a range of different scales, and for a range of areal units as diverse as census tracts, river catchments, land ownership parcels, travel-to-work areas, and market research surveys. Established procedures of statistical inference can only be used to reason from representative samples to the populations from which they were drawn. Yet these procedures do not regulate the assignment of inferred values to (usually smaller) zones, or their apportionment to *ad hoc* regional categorizations. There is an emergent tension within the socio-economic realm, for there is a limit to the uses of inferences drawn from conventional, scientifically valid data sources which are frequently out-of-date, zonally coarse, and irrelevant to what is happening in modern societies. Yet the alternative of using new rich sources of marketing data may be profoundly unscientific in its inferential procedures.

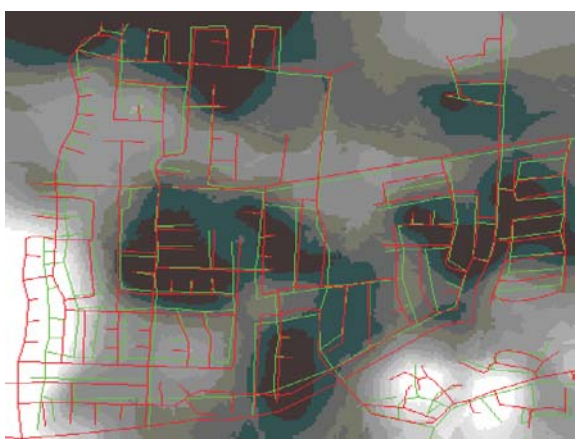


Figure 6.18 Overlay of two street databases for part of Goleta, California, USA. The red and green lines fail to match by as much as 100 m. Note also that in some cases streets in one dataset fail to appear in the other, or have different connections. The background is dark where the fit is best and white where it is poorest (it measures the average distance locally between matched intersections)

6.4.5 Internal and external validation; induction and deduction

Reformulation of the MAUP into a *geocomputational* (Box 1.9 and Section 16.1) approach to zone design has been one of the key contributions of geographer Stan Openshaw. Central to this is inductive use of GIS to seek patterns through repeated scaling and aggregation experiments, alongside much better external validation, deduced using the multitude of new datasets that are a hallmark of the information age.

The Modifiable Areal Unit Problem can be investigated through simulation of large numbers of alternative zoning schemes.

Neither of these approaches, used in isolation, is likely to resolve the uncertainties inherent in spatial analysis. Zone design experiments are merely playing with the MAUP, and most of the new sources of external validation are unlikely to sustain full scientific scrutiny, particularly if they were assembled through non-rigorous survey designs. The conception and measurement of elemental zones, the geographic individuals, may be *ad hoc*, but they are rarely wholly random either. Can our recognition and understanding of the empirical effects of the MAUP help us to neutralize its effects? Not really. In measuring the distribution of all possible zonally averaged outcomes ('simple random zoning' in analogy to simple random sampling in Section 4.4), there is no tenable analogy with the established procedures of statistical inference and its concepts of precision and error. And even if there were, as we have seen in Section 4.7 there are limits to the application of classic statistical inference to spatial data.

Zoning seems similar to sampling, but its effects are very different.

The way forward seems to be to complement our new-found abilities to customize zoning schemes in GIS with external validation of data and clearer application-centered thinking about the likely degree of within-zone heterogeneity that is concealed in our aggregated data. In this view, MAUP will disappear if GIS analysts understand the particular areal units that they wish to study. There is also a sense here that resolution of the MAUP requires acknowledgment of the uniqueness of places. There is also a practical recognition that the areal objects of study are ever-changing, and our perceptions of what constitutes their appropriate definition will change. And finally, within the socio-economic realm, the act of defining zones can also be self-validating if the allocation of individuals affects the interventions they receive (be they a mail-shot about a shopping opportunity or aid under an areal policy intervention). Spatial discrimination affects

Applications Box **6.5**

Uncertainty and town center definition

Although the locus of retail activity in many parts of the US long ago shifted to suburban locations, traditional town centers ('downtowns') remain vibrant and are cherished in most of the rest of the world. Indeed, many nations vigorously defend existing retail centers and through various planning devices seek to regulate 'out of center' development.

Therefore, many interests in the planning and retail sectors are naturally concerned with learning the precise extent of existing town centers. The pressure to devise standard definitions across its national jurisdiction led the UK's central government planning agency (the Office of the Deputy Prime Minister) to initiate a five-year research program to define and monitor changes in the shape, form, and internal geography of town centers across the nation. The work has been based at the Centre for Advanced Spatial Analysis (CASA) at University College London.

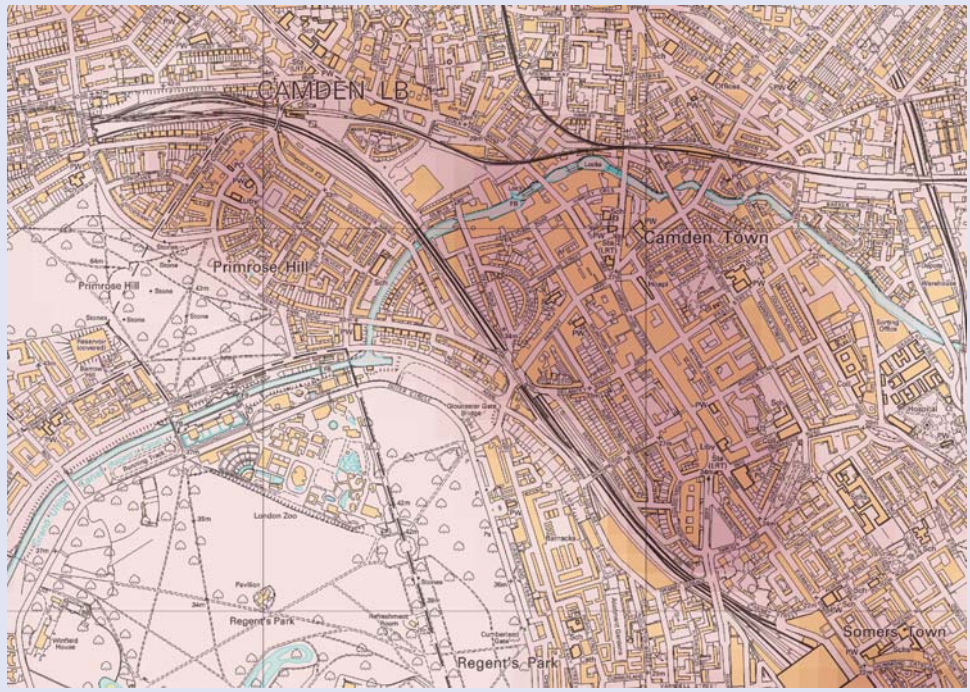
Town centers present classic examples of geographic phenomena with uncertain boundaries. Moreover the extent of any given town center is likely to change over time – for example, in response to economic fortunes consequent upon national business cycles. Candidate *indicator* variables of town centeredness might include tall buildings, pedestrian traffic, high levels of retail employment, and high retail floorspace figures.

After a consultation period with user groups (in the spirit of public participation in GIS, PPGIS: see Section 13.4) a set of the most pertinent indicators was agreed (the *conception* stage). These indicator variables measured retail and hospitality industry employment, shop and office floorspace, and retail, leisure, and service employment. The indicator measures were standardized, weighted, and summed into a summary index *measure*. This measure, mapped for all town centers, was the principal deliverable

(A)

Figure 6.19 (A) Camden Town Center, London (Reproduced by permission of Jamie G.A. Quinn); (B) a data surface representing the index of town center activity (the darker shades of red indicate greater levels of retail activity); and (C) the Camden Town Center report: Camden Town center boundary is blue, whilst the orange lines denote the retail core of Camden and of nearby town centers – the darker shades of red again indicate greater levels of retail activity

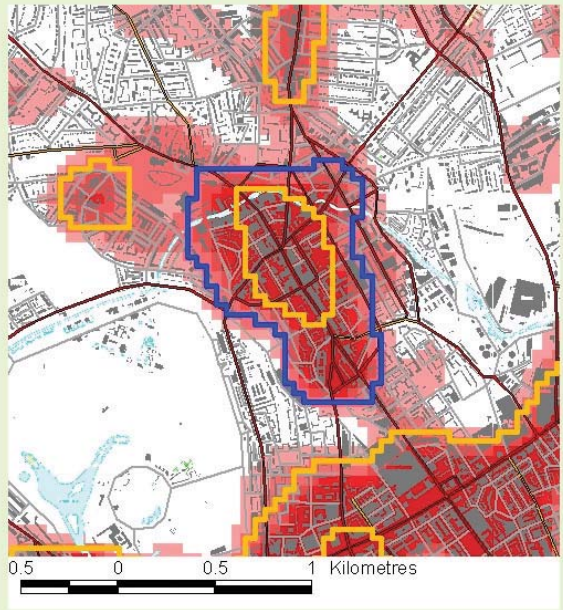
(B)



(C)

Town Centres

Area of Town Centre Activity: Camden Town, LB Camden



Employment (Persons)	
Convenience retail	1,143
Comparison retail	1,008
Service retail	1,750*
Offices	5,852
Civic and Public Administration	412
Restaurants & Licensed Premises	1,335
Arts, Culture and Entertainment	408

Turnover (£000s)	
Convenience retail	130,336
Comparison retail	89,475
Service retail	Disclosive
Restaurants & Licensed Premises	40,015
Arts, Culture and Entertainment	35,399

Floorspace (Sq m)	
A1	55,496
A2	6,726
A3	12,967
Retail	81,278
Offices	147,663

Back to Camden list
Back to Camden map

Printer friendly Version

* Indicates possible unreliability in the marked data

The map background is based upon the OS map by the Office of the Deputy Prime Minister (ODPM) with permission of Ordnance Survey on behalf of The Controller of Her Majesty's Stationery Office, © Crown Copyright. All rights reserved. Unauthorised reproduction infringes Crown Copyright and may lead to prosecution or civil proceedings. Licence Number: GD272671.

Production date: 21 Jun 2002

Figure 6.19 (*continued*)

of the research. After further consultation, the CASA team chose to represent the 'degree of town centeredness' as a field variable. This choice reflected various priorities, including the need to maintain confidentiality of those data that were not in the public domain, an attempt to avoid the worst effects of the MAUP, and the need to communicate the rather complex concept of the 'degree of town centeredness' to an audience of 'spatially *un*aware professionals' (see Section 1.4.3.2). The datasets used in the projects each represent populations (not samples), and so kernel density estimation (Section 14.4.4.4) was used to create the composite surfaces: the size of the kernel was subjectively set at 300 meters, because of the resolution of the data and on the basis of the empirical observation that this is the maximum distance that most shoppers are prepared to walk.

An example of the composite surface 'index of town center activity' for the town center of Camden, London (Figure 6.19A) is shown in Figure 6.19B. For reasons to be explored in our discussion of geovisualization (Chapter 13), most users prefer maps to have crisp and not graduated or uncertain boundaries. Thus crisp boundaries were subsequently interpolated as shown in Figure 6.19C.

spatial behavior, and so the principles of zone design are of much more than academic interest.

Many of issues of uncertainty in conception, measurement, representation, and analysis come together in the definition of town center boundaries (see Box 6.5).

6.5 Consolidation

Uncertainty is certainly much more than error. Just as the amount of available digital data and our abilities to process them have developed, so our understanding of the quality of digital depictions of reality has broadened. It is one of the supreme ironies of contemporary GIS that as we accrue more and better data and have more computational power at our disposal, so we seem to become more uncertain about the quality of our digital representations and the adequacy of our areal units of analysis. Richness of representation and computational power only make us more aware of the range and variety of established uncertainties, and challenge us to integrate new ones. The only way beyond this impasse is to advance hypotheses about the structure of data, in a spirit of humility rather than conviction. But this implies greater *a priori* understanding about the structure in spatial as well as attribute data. There are some general rules to guide us here and statistical measures such as spatial autocorrelation provide further structural clues (Section 4.3). The developing range of context-sensitive spatial analysis methods provides a bridge between such general statistics and methods of specifying place or local (natural) environment. Geocomputation helps too, by allowing us to assess the sensitivity of outputs to inputs, but, unaided, is unlikely to provide any unequivocal best solution. The fathoming of uncertainty requires a combination of the cumulative development of *a priori* knowledge (we should expect scientific research to be cumulative in its findings), external validation of data sources, and inductive generalization in the fluid, eclectic data-handling environment that is contemporary GIS.

More pragmatically, here are some rules for how to live with uncertainty: First, since there can be no such thing as perfectly accurate GIS analysis, it is essential to acknowledge that uncertainty is inevitable. It is better to take a positive approach, by learning what one can about uncertainty, than to pretend that it does not exist. To behave otherwise is unconscionable, and can also be very expensive in terms of lawsuits, bad decisions, and the unintended consequences of actions (see Chapter 19).

Second, GIS analysts often have to rely on others to provide data, through government-sponsored mapping programs like those of the US Geological Survey or the UK Ordnance Survey, or commercial sources. Data should never be taken as the truth, but instead it is essential to assemble all that is known about the quality of data, and to use this knowledge to assess whether, actually, the data are fit for use. Metadata (Section 11.2.1) are designed specifically for this purpose, and will often include assessments of quality. When these are not present, it is worth spending the extra effort to contact the creators of the data, or other people who have tried to use them, for advice on quality. Never trust data that have not been assessed for quality, or data from sources that do not have good reputations for quality.

Third, the uncertainties in the outputs of GIS analysis are often much greater than one might expect given knowledge of input uncertainties, because many GIS processes are highly non-linear. Other processes dampen uncertainty, rather than enhance it. Given this, it is important to gain some impression of the impacts of input uncertainty on output.

Fourth, rely on multiple sources of data whenever you can. It may be possible to obtain maps of an area at several different scales, or to obtain several different vendors' databases. Raster and vector datasets are often complementary (e.g., combine a remotely sensed image with a topographic map). Digital elevation models can often be augmented with spot elevations, or GPS measurements.

Finally, be honest and informative in reporting the results of GIS analysis. It is safe to assume that GIS designers will have done little to help in this respect – results will have been reported to high apparent precision, with more significant digits than are justified by actual accuracy, and lines will have been drawn on maps with widths that reflect relative importance, rather than uncertainty of position. It is up to you as the user to redress this imbalance, by finding ways of communicating what you know about accuracy, rather than relying on the GIS to do so. It is wise to put plenty of caveats into reported results, so that they reflect what you believe to be true, rather than what the GIS appears to be saying. As someone once said, when it comes to influencing people 'numbers beat no numbers every time, whether or not they are right', and the same is certainly true of maps (see Chapters 12 and 13).

Questions for further study

1. What tools do GIS designers build into their products to help users deal with uncertainty? Take a look at your favorite GIS from this perspective. Does it allow you to associate metadata about data quality with datasets? Is there any support for propagation of uncertainty? How does it determine the number of significant digits when it prints numbers? What are the pros and cons of including such tools?
2. Using aggregate data for Iowa counties, Openshaw found a strong positive correlation between the proportion of people over 65 and the proportion who were registered voters for the Republican party. What if anything does this tell us about the tendency for older people to register as Republicans?
3. Find out about the five components of data quality used in GIS standards, from the information available at **www.fdgc.gov**. How are the five components applied in the case of a standard mapping agency data product, such as the US Geological Survey's Digital Orthophoto Quarter–Quadrangle program? (Search the Web for the appropriate documents.)
4. You are a senior retail analyst for Safemart, which is contemplating expansion from its home US state to three others in the Union. Assess the relative merits of your own company's store loyalty card data (which you can assume are similar to those collected by any retail chain with which you are familiar) and of data from the 2000 Census in planning this strategic initiative. Pay, particular attention to issues of survey content, the representativeness of population characteristics, and problems of scale and aggregation. Suggest ways in which the two data sources might complement one another in an integrated analysis.

Further reading

Burrough P.A. and Frank A.U. (eds) 1996 *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis.

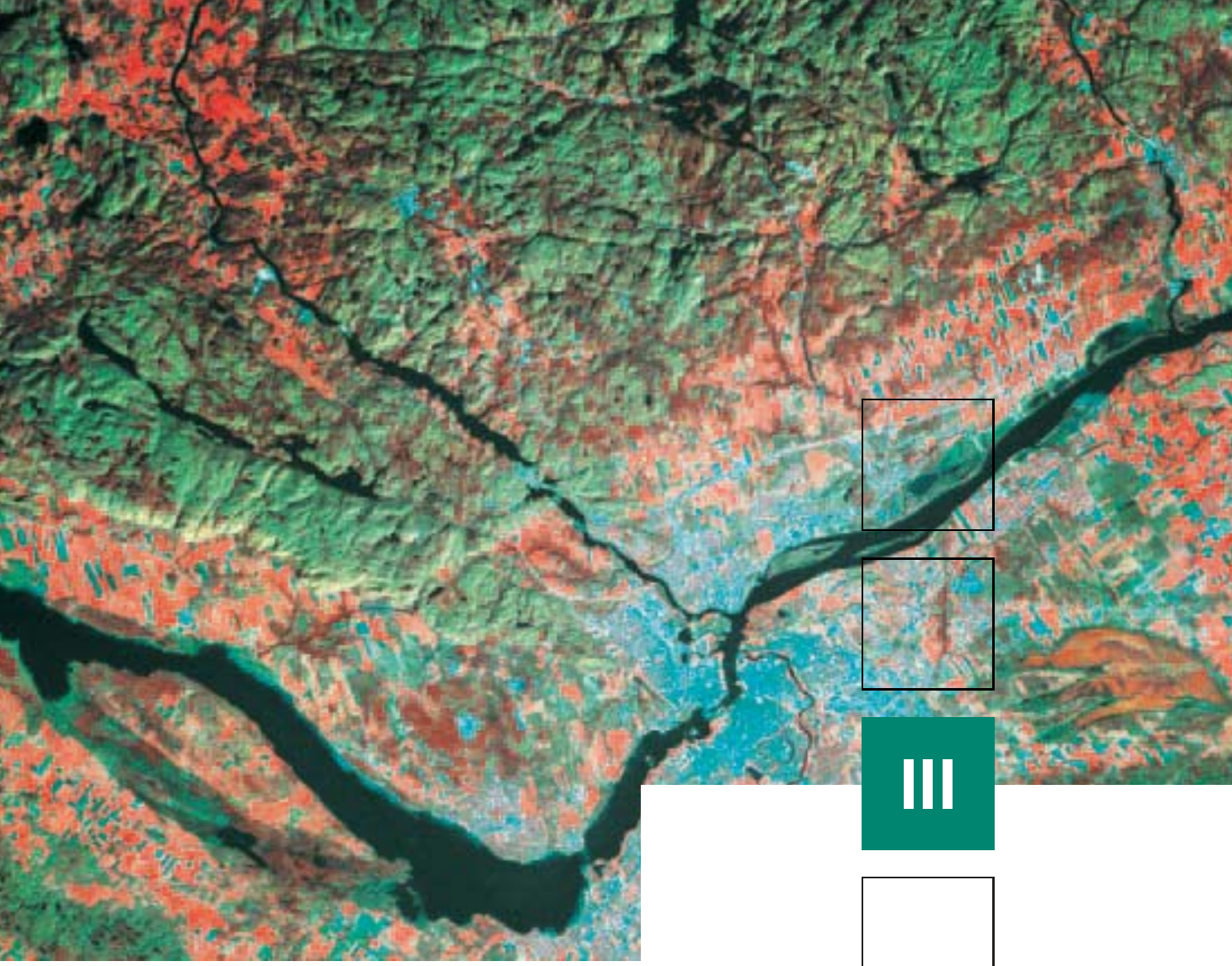
Fisher P.F. 2005 'Models of uncertainty in spatial data.' In Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, N.J.: Wiley, pp. 191–205.

Goodchild M.F. and Longley P.A. 2005 'The future of GIS and spatial analysis.' In Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, N.J.: Wiley, pp. 567–80.

Heuvelink G.B.M. 1998 *Error Propagation in Environmental Modelling with GIS*. London: Taylor and Francis.

Openshaw S. and Alvanides S. 2005 'Applying geocomputation to the analysis of spatial distributions'. In Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications* (abridged edition). Hoboken, N.J.: Wiley, pp. 267–282.

Zhang J.X. and Goodchild M.F. 2002 *Uncertainty in Geographical Information*. New York: Taylor and Francis.



III

Techniques

7 *GIS Software*
8 *Geographic data modeling*
9 *GIS data collection*
10 *Creating and maintaining geographic databases*
11 *Distributed GIS*

7 *GIS Software*

GIS software is the processing engine and a vital component of an operational GIS. It is made up of integrated collections of computer programs that implement geographic processing functions. The three key parts of any GIS software system are the user interface, the tools (functions), and the data manager. All three parts may be located on a single computer or they may be spread over multiple machines in a departmental or enterprise configuration. Four main types of computer system architecture configurations are used to build operational GIS implementations: desktop, client-server, centralized desktop, and centralized server. There are many different types of GIS software and this chapter uses five categories to organize the discussion: desktop, server (including Internet), developer, hand-held, and other. The market leading commercial GIS software vendors are ESRI, Intergraph, Autodesk, and GE Energy (Smallworld).

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will be able to:

- Understand the architecture of GIS software systems, specifically
  - Organization by project, department, or enterprise;
  - The three-tier architecture of software systems (graphical user interface; tools, and data access);
- Describe the process of GIS customization;
- Describe the main types of commercial software
  - Desktop
  - Server
  - Developer
  - Hand-held
  - Other;
- Outline the main types of commercial GIS software products currently available.

7.1 Introduction

In Chapter 1, the four technical parts of a geographic information system were defined as the network, the hardware, the software, and the data, which together functioned with reference to people and the procedural structures within which people work (Section 1.5). This chapter is concerned with GIS software, the geographic processing engine of a complete, working GIS. The functionality or capabilities of GIS software will be discussed later in this book (especially in Chapters 12–16). The focus here is on the different ways in which these capabilities are realized in GIS software products and implemented in operational GIS.

This chapter takes a fairly narrow view of GIS software, concentrating on systems with a range of generic capabilities to collect, store, manage, query, analyze, and present geographic information. It excludes atlases, simple graphics and mapping systems, route finding software, simple location-based services, image processing systems, and spatial extensions to database management systems (DBMS), which are not true GIS as defined here. The discussion is also restricted to GIS software products – well-defined collections of software and accompanying documentation, install scripts, etc. – that are subject to multi-versioned release control. By definition it excludes specific-purpose utilities, unsupported routines, and ephemeral codebases.

Earlier chapters, especially Chapter 3, introduced several fundamental computer concepts, including digital representations, data, and information. Two further concepts need to be introduced here. Programs are collections of instructions that are used to manipulate digital data in a computer. System software programs, such as a computer operating system, are used to support application software – the programs with which end users interact. Integrated collections of application programs are referred to as software packages or systems (or just software for short).

Software can be distributed to the market in several different ways. The dominant form of distribution is the sale of commercial-off-the-shelf (COTS) software products on hard copy media (CD/DVD). GIS software products of this type are developed with a view to providing users with a consistent and coherent model for interacting with geographic data. The product will usually comprise an integrated collection of software programs, an install script, on-line help files, sample data and maps, documentation and an associated website. Alternative distribution models that are becoming increasingly prevalent include *shareware* (usually intended for sale after a trial period), *liteware* (shareware with some capabilities disabled), *freeware* (free software but with copyright restrictions), *public domain software* (free with no restrictions), and *open source* software (where the *source code* is provided and users agree not to limit the distribution of improvements). The Internet is becoming the main medium for software distribution.

GIS software packages provide a unified approach to working with geographic information.

GIS software vendors – the companies that design, develop, and sell GIS software – build on top of basic computer operating system capabilities such as security, file management, peripheral drivers (controllers), printing, and display management. GIS software is constructed on these foundations to provide a controlled environment for geographic information collection, management, analysis, and interpretation. The unified architecture and consistent approach to representing and working with geographic information in a GIS software package aim to provide users with a standardized approach to GIS.

7.2 The evolution of GIS software

In the formative GIS years, GIS software consisted simply of collections of computer routines that a skilled programmer could use to build an operational GIS. During this period each and every GIS was unique in terms of its capabilities, and significant levels of resource

were required to create a working system. As software engineering techniques advanced and the GIS market grew in the 1970s and 1980s, demand increased for higher-level applications with a standard user interface. In the late 1970s and early 1980s the standard means of communicating with a GIS was to type in command lines. User interaction with a GIS entailed typing instructions to, for example, draw a topographic map, query the attributes of a forest stand object, or summarize the length of highways in a project area. Essentially, a GIS software package was a toolbox of geoprocessing operators or commands that could be applied to datasets to create new datasets. For example, three area-based data layers *Soil*, *Slope* and *Vegetation* could be combined using an overlay function to create an *IntegratedTerrainUnit* dataset.

To make the software easier to use and more generic, there were two key developments in the late 1980s. First, command line interfaces were supplemented and eventually largely replaced by graphical user interfaces (GUIs). These menu-driven, form-based interfaces greatly simplified user interaction with a GIS. Second, a customization capability was added to allow specific-purpose applications to be created from the generic toolboxes. Software developers and advanced technical users could make calls using a high level programming language (such as Visual Basic or Java) to published application programming interfaces (APIs) that exposed key functions. Together these stimulated enormous interest in GIS, and led to much wider adoption and expansion into new areas. In particular, the ability to create custom application solutions allowed developers to build focused applications for end users in specific market areas. This led to the creation of GIS applications specifically tailored to the needs of major markets (e.g., government, utilities, military, and environment). New terms were developed to distinguish these subtypes of GIS software: planning information systems, automated mapping/facility management (AM/FM) systems, land information systems, and more recently, location-based services systems.

In the last few years a new method of software interaction has evolved that allows software systems to communicate over the Web using a Web services paradigm. A Web service is an application that exposes its functions via a well-defined published interface that can be accessed over the Web from another program or Web service. This new software interaction paradigm will allow geographically distributed GIS functions to be linked together to create complete GIS applications. For example, a market analyst who wants to determine the suitability of a particular site for locating a new store can start a small browser-based program on their desktop computer that links to remote services over the Web that provide access to the latest population census and geodemographics data, as well as analytical models. Although these data and programs are remotely hosted and maintained they can be used for site suitability analysis as though they were resident on the market analyst's desktop. Chapter 11 explores Web services in more depth in the context of distributed GIS.

The GIS software of today still embodies the same principles of an easy-to-use menu-driven interface and a customization capability, but can now be distributed over the Web.

7.3 Architecture of GIS software

7.3.1 Project, departmental, and enterprise GIS

Usually, GIS is first introduced into organizations in the context of a single, fixed-term project (Figure 7.1). The technical components (network, hardware, software, and data) of an operational GIS are assembled for the duration of the project, which may be from several months to a few years. Data are collected specifically for the project and typically little thought is given to reuse of software, data, and human knowledge. In larger organizations, multiple projects may run one after another or even in parallel. The 'one off' nature of the projects, coupled with an absence of organizational vision, often leads to duplication, as each project develops using different hardware, software, data, people, and procedures. Sharing data and experience is usually a low priority.

As interest in GIS grows, to save money and encourage sharing and resource reuse, several projects in the same department may be amalgamated. This often leads to the creation of common standards, development of a focused GIS team, and procurement of new GIS capabilities. Yet it is also quite common for different departments to have different GIS software and data standards.

As GIS becomes more pervasive, organizations learn more about it and begin to become dependent on it. This leads to the realization that GIS is a useful way to structure many of the organization's assets, processes, and workflows. Through a process of natural growth, and possibly further major procurement (e.g., purchase of upgraded hardware, software, and data), GIS gradually becomes accepted as an important enterprise-wide information system. At this point GIS standards are accepted across multiple departments, and resources to support and manage the GIS are often centrally-funded and managed. A fourth type of societal implementation has additionally been identified in which hundreds or thousands of users become engaged in GIS and connected by a network. Today there are only a few examples of societal implementations with perhaps the best being the State of Qatar in the Middle East where more than 16 government departments have joined together to create a comprehensive and nationwide GIS with thousands of users.

7.3.2 The three-tier architecture

From an information systems perspective there are three key parts to a GIS: the user interface, the tools, and

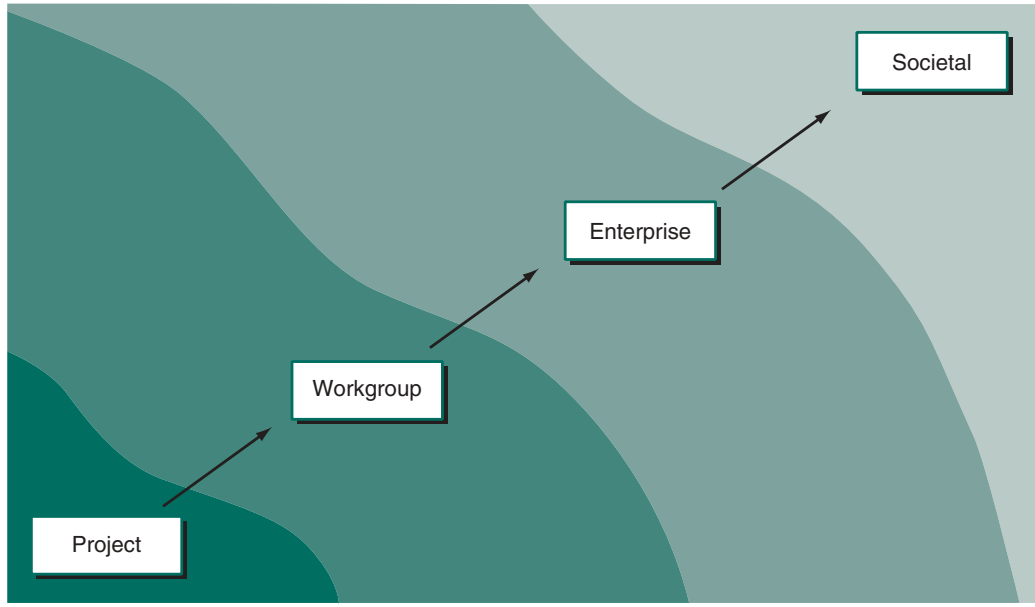


Figure 7.1 Types of GIS implementation

the data management system (Figure 7.2). The user's interaction with the system is via a graphical user interface (GUI), an integrated collection of menus, tool bars, and other controls. The GUI provides access to the GIS tools. The toolset defines the capabilities or functions that the GIS software has available for processing geographic data. The data are stored in files or databases organized by data management software. In standard information system terminology this is a three-tier architecture with the three tiers being called: presentation, business logic, and data server. Each of these software tiers is required to perform different types of independent tasks. The presentation tier must be adept at rendering (displaying) and interacting with graphic objects. The business logic tier is responsible for performing compute-intensive operations such as data overlay processing and raster analysis. It is here also that the GIS data model logic is implemented. The data server tier must import and export data and service requests for subsets of data from a database or file system. In order to maximize system performance it is useful to optimize hardware and operating systems settings differently for each of these types of task. For example, rendering maps requires large amounts of memory and fast CPU clock speeds, whereas database queries need fast disks and buses for moving large amounts of data around. By placing each tier on separate computers some tasks can be performed in parallel and greater overall system scalability can be achieved.

GIS software systems deal with user interfaces, tools, and data management.

Four types of computer system architecture configurations are used to build operational GIS implementations: desktop, client-server, centralized desktop, and centralized

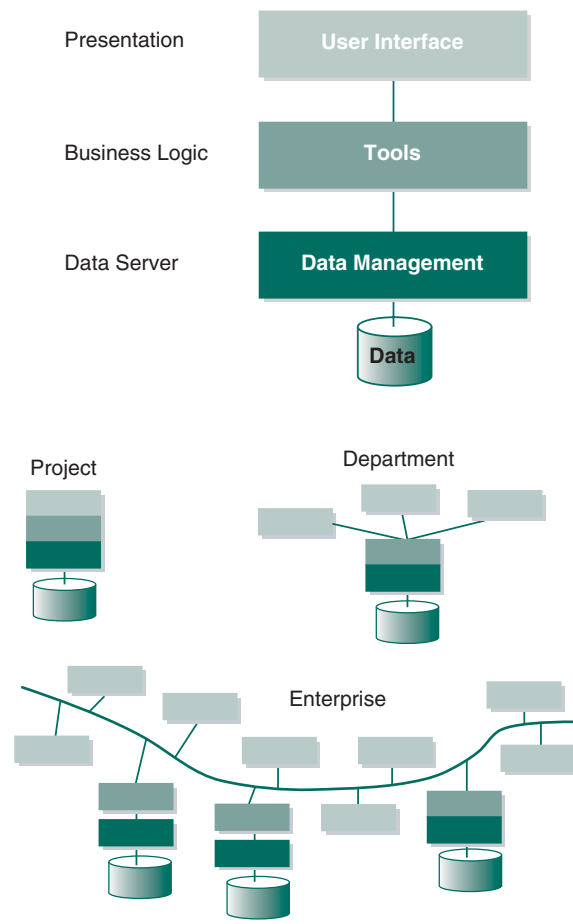


Figure 7.2 Classical three-tier architecture of a GIS software system

server. In the simplest desktop configuration, as used in single-user project GIS, the three software tiers are all installed on a single piece of hardware (most commonly a desktop PC) in the form of a desktop GIS software package and users are usually unaware of their existence (Figure 7.3A). In a variation on this theme, data files are held on a centralized file server, but the data server functionality is still part of the desktop GIS. This means that the entire contents of any accessed file must be pulled across the network even if only a small amount of it is required (Figure 7.3B)

In larger and more advanced multiuser workgroup or departmental GIS, the three tiers can be installed on multiple machines to improve flexibility and performance (Figure 7.4). In this type of configuration, the users in a single department (for example, the planning or public works department in a typical local government organization) still interact with a desktop GIS GUI (presentation layer) on their desktop computer which also contains all the business logic, but the database management software (data server layer) and data may be located on another machine connected over a network. This type of computing architecture is usually referred to as client-server, because clients request data or processing services from servers that perform work to satisfy client requests. The data server has local processing capabilities and is able to query and process data and thus return

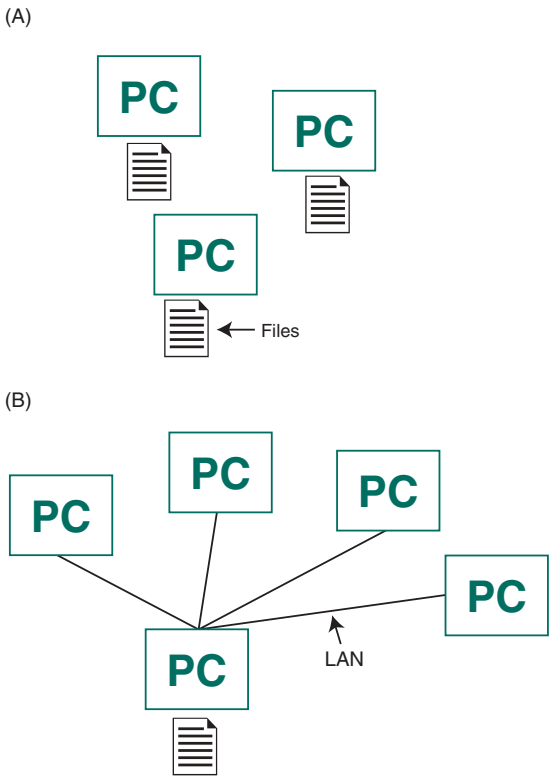


Figure 7.3 Desktop GIS software architecture used in project GIS: (A) stand alone desktop GIS on PCs each with own local files; (B) desktop GIS on PCs sharing files on a PC file server over a LAN (Local Area Network)

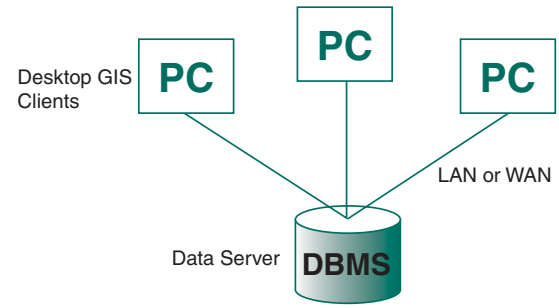


Figure 7.4 Client-server GIS: desktop GIS software and DBMS data server in a workgroup or departmental GIS configuration (LAN = Local Area Network, WAN = Wide Area Network)

part of the whole database to the client. Clients and servers can communicate over local area, wide area, or Internet networks, but given the large amount of data communication between the client and server, faster local area networks are most widely used.

In a client-server GIS, clients request data or processing services from servers that perform work to satisfy client requests.

Both the desktop and client-server architecture configurations discussed above have significant amounts of functionality on the desktop and are said to be thick clients (see also Section 7.3.4). In contrast, in the centralized desktop architecture configuration all the GUI and business logic is hosted on a centralized server, called an application (or middle tier) server (Figure 7.5). Typically this is in the form of a desktop GIS software package. An additional piece of software is also installed on the application server (Citrix or Windows Terminal Server) that allows users on remote machines full access to the software over a Local Area Network (LAN) or Wide Area Network (WAN) as though it were on the local desktop PC. Since the only application software that runs on the

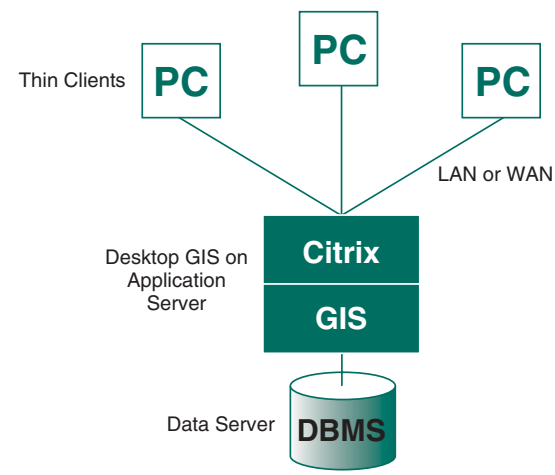


Figure 7.5 Centralized desktop GIS as used in advanced departmental and enterprise implementations

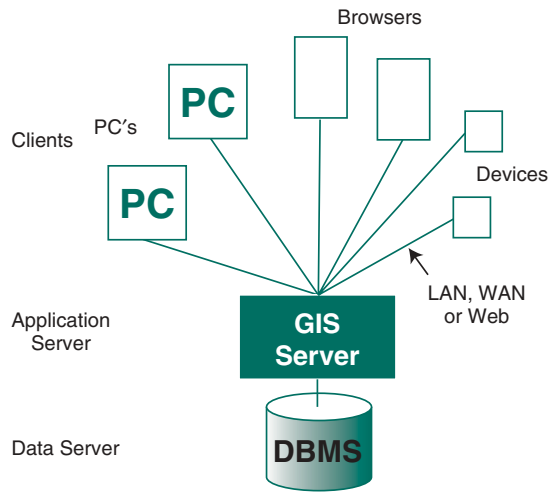


Figure 7.6 Centralized server GIS as used in advanced departmental and enterprise implementations

desktop PC is a small client library this is said to be a thin client. A data server (DBMS) is usually used for data management. This type of configuration is widely employed in large departmental and enterprise applications where high-end capabilities such as advanced editing, mapping and analysis are required.

In a more common variation of the centralized desktop implementation, the business logic is implemented as a true server system and runs on a middle tier server machine (Figure 7.6). In this configuration a range of thick and thin clients running on desktop PCs, Web browsers and specialist devices communicate with the middle tier server over a network connection. In the case of thin client access, the presentation tier (user interface) also runs on the server (although technically it is still presented on the desktop). The server machines may be connected over a local area network, but increasingly the Internet is used to connect widely distributed servers. This type of implementation is common in enterprise GIS. Large, enterprise GIS may involve more than ten servers and hundreds or even thousands of clients that are widely dispersed geographically.

Although organizations often standardize on either a project, departmental, or enterprise system, it is also common for large organizations to have project, department, and enterprise configurations all operating in parallel or as subparts of a full-scale system.

7.3.3 Software data models and customization

In addition to the three-tier model, two further topics are relevant to an understanding of software architecture: data models and customization.

GIS data models will be discussed in detail in Chapter 8, and so the discussion here will be brief. From a software perspective, a data model defines how the real world is represented in a GIS. It will also affect the type of software tools (functions or operators) that are available, how they are organized, and their mode of operation. A software data model defines how the different tools are grouped together, how they can be used, and how they interact with data. Although such software facets are largely transparent to end users whose interaction with a GIS is via a user interface, they become very important to software developers that are interested in customizing or extending software.

Customization is the process of modifying GIS software to, for example, add new functionality to applications, embed GIS functions in other applications, or create specific-purpose applications. It can be as simple as deleting unwanted controls (for example menu choices or buttons) from a GUI, or as sophisticated as adding a major new extension to a software package for such things as network analysis, high-quality cartographic production, or land parcel management.

To facilitate customization, GIS software products must provide access to the data model and expose capabilities to use, modify, and supplement existing functions. In the late 1980s when customization capabilities were first added to GIS software products, each vendor had to provide a proprietary customization capability simply because no standard customization systems existed. Nowadays, with the widespread adoption of the Microsoft .Net and Sun Java frameworks, as well as public domain languages, a number of industry standard programming languages (such as Visual Basic, Java, and Python) are available for customizing GIS software systems.

Modern programming languages are one component of larger developer-oriented software packages called integrated development environments (IDEs). The term IDE refers to the fact that the packages combine several software development tools including a visual programming language; an editor; a debugger, and a profiler. Many of the so-called visual programming languages, such as C, C#, Visual Basic and Java, support the development of Windows-based GUIs containing forms, dialogs, buttons, and other controls. Program code can be entered and attached to the GUI elements using the integrated code editor. An interactive debugger will help identify syntactic problems in the code, for example, misspelled commands and missing instructions. Finally, there are also tools to support profiling programs. These show where resources are being consumed and how programs can be speeded up or improved in other ways.

Contemporary GIS typically use an industry-standard programming language like Visual Basic or Java for customization.

A number of mainstream COTS GIS software vendors have licensed the right to include an IDE within their GIS software package. A particularly popular choice for desktop GIS is Microsoft's Visual Basic. Figure 7.7 shows a screenshot of the customization environment within ESRI's ArcGIS 9 Desktop GIS. For server-based products both C and Java are widely used, but because of the specialist nature of this market segment they must be obtained separately from the GIS software. To support customization using open, industry-standard IDEs, a GIS vendor must expose details of the software

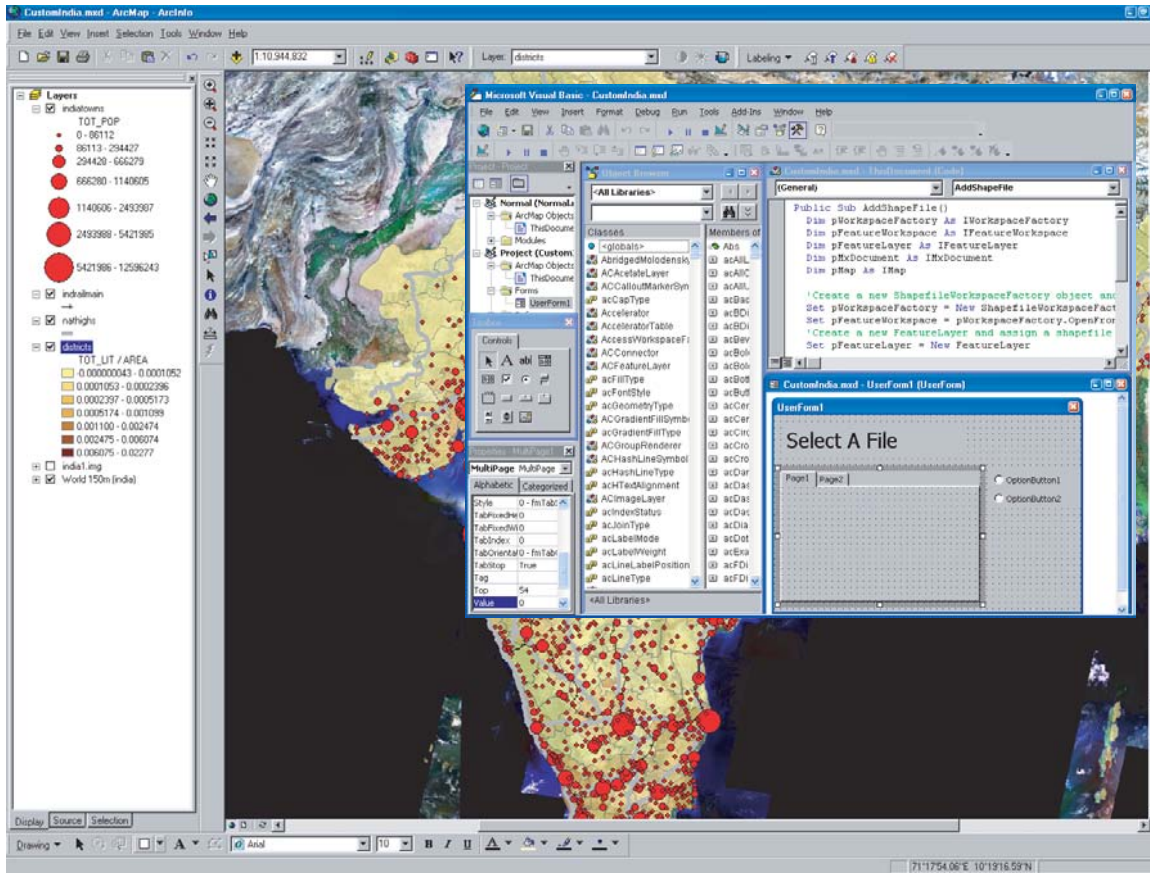


Figure 7.7 The customization capabilities of ESRI's ArcGIS 9. ESRI chose to embed Microsoft's Visual Basic for Applications as the scripting and GUI integrated development environment. The window at the front is the Visual Basic Integrated Development Environment (IDE). The window at the back is ArcMap, the main map-centric application of ArcInfo (see also Box 7.3)

package's functionality. This can be done by creating and documenting a set of application programming interfaces (APIs). These are interfaces that allow GIS functionality to be called by the programming tools in an IDE. In recent years, second-generation interfaces have been developed for accessing software functionality in the form of independent building blocks called software components.

In recent years, three technology standards have emerged for defining and reusing software functionality (components). For building interactive desktop applications, Microsoft's .Net framework is the de facto standard for high-performance, interactive applications that use fine-grained components (that is, a large number of small functionality blocks). For server-centric GIS both .Net and Sun Microsystems's Java framework are widely deployed in operational GIS applications. Although both .Net and Java work very well for building fine-grained client or server applications, they are less well suited for building applications that need to communicate over the Web. Because of the loosely-coupled, comparatively slow, heterogeneous nature of Web networks and applications, fine-grained programming models do not work well. As a consequence, coarse-grained messaging systems have been built on top of the fine-grained .Net and Java frameworks using the XML (extensible markup language) protocol that allow applications with Web services interfaces to interact over the Web.

Components are important to software developers because they are the mechanism by which reusable, self-contained, software building blocks are created. They allow many programmers to work together to develop a large software system incrementally. The standard, open (published) format of components means that they can easily be assembled into larger systems. In addition, because the functionality within components is exposed through interfaces, developers can reuse them in many different ways, even supplementing or replacing functions if they so wish. Users also benefit from this approach because GIS software can evolve incrementally and support multiple third party extensions. In the case of GIS products this includes, for example, tools for charting, reporting, and data table management.

7.3.4 GIS on the desktop and on the Web

Today mainstream, high-end GIS users work primarily with software that runs either on the desktop or over the

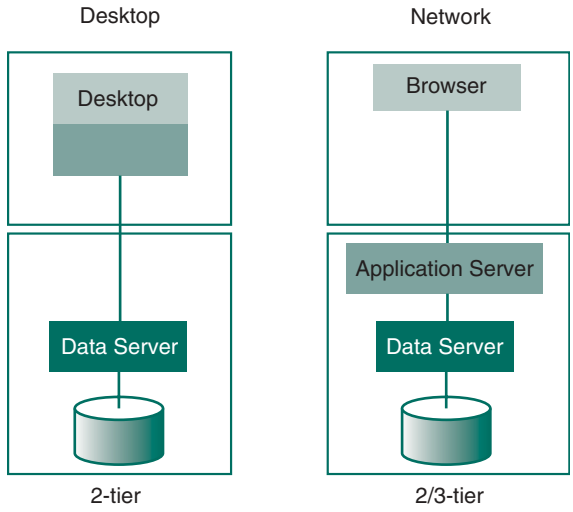


Figure 7.8 Desktop and network GIS paradigms. In the desktop top case the business logic is part of the client, but in the network case it runs on a server

Table 7.1 Comparison of desktop and network GIS

Feature	Desktop	Network
Client size	Thick	Thin
Client platform	Windows	Cross Platform Browser
Server size	Thin/thick	Thick
Server platform	Windows/Unix/ Linux	Windows/Unix/Linux
Component standard	.Net	.Net/Java
Network	LAN/WAN	LAN/WAN/Internet

Web. In the desktop case a PC (personal computer) is the main hardware platform and Microsoft Windows the operating system (Figure 7.8 and Table 7.1). In the desktop paradigm clients tend to be functionally rich and substantial in size, and are often referred to as thick or fat clients. Use of the Windows standard facilitates interoperability (interaction) with other desktop applications, such as word processors, spreadsheets, and databases. As noted earlier (Section 7.3.2), most sophisticated and mature GIS workgroups have adopted the client-server implementation approach by adding either a thin or thick server application running on the Windows, Linux or Unix operating system. The terms thin and thick are less widely used in the context of servers, but they mean essentially the same as when applied to clients. Thin servers perform relatively simple tasks, such as serving data from files or databases, whereas thick servers also offer more extensive analytical capabilities such as geocoding, routing, mapping, and spatial analysis. In desktop GIS implementations, LANs and WANs tend to be used for client-server communication. It is natural for developers to select Microsoft's .Net technology framework to build the underlying components making up these systems given the preponderance of the Windows operating system, although other component standards could also be used. The Windows-based client-server system architecture is a good platform for hosting interactive, high-performance GIS applications. Examples of applications well suited to this platform include those involving geographic data editing, map production, 2-D and 3-D visualization, spatial analysis, and modeling. It is currently the most practical platform for general-purpose systems because of its wide availability, good performance for a given price, and common usage in business, education, and government.

GIS users are standardizing their systems on the desktop and Internet implementation models.

In the last few years there has been increasing interest in harnessing the power of the Web for GIS. Although desktop GIS have been and continue to be very successful, users are constantly looking for lower costs of ownership and improved access to geographic information. Network-based (sometimes called distributed) GIS allow previously inaccessible information resources to be made more widely available. The network GIS model intrigues many organizations because it is based on centralized software and data management, which can dramatically reduce initial implementation and ongoing support and maintenance costs. It also provides the opportunity to link nodes of distributed user and processing resources using the medium of the Internet. The continued rise in network GIS will not signal the end of desktop GIS, indeed quite the reverse, since it is likely to stimulate the demand for content and professional GIS skills in geographic database automation and administration, and application development.

In contrast to desktop GIS, network GIS can use the cross-platform Web browser to host the viewer user-interface. Currently, clients are typically very thin, often with simple display and query capabilities, although there is an increasing trend for them to become more functionally rich. Server-side functionality may be encapsulated on a single server, although in medium and large systems it is more common to have two servers, one containing the business logic (a middleware application server), the other the data manager (data server). The server applications typically contain all the business logic and are comparatively thick. The server applications may run on a Windows, Unix or Linux platform.

Recently, there has been a move to combine the best elements of the desktop and network paradigms to create so-called rich clients. These are stored and managed on the server and dynamically downloaded to a client computer according to user demand. The business logic and presentation layers run on the server, with the client hardware simply used to render the data and manage use interaction. The new software capabilities in recent editions of the .Net and Java software development kits allow the development of applications with extensive user interaction that closely emulate the user experience of working with desktop software.

7.4 Building GIS software systems

Commercial GIS software systems are built and released as GIS software products by GIS-vendor software development and product teams. Such products are subject to carefully planned versioned release cycles that incrementally enhance and extend the pool of capabilities. The key parts of a GIS software architecture – user interface, business logic (tools), data manager, data model, and customization environment – were outlined in the previous section.

GIS software vendors start with a formal design for a software system and then build each part or component separately before assembling the whole system. Typically, development will be iterative with creation of an initial prototype framework containing a small number of partially functioning parts, followed by increasing refinement of the system. Core GIS software systems are usually written in a modern programming language like Visual C++, C# or Java, with Visual Basic or Java sometimes used for operations that do not involve significant amounts of computer processing like the GUI.

As standards for software development become more widely adopted, so the prospect of reusing software components becomes a reality. A key choice that then faces all software developers or customizers is whether to design a software system by buying in components, or to build it more or less from scratch. There are advantages to both options: building components gives greater control over system capabilities and enables specific-purpose optimization; and buying components can save time and money. Examples of components which have been purchased and licensed for use in GIS software systems include: Business Objects' Crystal Reports in MapInfo Professional, Microsoft Visual Basic for Applications in ESRI ArcGIS Desktop, and Safe Software's Feature Manipulation Engine for data conversion in Autodesk Map 3D.

A key GIS implementation issue is whether to buy a system or to build one.

A modern GIS software system comprises an integrated suite of software components of three basic types: a data management system for controlling access to data (Chapters 8, 9 and 10); a mapping system for display and interaction with maps and other geographic visualizations (Chapters 12 and 13); and a spatial analysis and modeling system for transforming geographic data using operators (Chapters 14, 15 and 16). The components for these parts may reside on the same computer or can be distributed widely (Chapter 11) over a network. The work of one of GIS's leading software developers is described in Box 7.1

7.5 GIS software vendors

Daratech Inc., an IT market research and technology assessment company, produces annual estimates about the size and characteristics of the GIS market. For 2003 (Figure 7.10) they list the main players in the worldwide GIS software market as ESRI, Intergraph, Autodesk, and GE Energy. Secondary players include Leica Geosystems, IBM, and MapInfo. In order to understand the current commercial GIS software market place and the direction in which it is likely to head, it is first necessary to examine the background, current product offerings and strategy of the main players.

7.5.1 ESRI Inc.

ESRI is a privately held company founded in 1969 by Jack and Laura Dangermond. Headquartered in Redlands, California, ESRI employs over 4000 people worldwide and has annual revenues of over US $500 m. Today it serves more than 130 000 organizations and more than 1 million users. ESRI focuses solely on the GIS market, primarily as a software product company, but also generates about a quarter of its revenue from project work such as advising clients on how to implement GIS. ESRI started building commercial software products in the late 1970s. Today, ESRI's product strategy is centered on an integrated family of products called ArcGIS. The ArcGIS family is aimed at both end users and technical developers and includes products that run on hand-held devices, desktop personal computers, and servers.

ESRI is the classic high-end GIS vendor. It has a wide range of mainstream products covering all the main technical and industry markets. ESRI is a technically-led geographic company focused squarely on the needs of hard-core GIS users. Box 7.3 describes ESRI's ArcGIS product.

7.5.2 Intergraph Inc.

Like ESRI, Intergraph was also founded in 1969 as a private company. The initial focus from their Huntsville, Alabama offices was the development of computer graphics systems. After going public in 1981, Intergraph grew rapidly and diversified into a range of graphics areas including CAD and mapping software, consulting services and hardware. After a series of reorganizations in the late 1990s and early 2000s, Intergraph is today structured into four main operating units: Process, Power and Marine; Public Safety; Solutions; and Mapping and Geospatial Solutions. The latter is the main GIS focus of the company. Mapping and Geospatial Solutions accounts for more than $200 m of the annual Intergraph total revenue which exceeds $500 m.

Intergraph has a large and diverse product line. From a GIS perspective the principal product family is GeoMedia which spans the desktop and network (Internet) server markets.

Historically Intergraph has been one of the top two global GIS companies. Today it is strongest in the military, infrastructure and utility market areas. Box 7.2 describes Intergraph GeoMedia.

Biographical Box **7.1**

Scott Morehouse, software developer and father of ArcGIS

Figure 7.9 Scott Morehouse, software architect

Scott Morehouse (Figure 7.9) learnt his programming and software development tradecraft at the Harvard Laboratory for Computer Graphics and Spatial Analysis in Massachusetts, in the 1970s. He was one of the lead developers of a system developed at the Lab called ODYSSEY which was the first general purpose vector-based GIS. It implemented many key GIS ideas such as digitizing, polygon overlay and choropleth mapping.

In 1981 he moved to Redlands, California to work at ESRI where he was the lead architect and developer for the initial ArcInfo release (see also Box 7.3). He was one of the programmers who implemented the first commercial vector polygon overlay algorithm which is still in use twenty-five years later. For the past two decades Scott has been the lead designer and manager for software development for all ESRI software products. He has overseen the development of ArcGIS, including ArcView, ArcGIS Engine, and ArcGIS Server.

Scott is very much a pragmatist who realized early on that successful software must not only be well designed and implemented with good algorithms and data structures, but also must be robust, well documented and widely applicable. He has always had a talent for synthesizing complex information and distilling out the essence so that non-experts can understand. He sees his role at ESRI as building software tools that apply geographic theory to help people solve real-world problems. Scott works alongside programmers and other product developers to design practical software architectures and robust solutions to user problems.

Although less well known in the GIS industry than ESRI's charismatic leader, Jack Dangermond, Scott is every bit as responsible for ESRI's success in the GIS software field. He has strong beliefs and a clear vision about what it takes to make things that really work. These are vital for maintaining the integrity of a large, complex system that is built by a diverse team of software engineers. He is also adept at managing the conflicting interests of supporting existing applications, while at the same time evolving GIS software products through innovation.

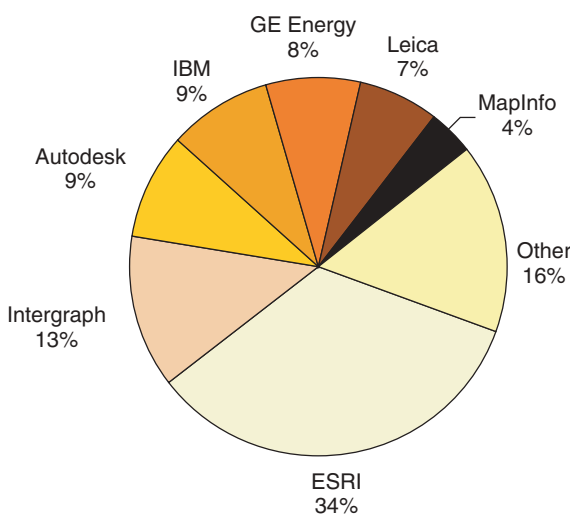


Figure 7.10 GIS 2003 software vendor market share (Courtesy Daratech)

7.5.3 Autodesk Inc.

Autodesk is a large and well known publicly traded company with headquarters in San Rafael, California. It is one of the world's leading digital design and content companies and serves customers in markets where design is critical to success: building, manufacturing, infrastructure, digital media, and location services. Autodesk is best known for its AutoCAD product family which is used worldwide by more than 4 million customers. The company was founded more than 20 years ago and has grown to become a publicly traded $1 bn entity employing over 3700 staff. The GIS division at Autodesk contributes over 10% of the company's revenue.

Autodesk's success in the GIS arena centers around three main product areas: desktop, where Autodesk Map 3D (based on AutoCad) is the flagship; an Internet server called MapGuide; and hand-held GIS, through the OnSite family of products.

Autodesk is classically thought of as a successful computer-aided design (CAD) company that has extended

itself into GIS. It has been especially successful in industries that have a strong engineering and design element. Autodesk MapGuide is described in Box 7.4.

7.5.4 GE Energy

GE Energy is very different to the other major GIS players on Daratech's GIS software vendor list. GE Energy's GIS software, which is referred to as a Geospatial Asset Management Solution, is based on the Smallworld GIS. The codebase was acquired in 2000 when Smallworld was purchased by GE Power Systems (now GE Energy). Smallworld was established in the late 1980s in Cambridge, UK by several entrepreneurs with a history in the CAD industry. From the outset the technology focused on complex utility network solutions, especially in the electric and gas industries. Smallworld grew rapidly to become one of the top three GIS utility software providers.

The Smallworld product suite offers an integrated workgroup and enterprise solution that spans the desktop and Internet, and is able to integrate with other IT business systems. Increasingly the product suite is focused on specific electric, gas and telco utility design and operational system solutions where it is used for network design and operation, and asset management.

7.6 Types of GIS software systems

Over 100 commercial software systems claim to have mapping and GIS capabilities. From the previous section on the product strategy of the major GIS software vendors it can be seen that four main types of generic GIS software dominate the market: desktop; server; developer; and hand-held. In this section, these four categories of mainstream GIS software will be discussed followed by a brief summary of other types of software. Reviews of currently popular GIS software packages can be found in the various GIS magazines (see Box 1.4).

7.6.1 Desktop GIS software

Since the mid-1990s, desktop GIS has been the mainstay of the majority of GIS implementations and the most widely used category of GIS software. Desktop GIS software owes its origins to the personal computer and the Microsoft Windows operating system. Initially, the major GIS vendors ported their workstation or minicomputer GIS software to the PC, but subsequently redeveloped their software specifically for the PC platform. Desktop GIS software provides personal productivity tools for a wide variety of users across a broad cross section of industries. PCs are commonly available, relatively inexpensive and offer a large collection of user-oriented tools including databases, word processors, and spreadsheets. The desktop GIS software category includes a range of options from simple viewers (such as ESRI ArcReader, Intergraph GeoMedia Viewer and MapInfo ProViewer) to desktop mapping and GIS software systems (such as Autodesk Map 3D, ESRI ArcView, GE Spatial Intelligence, Intergraph GeoMedia (see Box 7.2, Figure 7.11), and MapInfo Professional), and at the high-end, full-featured 'professional' editor/analysis systems (such as ESRI ArcInfo (see Box 7.3), Integraph GeoMedia Professional, and GE Energy Smallworld GIS).

Desktop GIS are the mainstream workhorses of GIS today.

In the late 1990s, a number of vendors released free GIS viewers that are able to display and query popular file formats. Today, the GIS viewer has developed into a significant product subcategory. The rationale behind the development of these products is that they help to establish market share, and can create de facto standards for specific vendor terminology and data formats. GIS users often work with viewers on a casual basis, often in conjunction with more sophisticated GIS software products. GIS viewers have limited functional capabilities restricted to display, query, and simple mapping. They do not support editing, sophisticated analysis, modeling, or customization.

With their focus on data use, rather than data creation, and their excellent tools for making maps, reports, and charts, desktop mapping and GIS software packages represent most people's experience of mainstream GIS today. The successful systems have all adopted the Microsoft standards for interoperability and user interface style. Users often see a desktop mapping and GIS package as simply a tool to enable them to do their full-time job faster, more easily, or more cheaply. Desktop mapping and GIS users work in planning, engineering, teaching, the army, marketing, and other similar professions; they are often not Spatially Aware Professionals (SAPs). Desktop GIS software prices typically range from $1000–2000 (these and other prices mentioned later typically have discounts for multiple purchases).

The term 'professional' relates to the full-featured nature of this subcategory of software. The distinctive features of professional GIS include data collection and editing, database administration, advanced geoprocessing and analysis, and other specialist tools. Professional GIS offer a superset of the capabilities of the systems in other classes. The people who use these systems are typically technically literate and think of themselves as SAPs – GIS professionals (career GIS staff) with degrees and, in many cases, advanced degrees in GIS or related disciplines. Prices for professional GIS are typically in the range $7000–20 000 per user.

Professional GIS are high-end, fully functional systems.

7.6.2 Server GIS

The last decade of GIS has been dominated by the desktop GIS architecture running on PCs. It is a fair bet that the

Applications Box **7.2**

Desktop GIS: Intergraph GeoMedia

GeoMedia is an archetypal example of a mainstream commercial desktop GIS software product (Figure 7.11). First released in the late 1990s from a new codebase, it was built from the ground up to run on the Windows desktop operating system. Like other products in the desktop GIS category it is primarily designed with the end user in mind. It has a Windows-based graphical user interface and many tools for editing, querying, mapping and spatial analysis. Data can be stored in proprietary GeoMedia files or in DBMS such as Oracle and Microsoft Access and SQL Server. GeoMedia enables data from multiple disparate databases to be brought into a single GIS environment for viewing, analysis, and presentation. The data are read and translated on the fly directly into memory. GeoMedia provides access to all major geospatial/CAD data file formats and to industry-standard relational databases.

GeoMedia is built as a collection of software object components. These underlying objects are exposed to developers who can customize the software using a programming language such as Visual Basic or C#.

GeoMedia offers a suite of analysis tools, including attribute and spatial query, buffer zones, spatial overlays, and thematic analysis. The product's layout composition tools provide the flexibility to design a range of types of maps that can be distributed on the Web, printed, or exported as files.

GeoMedia is the entry to a family of products. Several extensions (add-ons) offer additional functionality (e.g., image, grid and terrain analysis, and transaction management) and

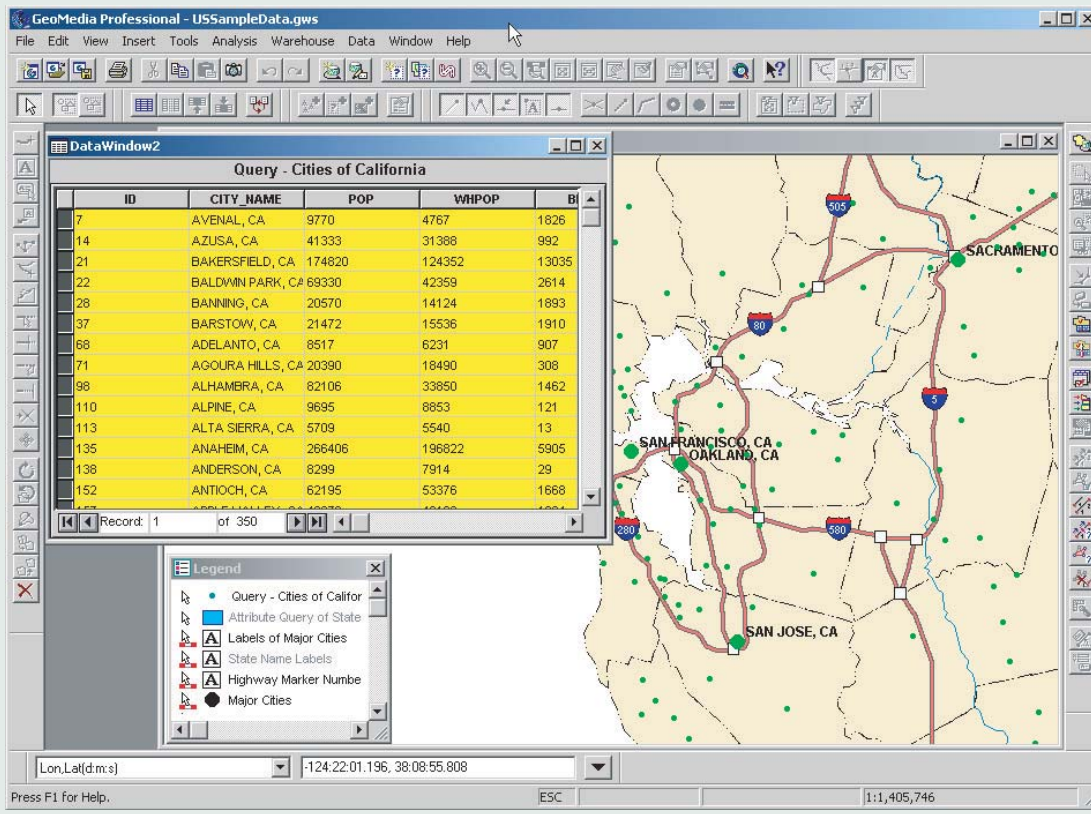


Figure 7.11 Screenshot of Intergraph GeoMedia – Desktop GIS

support for industry-specific workflows (e.g., transportation, parcel and public works). Other members of the product family include GeoMedia Viewer (free data viewer), GeoMedia Pro (high-end 'professional' functionality) and GeoMedia WebMap (Internet publishing).

All in all the GeoMedia product family offers a wide range of capabilities for core GIS activities in the mainstream markets of government, education, and private companies. It is a modern and integrated product line with strengths in the areas of data access and user productivity.

Applications Box **7.3**

Desktop GIS: ESRI ArcGIS ArcInfo

ArcInfo is ESRI's full-featured professional GIS software product (Figure 7.12). It supports the full range of GIS functions including: data collection and import; editing, restructuring, and transformation; display; query; and analysis. It is also the platform for a suite of analytic extensions for 3-D analysis, network routing, geostatistical and spatial (raster) analysis, among others. ArcInfo's strengths are the comprehensive portfolio of capabilities, the

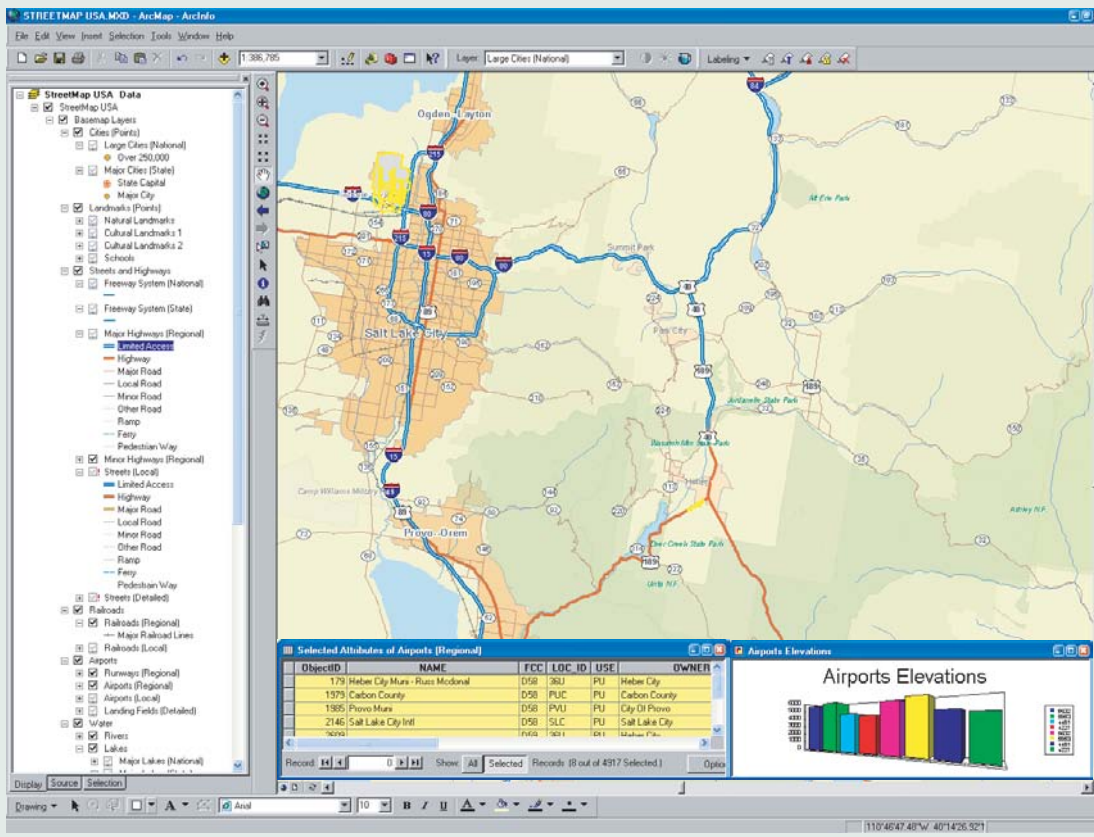


Figure 7.12 Screenshot of ESRI ArcGIS ArcInfo – Desktop GIS

sophisticated tools for data management and analysis, the customization options, and the vast array of third party tools and interfaces.

ArcInfo was originally released in 1981 on minicomputers. The early releases offered very limited functionality by today's standards and the software was basically a collection of subroutines that a programmer could use to build a working GIS software application. A major breakthrough came in 1987 when ArcInfo 4 was released with AML (Arc Macro Language), a scripting language that allowed ArcInfo to be easily customized. This release also saw the introduction of a port to Unix workstations (the software was adapted to function on this new platform) and the ability to work with data in external databases like Oracle, Informix, and Sybase. In 1991, with the release of ArcInfo 6, ESRI again re-engineered ArcInfo to take better advantage of Unix and the X-Windows windowing standard. The next major milestone was the development of a menu-driven user interface in 1993 called ArcTools. This made the software considerably easier to use and also defined a standard for how developers could write ArcInfo-based applications. ArcInfo was ported to Windows NT at the 7.1 release in 1996. About this time ESRI also took the decision to re-engineer ArcInfo from first principles. This vision was realized in the form of ArcInfo 8, released in 1999.

ArcInfo 8 was quite unlike earlier versions of the software because it was designed from the outset as a collection of reusable, self-contained software components, based on Microsoft's COM standard. ESRI used these components to create an integrated suite of menu-driven, end-user applications: ArcMap – a map-centric application supporting integrated editing and viewing; ArcCatalog – a data-centric application for browsing and managing geographic data in files and databases; and ArcToolbox – a tool-oriented application for performing geoprocessing tasks such as proximity analysis, map overlay, and data conversion. ArcInfo is customizable using either the in-built Microsoft Visual Basic for Applications (VBA) or any other COM-compliant programming language. The software is also notable because of the ability to store and manage all data (geographic and attribute) in standard commercial off-the-shelf DBMS (e.g., DB2, Informix, SQL Server, and Oracle). In 2004, ESRI released version 9 which builds on the foundations of version 8.

Another interesting aspect of ArcInfo is that for compatibility reasons since Release 8 ESRI has included a fully working version of the original ArcInfo workstation technology and applications. This has allowed ESRI users to migrate their existing databases and applications to the new version in their own time.

next decade will in turn come to be dominated by server GIS products. In simple terms, a server GIS is a GIS that runs on a computer server that can handle concurrent processing requests from a range of networked clients. Server GIS products have the potential for the largest user base and lowest cost per user. Stimulated by advances in server hardware and networks, the widespread availability of the Internet and market demand for greater access to geographic information, GIS software vendors have been quick to release server-based products. Examples of server GIS include Autodesk MapGuide, ESRI ArcGIS Server, GE Spatial Application Server, Intergraph GeoMedia Webmap, and MapInfo MapXtreme. The cost of server GIS products varies from around $5000–25 000, for small to medium-sized systems, to well beyond for large multifunction, multiuser systems.

Internet GIS have the highest number of users, although typically Internet users focus on simple display and query tasks.

Initially, such products were nothing more than ports of desktop GIS products, but second generation systems were subsequently built using a multiuser services-based architecture that allows them to run unattended and to handle many concurrent requests from remote networked users. These software systems initially focused on display and query applications – making simple things simple and cost-effective – with more advanced applications becoming available as user awareness and technology expanded. Today, it is routinely possible to perform standard operations like making maps (Microsoft Expedia has online interactive maps; **www.expediamaps.com**), routing (MapQuest offers pathfinding with directions (**www.mapquest.com**: see Section 1.4.3), publishing census data (US Census Bureau, AmericanFactFinder has online census data and maps for the whole US at **www.census.gov**), and suitability analysis (the US National Association of Realtors has a site for locating homes for purchase based on user-supplied criteria at **www.realtor.com**). A recent trend has been the development of Internet-based online data networks such as the Geography Network (**www.GeographyNetwork.com**) and ESRI ArcWeb Services and Microsoft MapPoint.Net.

Applications Box **7.4**

Server GIS: Autodesk MapGuide

In the late 1990s, a time when desktop GIS had become dominant, a small Canadian company called Argus released an Internet GIS product called MapGuide (Figure 7.13). Subsequently purchased by Autodesk, MapGuide marked the start of another chapter in the history of GIS software. The emphasis in MapGuide and other Internet (now called server) GIS products is very much on map display and use. Indeed, these systems have few if any data editing capabilities.

MapGuide is an important innovation for the many users who have spent considerable amounts of time and money creating valuable databases, and who want to make them available to other users inside or outside their organization. Autodesk MapGuide allows users to leverage their existing GIS investment by publishing dynamic, intelligent maps at the point at which they are most valuable – in the field, at the job site, or on the desks of colleagues, clients, and the public.

There are three key components of MapGuide: the viewer – a relatively easy-to-use Web application with a browser style interface; the author – a menu-driven authoring environment used to create and publish a site for client access; and the server – the administrative software that monitors site usage, and manages requests from multiple clients, and to external databases. MapGuide works directly with Internet browsers and servers, and uses the HTTP protocol for communication. It makes good use

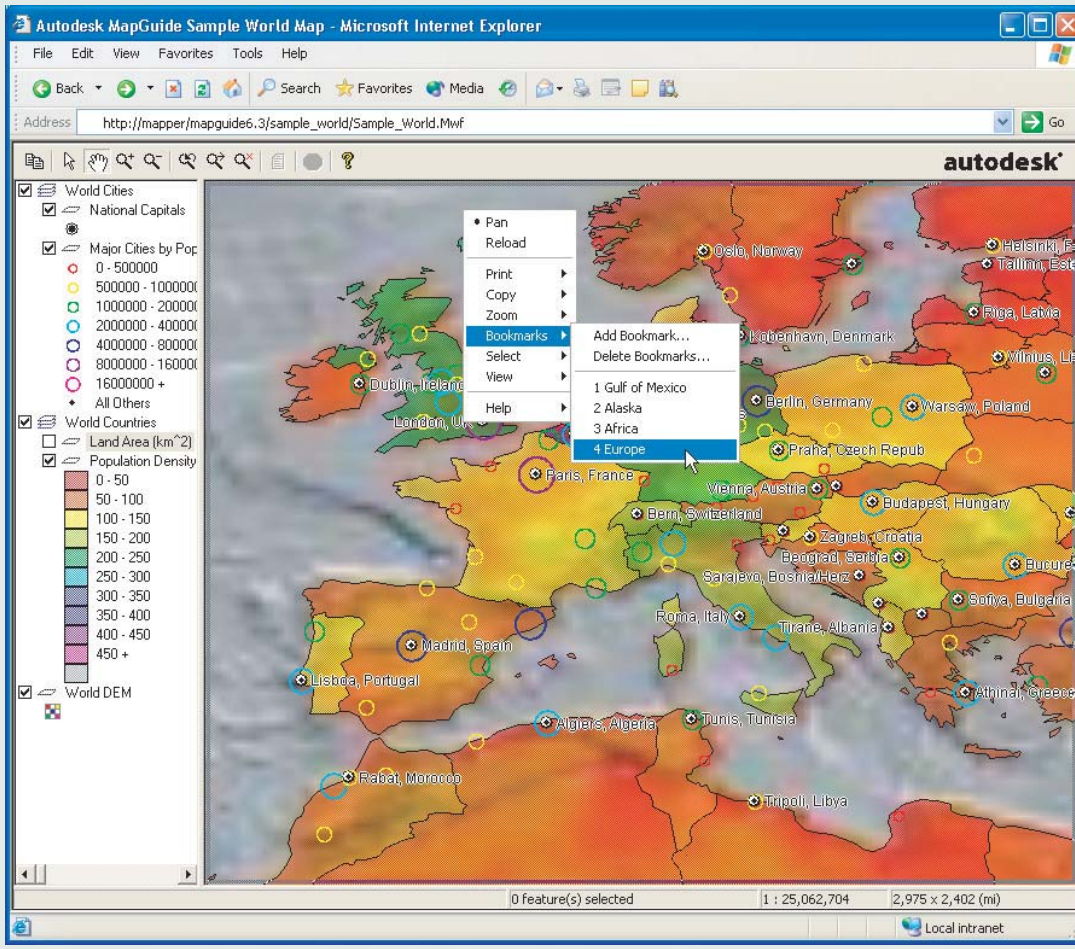


Figure 7.13 Screenshot of Autodesk MapGuide (Reproduced by permission of Autodesk MapGuide ®. Autodes, the Autodesk logo, and Autodesk MapGuide are registered trademarks of Autodesk, Inc. in the USA and/or other countries.)

of standard Internet tools like HTML (hypertext markup language) and JScript (Java Script) for building client applications and ColdFusion (an Internet site generation and management tool from Allaire Corp.) and ASP (Active Server Pages) / JSP (Java Server Pages) for managing data and queries on the server.

Typical features of MapGuide sites include the display of raster and vector maps, map navigation (pan and zoom), geographic and attribute queries, simple buffering, report generation, and printing. Like other advanced Internet server GIS, MapGuide has tools for redlining (drawing on maps) and editing geographic objects, although in MapGuide's case the editing tools are somewhat limited.

To date MapGuide has been used most widely in existing mature GIS sites that want to publish their data internally or externally, and in new sites that want a way to publish dynamic maps quickly to a widely dispersed collection of users (for example, maps showing election results, or transportation network status).

In conclusion, MapGuide and the other server GIS products are growing in importance. Their cost-effective nature, ability to be centrally managed, and focus on ease of use will help to disseminate geographic information even more widely and will introduce many new users to the field of GIS.

The second generation server GIS products were strong on architecture and exploited the unique characteristics of the Web by developing GIS technology that integrates with Web browsers and servers. Unfortunately, these gains were at the expense of reduced functionality. In the past few years this problem has been rectified and there is a new breed of true GIS server that offers 'complete' GIS functionality in a multiuser server environment. These server GIS products have functions for editing, mapping, data management, and spatial analysis, and support state of the art customization.

7.6.3 Developer GIS

With the advent of component-based software development (see Section 7.4), a number of GIS vendors have released collections of GIS software components oriented toward the needs of developers. These are really tool kits of GIS functions (components) that a reasonably knowledgeable programmer can use to build a specific-purpose GIS application. They are of interest to developers because such components can be used to create highly customized and optimized applications that can either stand alone or can be embedded within other software systems. Typically, component GIS packages offer strong display and query capabilities, but only limited editing and analysis tools.

Developer GIS products are collections of components used by developers to create focused applications.

Examples of component GIS products include Blue Marble Geographics GeoObjects, ESRI ArcGIS Engine, and MapInfo MapX. Most of the developer GIS products from mainstream vendors are built on top of Microsoft's .Net technology standards, but there are several cross platform choices (e.g., ESRI ArcGIS Engine) and several Java-based toolkits (e.g., ObjectFX SpatialFX and Engenuity JLOOX). The typical cost for a developer GIS product is $1000–5000 for the developer kit and $100–500 per deployed application. The people who use deployed applications may not even realize that they are using a GIS, because often the run-time deployment is embedded in other applications (e.g., customer care systems, routing systems, or interactive atlases).

7.6.4 Hand-held GIS

As hardware design and miniaturization have improved dramatically over the past few years, so it has become possible to develop GIS software for mobile and personal use on hand-held systems. The development of low cost, lightweight location positioning technologies (primarily based on the Global Positioning System, GPS, see Section 5.8) and wireless networking has further stimulated this market. With capabilities similar to the desktop systems of just a few years ago, these palm and pocket devices can support many display, query, and simple analytical applications, even on displays of 320 by 240 pixels (a quarter of the VGA (640 by 480) pixel screen resolution standard). An interesting characteristic of these systems at the present time is that all programs and data are held in memory because of the lack of a hard disk. This provides fast access, but because of the cost of memory compared to disk systems, designers have had to develop compact data storage structures.

Hand-held GIS are lightweight systems designed for mobile and field use.

A very recent development is the availability of hand-held software on high-end so-called 'smartphones'.

Figure 7.14 A hand-held GIS for a smartphone

In spite of their compact size they are able to deal with comparatively large amounts of data (up to 1 GB) and surprisingly sophisticated software applications (Figure 7.14). The systems usually operate in a mixed connected/disconnected environment and so can make active use of data and software applications held on the server and accessed over a wireless telephone network.

Hand-held GIS are now available from many vendors, and include Autodesk OnSite, ESRI ArcPad, and Intergraph Intelliwhere. Many of these systems are designed to work with server GIS products (see above). Costs are typically around $400–600.

7.6.5 Other types of GIS software

The previous section has focused on mainstream GIS software from the major commercial vendors. There are many other types of commercial and non-commercial software that provide valuable GIS capabilities. This section will briefly review some of the main types of other software.

Raster-based GIS, as the name suggests, focus primarily on raster (image) data and raster analysis. Chapters 3 and 9 provide a discussion of the principles and techniques associated with raster and other data models, while Chapters 13 and 14 review their specific capabilities. Just as many vector-based systems have raster analysis extensions (for example, ESRI ArcGIS has Spatial Analyst, and GeoMedia has Image and Grid), in recent years raster systems have added vector capabilities (for example, Leica Geosystems EROAS IMAGINE and Clark Labs' Idrisi now have vector capabilities built in). The distinction between raster-based and other software system categories is becoming increasingly blurred as a consequence. The users of raster-based GIS are primarily interested in working with imagery and undertaking spatial analysis and modeling activities. The prices for raster-based GIS range from $500–10 000.

Computer-Aided Design (CAD)-based GIS are systems that started life as CAD packages and then had GIS capabilities added. Typically, a CAD system is supplemented with database, spatial analysis, and cartography capabilities. Not surprisingly, these systems appeal mainly to users whose primary focus is in typical CAD application areas such as architecture, engineering, and construction, but who also want to use geographic information and geographic analysis in their projects. The best-known examples of CAD-based GIS are Autodesk Map 3D and Bentley GeoGraphics. CAD-based GIS typically cost $3000–5000.

Many enterprise-wide GIS incorporate middleware (middle tier) GIS data and application servers. Their purpose is to manage multiple users accessing continuous geographic databases, which are stored and managed in commercial-off-the-shelf (COTS) database management systems (DBMS). GIS middleware products offer centralized management of data, the ability to process data on a server (which delivers good performance for certain types of applications), and control over database editing and update (see Chapter 11 for further details). A number of GIS vendors have developed technology that fulfills this function. Examples of GIS application servers include: Autodesk GIS Design Server, ESRI ArcSDE, and MapInfo SpatialWare. These systems typically cost $10 000–25 000 or more depending on the number of users.

To assist in managing data in standard DBMS, some vendors – notably IBM and Oracle – have developed technology to extend their DBMS servers so that they are able to store and process geographic information efficiently. Although not strictly GIS in their own right (due to the absence of editing, mapping and analysis tools) they are included here for completeness. Box 10.1. provides an overview of Oracle's Spatial DBMS extension.

In addition to the commercial-off-the-shelf GIS software products that have been the mainstay of this chapter, it is also important to acknowledge that there is a growing movement that is creating public-domain, open source and free software. In the early days such software products provided only rather simple, poorly engineered tools with no user support. Today there are several high-quality, feature rich software products. Some noteworthy examples include: GeoDa for spatial analysis and visualization (**www.csiss.org/clearinghouse/GeoDa/**), Minnesota Map Server for serving maps over the Web (**http://mapserver.gis.umn.edu/**), PostGIS for storing

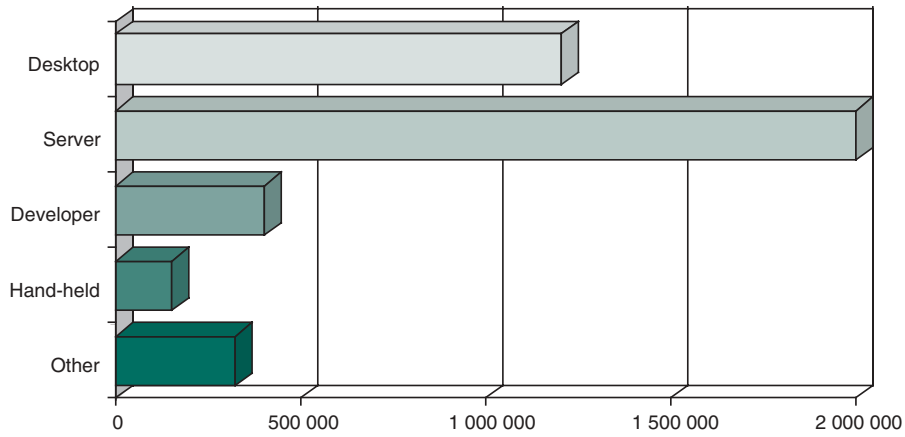


Figure 7.15 Estimated size (number of users) of the different GIS software sectors

data in a DBMS (PostgresSQL) (**http://postgis.refractions.net/**) and GRASS (**http://grass.itc.it/**) for desktop GIS tasks.

Looking forward, it is interesting to consider a new trend in GIS software. GIS functionality is now being delivered packaged along with data as seamless GIServices (Section 1.5.3). To date the most developed services have centered on simple location-based services, street mapping and routing, such as ESRI ArcWeb Services and Microsoft MapPoint, but more advanced services are now becoming available.

7.7 GIS software usage

Estimates of the size of the GIS market by type of system, based on the authors' knowledge, is shown in Figure 7.15. In 2005, the total size of the GIS market, measured in terms of the numbers of regular GIS software users (using the high-end definition adopted in this chapter), is about 4 million spread over 2 million sites. If the number of Internet GIS users is also taken into consideration then the GIS user population rises to about 10 million in total. This excludes users of GIS products such as hard copy maps, charts, and reports.

7.8 Conclusion

GIS software is a fundamental and critical part of any operational GIS. The software employed in a GIS project has a controlling impact on the type of studies that can be undertaken and the results that can be obtained. There are also far reaching implications for user productivity and project costs. Today, there are many types of GIS software product to choose from and a number of ways to configure implementations. One of the exciting and at times unnerving characteristics of GIS software is its very rapid rate of development. This is a trend that seems set to continue as the software industry pushes ahead with significant research and development efforts. The following chapters will explore in more detail the functionality of GIS software and how it can be applied in real-world contexts.

Questions for further study

1. Design a GIS architecture that 25 users in 3 cities could use to create an inventory of recreation facilities.
2. Discuss the role of each of the three tiers of software architecture in an enterprise GIS implementation.
3. With reference to a large organization that is familiar to you, describe the ways in which its staff might use GIS, and evaluate the different types of GIS software systems that might be implemented to fulfill these needs.
4. Go to the websites of the main GIS software vendors and compare their product strategies with open source GIS products. In what ways are they different?
   - Autodesk: **www.autodesk.com**
   - ESRI: **www.esri.com**
   - Intergraph: **www.intergraph.com**
   - GE Energy **www.gepower.com**

Further reading

Bishr Y. and Radwan M. 2000 'GDI Architectures'. In Groot R. and McLaughlin J. (eds) *Geospatial Data Infrastructure: Concepts, Cases, and Good Practice*. New York: Oxford University Press, pp. 135–50.

Coleman D.J. 2005 'GIS in networked environments'. In Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley, pp. 317–29.

Elshaw Thrall S. and Thrall G.I. 2005 'Desktop GIS software'. In Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley, pp. 331–45

Maguire D.J. 2005 'GIS customization'. In Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley, pp. 359–69

Peng Z-H. and Tsou M-H. 2003 *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ: Wiley.

8 *Geographic data modeling*

This chapter discusses the technical issues involved in modeling the real world in a GIS. It describes the process of data modeling and the various data models that have been used in GIS. A data model is a set of constructs for describing and representing parts of the real world in a digital computer system. Data models are vitally important to GIS because they control the way that data are stored and have a major impact on the type of analytical operations that can be performed. Early GIS were based on extended CAD, simple graphical, and image data models. In the 1980s and 1990s, the hybrid georelational model came to dominate GIS. In the last few years major software systems have been developed on more advanced and standards-based geographic object models that include elements of all earlier models.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will be able to:

- Define what geographic data models are and discuss their importance in GIS;
- Understand how to undertake GIS data modeling;
- Outline the main geographic models used in GIS today and their strengths and weaknesses;
- Understand key topology concepts and why topology is useful for data validation, analysis, and editing;
- Read data model notation;
- Describe how to model the world and create a useful geographic database.

8.1 Introduction

This chapter builds on the material on geographic representation presented in Chapter 3. By way of introduction it should be noted that the terms representation and model as used here overlap considerably (we will return to a more detailed discussion of models in Chapter 16). For present purposes, representation can be considered to denote the conceptual and scientific issues, whereas model is used in practical and database circles. In this chapter, given the technical and practical approach, *data model* will be used to distinguish it from process models as discussed in Chapter 16. This chapter focuses on how geographic reality is modeled (abstracted or simplified) in a GIS, with particular emphasis on choosing one particular style of data model over another. A data model is an essential ingredient of any operational GIS and, as the discussion will show, has important implications for the types of operations that can be performed and the results that can be obtained.

8.1.1 Data model overview

The heart of any GIS is the data model, which is a set of constructs for representing objects and processes in the digital environment of the computer (Figure 8.1). People (GIS users) interact with operational GIS in order to undertake tasks like making maps, querying databases, and performing site suitability analyses. Because the types of analyses that can be undertaken are strongly influenced by the way the real-world is modeled, decisions about the type of data model to be adopted are vital to the success of a GIS project.

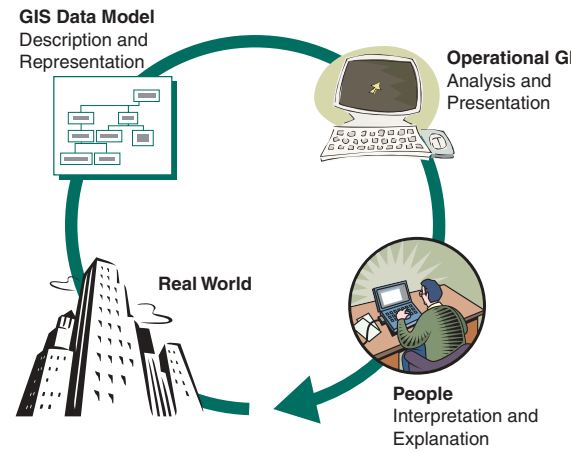


Figure 8.1 The role of a data model in GIS

A data model is a set of constructs for describing and representing selected aspects of the real-world in a computer.

As described in Chapter 3, geographic reality is continuous and infinitely complex, but computers are finite, comparatively simple, and can only deal with digital data. Therefore, difficult choices have to be made about what things are modeled in a GIS and how they are represented. Because different types of people use GIS for different purposes, and the phenomena these people study have different characteristics, there is no single type of all-encompassing GIS data model that is best for all circumstances.

8.1.2 Levels of data model abstraction

When representing the real-world in a computer, it is helpful to think in terms of four different levels of abstraction (levels of generalization or simplification) and these are shown in Figure 8.2. First, *reality* is made up of real-world phenomena (buildings, streets, wells, lakes, people, etc.), and includes all aspects that may or may not be perceived by individuals, or deemed relevant to a particular application. Second, the *conceptual model* is a human-oriented, often partially structured, model of selected objects and processes that are thought relevant to a particular problem domain. Third, the *logical model* is an implementation-oriented representation of reality that is often expressed in the form of diagrams and lists. Lastly, the *physical model* portrays the actual implementation in a GIS, and often comprises tables stored as files or databases (see Chapter 10). In terms of the discussion of uncertainty in Chapter 6 (Figure 6.1), the conceptual and logical models are found beyond the U1 filter, and the physical model is that upon which

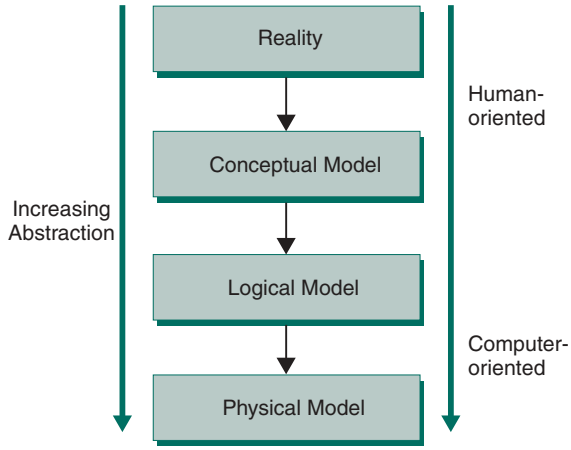


Figure 8.2 Levels of abstraction relevant to GIS data models

analysis may be performed (beyond the U2 filter). Use of the term 'physical' here is actually misleading because the models are not physical, they only exist digitally in computers, but this is the generally accepted use of the term.

In data modeling, users and system developers participate in a process that successively engages with each of these levels. The first phase of modeling begins with definition of the main types of objects to be represented in the GIS and concludes with a conceptual description of the main types of objects and relationships between them. Once this phase is complete, further work will lead to the creation of diagrams and lists describing the names of objects, their behavior, and the type of interaction between objects. This type of logical data model is very valuable for defining what a GIS will be able to do and the type of domain over which it will extend. Logical models are implementation independent, and can be created in any GIS with appropriate capabilities. The final data modeling phase involves creating a model showing how the objects under study can be digitally implemented in a GIS. Physical models describe the exact files or database tables used to store data, the relationships between object types, and the precise operations that can be performed. For more details about the practical steps involved in data modeling see Sections 8.3 and 8.4.

A data model provides system developers and users with a common understanding and reference point. For developers, a data model is the means to represent an application domain in terms that may be translated into a design and then implemented in a system. For users, it provides a description of the structure of the system, independent of specific items of data or details of the particular application. A data model controls the types of things that can be handled by a GIS and the range of operations that can be performed on them.

The discussion of geographic representation in Chapter 3 introduced discrete objects and fields, the two fundamental conceptual models for representing things geographically in the real world. In the same chapter the raster and vector logical models were also introduced. Figure 8.3 shows two representations of the same area in a GIS, one raster and the other vector. Notice the difference in the objects represented. Major roads, in green, and areas cleared of vegetation in red, are more clearly visible in the vector representation, whereas smaller roads and built-up areas in lighter shades of gray can best be seen on the scanned raster aerial photograph. The next sections in this chapter focus on the logical and physical representation of raster, vector, and related models in GIS software systems.

8.2 GIS data models

In the past half-century, many GIS data models have been developed and deployed in GIS software systems.

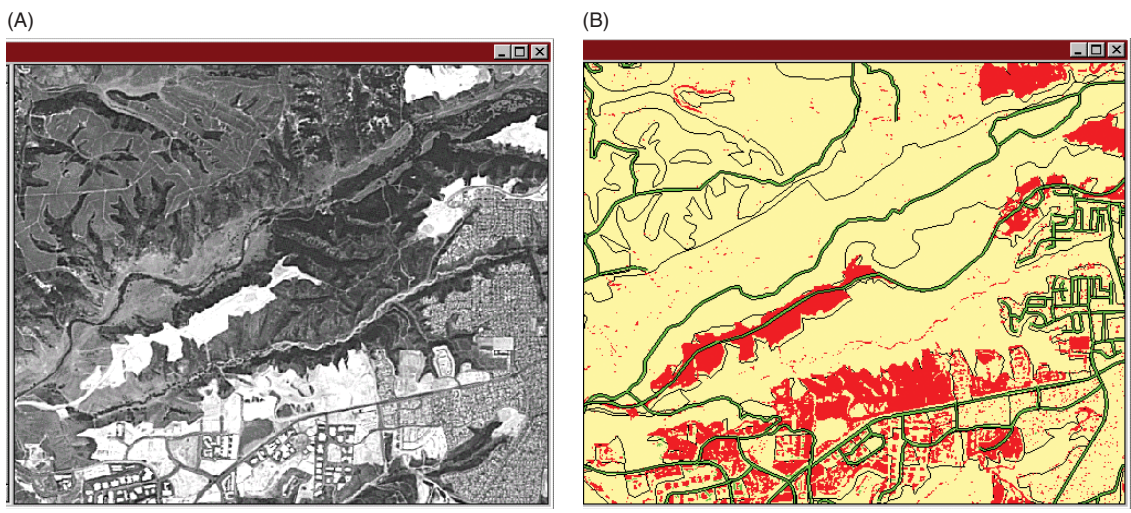


Figure 8.3 Two representations of San Diego, California (Courtesy Leica Geosystems): (A) panchromatic SPOT raster satellite image collected in 1990 at 10 m resolution; (B) vector objects digitized from the image

Table 8.1 Geographic data models used in GIS

Data model	Example applications
Computer-Aided Design (CAD)	Automating engineering design and drafting.
Graphical (non-topological)	Simple mapping.
Image	Image processing and simple grid analysis.
Raster/Grid	Spatial analysis and modeling especially in environmental and natural resources applications.
Vector/Georelational topological	Many operations on vector geometric features in cartography, socio-economic and resource analysis, and modeling.
Network	Network analysis in transportation, hydrology, and utilities.
Triangulated Irregular Network (TIN)	Surface/terrain visualization, analysis, and modeling.
Object	Many operations on all types of entities (raster/vector/TIN, etc.) in all types of applications.

The key types of geographic data models and their main areas of application are listed in Table 8.1. All are based in some way on the conceptual discrete object/field and logical vector/raster geographic data models (see Chapter 3 for more details). All GIS software systems include a core data model that is built on one or more of these GIS data models. In practice, any modern comprehensive GIS supports at least some elements of all these models. As discussed earlier, the GIS software core system data model is the means to represent geographic aspects of the real world and defines the type of geographic operations that can be performed. It is the responsibility of the GIS implementation team to populate this generic model with information about a particular problem (e.g., utility outage management, military mapping, or natural resource planning). Some GIS software packages come with a fixed data model, while others have models that can be easily extended. Those that can easily be extended are better able to model the richness of geographic domains, and in general are the easiest to use and the most productive systems.

When modeling the real world for representation inside a GIS, it is convenient to group entities of the same geometric type together (for example, all point entities such as lights, garbage cans, dumpsters, etc., might be stored together). A collection of entities of the same geometric type (dimensionality) is referred to as a class or layer. It should also be noted that the term layer is quite widely used in GIS as a general term for a specific dataset. It derived from the process of entering different types of data into a GIS from paper maps, which was undertaken one plate at a time (all entities of the same type were represented in the same color and, using printing technology, were reproduced together on film or printing plates). Grouping entities of the same geographic type together makes the storage of geographic databases more efficient (for further discussion see Section 10.3). It also makes it much easier to implement rules for validating edit operations (for example, the addition of a new building or census administrative area) and for building relationships between entities. All of the data models discussed below use layers in some way to handle geographic entities.

A layer is a collection of geographic entities of the same geometric type (e.g., points, lines, or polygons). Grouped layers may combine layers of different geometric types.

8.2.1 CAD, graphical, and image GIS data models

The earliest GIS were based on very simple models derived from work in the fields of CAD (computer-aided design and drafting), computer cartography, and image analysis. In a CAD system, real-world entities are represented symbolically as simple point, line, and polygon vectors. This basic CAD data model never became widely popular in GIS because of three severe problems for most applications at geographic scales. First, because CAD models typically use local drawing coordinates instead of real world coordinates for representing objects, they are of little use for map-centric applications. Second, because individual objects do not have unique identifiers it is difficult to tag them with attributes. As the following discussion shows this is a key requirement for GIS applications. Third, because CAD data models are focused on graphical representation of objects they cannot store details of any relationships between objects (e.g., topology or networks), the type of information essential in many spatial analytical operations.

A second type of simple GIS geometry model was derived from work in the field of computer cartography. The main requirement for this field in the 1960s was the automated reproduction of paper topographic maps and the creation of simple thematic maps. Techniques were developed to digitize maps and store them in a computer for subsequent plotting and printing. All paper map entities were stored as points, lines, and polygons, with annotation used for placenames. Like CAD systems, there was no requirement to tag objects with attributes or to work with object relationships.

At about the same time that CAD and computer cartography systems were being developed, a third type of data model emerged in the field of image processing. Because the main data source for geographic image processing is scanned aerial photographs and digital satellite images it was natural that these systems would use rasters or grids to represent the patterning of real-world objects on the Earth's surface. The image data model is also well suited to working with pictures of real-world objects, such as photographs of water valves and scanned building floor plans that are held as attributes of geographically referenced entities in a database (Figure 8.4).

Figure 8.4 An image used as a hydrant object attribute in a water-facility system

In spite of their many limitations, GIS still exist based on these simple data models. This is partly for historical reasons – the GIS may have been built before newer, more advanced models became available – but also because of lack of knowledge about the newer approaches described below.

8.2.2 Raster data model

The raster data model uses an array of cells, or pixels, to represent real-world objects (Figure 3.7). The cells can hold any attribute values based on one of several encoding schemes including categories, and integer and floating-point numbers (see Box 3.3 for details). In the simplest case a binary representation is used (for example, presence or absence of vegetation), but in more advanced cases floating-point values are preferred (for example, height of terrain above sea level in meters). In some systems, multiple attributes can be stored for each cell in a type of value attribute table where each column is an attribute and each row either a pixel, or a pixel class (Figure 8.5).

Usually, raster data are stored as an array of grid values, with metadata (data about data: see Section 11.2.1) about the array held in a file header. Typical metadata include the geographic coordinate of the upper-left corner of the grid, the cell size, the number of row and column elements, and the projection. The data array itself is usually stored as a compressed file or record in a database management system (see Section 10.3). Techniques for compressing rasters are described in Box 8.1 (see also Section 3.6.1 for the general principles involved).

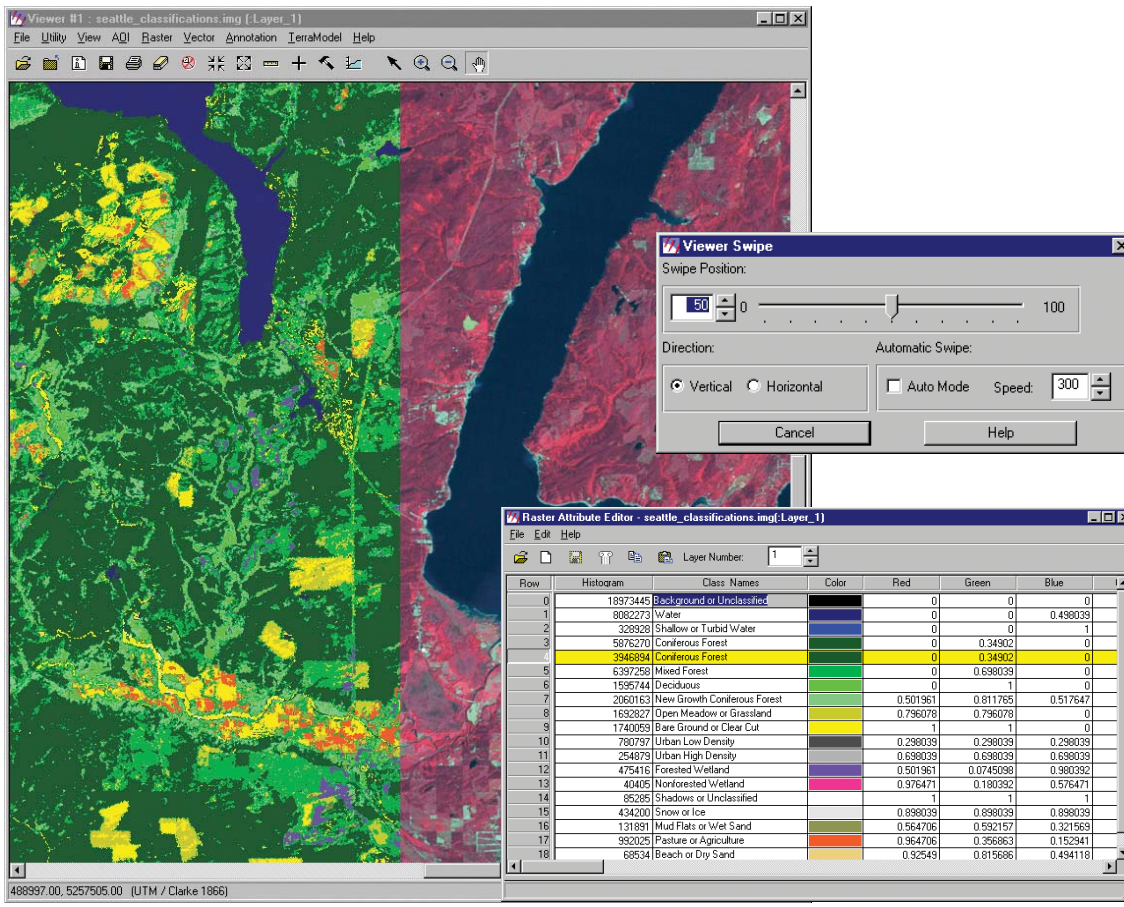


Figure 8.5 Raster data of the Olympic Peninsula, Washington State, USA, with associated value attribute table. Bands 4, 3, 2 from Landsat 5 satellite with land cover classification overlaid (Screenshot courtesy Leica Geosystems; data courtesy of US Geological Survey. Data available from US Geological Survey, EROS Data Center, Sioux Falls, SD)

Technical Box **8.1**

Raster compression techniques

Although the raster data model has many uses in GIS, one of the main operational problems associated with it is the sheer amount of raw data that must be stored. To improve storage efficiency many types of raster compression technique have been developed such as run-length encoding, block encoding, wavelet compression, and quadtrees (see Section 10.7.2.2 for another use of quadtrees as a means to index geographic data). Table 8.2 presents a comparison of file sizes and compression rates for three compression techniques based on the image in Figure 8.6. It can be seen that even the comparatively simple run-length encoding technique compresses the file size by a factor of 5. The more sophisticated wavelet compression technique results in a compression rate of almost 40, reducing the file from 80.5 to 2.3 MB.

Table 8.2 Comparison of file sizes and compression rates for raster compression techniques (using image shown in Figure 8.6)

Compression technique	File size (MB)	Compression rate
Uncompressed original	80.5	
Run-length	17.7	5.1
Wavelet	2.3	38.3

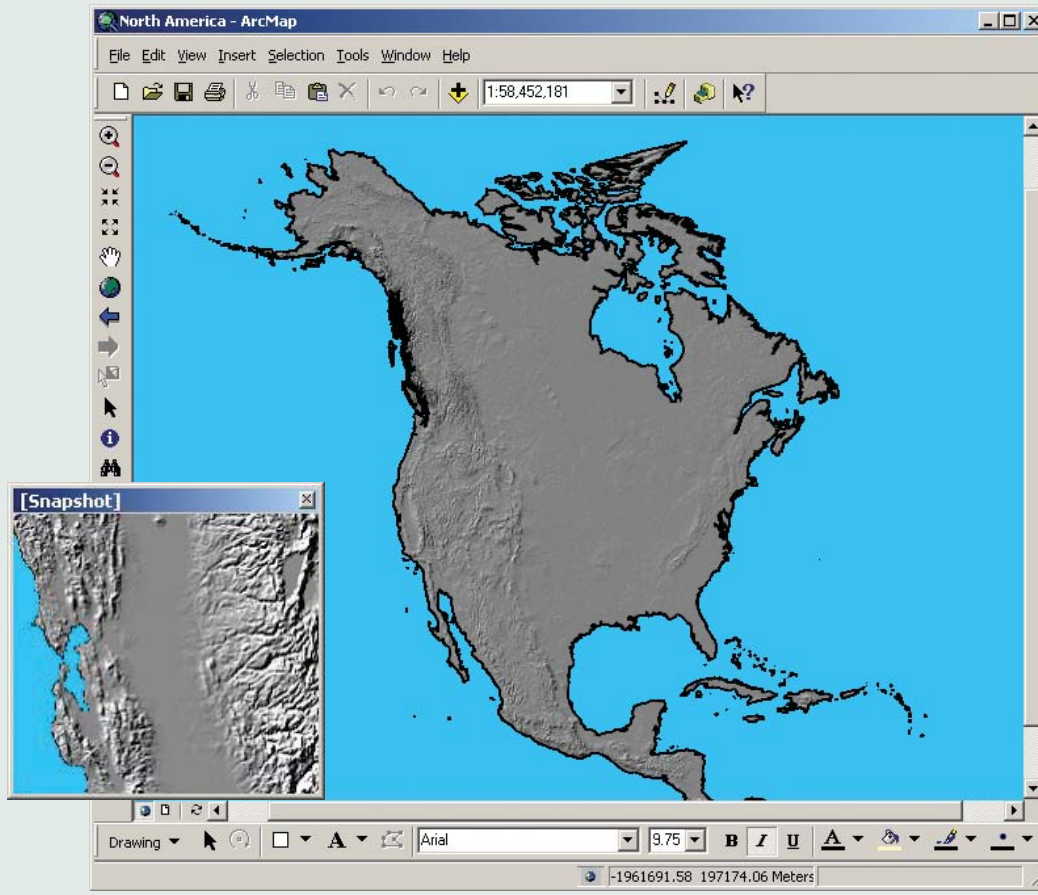


Figure 8.6 Shaded digital elevation model of North America used for comparison of image compression techniques in Table 8.2. Original image is 8726 by 10 618 pixels, 8 bits per pixel. The inset shows part of the image at a zoom factor of 1000 for the San Francisco Bay area

Run-length encoding

Run-length encoding is perhaps the simplest compression method and is very widely used. It involves encoding adjacent row cells that have the same value, with a pair of values indicating the number of cells with the same value, and the actual value.

Block encoding

Block encoding is a two-dimensional version of run-length encoding in which areas of common cell values are represented with a single value. An array is defined as a series of square blocks of the largest size possible. Recursively, the array is divided using blocks of smaller and smaller size. It is sometimes described as a quadtree data structure (see also Section 10.7.2.2).

Wavelet

Wavelet compression techniques invoke principles similar to those discussed in the treatment of fractals (Section 4.8). They remove information by recursively examining patterns in datasets at different scales, always trying to reproduce a faithful representation of the original. A useful byproduct of this for geographic applications is that wavelet-compressed raster layers can be quickly viewed at different scales with appropriate amounts of detail. MrSID (Multiresolution Seamless Image Database) from LizardTech is an example of a wavelet compression technique that is widely used in geographic applications, especially for compressing aerial photographs. Similar wavelet compression algorithms are available from other public and private sources and have been incorporated into the JPEG 2000 standard which is increasingly being used for image compression.

Run-length and block encoding both result in lossless compression of raster layers, that is, a layer can be compressed and decompressed without degradation of information. In contrast, the MrSID wavelet compression technique is lossy since information is irrevocably discarded during compression. Although MrSID compression results in very high compression ratios, because information is lost its use is limited to applications that do not need to use the raw digital numbers for processing or analysis. It is not appropriate for compressing DEMs for example, but many organizations use it to compress scanned maps and aerial photographs when access to the original data is not necessary.

Datasets encoded using the raster data model are particularly useful as a backdrop map display because they look like conventional maps and can communicate a lot of information quickly. They are also widely used for analytical applications such as disease dispersion modeling, surface water flow analysis, and store location modeling.

8.2.3 Vector data model

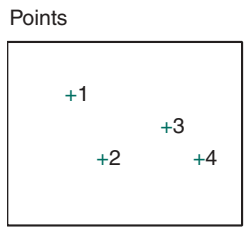
The raster data model discussed above is most commonly associated with the field conceptual data model. The vector data model on the other hand is closely linked with the discrete object view. Both of these conceptual perspectives were introduced in Section 3.5. The vector data model is used in GIS because of the precise nature of its representation method, its storage efficiency, the quality of its cartographic output, and the availability of functional tools for operations like map projection, overlay, and analysis.

In the vector data model each object in the real world is first classified into a geometric type: in the 2-D case point, line, or polygon (Figure 8.7). Points (e.g., wells, soil pits, and retail stores) are recoded as single coordinate pairs, lines (e.g., roads, streams, and geologic faults) as a series of ordered coordinate pairs (also called polylines – Section 3.6.2), and polygons (e.g., census tracts, soil areas, and oil license zones) as one or more line segments that close to form a polygon area. The coordinates that define the geometry of each object may have 2, 3, or 4 dimensions: 2 (x, y: row and column, or latitude and longitude), 3 (x, y, z: the addition of a height value), or 4 (x, y, z, m: the addition of another value to represent time or some other property – perhaps the offset of road signs from a road centerline, or an attribute).

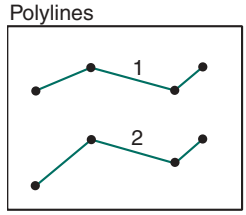
For completeness, it should also be said that in some data models linear features can be represented not only as a series of ordered coordinates, but also as curves defined by a mathematical function (e.g., a spline or Bézier curve). These are particularly useful for representing built environment entities like road curbs and some buildings.

8.2.3.1 Simple features

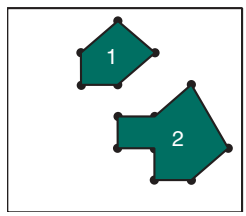
Geographic entities encoded using the vector data model are usually called features and this will be the convention adopted here. Features of the same geometric type are stored in a geographic database as a feature class, or when speaking about the physical (database) representation the



Point number	(x,y) coordinates
1	(2,4)
2	(3,2)
3	(5,3)
4	(6,2)



Polyline number	(x,y) coordinates
1	(1,5) (3,6) (6,5) (7,6)
2	(1,1) (3,3) (6,2) (7,3)



Polygon number	(x,y) coordinates
1	(2,4) (2,5) (3,6) (4,5) (3,4) (2,4)
2	(3,2) (3,3) (4,3) (5,4) (6,2) (5,1) (4,1) (4,2) (3,2)

Figure 8.7 Representation of point, line, and polygon objects using the vector data model

term *feature table* is preferred. Here each feature occupies a row and each property of the feature occupies a column. GIS commonly deal with two types of feature: simple and topological. The structure of simple feature polyline and polygon datasets is sometimes called *spaghetti* because, like a plate of cooked spaghetti, lines (strands of spaghetti) and polygons (spaghetti hoops) can overlap and there are no relationships between any of the objects.

Features are vector objects of type point, polyline, or polygon.

Simple feature datasets are useful in GIS applications because they are easy to create and store, and because they can be retrieved and rendered on screen very quickly. On the other hand because simple features lack more advanced data structure characteristics, such as topology (see next section), operations like shortest-path network analysis and polygon adjacency cannot be performed without additional calculations.

8.2.3.2 Topological features

Topological features are essentially simple features structured using topological rules. Topology is the mathematics and science of geometrical relationships. Topological relationships are non-metric (qualitative) properties of geographic objects that remain constant when the geographic space of objects is distorted. For example, when a map is stretched properties such as distance and angle change, whereas topological properties such as adjacency and containment do not. Topological structuring of vector layers introduces some interesting and very useful properties, especially for polyline (also called 1-cell, arc, edge, line, and link) and polygon (also called 2-cell, area, and face) data. Topological structuring of line, layers forces all line ends that are within a user-defined distance to be snapped together so that they are given exactly the same coordinate value. A node is placed wherever the ends of lines meet or cross. Following on from the earlier analogy this type of data model is sometimes referred to as spaghetti with meatballs (the nodes being the meatballs on the spaghetti lines). Topology is important in GIS because of its role in data validation, modeling integrated feature behavior, editing, and query optimization.

Topology is the science and mathematics of relationships used to validate the geometry of vector entities, and for operations such as network tracing and tests of polygon adjacency.

Data validation

Many of the geographic data collected from basic digitizing, field data collection devices, photogrammetry, and CAD systems comprise simple features of type point, polyline, or polygon, with limited structural intelligence (for example, no topology). Testing the topological integrity of a dataset is a useful way to validate the geometric quality of the data and to assess their suitability for geographic analysis. Some useful data validation topology tests include:

- Network connectivity – do all network elements connect to form a graph (i.e., are all the pipes in a water network joined to form a wastewater system)? Network elements that connect must be 'snapped' together (that is, given the same coordinate value) at junctions (intersections).
- Line intersection – are there junctions at intersecting polylines, but not at crossing polylines? It is, for example, perfectly possible for roads to cross in planimetric 2-D view, but not intersect in 3-D (for example, at a bridge or underpass; Figure 3.4).
- Overlap – do adjacent polygons overlap? In many applications (e.g., land ownership) it is important to build databases free from overlaps and gaps so that ownership is unambiguous.
- Duplicate lines – are there multiple copies of network elements or polygons? Duplicate polylines often occur during data capture. During the topological creation process it is necessary to detect and remove duplicate polylines to ensure that topology can be built for a dataset.

Modeling the integrated behavior of different feature types

In the real world many objects share common locations and partial identities. For example, water distribution areas often coincide with water catchments, electric distribution areas often share common boundaries with building sub-divisions, and multiple telecommunications fibers are often run down the same conduit. These situations can be modeled in a GIS database as either single objects with multiple geometry representations, or multiple objects with separate geometry integrated for editing, analysis, and representation. There are advantages and disadvantages to both approaches. Multiple objects with separate geometries are certainly easier to implement in commercially available databases and information systems. If one feature is moved during editing then logically both features should move. This is achieved by storing both objects separately in the database, each with their own geometry, but integrating them inside the GIS editor application so that they are treated as single features. When the geometry of one is moved the other automatically moves with it. There is further discussion of shared editing in the next section.

Editing productivity

Topology improves editor productivity by simplifying the editing process and providing additional capabilities to manipulate feature geometries. Editing requires both topological data structuring and a set of topologically aware tools. The productivity of editors can be improved in several ways:

- Topology provides editors with the ability to manipulate common, shared polylines and nodes as single geometric objects to ensure that no differences are introduced into the common geometries.
- Rubberbanding is the process of moving a node, polyline, or polygon boundary and receiving interactive feedback on screen about the location of all topologically connected geometry.
- Snapping is a useful technique to both speed up editing and maintain a high standard of data quality.
- Auto-closure is the process of completing a polygon by snapping the last point to the first digitized point.
- Tracing is a type of network analysis technique that is used, especially in utility applications, to test the connectivity of linear features (is the newly designed service going to receive power?).

Optimized queries

There are many GIS queries that can be optimized by pre-computing and storing information about topological relationships. Some common examples include:

- Network tracing (e.g., find all connected water pipes and fittings).
- Polygon adjacency (e.g., who owns the parcels adjoining those owned by a specific owner?).
- Containment (e.g., which manholes lie within the pavement area of a given street?).
- Intersection (e.g., which census tracts intersect with a set of health areas?).

In the remainder of this section the discussion concentrates on a conceptual understanding of GIS topology focusing on the more complex polygon case. The network case is considered in the next section. The relative merits and implementations of two approaches to GIS topology are discussed later in Section 10.7.1. These two implementation approaches differ because in one case relationships are batch built and stored along with feature geometry, and in the other relationships are calculated interactively when they are needed.

Conceptually speaking, in a topologically structured polygon data layer each polygon is defined as a collection of polylines that in turn are made up of an ordered list of coordinates (vertices). Figure 8.8 shows an example of a polygon dataset comprising six polygons (including the 'outside world': polygon 1). A number in a circle identifies a polygon. The lines that make up the polygons are shown in the polygon-polyline list. For example, polygon 2 can be assembled from lines 4, 6, 7, 10, and 8. In this particular implementation example the 0 before the 8 is used to indicate that line 8 actually defines an 'island' inside polygon 2. The list of coordinates for each line is also shown in Figure 8.8. For example, line 5 begins with coordinates 7,4 and 6,3 – other coordinates have been omitted for brevity. A line may appear in the polygon-polyline list more than once (for example, line 6 is used in the definition of both polygons 2 and 5), but the actual coordinates for each polyline are only stored once in the polyline-coordinate list. Storing common boundaries between adjacent polygons avoids the potential problems of gaps (slivers) or overlaps between adjacent polygons. It has the added bonus that there are fewer coordinates in a topologically structured polygon feature layer compared with a simple feature layer representation of the same entities. The downside, however, is that drawing a polygon requires that multiple polylines must be retrieved

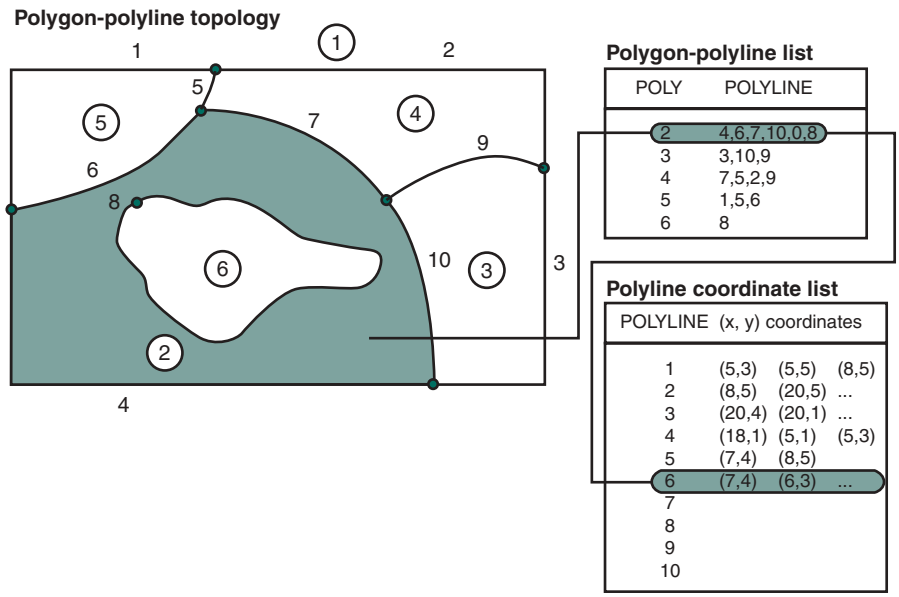


Figure 8.8 A topologically structured polygon data layer. The polygons are made up of the polylines shown in the polygon-polyline list. The lines are made up of the coordinates shown in the line coordinate list (*Source*: after ESRI 1997)

from the database and then assembled into a boundary. This process can be time consuming when repeated for each polygon in a large dataset.

Planar enforcement is a very important property of topologically structured polygons. In simple terms, planar enforcement means that all the space on a map must be filled and that any point must fall in one polygon alone, that is, polygons must not overlap. Planar enforcement implies that the phenomenon being represented is conceptualized as a field.

The contiguity (adjacency) relationship between polygons is also defined during the process of topological structuring. This information is used to define the polygons on the left- and right-hand side of each polyline, in the direction defined by the list of coordinates (Figure 8.9). In Figure 8.9, Polygon 2 is on the left of Polyline 6 and Polygon 5 is on the right. Thus we can deduce from a simple look-up operation that Polygons 2 and 5 are adjacent.

Software systems based on the vector topological data model have become popular over the years. A special case of the vector topological model is the *georelational* model. In this model derivative, the feature geometries and associated topological information are stored in regular computer files, whereas the associated attribute information is held in relational database management system (RDBMS) tables. The GIS software maintains the intimate linkage between the geometry, topology, and attribute information. This hybrid data management solution was developed to take advantage of RDBMS to store and manipulate attribute information. Geometry and topology were not placed in RDBMS because, until relatively recently, RDBMS were unable to store and retrieve geographic data efficiently. Figure 8.10 is an example of a georelational model as implemented in ESRI's ArcInfo coverage polygon dataset. It shows file-based geometry and topology information linked to attributes in an RDBMS table. The ID (identifier) of the polygon, the label point, is linked (related or joined) to the ID column in the attribute table (see also Chapter 10). Thus, in this soils dataset polygon 3 is soil B7, of class 212, and its suitability is moderate.

The topological feature geographic data model has been extensively used in GIS applications over the last 20 years, especially in government and natural resources applications based on polygon representations (see Sections 2.3.2 and 2.3.5). Typical government applications include: cadastral management, tax assessment, parcel management, zoning, planning, and building control. In the areas of natural resources and environment, key applications include site suitability analysis, integrated land use modeling, license mapping, natural resource management, and conservation. The tax appraisal case study discussed in Section 2.3.2.2 is an example of a GIS based on the topological feature data model. The developers of this system chose this model because they wanted to avoid overlaps and gaps in tax parcels (polygons), to ensure that all parcel boundaries closed (were validated), and to store data in an efficient way. This is in spite of the fact that there is an overhead in creating and maintaining parcel topology, as well as degradation in draw and query performance for large databases.

8.2.3.3 Network data model

The network data model is really a special type of topological feature model. It is discussed here separately because it raises several new issues and has been widely applied in GIS studies.

Networks can be used to model the flow of goods and services. There are two primary types of networks: *radial* and *looped*. In radial or tree networks flow

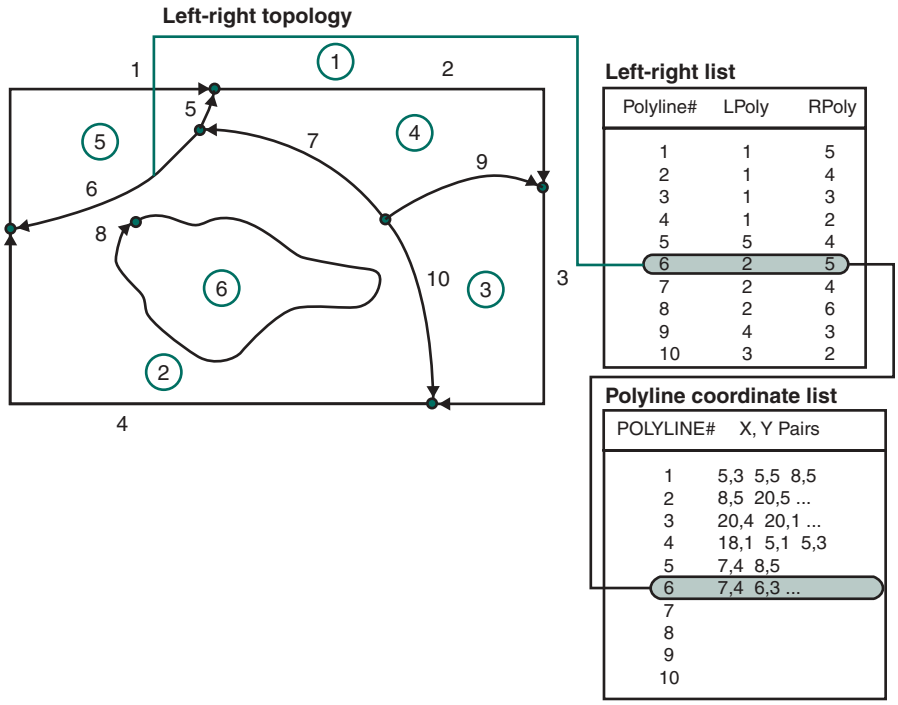


Figure 8.9 The contiguity of a topologically structured polygon data layer. For each polyline the left and right polygon is stored with the geometry data (*Source*: after ESRI 1997)

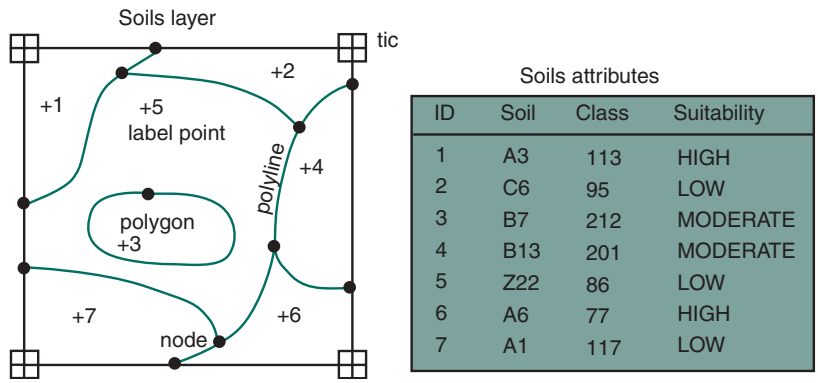


Figure 8.10 An example of a georelational polygon dataset. Each of the polygons is linked to a row in an RDBMS table. The table has multiple attributes, one in each column (*Source*: after ESRI 1997)

always has an upstream and downstream direction. Stream and storm drainage systems are examples of radial networks. In looped networks, self-intersections are common occurrences. Water distribution networks are looped by design to ensure that service interruptions affect the fewest customers.

In GIS software systems, networks are modeled as points (for example, street intersections, fuses, switches, water valves, and the confluence of stream reaches: usually referred to as nodes in topological models), and lines (for example, streets, transmission lines, pipes, and stream reaches). Network topological relationships define how lines connect with each other at nodes. For the purpose of network analysis it is also useful to define rules about how flows can move through a network. For example, in a sewer network, flow is directional from a customer (source) to a treatment plant (sink), but in a pressurized gas network flow can be in any direction. The rate of flow is modeled as impedances (weights) on the nodes and lines. Figure 8.11 shows an example of a street network. The network comprises a collection of nodes (types of street intersection) and lines (types of street), as well as the topological relationships between them. The topological information makes it possible, for example, to trace the flow of traffic through the network and to examine the impact of street closures. The impedance on the intersections and streets determines the speed at which traffic flows. Typically, the rate of

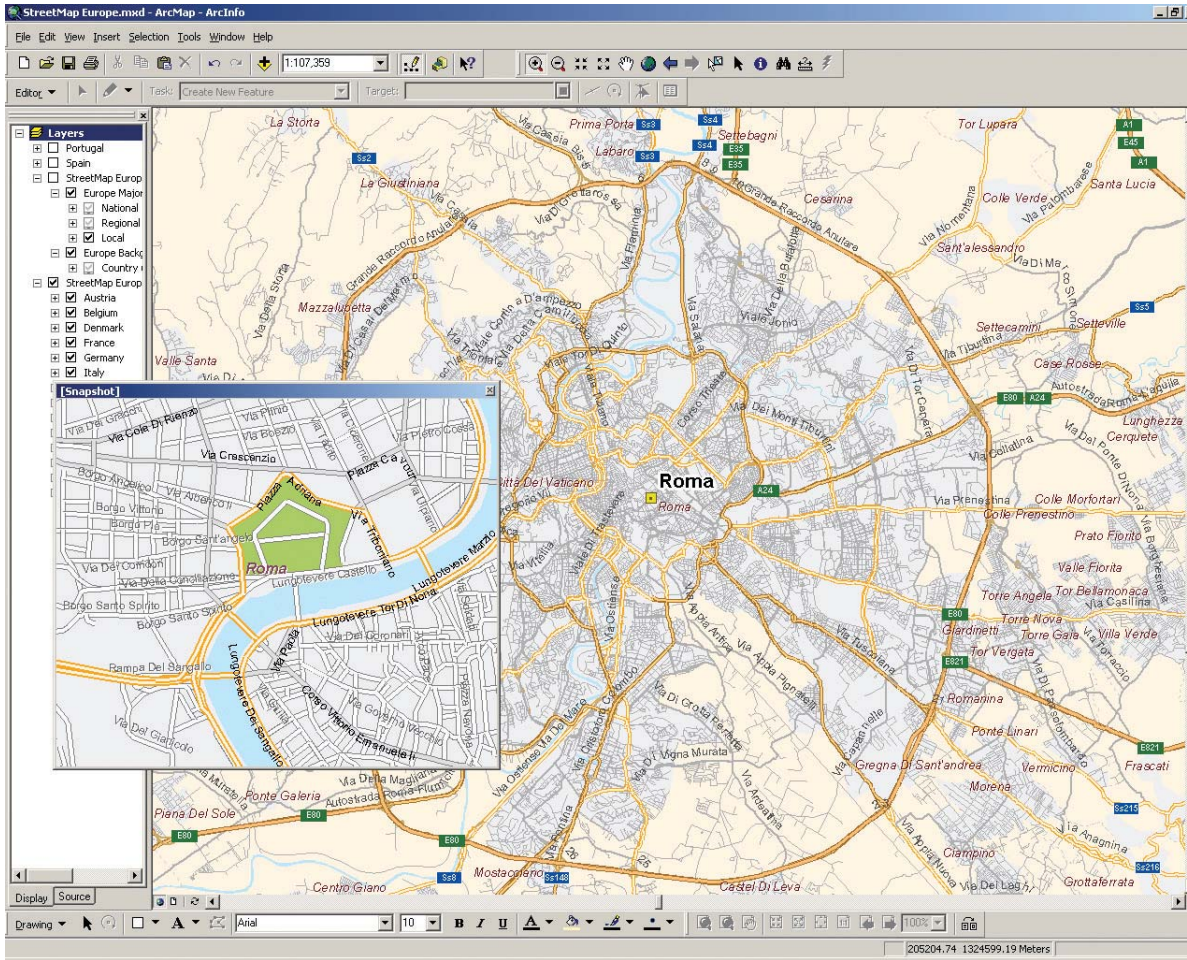


Figure 8.11 An example of a street network

flow is proportional to the street speed limit and number of lanes, and the timing of stoplights at intersections. Although this example relates to streets, the same basic principles also apply to, for example, electric, water, and railroad networks.

In georelational implementations of the topological network feature model, the geometry and topology information is typically held in ordinary computer files and the attributes in a linked database. The GIS software tools are responsible for creating and maintaining the topological information each time there is a change in the feature geometry. In more modern object models the geometry, attributes, and topology may be stored together in a DBMS, or topology may be computed on the fly.

There are many applications that utilize networks. Prominent examples include: calculating power load drops over an electricity network; routing emergency response vehicles over a street network; optimizing the route of mail deliveries over a street network; and tracing pollution upstream to a source over a stream network.

Network data models are also used to support another data model variant called *linear referencing* (Section 5.4). The basic principle of linear referencing is quite simple. Instead of recording the location of geographic entities as explicit x, y, z coordinates, they are stored as distances along a network (called a route system) from a point of origin. This is a very efficient way of storing information such as road pavement (surface) wear characteristics (e.g., the location of pot holes and degraded asphalt), geological seismic data (e.g., shockwave measurements at sensors along seismic lines), and pipeline corrosion data. An interesting aspect of this is that a two-dimensional network is reduced to a one-dimensional linear route list. The location of each entity (often called an event) is simply a distance along the route from the origin. Offsets are also often stored to indicate the distance from a network centerline. For example, when recording the surface characteristics of a multi-carriageway road several readings may be taken for each carriageway at the same linear distance along the route. The offset value will allow the data to be related to the correct carriageway. Dynamic segmentation is a special type of linear referencing. The term derives from the fact that event data values are held separately from the actual network route in database tables (still as linear distances and offsets) and then dynamically added to the route (segmented) each time the user queries the database. This approach is especially useful in situations in which the

event data change frequently and need to be stored in a database due to access from other applications (e.g., traffic volumes or rate of pipe corrosion).

8.2.3.4 TIN data model

The geographic data models discussed so far have concentrated on one- and two-dimensional data. There are several ways to model three-dimensional data, such as terrain models, sales cost surfaces, and geologic strata. The term 2.5-D is sometimes used to describe surface structure because they have dimensional properties between 2-D and 3-D (Box 3.4). A true 3-D structure will contain multiple z values at the same x, y location and thus is able to model overhangs and tunnels, and support accurate volumetric calculations like cut and fill (a term derived from civil engineering applications that describes cutting earth from high areas and placing it in low areas to construct a flat surface, as is required in, for example, railroad construction). Both grids and triangulated irregular networks (TINs) are used to create and represent surfaces in GIS. A regular grid surface is really a type of raster dataset as discussed earlier in Section 8.2.2. Each grid cell stores the height of the surface at a given location. The TIN structure, as the name suggests, represents a surface as contiguous non-overlapping triangular elements (Figure 8.12). A TIN is created from a set of points, that is, points with x, y, and z coordinate values. A key advantage of the TIN structure is that the density of sampled points, and therefore the size of triangles, can be adjusted to reflect the relief of the surface being modeled, with more points sampled in areas of variable relief (see Section 4.4). TIN surfaces can be created by performing what is called a Delaunay triangulation (Figure 8.13, Section 14.4.4.1). First, a convex hull is created for a dataset – the smallest convex polygon that contains the set of points. Next, straight lines that do not cross each other are drawn from interior points to points on the boundary of the convex hull and to each other. This divides the convex hull into a set of polygons which are then divided into triangles by drawing more lines between vertices of the polygons.

A TIN is a topological data structure that manages information about the nodes comprising each triangle and the neighbors of each triangle. Figure 8.13 shows the topology of a simple TIN. As with other topological data structures, information about a TIN may be conveniently stored in a file or database table, or computed on the fly. TINs offer many advantages for surface analysis. First, they incorporate the original sample points, providing a useful check on the accuracy of the model. Second, the variable density of triangles means that a TIN is an efficient way of storing surface representations such as terrains that have substantial variations in topography. Third, the data structure makes it easy to calculate elevation, slope, aspect, and line-of-sight between points. The combination of these factors has led to the widespread use of the TIN data structure in applications such as volumetric calculations for roadway design, drainage studies for land development, and visualization of urban forms. Figure 8.14 shows two example applications of TINs. Figure 8.14A is a shaded landslide risk TIN of the Pisa district, Italy with building objects draped on top

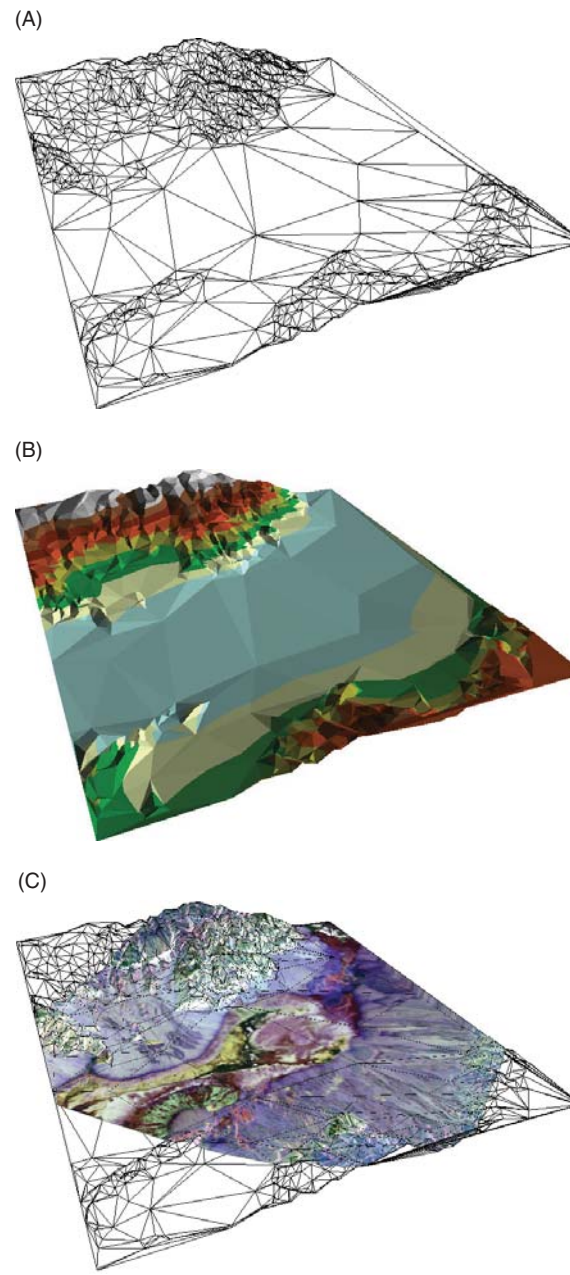
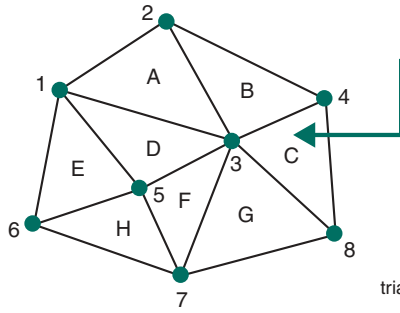


Figure 8.12 TIN surface of Death Valley, California: (A) 'wireframe' showing all triangles; (B) shaded by elevation; (C) draped with satellite image

to give a sense of landscape. Figure 8.14B is a TIN of the Yangtse River, China, greatly exaggerated in the z dimension. It shows how TINs draped with images can provide photo-realistic views of landscapes.

Like all 2.5-D and 3-D models, TINs are only as good as the input sample data. They are especially susceptible to extreme high and low values because there is no smoothing of original data. Other limitations of TINs include their inability to deal with discontinuity of slope across triangle boundaries, the difficulty of calculating

A TIN is a topologic data structure that manages information about the nodes that comprise each triangle and the neighbors to each triangle

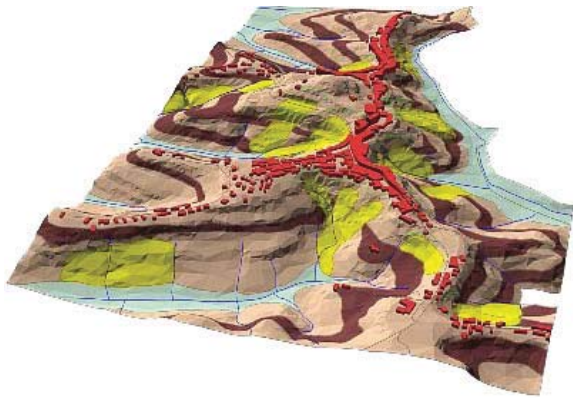


Triangle	Node list	Neighbors
A	1, 2, 3	-, B, D
B	2, 4, 3	-, C, A
C	4, 8, 3	-, G, B
D	1, 3, 5	A, F, E
E	1, 5, 6	D, H, -
F	3, 7, 5	G, H, D
G	3, 8, 7	C, -, F
H	5, 7, 6	F, -, E

Triangles always have three nodes and usually have three neighboring triangles. Triangles on the periphery of the TIN can have one or two neighbors.

Figure 8.13 The topology of a TIN (*Source*: after Zeiler 1999)

(A)



(B)

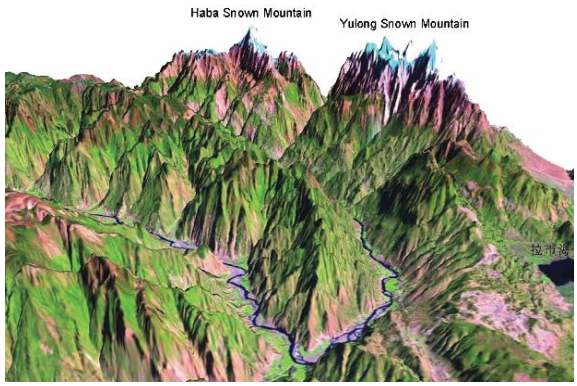


Figure 8.14 Examples of applications that use the TIN data model: (A) Landslide risk map for Pisa, Italy (Courtesy: Earth Science Department, University of Siena, Italy); (B) Yangtse River, China (Courtesy: Human Settlements Research Center, Tsinghua University, China)

optimum routes, and the need to ensure that peaks, pits, ridges, and channels are captured if a drainage network TIN is to be accurate.

8.2.4 Object data model

All the geographic data models described so far are geometry-centric, that is they model the world as collections of points, lines, and areas, TINs, or rasters. Any operations to be performed on the geometry (and, in some cases, associated topology) are created as separate procedures (programs or scripts). Unfortunately, this approach can present several limitations for modeling geographic systems. All but the simplest of geographic systems contain many entities with large numbers of properties, complex relationships, and sophisticated behavior. Modeling such entities as simple geometry types is overly simplistic and does not easily support the sophisticated characteristics required for modern analysis. Additionally, separating the state of an entity (attributes or properties defining what it *is*) from the behavior of an entity (methods defining what it *does*) makes software and database development tedious, time-consuming, and error prone. To try to address these problems geographic object data models were developed. These allow the full richness of geographic systems to be modeled in an integrated way in a GIS.

The central focus of a GIS object data model is the collection of geographic objects and the relationships between the objects (see Box 8.2). Each geographic object is an integrated package of geometry, properties, and methods. In the object data model, geometry is treated like any other attribute of the object and not as its primary characteristic (although clearly from an application perspective it is often the major property of interest). Geographic objects of the same type are grouped together as object classes, with individual objects in the class referred to as 'instances'. In many GIS software systems each object class is stored physically as a database table, with each row an object and each property a column. The methods that apply are attached to the object instances when they are created in memory for use in the application.

An object is the basic atomic unit in an object data model and comprises all the properties that define the state of an object, together with the methods that define its behavior.

Technical Box **8.2**

Object-oriented concepts in GIS

An *object* is a self-contained package of information describing the characteristics and capabilities of an entity under study. An interaction between two objects is called a *relationship*. In a geographic object data model the real world is modeled as a collection of objects and the relationships between the objects. Each entity in the real world to be included in the GIS is an object. A collection of objects of the same type is called a *class*. In fact, classes are a more central concept than objects from the implementation point of view because many of the object-oriented characteristics are built at the class level. A class can be thought of as a template for objects. When creating an object data model the data model designer specifies classes and the relationships between classes. Only when the data model is used to create a database are objects (instances or examples of classes) actually created.

Examples of objects include oil wells, soil bodies, stream catchments, and aircraft flight paths. In the case of an oil well class, each oil well object might include *properties* defining its *state* – annual production, owner name, date of construction, and type of geometry used for representation at a given scale (perhaps a point on a small-scale map and a polygon on a large-scale one). The oil well class could have connectivity relationships with a pipeline class that represents the pipeline used to transfer oil to a refinery. There could also be a relationship defining the fact that each well must be located on a drilling platform. Finally, each oil well object might also have *methods* defining the *behavior* or what it can do. Example behavior might include how objects draw themselves on a computer screen, how objects can be created and deleted, and editing rules about how oil wells snap to pipelines.

There are three key facets of object data models that make them especially good for modeling geographic systems: encapsulation, inheritance, and polymorphism.

Encapsulation describes the fact that each object packages together a description of its state and behavior. The state of an object can be thought of as its properties or attributes (e.g., for a forest object it could be the dominant tree type, average tree age, and soil pH). The behavior is the methods or operations that can be performed on an object (for a forest object these could be create, delete, draw, query, split, and merge). For example, when splitting a forest polygon into two parts, perhaps following a part sale, it is useful to get the GIS to automatically calculate the areas of the two new parts. Combining the state and behavior of an object together in a single package is a natural way to think of geographic entities and a useful way to support the reuse of objects.

Inheritance is the ability to reuse some or all of the characteristics of one object in another object. For example, in a gas facility system a new type of gas valve could easily be created by overwriting or adding a few properties or methods to a similar existing type of valve. Inheritance provides an efficient way to create models of geographic systems by reusing objects and also a mechanism to extend models easily. New object classes can be built to reuse parts of one or more existing object classes and add some new unique properties and methods. The example described in Section 8.3 shows how inheritance and other object characteristics can be used in practice.

Polymorphism describes the process whereby each object has its own specific implementation for operations like draw, create, and delete. One example of the benefit of polymorphism is that a geographic database can have a generic object creation component that issues requests to be processed in a specific way by each type of object class. A utility system's editor software can send a generic create request to all objects (e.g., gas pipes, valves, and service lines) each of which has specific create algorithms. If a new object class is added to the system (e.g., landbase) then this mechanism will work because the new class is responsible for implementing the create method. Polymorphism is essential for isolating parts of software as self-contained components (see Chapter 10).

All geographic objects have some type of relationship to other objects in the same object class and, possibly, to objects in other object classes. Some of these relationships are inherent in the class definition (for example, some GIS remove overlapping polygons) while other interclass relationships are user-definable. Three types of relationships are commonly used in geographic object data models: topological, geographic, and general.

A class is a template for creating objects.

Generally, topological relationships are built into the class definition. For example, modeling real-world entities as a network class will cause network topology to be built for the nodes and lines participating in the network. Similarly, real-world entities modeled as topological polygon classes will be structured using the node–polyline model described in Section 8.2.3.2.

Geographic relationships between object classes are based on geographic operators (such as overlap, adjacency, inside, and touching) that determine the interaction between objects. In a model of an agricultural system, for example, it might be useful to ensure that all farm buildings are within a farm boundary using a test for geographic containment.

General relationships are useful to define other types of relationship between objects. In a tax assessment system, for example, it is advantageous to define a relationship between land parcels (polygons) and ownership data that is stored in an associated DBMS table. Similarly, an electric distribution system relating light poles (points) to text strings (called annotation) allows depiction of pole height and material of construction on a map display. This type of information is very valuable for creating work orders (requests for change) that alter the facilities. Establishing relationships between objects in this way is useful because if one object is moved then the other will move as well, or if one is deleted then the other is also deleted. This makes maintaining databases much easier and safer.

In addition to supporting relationships between objects (strictly speaking, between object classes), object data models also allow several types of rules to be defined. Rules are a valuable means of maintaining database integrity during editing tasks. The most popular types of rules used in object data models are attribute, connectivity, relationship, and geographic.

Attribute rules are used to define the possible attribute values that can be entered for any object. Both range and coded value attribute rules are widely employed. A range attribute rule defines the range of valid values that can be entered. Examples of range rules include: highway traffic speed must be in the range 25–70 miles (40–120 km) per hour; forest compartment average tree height must be in the range 0–50 meters. Coded attribute rules are used for categorical data types. Examples include: land use must be of type commercial, residential, park, or other; or pipe material must be of type steel, copper, lead, or concrete.

Connectivity rules are based on the specification of valid combinations of features, derived from the geometry, topology, and attribute properties. For example, in an electric distribution system a 28.8 kV conductor can only connect to a 14.4 kV conductor via a transformer. Similarly, in a gas distribution system it should not be possible to add pipes with free ends (that is, with no fitting or cap) to a database.

Geographic rules define what happens to the properties of objects when an editor splits or merges them (Figure 8.15). In the case of a land parcel split following the sale of part of the parcel, it is useful to define rules to determine the impact on properties like area, land use code, and owner. In this example, the original parcel area value should be divided in proportion to the size of the

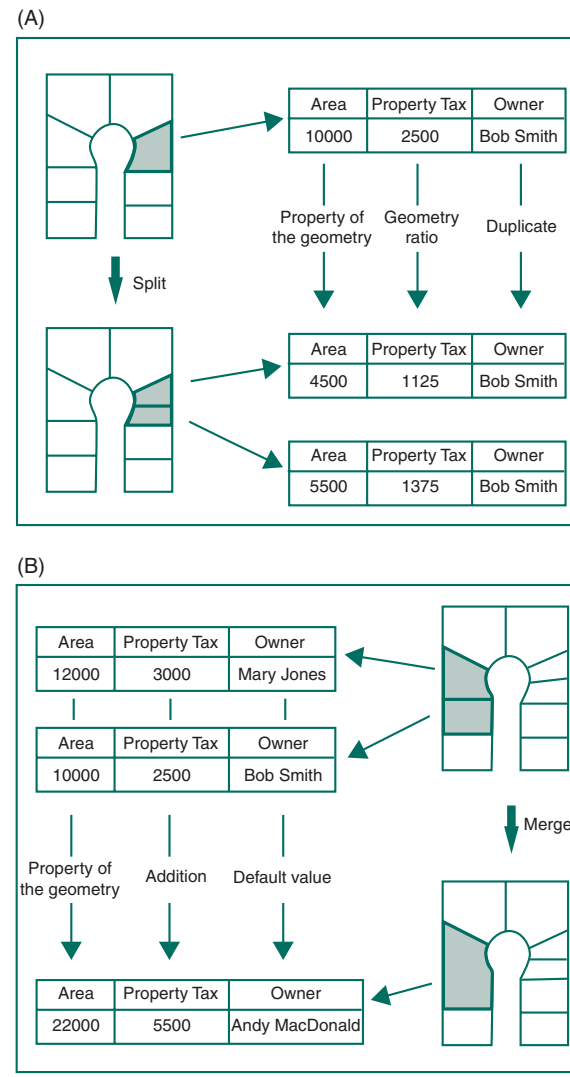


Figure 8.15 Example of split and merge rules for parcel objects: (A) split; (B) merge (*Source*: after MacDonald 1999)

two new parcels, the land use code should be transferred to both parcels, and the owner name should remain for one parcel, but a new one should be added for the part that was sold off. In the case of a merge of two adjacent water pipes, decisions need to be made about what happens to attributes like material, length, and corrosion rate. In this example, the two pipe materials should be the same, the lengths should be summed, and the new corrosion rate determined by a weighted average of both pipes.

8.3 Example of a water-facility object data model

The goal of this section is to describe an example of a geographic object model. It will discuss how many of

the concepts introduced earlier in this chapter are used in practice. The example selected is that of an urban water-facility model. The types of issues raised in this example apply to all geographic object models, although of course the actual objects, object classes, and relationships under consideration will differ. The role of data modeling, as discussed in Section 8.1, is to represent the key aspects of the real world inside the digital computer for management, analysis, and display purposes.

Figure 8.16 is a diagram of part of a water distribution system, a type of pressurized network controlled by several devices. A pump is responsible for moving water through pipes (mains and laterals) connected together by fittings. Meters measure the rate of water consumption at houses. Valves and hydrants control the flow of water.

The purpose of the example object model is to support asset management, mapping, and network analysis applications. Based on this it is useful to classify the objects into two types: the landbase and the water-facilities. Landbase is a general term for objects like houses and streets that provide geographic context but are not used in network analysis. The landbase object types are Pump House, House, and Street. The water-facilities object types are: Main, Lateral (a smaller type of WaterLine), Fitting (water line connectors), Meter, Valve, and Hydrant. All of these object types need to be modeled as a network in order to support network analysis operations like network isolation traces and flow prediction. A network isolation trace is used to find all parts of a network that are unconnected (isolated). Using the topological connectivity of the network and information about whether pipes and fittings support water flow, it is possible to determine connectivity. Flow prediction is used to estimate the flow of water through the network based on network connectivity and data about water availability and consumption. Figure 8.17 shows all the main object types and the implicit geographic relationships to be incorporated into the model. The arrows indicate the direction of flow in the network. When digitizing this network using a GIS editor it will be useful to specify topological connectivity and attribute rules to control how objects can be connected (see Section 8.2.3.3 above). Before this network can be used for analysis it will also be necessary to add flow impedances to each link (for example, pipe diameter).

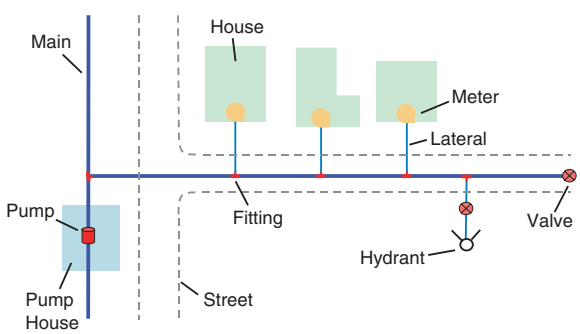


Figure 8.16 Water distribution system water-facility object types and geographic relationships

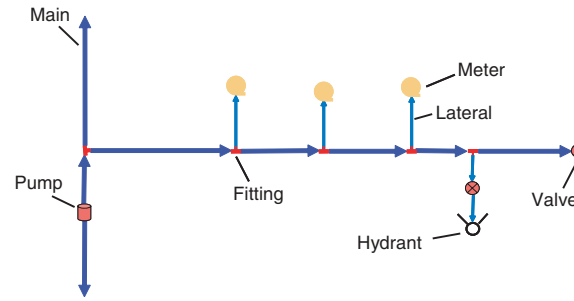


Figure 8.17 Water distribution system network

Having identified the main object types, the next step is to decide how objects relate to each other and the most efficient way to implement them. Figure 8.18 shows one possible object model that uses the Unified Modeling Language (UML) to show objects and the relationships between them. Some additional color-coding has been added to help interpret the model. In UML models each box is an object class and the lines define how one class reuses (inherits) part of the class above it in a hierarchy.

Object class names in an italic font are abstract classes; those with regular font names are used to create (instantiate) actual object instances. Abstract classes do not have instances and exist for efficiency reasons. It is sometimes useful to have a class that implements some capabilities once, so that several other classes can then be reused. For example, Main and Lateral are both types of *Line*, as is Street. Because Main and Lateral share several things in common – such as ConstructionMaterial, Diameter, and InstallDate properties, and connectivity and draw behavior – it is efficient to implement these in a separate abstract class, called *WaterLine*. The triangles indicate that one class is a type of another class. For example, Pump House and House are types of *Building*, and Street and *WaterLine* are types of *Line*. The diamonds indicate composition. For example, a network is composed of a collection of *Line* and *Node* objects. In the water-facility object model, object classes without any geometry are colored pink. The Equipment and OperationsRecord object classes have their location determined by being associated with other objects (e.g., valves and mains). The Equipment and OperationsRecord classes are useful places to store properties common to many facilities, such as EquipmentID, InstallDate, ModelNumber, and SerialNumber.

Once this logical geographic object model has been created it can be used to generate a physical data model. One way to do this is to create the model using a computer-aided software engineering (CASE) tool. A CASE tool is a software application that has graphical tools to draw and specify a logical model (Figure 8.19). A further advantage of a CASE tool is that physical models can be generated directly from the logical models, including all the database tables and much of the supporting code for implementing behavior. Once a database structure (schema) has been created, it can be populated with objects and the intended applications put into operation.

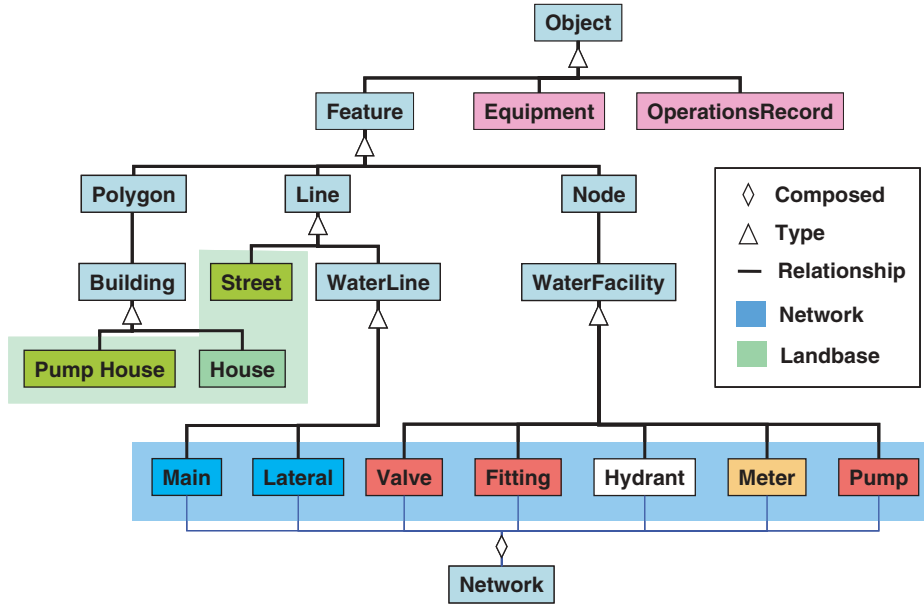


Figure 8.18 A water-facility object model

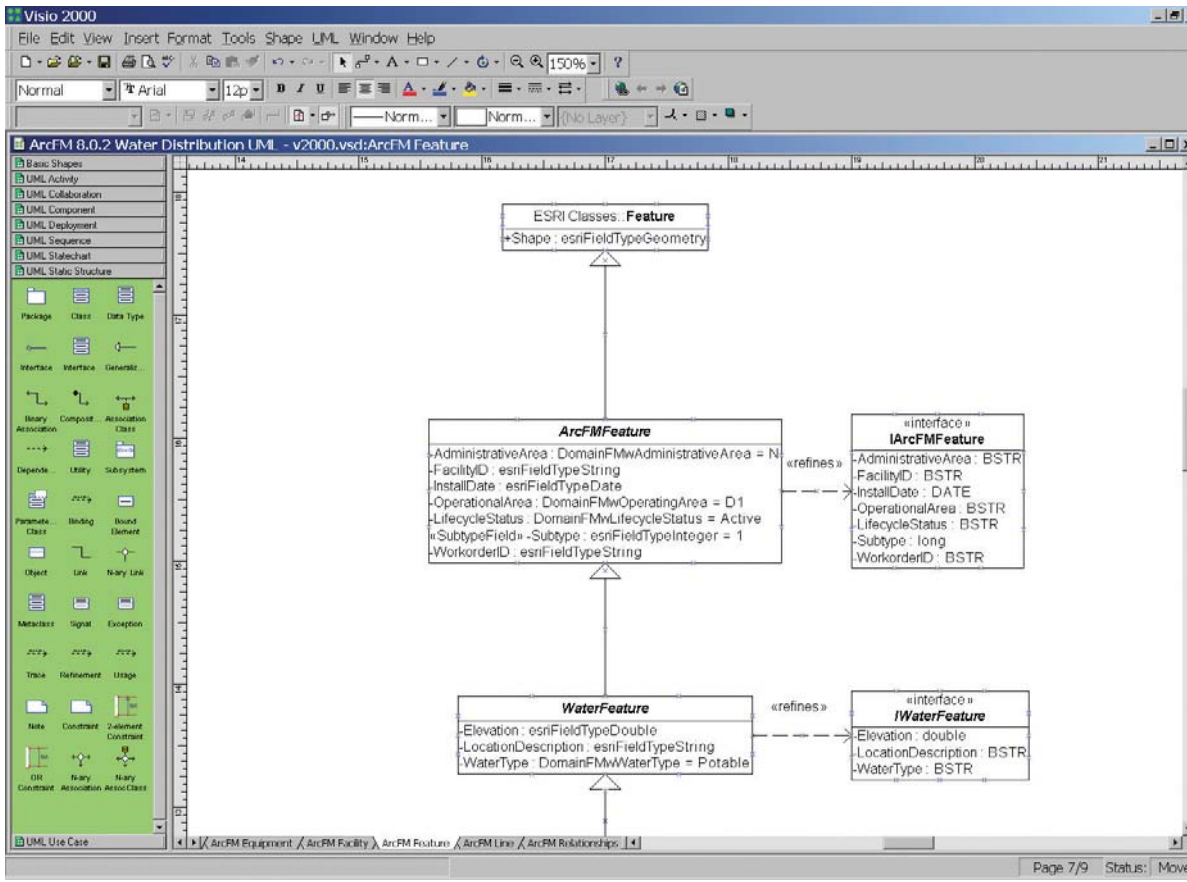


Figure 8.19 An example of a CASE tool (Microsoft Visio). The UML model is for a utility water system

8.4 Geographic data modeling in practice

Geographic analysis is only as good as the geographic database on which it is based and a geographic database is only as good as the geographic data model from which it is derived. Geographic data modeling begins with a clear definition of the project goals and progresses through an understanding of user requirements, a definition of the objects and relationships, formulation of a logical model, and then creation of a physical model. These steps are a prelude to database creation and, finally, database use. In Box 8.3 Leslie Cone of the US Bureau of Land Management describes her experience of creating a land management system with a parcel data model.

No step in data modeling is more important than understanding the purpose of the data modeling exercise. This understanding can be gained by collecting user requirements from the main users. Initially, user requirements will be vague and ill-defined, but over time they will become clearer. Project goals and user requirements should be precisely specified in a list or narrative.

Biographical Box **8.3**

Leslie M. Cone, Project Manager, BLM Land and Resources Project Office

The BLM (US Bureau of Land Management) administers some 261 million surface acres of America's public lands, located primarily in 12 Western States. The BLM mission is 'to sustain the health, diversity, and productivity of the public lands for the use and enjoyment of present and future generations'.

Figure 8.20 Leslie Cone, project manager (courtesy Leslie Cone)

Those who use the Internet to access BLM's land resources and data owe many thanks to Leslie Cone, a 31-year employee of the BLM. As Project Manager for the Land and Resources Project Office, she leads the team of BLM employees, interns, and contractors who manage 24 national-level projects that include the National Integrated Land System (NILS), BLM's contribution to Geospatial One-Stop Portal, and the Automated Fluid Minerals Support System. According to Leslie 'The mission of the Land and Resources Project Office is to provide automation of BLM's land and resources data and the tools to use and manage them'.

Addressing her contributions to the GIS, Leslie says, 'NILS was initiated to create a business solution for land managers who face an increasingly complex environment of land transactions, legal challenges, and deteriorating and difficult-to-access records'. She adds, 'This complex task was made even more challenging because the Bureau's existing land records date back over 200 years'.

The NILS project was the first step towards providing a common parcel-based solution for sharing land record information within the US government and the private sector. NILS is a joint development project of the BLM and the US Forest Service, in partnership with states, counties, and private industry. The project developed a common parcel data model and a set of software tools for the collection, management, and sharing of survey-based data, and land record information. The data model resulted from an extensive analysis of user requirements, the accumulated experience of previous-generation systems, and several prototypes. It has been crucial to the success of the project.

To government agencies, commercial businesses, students, and the general public, this provides an easy way to perform research and other tasks that were formerly a major challenge. For the public, all that is needed to access the public data is an Internet connection. The NILS GeoCommunicator application is an interactive Web-based land information portal that allows users to share, search, locate, view, and access geographic information, data, maps, and images. In addition, NILS includes proprietary applications for the government and private business sectors.

Since she became the Project Manager six years ago, the BLM Land and Resources Project Office has grown from a few individuals to approximately 60 managers, engineers, programmers, analysts, Web developers, technical writers, and student interns. Leslie Cone holds a Bachelor of Science degree in Forestry and Outdoor Recreation from Colorado State University and a Master's degree in Public Administration from the University of New Mexico.

Formulation of a logical model necessitates identification of the objects and relationships to be modeled. Both the attributes and behavior of objects are required for an object model. A useful graphic tool for creating logical data models is a CASE tool and a useful language for specifying models is UML. It is not essential that all objects and relationships be identified at the first attempt because logical models can be refined over time. The key objects and relationships for the water distribution system object model are shown in Figure 8.18.

Once an implementation-independent logical model has been created, this model can be turned into a system-dependent physical model. A physical model will result in an empty database schema – a collection of database tables and the relationships between them. Sometimes, for performance optimization reasons or because of changing requirements, it is necessary to alter the physical data model. Even at this relatively late stage in the process, flexibility is still necessary.

It is important to realize that there is no such thing as the correct geographic data model. Every problem can be represented with many possible data models. Each data model is designed with a specific purpose in mind and is sub-optimal for other purposes. A classic dilemma is whether to define a general-purpose data model that has wide applicability, but that can, potentially, be complex and inefficient, or to focus on a narrower highly optimized model. A small prototype can often help resolve some of these issues.

Geographic data modeling is both an art and a science. It requires a scientific understanding of the key geographic characteristics of real-world systems, including the state and behavior of objects, and the relationships between them. Geographic data models are of critical importance because they have a controlling influence over the type of data that can be represented and the operations that can be performed. As we have seen, object models are the best type of data model for representing rich object types and relationships in facility systems, whereas simple feature models are sufficient for elementary applications such as a map of the body. In a similar vein, so to speak, raster models are good for data represented as fields such as soils, vegetation, pollution, and population counts.

Questions for further study

1. Figure 8.21 is an oblique aerial photograph of part of the city of Kfar-Saba, Israel. Take ten minutes to list all the object classes (including their attributes and behavior) and the relationships between the classes that you can see in this picture that would be appropriate for a city information system study.
2. Why is it useful to include the conceptual, logical, and physical levels in geographic data modeling?
3. Describe, with examples, five key differences between the topological vector and raster geographic data models. It may be useful to consult Figure 8.3 and Chapter 3.
4. Review the terms encapsulation, inheritance, and polymorphism and explain with geographic examples why they make object data models superior for representing geographic systems.

Figure 8.21 Oblique aerial view of Kfar-Saba, Israel (Courtesy: ESRI)

Further reading

Arctur D. and Zeiler M. 2004 *Designing Geodatabases: Case Studies in GIS Data Modeling*. Redlands, CA: ESRI Press.

ESRI 1997 *Understanding GIS: the ArcInfo Method*. Redlands, CA: ESRI Press.

MacDonald A. 1999 *Building a Geodatabase*. Redlands, CA: ESRI Press.

Worboys M.F. and Duckham M. 2004 *GIS: A Computing Perspective* (2nd edn). Boca Raton, FL: CRC Press.

Zeiler M. 1999 *Modeling Our World: The ESRI Guide to Geodatabase Design*. Redlands, CA: ESRI Press.

9 *GIS data collection*

Data collection is one of the most time-consuming and expensive, yet important, of GIS tasks. There are many diverse sources of geographic data and many methods available to enter them into a GIS. The two main methods of data collection are data capture and data transfer. It is useful to distinguish between primary (direct measurement) and secondary (derivation from other sources) data capture for both raster and vector data types. Data transfer involves importing digital data from other sources. There are many practical issues associated with planning and executing an effective GIS data collection plan. This chapter reviews the main methods of GIS data capture and transfer and introduces key practical management issues.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will be able to:

- Describe data collection workflows;
- Understand the primary data capture techniques in remote sensing and surveying;
- Be familiar with the secondary data capture techniques of scanning, manual digitizing, vectorization, photogrammetry, and COGO feature construction;
- Understand the principles of data transfer, sources of digital geographic data, and geographic data formats;
- Analyze practical issues associated with managing data capture projects.

9.1 Introduction

GIS can contain a wide variety of geographic data types originating from many diverse sources. Data collection activities for the purposes of organizing the material in this chapter are split into data capture (direct data input) and data transfer (input of data from other systems). From the perspective of creating geographic databases, it is convenient to classify raster and vector geographic data as primary and secondary (Table 9.1). Primary data sources are those collected in digital format specifically for use in a GIS project. Typical examples of primary GIS sources include raster SPOT and IKONOS Earth satellite images, and vector building-survey measurements captured using a total survey station. Secondary sources are digital and analog datasets that were originally captured for another purpose and need to be converted into a suitable digital format for use in a GIS project. Typical secondary sources include raster scanned color aerial photographs of urban areas and United States Geological Survey (USGS) or Institut Géographique National, France (IGN) paper maps that can be scanned and vectorized. This classification scheme is a useful organizing framework for this chapter and, more importantly, it highlights the number of processing-stage transformations that a dataset goes through, and therefore the opportunities for errors to be introduced. However, the distinctions between primary and secondary, and raster and vector, are not always easy to determine. For example, is digital satellite remote sensing data obtained on a DVD primary or secondary? Clearly the commercial satellite sensor feeds do not run straight into GIS databases, but to ground stations where the data are pre-processed onto digital media. Here it is considered primary because usually the data has undergone only minimal transformation since being collected by the satellite sensors and because the characteristics of the data make them suitable for virtually direct use in GIS projects.

Table 9.1 Classification of geographic data for data collection purposes with examples of each type

	Raster	Vector
Primary	Digital satellite remote-sensing images	GPS measurements
	Digital aerial photographs	Survey measurements
Secondary	Scanned maps or photographs	Topographic maps
	Digital elevation models from topographic map contours	Toponymy (placename) databases

Primary geographic data sources are captured specifically for use in GIS by direct measurement. Secondary sources are those reused from earlier studies or obtained from other systems.

Both primary and secondary geographic data may be obtained in either digital or analog format (see Section 3.7 for a definition of analog). Analog data must always be digitized before being added to a geographic database. Analog to digital transformation may involve the scanning of paper maps or photographs, optical character recognition (OCR) of text describing geographic object properties, or the vectorization of selected features from an image. Depending on the format and characteristics of the digital data, considerable reformatting and restructuring may be required prior to importing into a GIS. Each of these transformations alters the original data and will introduce further uncertainty into the data (see Chapter 6 for discussion of uncertainty).

This chapter describes the data sources, techniques, and workflows involved in GIS data collection. The processes of data collection are also variously referred to as data capture, data automation, data conversion, data transfer, data translation, and digitizing. Although there are subtle differences between these terms, they essentially describe the same thing, namely, adding geographic data to a database. Data capture refers to direct entry. Data transfer is the importing of existing digital data across a network connection (Internet, wide area network (WAN), or local area network (LAN)) or from physical media such as CD ROMs, zip disks, or diskettes. This chapter focuses on the techniques of data collection; of equal, perhaps more, importance to a real-world GIS implementation are project management, cost, legal, and organization issues. These are covered briefly in Section 9.6 of this chapter as a prelude to more detailed treatment in Chapters 17 through 20.

Table 9.2 Breakdown of costs (in $1000s) for two typical client-server GIS as estimated by the authors

	10 seats		100 seats	
	$	%	$	%
Hardware	30	3.4	250	8.6
Software	25	2.8	200	6.9
Data	400	44.7	450	15.5
Staff	440	49.1	2000	69.0
Total	895	100	2900	100

Table 9.2 shows a breakdown of costs (in $1000s) for two typical client-server GIS implementations: one with 10 seats (systems) and the other with 100. The hardware costs include desktop clients and servers only (i.e., not network infrastructure). The data costs assume the purchase of a landbase (e.g., streets, parcels, and land marks) and digitizing assets such as pipes and fittings (water utility), conductors and devices (electrical utility), or land and property parcels (local government). Staff costs assume that all core GIS staff will be full-time, but that users will be part-time.

In the early days of GIS, when geographic data were very scarce, data collection was the main project task and typically it consumed the majority of the available resources. Even today data collection still remains a time-consuming, tedious, and expensive process. Typically it accounts for 15–50% of the total cost of a GIS project (Table 9.2). Data capture costs can in fact be much more significant because in many organizations (especially those that are government funded) staff costs are often assumed to be fixed and are not used in budget accounting. Furthermore, as the majority of data capture effort and expense tends to fall at the start of projects, data capture costs often receive greater scrutiny from senior managers. If staff costs are excluded from a GIS budget then in cash expenditure terms data collection can be as much as 60–85% of costs.

Data capture costs can account for up to 85% of the cost of a GIS.

After an organization has completed basic data collection tasks, the focus of a GIS project moves on to data maintenance. Over the multi-year lifetime of a GIS project, data maintenance can turn out to be a far more complex and expensive activity than initial data collection. This is because of the high volume of update transactions in many systems (for example, changes in land parcel ownership, maintenance work orders on a highway transport network, or logging military operational activities) and the need to manage multi-user access to operational databases. For more information about data maintenance, see Chapter 10.

9.1.1 Data collection workflow

In all but the simplest of projects, data collection involves a series of sequential stages (Figure 9.1). The workflow

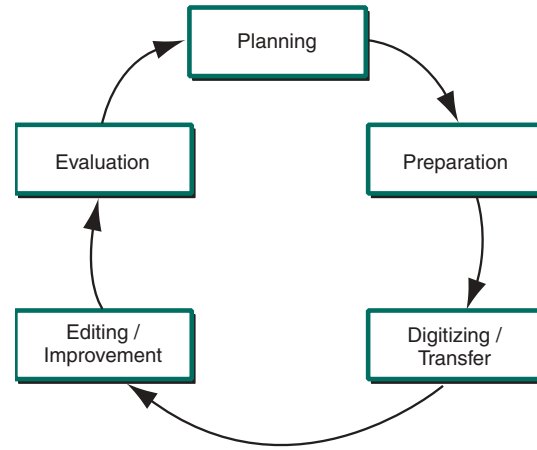


Figure 9.1 Stages in data collection projects

commences with planning, followed by preparation, digitizing/transfer (here taken to mean a range of primary and secondary techniques such as table digitizing, survey entry, scanning, and photogrammetry), editing and improvement and, finally, evaluation.

Planning is obviously important to any project and data collection is no exception. It includes establishing user requirements, garnering resources (staff, hardware, and software), and developing a project plan. Preparation is especially important in data collection projects. It involves many tasks such as obtaining data, redrafting poor-quality map sources, editing scanned map images, and removing noise (unwanted data such as speckles on a scanned map image). It may also involve setting up appropriate GIS hardware and software systems to accept data. Digitizing and transfer are the stages where the majority of the effort will be expended. It is naïve to think that data capture is really just digitizing, when in fact it involves very much more as discussed below. Editing and improvement follows digitizing/transfer. This covers many techniques designed to validate data, as well as correct errors and improve quality. Evaluation, as the name suggests, is the process of identifying project successes and failures. These may be qualitative or quantitative. Since all large data projects involve multiple stages, this workflow is iterative with earlier phases (especially a first, pilot, phase) helping to improve subsequent parts of the overall project.

9.2 Primary geographic data capture

Primary geographic capture involves the direct measurement of objects. Digital data measurements may be input directly into the GIS database, or can reside in a temporary file prior to input. Although the former is preferable as it minimizes the amount of time and the possibility of errors, close coupling of data collection devices and GIS databases is not always possible. Both raster and vector GIS primary data capture methods are available.

9.2.1 Raster data capture

Much the most popular form of primary raster data capture is remote sensing. Broadly speaking, remote sensing is a technique used to derive information about the physical, chemical, and biological properties of objects without direct physical contact (Section 3.6). Information is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects. A variety of sensors, operating throughout the electromagnetic spectrum from visible to microwave wavelengths, are commonly employed to obtain measurements (see Section 3.6.1). Passive sensors are reliant on reflected solar radiation or emitted terrestrial radiation; active sensors (such as synthetic aperture radar) generate their own source of electromagnetic radiation. The platforms on which these instruments are mounted are similarly diverse. Although Earth-orbiting satellites and fixed-wing aircraft are by far the most common, helicopters, balloons, masts, and booms are also employed (Figure 9.2). As used here, the term *remote sensing* subsumes the fields of satellite remote sensing and aerial photography.

Remote sensing is the measurement of physical, chemical, and biological properties of objects without direct contact.

From the GIS perspective, resolution is a key physical characteristic of remote sensing systems. There are three aspects to resolution: spatial, spectral, and temporal. All sensors need to trade off spatial, spectral, and temporal properties because of storage, processing, and bandwidth considerations. For further discussion of the important topic of resolution see also Sections 3.4, 3.6.1, 4.1, 6.4.2, 7.1, and 16.1.

Three key aspects of resolution are: spatial, spectral, and temporal.

Spatial resolution refers to the size of object that can be resolved and the most usual measure is the pixel size. Satellite remote sensing systems typically provide data with pixel sizes in the range 0.5 m–1 km. The resolution of cameras used for capturing aerial photographs usually ranges from 0.1 m–5 m. Image (scene) sizes vary quite widely between sensors – typical ranges include 900 by 900 to 3000 by 3000 pixels. The total coverage of remote sensing images is usually in the range 9 by 9 to 200 by 200 km.

Spectral resolution refers to the parts of the electromagnetic spectrum that are measured. Since different objects emit and reflect different types and amounts of radiation, selecting which part of the electromagnetic spectrum to measure is critical for each application area. Figure 9.3 shows the spectral signatures of water, green vegetation, and dry soil. Remote sensing systems may capture data in one part of the spectrum (referred to as a single band) or simultaneously from several parts (multi-band or multi-spectral). The radiation values are usually normalized and resampled to give a range of integers from 0–255 for each band (part of the electromagnetic spectrum measured), for each pixel, in each image. Until recently, remote sensing satellites typically measured a small number of bands, in the visible part of the spectrum. More recently a number of hyperspectral systems have come into operation that measure very large numbers of bands across a much wider part of the spectrum.

Temporal resolution, or repeat cycle, describes the frequency with which images are collected for the same area. There are essentially two types of commercial remote sensing satellite: Earth-orbiting and geostationary. Earth-orbiting satellites collect information about different parts of the Earth surface at regular intervals. To maximize utility, typically orbits are polar, at a fixed altitude and speed, and are Sun synchronous.

The French SPOT (Système Probatoire d'Observation de la Terre) 5 satellite launched in 2002, for example, passes virtually over the poles at an altitude of 822 km sensing the same location on the Earth surface during daylight every 26 days. The SPOT platform carries multiple sensors: a panchromatic sensor measuring radiation in the visible part of the electromagnetic spectrum at a spatial resolution of 2.5 by 2.5 m; a multi-spectral sensor measuring green, red, and reflected infrared radiation at a spatial resolution of 10 by 10 m; a shortwave near-infrared sensor with a resolution of 20 by 20 m; and a vegetation sensor measuring four bands at a spatial resolution of 1000 m. The SPOT system is also able to provide stereo images from which digital terrain models and 3-D measurements can be obtained. Each SPOT scene covers an area of about 60 by 60 km.

Much of the discussion so far has focused on commercial satellite remote sensing systems. Of equal importance, especially in medium- to large (coarse)-scale GIS projects, is aerial photography. Although the data products resulting from remote sensing satellites and aerial photography systems are technically very similar (i.e., they are both images) there are some significant differences in the way data are captured and can, therefore, be interpreted. The most notable difference is that aerial photographs are normally collected using analog optical cameras (although digital cameras are becoming more widely used) and then later rasterized, usually by scanning a film negative. The quality of the optics of the camera and the mechanics of the scanning process both affect the spatial and spectral characteristics of the resulting images. Most aerial photographs are collected on an *ad hoc* basis using cameras mounted in airplanes flying at low altitudes (3000–9000 m) and are either panchromatic (black and white) or color, although multi-spectral cameras/sensors operating in the non-visible parts of the electromagnetic spectrum are also used. Aerial photographs are very suitable for detailed surveying and mapping projects.

An important feature of satellite and aerial photography systems is that they can provide stereo imagery from overlapping pairs of images. These images are used to create a 3-D analog or digital model from which 3-D coordinates, contours, and digital elevation models can be created (see Section 9.3.2.4).

Satellite and aerial photograph data offer a number of advantages for GIS projects. The consistency of the data and the availability of systematic global coverage

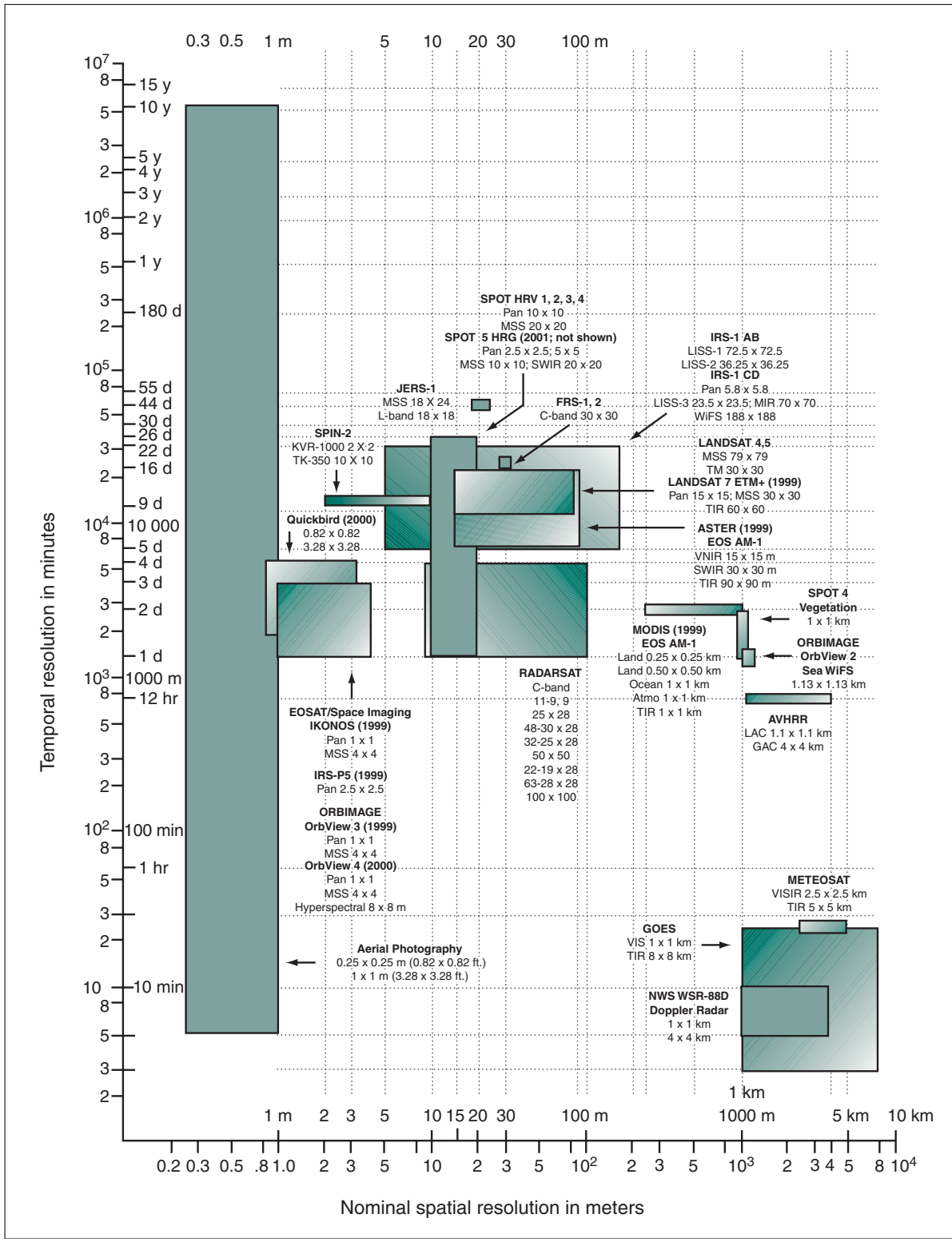


Figure 9.2 Spatial and temporal characteristics of commonly used remote sensing systems and their sensors (*Source*: after Jensen J.R. and Cowen D.C. 1999 'Remote sensing of urban/suburban infrastructure and socioeconomic attributes', *Photogrammetric Engineering and Remote Sensing* **65**, 611–622)

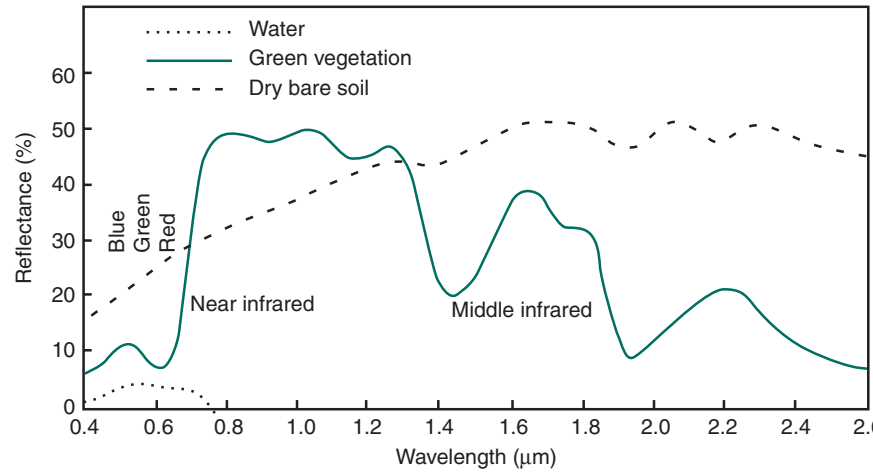


Figure 9.3 Typical reflectance signatures for water, green vegetation, and dry soil (*Source*: after Jones C. 1997 *Geographic Information Systems and Computer Cartography*. Reading, MA: Addison-Wesley Longman)

make satellite data especially useful for large-area, small-scale projects (for example, mapping landforms and geology at the river catchment-area level) and for mapping inaccessible areas. The regular repeat cycles of commercial systems and the fact that they record radiation in many parts of the spectrum make such data especially suitable for assessing the condition of vegetation (for example, the moisture stress of wheat crops). Aerial photographs in particular are very useful for detailed surveying and mapping of, for example, urban areas and archaeological sites, especially those applications requiring 3-D data (see Chapter 12).

On the other hand, the spatial resolution of commercial satellites is too coarse for many large-scale projects and the data collection capability of many sensors is restricted by cloud cover. Some of this is changing, however, as the new generation of satellite sensors now provide data at 0.6 m spatial resolution and better, and radar data can be obtained that are not affected by cloud cover. The data volumes from both satellites and aerial cameras can be very large and create storage and processing problems for all but the most modern systems. The cost of data can also be prohibitive for a single project or organization.

9.2.2 Vector data capture

Primary vector data capture is a major source of geographic data. The two main branches of vector data capture are ground surveying and GPS – which is covered in Section 5.8 – although as more surveyors use GPS routinely the distinction between the two is becoming increasingly blurred.

Ground surveying is based on the principle that the 3-D location of any point can be determined by measuring angles and distances from other known points. Surveys begin from a benchmark point. If the coordinate system of this point is known, all subsequent points can be collected in this coordinate system. If it is unknown then the survey will use a local or relative coordinate system (see Section 5.7).

Since all survey points are obtained from survey measurements, their known locations are always relative to other points. Any measurement errors need to be apportioned between multiple points in a survey. For example, when surveying a field boundary, if the last and first points are not identical in survey terms (within the tolerance employed in the survey) then errors need to be apportioned between all points that define the boundary (see Section 6.3.4). As new measurements are obtained these may change the locations of points.

Traditionally, surveyors used equipment like transits and theodolites to measure angles, and tapes and chains to measure distances. Today these have been replaced by electro-optical devices called total stations that can measure both angles and distances to an accuracy of 1 mm (Figure 9.4). Total stations automatically log data and the most sophisticated can create vector point, line, and area objects in the field, thus providing direct validation.

The basic principles of surveying have changed very little in the past 100 years, although new technology has considerably improved accuracy and productivity. Two people are usually required to perform a survey, one to operate the total station and the other to hold a reflective prism that is placed at the object being measured. On some remote-controlled systems a single person can control both the total station and the prism.

Ground survey is a very time-consuming and expensive activity, but it is still the best way to obtain highly accurate point locations. Surveying is typically used for capturing buildings, land and property boundaries, manholes, and other objects that need to be located accurately. It is also employed to obtain reference marks for use in other data capture projects. For example, large-scale aerial photographs and satellite images are frequently georeferenced using points obtained from ground survey.

Figure 9.4 A tripod-mounted Leica TPS1100 Total Station (Courtesy: Leica Geosystems)

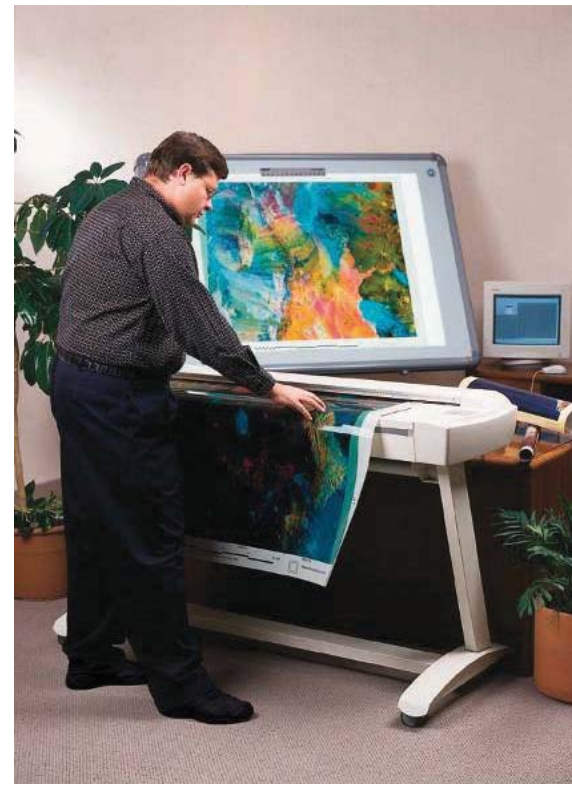


Figure 9.5 A large-format roll-feed image scanner (Reproduced by permission of GTCO Calcomp, Inc.)

9.3 Secondary geographic data capture

Geographic data capture from secondary sources is the process of creating raster and vector files and databases from maps, photographs, and other hard-copy documents. Scanning is used to capture raster data. Table digitizing, heads-up digitizing, stereo-photogrammetry, and COGO data entry are used for vector data.

9.3.1 Raster data capture using scanners

A scanner is a device that converts hard-copy analog media into digital images by scanning successive lines across a map or document and recording the amount of light reflected from a local data source (Figure 9.5). The differences in reflected light are normally scaled into bi-level black and white (1 bit per pixel), or multiple gray levels (8, 16, or 32 bits). Color scanners output data into 8-bit red, green, and blue color bands. The spatial resolution of scanners varies widely from as little as 200 dpi (8 dots per mm) to 2400 dpi (96 dots per mm) and beyond. Most GIS scanning is in the range 400–900 dpi (16–40 dots per mm). Depending on the type of scanner and the resolution required, it can take from 30 seconds to 30 minutes or more to scan a map.

Scanned maps and documents are used extensively in GIS as background maps and data stores.

There are three main reasons to scan hardcopy media for use in GIS:

- Documents, such as building plans, CAD drawings, property deeds, and equipment photographs are scanned to reduced wear and tear, improve access, provide integrated database storage, and to index them geographically (e.g., building plans can be attached to building objects in geographic space).
- Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for other data (typically vector layers). This type of unintelligent image or background geographic wall-paper is very popular in systems that manage equipment and land and property assets (Figure 9.6).
- Maps, aerial photographs, and images are scanned prior to vectorization (see below), and sometimes as a prelude to spatial analysis.

An 8 bit (256 gray level) 400 dpi (16 dots per mm) scanner is a good choice for scanning maps for use as a background GIS reference layer. For a color aerial

Figure 9.6 An example of raster background data (black and white aerial photography) underneath vector data (land parcels)

photograph that is to be used for subsequent photo-interpretation and analysis, a color (8 bit for each of three bands) 900 dpi (40 dots per mm) scanner is more appropriate. The quality of data output from a scanner is determined by the nature of the original source material, the quality of the scanning device, and the type of preparation prior to scanning (e.g., redrafting key features or removing unwanted marks will improve output quality).

9.3.2 Vector data capture

Secondary vector data capture involves digitizing vector objects from maps and other geographic data sources. The most popular methods are manual digitizing, heads-up digitizing and vectorization, photogrammetry, and COGO data entry.

9.3.2.1 Manual digitizing

Manually operated digitizers are much the simplest, cheapest, and most commonly used means of capturing vector objects from hardcopy maps. Digitizers come in several designs, sizes, and shapes. They operate on the principle that it is possible to detect the location of a cursor or puck passed over a table inlaid with a fine mesh of wires. Digitizing table accuracies typically range from 0.0004 inch (0.01 mm) to 0.01 inch (0.25 mm). Small digitizing tablets up to 12 by 24 inches (30 by 60 cm)

(A)

(B)

Figure 9.7 Digitizing equipment: (A) Digitizing table, (B) cursor (Reproduced by permission of GTCO Calcomp, Inc.)

are used for small tasks, but bigger (typically 44 by 60 inches (112 by 152 cm)) freestanding table digitizers are preferred for larger tasks (Figure 9.7). Both types of digitizer usually have cursors with cross hairs mounted in glass and buttons to control capture. Box 9.1 describes the process of table digitizing.

Manual digitizing is still the simplest, easiest, and cheapest method of capturing vector data from existing maps.

Technical Box **9.1**

Manual digitizing

Manual digitizing involves five basic steps.

1. The map document is attached to the center of the digitizing table using sticky tape.
2. Because a digitizing table uses a local rectilinear coordinate system, the map and the digitizer must be registered so that vector data can be captured in real-world coordinates. This is achieved by digitizing a series of four or more well-distributed control points (also called reference points or tick marks) and then entering their real-world values. The digitizer control software (usually the GIS) will calculate a transformation and then automatically apply this to any future coordinates that are captured.
3. Before proceeding with data capture it is useful to spend some time examining a map to determine rules about which features are to be captured at what level of generalization. This type of information is often defined in a data capture project specification.
4. Data capture involves recording the shape of vector objects using manual or stream mode digitizing as described in Section 9.3.2.1. A common rule for vector GIS is to press Button 2 on the digitizing cursor to start a line, Button 1 for each intermediate vertex, and Button 2 to finish a line. There are other similar rules to control how points and polygons are captured.
5. Finally, after all objects have been captured it is necessary to check for any errors. Easy ways to do this include using software to identify geometric errors (such as polygons that do not close or lines that do not intersect – see Figure 9.9), and producing a test plot that can be overlaid on the original document.

Vertices defining point, line, and polygon objects are captured using manual or stream digitizing methods. Manual digitizing involves placing the center point of the cursor cross hairs at the location for each object vertex and then clicking a button on the cursor to record the location of the vertex. Stream-mode digitizing partially automates this process by instructing the digitizer control to collect vertices automatically every time a distance or time threshold is crossed (e.g., every 0.02 inch (0.5 mm) or 0.25 second). Stream-mode digitizing is a much faster method, but it typically produces larger files with many redundant coordinates.

9.3.2.2 Heads-up digitizing and vectorization

One of the main reasons for scanning maps (see Section 9.3.1) is as a prelude to vectorization – the process of converting raster data into vector data. The simplest way to create vectors from raster layers is to digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor. This method is called heads-up digitizing because the map is vertical and can be viewed without bending the head down. It is widely used for the selective capture of, for example, land parcels, buildings, and utility assets.

Vectorization is the process of converting raster data into vector data. The reverse is called rasterization.

A faster and more consistent approach is to use software to perform automated vectorization in either batch or semi-interactive mode. Batch vectorization takes an entire raster file and converts it to vector objects in a single operation. Vector objects are created using software algorithms that build simple (spaghetti) line strings from the original pixel values. The lines can then be further processed to create topologically correct polygons (Figure 9.8). A typical map will take only a few minutes to vectorize using modern hardware and software systems. See Section 10.7.1 for further discussion on structuring geographic data.

Unfortunately, batch vectorization software is far from perfect and post-vectorization editing is usually required to clean up errors. To avoid large amounts of vector editing, it is useful to undertake a little raster editing of the original raster file prior to vectorization to remove unwanted noise that may affect the vectorization process. For example, text that overlaps lines should be deleted and dashed lines are best converted into solid lines. Following vectorization, topological relationships are usually created for the vector objects. This process may also highlight some previously unnoticed errors that require additional editing.

Batch vectorization is best suited to simple bi-level maps of, for example, contours, streams, and highways. For more complicated maps and where selective vectorization is required (for example, digitizing electric conductors and devices, or water mains and fittings off topographic maps), interactive vectorization (also called semi-automatic vectorization, line following, or tracing) is preferred. In interactive vectorization, software is used to automate digitizing. The operator snaps the cursor to a pixel, indicates a direction for line following, and

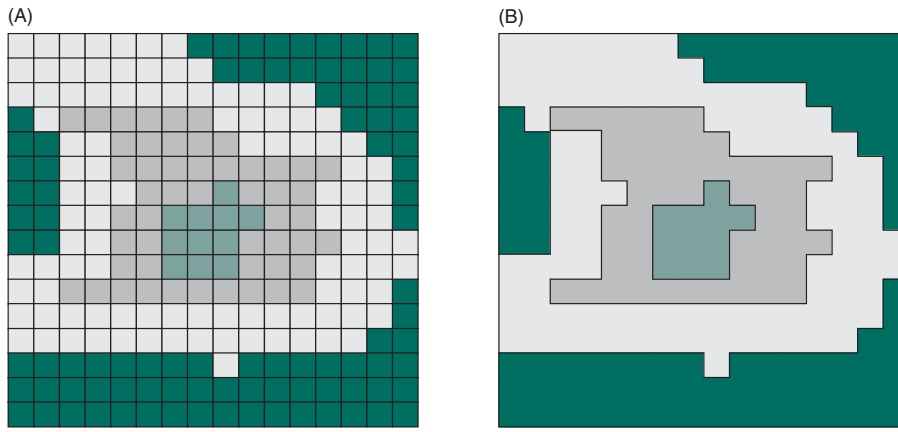


Figure 9.8 Batch vectorization of a scanned map: (A) original raster file; (B) vectorized polygons. Adjacent raster cells with the same attribute values are aggregated. Class boundaries are then created at the intersection between adjacent classes in the form of vector lines

the software then automatically digitizes lines. Typically, many parameters can be tuned to control the density of points (level of generalization), the size of gaps (blank pixels in a line) that will be jumped, and whether to pause at junctions for operator intervention or always to trace in a specific direction (most systems require that all polygons are ordered either clockwise or counterclockwise). Interactive vectorization is still quite labor intensive, but generally it results in much greater productivity than manual or heads-up digitizing. It also produces high-quality data, as software is able to represent lines more accurately and consistently than can humans. It is for these reasons that specialized data capture groups much prefer vectorization to manual digitizing.

9.3.2.3 Measurement error

Data capture, like all geographic workflows, is likely to generate errors. Because digitizing is a tedious and hence error-prone practice, it presents a source of measurement errors – as when the operator fails to position the cursor correctly, or fails to record line segments. Figure 9.9 presents some examples of human errors that are commonly introduced in the digitizing procedure. They are: overshoots and undershoots where line intersections are inexact (Figure 9.9A); invalid polygons which are topologically inconsistent because of omission of one or more lines, or omission of tag data (Figure 9.9B); and sliver polygons, in which multiple digitizing of the common boundary between adjacent polygons leads to the creation of additional polygons (Figure 9.9C).

Most GIS packages include standard software functions, which can be used to restore integrity and clean (or rather obscure, depending upon your viewpoint!) obvious measurement errors. Such operations are best carried out immediately after digitizing, in order that omissions may be easily rectified. Data cleaning operations require sensitive setting of threshold values, or else damage can be done to real-world features, as Figure 9.10 shows.

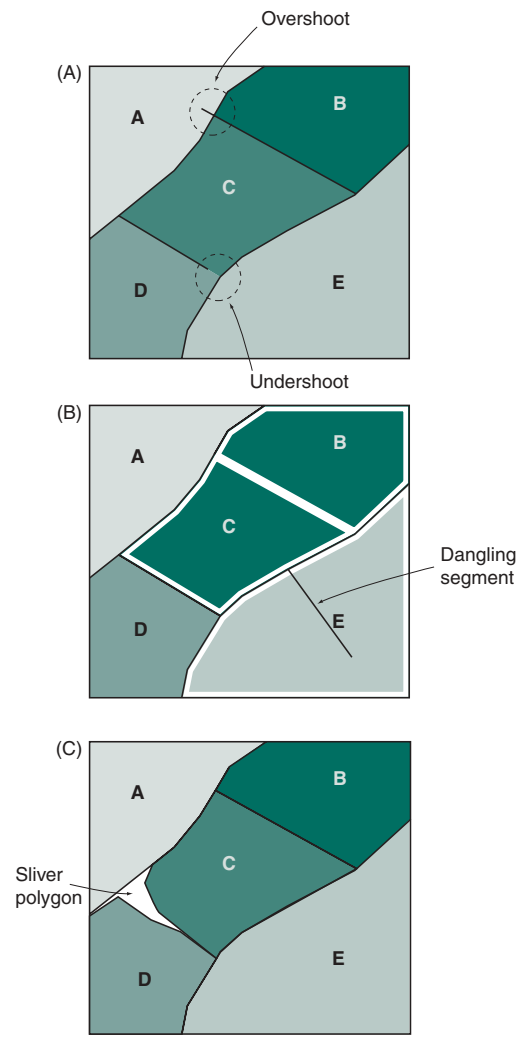


Figure 9.9 Examples of human errors in digitizing: (A) undershoots and overshoots; (B) invalid polygons; and (C) sliver polygons

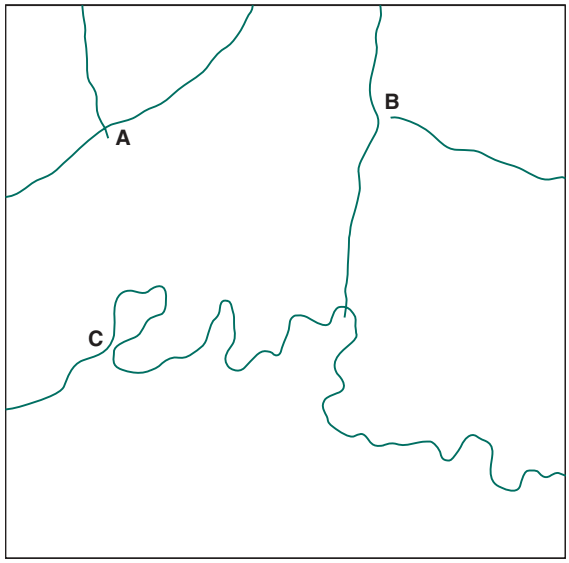


Figure 9.10 Error induced by data cleaning. If the tolerance level is set large enough to correct the errors at A and B, the loop at C will also (incorrectly) be closed

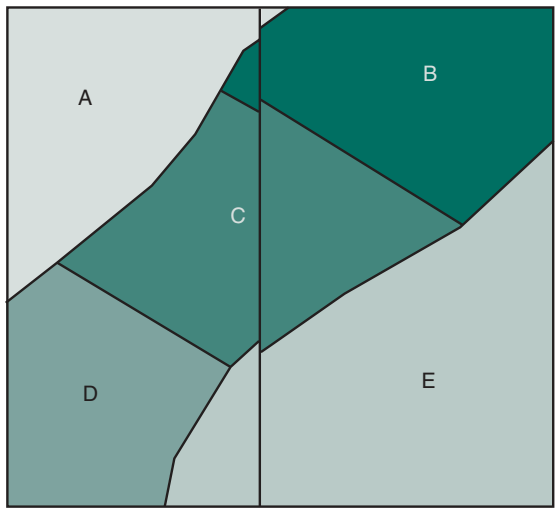


Figure 9.11 Mismatches of adjacent spatial data sources that require rubber-sheeting

Many errors in digitizing can be remedied by appropriately designed software.

Further classes of problems arise when the products of digitizing adjacent map sheets are merged together. Stretching of paper base maps, coupled with errors in rectifying them on a digitizing table, give rise to the kinds of mismatches shown in Figure 9.11. Rubber-sheeting is the term used to describe methods for removing such errors on the assumption that strong spatial autocorrelation exists among errors. If errors tend to be spatially autocorrelated up to a distance of x, say, then rubber-sheeting will be successful at removing them, at least partially, provided control points can be found that are spaced less than x apart. For the same reason, the shapes of features that are less than x across will tend to have little distortion, while very large shapes may be badly distorted. The results of calculating areas (Section 14.3), or other geometric operations that rely only on relative position, will be accurate as long as the areas are small, but will grow rapidly with feature size. Thus it is important for the user of a GIS to know which operations depend on relative position, and over what distance; and where absolute position is important (of course, the term *absolute* simply means relative to the Earth frame, defined by the Equator and the Greenwich Meridian, or relative over a very long distance: see Section 5.6). Analogous procedures and problems characterize the rectification of raster datasets – be they scanned images of paper maps or satellite measurements of the curved Earth surface.

9.3.2.4 Photogrammetry

Photogrammetry is the science and technology of making measurements from pictures, aerial photographs, and images. Although in the strict sense it includes 2-D measurements taken from single aerial photographs, today in GIS it is almost exclusively concerned with capturing 2.5-D and 3-D measurements from models derived from stereo-pairs of photographs and images. In the case of aerial photographs, it is usual to have 60% overlap along each flight line and 30% overlap between flight lines. Similar layouts are used by remote sensing satellites. The amount of overlap defines the area for which a 3-D model can be created.

Photogrammetry is used to capture measurements from photographs and other image sources.

To obtain true georeferenced Earth coordinates from a model, it is necessary to georeference photographs using control points (the procedure is essentially analogous to that described for manual digitizing in Box 9.1). Control points can be defined by ground survey or nowadays more usually with GPS (see Section 9.2.2.1 for discussion of these techniques).

Measurements are captured from overlapping pairs of photographs using stereoplotters. These build a model and allow 3-D measurements to be captured, edited, stored, and plotted. Stereoplotters have undergone three major generations of development: analog (optical), analytic, and digital. Mechanical analog devices are seldom used today, whereas analytical (combined mechanical and digital) and digital (entirely computer-based) are much more common. It is likely that digital (soft-copy) photogrammetry will eventually replace mechanical devices entirely.

There are many ways to view stereo models, including a split screen with a simple stereoscope, and the use of special glasses to observe a red/green display or polarized light. To manipulate 3-D cursors in the x, y, and z planes, photogrammetry systems offer free-moving hand controllers, hand wheels and foot disks, and 3-D mice. The options for extracting vector objects from 3-D models are directly analogous to those available for manual digitizing as described above: namely batch, interactive, and manual (Sections 9.3.2.1 and 9.3.2.2). The obvious

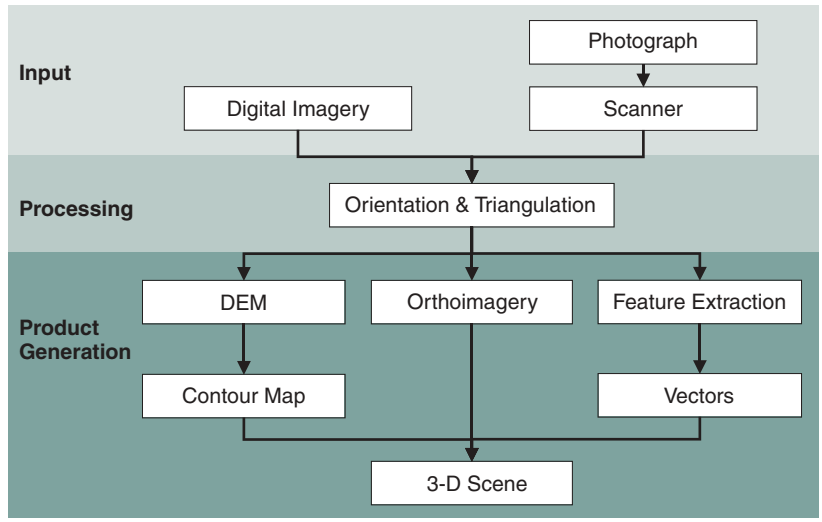


Figure 9.12 Typical photogrammetry workflow (after Tao C.V. 2002 'Digital photogrammetry: the future of spatial data collection', *GeoWorld*. **www.geoplace.com/gw/2002/0205/0205dp.asp**). (Reproduced by permission of GeoTec Media)

difference, however, is that there is a requirement for capturing z (elevation) values.

9.3.2.4.1 Digital photogrammetry workflow

Figure 9.12 shows a typical workflow in digital photogrammetry. There are three main parts to digital photogrammetry workflows: data input, processing, and product generation. Data can be obtained directly from sensors or by scanning secondary sources.

Orientation and triangulation are fundamental photogrammetry processing tasks. Orientation is the process of creating a stereo model suitable for viewing and extracting 3-D vector coordinates that describe geographic objects. Triangulation (also called 'block adjustment') is used to assemble a collection of images into a single model so that accurate and consistent information can be obtained from large areas.

Photogrammetry workflows yield several important product outputs including digital elevation models (DEMs), contours, orthoimages, vector features, and 3-D scenes. DEMs – regular arrays of height values – are created by 'matching' stereo image pairs together using a series of control points. Once a DEM has been created it is relatively straightforward to derive contours using a choice of algorithms. Orthoimages are images corrected for variations in terrain using a DEM. They have become popular because of their relatively low cost of creation (when compared with topographic maps) and ease of interpretation as base maps. They can also be used as accurate data sources for heads-up digitizing (see Section 9.3.2.2). Vector feature extraction is still an evolving field and there are no widely applicable fully automated methods. The most successful methods use a combination of spectral analysis and spatial rules that define context, shape, proximity, etc. Finally, 3-D scenes can be created by merging vector features with a DEM and an orthoimage (Figure 9.13).

In summary, photogrammetry is a very cost-effective data capture technique that is sometimes the only practical method of obtaining detailed topographic data about an area of interest. Unfortunately, the complexity and high cost of equipment have restricted its use to large-scale primary data capture projects and specialist data capture organizations.

9.3.2.5 COGO data entry

COGO, a contraction of the term *coordinate geometry*, is a methodology for capturing and representing geographic data. COGO uses survey-style bearings and distances to define each part of an object in much the same way as described in Section 9.2.2.1. Some examples of COGO object construction tools are shown in Figure 9.14. The Construct Along tool creates a point along a curve using a distance along the curve. The Line Construct Angle Bisector tool constructs a line that bisects an angle defined by a from-point, through-point, to-point, and a length. The Construct Fillet tool creates a circular arc tangent from two segments and a radius.

The COGO system is widely used in North America to represent land records and property parcels (also called lots). Coordinates can be obtained from COGO measurements by geometric transformation (i.e., bearings and distances are converted into x, y coordinates). Although COGO data obtained as part of a primary data capture activity are used in some projects, it is more often the case that secondary measurements are captured from hardcopy maps and documents. Source data may be in the form of legal descriptions, records of survey, tract (housing estate) maps, or similar documents.

COGO stands for coordinate geometry. It is a vector data structure and method of data entry.

COGO data are very precise measurements and are often regarded as the only legally acceptable definition of land parcels. Measurements are usually very detailed and data capture is often time consuming. Furthermore, commonly occurring discrepancies in the data must be manually resolved by highly qualified individuals.

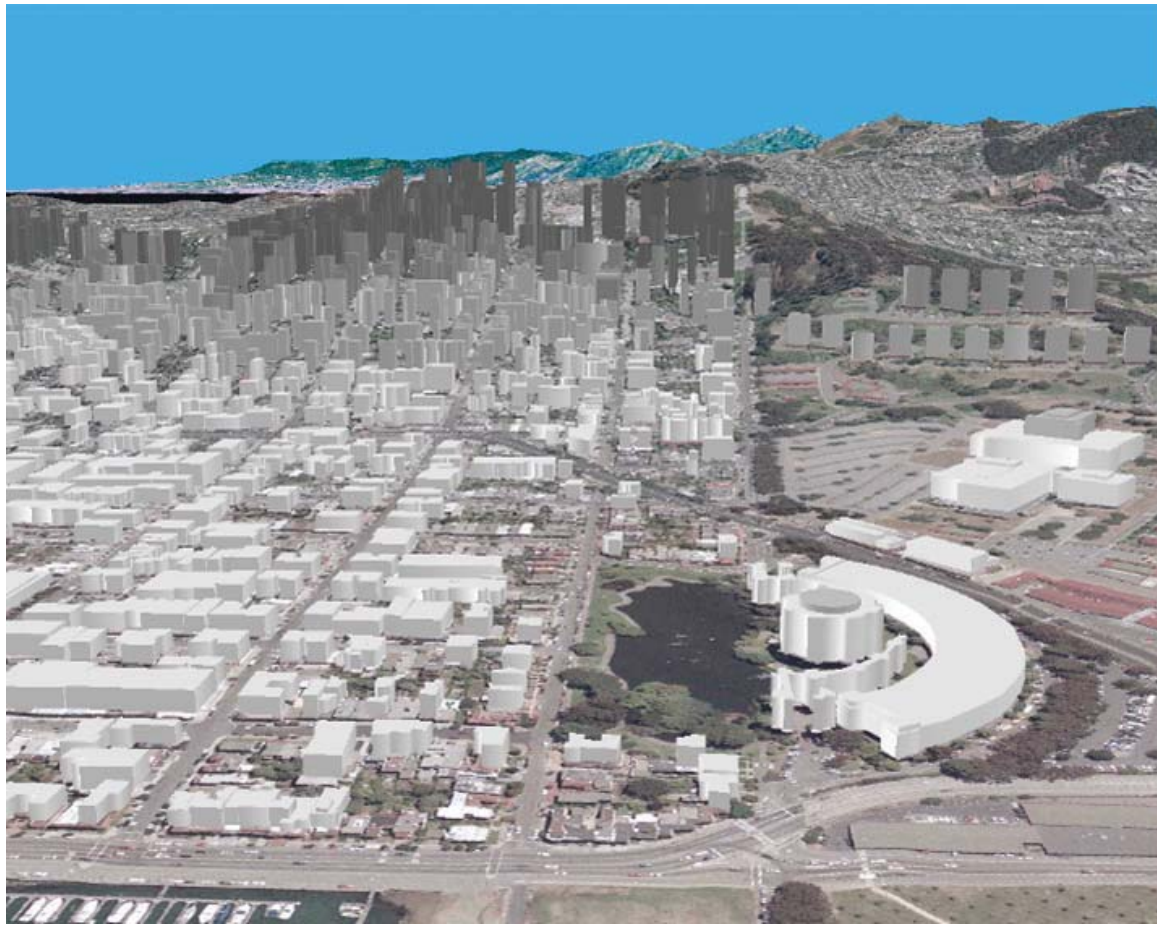


Figure 9.13 Example 3-D scene as generated from a photogrammetry workflow

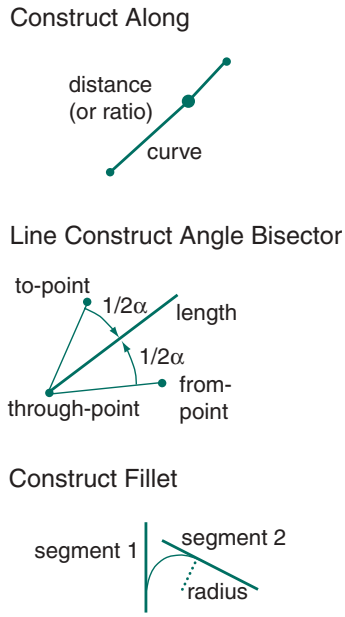


Figure 9.14 Example COGO construction tools used to represent geographic features

9.4 Obtaining data from external sources (data transfer)

One major decision that needs to be faced at the start of a GIS project is whether to build or buy part or all of a database. All the preceding discussion has been concerned with techniques for building databases from primary and secondary sources. This section focuses on how to import or transfer data into a GIS that has been captured by others. Some datasets are freely available, but many of them are sold as a commodity from a variety of outlets including, increasingly, Internet sites.

There are many sources and types of geographic data. Space does not permit a comprehensive review of all geographic data sources here, but a small selection of key sources is listed in Table 9.3. In any case, the characteristics and availability of datasets are constantly changing so those seeking an up-to-date list should consult one of the good online sources described below. Section 18.4.3 also discusses the characteristics of geographic information and highlights several issues to bear in mind when using data collected by others.

Table 9.3 Examples of some digital data sources that can be imported into a GIS. NMOs = National Mapping Organizations, USGS = United States Geologic Survey, NGA = US National Geospatial-Intelligence Agency, NASA = National Aeronautics and Space Administration, DEM = Digital Elevation Model, EPS = US Environmental Protection Agency, WWF = World Wildlife Fund for Nature, FEMA = Federal Emergency Management Agency, EBIS = ESRI Business Information Solutions

Type	Source	Details
	Basemaps	
Geodetic framework	Many NMOs, e.g., USGS and Ordnance Survey	Definition of framework, map projections, and geodetic transformations
General topographic map data	NMOs and military agencies, e.g., NGA	Many types of data at detailed to medium scales
Elevation	NMOs, military agencies, and several commercial providers, e.g., USGS, SPOT Image, NASA	DEMs, contours at local, regional, and global levels
Transportation	National governments, and several commercial vendors, e.g., TeleAtlas and NAVTEQ	Highway/street centerline databases at national levels
Hydrology	NMOs and government agencies	National hydrological databases are available for many countries
Toponymy	NMOs, other government agencies and commercial providers	Gazetteers of placenames at global and national levels
Satellite images	Commercial and military providers, e.g., Landsat, SPOT, IRS, IKONOS, Quickbird	See Figure 9.2 for further details
Aerial photographs	Many private and public agencies	Scales vary widely, typically from 1:500–1:20 000
	Environmental	
Wetlands	National agencies, e.g., US National Wetlands Inventory	Government wetlands inventory
Toxic release sites	National Environmental Protection Agencies, e.g., EPA	Details of thousands of toxic sites
World eco-regions	World Wildlife Fund for Nature (WWF)	Habitat types, threatened areas, biological distinctiveness
Flood zones	Many national and regional government agencies, e.g., FEMA	National flood risk areas
	Socio-economic	
Population census	National governments, with value added by commercial providers	Typically every 10 years with annual estimates
Lifestyle classifications	Private agencies (e.g., CACI and Experian)	Derived from population censuses and other socio-economic data
Geodemographics	Private agencies (e.g., Claritas and EBIS)	Many types of data at many scales and prices
Land and property ownership	National governments	Street, property, and cadastral data
Administrative areas	National governments	Obtained from maps at scales of 1:5000–1:750 000

The best way to find geographic data is to search the Internet. Several types of resources and technologies are available to assist searching, and are described in detail in Section 11.2. These include specialist geographic data catalogs and stores, as well as the sites of specific geographic data vendors (some websites are shown in Table 9.4 and the history of one vendor is described in Box 9.2). Particularly good sites are the Data Store (**www.datastore.co.uk/**) and the AGI (Association for Geographic Information) Resource List (**www.geo.ed.ac.uk/home/giswww.html**). These sites provide access to information about the characteristics and availability of geographic data. Some also have facilities to purchase and download data directly. Probably the most useful resources for locating geographic data are the geolibraries and geoportals (see Section 11.2) that have been created as part of national and global spatial data infrastructure initiatives (SDI).

The best way to find geographic data is to search the Internet using one of the specialist geolibraries or SDI geographic data geoportals.

9.4.1 Geographic data formats

One of the biggest problems with data obtained from external sources is that they can be encoded in many different formats. There are so many different geographic data formats because no single format is appropriate for all tasks and applications. It is not possible to design a format that supports, for example, both fast rendering in police

Biographical Box **9.2**

Don Cooke, geographic data provider

Figure 9.15 Don Cooke, geographic data provider

Don Cooke (Figure 9.15) took a part-time job with the New Haven Census Use Study while finishing his senior year at Yale in 1967. Cooke's three years of Army artillery survey plus an introductory Fortran class gave him GIS credentials typical of most people in the field at the time. Cooke and Bill Maxfield were charged with making computer maps of census and local data. It quickly became apparent that computerized base maps linking census geometry, street addressing, and coordinates were a prerequisite to computer mapping. DIME (Dual Independent Map Encoding) was their solution, probably the first implementation of a topological data structure with redundant encoding for error correction. Cooke, Maxfield, and Jack Sweeney founded Urban Data Processing, Inc. (UDP) in 1968 to bring geocoding, computer mapping, and demographic analysis to the private sector. The Census Bureau adopted DIME which evolved into the nationwide TIGER database during the 1980s.

When Harte-Hanks bought UDP in 1980, Cooke founded Geographic Data Technology (GDT) to commercialize Census DIME and later TIGER files. By the late 1990s, GDT had grown to 500 employees and in 2004 was acquired by TeleAtlas. Cooke remains in his role as Founder, and the TeleAtlas North America operation (effectively a combination of ETAK and GDT) faces NAVTEQ as a competitor in GIS and Navigation markets.

Cooke served on the National Academy of Sciences Mapping Science Committee for four years, and on the Board of the Urban and Regional Information Systems Association (URISA), where he founded the first Special Interest Group (SIG) focusing on GIS. He is an active proponent of GIS and GPS technology in education at all levels. His leadership in this area helped GDT win 'School-to-Careers Company of the Year' recognition from the National Alliance of Business. He is the author of 'Fun with GPS', a GIS primer written for owners of consumer GPS receivers.

On the subject of the current state of GIS Don says: 'Suddenly it seems really easy to explain what GIS is. Most people have some contact or context; they've used MapQuest or know someone who has a GPS. People with GPS think nothing of finding mapping services through Google; they overlay their GPS tracks and points on USGS Digital Raster Graphics and Digital Ortho Quarter Quads without even knowing those terms or messing with GeoSpatial One-Stop (see Box 11.4).'

Thinking about the future, he muses: 'I like to picture a near-term future where every high-school graduate has collected GPS data for a project and mapped it with GIS; we're already there in some schools. The best thing about this is more often than not their mapping has been for a community project and they've seen through the experience how they can participate in and contribute to their community.'

command and control systems, and sophisticated topological analysis in natural resource information systems: the two are mutually incompatible. Also, given the great diversity of geographic information a single comprehensive format would simply be too large and cumbersome. The many different formats that are in use today have evolved in response to diverse user requirements.

Given the high cost of creating databases, many tools have been developed to move data between systems and to reuse data through open application programming interfaces (APIs). In the former case, the approach has been to develop software that is able to translate data (Figure 9.16), either by a direct read into memory, or via an intermediate file format. In the latter case, software developers have created open interfaces to allow access to data.

Many GIS software systems are now able to read directly AutoCAD DWG and DXF, Microstation DGN, and Shapefile, VPF, and many image formats. Unfortunately, direct read support can only easily be provided for relatively simple product-oriented formats. Complex formats, such as SDTS, were designed for exchange purposes and require more advanced processing before they can be viewed (e.g., multi-pass read and feature assembly from several parts).

Data can be transferred between systems by direct read into memory or via an intermediate file format.

More than 25 organizations are involved in the standardization of various aspects of geographic data and geoprocessing; several of them are country and domain

Table 9.4 Selected websites containing information about geographic data sources

Source	URL	Description
AGI GIS Resource List	**www.geo.ed.ac.uk/home/giswww.html**	Indexed list of several hundred sites
The Data Store	**www.data-store.co.uk/**	UK, European, and worldwide data catalog
Geospatial One-Stop	**www.geodata.gov**	Geoportal providing metadata and direct access to over 50 000 datasets
MapMart	**www.mapmart.com/**	Extensive data and imagery provider
EROS Data Center	**edc.usgs.gov/**	US government data archive
Terraserver	**www.terraserver-usa.com/**	High-resolution aerial imagery and topo maps
Geography Network	**www.GeographyNetwork.com**	Global online data and map services
National Geographic Society	**www.nationalgeographic.com**	Worldwide maps
GeoConnections	**www.connect.gc.ca/en/692-e.asp**	Canadian government's geographic data over the Web
EuroGeographics	**www.eurogeographics.org/eng/01_about.asp**	Coalition of European NMOs offering topographic map data
GEOWorld Data Directory	**www.geoplace.com**	List of GIS data companies
The Data Depot	**www.gisdatadepot.com**	Extensive collection of mainly free geographic data depot

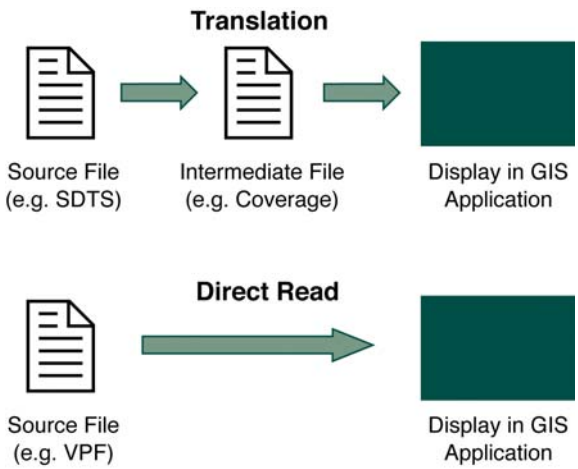


Figure 9.16 Comparison of data access by translation and direct read

specific. At the global level, the ISO (International Standards Organization) is responsible for coordinating efforts through the work of technical committees TC 211 and 287. In Europe, CEN (Comité Européen de Normalisation) is engaged in geographic standardization. At the national level, there are many complementary bodies. One other standards-forming organization of particular note is OGC (Open Geospatial Consortium: **www.opengeospatial.org**), a group of vendors, academics, and users interested in the interoperability of geographic systems (see Box 11.1). To date there have been promising OGC-coordinated efforts to standardize on simple feature access (simple geometric object types), metadata catalogs, and Web access.

The most efficient way to translate data between systems is usually via a common intermediate file format.

Having obtained a potentially useful source of geographic information the next task is to import it into a GIS database. If the data are already in the native format of the target GIS software system, or the software has a direct read capability for the format in question, then this is a relatively straightforward task. If the data are not compatible with the target GIS software then the alternatives are to ask the data supplier to convert the data to a compatible format, or to use a third-party translation software system, such as the Feature Manipulation Engine from Safe Software (**www.safe.com** lists over 60 supported geographic data formats) to convert the data. Geographic data translation software must address both syntactic and semantic translation issues. Syntactic translation involves converting specific digital symbols (letters and numbers) between systems. Semantic translation is concerned with converting the meaning inherent in geographic information. While the former is relatively simple to encode and decode, the latter is much more difficult and has seldom met with much success to date.

Although the task of translating geographic information between systems was described earlier as relatively straightforward, those that have tried this in practice will realize that things on the ground are seldom quite so simple. Any number of things can (and do!) go wrong. These range from corrupted media, to incomplete data files, wrong versions of translators, and different interpretations of a format specification, to basic user error.

There are two basic strategies used for data translation: one is direct and the other uses a neutral intermediate format. For small systems that involve the translation of a small number of formats, the first is the simplest. Directly translating data back and forth between the internal structures of two systems requires two new translators (A to B, B to A). Adding two further systems will require 12 translators to share data between all systems (A to B, A to C, A to D, B to A, B to C, B to D, C to A, C to B, C to D, D to A, D to B, and D to C). A more efficient way of solving this problem is to use the

concept of a data *switchyard* and a common intermediate file format. Systems now need only to translate to and from the common format. The four systems will now need only eight translators instead of 12 (A to Neutral, B to Neutral, C to Neutral, D to Neutral, Neutral to A, Neutral to B, Neutral to C, and Neutral to D). The more systems there are the more efficient this becomes. This is one of the key principles underlying the need for common file interchange formats.

9.5 Capturing attribute data

All geographic objects have attributes of one type or another. Although attributes can be collected at the same time as vector geometry, it is usually more cost-effective to capture attributes separately. In part, this is because attribute data capture is a relatively simple task that can be undertaken by lower-cost clerical staff. It is also because attributes can be entered by direct data loggers, manual keyboard entry, optical character recognition (OCR) or, increasingly, voice recognition, which do not require expensive hardware and software systems. Much the most common method is direct keyboard data entry into a spreadsheet or database. For some projects, a custom data entry form with in-built validation is preferred. On small projects single entry is used, but for larger, more complex projects data are entered twice and then compared as a validation check.

An essential requirement for separate data entry is a common identifier (also called a key) that can be used to relate object geometry and attributes together following data capture (see Figure 10.2 for a diagrammatic explanation of relating geometry and attributes).

Metadata are a special type of non-geometric data that are increasingly being collected. Some metadata are derived automatically by the GIS software system (for example, length and area, extent of data layer, and count of features), but some must be explicitly collected (for example, owner name, quality estimate, and original source). Explicitly collected metadata can be entered in the same way as other attributes, as described above. For further information about metadata see Section 11.2.

9.6 Managing a data collection project

The subject of managing a GIS project is given extensive treatment later in this book in Chapters 17–20. The management of data capture projects is discussed briefly here both because of its critical importance and because there are several unique issues. That said, most of the general principles for any GIS project apply to data collection: the need for a clearly articulated plan, adequate resources, appropriate funding, and sufficient time.

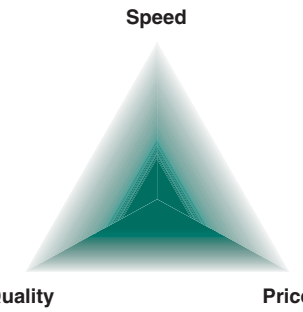


Figure 9.17 Relationship between quality, speed, and price in data collection (*Source*: after Hohl 1997)

In any data collection project there is a fundamental tradeoff between quality, speed, and price. Collecting high-quality data quickly is possible, but it is also very expensive. If price is a key consideration then lower-quality data can be collected over a longer period (Figure 9.17).

GIS data collection projects can be carried out intensively or over a longer period. A key decision facing managers of such projects is whether to pursue a strategy of incremental or very rapid collection. Incremental data collection involves breaking the data collection project into small manageable subprojects. This allows data collection to be undertaken with lower annual resource and funding levels (although total project resource requirements may be larger). It is a good approach for inexperienced organizations that are embarking on their first data collection project because they can learn and adapt as the project proceeds. On the other hand, these longer-term projects run the risk of employee turnover and burnout, as well as changing data, technology, and organizational priorities.

Whichever approach is preferred, a pilot project carried out on part of the study area and a selection of the data types can prove to be invaluable. A pilot project can identify problems in workflow, database design, personnel, and equipment. A pilot database can also be used to test equipment and to develop procedures for quality assurance. Many projects require a test database for hardware and software acceptance tests, as well as to facilitate software customization. It is essential that project managers are prepared to discard all the data obtained during a pilot data collection project, so that the main phase can proceed unconstrained.

A further important decision is whether data collection should use in-house or external resources. It is now increasingly common to outsource geographic data collection to specialist companies that usually undertake the work in areas of the world with very low labor costs (e.g., India and Thailand). Three factors influencing this decision are: cost/schedule, quality, and long-term ramifications. Specialist external data collection agencies can often perform work faster, cheaper, with higher quality than in-house staff, but because of the need for real cash to pay external agencies this may not be possible. In the short term, project costs, quality, and time are the main considerations, but over time dependency on external groups may become a problem.

Questions for further study

1. Using the websites listed in Table 9.4 as a starting point, evaluate the suitability of free geographic data for your home region or country for use in a GIS project of your choice.
2. What are the advantages of batch vectorization over manual table digitizing?
3. What quality assurance steps would you build into a data collection project designed to construct a database of land parcels for tax assessment?
4. Why do so many geographic data formats exist? Which ones are most suitable for selling vector data?

Further reading

Hohl P. (ed) 1997 *GIS Data Conversion: Strategies, Techniques and Management*. Santa Fe, NM: OnWord Press.

Jones C. 1997 *Geographic Information Systems and Computer Cartography*. Reading, MA: Addison-Wesley Longman.

Lillesand T.M., Kiefer R.W. and Chipman R.W. 2003 *Remote Sensing and Image Interpretation* (5th edn). Hoboken, NJ: Wiley.

Paine D.P. and Kiser J.D. 2003 *Aerial Photography and Image Interpretation* (2nd edn). Hoboken, NJ: Wiley.

Walford N. 2002 *Geographical Data: Characteristics and Sources*. Hoboken, NJ: Wiley.

10 *Creating and maintaining geographic databases*

All large operational GIS are built on the foundation of a geographic database. After people, the database is arguably the most important part of a GIS because of the costs of collection and maintenance, and because the database forms the basis of all queries, analysis, and decision making. Today, virtually all large GIS implementations store data in a database management system (DBMS), a specialist piece of software designed to handle multi-user access to an integrated set of data. Extending standard DBMS to store geographic data raises several interesting challenges. Databases need to be designed with great care, and to be structured and indexed to provide efficient query and transaction performance. A comprehensive security and transactional access model is necessary to ensure that multiple users can access the database at the same time. On-going maintenance is also an essential, but very resource-intensive, activity.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will:

- Understand the role of database management systems in GIS;
- Recognize structured query language (SQL) statements;
- Understand the key geographic database data types and functions;
- Be familiar with the stages of geographic database design;
- Understand the key techniques for structuring geographic information, specifically creating topology and indexing;
- Understand the issues associated with multi-user editing and versioning.

10.1 Introduction

A database can be thought of as an integrated set of data on a particular subject. Geographic databases are simply databases containing geographic data for a particular area and subject. It is quite common to encounter the term 'spatial' in the database world. As discussed in Section 1.1.1, 'spatial' refers to data about space at both geographic and non-geographic scales. A geographic database is a critical part of an operational GIS. This is both because of the cost of creation and maintenance, and because of the impact of a geographic database on all analysis, modeling, and decision-making activities. Databases can be physically stored in files or in specialist software programs called database management systems (DBMS). Today, most large organizations use a combination of files and DBMS for storing data assets.

A database is an integrated set of data on a particular subject.

The database approach to storing geographic data offers a number of advantages over traditional file-based datasets:

- Assembling all data at a single location reduces redundancy.
- Maintenance costs decrease because of better organization and reduced data duplication.
- Applications become data independent so that multiple applications can use the same data and can evolve separately over time.
- User knowledge can be transferred between applications more easily because the database remains constant.
- Data sharing is facilitated and a corporate view of data can be provided to all managers and users.
- Security and standards for data and data access can be established and enforced.
- DBMS are better suited to managing large numbers of concurrent users working with vast amounts of data.

On the other hand there are some disadvantages to using databases when compared to files:

- The cost of acquiring and maintaining DBMS software can be quite high.
- A DBMS adds complexity to the problem of managing data, especially in small projects
- Single user performance will often be better for files, especially for more complex data types and structures where specialist indexes and access algorithms can be implemented.

In recent years geographic databases have become increasingly large and complex. For example, AirPhoto USA's US National Image Mosaic is 25 terabytes (TB) in size, EarthSat's global Landsat mosaic at 15 m resolution is 6.5 TB, and Ordnance Survey of Great Britain has approximately 450 million vector features in its MasterMap database covering all of Britain. This chapter describes how to create and maintain geographic databases, and the concepts, tools, and techniques that are available to manage geographic data in databases. Several other chapters provide additional information that is relevant to this discussion. In particular, the nature of geographic data and how to represent them in GIS were described in Chapters 3, 4, and 5, and data modeling and data collection were discussed in Chapters 8 and 9 respectively. Later chapters introduce the tools and techniques that are available to query, model, and analyze geographic databases (Chapters 14, 15, and 16). Finally, Chapters 17 through 20 discuss the important management issues associated with creating and maintaining geographic databases.

10.2 Database management systems

A DBMS is a software application designed to organize the efficient and effective storage and access of data.

Small, simple databases that are used by a small number of people can be stored on computer disk in

standard files. However, larger, more complex databases with many tens, hundreds, or thousands of users require specialist database management system (DBMS) software to ensure database integrity and longevity. A DBMS is a software application designed to organize the efficient and effective storage of and access to data. To carry out this function DBMS provide a number of important capabilities. These are introduced briefly here and are discussed further in this and other chapters. DBMS provide:

- A data model. As discussed in Section 8.4, a data model is the mechanism used to represent real-world objects digitally in a computer system. All DBMS include standard general-purpose core data models suitable for representing several types of object (e.g., integer and floating-point numbers, dates, and text). In most cases DBMS can be extended to support geographic object types.
- A data load capability. DBMS provide tools to load data into databases. Simple tools are available to load standard supported data types (e.g., character, number, and date) in well-structured formats. Other non-standard data formats can be loaded by writing custom software programs that convert the data into a structure that can be read by the standard loaders.
- Indexes. An index is a data structure used to speed up searching. All databases include tools to index standard database data types.
- A query language. One of the major advantages of DBMS is that they support a standard data query/manipulation language called SQL (Structured/Standard Query Language).
- Security. A key characteristic of DBMS is that they provide controlled access to data. This includes restricting user access to all or part of a database. For example, a casual GIS user might have read-only access to just part of a database, but a specialist user might have read and write (create, update, and delete) access to the entire database.
- Controlled update. Updates to databases are controlled through a transaction manager responsible for managing multi-user access and ensuring that updates affecting more than one part of the database are coordinated.
- Backup and recovery. It is important that the valuable data in a database are protected from system failure and incorrect (accidental or deliberate) update. Software utilities are provided to back up all or part of a database and to recover the database in the event of a problem.
- Database administration tools. The task of setting up the structure of a database (the schema), creating and maintaining indexes, tuning to improve performance, backing up and recovering, and allocating user access rights is performed by a database administrator (DBA). A specialized collection of tools and a user interface are provided for this purpose.
- Applications. Modern DBMS are equipped with standard, general-purpose tools for creating, using, and maintaining databases. These include applications for designing databases (CASE tools) and for building user interfaces for data access and presentations (forms and reports).
- Application programming interfaces (APIs). Although most DBMS have good general-purpose applications for standard use, most large, specialist applications will require further customization using a commercial off-the-shelf programming language and a DBMS programmable API.

This list of DBMS capabilities is very attractive to GIS users and so, not surprisingly, virtually all large GIS databases are based on DBMS technology. Indeed, most GIS software vendors include DBMS software within their GIS software products, or provide an interface that supports very close coupling to a DBMS. For further discussion of this see Chapter 8.

Today, virtually all large GIS use DBMS technology for data management.

10.2.1 Types of DBMS

DBMS can be classified according to the way they store and manipulate data. Three main types of DBMS are available to GIS users today: relational (RDBMS), object (ODBMS), and object-relational (ORDBMS).

A relational database comprises a set of tables, each a two-dimensional list (or array) of records containing attributes about the objects under study. This apparently simple structure has proven to be remarkably flexible and useful in a wide range of application areas, such that today over 95% of the data in DBMS are stored in RDBMS.

Object database management systems (ODBMS) were initially designed to address several of the weaknesses of RDBMS. These include the inability to store complete objects directly in the database (both object state and behavior: see Box 8.2 for an introduction to objects and object technology). Because RDBMS were focused primarily on business applications such as banking, human resource management, and stock control and inventory, they were never designed to deal with rich data types, such as geographic objects, sound, and video. A further difficulty is the poor performance of RDBMS for many types of geographic query. These problems are compounded by the difficulty of extending RDBMS to support geographic data types and processing functions, which obviously limits their adoption for geographic applications. ODBMS can store objects persistently (semi-permanently on disk or other media) and provide object-oriented query tools. A number of commercial ODBMS have been developed including GemStone/S Object Server from GemStone Systems Inc., Objectivity/DB from Objectivity Inc., ObjectStore from Progress Software, and Versant from Versant Object Technology Corp.

In spite of the technical elegance of ODBMS, they have not proven to be as commercially successful as some predicted. This is largely because of the massive installed base of RDBMS and the fact that RDBMS vendors have now added many of the important ODBMS

capabilities to their standard RDBMS software systems to create hybrid object-relational DBMS (ORDBMS). An ORDBMS can be thought of as an RDBMS engine with an extensibility framework for handling objects. They can handle both the data describing what an object is (object attributes such as color, size, and age) and the behavior that determines what an object does (object methods or functions such as drawing instructions, query interfaces, and interpolation algorithms) and these can be managed and stored together as an integrated whole. Examples of ORDBMS software include IBM DB2 and Informix Dynamic Server, Microsoft SQL Server, and Oracle. As ORDBMS and the underlying relational model are so important in GIS, these topics are discussed at length in Section 10.3.

The ideal geographic ORDBMS is one that has been extended to support geographic object types and functions through the addition of the following (these topics are introduced here and discussed further later in this chapter):

- A query parser – the engine used to interpret SQL queries is extended to deal with geographic types and functions.
- A query optimizer – the software query optimizer is able to handle geographic queries efficiently. Consider a query to find all potential users of a new brand of premier wine to be marketed to wealthy households from a network of retail stores. The objective is to select all households within 3 km of a store that have an income greater than $110 000. This could be carried out in two ways:
  1. Select all households with an income greater than $110 000; from this selected set, select all households within 3 km of a store.
  2. Select all households within 3 km of a store; from this selected set select all households with an income greater than $110 000.

  Selecting households with an income greater than $110 000 is an attribute query that can be performed very quickly. Selecting households within 3 km of a store is a geometric query that takes much longer. Executing the attribute query first (option 1 above) will result in fewer tests for store proximity and therefore the whole query will be completed much more quickly.
- A query language – the query language is able to handle geographic types (e.g., points and polygons) and functions (e.g., select polygons that touch each other).
- Indexing services – the standard unidimensional DBMS data index service is extended to support multidimensional (i.e., x, y, z coordinates) geographic data types.
- Storage management – the large volume of geographic records with different sizes (especially geometric and topological relationships) is accommodated through specialized storage structures.
- Transaction services – standard DBMS are designed to handle short (sub-second) transactions and are extended to deal with the long transactions common in many geographic applications.
- Replication – services for replicating databases are extended to deal with geographic types, and problems of reconciling changes made by distributed users.

10.2.2 Geographic DBMS extensions

Two of the major commercial DBMS vendors have released spatial database extensions to their standard ORDBMS products: IBM offers two solutions – DB2 Spatial Extender and Informix Spatial Datablade – and Oracle has a Spatial option (see Box 10.1).

Although there are differences in the technology, scope, and capabilities of these systems, they all provide basic functions to store, manage, and query geographic objects. This is achieved by implementing the seven key database extensions described in the previous section. It

Technical Box **10.1**

Oracle Spatial

Oracle Spatial is an extension to the Oracle DBMS that provides the foundation for the management of spatial (including geographic) data inside an Oracle database. The standard types and functions in Oracle (CHAR, DATE or INTEGER, etc.) are extended with geographic equivalents. Oracle Spatial supports three basic geometric forms:

- *Points*: points can represent locations such as buildings, fire hydrants, utility poles, oil rigs, boxcars, or roaming vehicles.
- *Lines*: lines can represent things like roads, railroad lines, utility lines, or fault lines.
- *Polygons and complex polygons with holes*: polygons can represent things like outlines of cities, districts, flood plains, or oil and gas fields. A polygon with a hole might, for example, geographically represent a parcel of land surrounding a patch of wetlands.

These simple feature types can be aggregated to richer types using topology and linear referencing capabilities. Additionally, Oracle Spatial can store and manage georaster (image) data.

Oracle Spatial extends the Oracle DBMS query engine to support geographic queries. There is

a set of spatial operators to perform: area-of-interest and spatial-join queries; length, area, and distance calculations; buffer and union queries; and administrative tasks. The Oracle Spatial SQL used to create a table and populate it with a single record is shown below. The characters after -- on each line are comments to describe the operations. The discussion of SQL syntax in Section 10.4 will help decode this program.

```
-- Create a table for routes (highways).
CREATE TABLE lrs_routes (
 route_id  NUMBER PRIMARY KEY,
 route_name  VARCHAR2(32),
 route_geometry  MDSYS.SDO_GEOMETRY);
 -- Populate table with just one route
    for this example.
INSERT INTO lrs_routes VALUES(
 1,
 'Route1',
MDSYS.SDO_GEOMETRY(
 3002,   -- line string, 3 dimensions:
    X,Y,M
 NULL,
 NULL,
 MDSYS.SDO_ELEM_INFO_ARRAY(1,2,1), -- one
    line string, straight segments
 MDSYS.SDO_ORDINATE_ARRAY(
  2,2,0,   -- Starting point - Exit1; 0
     is measure from start.
  2,4,2,   -- Exit2; 2 is measure from
     start.
  8,4,8,   -- Exit3; 8 is measure from
     start.
  12,4,12,  -- Exit4; 12 is measure from
     start.
  12,10,NULL, -- Not an exit; measure
     will be automatically calculated &
     filled.
  8,10,22,  -- Exit5; 22 is measure from
     start.
  5,14,27)  -- Ending point (Exit6); 27
     is measure from start.
  )
);
```

Geographic data in Oracle Spatial can be indexed using R-tree and quadtree indexing methods (these terms are defined in Section 10.7.2). There are also capabilities for managing projections and coordinate systems, as well as long transactions (see discussion in 10.9.1). Finally, there are also some tools for elementary spatial data mining (Section 15.1).

Oracle Spatial can be used with all major GIS software products and developers can create specific-purpose applications that embed SQL commands for manipulating and querying data.

As a new player in the GIS market space, Oracle has generated interest among larger IT-focused organizations. IBM has approached this market in a similar way with its Spatial Extender for the DB2 DBMS and Spatial Datablade for Informix.

is important to realize, however, that none of these is a complete GIS software system in itself. The focus of these extensions is data storage, retrieval, and management, and they have no real capabilities for geographic editing, mapping, and analysis. Consequently, they must be used in conjunction with a GIS except in the case of the simplest query-focused applications. Figure 10.1 shows how GIS and DBMS software can work together and some of the tasks best carried out by each system.

ORDBMS can support geographic data types and functions.

10.2.3 Geographic middleware extensions

An alternative to extending the DBMS software kernel to manage geographic data is to build support for spatial data types and functions into a middle-tier (or middleware) application server. Section 7.3.2 provides a description of this system architecture concept. This type of configuration offers many of the same capabilities as the core DBMS extensions, but can also support a wider range of data types and processing functions. A

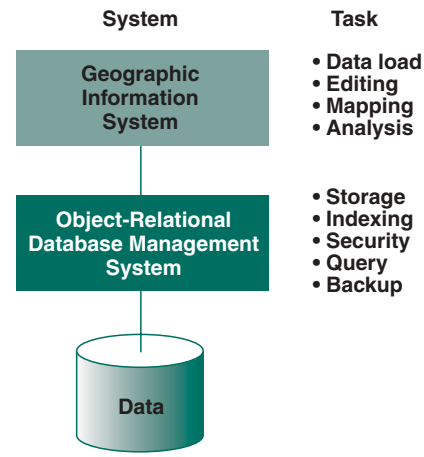


Figure 10.1 The roles of GIS and DBMS

middleware solution can also deliver better performance especially in the case of the more complex queries used in high-end GIS applications, because both the DBMS and the application server hardware resources can be used in parallel. Geographic middleware is available from ESRI in the form of ArcSDE, Intergraph in the form of

GeoMedia Transaction Server and MapInfo in the form of SpatialWare.

10.3 Storing data in DBMS tables

The lowest level of user interaction with a geographic database is usually the object class (also called a layer or feature class), which is an organized collection of data on a particular theme (e.g., all pipes in a water network, all soil polygons in a river basin, or all elevation values in a terrain surface). Object classes are stored in standard database tables. A table is a two-dimensional array of rows and columns. Each object class is stored as a single database table in a database management system (DBMS). Table rows contain objects (instances of object classes, e.g., data for a single pipe) and the columns contain object properties or attributes as they are frequently called (Figure 10.2). The data stored at individual row, column intersections are usually referred to as values. Geographic database tables are distinguished from non-geographic tables by the presence of a geometry column (often called the shape column). To save space and improve performance, the actual coordinate values may be stored in a highly compressed binary form.

Relational databases are made up of tables. Each geographic class (layer) is stored as a table.

Tables are joined together using common row/column values or keys as they are known in the database world. Figure 10.2 shows parts of tables containing data about US states. The STATES table (Figure 10.2A) contains the geometry and some basic attributes, an important one being a unique STATE_FIPS (State FIPS [Federal Information Processing Standard] code) identifier. The POPULATION table (Figure 10.2B) was created entirely independently, but also has a unique identifier column called STATE_FIPS. Using standard database tools the two tables can be joined together based on the common STATE_FIPS identifier column (the key) to create a third table, COMBINED STATES and POPULATION (Figure 10.2C). Following the join these can be treated as a single table for all GIS operations such as query, display, and analysis.

Database tables can be joined together to create new views of the database.

In a groundbreaking description of the relational model that underlies the vast majority of the world's geographic databases, In 1970 Ted Codd of IBM defined a series of rules for the efficient and effective design of database table structures. The heart of Codd's idea was that the best relational databases are made up of simple, stable tables that follow five principles:

1. Only one value is in each cell at the intersection of a row and column.
2. All values in a column are about the same subject.
3. Each row is unique (there are no duplicate records).

(A)

Attributes of STATES

FID	Shape*	AREA	STATE_NAME	STATE_FIPS
41	Polygon	51715.656	Alabama	01
49	Polygon	576556.687	Alaska	02
35	Polygon	113711.523	Arizona	04
45	Polygon	52912.797	Arkansas	05
23	Polygon	157774.187	California	06
30	Polygon	104099.109	Colorado	08
17	Polygon	4976.434	Connecticut	09
27	Polygon	2054.506	Delaware	10
26	Polygon	66.063	District of Columbia	11
47	Polygon	55815.051	Florida	12
43	Polygon	58629.195	Georgia	13
48	Polygon	6381.435	Hawaii	15
7	Polygon	83340.594	Idaho	16
25	Polygon	56297.953	Illinois	17
20	Polygon	36399.516	Indiana	18
12	Polygon	56257.219	Iowa	19
32	Polygon	82195.437	Kansas	20
31	Polygon	40318.777	Kentucky	21
46	Polygon	45835.898	Louisiana	22
2	Polygon	32161.664	Maine	23
29	Polygon	9739.753	Maryland	24
13	Polygon	8172.482	Massachusetts	25
50	Polygon	57898.367	Michigan	26
9	Polygon	84517.469	Minnesota	27
42	Polygon	47618.723	Mississippi	28
34	Polygon	69831.625	Missouri	29
1	Polygon	147236.031	Montana	30

Record: 0 Show: All Selected Records (0 out of 51 Selected.)

(B)

Attributes of POPULATION

STATE_FIPS	SUB_REGION	STATE_ABBR	POP1990	POP1996
53	Pacific	WA	4866692	5629613
30	Mtn	MT	799065	885762
23	N Eng	ME	1227928	1254465
38	W N Cen	ND	638800	633534
46	W N Cen	SD	696004	721374
56	Mtn	WY	453588	487142
55	E N Cen	WI	4891769	5144123
16	Mtn	ID	1006749	1201327
50	N Eng	VT	562758	587726
27	W N Cen	MN	4375099	4639933
41	Pacific	OR	2842321	3203820
33	N Eng	NH	1109252	1156932
19	W N Cen	IA	2776755	2831890
25	N Eng	MA	6016425	6066573
31	W N Cen	NE	1578385	1622272
36	Mid Atl	NY	17990455	18293435
42	Mid Atl	PA	11881643	12077607
09	N Eng	CT	3287116	3287604
44	N Eng	RI	1003464	993306
34	Mid Atl	NJ	7730188	7956917
18	E N Cen	IN	5544159	5801023
32	Mtn	NV	1201833	1532295
49	Mtn	UT	1722850	2000630
06	Pacific	CA	29760021	32218713
39	E N Cen	OH	10847115	11123416
17	E N Cen	IL	11430602	11731783
11	S Atl	DC	606900	550076

Record: 0 Show: All Selected Records (0 out of 51 Selec

Figure 10.2 GIS database tables for US States: (A) STATES table; (B) POPULATION table; (C) joined table – COMBINED STATES and POPULATION

(C)

Attributes of COMBINED STATES and POPULATION

FID	Shape*	AREA	STATE_NAME	STATE_FIPS	SUB_REGION	STATE_ABBR	POP1990	POP1996
0	Polygon	67286.875	Washington	53	Pacific	WA	4866692	5629613
1	Polygon	147236.031	Montana	30	Mtn	MT	799065	885762
2	Polygon	32161.664	Maine	23	N Eng	ME	1227928	1254465
3	Polygon	70810.156	North Dakota	38	W N Cen	ND	638800	633534
4	Polygon	77193.625	South Dakota	46	W N Cen	SD	696004	721374
5	Polygon	97799.492	Wyoming	56	Mtn	WY	453588	487142
6	Polygon	56088.066	Wisconsin	55	E N Cen	WI	4891769	5144123
7	Polygon	83340.594	Idaho	16	Mtn	ID	1006749	1201327
8	Polygon	9603.218	Vermont	50	N Eng	VT	562758	587726
9	Polygon	84517.469	Minnesota	27	W N Cen	MN	4375099	4639933
10	Polygon	97070.750	Oregon	41	Pacific	OR	2842321	3203820
11	Polygon	9259.514	New Hampshire	33	N Eng	NH	1109252	1156932
12	Polygon	56257.219	Iowa	19	W N Cen	IA	2776755	2831890
13	Polygon	8172.482	Massachusetts	25	N Eng	MA	6016425	6066573
14	Polygon	77328.336	Nebraska	31	W N Cen	NE	1578385	1622272
15	Polygon	48560.578	New York	36	Mid Atl	NY	17990455	18293435
16	Polygon	45359.238	Pennsylvania	42	Mid Atl	PA	11881643	12077607
17	Polygon	4976.434	Connecticut	09	N Eng	CT	3287116	3287604
18	Polygon	1044.850	Rhode Island	44	N Eng	RI	1003464	993306
19	Polygon	7507.302	New Jersey	34	Mid Atl	NJ	7730188	7956917
20	Polygon	36399.516	Indiana	18	E N Cen	IN	5544159	5801023
21	Polygon	110667.297	Nevada	32	Mtn	NV	1201833	1532295
22	Polygon	84870.187	Utah	49	Mtn	UT	1722850	2000630
23	Polygon	157774.187	California	06	Pacific	CA	29760021	32218713
24	Polygon	41192.863	Ohio	39	E N Cen	OH	10847115	11123416
25	Polygon	56297.953	Illinois	17	E N Cen	IL	11430602	11731783
26	Polygon	66.063	District of Columbia	11	S Atl	DC	606900	550076
27	Polygon	2054.506	Delaware	10	S Atl	DE	666168	724890

Record: 0 Show: All Selected Records (0 out of 51 Selected.) Options

Figure 10.2 (*continued*)

4. There is no significance to the sequence of columns.
5. There is no significance to the sequence of rows.

Figure 10.3A shows a land parcel tax assessment database table that contradicts several of Codd's principles. Codd suggests a series of transformations called normal forms that successively improve the simplicity and stability, and reduce redundancy of database tables (thus reducing the risk of editing conflicts) by splitting them into sub-tables that are re-joined at query time. Unfortunately, joining large tables is computationally expensive and can result in complex database designs that are difficult to maintain. For this reason, non-normalized table designs are often used in GIS.

Figure 10.3B is a cleansed version of 10.3A that has been entered into a GIS DBMS: there is now only one value in each cell (Date and AssessedValue are now separate columns); missing values have been added; an OBJECTID (unique system identifier) column has been added; and the potential confusion between Dave Widseler and D Widseler has been resolved. Figure 10.3C shows the same data after some normalization to make it suitable for use in a GIS tax assessment application. The database now consists of three tables that can be joined together using common keys. Figure 10.3C `Attributes of Tab10_3b` can be joined to Figure 10.3C `Attributes of Tab10_3a`

(A)

ParcelNumb	OwnerNam	OwnerAddress	PostalCode	ZoningCode	ZoningType	Date / AssessedValue
673/100	Jeff Peters	10 Railway Cuttings	114390	2	Residential	2002 220000
673-101	Joel Campbell	1115 Center Place	114390	2	Residential	2003 545500
674-100	Dave Widseler		114391	3	Commercial	99 249000
674-100		452 Diamond Plaza	114391	3	Commercial	2000 275500
674-100	D Widseler	452 Diamond Plaza	114391	3	Commercial	2001 290000
670-231	Sam Camarata	19 Big Bend Bld	114391	2	Residential	2004 450575
674-112	Chris Capelli	Hastings Barracks	114392	2	Residential	2004 350000
674-113	Sheila Sullivan	10034 Endin Mansions	114391	2	Residential	02 1005425

Figure 10.3 Tax assessment database: (A) raw data; (B) cleansed data in a GIS DBMS; (C) data partially normalized into three sub-tables; (D) joined table

(B)

Attributes of Tab10_2

OBJECTID*	ParcelNumb	OwnerNam	OwnerAddress	PostalCode	ZoningCode	ZoningType	DateAssessed	AssessedValue
1	673-100	Jeff Peters	10 Railway Cuttings	114390	2	Residential	2002	220000
2	673-101	Joel Campbell	1115 Center Place	114390	2	Residential	2003	545500
3	674-100	Dave Widseler	452 Diamond Plaza	114391	3	Commercial	1999	249000
4	674-100	Dave Widseler	452 Diamond Plaza	114391	3	Commercial	2000	275500
5	674-100	Dave Widseler	452 Diamond Plaza	114391	3	Commercial	2001	290000
6	670-231	Sam Camarata	19 Big Bend Bld	114391	2	Residential	2004	450575
7	674-112	Chris Capelli	Hastings Barracks	114392	2	Residential	2004	350000
8	674-113	Sheila Sullivan	10034 Endin Mansions	114391	2	Residential	2002	1005425

Record: 0 Show: All Selected Records (0 out of 8 Selected.) Options

(C)

Attributes of Tab10_3a

OBJECTID*	ParcelNumb	ZoningCode	DateAssessed	AssessedValue	OwnersName
1	673-100	2	2002	222000	Jeff Peters
2	673-101	2	2003	545500	Joel Campbell
3	674-100	3	1999	249000	Dave Widseler
4	674-100	3	2000	275500	Dave Widseler
5	674-100	3	2001	290000	Dave Widseler
6	670-231	2	2004	450575	Sam Camarata
7	674-112	2	2004	350000	Chris Capelli
8	674-113	2	2002	1005425	Sheila Sullivan

Record: 8 Show: All Selected Records (0 out of 8 Selected.) Options

Attributes of Tab10_3b

OBJECTID*	ZoningCode	ZoningType
1	2	Residential
2	3	Commercial

Record: 1 Show: All Selected Records (0 out of 2 Selected.)

Attributes of Tab10_3c

OBJECTID*	OwnerName*	Address	PostalCode
2	Jeff Peters	10 Railway Cuttings	114390
3	Joel Campbell	1115 Center Place	114390
4	Dave Widseler	452 Diamond Plaza	114391
5	Sam Camarata	19 Big Bend Bld	114391
6	Chris Capelli	Hastings Barracks	114392
7	Sheila Sullivan	10034 Endin Mansions	114391

Record: 0 Show: All Selected Records (0 out of 6 Selected.)

(D)

Attributes of Tab10_3a_Tab10_3b_Tab10_3c

OBJECTID	ParcelNumb	DateAssessed	AssessedValue	ZoningType	ZoingCode	OwnerName	OwnerAddress	PostalCode
1	673-100	2002	222000	Residential	2	Jeff Peters	10 Railway Cuttings	114390
2	673-101	2003	545500	Residential	2	Joel Campbell	1115 Center Place	114390
3	674-100	1999	249000	Commercial	3	Dave Widseler	452 Diamond Plaza	114391
4	674-100	2000	275500	Commercial	3	Dave Widseler	452 Diamond Plaza	114391
5	674-102	2004	290000	Residential	2	Dave Widseler	452 Diamond Plaza	114391
6	670-231	2004	450575	Residential	2	Sam Camarata	19 Big Bend Bld	114391
7	674-112	2004	350000	Residential	2	Chris Capelli	Hastings Barracks	114392
8	674-113	2002	1005425	Residential	2	Sheila Sullivan	10034 Endin Mansions	114391

Record: 0 Show: All Selected Records (0 out of 8 Selected.) Options

Figure 10.3 (*continued*)

using the common ZoningCode column, and Figure 10.3C `Attributes of Tab10_3c` can be joined using OwnersName to create Figure 10.3D. It is now possible to execute SQL queries against these joined tables as discussed in the next section.

10.4 SQL

The standard database query language adopted by virtually all mainstream databases is SQL (Structured or Standard Query Language: ISO Standard ISO/IEC 9075). There are many good background books and system implementation manuals on SQL and so only brief details will be presented here. SQL may be used directly via an interactive command line interface; it may be compiled in a general-purpose programming language (e.g., C/C++/C#, Java, or Visual Basic); or it may be embedded in a graphical user interface (GUI). SQL is a set based, rather than a procedural (e.g., Visual Basic) or object-oriented (e.g., Java or C#), programming language designed to retrieve sets (row and column combinations) of data from tables. There are three key types of SQL statements: DDL (data definition language), DML (data manipulation language) and DCL (data control language). The third major revision of SQL (SQL 3) which came out in 2004 defines spatial types and functions as part of a multi-media extension called SQL/MM.

The data in Figure 10.3C may be queried to find parcels where the AssessedValue is greater than $300 000 and the ZoningType is Residential. This is an apparently simple query, but it requires three table joins to execute it. The SQL statements in Microsoft Access are as follows:

```
SELECT Tab10_3a.ParcelNumb, Tab10_3c.Address,
    Tab10_3a.AssessedValue
FROM (Tab10_3b INNER JOIN Tab10_3a ON
    Tab10_3b.ZoningCode =
    Tab10_3a.ZoningCode) INNER JOIN Tab10_3c
    ON Tab10_3a.OwnersName =
    Tab10_3c.OwnerName
WHERE (((Tab10_3a.AssessedValue)>300000) AND
    ((Tab10_3b.ZoningType)="Residential"));
```

The `SELECT` statement defines the columns to be displayed (the syntax is TableName.ColumnName). The `FROM` statement is used to identify and join the three tables (`INNER JOIN` is a type of join that signifies that only matching records in the two tables will be considered). The `WHERE` clause is used to select the rows from the columns using the constraints `(((Tab10_3a.AssessedValue)>300000) AND ((Tab10_3b.ZoningType)="Residential"))`. The result of this query is shown in Figure 10.4. This triplet of `SELECT`, `FROM`, `WHERE` is the staple of SQL queries.

SQL is the standard database query language. Today it has geographic capabilities.

Join3a3b3c : Select Query

ParcelNumb	Address	AssessedValue
673-101	1115 Center Place	545500
670-231	19 Big Bend Bld	450575
674-112	Hastings Barracks	350000
674-113	10034 Endin Mansions	1005425

Record: 1 of 4

Figure 10.4 Results of a SQL query against the tables in Figure 10.3C (see text for query and further explanation)

In SQL, data definition language statements are used to create, alter, and delete relational database structures. The `CREATE TABLE` command is used to define a table, the attributes it will contain, and the primary key (the column used to identify records uniquely). For example, the SQL statement to create a table to store data about Countries, with two columns (name and shape (geometry)) is as follows:

```
CREATE TABLE Countries (
name            VARCHAR(200) NOT NULL PRIMARY
    KEY,
shape           POLYGON NOT NULL
CONSTRAINT  spatial reference
CHECK       (SpatialReference(shape) = 14)
)
```

This SQL statement defines several table parameters. The name column is of type `VARCHAR` (variable character) and can store values up to 200 characters. Name cannot be null (`NOT NULL`), that is, it must have a value, and it is defined as the `PRIMARY KEY`, which means that its entries must be unique. The shape column is of type `POLYGON`, and it is defined as `NOT NULL`. It has an additional spatial reference constraint (projection), meaning that a spatial reference is enforced for all shapes (Type `14` – this will vary by system, but could be Universal Transverse Mercator (UTM) – see Section 5.7.2).

Data can be inserted into this table using the SQL `INSERT` command:

```
INSERT INTO Countries
(Name, Shape) VALUES ('Kenya', Polygon('((x
    y, x y, x y, x y)) ,2))
```

Actual coordinates would need to be substituted for the x, y values. Several additions of this type would result in a table like this:

Name	Shape
Kenya	Polygon geometry
South Africa	Polygon geometry
Egypt	Polygon geometry

Data manipulation language statements are used to retrieve and manipulate data. Objects with a size greater

than 11 000 can be retrieved from the countries table using a SELECT statement:

```
SELECT Countries.Name,
FROM Countries
WHERE Area(Countries.Shape) > 11000
```

In this system implementation, area is computed automatically from the shape field using a DBMS function and does not need to be stored.

Data control language statements handle authorization access. The two main DCL keywords are GRANT and REVOKE that authorize and rescind access privileges respectively.

10.5 Geographic database types and functions

There have been several attempts to define a superset of geographic data types that can represent and process geographic data in databases. Unfortunately space does not permit a review of them all. This discussion will focus on the practical aspects of this problem and will be based on the recently developed International Standards Organization (ISO) and the Open Geospatial Consortium (OGC) standards.

The GIS community working under the auspices of ISO and OGC has defined the core geographic types and functions to be used in a DBMS and accessed using the SQL language. The geometry types are shown in Figure 10.5. The Geometry class is the root class. It has an associated spatial reference (coordinate system and projection, for example, Lambert_Azimuthal_Equal_Area). The Point, Curve, Surface, and GeometryCollection classes are all subtypes of Geometry. The other classes (boxes) and relationships (lines) show how geometries of one type are aggregated from others (e.g., a LineString is a collection of Points). For further explanation of how to interpret this object model diagram see the discussion in Section 8.3.

According to this ISO/OGC standard, there are nine methods for testing spatial relationships between these geometric objects. Each takes as input two geometries (collections of one or more geometric objects) and evaluates whether the relationship is true or not. Two examples of possible relations for all point, line, and polygon combinations are shown in Figure 10.6. In the case of the point-point Contain combination (northeast square in Figure 10.6A), two comparison geometry points (big circles) are contained in the set of base geometry points (small circles). In other words, the base geometry is a superset of the comparison geometry. In the case of line-polygon Touches combination (east square in Figure 10.6B) the two lines touch the polygon because they intersect the polygon boundary. The full set of Boolean operators to test the spatial relationships between geometries is:

- Equals – are the geometries the same?
- Disjoint – do the geometries share a common point?
- Intersects – do the geometries intersect?
- Touches – do the geometries intersect at their boundaries?
- Crosses – do the geometries overlap (can be geometries of different dimensions, for example, lines and polygons)?

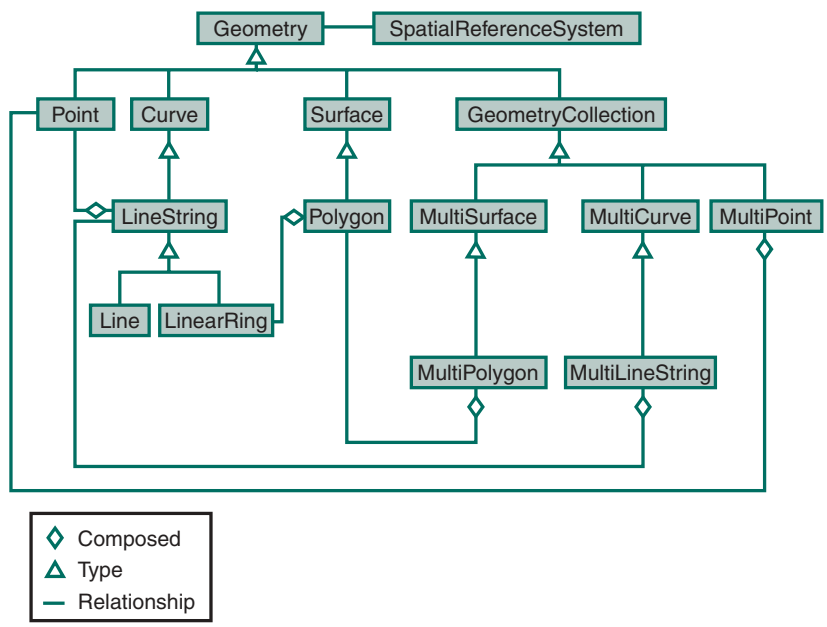
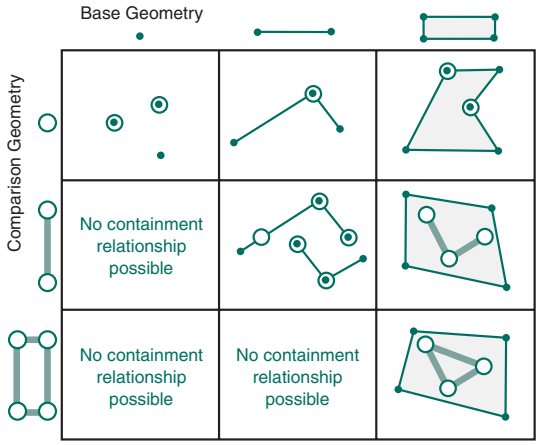


Figure 10.5 Geometry class hierarchy (*Source*: after OGC 1999) (Reproduced by permission of Open Geospatial Consortium, Inc.)

(A) **Contains**

Does the base geometry contain the comparison geometry?

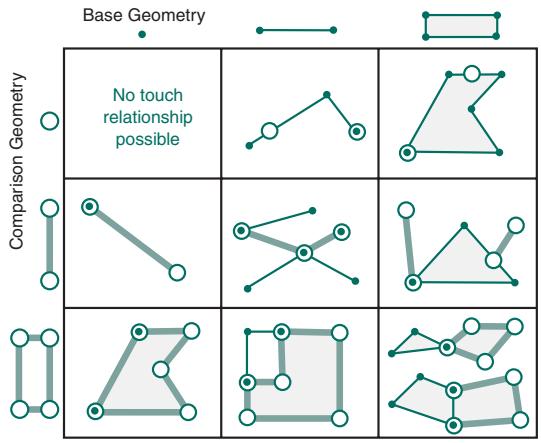


For the base geometry to contain the comparison geometry, it must be a superset of that geometry.

A geometry cannot contain another geometry of higher dimension.

(B) **Touches**

Does the base geometry touch the comparison geometry?



Two geometries touch when only their boundaries intersect.

Figure 10.6 Examples of possible relations for two geographic database operators: (A) Contains; and (B) Touches operators (*Source*: after Zeiler 1999)

- Within – is one geometry within another?
- Contains – does one geometry completely contain another?
- Overlaps – do the geometries overlap (must be geometries of the same dimension)?
- Relate – are there intersections between the interior, boundary, or exterior of the geometries?

Seven methods support spatial analysis on these geometries. Four examples of these methods are shown in Figure 10.7.

- Distance – determines the shortest distance between any two points in two geometries (Section 14.3.1).
- Buffer – returns a geometry that represents all the points whose distance from the geometry is less than or equal to a user-defined distance (Section 14.4.1).
- ConvexHull – returns a geometry representing the convex hull of a geometry (a convex hull is the smallest polygon that can enclose another geometry without any concave areas).
- Intersection – returns a geometry that contains just the points common to both input geometries.
- Union – returns a geometry that contains all the points in both input geometries.
- Difference – returns a geometry containing the points that are different between the two geometries.
- SymDifference – returns a geometry containing the points that are in either of the input geometries, but not both.

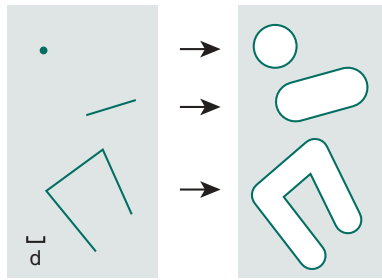
10.6 Geographic database design

This section is concerned with the technical aspects of logical and physical geographic database design. Chapter 8 provides an overview of these subjects and Chapters 17 to 20 discuss the organizational, strategic, and business issues associated with designing and maintaining a database.

10.6.1 The database design process

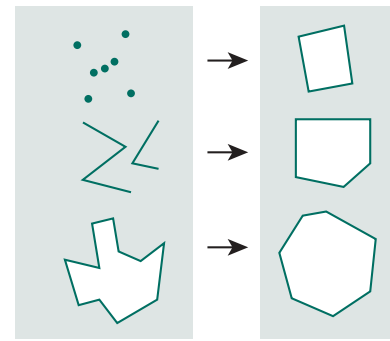
All GIS and DBMS packages have their own core data model that defines the object types and relationships that can be used in an application (Figure 10.8). The DBMS package will define and implement a model for data types and access functions, such as those in SQL discussed in Section 10.4. DBMS are capable of dealing with simple features and types (e.g., points, lines, and polygons) and relationships. A GIS can build on top of these simple feature types to create more advanced types and relationships (e.g., TINs, topologies, and feature-linked annotation geographic relationships: see Chapter 8 for a definition of these terms). The GIS types can be combined with domain data models that define specific object classes and relationships for specialist domains (e.g., water utilities, city parcel maps, and census geographies). Lastly, individual projects create specific physical data model instances that are populated with data (objects for the specified object classes). For example, a City Planning department may build a database of sewer lines that uses

(A) Buffer



Given a geometry and a buffer distance, the buffer operator returns a polygon that covers all points whose distance from the geometry is less than or equal to the buffer distance.

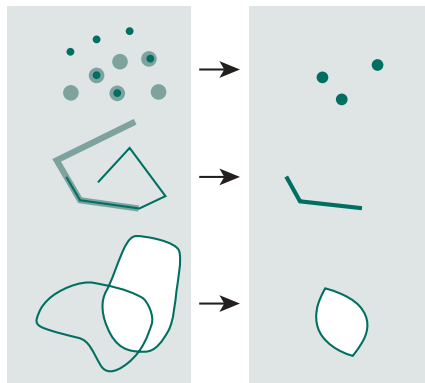
(B) Convex Hull



Given an input geometry, the convex hull operator returns a geometry that represents all points that are within all lines between all points in the input geometry.

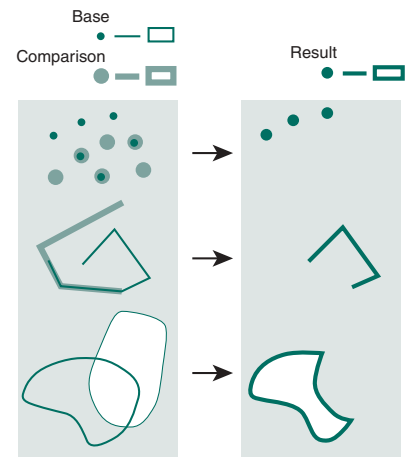
A convex hull is the smallest polygon that wraps another geometry without any concave areas.

(C) Intersection



The intersect operator compares a base geometry (the object from which the operator is called) with another geometry of the same dimension and returns a geometry that contains the points that are in both the base geometry and the comparison geometry.

(D) Difference



The difference operator returns a geometry that contains points that are in the base geometry and subtracts points that are in the comparison geometry.

Figure 10.7 Examples of spatial analysis methods on geometries: (A) Buffer; (B) Convex Hull; (C) Intersection; (D) Difference (*Source*: after Zeiler 1999)

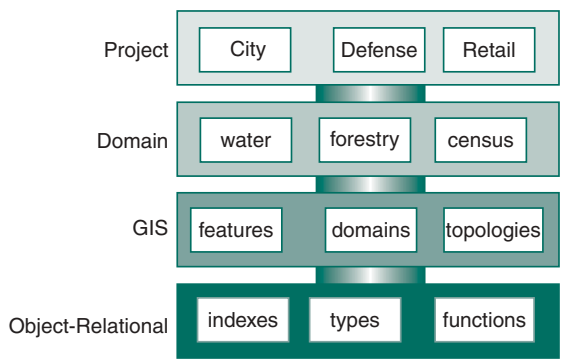


Figure 10.8 Four levels of data model available for use in GIS projects with examples of constructs used

a water/wastewater (sewer) domain data model template which is built on top of core GIS and DBMS models.

Figure 8.2 and Section 8.1.2 show three increasingly abstract stages in data modeling: conceptual, logical, and physical. The result of a data modeling exercise is a physical database design. This design will include specification of all data types and relationships, as well as the actual database configuration required to store them.

Database design involves three key stages: conceptual, logical, and physical.

Database design involves the creation of conceptual, logical, and physical models in the six practical steps shown in Figure 10.9:

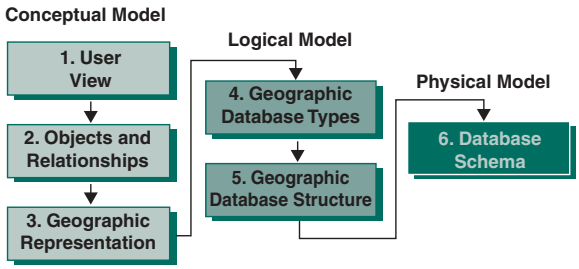


Figure 10.9 Stages in database design (*Source*: after Zeiler 1999)

10.6.1.1 Conceptual model

Model the user's view. This involves tasks such as identifying organizational functions (e.g., controlling forestry resources, finding vacant land for new building, and maintaining highways), determining the data required to support these functions, and organizing the data into groups to facilitate data management. This information can be represented in many ways – a report with accompanying tables is often used.

Define objects and their relationships. Once the functions of an organization have been defined, the object types (classes) and functions can be specified. The relationships between object types must also be described. This process usually benefits from the rigor of using object models and diagrams to describe a set of object classes and the relationships between them.

Select geographic representation. Choosing the types of geographic representation (discrete object – point, line, polygon – or field: see Section 3.6) will have profound implications for the way a database is used and so it is a critical database design task. It is, of course, possible to change between representation types, but this is computationally expensive and results in loss of information.

10.6.1.2 Logical model

Match to geographic database types. This involves matching the object types to be studied to specific data types supported by the GIS that will be used to create and maintain the database. Because the data model of the GIS is usually independent of the actual storage mechanism (i.e., it could be implemented in Oracle, Microsoft Access, or a proprietary file system), this activity is defined as a logical modeling task.

Organize geographic database structure. This includes tasks such as defining topological associations, specifying rules and relationships, and assigning coordinate systems.

10.6.1.3 Physical model

Define database schema. The final stage is definition of the actual physical database schema that will hold the database data values. This is usually created using the DBMS software's data definition language. The most popular of these is SQL with geographic extensions (see Section 10.4), although some non-standard variants also exist in older GIS/DBMS.

10.7 Structuring geographic information

Once data have been captured in a geographic database according to a schema defined in a geographic data model, it is often desirable to perform some structuring and organization in order to support efficient query, analysis, and mapping. There are two main structuring techniques relevant to geographic databases: topology creation and indexing.

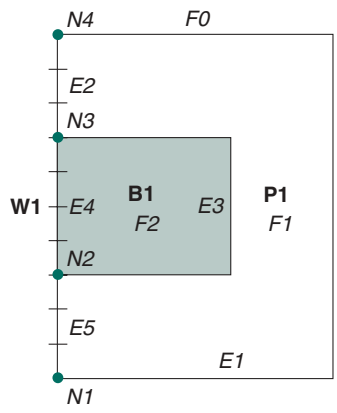
10.7.1 Topology creation

The subject of building topology was covered in Section 8.2.3.2 from a conceptual data modeling perspective, and is revisited here in the context of databases where the discussion focuses on the two main approaches to structuring and storing topology in a DBMS.

Topology can be created for vector datasets using either batch or interactive techniques. Batch topology builders are required to handle CAD, survey, simple feature, and other unstructured vector data imported from non-topological systems. Creating topology is usually an iterative process because it is seldom possible to resolve all data problems during the first pass and manual editing is required to make corrections. Some typical problems that need to be fixed are shown in Figures 9.9, 9.10, and 9.11 and discussed in Section 9.3.2.3. Interactive topology creation is performed dynamically at the time objects are added to a database using GIS editing software. For example, when adding water pipes using interactive vectorization tools (see Section 9.3.2.2), before each object is committed to the database topological connectivity can be checked to see if the object is valid (that is, it conforms to some pre-established database object connectivity rules).

Two database-oriented approaches have emerged in recent years for storing and managing topology: Normalized and Physical. The Normalized Model focuses on the storage of an arc-node data structure. It is said to be normalized because each object is decomposed into individual topological primitives for storage in a database and then subsequent reassembly when a query is posed. For example, polygon objects are assembled at query time by joining together tables containing the line segment geometries and topological primitives that define topological relationships (see Section 8.2.3.2 for a conceptual description of this process). In the Physical Model topological primitives are not stored in the database and the entire geometry is stored together for each object. Topological relationships are then computed on-the-fly whenever they are required by client applications.

Figure 10.10 is a simple example of a set of database tables that store a dataset according to the Normalized Model for topology. The dataset (sketch in top left corner) comprises three feature classes (Parcels, Buildings, and Walls) and is implemented in three tables. In this example the three feature class tables have only one column



Parcels

ID
P1

Buildings

ID
B1

Walls

ID
W1

Parcel x Face

Parcel	Face
P1	F1
P1	F2

Wall x Edge

Wall	Edge	Order	Orientation
W1	E2	1	+
W1	E4	2	-
W1	E5	3	+

Building x Face

Parcel	Face
B1	F2

Nodes

ID
N1
N2
N3
N4

Edges

ID	Vertices	From	To	Left	Right
E1	(0,10),(8,10),(8,0),(0,0)	N4	N1	F0	F1
E2	(0,10),(0,7)	N4	N3	F1	F0
E3	(0,7),(5,7),(5,3),(0,3)	N3	N2	F1	F2
E4	(0,3),(0,7)	N2	N3	F0	F2
E5	(0,3),(0,0)	N2	N1	F1	F0

Faces

ID
F0
F1
F2

Figure 10.10 Normalized database topology model

(ID) and one row (a single instance of each feature in a feature class). The Nodes, Edges, and Faces tables store the points, lines, and polygons for the dataset and some of the topology (for each Edge the From-To connectivity and the Faces on the Left-Right in the direction of digitizing). Three other tables (Parcel × Face, Wall × Edge, and Building × Face) store the cross references for assembling Parcels, Buildings, and Walls from the topological primitives.

The Normalized approach offers a number of advantages to GIS users. Many people find it comforting to see topological primitives actually stored in the database. This model has many similarities to the arc-node conceptual topology model (see Section 8.2.3.2) and so it is familiar to many users and easy to understand. The geometry is only stored once thus minimizing database size and avoiding 'double digitizing' slivers (Section 9.3.2.3). Finally, the normalized approach easily lends itself to access via a SQL Application Programming Interface (API). Unfortunately, there are three main disadvantages associated with the Normalized approach to database topology: query performance, integrity checking, and update performance/complexity.

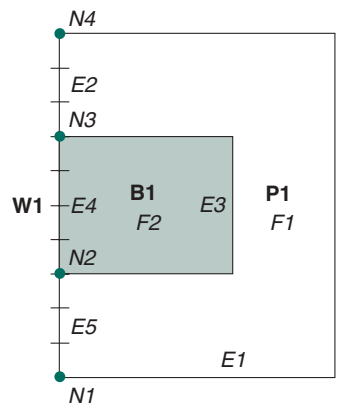
Query performance suffers because queries to retrieve features from the database (much the most common type of query) must combine data from multiple tables. For example, to fetch the geometry of Parcel P1, a query must combine data from four tables (Parcels, Parcel × Face, Faces, and Edges) using complex geometry/topology logic. The more tables that have to be visited, and especially the more that have to be joined, the longer it will take to process a query.

The standard referential integrity rules in DBMS are very simple and have no provision for complex topological relationships of the type defined here. There are many pitfalls associated with implementing topological structuring using DBMS techniques such as stored procedures (program code stored in a database) and in practice systems have resorted to implementing the business logic to manage things like topology external to the DBMS in a middle-tier application server (see Section 7.3.2).

Updates are similarly problematic because changes to a single feature will have cascading effects on many tables. This raises attendant performance (especially scalability, that is, large numbers of concurrent queries) and integrity issues. Moreover, it is uncertain how multi-user updates will be handled that entail long transactions with design alternatives (see Sections 10.9 and 10.9.1 for coverage of these two important topics).

For comparative purposes the same dataset used in the Normalized model (Figure 10.10) is implemented using the Physical model in Figure 10.11. In the physical model the three feature classes (Parcels, Buildings, and Walls) contain the same IDs, but differ significantly in that they also contain the geometry for each feature. The only other things required to be stored in the database are the specific set of topology rules that have been applied to the dataset (e.g., parcels should not overlap each other, and buildings should not overlap with each other), together with information about known errors (sometimes users defer topology clean-up and commit data with known errors to a database) and areas that have been edited, but not yet been validated (had their topology (re-)built).

The Physical model requires that an external client or middle-tier application server is responsible for validating the topological integrity of datasets. Topologically correct features are then stored in the database using a much simpler structure than the Normalized model. When compared to the Normalized model, the Physical model offers



Parcels

ID	Vertices
P1	(0,0),(0,3),(0,7),(0,10),(8,10),(8,0),(0,0)

Buildings

ID	Vertices
B1	(0,3),(0,7),(5,7),(5,3),(0,3)

Walls

ID	Vertices
W1	(0,10),(0,7),(0,3),(0,0)

Topology Rules

ID	Rule Type
R1	Parcels no overlap
R2	Buildings no overlap

Topology Errors

ID	Vertices	Rule Type	Feature	IsException
E1	…	-	-	-

Dirty Area

Vertices
…

Figure 10.11 Physical database topology model

two main advantages of simplicity and performance. Since all the geometry for each feature is stored in the same table column/row value, and there is no need to store topological primitives and cross references, this is a very simple model. The biggest advantage and the reason for the appeal of this approach is query performance. Most DBMS queries do not need access to the topology primitives of features and so the overhead of visiting and joining multiple tables (four) is unnecessary. Even when topology is required it is faster to retrieve feature geometries and re-compute topology outside the database, than to retrieve geometry and topology from the database.

In summary, there are advantages and disadvantages to both the Normalized and Physical topology models. The Normalized model is implemented in Oracle Spatial and can be accessed via a SQL API making it easily available to a wide variety of users. The Physical model is implemented in ESRI ArcGIS and offers fast update and query performance for high-end GIS applications. ESRI has also implemented a long transaction and versioning model based on the physical database topology model.

10.7.2 Indexing

Geographic databases tend to be very large and geographic queries computationally expensive. Because of this, geographic queries, such as finding all the customers (points) within a store trade area (polygon), can take a very long time (perhaps 10–100 seconds or more for a 50 million customer database). The point has already been made in Section 8.2.3.2 that topological structuring can help speed up certain types of queries such as adjacency and network connectivity. A second way to speed up queries is to index a database and use the index to find data records (database table rows) more quickly. A database index is logically similar to a book index; both are special organizations that speed up searching by allowing random instead of sequential access. A database index is, conceptually speaking, an ordered list derived from the data in a table. Using an index to find data reduces the number of computational tests that have to be performed to locate a given set of records. In DBMS jargon, indexes avoid expensive full-table scans (reading every row in a table) by creating an index and storing it as a table column.

A database index is a special representation of information about objects that improves searching.

Figure 10.12 is a simple example of the standard DBMS one-dimensional B-tree (Balanced Tree) index that is found in most major commercial DBMS. Without an index a search of the original data to guarantee finding any given value will involve 16 tests/read operations (one for each data point). The B-tree index orders the data and splits the ordered list into buckets of a given size (in this example it is four and then two) and the upper value for the bucket is stored (it is not necessary to store the uppermost value). To find a specific value, such as 72, using the index involves a maximum of six tests: one at level 1 (less than or greater than 36), one at level 2 (less than or greater than 68), and a sequential read of four records at Level 3. The number of levels and buckets for each level can be optimized for each type of data. Typically, the larger the dataset the more effective indexes are in retrieving performance.

Unfortunately, creating and maintaining indexes can be quite time consuming and this is especially an issue when the data are very frequently updated. Since indexes can occupy considerable amounts of disk space, requirements can be very demanding for large datasets. As a consequence, many different types of index have been developed to try and alleviate these problems for both geographic and non-geographic data. Some indexes exploit specific characteristics of data to deliver optimum

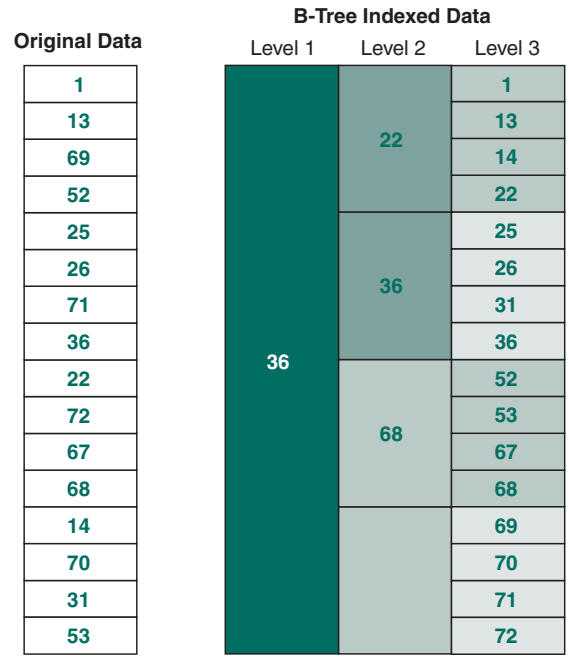


Figure 10.12 An example B-tree index

query performance, some are fast to update, and others are robust across widely different data types.

The standard indexes in DBMS, such as the B-tree, are one-dimensional and are very poor at indexing geographic objects. Many types of geographic indexing techniques have been developed. Some of these are experimental and have been highly optimized for specific types of geographic data. Research shows that even a basic spatial index will yield very significant improvements in spatial data access and that further refinements often yield only marginal improvements at the costs of simplicity, and speed of generation and update. Three main methods of general practical importance have emerged in GIS: grid indexes, quadtrees, and R-trees.

10.7.2.1 Grid index

A grid index can be thought of as a regular mesh placed over a layer of geographic objects. Figure 10.13 shows a layer that has three features indexed using two grid levels. The highest (coarsest) grid (Index 1) splits the layer into four equal sized cells. Cell A includes parts of Features 1, 2, and 3, Cell B includes a part of Feature 3 and Cell C has part of Feature 1. There are no features on Cell D. The same process is repeated for the second level index (Index 2). A query to locate an object searches the indexed list first to find the object and then retrieves the object geometry or attributes for further analysis (e.g., tests for overlap, adjacency, or containment with other objects on the same or another layer). These two tests are often referred to as primary and secondary filters. Secondary filtering, which involves geometric processing, is much more computationally expensive. The performance of an index is clearly related to the relationship between grid and object size, and object density. If the grid size is too large relative to the size of object, too many objects will be retrieved by the primary filter, and therefore a lot of expensive secondary processing will be needed. If the grid size is too small any large objects will be spread across many grid cells, which is inefficient for draw queries (queries made to the database for the purpose of displaying objects on a screen).

For data layers that have a highly variable object density (for example, administrative areas tend to be smaller and more numerous in urban areas in order to equalize population counts) multiple levels can be used to optimize performance. Experience suggests that three grid levels are normally sufficient for good all-round performance. Grid indexes are one of the simplest and most robust indexing methods. They are fast to create and update and can handle a wide range of types and densities of data. For this reason they have been quite widely used in commercial GIS software systems.

Grid indexes are easy to create, can deal with a wide range of object types, and offer good performance.

10.7.2.2 Quadtree indexes

Quadtree is a generic name for several kinds of index that are built by recursive division of space into quadrants. In many respects quadtrees are a special types of grid index. The difference here is that in quadtrees space is

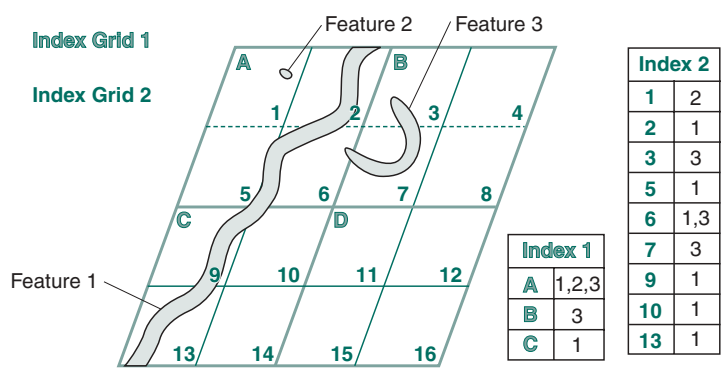


Figure 10.13 A multi-level grid geographic database index

recursively split into four quadrants based on data density. Quadtrees are data structures used for both indexing and compressing geographic database layers, although the discussion here will relate only to indexing. The many types of quadtree can be classified according to the types of data that are indexed (points, lines, areas, surfaces, or rasters), the algorithm that is used to decompose (divide) the layer being indexed, and whether fixed or variable resolution decomposition is used.

In a point quadtree, space is divided successively into four rectangles based on the location of the points (Figure 10.14). The root of the tree corresponds to the region as a whole. The rectangular region is divided into four usually irregular parts based on the x, y coordinate of the first point. Successive points subdivide each new sub-region into quadrants until all the points are indexed.

Region quadtrees are commonly used to index lines, polygons, and rasters. The quadtree index is created by successively dividing a layer into quadrants. If a quadrant cell is not completely filled by an object, then it is sub-divided again. Figure 10.15 is a quadtree of a woodland (red) and water (white) layer. Once a layer has been decomposed in this way, a linear index can be created using the search order shown in Figure 10.16. By reducing two-dimensional geographic data to a single

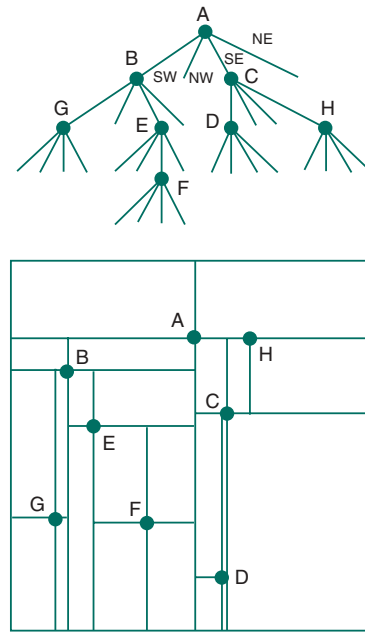


Figure 10.14 The point quadtree geographic database index (*Source*: after van Oosterom 2005) (Reproduced by permission of John Wiley & Sons Inc.)

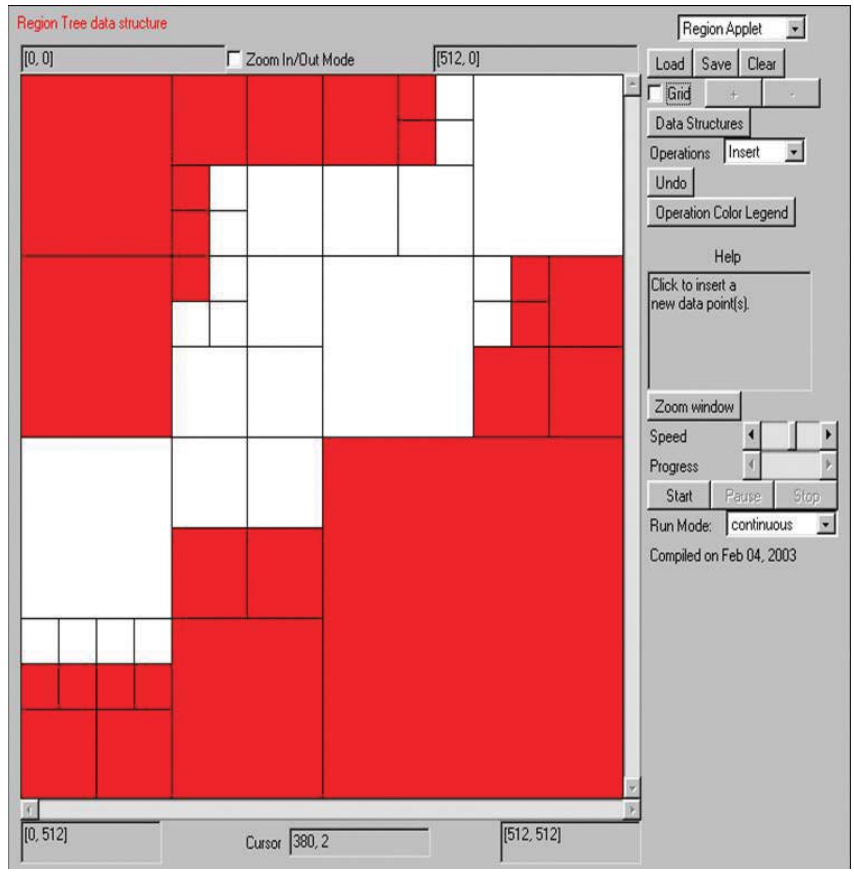


Figure 10.15 The region quadtree geographic database index (from **www.cs.umd.edu/~brabec/quadtree/**) (Reproduced by permission of Hanan Samet and Frantisek Brabec)

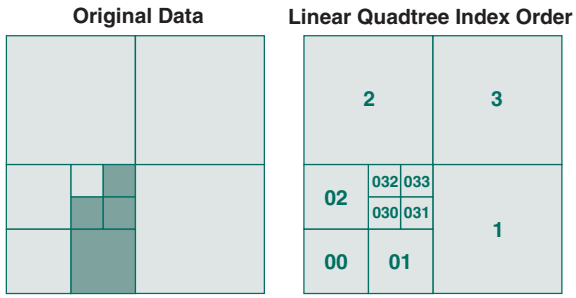


Figure 10.16 Linear quadtree search order

linear dimension, a standard B-tree can be used to find data quickly.

Quadtrees have found favor in GIS software systems because of their applicability to many types of data (both raster and vector), their ease of implementation, and their relatively good performance.

10.7.2.3 R-tree indexes

R-trees group objects using a rectangular approximation of their location called a minimum bounding rectangle (MBR) or minimum enclosing rectangle (see Box 10.2). Groups of point, line, or polygon objects are indexed based on their MBR. Objects are added to the index by choosing the MBR that would require the least expansion to accommodate each new object. If the object causes the MBR to be expanded beyond some preset parameter then the MBR is split into two new MBR. This may also cause the parent MBR to be become too large, resulting in this also being split. The R-tree shown in Figure 10.17 has two levels. The lowest level contains three 'leaf nodes';

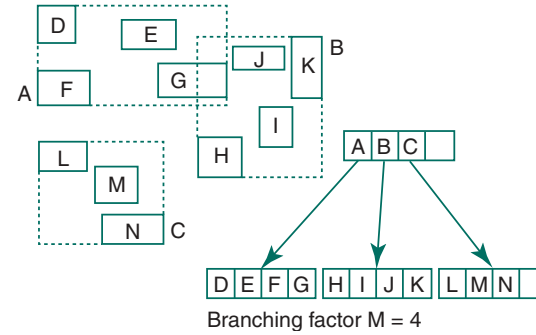


Figure 10.17 The R-tree geographic database index (*Source*: after van Oosterom 2005) (Reproduced by permission of John Wiley & Sons Inc.)

Technical Box **10.2**

Minimum Bounding Rectangle

Minimum Bounding Rectangles (MBR) are very useful structures that are widely implemented in GIS. An MBR essentially defines the smallest box whose sides are parallel to the axes of the coordinate system that encloses a set of one or more geographic objects. It is defined by the two coordinates at the bottom left (minimum *x*, minimum *y*) and the top right (maximum *x*, maximum *y*) as is shown in Figure 10.18.

MBR can be used to generalize a set of data by replacing the geometry of the objects in the box with two coordinates defining the box. A second use is for fast searching. For example, all the polygon objects in a database layer that are within a given study area can be found by performing a polygon on polygon contains test (see Figure 10.18) with each object and the study area boundary. If the polygon objects have complex boundaries (as is normally the case in GIS) this can be a very time-consuming task. A quicker approach is to split the task into two parts. First, screen out all the objects that are definitely in and definitely out by comparing their MBR. Because very few coordinate comparisons are required this is very fast. Then use the full geometry outline of the remaining polygon objects to determine containment. This is computationally expensive for polygons with complex geometries.

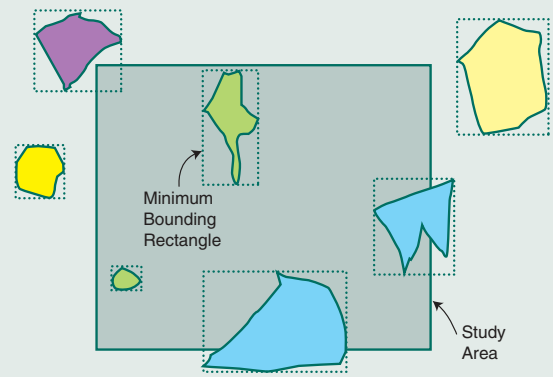


Figure 10.18 Polygon in polygon test using MBR. A MBR can be used to determine objects definitely within the study area (green) because of no overlap, definitely out (yellow), or possibly in (blue). Objects possibly in can then be analyzed further using their exact geometries. Note the purple object that is actually completely outside, although the MBR suggests it is partially within the study area

the highest has one node with pointers to the MBR of the leaf nodes. The MBR is used to reduce the number of objects that need to be examined to satisfy a query.

R-trees are popular methods of indexing geographic data because of their flexibility and excellent performance.

R-trees are suitable for a range of geographic object types and can be used for multi-dimensional data. They offer good query performance, but the speed of update is not as fast as grids and quadtrees. The Spatial Datablade extensions to the IBM Informix DBMS and Oracle Spatial both use R-tree indexes.

10.8 Editing and data maintenance

Editing is the process of making changes to a geographic database by adding new objects or changing existing objects as part of data load or database update and maintenance operations. A database update is any change to the geometry and/or attributes of one or more objects or any change to the database schema. A general-purpose geographic database will require many tools for tasks such as geometry and attribute editing, database maintenance (e.g., system administration and tuning), creating and updating indexes and topology, importing and exporting data, and georeferencing objects.

Contemporary GIS come equipped with an extensive array of tools for creating and editing geographic object geometries and attributes. These tools form workflow tasks that are exposed within the framework of a WYSIWYG (what you see is what you get) editing environment. The objects are displayed using map symbology usually in a projected coordinate system space and frequently on top of 'background' layers such as aerial photographs or satellite images, or street centerline files. Object coordinates can be digitized into a geographic database using many methods including freehand digitizing on a digitizing table, on screen heads-up vector digitizing by copying existing raster and/or vector sources, on screen semi-automatic line following, automated feature recognition, and reading survey measurements from an instrument (e.g., GPS or Total Station) file (see Sections 5.8 and 9.2.2.1). The end result is always a layer of x, y coordinates with optional z and m (height and attribute) values. Similar tools also exist for loading and editing raster data.

Data entered into the editor must be stored persistently in a file system or database and access to the database must be carefully managed to ensure continued security and quality is maintained. The mechanism for managing edits to a file or database is a called a transaction. There are many challenging issues associated with implementing multi-user access geographic data stored in a DBMS as discussed in the next section.

10.9 Multi-user editing of continuous databases

For many years, one of the most challenging problems in GIS data management was how to allow multiple users to edit the same continuous geographic database at the same time. In GIS applications the objects of interest (geographic features) do not exist in isolation and are usually closely related to surrounding objects. For example, a tax parcel will share a common boundary with an adjacent parcel and changes in one will directly affect the other; similarly connected road segments in a highway network need to be edited together to ensure continued connectivity. It is relatively easy to provide multiple users with concurrent read and query access to a continuous shared database, but more difficult to deal with conflicts and avoid potential database corruption when multiple users want write (update) access. However, solutions to both these problems have been implemented in mainstream GIS and DBMS. These solutions extend standard DBMS transaction models and provide a multi-user framework called versioning.

10.9.1 Transactions

A group of edits to a database, such as the addition of three new land parcels and changes to the attributes of a sewer line, is referred to as a 'transaction'. In order to protect the integrity of databases, transactions are atomic, that is, transactions are either completely committed to the database or they are rolled back (not committed at all). Many of the world's GIS and non-GIS databases are multi-user and transactional, that is, they have multiple users performing edit/update operations at the same time. For most types of database, transactions take a very short time (sub-second). For example, in the case of a banking system a transfer from a savings account to a checking account takes perhaps 0.01 second. It is important that the transaction is coordinated between the accounts and that it is atomic, otherwise one account might be debited and the other not credited. Multi-user access to banking and similar systems is handled simply by locking (preventing access to) affected database records (table rows) during the course of the transaction. Any attempt to write to the same record is simply postponed until the record lock is removed after the transaction is completed. Because banking transactions, like many others, take only a very short amount of time, users never even notice if a transaction is deferred.

A transaction is a group of changes that are made to a database as a coherent group. All the changes that form part of a transaction are either committed or the database is rolled back to its initial state.

Although some geographic transactions have a short duration (short transactions), many extend to hours,

weeks, and months, and are called long transactions. Consider, for example, the amount of time necessary to capture all the land parcels in a city subdivision (housing estate). This might take a few hours for an efficient operator, for a small subdivision, but an inexperienced operator working on a large subdivision might take days or weeks. This may cause two multi-user update problems. First, locking the whole or even part of a database to other updates for this length of time during a long transaction is unacceptable in many types of application, especially those involving frequent maintenance changes (e.g., utilities and land administration). Second, if a system failure occurs during the editing, work may be lost unless there is a procedure for storing updates in the database. Also, unless the data are stored in the database they are not easily accessible to others that would like to use them.

10.9.2 Versioning

Short transactions use what is called a pessimistic locking concurrency strategy. That is, it is assumed that conflicts will occur in a multi-user database with concurrent users and that the only way to avoid database corruption is to lock out all but one user during an update operation. The term 'pessimistic' is used because this is a very conservative strategy assuming that update conflicts will occur and that they must be avoided at all cost. An alternative to pessimistic locking is optimistic versioning, which allows multiple users to update a database at the same time. Optimistic versioning is based on the assumption that conflicts are very unlikely to occur, but if they do occur then software can be used to resolve them.

The two strategies for providing multi-user access to geographic databases are pessimistic locking and optimistic versioning.

Versioning sets out to solve the long transaction and pessimistic locking concurrency problem described above. It also addresses a second key requirement peculiar to geographic databases – the need to support alternative representations of the same objects in the database. In many applications, it is a requirement to allow designers to create and maintain multiple object designs. For example, when designing a new subdivision, the water department manager may ask two designers to lay out alternative designs for a water system. The two designers would work concurrently to add objects to the same database layers, snapping to the same objects. At some point, they may wish to compare designs and perhaps create a third design based on parts of their two designs. While this design process is taking place, operational maintenance editors could be changing the same objects they are working with. For example, maintenance updates resulting from new service connections or repairs to broken pipes will change the database and may affect the objects used in the new designs.

Figure 10.19 compares linear short transactions and branching long transactions as implemented in a versioned database. Within a versioned database, the different

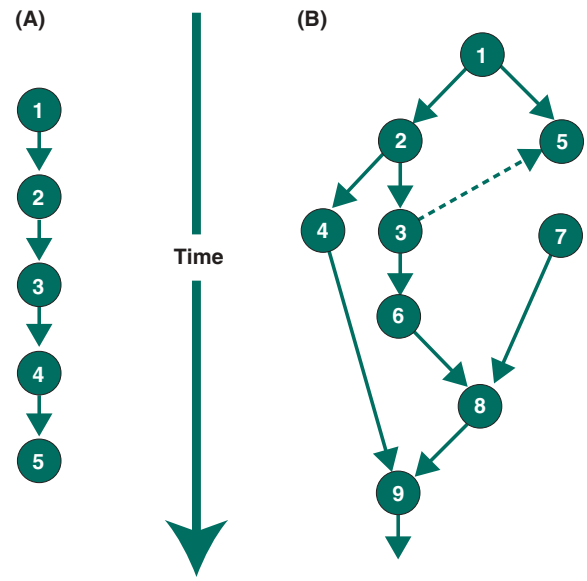


Figure 10.19 Database transactions: (A) linear short transactions; (B) branching version tree

database versions are logical copies of their parents (base tables), i.e., only the modifications (additions and deletions) are stored in the database (in version change tables). A query against a database version combines data from the base table with data in the version change tables. The process of creating two versions based on the same parent version is called branching. In Figure 10.19B, Version 4 is a branch from Version 2. Conversely, the process of combining two versions into one version is called merging. Figure 10.19B also illustrates the merging of Versions 6 and 7 into Version 8. A version can be updated at any time with any changes made in another version. Version reconciliation can be seen between Version 3 and 5. Since the edits contained within Version 5 were reconciled with 3, only the edits in Versions 6 and 7 are considered when merging to create Version 8.

There are no database restrictions or locks placed on the operations performed on each version in the database. The versioning database schema isolates changes made during the edit process. With optimistic versioning, it is possible for conflicting edits to be made within two separate versions, although normal working practice will ensure that the vast majority of edits made will not result in any conflicts (Figure 10.20). In the event that conflicts are detected, the data management software will handle them either automatically or interactively. If interactive conflict resolution is chosen, the user is directed to each feature that is in conflict and must decide how to reconcile the conflict. The GUI will provide information about the conflict and display the objects in their various states. For example, if the geometry of an object has been edited in the two versions, the user can display the geometry as it was in any of its previous states and then select the geometry to be added to the new database state.

An example will help to illustrate how versioning works in practice. In this scenario, a geographic database

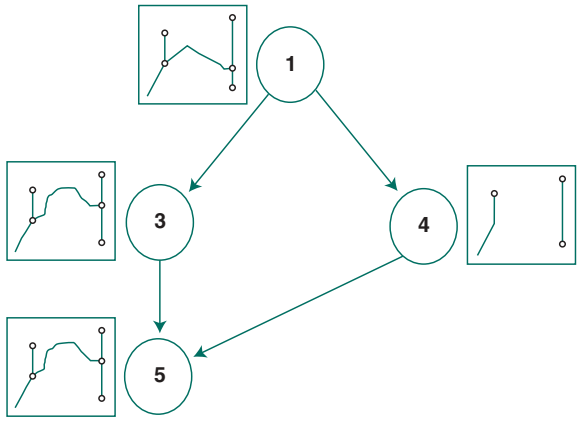


Figure 10.20 Version reconciliation. For Version 5 the user chooses via the GUI the geometry edit made in Version 3 instead of 1 or 4

of a water network similar to that shown in Figure 8.16, called *Main Plant* (Figure 10.21), is edited in parallel by four users each with different tasks. Users 1 and 2 are updating the *Main Plant* database with recent survey results (*Update 1* and *Update 2*), User 3 is performing some initial design prototype work (*Proposal 1*), and User 4 is following the progress of a construction project (*Plan A*).

The current state of the *Main Plant* database is represented by the *As Built* layer (this term is used in utility applications to denote the actual state of objects in the real world). This version is marked as the default (base table) version of the database. Edits made to this version (1) are saved as a second version (2) which now becomes the *As Built* layer. User 4 branches the database, basing the new version (3), *Plan A*, on the *As Built* version (2). As work progresses, edits are made to *Plan A* (3) and in so doing two further states of *Plan A* are created (Versions 4 and 9). Meanwhile, User 3 creates a version (5) *Proposal 1*, which is based on User 4's *Plan A* (3). After initial design work, no further action is taken on *Proposal 1* (5), but for a historical record the version is left undeleted. Users 1 and 2 create versions of *As Built* (2) in order for their edits to be made (6 and 7). When Users 1 and 2 have completed editing, User 1 reconciles the edits in versions *Update 1* (6) and *Update 2* (7) by merging them together, so that version *Update 1* (8) contains all the edits. If any conflicts occur in the edits in the two versions then the versioning software will highlight these and support either automatic or manual conflict resolution. On completion of the work by User 4 the changes made in *Plan A* (9) are merged into the *As Built* version (10) and the intermediate versions (3, 4, and 9) can be deleted from the database (unless they are to be saved for historical reasons). Finally, the updates held in version *Update 1* (8) are merged into the *As Built* version (11).

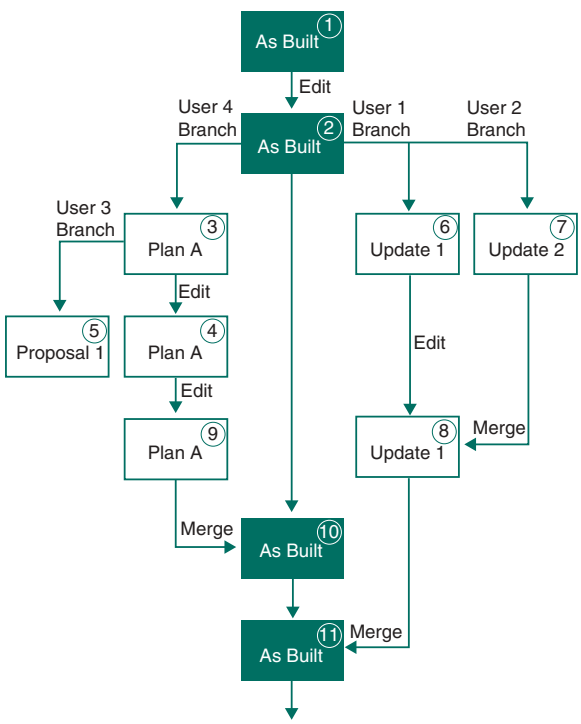


Figure 10.21 A version tree for the Main Plant geographic database. Version numbers are in circles

Users performing read-only queries to the database and even simple edits (such as updating attributes) always work against the current state of the *As Built* layer. They are unaware of the other versions of *Main Plant* until they are made public, or are merged into the *Main Plant As Built* database layer. The *Main Plant* database is not directly edited by end users, conceptually each user is editing their own version of the *As Built* layer. Since the child versions of the *As Built* version are only logical copies, the system can be scaled to many concurrent users without the need for vast increases in processing and disk storage capacity. The changes to the database are actually performed with a GIS map editor using standard editing tools.

Versioning is a very useful mechanism for solving the problem of allowing multiple users to edit a shared continuous geographic database at the same time. In this section the discussion has considered the use of versioning in long transactions, design alternatives, and history (using versions to store the state of a database at specific points in time). In recent years the same versioning technology has been used to underpin replication between multiple databases. In version replication the version becomes the payload for transferring data between systems that need to be synchronized, and the versioning reconciliation framework is used to integrate changes.

10.10 Conclusion

Database management systems are now a vital part of large modern operational GIS. They bring with them standardized approaches for storing and, more

Biographical Box **10.3**

Mike Worboys, mathematician and computer scientist

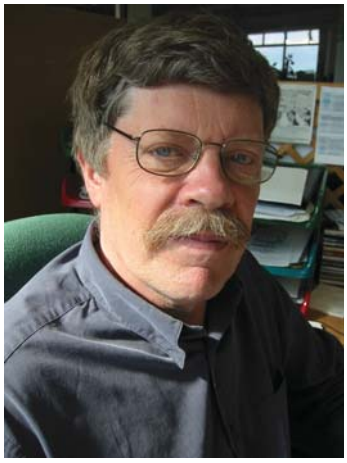


Figure 10.22 Mike Worboys, mathematician and computer scientist

Mike Worboys is a graduate in mathematics and holds a masters degree in mathematical logic and the foundations of mathematics. His Ph.D. in pure mathematics developed computational approaches to very large algebraic structures arising from symmetries in high-dimensional spaces. Mike has worked as a mathematician and computer scientist and has had a long-standing interest in databases. His introduction to GIS was as a member of the newly created UK Midlands Regional Research Laboratory at Leicester in 1989. Mike's interests were in computational data modeling, and he realized that geographic data provided an interesting test-bed for the new object-oriented approaches being developed at that time. In his textbook *GIS: A Computing Perspective* (Worboys and Duckham 2004), Mike describes the database as the heart of a GIS. Much of Mike's work has been to bring more expressive approaches from the formal and computing communities to bear on the representation and management of geographic phenomena. He has subsequently been involved in research on computational models for spatio-temporal information and uncertainty in geographic information. He is presently Professor of Spatial Information Science and Engineering and a member of the National Center for Geographic Information and Analysis at the University of Maine.

Reflecting on the contributions of mathematics and computing to geographic information science, Mike says: 'I increasingly believe that database technology can only take us so far along the road to effective systems. Understanding human conceptions and representations of geographic space, and translating these into formal and computational models, provide the real key. As for the future, I think the challenge is to build systems supporting truly user-centric views of information about both static and dynamic geographic phenomena.'

importantly, accessing and manipulating geographic data using the SQL query language. GIS provide the necessary tools to load, edit, query, analyze, and display geographic data. DBMS require a database administrator (DBA) to control database structure and security, and to tune the database to achieve maximum performance. Innovative work in the GIS field has extended standard DBMS to store and manage geographic data and has led to the development of long transactions and versioning that have application across several fields. Mike Worboys, a leading authority on spatial DBMS, offers his insights into the future of this important area of GIS in Box 10.3.

Questions for further study

1. Identify a geographic database with multiple layers and draw a diagram showing the tables and the relationships between them. Which are the primary keys, and which keys are used to join tables? Does the database have a good relational design?
2. What are the advantages and disadvantages of storing geographic data in a DBMS?
3. Is SQL a good language for querying geographic databases?
4. Why are there multiple methods of indexing geographic databases?

Further reading

Date C.J. 2003 *Introduction to Database Systems* (8th edn). Reading, MA: Addison-Wesley.

Hoel E., Menon S. and Morehouse S. 2003 'Building a robust relational implementation of topology.' In Hadzilacos T., Manolopoulos Y., Roddick J.F. and Theodoridis Y. (eds) *Advances in Spatial and Temporal Databases. Proceedings of 8th International Symposium, SSTD 2003 Lecture Notes in Computer Science*, Vol. 2750.

OGC 1999 OpenGIS simple features specification for SQL, Revision 1.1. Available at **www.opengis.org**

Samet H. 1990 *The Design and Analysis of Spatial Data Structures*. Reading, MA: Addison-Wesley.

van Oosterom P. 2005 'Spatial access methods.' In Longley P.A., Goodchild M.F., Maguire D.J. and Rhind D.W. (eds) *Geographic Information Systems: Principles, Techniques, Applications and Management (abridged edition)*. Hoboken, NJ: Wiley, 385–400.

Worboys M.F. and Duckham M. 2004 *GIS: A Computing Perspective* (2nd edn). Boca Raton, FL: CRC Press.

Zeiler M. 1999 *Modeling our World: The ESRI Guide to Geodatabase Design*. Redlands, CA: ESRI Press.

11 *Distributed GIS*

Until recently, the only practical way to apply GIS to a problem was to assemble all of the necessary parts in one place, on the user's desktop. But recent advances now allow all of the parts – the data and the software – to be accessed remotely, and moreover they allow the user to move away from the desktop and hence to apply GIS anywhere. Limited GI services are already available in common mobile devices such as cellphones, and are increasingly being installed in vehicles. This chapter describes current capabilities in distributed GIS, and looks to a future in which GIS is increasingly mobile and available everywhere. It is organized into three major sections, dealing with distributed data, distributed users, and distributed software.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will understand:

- How the parts of GIS can be distributed instead of centralized;
- The concept of a geolibrary, and the standards and protocols that allow remotely stored data to be discovered and accessed;
- The capabilities of mobile devices, including cellphones and wearable computers;
- The concept of augmented reality, and its relationship to location-based services;
- The concept of remotely invoked services, and their applications in GIS.

11.1 Introduction

Early computers were extremely expensive, forcing organizations to provide computing services centrally, from a single site, and to require users to come to the computing center to access its services. As the cost of computers fell, from millions of dollars in the 1960s, to hundreds of thousands in the late 1970s, and now to a thousand or so, it became possible for departments, small groups, and finally individuals to own computers, and to install them on their desktops (see Chapter 7). Today, of course, the average professional worker expects to have a computer in the office, and a substantial proportion of households in industrialized countries have computers in the home.

In Chapter 1 we identified the six component parts of a GIS as its hardware, software, data, users, procedures, and network. This chapter describes how the network, to which almost all computers are now connected, has enabled a new vision of *distributed* GIS, in which the component parts no longer need to be co-located. New technologies are moving us rapidly to the point where it will be possible for a GIS project to be conducted not only at the desktop but anywhere the user chooses to be, using data located anywhere on the network, and using software services provided by remote sites on the network.

In distributed GIS, the six component parts may be at different locations.

In Chapter 7 we already discussed some aspects of this new concept of distributed GIS. Many GIS vendors are now offering limited-functionality software that is capable of running in mobile devices, such as PDAs and tablet computers, which we have termed Hand-Held GIS (Section 7.6.4). Such devices are now widely used for field data collection, by statistical agencies such as the US Bureau of the Census or the US Department of Agriculture and by utilities and environmental companies. Such organizations now routinely equip their field crews with PDAs, allowing them to record data in the field and to upload the results back in the office (Figure 11.1); and similar systems are used by the field crews of utility companies to record the locations and status of transformers, poles, and switches (Figure 11.2). Vendors are also offering various forms of what we have termed Server or Internet GIS (Section 7.6.2). These allow GIS users to access processing services at remote sites, using little more than a standard Web browser, thus avoiding the cost of installing a GIS locally.

Certain concepts need to be clarified at the outset. First, there are four distinct locations of significance to distributed GIS:

- The location of the user and the interface from which the user obtains GIS-created information, denoted by *U* (e.g., the doorway in Figure 11.2).
- The location of the data being accessed by the user, denoted by *D*. Traditionally, data had to be moved to the user's computer before being used, but new technology is allowing data to be accessed directly from data warehouses and archives.

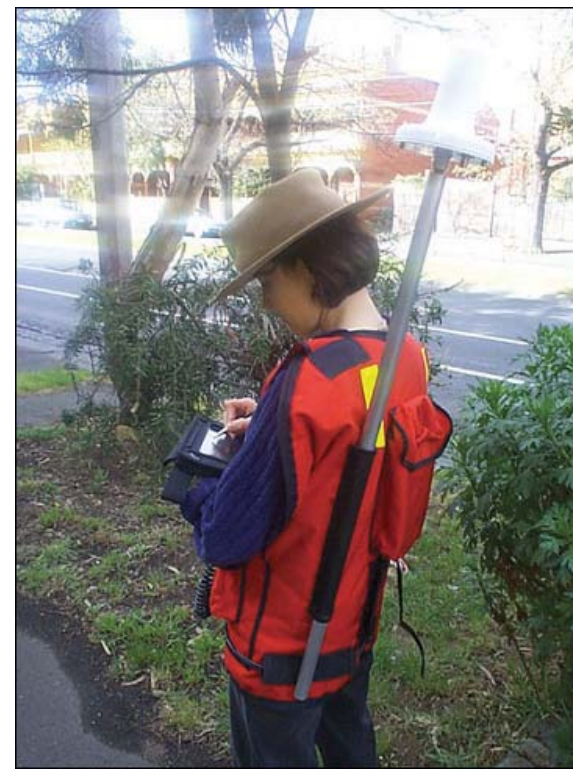


Figure 11.1 Using a simple GIS in the field to collect data. The device on the pole is a GPS antenna, used to georeference data as they are collected

Figure 11.2 Collecting survey data in the field, on behalf of a government statistical agency (Courtesy: Julie Dillemuth, João Hespanha, Stacy Rebich)

- The location where the data are processed, denoted by P. In Section 1.5.3 we introduced the concept of a GIService, a processing capability accessed at a remote site rather than provided locally by the user's desktop GIS.
- The area that is the focus of the GIS project, or the *subject* location, denoted by S. All GIS projects necessarily study some area, obtain data as a representation of the area, and apply GIS processes to those data.

In traditional GIS, three of these locations – U, D, and P – are the same, because the data and processing both occur at the user's desktop. The subject location could be anywhere in the world, depending on the project. But in distributed GIS there is no longer any need for D and P to be the same as U, and moreover it is possible for the user to be located in the subject area S, and able to see, touch, feel, and even smell it, rather than being in a distant office. The GIS might be held in the user's hand, or stuffed in a backpack, or mounted in a vehicle.

In distributed GIS the user location and the subject location can be the same.

Distributed GIS has many potential benefits, and these are discussed in detail in the following sections. Section 11.2 discusses the feasibility of distributing data, and the benefits of doing so, and introduces the technologies of distributed data access. Section 11.3 discusses the technologies of mobile computing, and the benefits of allowing the user to do GIS anywhere, and particularly within the subject area S. Section 11.4 reviews the status of distributed GIServices, and the benefits of obtaining processing from remote locations. It is important to distinguish at the outset between the *vision* of distributed GIS, and what is *practical* at this time, and so both perspectives are provided throughout the chapter.

Critical to distributed GIS are the standards and specifications that make it possible for devices, data, and processes to operate together, or *interoperate*. Some of these are universal, such as ASCII, the standard for

Biographical Box **11.1**

David Schell, proponent of open GIS

David Schell earned degrees in mathematics and literature from Brown University and the University of North Carolina. Initially, he worked in system development at IBM. Later, during the 1980s, David managed marketing and business development programs for various Massachusetts workstation startups. In these positions, he was actively involved in the development of Unix software libraries and the integration of productivity tools for office automation 'measurement and control' and scientific applications. In the late 1980s, while working at real-time computing pioneer Masscomp (later Concurrent Computer), David had the opportunity to work closely with the Corps of Engineers Geographic Resources Analysis Support System (GRASS) development team on technology transfer and the integration of GRASS open source with commercial GIS and imaging systems. This integration work resulted in an early prototype of the principles of 'Web mapping'. Carl Reed and John Davidson of Genasys developed an operational system in which a user could access both GenaMap and GRASS seamlessly from the same command line and where map graphics from both systems displayed into the same X-Window. This resulted in the development of some of the earliest interface-based interoperability models for geospatial processing.

Figure 11.3 David Schell, President, Open Geospatial Consortium

In 1992, with the aid of a Cooperative Research and Development Agreement with the US Army Corps of Engineers USACERL, David founded the Open GIS Foundation (OGF) to facilitate technology transfer of COE technology to the private sector as well as continuing the commercial interoperability initiatives begun with GenaMap. In 1994, responding to the need to engage geospatial technology users and providers in a formal consensus standards process, he reorganized OGF to become the Open GIS Consortium (now the Open Geospatial Consortium). With the support of both public and private sector sponsors, the OpenGIS Project began an intense and highly focused industry-wide consensus process to create an architecture to support interoperable geoprocessing. David was cited in 2003 as one of the 20 most influential innovators in the IT community by the editors of *CIO Magazine* and received the CIO 20/20 Award for visionary achievement.

Speaking about the industry's evolution, David Schell said: 'It is most important for us to realize the far-reaching cultural impact of geospatial interoperability, in particular its influence on the way we conceptualize the challenges of policy and management in the modern industrialized world. Without such a capability, we would still be wasting most of our creative energies laboring in the dark ages of time-consuming and laborious data conversion and hand-made application stove-pipes. Now it is possible for scientists and thinkers in every field to focus their energies on problems of real intellectual merit, instead of having to wrestle exhaustingly to reconcile the peculiarities of data imposed by diverse spatial constructs and vendor-limited software architectures. As our open interfaces make geospatial data and services discoverable and accessible in the context of the World Wide Web, geospatial information becomes truly more useful to more people and its potential for enabling progress and enlightenment in many domains of human activity can be more fully realized.'

coding of characters (Box 3.1), and XML, the extensible markup language that is used by many services on the Web. Others are specific to GIS, and many of these have been developed through the Open Geospatial Consortium (OGC, **www.opengeospatial.org**), an organization set up to promote openness and interoperability in GIS (see Box 11.1). Among the many successes of OGC over the past decade are the simple feature specification, which standardizes many of the terms associated with primitive elements of GIS databases (polygons, polylines, points, etc.; see also Chapter 9); Geography Markup Language (GML), a version of XML that handles geographic features and enables open-format communication of geographic data; and specifications for Web services (Web Map Service, Web Feature Service, Web Coverage Services) that allow users to request data automatically from remote servers.

Distributed GIS reinforces the notion that today's computer is not simply the device on the desk, with its hard drive, processor, and peripherals, but something more extended. The slogan 'The Network is the Computer' has provided a vision for at least one company – Sun Microsystems (**www.sun.com**) – and propels many major developments in computing. The term *cyberinfrastructure* describes a new approach to the conduct of science, relying on high-speed networks, massive processors, and distributed networks of sensors and data archives (**www.cise.nsf.gov/sci/reports/toc.cfm**). Efforts are being made to integrate the world's computers to provide the kinds of massive computing power that are needed by such projects as SETI (the Search for Extra-Terrestrial Intelligence, **www.seti.org**), which processes terabytes of data per day in the search for anomalies that might indicate life elsewhere in the universe, and makes use of computer power wherever it can find it on the Internet (Figure 11.4). *Grid* computing is a generic term for such a fully integrated worldwide network of computers and data.

Figure 11.4 The power of this home computer would be wasted at night when its owner is sleeping, so instead its power has been 'harvested' by a remote server and used to process signals from radio telescopes as part of a search for extra-terrestrial intelligence

11.2 Distributing the data

Since its popularization in the early 1990s the Internet has had a tremendous and far-reaching impact on

the accessibility of GIS data, and on the ability of GIS users to share datasets. As we saw in Chapter 9, a large and increasing number of websites offer GIS data, for free, for sale, or for temporary use, and also provide services that allow users to search for datasets satisfying certain requirements. In effect, we have gone in a period of little more than ten years from a situation in which geographic data were available only in the form of printed maps from map libraries and retailers, to one in which petabytes (Table 1.1) of information are available for download and use at electronic speed (about 1.5 million CDs would be required to store 1 petabyte). For example, the NASA-sponsored EOSDIS (Earth Observing System Data and Information System; **spsosun.gsfc.nasa.gov/New_EOSDIS.html**) archives and distributes the geographic data from the EOS series of satellites, acquiring new data at over a terabyte per day, with an accumulated total of more than 1 petabyte at this site alone.

Some GIS archives contain petabytes of data.

The vision of distributed GIS goes well beyond the ability to access and retrieve remotely located data, however, because it includes the concepts of *search*, *discovery*, and *assessment*: in the world of distributed GIS, how do users search for data, discover their existence at remote sites, and assess their fitness for use? Three concepts are important in this respect: *object-level metadata*, *geolibraries*, and *collection-level metadata*.

11.2.1 Object-level metadata

Strictly defined, metadata are data about data, and *object-level* metadata (OLM) describe the contents of a single dataset by providing essential documentation. We need information about data for many purposes, and OLM try to satisfy them all. First, we need OLM to automate the process of search and discovery over archives. In that sense OLM are similar to a library's catalog, which organizes the library's contents by author, title, and subject, and makes it easy for a user to find a book. But OLM are potentially much more powerful, because a computer is more versatile than the traditional catalog in its potential for re-sorting items by a large number of properties, going well beyond author, title, and subject, and including geographic location. Second, we need OLM to determine whether a dataset, once discovered, will satisfy the user's requirements – in other words, to assess the fitness of a dataset for a given use. Does it have sufficient spatial resolution, and acceptable quality? Such metadata may include comments provided by others who tried to use the data, or contact information for such previous users (users often comment that the most useful item of metadata is the phone number of the person who last tried to use the data). Third, OLM must provide the information needed to handle the dataset effectively. This may include technical specifications of format, or the names of software packages that are compatible with the data, along with information about the dataset's location, and its volume. Finally, OLM may provide useful information on the dataset's contents. In the case of remotely sensed images, this may include the percentage of cloud obscuring the scene, or whether the scene contains particularly useful instances of specific phenomena, such as hurricanes.

Object-level metadata are formal descriptions of datasets that satisfy many different requirements.

OLM generalize and abstract the contents of datasets, and therefore we would expect that they would be smaller in volume than the data they describe. In reality, however, it is easy for the complete description of a dataset to generate a greater volume of information than the actual contents. OLM are also expensive to generate, because they represent a level of understanding of the data that is difficult to assemble, and require a high level of professional expertise. Generation of OLM for a geographic dataset can easily take much longer than it takes to catalog a book, particularly if it has to deal with technical issues such as the precise geographic coverage of the dataset, its projection and datum details (Chapter 5), and other properties that may not be easily accessible. Thus the cost of OLM generation, and the incentives that motivate people to provide OLM, are important issues.

For metadata to be useful, it is essential that they follow widely accepted standards. If two users are to be able to share a dataset, they must both understand the rules used to create its OLM, so that the custodian of the dataset can first create the description, and so that the potential user can understand it. The most widely used standard for OLM is the US Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata, or CSDGM, first published in 1993 and now the basis for many other standards worldwide. Box 11.2 lists some of its major features. As a *content standard* CSDGM describes the items that should be in an OLM archive, but does not prescribe exactly how they should be formatted or structured. This allows developers to implement the standard in ways that suit their own software environments, but guarantees that one implementation will be understandable to another – in other words, that the implementations will be *interoperable*. For example, ESRI's ArcGIS provides two formats for OLM, one using the widely recognized XML standard and the other using ESRI's own format.

CSGDM was devised as a system for describing geographic datasets, and most of its elements make sense only for data that are accurately georeferenced and represent the spatial variation of phenomena over the Earth's surface. As such, its designers did not attempt to place CSGDM within any wider framework. But in the past decade a number of more broadly based efforts have also been directed at the metadata problem, and at the extension of traditional library cataloging in ways that make sense in the evolving world of digital technology.

One of the best known of these is the Dublin Core (see Box 11.3), the outcome of an effort to find the minimum set of properties needed to support search and discovery for datasets in general, not only geographic datasets. Dublin Core treats both space and time as instances of a single property, coverage, and unlike CSGDM

Technical Box **11.2**

Major features of the US Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata

1. Identification Information – basic information about the dataset.
2. Data Quality Information – a general assessment of the quality of the dataset.
3. Spatial Data Organization Information – the mechanism used to represent spatial information in the dataset.
4. Spatial Reference Information – the description of the reference frame for, and the means to encode, coordinates in the dataset.
5. Entity and Attribute Information – details about the information content of the dataset, including the entity types, their attributes, and the domains from which attribute values may be assigned.
6. Distribution Information – information about the distributor of and options for obtaining the dataset.
7. Metadata Reference Information – information on the currentness of the metadata information, and the responsible party.
8. Citation Information – the recommended reference to be used for the dataset.
9. Time Period Information – information about the date and time of an event.
10. Contact Information – identity of, and means to communicate with, person(s) and organization(s) associated with the dataset.

Technical Box **11.3**

The fifteen basic elements of the Dublin Core metadata standard

1. TITLE. The name given to the resource by the CREATOR or PUBLISHER.
2. AUTHOR or CREATOR. The person(s) or organization(s) primarily responsible for the intellectual content of the resource.
3. SUBJECT or KEYWORDS. The topic of the resource, or keywords, phrases, or classification descriptors that describe the subject or content of the resource.
4. DESCRIPTION. A textual description of the content of the resource, including abstracts in the case of document-like objects or content description in the case of visual resources.
5. PUBLISHER. The entity responsible for making the resource available in its present form, such as a publisher, a university department, or a corporate entity.
6. OTHER CONTRIBUTORS. Person(s) or organization(s) in addition to those specified in the CREATOR element who have made significant intellectual contributions to the resource, but whose contribution is secondary to the individuals or entities specified in the CREATOR element.
7. DATE. The date the resource was made available in its present form.
8. RESOURCE TYPE. The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary.
9. FORMAT. The data representation of the resource, such as text/html, ASCII, Postscript file, executable application, or JPEG image.
10. RESOURCE IDENTIFIER. String or number used to uniquely identify the resource.
11. SOURCE. The work, either print or electronic, from which this resource is delivered, if applicable.
12. LANGUAGE. Language(s) of the intellectual content of the resource.
13. RELATION. Relationship to other resources.
14. COVERAGE. The spatial locations and temporal duration characteristic of the resource.
15. RIGHTS MANAGEMENT. The content of this element is intended to be a link (a URL or other suitable URI as appropriate) to a copyright notice, a rights-management statement, or perhaps a server that would provide such information in a dynamic way.

does not lay down how such specific properties as spatial resolution, accuracy, projection, or datum should be described.

The principle of establishing a minimum set of properties is sharply distinct from the design of CSGDM, which was oriented more toward the capture of all knowable and potentially important properties of geographic datasets. Of direct relevance here is the problem of cost, and specifically the cost of capturing a full CSGDM metadata record. While many organizations have wanted to make their data more widely available, and have been driven to create OLM for their datasets, the cost of determining the full set of CSDGM elements is often highly discouraging. There is interest therefore in a concept of *light metadata*, a limited set of properties that is both comparatively cheap to capture, and still useful to support search and discovery. Dublin Core represents this approach, and thus sits at the opposite end of a spectrum from CSGDM. Every organization must somehow determine where its needs lie on this spectrum, which ranges from light and cheap to heavy and expensive.

Light, or stripped-down, OLM provide a short but useful description of a dataset that is cheaper to create.

11.2.2 Geolibraries

The use of digital technology to support search and discovery opens up many options that were not available in the earlier world of library catalogs and bookshelves. Books must be placed in a library on a permanent basis, and there is no possibility of reordering their sequence – but in a digital catalog it is possible to reorder the sequence of holdings in a collection almost instantaneously. So while a library's shelves are traditionally sorted by subject, it would be possible to re-sort them digitally by author name or title, or by any property in the OLM catalog. Similarly the traditional card catalog allowed only three properties to be sorted – author, title, and subject – and discouraged sorting by multiple subjects. But the digital catalog can support any number of subjects.

Of particular relevance to GIS users is the possibility of sorting a collection by the coverage properties: location and time. Both the spatial and temporal dimensions are continuous, so it is impossible to capture them in a single property analogous to author that can then be sorted numerically or alphabetically. But in a digital system this is not a serious problem, and it is straightforward to capture the coverage of a dataset, and to allow the user to search for datasets that cover an area or time of interest defined by the user. Moreover, the properties of location and time are not limited to geographic datasets, since there are many types of information that are associated with specific areas on the Earth's surface, or with specific time periods. Searching based on location or time would enable users to find information about any place on the Earth's surface, or any time period – and to find reports, photographs, or even pieces of music, as long as they possessed geographical and temporal *footprints*.

The term *geolibrary* has been coined to describe digital libraries that can be searched for information about any user-defined geographic location. A US National Research Council report (see Further Reading) describes the concept and its implementation, and many instances of geolibraries can be found on the Web (see Box 11.4). Digital Earth is a vision of a geolibrary in which the user interacts with a visual representation of the surface of the Earth, rather than with a digital representation of the traditional card catalog.

A geolibrary can be searched for information about a specific geographic location.

11.2.3 Collection-level metadata

Thus far we have seen how metadata can be useful in the description of individual datasets. There are several other forms of metadata of potential interest to GIS users, and this section describes the issues associated with one of them – metadata that describe entire collections of datasets, rather than individual datasets.

Many different collections or archives of geographic datasets exist on the Internet, and Box 11.4 and Table 9.4 identify several examples. The Geospatial One-Stop is structured as a mechanism that allows custodians to make their data visible to users who access the portal, and that allows users to search across multiple servers using a simple protocol, assess the appropriateness of datasets, and access them through their own GIS. The US Geological Survey supports several major collections of geographic data, notably at the EROS Data Center in Sioux Falls, South Dakota (**edcwww.cr.usgs.gov**), and major commercial collections of data can be found at the sites of Space Imaging (**www.spaceimaging.com**), Terraserver (**www.terraserver.com**), GlobeXplorer (**www.globexplorer.com**), and at many other locations. Clearinghouses are also sponsored by many other countries and by several US states.

But this abundance of readily accessible data presents a problem – how does the user know which collection to search for a given dataset? Although the Geospatial One-Stop hopes to be a single portal for a large proportion of geographic data, its emphasis is on the US, and no site, however successful, will ever have a monopoly of all data for all parts of the world. The term *collection-level metadata* (CLM) defines the information needed to make an intelligent choice, based on general knowledge of each collection's contents. For example, the EROS Data Center is a good site to search for images and maps of the US. Microsoft's Terraserver (**terraserver.microsoft.com**) is a good source for imagery of any part of the world. In looking for maps of a specific country, it makes sense to try to find a site sponsored by that country's national mapping agency, and similarly, information about a state may exist in a clearinghouse sponsored by that state.

Collection-level metadata describe the entire collection available through a single website.

Unfortunately there has been little consistent effort to develop standards for CLM, unlike its individual dataset counterpart (OLM). Knowledge of CLM is still

Technical Box **11.4**

Exemplar geolibraries

The Alexandria Digital Library (**www.alexandria.ucsb.edu**) provides access to over 2 million maps and images using a simple method of search based on user interaction with a world map, placenames, or coordinates. Many of the maps and images are stored in the Map and Imagery Laboratory at the University of California, Santa Barbara, but the Alexandria catalog also identifies materials in other collections. Figure 11.5 shows a screen shot of the library's Web interface.

The US Geospatial One-Stop (**www.geo-one-stop.gov**; Figure 11.6) is intended to provide a single portal (a *geoportal*) to vast and distributed resources of geographic data. Managed by the Department of the Interior, it presents a catalog of datasets residing in federal, state, local, tribal, and private archives, and allows users to search across these resources independently of their storage locations. Its benefits include reduced duplication, by encouraging users to share the data of others rather than creating them themselves, and increased collaboration between agencies. Users can volunteer their own datasets by registering ('publishing') them to the portal. Geoportals can also act as directories to remotely accessible GIServices (Section 11.4).

(A)

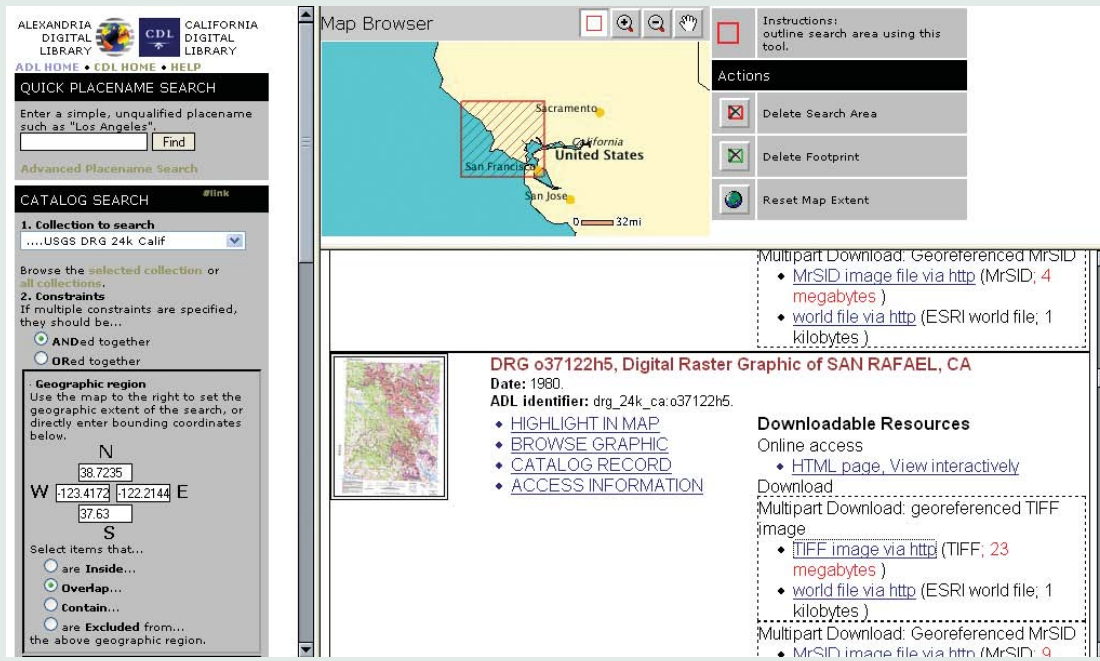


Figure 11.5 Example of a geolibrary: the Alexandria Digital Library (**alexandria.ucsb.edu**). (A) shows the interface to the library in a standard Netscape browser. The user has selected an area of California to search, and the system has returned the first 50 items in the library whose metadata match the search area footprint and the user's request for scanned images of 1:24 000 topographic maps. (B) shows one of the 'hits', part of the San Rafael quadrangle centered on Muir Woods National Monument

(B)

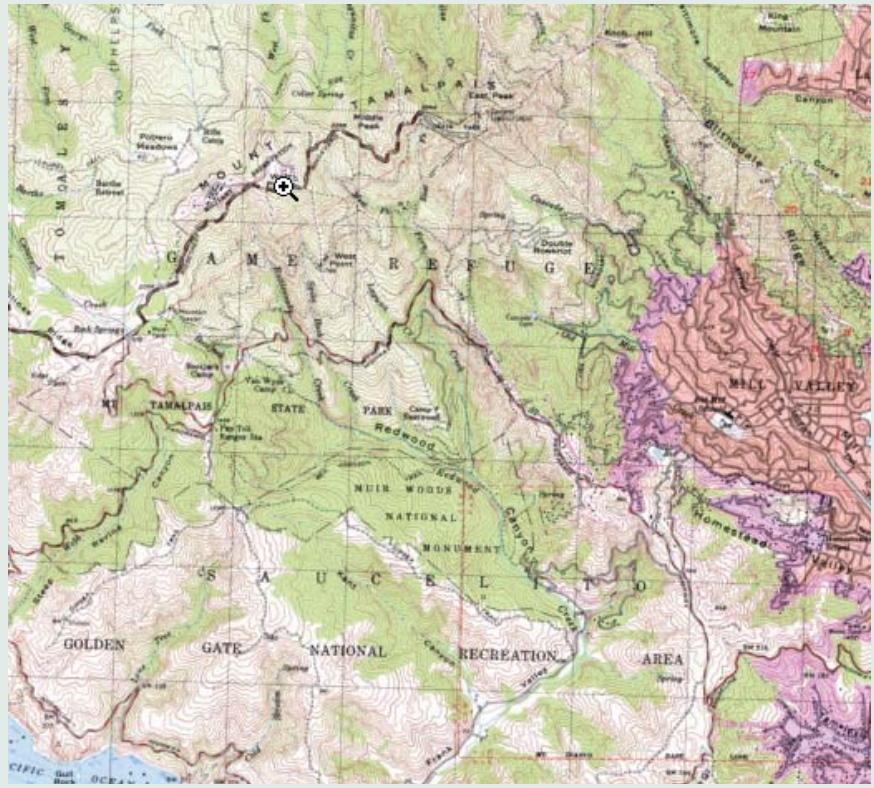


Figure 11.5 (*continued*)

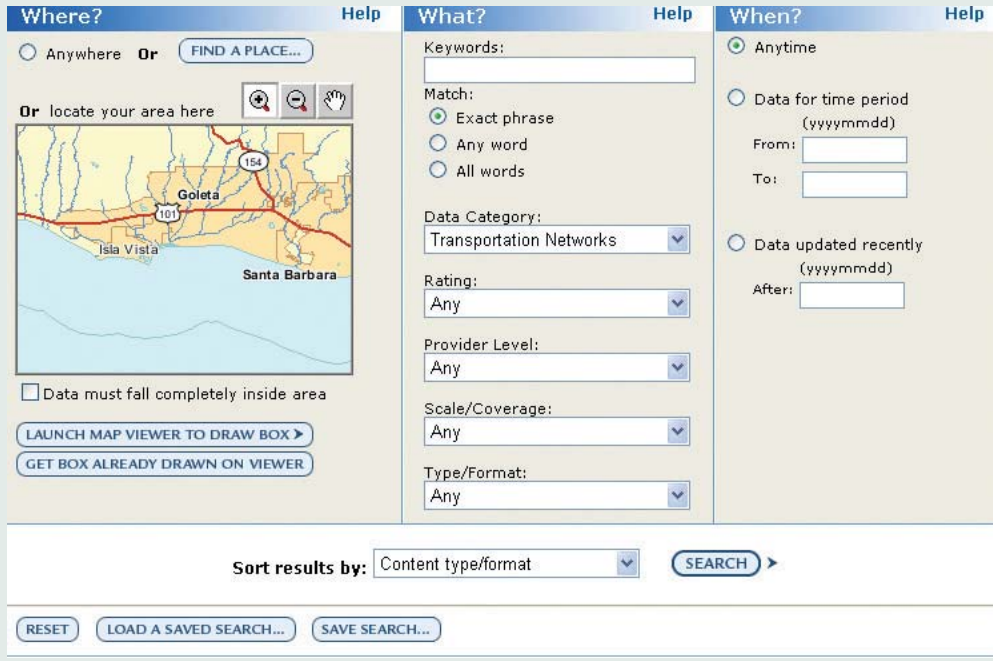


Figure 11.6 The US Department of the Interior's Geospatial One-Stop, another example of a geolibrary. In this case the user has zoomed in to the area of Goleta, California, and requested data on transportation networks. A total of 66 'hits' were identified in this search, indicating 66 possible sources of suitable information

largely unorganized, and tends to accumulate slowly in the minds of geographic information specialists (or SAPs, see Section 1.4.3.2). Knowing where to look is still largely a matter of personal knowledge and luck.

Various possible solutions to this problem can be identified. In the future, we may find the means to develop a new generation of Web search engines that are able to discover geographic datasets automatically, and catalog their OLM. Then a user needing information on a particular area could simply access the search engine's results, and be provided with a list of appropriate collections. That capability is well beyond the reach of today's generation of search engines, although several format standards, including GeoTIFF, already add the tags to geographic datasets that clever search engines could recognize. Alternatively, some scheme might be developed that authorizes a uniform system of servers, such that each server's contents are defined precisely, using geographic, thematic, or other criteria. In such a system all data about locations within the server's nominal coverage area, such as a state or a range of latitude and longitude, would be found on that server and no other. Such a system would need to be hierarchical, with very detailed data in collections with small coverage areas, and coarser data in collections with larger coverage areas. But at this time the CLM problem remains a major impediment to effective sharing of geographic data.

11.3 The mobile user

Computing has become so much a part of our lives that for many people it is difficult to imagine life without it. Increasingly, we need computers to shop, to communicate with friends, to obtain the latest news, and to entertain ourselves. In the early days, the only place one could compute was in a computing center, within a few meters of the central processor. Computing had extended to the office by the 1970s, and to the home by the 1990s. The portable computers of the 1980s opened the possibility of computing in the garden, at the beach, or 'on the road' in airports and airplanes. Wireless communication services such as WiFi (the wireless access technology based on the 802.11 family of standards) now allow broadband communication to the Internet from 'hot-spots' in many hotels, restaurants, airports, office buildings, and private homes (Figure 11.7). The range of mobile computing devices is also multiplying rapidly, from the relatively cumbersome but powerful laptop weighing several kg to the PDA, tablet computer, and the cellphone weighing a hundred or so grams. Within a few years we will likely see the convergence of these devices into a single, low-weight, and powerful mobile personal device that acts as computer, storage device, and cellphone.

In many ways the ultimate end point of this progression is the *wearable* computer, a device that is fully embedded

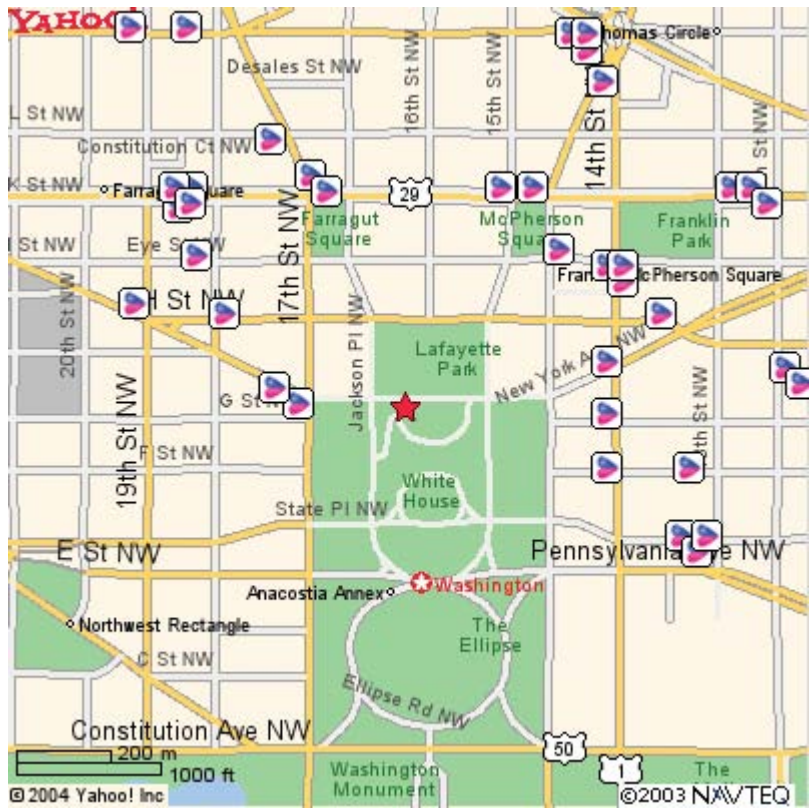


Figure 11.7 Map of WiFi (802.11) wireless broadband 'hotspots' within 1 mile of the White House (1600 Pennsylvania Ave, Washington, DC) and using the T-Mobile Internet provider

Figure 11.8 A wearable computer in use. The outfit consists of a processor and storage unit hung on the user's waist belt; an output unit clipped to the eyeglasses with a screen approximately 1 cm across and VGA resolution; an input device in the hand; and a GPS antenna on the shoulder. The batteries are in a jacket pocket (Courtesy: Keith Clarke)

in the user's clothing, goes everywhere, and provides ubiquitous computing service. Such devices are already obtainable, though in fairly cumbersome and expensive form. They include a small box worn on the belt and containing the processor and storage, and an output display clipped to the user's eyeglasses. Figure 11.8 shows such a device in use.

Computing has moved from the computer center and is now close to, even inside, the human body. The ultimate form of mobile computing is the wearable computer.

What possibilities do such systems create for GIS? Of greatest interest is the case where the user is situated in the subject location, that is, *U* is contained within *S*, and the GIS is used to analyze the immediate surroundings. A convenient way to think about the possibilities is by comparing *virtual* reality with *augmented* reality.

11.3.1 Virtual reality and augmented reality

One of the great strengths of GIS is the window it provides on the world. A researcher in an office in Nairobi, Kenya might use GIS to obtain and analyze data on the effects of salinization in the Murray Basin of Australia, combining images from satellites with base topographic data, data on roads, soils, and population distributions, all obtained from different sources via the Internet. In doing so the researcher would build up a comprehensive picture of a part of the world he or she might never have visited, all through the medium of digital geographic data and GIS, and might learn almost as much through this GIS window as from actually being there. The expense and time of traveling to Australia would be avoided, and the analysis could proceed almost instantaneously. Some aspects of the study area would be missing, of course – aspects of culture, for example, that can best be experienced by meeting with the local people.

Research environments such as this are termed *virtual realities*, because they replace what humans normally gather through their senses – sight, sound, touch, smell, and taste – by presenting information from a database. In most GIS applications only one of the senses, sight, is used to create this virtual reality, or VR. In principle, it is possible to record sounds and store them in GIS as attributes of features, but in practice very little use is made of any sensory channel other than vision. Moreover, in most GIS applications the view presented to the user is the view from *above*, even though our experience at looking at the world from this perspective is limited (for most of us, to times when we requested a window seat in an airplane). GIS has been criticized for what has been termed the *God's eye view* by some writers, on the basis that it distances the researcher from the real conditions experienced by people on the ground.

Virtual environments attempt to place the user in distant locations.

More elaborate VR systems are capable of *immersing* the user, by presenting the contents of a database in a three-dimensional environment, using special eyeglasses or by projecting information onto walls surrounding the user, and effectively *transporting* the user into the environment represented in the database. Virtual London (Box 13.7 and see **www.casa.ucl.ac.uk/research/virtuallondon.htm**) and the Virtual Field Course (**www.geog.le.ac.uk/vfc/**) are examples of projects to build databases to support near-immersion in particular environments. Some of the most interesting of these are projects that recreate historic environments, such as those of the Cultural VR Lab at the University of California, Los Angeles (**www.cvrlab.org**), which supports roaming through three-dimensional visualizations of the Roman church of Santa Maria Maggiore (Figure 11.9) and other classical structures.

Roger Downs, Professor of Geography at Pennsylvania State University, uses Johannes Vermeer's famous painting *The Geographer* (Figure 11.10) to make an important point about virtual realities. On the table in front of the figure is a map, representing the geographer's window on a part of the world that happens to be of interest. But the subject figure is shown looking out of the window, at the real world, perhaps because he needs the information he derives from his senses to understand the world as shown on the map. The idea of combining information from a database with information derived directly through the senses is termed *augmented reality*, or AR. In terms of the locations of computing discussed earlier, AR is clearly of most value when the location of the user *U* is contained within the subject area *S*, allowing the user to augment what can be seen directly with information retrieved about the same area from a database. This might include historic information, or predictions about

Figure 11.9 Image generated from a three-dimensional digital reconstruction of the Basilica of Santa Maria Maggiore by the Virtual Reality Laboratory of the University of California, Los Angeles (**www.cvrlab.org**). The basilica was built in Rome around 440 AD and has long since been destroyed (Reproduced by permission of Regents of the University of California

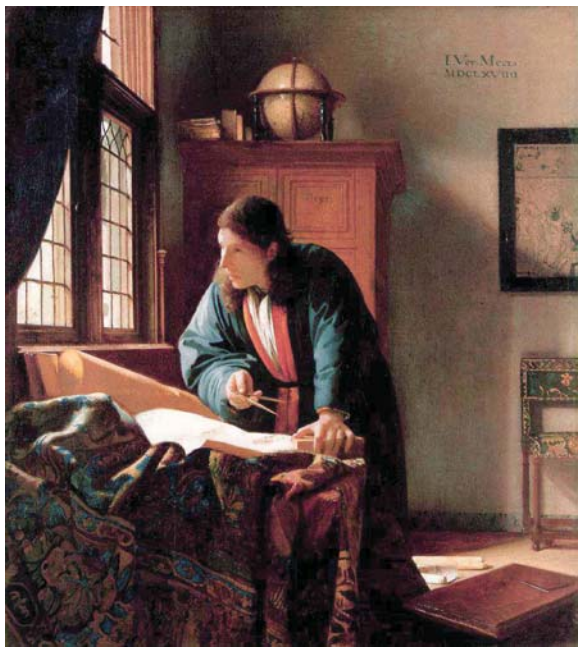


Figure 11.10 Johannes Vermeer's painting *The Geographer* from 1669

the future, or information that is for some other reason invisible to the user.

AR can be used to augment or replace the work of senses that are impaired or absent. A team led by Reginald Golledge, Professor of Geography at the University of California, Santa Barbara (see Box 1.8), has been experimenting with AR as a means of helping visually impaired people to perform the simple task of navigation. The system, which is worn by the user (Figure 11.11), includes a differential GPS for accurate positioning, a GIS database that includes very detailed information on the immediate environment, a compass to determine the position of the user's head, and a pair of earphones. The user gives the system information to identify the desired destination, and the system then generates sufficient information to replace the normal role of sight. This might be through verbal instructions, or by generating stereo sounds that appear to come from the appropriate direction.

Augmented reality combines information from the database with information from the senses.

Steven Feiner, Professor of Computer Science at Columbia University, has demonstrated another form of AR that superimposes historic images and other information directly on the user's field of view. For example, in Figures 11.12 and 11.13, a user wearing a head-mounted device coupled to a wearable computer is seeing both the Columbia University main library building, and also an image generated from a database showing the building that occupied the library's position prior to the University's move to this site in 1896: the Bloomingdale Insane Asylum.

Figure 11.11 The system worn here by Reg Golledge (a leader of the development team) uses GIS and GPS to augment the senses of a visually impaired person navigating through a complex space such as a university campus (Reproduced by permission of Reg Golledge)

Figure 11.12 An augmented reality system developed at Columbia University by Prof. Steve Feiner and his group © 1999, T. Höllerer, S. Feiner & J. Pavlik, Computer Graphics & User Interfaces Lab., Columbia University. Reproduced with permission

11.3.2 Location-based services

One of the four big trends in software is location-based applications.
Bill Gates, Wireless 2000 Conference, March 2000

A location-based service (LBS) is defined as an information service provided by a device *that knows where it is* and is capable of *modifying* the information it provides based on that knowledge. Traditional computing devices, such as desktops or laptops, have no way of knowing where they are, and their functions are in no way changed when they are moved. But increasingly the essential information about a device's location is available, and is being used for a wide range of purposes, some beneficial and some distinctly questionable (see Box 5.2).

A location-based service is provided by a computing device that knows where it is.

The simplest and most obvious form of locationally enabled device is the GPS receiver, and any computer that includes a GPS capability, such as a laptop or PDA with an added PCMCIA card or chip, is capable of providing LBS. But the most ubiquitous LBS-capable device is the modern cellphone. A variety of methods exist for determining cellphone locations, including GPS embedded in the cellphone itself, measurements made by the cellphone of signals received from cellphone towers, and measurements made by the towers of signals received

Figure 11.13 The field of view of the user of Feiner's AR system, showing the Columbia University main library and the insane asylum that occupied the site in the 1800s © 1999, T. Höllerer, S. Feiner & J. Pavlik, Computer Graphics & User Interfaces Lab., Columbia University. Reproduced with permission

from the phone. As long as the phone is on, the operator of the cellphone network is able to pinpoint its location at least to the accuracy represented by the size of the cell, of the order of 10 km, and frequently much more accurately. Within a few years it is likely that the locations of the vast majority of active phones will be known to an accuracy of 10 m.

One of the strongest motives driving this process is emergency response. A large and growing proportion of emergency calls come from cellphones, and while the location of each land-line phone is likely to be recorded in a database available to the emergency responder, in a significant proportion of cases the user of a cellphone is unable to report his or her current location to sufficient accuracy to enable effective response. Several well-publicized cases have drawn attention to the problem. The magazine *Popular Science*, for example, reported a case of a woman who lost control of her car in Florida in February 2001, skidding into a canal. Although she called 911 (the emergency number standard in the US and Canada), she was unable to report her location accurately and died before the car was found.

One solution to the 911 problem is to install GPS in the vehicle, communicating location directly to the dispatcher. The Onstar system (**www.onstar.com**) is one such system, combining a GPS device and wireless communication. The system has been offered in the US to Cadillac purchasers for several years, and provides various advisory services in addition to emergency response, such as advice on local attractions and directions to local services. When the vehicle's airbags inflate, indicating a likely accident, the system automatically measures and radios its location to the dispatcher, who relays the necessary information to the emergency services.

Emergency services provide one of the strongest motivations for LBS.

There are many other examples of LBS that take advantage of locationally enabled cellphones. A *yellow-page* service responds to a user who requests information on businesses that are close to his or her current location (Where is the nearest pizza restaurant? Where is the nearest hospital? Where is the nearest WiFi hotspot?) by sending a request that includes the location to a suitable Web server. The response might consist of an ordered list presented on the cellphone screen, or a simple map centered on the user's current location (Figure 11.7). A *trip planner* gives the user the ability to find an optimum driving route from the current location to some defined destination. Similar services are now being provided by public transport operators, and in some cases these services make use of GPS transponders on buses and trains to provide information on actual, as distinct from scheduled, arrival and departure times. AT&T's Find Friends service allows a cellphone user to display a map showing the current locations of nearby friends (provided their cellphones are active and they are also registered for this service). Undercover (**www.playundercover.com**) is an example of a *location-based game* that involves the actual locations of players (see Box 11.5) moving around a real environment.

Direct determination of location, using GPS or measurement to or from towers, is only one basis on which a computing device might know its location, however. Other forms of LBS are provided by fixed devices, and rely on the determination of location when the device was installed. For example, many *point-of-sale* systems that are used by retailers record the location of the sale, combining it with other information about the buyer obtained by accessing credit-card or store-affinity-card records (Figure 11.18). In exchange for the convenience

Biographical Box **11.5**

Antonio Câmara, designer of location-based services

Antonio Câmara (Figure 11.14) is the CEO and co-founder of YDreams (**www.ydreams.com**), a company specializing in location-based games and novel applications of computer graphics. A Professor at the New University of Lisbon in Portugal, he has been involved in GIS since 1977. Antonio Câmara was a member of the team that developed SNIG, the first spatial data infrastructure on the Internet, in 1995.

The mobile-phone-based entertainment market is expected to reach between 8 and 25 billion Euros a year in revenues by 2008. Games where the player's location is known are becoming a prominent component of that market. Undercover (**www.playundercover.com**), developed by YDreams, was the first location-based game with a visual interface. Undercover is a multi-player game where players can act solely or in a group. The purpose is to eliminate terrorists, which are either virtual or result from an infection acquired by a player. The positioning of players and terrorists can be seen on the mobile-phone screen (Figure 11.15).

Figure 11.14 Antonio Câmara, CEO and co-founder of YDreams, a company specializing in location-based games (Reproduced by permission of Antonio Câmara)

Figure 11.15 Cellphone display for a user playing Undercover: scanning for friends (blue) and foes (red) (Reproduced by permission of Antonio Câmara)

Figure 11.16 The Lex Ferrum location-based game developed by YDreams relies on the display of three-dimensional characters on the cellphone screen (Reproduced by permission of Antonio Câmara)

Figure 11.17 A Formula 1 race played on a large screen using mobile phones. The project was developed by IDEO and YDreams for Vodafone (Reproduced by permission of Antonio Câmara)

In the first version, Undercover relied on simplified maps where only main points of interest were considered. Positioning was based on Cell-ID techniques provided by the operator. Undercover's second version will use city maps in the background. Positioning may use a range of systems including GPS units communicating via Bluetooth with the mobile phones. YDreams has also developed a proximity-based location game called Lex Ferrum (Figure 11.16) for the NOKIA N-Gage. In this game, the player's location is scanned via Bluetooth or GPRS/UMTS. Local combat is done using Bluetooth. Remote combat may use the cellular network.

Such games will also benefit from the widespread deployment of large screens in cities. Mobile phones can then be used as controllers of such screens. Imagine a user visualizing a three-dimensional representation of the city, querying the underlying GIS and playing games such as an imaginary Formula 1 race (Figure 11.17).

Figure 11.18 Credit card in use (Reproduced with permission of PhotoDisc, Inc.)

of a credit card, the user effectively surrenders some degree of location privacy to the company whenever the card is used. One benefit is that it is possible for the company to analyze transactions, looking for patterns of purchase that are outside the card user's normal buying habits, perhaps because they occur in locations that the user does not normally frequent, or at anomalous times. Many of us will have experienced the embarrassment of having a credit card transaction refused in an unfamiliar city, because the techniques used by the company have flagged the transaction as an indicator the card might have been stolen. In principle, a store-affinity card gives the company access to information about buying habits and locations, in return for a modest discount.

Location is revealed every time a credit, debit, or store affinity card is used.

Another technique relies on the locations given when an Internet IP address is registered, augmented by location information that may appear on Web pages. Companies such as InfoSplit (**www.infosplit.com**; Figure 11.19) specialize in gathering and providing lists of such locations to customers. A search-engine company such as Google would be interested in knowing the locations of computers that send it queries, because it would then be possible to sort responses by distance from the computer – ranking in this way would be of commercial interest to retailers, for example, because it would make their sites more conspicuous to local customers.

11.3.3 Issues in mobile GIS

GIS in the field or 'on the road' is very different from GIS in the office. First, the location of the user is important, and directly relevant to the application. It makes good sense to center maps on the user's location, to provide the capability for maps that show the view from the user's location rather than from above, and to offer maps that are oriented to the user's direction of travel, rather than north. Second, the field environment may make certain kinds of interaction impractical, or less desirable. In a moving vehicle, for example, it would be dangerous to present the driver with visual displays, unless perhaps these are directly superimposed on the field of view (on the windshield) – instead, such systems often provide instructions through computer-generated speech, and use speech recognition to receive instructions. With wearable devices that provide output on minute screens attached to the user's eyeglasses there is no prospect of conventional point-and-click with a mouse, so again voice communication may be more appropriate. On the other hand, many environments are noisy, creating problems for voice recognition.

One of the most important limitations to mobility remains the battery, and although great strides have been made in recent years, battery technology has not advanced as rapidly as other components of mobile systems, such as processors and storage devices. Batteries typically account for the majority of a mobile system's weight, and limit its operating time.

The battery remains the major limitation to LBS and to mobile computing in general.

Figure 11.19 Companies such as InfoSplit (**www.infosplit.com**) specialize in determining the locations of computers from their IP addresses. In this case the server has correctly determined that the 'hit' comes from Santa Barbara, California

Although broad band wireless communication is possible using WiFi, connectivity remains a major issue. Wireless communication techniques tend to be:

- Limited in spatial coverage. WiFi hotspots are limited to a single building or perhaps a hundred meters from the router; cellphone-based techniques have wider coverage but lower bandwidth (communication speeds), and only the comparatively slow satellite-based systems approach global coverage.
- Noisy. Cellphone and WiFi communications tend to 'break up' at the edges of coverage areas, or as devices move between cells, leading to errors in communication.
- Limited in temporal coverage. A moving device is likely to lose signal from time to time, while some devices, particularly recorders installed in commercial vehicles and PDAs used for surveys, are unable to upload data until within range of a home system.
- Insecure. Wireless communications often lack adequate security to prevent unwanted intrusion. It is comparatively easy, for example, for someone to tap into a wireless session and obtain sensitive information.

Progress is being made on all of these fronts, but it will be some years before it becomes possible to 'do GIS' as efficiently, effectively, and safely anywhere in the field as it currently is in the office.

11.4 Distributing the software: GIServices

This final section addresses distributed processing, the notion that the actual operations of GIS might be provided from remote sites, rather than by the user's own computer. Despite the move to component-based software (Section 7.3.3), it is still true that almost all of the operations performed on a user's data are executed in the same computer, the one on the user's desk. Each copy of a popular GIS includes the same functions, which are replicated in every installation of that GIS around the world. When new versions are released, they must be copied and installed at each site, and because not all copies are replaced there is no guarantee that the GIS used by one user is identical to the GIS used by another, even though the vendor and product name are the same.

A GIService is defined as a program executed at a remote site that performs some specific GIS task (see also Section 7.2). The execution of the program is initiated remotely by the user, who may have supplied data, or may rely on data provided by the service, or both. A simple example of a GIService is that provided by wayfinding sites like MapQuest (**www.mapquest.com**) or Yell.com (see Section 1.4.3.1). The user's current location and desired destination are provided to the service (the

current location may be entered by the user, or provided automatically from GPS), but the data representing the travel network are provided by the GIService. The results are obtained by solving for the shortest path between the origin and the destination (often a compromise between minimizing distance and minimizing expected travel time), a function that exists in many GIS but in this case is provided remotely by the GIService. Finally, the results are returned to the user in the form of driving instructions, a map, or both.

A GIService replaces a local GIS function with one provided remotely by a server.

In principle, any GIS function could be provided in this way, based on GIS server software (Section 7.6.2). In practice, however, certain functions tend to have attracted more attention than others. One obvious problem is commercial: how would a GIService pay for itself, would it charge for each transaction, and how would this compare to the normal sources of income for GIS vendors based on software sales? Some services are offered free, and generate their revenue by sales of advertising space or by offering the service as an add-on to some other service – MapQuest and Yell are good examples, generating much of their revenue through direct advertising and through embedding their service in hotel and other sites. In general, however, the characteristics that make a GIS function suitable for offering as a service appear to be:

- Reliance on a database that must be updated frequently, and is too expensive for the average user to acquire. Both geocoding (Box 5.3) and wayfinding services fall into this category, as do gazetteer services (Section 5.9).
- Reliance on GIS operations that are complex and can be performed better by a specialized service than by a generic GIS.

The number of available GIServices is growing steadily, creating a need for directories, portals, and other mechanisms to help users find and access them, and standards and protocols for interacting with them. The Geography Network (**www.geographynetwork.com**; Figure 11.20) provides such a directory, in addition to its role as a geolibrary, as does the Geospatial One-Stop (Box 11.4). Generic standards for remote services are emerging, such as WSDL (Web Services Definition Language, **www.w3.org/TR/wsdl**) and UDDI (Universal Description, Discovery, and Integration, **www.uddi.org**), though to date these have found little application in GIS.

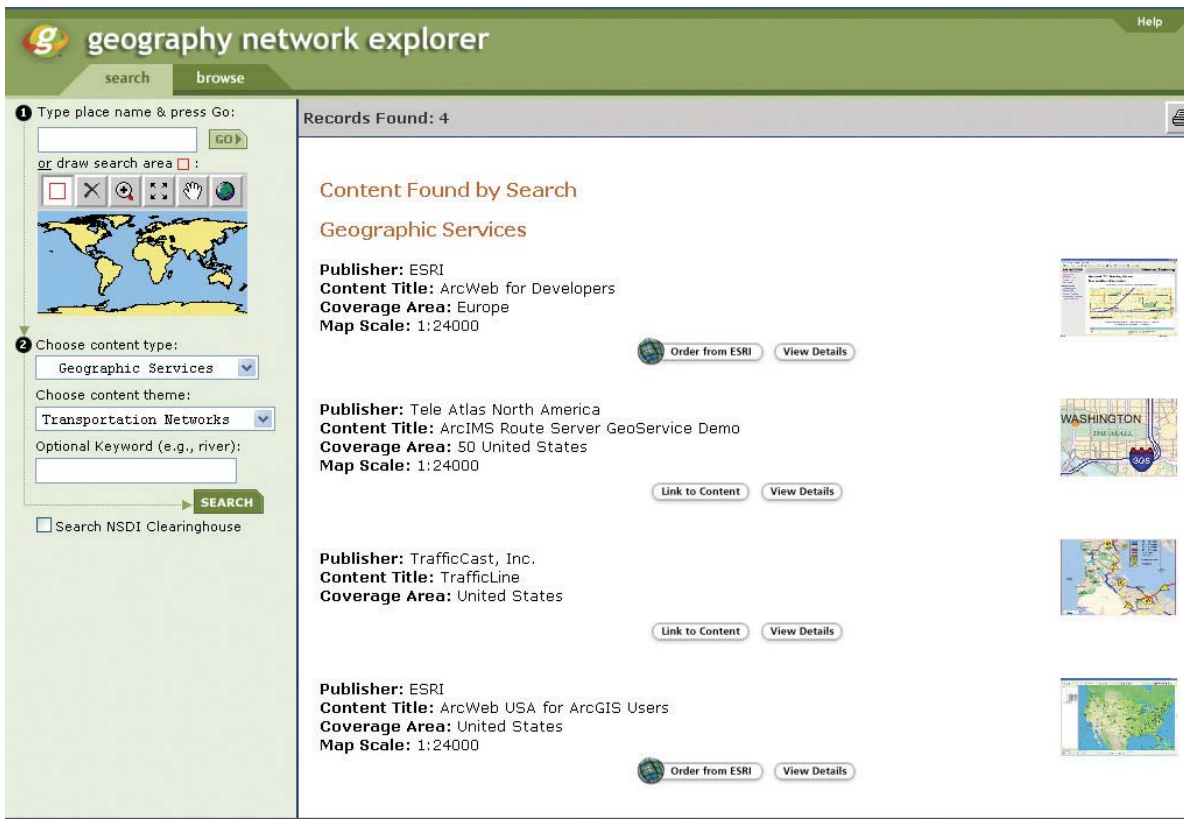


Figure 11.20 The Geography Network provides a directory of remote GIServices. In this case a search for services related to transportation networks has identified four

11.5 Prospects

Distributed GIS offers enormous advantages, in reducing duplication of effort, allowing users to access remotely located data and services through simple devices, and providing ways of combining information gathered through the senses with information provided from digital sources. Many issues continue to impede progress, however: complications resulting from the difficulties of interacting with devices in field settings; limitations placed on communication bandwidth and reliability; and limitations inherent in battery technology. Perhaps, at this time, more problematic than any of these is the difficulty of imagining the full potential of distributed GIS. We are used to associating GIS with the desktop, and conscious that we have not fully exploited its potential – so it is hard to imagine what might be possible when the GIS can be carried anywhere and its information combined with the window on the world provided by our senses.

Questions for further study

1. Design a location-based game based on GPS-enabled cellphones.
2. Find a selection of geolibraries on the Web, and identify their common characteristics. How does each of them: 1) allow the user to specify locations of interest; 2) allow the user to browse through datasets that have similar characteristics; and 3) allow the user to examine the contents of datasets before acquiring them?
3. To what extent do citizens have a right to locational privacy? What laws and regulations control the use of locational data on individuals?
4. Is an LBS a type of GIS, or is a GIS a type of LBS, or are they quite different forms of computer application? If they are different, are they likely to merge in the future, and what will the result look like?

Further reading

Câmara A.S. 2002 *Environmental Systems: A Multidimensional Approach*. New York: Oxford University Press.

National Research Council 1999 *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academy Press.

Peng Z.-H. and Tsou M.-H. 2003 *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ: Wiley.

IV

Analysis

12 *Cartography and map production*
13 *Geovisualization*
14 *Query, measurement, and transformation*
15 *Descriptive summary, design, and inference*
16 *Spatial modeling with GIS*

12 *Cartography and map production*

This chapter on cartography and map production is the first of a series of chapters that examine GIS output. The next, Chapter 13, deals with the closely related but distinct subject of visualization. The current chapter reviews the nature of cartography and the ways that users interact with GIS in order to produce digital and hard-copy reference and thematic maps. Standard cartographic conventions and graphic symbology are discussed, as is the range of transformations that are used in map design. Map production is reviewed in the context of creating maps for specific applications and also map series. Some specialized types of mapping are introduced that are appropriate for particular applications areas.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

At the end of this chapter you will understand:

- The nature of maps and cartography;
- Key map design principles;
- Choices that are available to compose maps;
- The many types of map symbology;
- Concepts of map production flow-lines.

12.1 Introduction

GIS output represents the pinnacle of many GIS projects. Since the purpose of information systems is to produce results, this aspect of GIS is vitally important to many managers, technicians, and scientists. Maps are a very effective way of summarizing and communicating the results of GIS operations to a wide audience. The importance of map output is further highlighted by the fact that many consumers of geographic information only interact with GIS through their use of map products.

For the purposes of organizing the discussion in this book, it is useful to distinguish between two types of GIS output: formal maps, created according to well-established cartographic conventions, that are used as a reference or communication product (e.g., a military mapping agency 1:250 000 scale topographic map, or a geological survey 1:50 000 scale paper map; Box 12.3); and transitory map and map-like visualizations used simply to display, analyze, edit, and query geographic information (e.g., results of a database query to retrieve areas with poor health indicators viewed on a desktop computer, or routing information displayed on a PDA device Figure 15.18). Both can exist in digital form on interactive display devices (see Box 13.3) or in hard-copy form on paper and other media. In practice, this distinction is somewhat arbitrary and there is considerable overlap, but

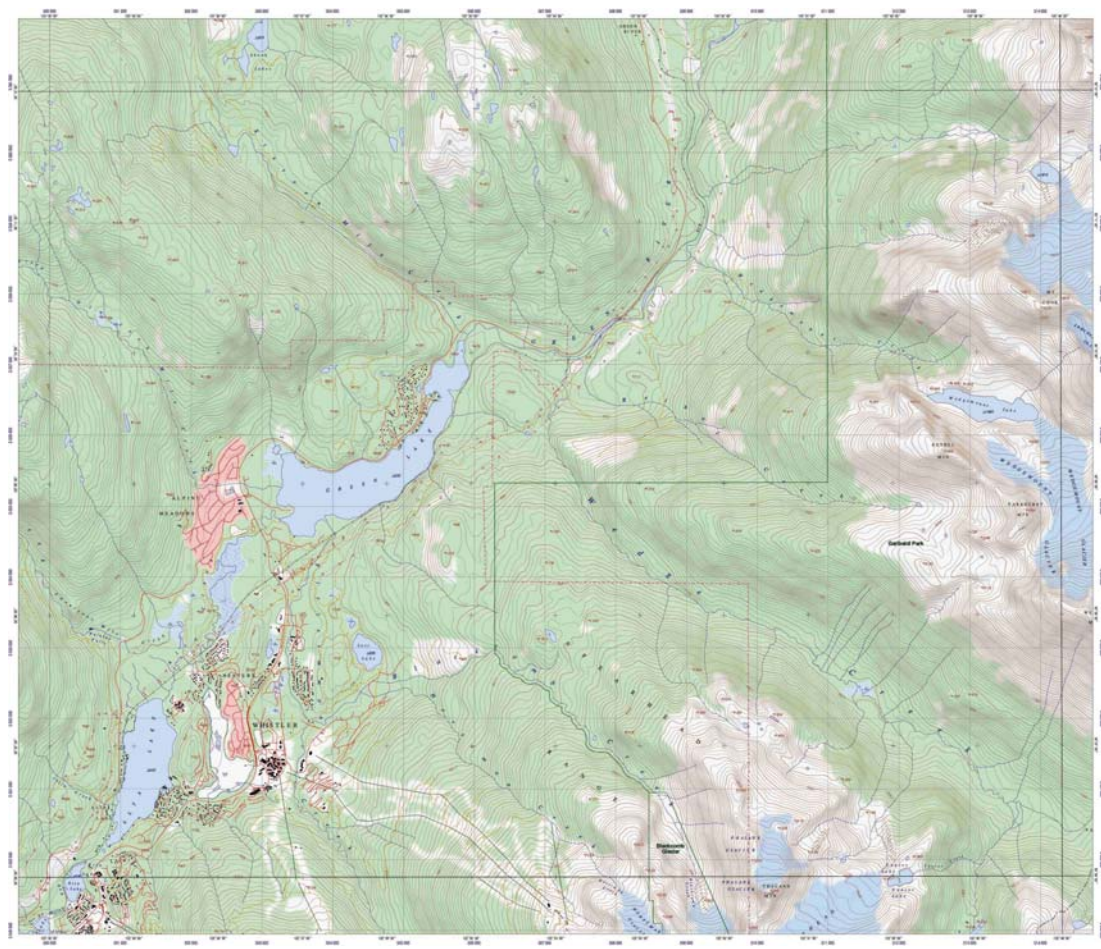


Figure 12.1 Terrestrial topographic map of Whistler, British Columbia, Canada. This is one of a collection of 7016 commercial maps at 1:20 000 scale covering the province (Courtesy: ESRI)

at the core the motivations, tools, and techniques for map production and map visualization are quite different. The current chapter will focus on maps and the next on visualization.

Cartography concerns the art, science, and techniques of making maps or charts. Conventionally, the term map is used for terrestrial areas (Figure 12.1) and chart for marine areas (Figure 12.2; see also Box 4.5), but they are both maps in the sense the word is used here. In statistical or analytical fields, charts provide the pictorial representation of statistical data, but this does not form part of the discussion in this chapter. Note, however, that statistical charts can be used on maps (e.g., Figure 12.3(B)). Cartography dates back thousands of years to a time before paper, but the main visual display principles were developed during the paper era and thus many of today's digital cartographers still use the terminology, conventions, and techniques from the paper era. Box 12.1 illustrates the importance of maps in a historic military context.

Maps are important communication and decision support tools.

Historically, the origins of many national mapping organizations can be traced to the need for mapping for 'geographical campaigns' of infantry warfare, for colonial administration, and for defense. Today such organizations fulfill a far wider range of needs of many more user types (see Chapter 19). Although the military remains a heavy user of mapping, such territorial changes as arise out of today's conflicts reflect a more subtle interplay of economic, political, and historical considerations – though, of course, the threat or actual deployment of force remains a pivotal consideration. Today, GIS-based terrestrial mapping serves a wide range of purposes – such as the support of humanitarian relief efforts (Figure 12.3) and the partitioning of territory through negotiation rather than force (e.g., GIS was used by senior decision makers in the partitioning of Bosnia

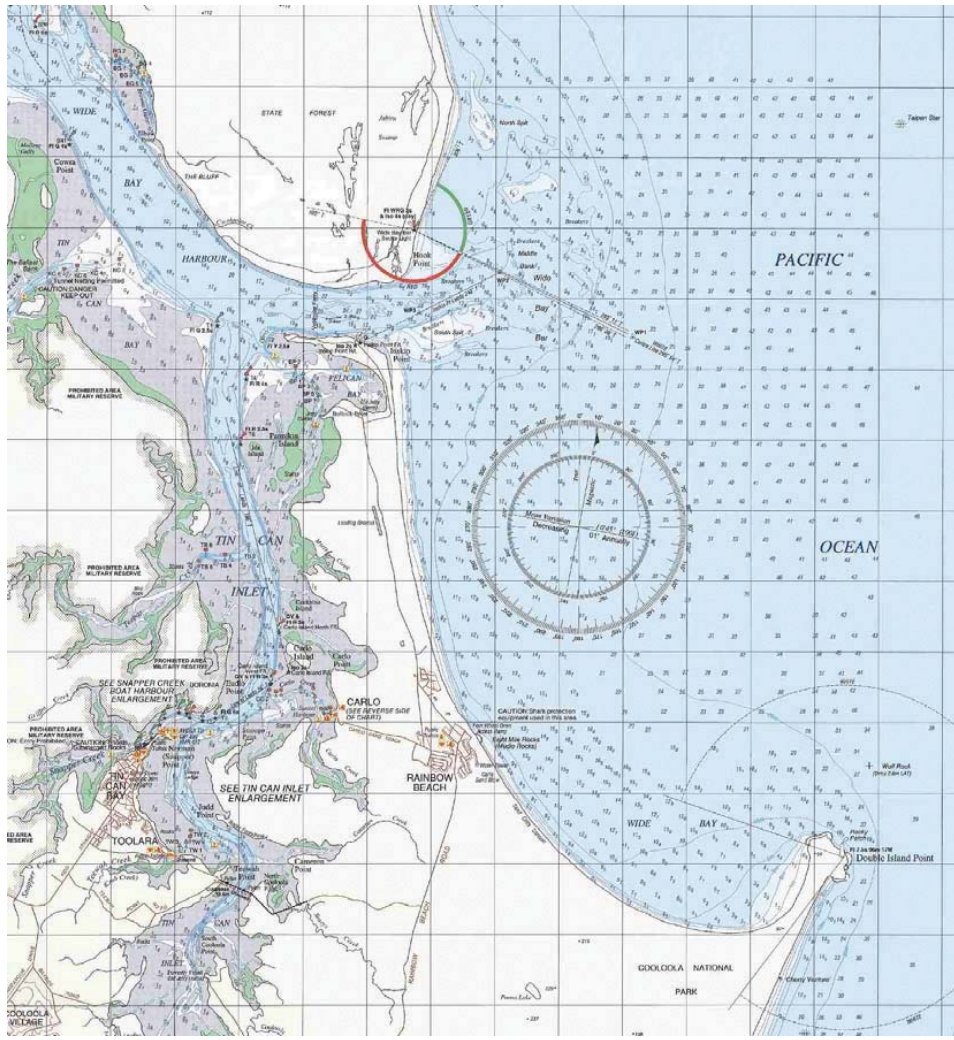


Figure 12.2 Great Sandy Strait (South), Queensland, Australia Boating Safety Chart. This marine chart conforms to international charting standards (Courtesy: ESRI) (Reproduced by permission of Maritime Safety Queensland.)

Applications Box **12.1**

Military maps in history

'Roll up that map; it will not be wanted these ten years.' British Prime Minister William Pitt the Younger made this remark after hearing of the defeat of British forces at the Battle of Austerlitz in 1805, where it became clear that his country's military campaign in Continental Europe had been thwarted for the foreseeable future. The quote illustrates the crucial historic role of mapping as a tool of decision support in warfare, in a world in which nation states were far more insular than they are today. It also identifies two other defining characteristics of the use of geographic information in 19th century society. First, the principal, straightforward purpose of much terrestrial mapping was to further national interests by infantry warfare. Second, the timeframe over which changes in geographic activity patterns unfolded was, by today's standards, incredibly slow – Pitt envisaged that no British citizen would revisit this territory for a quarter of a (then) average lifetime!

In the 19th century, printed maps became available in many countries for local, regional, and national areas, and their uses extended to land ownership, tax assessment, and navigation, among other activities.

(A)

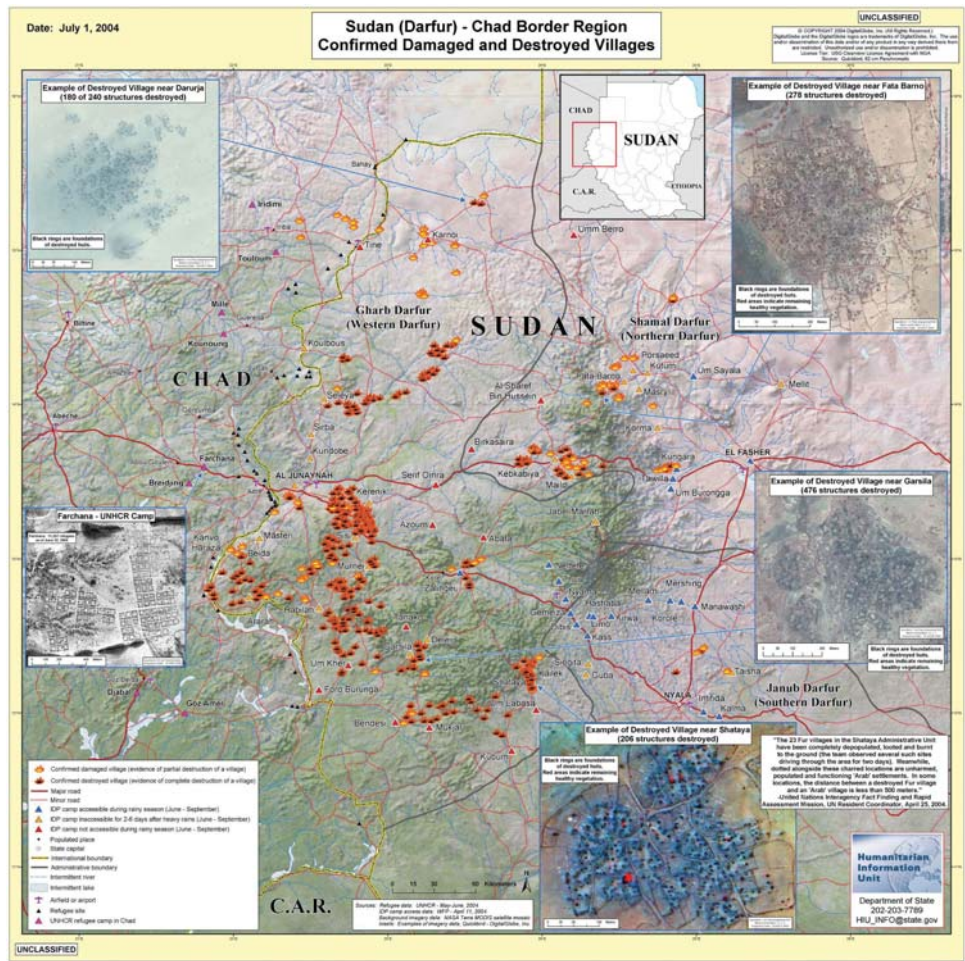


Figure 12.3 Humanitarian relief in Sudan–Chad border region: (A) damaged villages in July 2004 resulting from civil war (Courtesy Humanitarian Information Unit, Dept of State, US Government); (B) malaria vulnerability (Courtesy: World Health Organization)

(B)

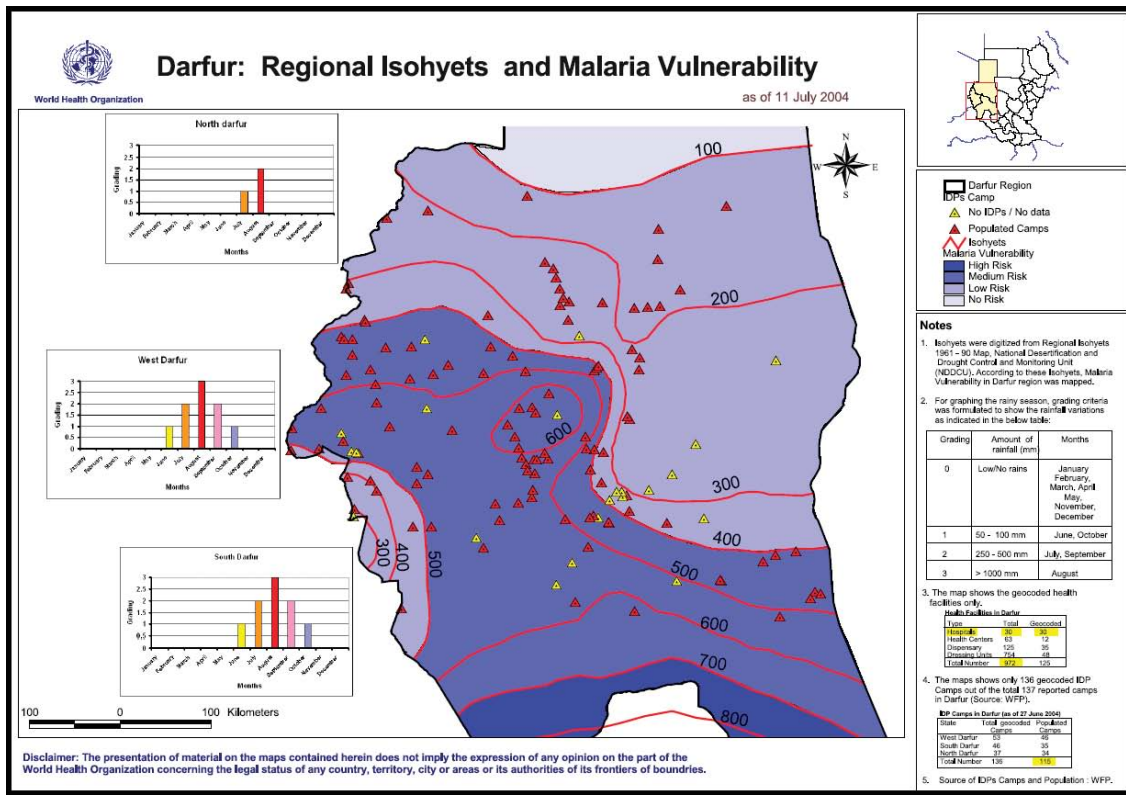


Figure 12.3 (*continued*)

in 1999). The timeframe over which events unfold is also much more rapid – it is inconceivable to think of politicians, managers, and officials being able to neglect geographic space for months, weeks, or even days, never mind years.

Paper maps remain in widespread use because of their transportability, their reliability, ease of use, and the straightforward application of printing technology that they entail. They are also amenable to conveying straightforward messages and supporting decision making. Yet the increasing detail of our understanding of the natural environment and the accelerating complexity of society mean that the messages that mapping can convey are increasingly sophisticated and, for the reasons set out in Chapter 6, uncertain. Greater democracy and accountability, coupled with the increased spatial reasoning abilities that better education brings, mean that more people than ever feel motivated and able to contribute to all kinds of spatial policy. This makes the decision support role immeasurably more challenging, varied, and demanding of visual media. Today's mapping must be capable of communicating an extensive array of messages and emulating the widest range of 'what if' scenarios (Section 16.1.1).

Both paper and digital maps have an important role to play in many economic, environmental, and social activities.

The visual medium of a given application must also be open to the widest community of users. Technology has led to the development of an enormous range of devices to bring mapping to the greatest range of users in the widest spectrum of decision environments. In-vehicle displays, palm top devices, and wearable computers are all important in this regard (Figures 11.1, 11.2 and 11.8). Most important of all, the innovation of the Internet makes 'societal representations' of space a real possibility for the first time.

12.2 Maps and cartography

There are many possible definitions of a map; here we use the term to describe digital or analog (soft- or hard-copy) output from a GIS that shows geographic information using well-established cartographic conventions. A map is the final outcome of a series of GIS data processing steps (Figure 12.4) beginning with data collection (Chapter 9), editing and maintenance, through data management (Chapter 10), analysis (Chapters 14–16) and concluding with a map. Each of these activities successively transforms a database of geographic information until it is in the form appropriate to display on a given technology.

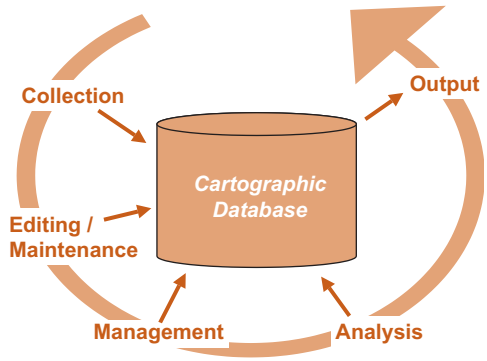


Figure 12.4 GIS processing transformations needed to create a map

Central to any GIS is the creation of a data model that defines the scope and capabilities of its operation (Chapter 8), and the management context in which it operates (Chapters 17–20). There are two basic types of map (Figure 12.5): reference maps, such as topographic maps from national mapping agencies (see also Figure 12.1), that convey general information; and thematic maps that depict specific geographic themes, such as population census statistics, soils, or climate zones (see Figures 12.7 and 12.15 for other examples).

Maps fulfill two very useful functions, acting as both storage and communication mechanisms for geographic information. The old adage 'a picture is worth a thousand words' connotes something of the efficiency of maps as a storage container. The modern equivalent of this is 'a map is worth a million bytes'. Before the advent of GIS, the paper map was the database, but a map can now be considered a single product generated from a digital database. Maps are also a mechanism to communicate information to viewers. Maps can present the results of analyses (e.g., the optimum site suitable for locating a new store, or analysis of the impact of an oil spill). They can communicate spatial relationships between phenomena across the same map, or between maps of the same or different areas. As such they can assist in the identification of spatial order and differentiation. Effective decision support requires that the message of the map is readily interpretable in the mind of the decision maker. A major function of a map is not simply to marshal and transmit known information about the world, but also to create or reinforce a particular message. Menno-Jan Kraak, an academic cartographer, speaks about his personal experiences of the development of cartography in Box 12.2

Maps are both storage and communication mechanisms.

(A)

Figure 12.5 Two maps of the Southern United States: (A) reference map showing topographic information; (B) thematic map showing population density in 1996

(B)

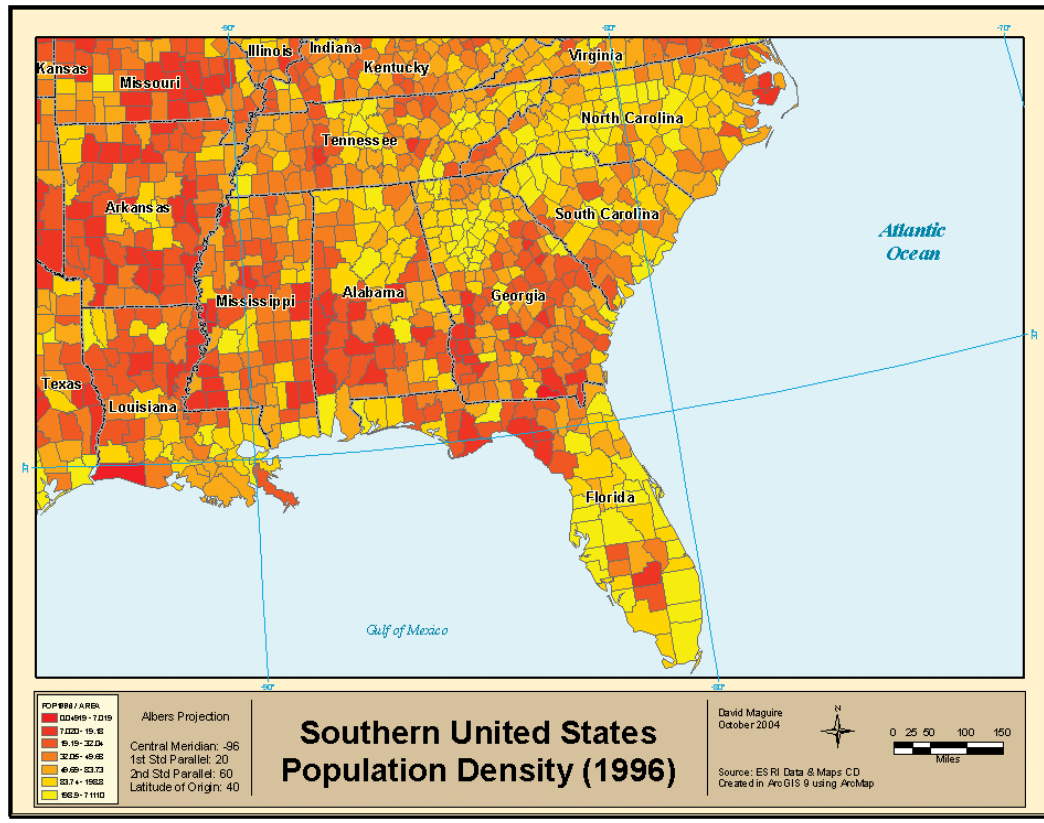


Figure 12.5 (*continued*)

Maps also have several limitations:

- Maps can be used to miscommunicate (lie) accidentally or on purpose. For example, incorrect use of symbols can convey the wrong message to users by highlighting one type of feature at the expense of another (see Figure 12.13 for an example of different choropleth map classifications).
- Maps are a single realization of a spatial process. If we think for a moment about maps from a statistical perspective, then each map instance represents the outcome of a sampling trial and is therefore a single occurrence generated from all possible maps on the same subject for the same area. The significance of this is that other sample maps drawn from the same population would exhibit variations and, consequently, we need to be careful in drawing inferences from a single map sample. For example, a map of soil textures is derived by interpolating soil sample texture measurements. Repeated sampling of soils will show natural variation in the texture measurements.
- Maps are often created using complex rules, symbology, and conventions, and can be difficult to understand and interpret by the untrained viewer. This is particularly the case, for example, in multivariate statistical thematic mapping where the idiosyncrasies of classification schemes and color symbology can be challenging to comprehend.

Uncertainty pertains to maps just as it does to other geographic information.

12.2.1 Maps and media

Without question, GIS has fundamentally changed cartography and the way we create, use, and think about maps. The digital cartography of GIS frees map-makers from many of the constraints inherent in traditional (non-GIS) paper mapping (see also Section 3.7). This is because:

- The paper map is of fixed scale. Generalization procedures (Section 3.8) can be invoked in order to maintain clarity during map creation. This detail is not recoverable, except by reference back to the data from which the map was compiled. The zoom facility of GIS can allow mapping to be viewed at a range of scales, and detail to be filtered out as appropriate at a given scale.
- The paper map is of fixed extent and adjoining map sheets must be used if a single map sheet does not cover the entire area of interest. (An unwritten law of paper map usage is that the most important map features always lie at the intersection of four paper map sheets!) GIS, by contrast, can provide a seamless medium for viewing space, and users are able to pan across wide swathes of territory.

Biographical Box **12.2**

Menno-Jan Kraak, cartographer

Figure 12.6 Menno-Jan Kraak, cartographer

Menno-Jan Kraak graduated in geography and cartography from the Department of Geosciences, Utrecht University, where he also started his academic career. After a short interruption for the obligatory conscript period, which he spent at the military geography unit, he moved to the Department of Geodesy, Delft University of Technology. In this geo-technical environment he was awarded a Ph.D. on three-dimensional mapping. In Delft, the foundations of a modern cartographic text book co-authored with Ferjan Ormeling were laid. *Cartography: Visualization of Spatial Data* was translated into several languages and is still available in its second edition. In 1996, Menno-Jan moved to Enschede to become Professor in Geovisualization at the ITC. This International Institute for Geo-Information Science and Earth Observation concentrates its activities on students from the developing world. His own international orientation is further expanded by the exotic traveling for ITC which offers an interesting perspective on the world of maps. Menno-Jan also has a keen interest in how cartography is depicted in world stamps (**www.itc.nl/personal/kraak/stamps/**).

Reflecting on the present state of cartography, Menno-Jan says:

During my career which started as a 'real' cartographer I have been part of the process that evolved cartography via computer cartography and GIS into an integral part of what is today called the GIScience world. However, during this process GIS took on board methods and techniques from non-geo-related disciplines such as scientific and information visualization. In parallel I have witnessed the same integrative trend in the Dutch geo-community. In my early Delft years I was a stranger in an (interesting) geodetic world and we all had our own professional organizations, but trends in our professions brought the different disciplines closer together. In 2003 I was involved in the merger of the geo-societies in the Netherlands into a single new geo-information society. Another interesting observation I would like to share is that one should always be open for new methods and techniques especially from other disciplines. You should certainly try them out in your own environment, but while doing so do not forget the knowledge already obtained. It can be very tempting to concentrate on the toys instead of trying to incorporate derived new knowledge in the theory of mapping.

- Most paper maps present a static view of the world whereas conventional paper maps and charts are not adept at portraying dynamics. GIS-based representations are able to achieve this through animation.
- The paper map is flat and hence limited in the number of perspectives that it can offer on three-dimensional data. 3-D visualization is much more effective within GIS which can support interactive pan and zoom operations (see Figure 9.13 for a 3-D view example).
- Paper maps provide a view of the world as essentially complete. GIS-based mapping allows the supplementation of base-map material with further data. Data layers can be turned on and off to examine data combinations.
- Paper maps provide a single, map producer-centric, view of the world. GIS users are able to create their own, user-centric, map images in an interactive way. Side-by-side map comparison is also possible in GIS.

GIS is a flexible medium for the production of many types of maps.

12.3 Principles of map design

Map design is a creative process during which the cartographer, or map-maker, tries to convey the message of the map's objective. Primary goals in map design are to share information, highlight patterns and processes, and illustrate results. A secondary objective is to create a pleasing and interesting picture, but this must not be

Applications Box **12.3**

Czech Geological Survey (CGS) Map Production

In 1994, the Czech Geological Survey (CGS) began extensive use of GIS technology to meet the increasing demand for digital information about the environment. GIS in the CGS is focused on the methods of spatial data processing, unification, and dissemination. Digital processing of geological maps and the development of GIS follows standardized procedures using common geological dictionaries and graphic elements. Recently, the main objective has been to create and implement a uniform geological data model and provide the public and the scientific community with easy access to geographic data, via a Web map server (**www.geology.cz**).

CGS has a unique geographic information system, containing more than 260 000 mapped geological objects from the entire Czech Republic. The fundamental part of this geographic database is the unified national geological index (legend), which consists of four main types of information – chronostratigraphical units, regional units, lithostratigraphical units, and lithological description of rocks. The database has been under revision since 1998 leading to the creation of a seamless digital geological map of the Czech Republic. This database has already been used for land use planning by government and local authorities.

The geological map of the Krkonose–Jizera mountains shown in Figure 12.7 is a cartographic presentation of one part of the CGS database. The overview map in the bottom left corner shows the extents of all maps in the series.

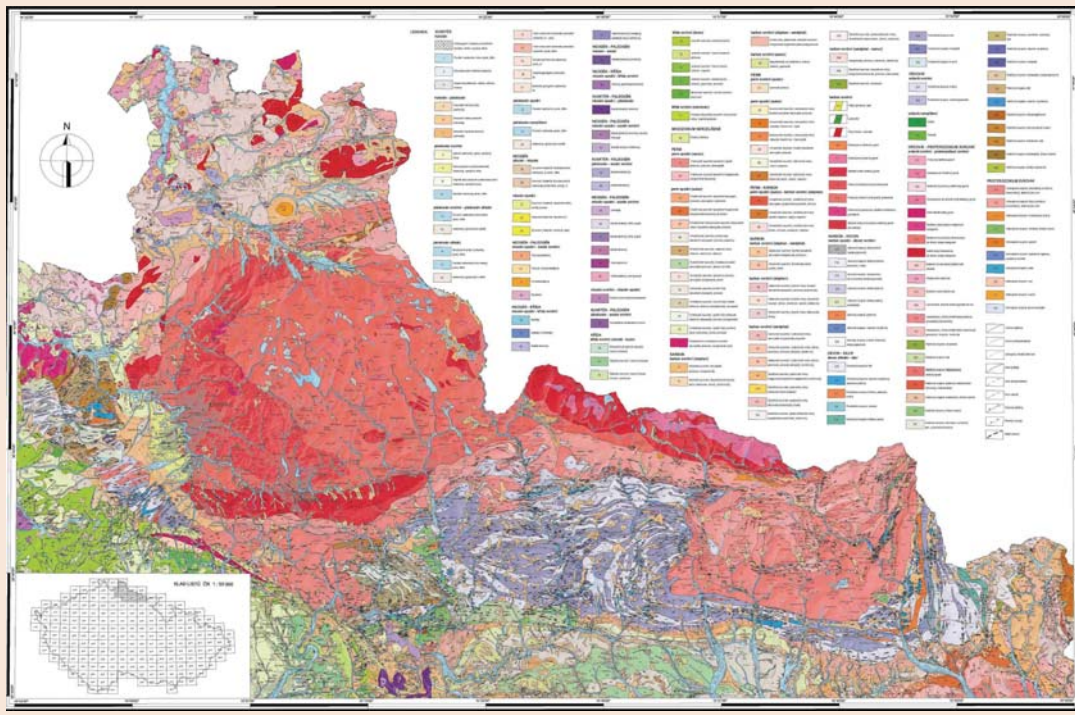


Figure 12.7 A 1:50 000 geological map of the Krkonose–Jizera mountains, Czech Republic (Courtesy: ESRI)

at the expense of fidelity to reality and meeting the primary goals. Map design is quite a complex procedure requiring the simultaneous optimization of many variables and harmonization of multiple methods. Cartographers must be prepared to compromise and balance choices. It is difficult to define exactly what constitutes a 'good design'. The general consensus is that a good design is one that looks good, is simple and elegant, and most importantly, leads to a map that is fit for the intended purpose.

Robinson et al (1995) define seven controls on the map design process:

- *Purpose.* The purpose for which a map is being made will determine what is to be mapped and how the information is to be portrayed. Reference maps are multi-purpose, whereas thematic maps tend to be single purpose (Figure 12.5). With the digital technology of a GIS, it is easier to create maps, and many more are digital and interactive. As a consequence today maps are increasingly single purpose.
- *Reality.* The phenomena being mapped will usually impose some constraints on map design. For example, the orientation of the country – whether it be predominantly east-west (Russia) or north-south (Chile) – will determine layout in no small part.
- *Available data.* The specific characteristics of data (e.g., raster or vector, continuous or discrete, or point, line, or area) will affect the design. There are many different ways to symbolize map data of all types, as discussed in Section 12.3.2.
- *Map scale.* Scale is an apparently simple concept, but it has many ramifications for mapping (see Box 4.2 for further discussion). It will control how many data can appear in a map frame, the size of symbols, the overlap of symbols, and much more. Although one of the early promises of digital cartography and GIS was 'scale-free' databases that could be used to create multiple maps at different scales, this has never been realized because of technical complexities.
- *Audience.* Different audiences want different types of information on a map and expect to see information presented in different ways. Usually, executives (and small children!) are interested in summary information that can be assimilated quickly, whereas advanced users often want to see more information. Similarly, those with restricted eyesight find it easier to read bigger symbols.
- *Conditions of use.* The environment in which a map is to be used will impose significant constraints. Maps for outside use in poor or very bright light will need to be designed differently than maps for use indoors where the light levels are less extreme.
- *Technical limits.* The display medium, be it digital or hardcopy, will impact the design process in several ways. For example, maps to be viewed in an Internet browser, where resolution and bandwidth are limited, should be simpler and based on less data than equivalents to be displayed on a desktop PC monitor.

12.3.1 Map composition

Map composition is the process of creating a map comprising several closely interrelated elements (Figure 12.8):

- Map body. The principal focus of the map is the main map body, or in the case of comparative maps there will be two or more map bodies. It should be given space and use symbology appropriate to its significance.
- Inset/overview map. Inset and overview maps may be used to show, respectively, an area of the main map body in more detail (at a larger scale) and the general location or context of the main body.

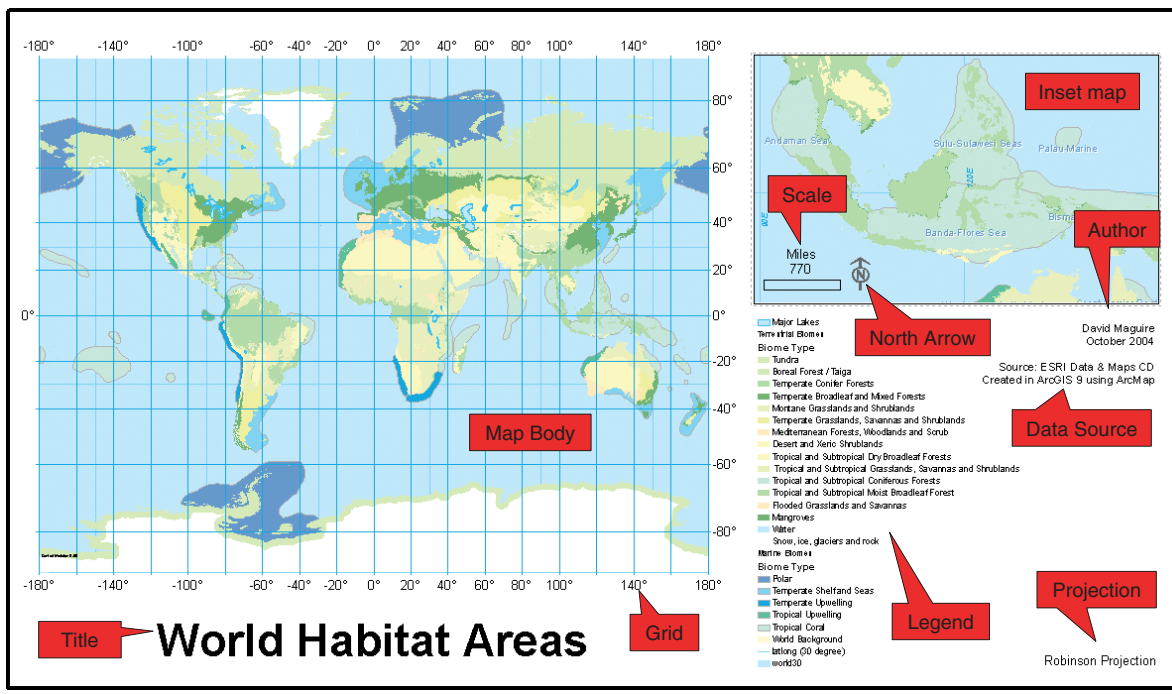


Figure 12.8 The principal components of a map composition layout

- Title. One or more map titles are used to identify the map and to inform the reader about its content.
- Legend. This lists the items represented on the map and how they are symbolized. Many different layout designs are available and there is a considerable body of information available about legend design.
- Scale. The map scale provides an indication of the size of objects and the distances between them. A paper map scale is a ratio, where one unit on the map represents some multiple of that value in the real world. The scale can be symbolized numerically (1:1000), graphically (a scalebar), or texturally ('one inch equals 1000 inches'). The scale is a representative fraction and so a 1:1000 scale is larger (finer) than a 1:100 000. A small (coarse) scale map displays a larger area than a large (fine) scale map, but with less detail. See also Box 4.2 for more details about scale.
- Direction indicator. The direction and orientation of a map can be conveyed in one of several ways including grids, graticules, and directional symbols (usually north arrows). A grid is a network of parallel and perpendicular lines superimposed on a map. A graticule is a network of longitude and latitude lines on a map that relates points on a map to their true location on the Earth.
- Map metadata. Map compositions can contain many other types of information including the map projection, date of creation, data sources, and authorship.

A key requirement for a good map is that all map elements are composed into a layout that has good visual balance. On large-scale maps, such as 1:50 000 national mapping agency topographic maps, all the contextual items (everything listed above except the map body) usually appear as marginal notations (or marginalia). In the case of map series or atlases (see Section 12.4), some of the common information may be in a separate document. On small- or medium-scale maps this information usually appears within the map border (e.g., Figure 12.7).

12.3.2 Map symbolization

The data to be displayed on a map must be classified and represented using graphic symbols that conform to well-defined and accepted conventions. The choice of symbolization is critical to the usefulness of any map. Unfortunately, the seven controls on the design process listed in Section 12.3.1 also conspire to mean that there is not a single universal symbology model applicable everywhere, but rather one for each combination of factors. Again, we see that cartographic design is a compromise reached by simultaneously optimizing several factors.

Good mapping requires that spatial objects and their attributes can be readily interpreted in applications. In Chapter 3, attributes were classified as being measured on the nominal, ordinal, interval, or ratio scales (Box 3.3), while in Chapter 4 spatial objects were classified into points, lines, areas, and surfaces (Box 4.1). We have already seen how attribute measures that we think of as continuous are actually discretized to levels of precision imposed by measurement or design (Section 6.3.). The representation of spatial objects is similarly imposed – cities might be captured as points, areas, mixtures of points, lines, and areas (as in a street map: Figure 12.19) or 3-D 'walk-throughs' (see Section 13.4), depending on the base-scale of a representation and the importance of city objects to the application. Measurement scales and spatial object types are thus one set of conventions that are used to abstract reality. Whether using GIS or paper, mapping may entail reclassification or transformation of attribute measures.

The process of mapping attributes frequently entails further problems of classification because many spatial attributes are inherently uncertain (Chapter 6). For example, in order to create a map of occupational type, individuals' occupations will be classified first into socio-economic groups (e.g., 'factory worker') and perhaps then into super-groups, such as 'blue collar'. At every stage in the aggregation process we inevitably do injustice to many individuals who perform a mix of white and blue collar, intermediate and skilled functions by lumping them into a single group (what social class is a frogman?). In practice, the validity and usefulness of an occupational classification will have become established over repeated applications, and the task of mapping is to convey thematic variation in as efficient a way as possible.

12.3.2.1 Attribute representation and transformation

Humans are good at interpreting visual data – much more so than interpreting numbers, for example – but conventions are still necessary to convey the message that the map-maker wants the data to impart. Many of these conventions relate to use of symbols (such as the way highway shields denote route numbers on many US medium- and fine-scale maps; Figure 12.19) and colors (blue for rivers, green for forested areas, etc.), and have been developed over the past few hundred years. Mapping of different themes (such as vegetation cover, surface geology, and socio-economic characteristics of human populations) has a more recent history. Here too, however, mapping conventions have developed, and sometimes they are specific to particular applications.

Attribute mapping entails use of graphic symbols, which (in two dimensions) may be referenced by points (e.g., historic monuments and telecoms antennae), lines (e.g., roads and water pipes) or areas (e.g., forests and urban areas). Basic point, line, and area symbols are modified in different ways in order to communicate different types of information. The ways in which these modifications take place adhere to cognitive principles and the accumulated experience of application implementations. The nature of these modifications was first explored by Bertin in 1967, and was extended to the typology illustrated in Figure 12.9 by MacEachren. The size and orientation of point and line symbols is varied principally to

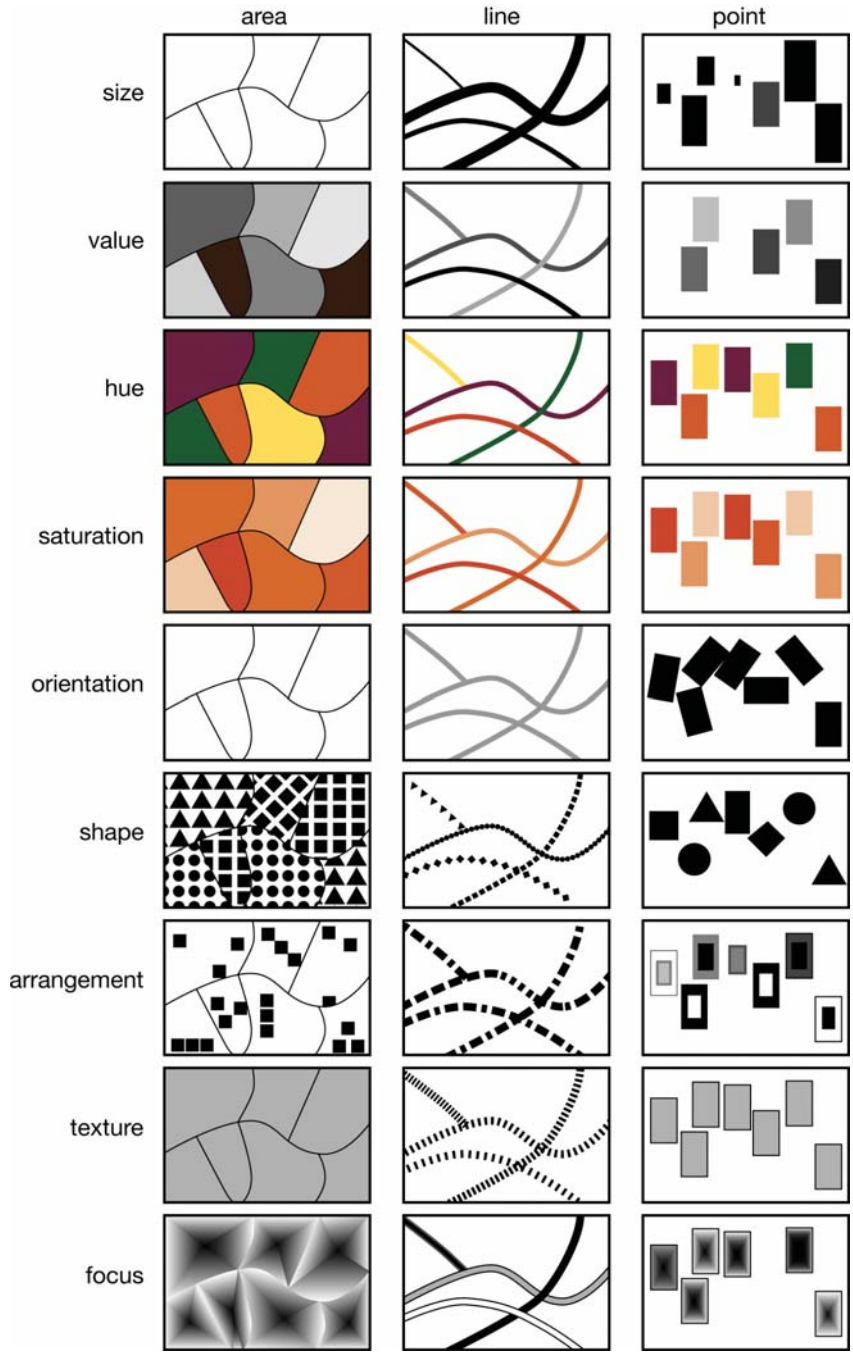


Figure 12.9 Bertin's graphic primitives, extended from seven to ten variables (the variable location is not depicted). *Source*: MacEachren 1994, from *Visualization in Geographical Information Systems,* Hearnshaw H.M. and Unwin D.J. (eds), Plate B. (Reproduced by permission of John Wiley & Sons, Ltd.)

distinguish between the values of ordinal and interval/ratio data using graduated symbols (such as the proportional pie symbols shown in Figure 12.10). Figure 12.11 illustrates how orientation and color can be used to depict the properties of locations, such as ocean current strength and direction.

Hue refers to the use of color, principally to discriminate between nominal categories, as in agricultural or urban land-use maps (Figure 12.12). Different hues may be combined with different textures or shapes (see below) if there are a large number of categories in order to avoid difficulties of interpretation. The shape of map symbols can be used either to communicate information about a spatial attribute (e.g., a viewpoint or the start of a walking trail), or its spatial location (e.g., the location of a road or boundary of a particular type: Figure 12.12), or spatial

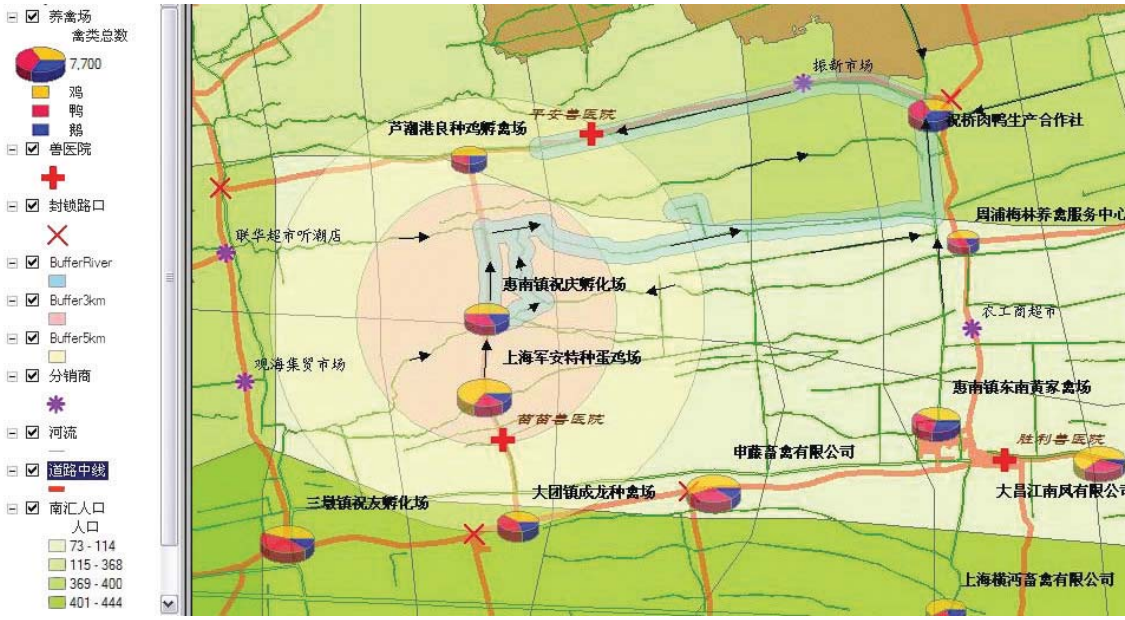


Figure 12.10 Incidence of Asian Bird Flu in Shanghai City. Pie size is proportional to the number of cases; red = duck, yellow = chicken, blue = goose

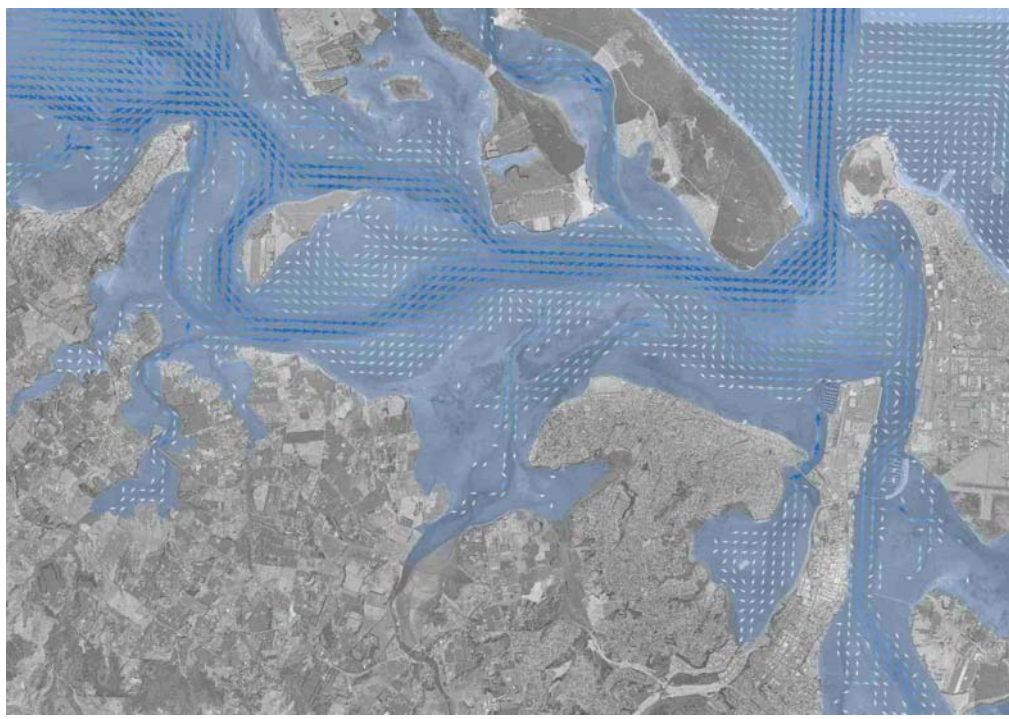
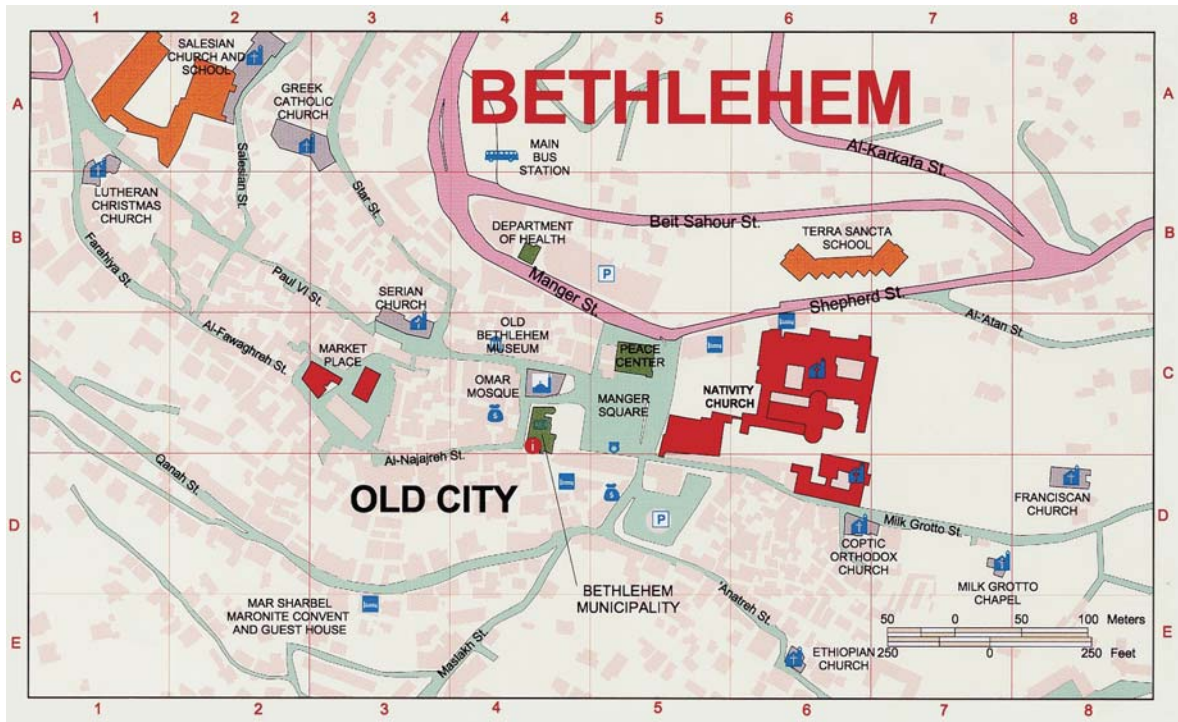


Figure 12.11 Tauranga Harbour tidal movements, Bay of Plenty, NZ. The arrows indicate speed (color) and direction (orientation) (Courtesy: ESRI)

relationships (e.g., the relationship between sub-surface topography and ocean currents). Arrangement, texture, and focus refer to within- and between-symbol properties that are used to signify pattern. A final graphic variable in the typologies of MacEachren and Bertin is location (not shown in Figure 12.9), which refers to the practice of offsetting the true coordinates of objects in order to improve map intelligibility, or changes in map projection. We discuss this in more detail below. Some of the common ways in which these graphic variables are used to visualize spatial object types and attributes are shown in Table 12.1.

The selection of appropriate graphic variables to depict spatial locations and distributions presents one



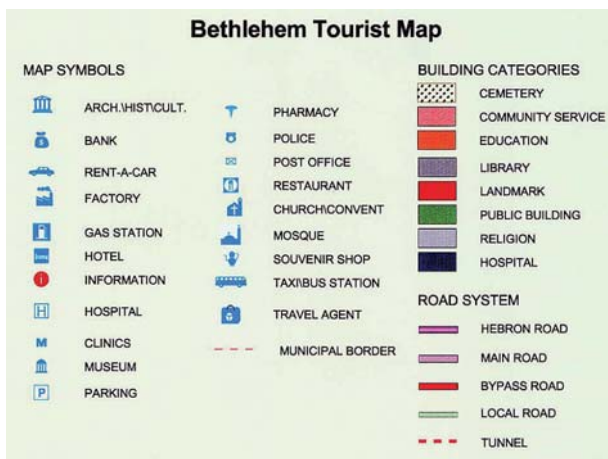


Figure 12.12 Use of hue (color) to discriminate between Bethlehem, Israel urban land-use categories, and of symbols to communicate location and other attribute information (Courtesy: ESRI)

set of problems in mapping. A related task is how best to position symbols on the map, so as to optimize map interpretability. The representation of nominal data by graphic symbols and icons is apparently trivial, although in practice automating placement presents some challenging analytical problems. Most GIS packages include generic algorithms for positioning labels and symbols in relation to geographic objects. Point labels are positioned to avoid overlap by creating a window, or mask (often invisible to the user), around text or symbols. Linear features, such as rivers, roads, and contours, are labeled by placing the text using a spline function to give a smooth even distribution, or distinguished by use of color. Area labels are assigned to central points (Figure 12.16), using geometric algorithms similar to those used to calculate geometric centroids (Section 15.2.1). These generic algorithms are frequently customized to accommodate common conventions and rules for particular classes of application – such as topographic (Figure 12.1), utility, transportation, and seismic maps, for example.

Generic and customized algorithms also include color conventions for map symbolization and lettering.

Ordinal attribute data are assigned to point, line, and area objects in the same rule-based manner, with the ordinal property of the data accommodated through use of a hierarchy of graphic variables (symbol and lettering sizes, types, colors, intensities, etc.). As a general rule,

Table 12.1 Common methods of mapping spatial object types and attribute data with examples (2-D = two-dimensional)

Spatial object type	Attribute type		
	Nominal	Ordinal	Interval/Ratio
Point (0-D)	Symbol map (each category a different class of symbol – color, shape, orientation), and/or use of lettering: e.g., presence/absence of city (Figure 12.12)	Hierarchy of symbols or lettering (color and size): e.g., small/medium/large depots (Figure 12.22)	Graduated symbols (color and size): e.g., pollution concentrations (Figure 12.10)
Line (1-D)	Network connectivity map (color, shape, orientation): e.g., presence/absence of connection (Figure 12.20)	Graduated line symbology (color and size): e.g., road classifications (Figure 12.19)	Flow map with width or color lines proportional to flows (color and size): e.g., traffic flows (Figure 12.11)
Area (2-D)	Unique category map (color, shape, orientation, pattern): e.g., soil types (Figure 12.7)	Graduated color or shading map: e.g., timber yield low/medium/high (Figure 12.3)	Continuous hue/shading, e.g., dot-density or choropleth map: e.g., percentage of retired population (Figure 12.5B)
Surface (2.5-D)	One color per category (color, shape, orientation, pattern), e.g., relief classes: mountain/valley (Figure 6.7)	Ordered color map, e.g., areas of gentle/steep/very steep slopes	Contour map, e.g., isobars/isohyets: e.g., topography contours (Figure 12.3B)

the typical user is unable to differentiate between more than seven (plus or minus two) ordinal categories, and this provides an upper limit on the normal extent of the hierarchy.

A wide range of conventions is used to visualize interval- and ratio-scale attribute data. Proportional circles and bar charts are often used to assign interval- or ratio-scale data to point locations (Figures 12.10 and 12.14). Variable line width (with increments that correspond to the precision of the interval measure) is a standard convention for representing continuous variation in flow diagrams.

There is a variety of ways of ascribing interval or ratio scale attribute data to areal entities that are pre-defined. In practice, however, none is unproblematic. The standard method of depicting areal data is in zones (Figure 12.5B). However, as was discussed in Section 6.2, the choropleth map brings the dubious visual implication of within-zone uniformity of attribute value. Moreover, conventional choropleth mapping also allows any large (but possibly uninteresting) areas to dominate the map visually. A variant on the conventional choropleth map is the dot-density map, which uses points as a more aesthetically pleasing means of representing the relative density of zonally averaged data – but not as a means of depicting the precise locations of point events. Proportional circles provide one way around this problem; here the circle is scaled in proportion to the size of the quality being mapped and the circle can be centered on any convenient point within a zone (Figure 12.10). However, there is a tension between using circles that are of sufficient size to convey the variability in the data and the problems of overlapping circles on 'busy' areas of maps that have large numbers of symbols. Circle positioning also entails the same kind of positioning problem as that of name and symbol placement outlined above.

If the richness of the map presentation is to be equivalent to that of the representation from which it was derived, the intensity of color or shading should directly mirror the intensity or magnitude of attributes. The human eye is adept at discerning continuous variations in color and shading, and Waldo Tobler has advanced the view that continuous scales present the best means of representing geographic variation. There is no natural ordering implied by use of different colors and the common convention is to represent continuous variation on the red-green-blue (RGB) spectrum. In a similar fashion, difference in the hue, lightness, and saturation (HLS) of shading is used in color maps to represent continuous variation. International standards on intensity and shading have been formalized.

At least four basic classification schemes have been developed to divide interval and ratio data into categories (Figure 12.13):

1. Natural (Jenks) breaks, in which classes are defined according to apparently natural groupings of data values. The breaks may be imposed on the basis of break points that are known to be relevant to a particular application, such as fractions and multiples of mean income levels, or rainfall thresholds known to support different thresholds of vegetation ('arid', 'semi-arid', 'temperate', etc.). This is 'top down' or deductive assignment of breaks. Inductive ('bottom up') classification of data values may be carried out by using GIS software to look for relatively large jumps in data values.
2. Quantile breaks, in which each of a predetermined number of classes contains an equal number of observations. Quartile (four-category) classifications are widely used in statistical analysis, while quintile (five-category: Figure 12.13) classifications are well suited to the spatial display of uniformly distributed

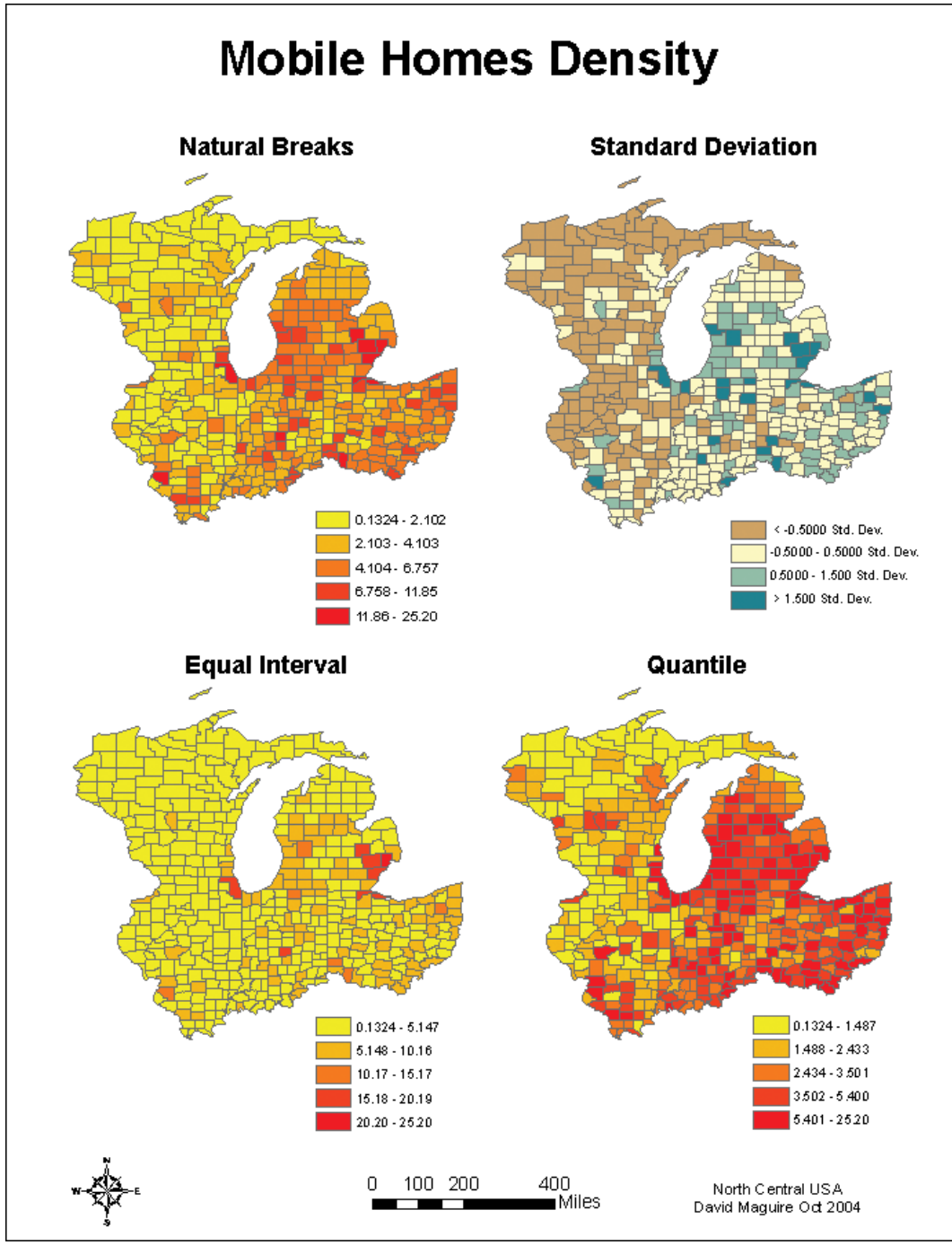


Figure 12.13 Comparison of choropleth class definition schemes: natural breaks, standard deviation, equal interval, quantile

data. Yet because the numeric size of each class is rigidly imposed, the result can be misleading. The placing of the boundaries may assign almost identical attributes to adjacent classes, or features with quite widely different values in the same class. The resulting visual distortion can be minimized by increasing the number of classes – assuming the user can assimilate the extra detail that this creates.

3. Equal interval breaks. These are best applied if the data ranges are familiar to the user of the map, such as temperature bands.

4. Standard deviation classifications show the distance of an observation from the mean. The GIS calculates the mean value and then generates class breaks in standard deviation measures above and below it. Use

of a two-color ramp helps to emphasize values above and below the mean (Figure 12.13).

Classification procedures are used in map production in order to ease user interpretation.

The choice of classification is very much the outcome of choice, convenience, and the accumulated experience of the cartographer. The automation of mapping in GIS has made it possible to evaluate different possible classifications. Looking at distributions (Figure 14.8) allows us to see if the distribution is strongly skewed – which might justify using unequal class intervals in a particular application. A study of poverty, for example, could quite happily class millionaires along with all those earning over $50 000, as they would be equally irrelevant to the study.

12.3.2.2 Multivariate mapping

Multivariate maps show two or more variables for comparative purposes. For example, Figure 12.10 shows the incidence of Asian Bird Flu as proportional symbols, and Figure 12.3B illustrates rainfall and malaria risk. Many maps are a compilation of composite maps based upon a range of constituent indicators. Climate maps, for example, are compiled from direct measures such as amount and distribution of precipitation, diurnal temperature variation, humidity, and hours of sunshine, plus indirect measures such as vegetation coverage and type. It is unlikely that there will ever be a perfect correspondence between each of these components, and historically it has been the role of the cartographer to integrate disparate data. In some instances, data may be averaged over zones which are devoid of any strong meaning, while the different components that make up a composite index may have been measured at a range of scales. Mapping can mask scale and aggregation problems where composite indicators are created at large scales using components that were only intended for use at small scales.

Figure 12.14 illustrates how multivariate data can be displayed on a shaded area map. In this soil texture map of Mexico three variables are displayed simultaneously, as indicated in the legend, using color variations to display combinations of per cent sand (base), silt (right), and clay (left).

Box 12.4 shows an innovative method for representing multivariate data on maps as clock diagrams.

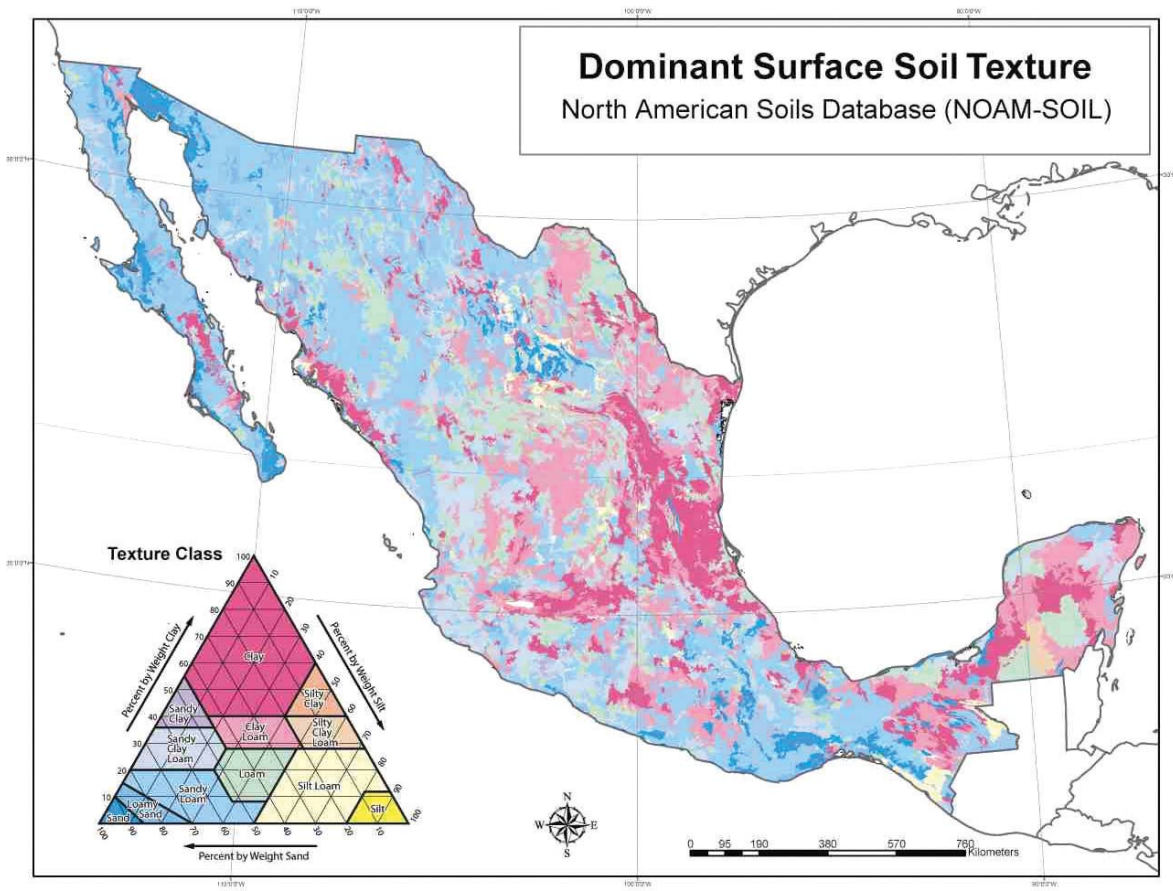


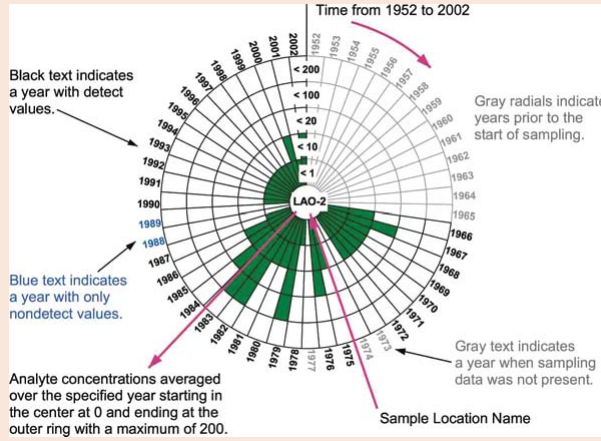
Figure 12.14 North America Soils (Mexico) – Dominant Surface Soil Texture (Courtesy: ESRI)

Applications Box **12.4**

Multivariate mapping of groundwater analyte over time

An increasing number of GIS users need to represent time dependent (temporal) data on maps in a meaningful and concise manner. The ability to analyze data for both spatial and temporal patterns in a single presentation provides a powerful incentive to develop tools for communication to map users and decision makers. Although there are many different methods to display scientific data to help discern either spatial or temporal trends, few visualization software packages allow for a single graphical presentation within a geographic context.

Chemical concentration data for many different groundwater monitoring wells is widely available for one or more chemical



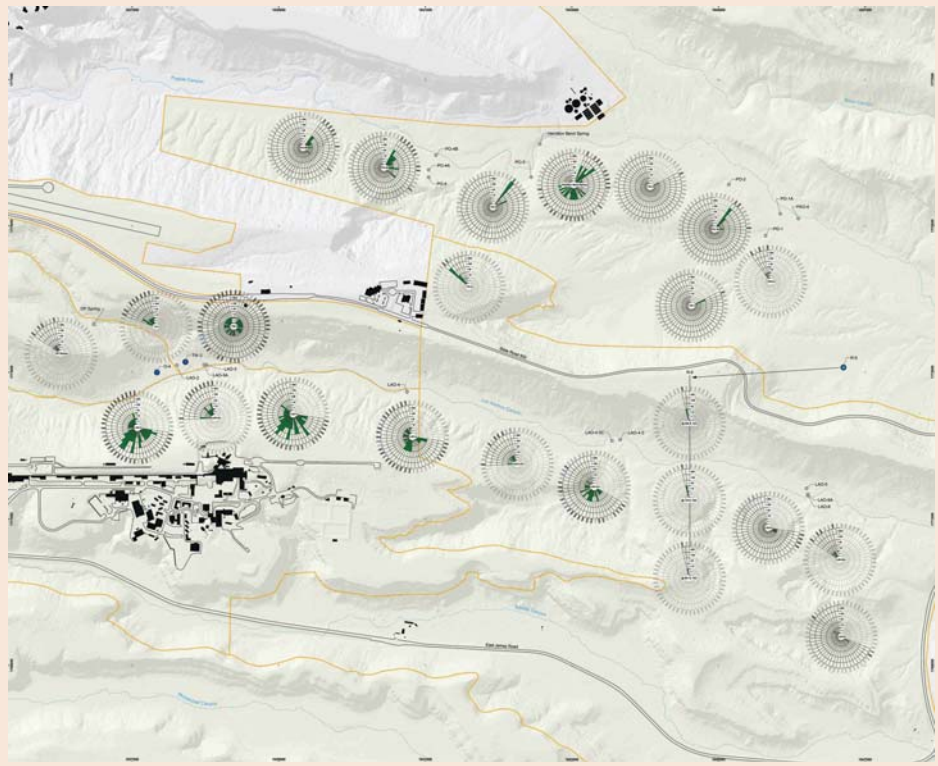


Figure 12.15 Representing temporal data as clock diagrams on maps (Courtesy: ESRI)

or analyte samples over time. Figure 12.15 shows how this complex temporal data can be represented on a map so that the movement and change in concentration of the analyte can be observed, both horizontally and vertically, and the potential for the analyte reaching the water table after many years can be assessed.

This map is innovative because it uses a clock diagram code for temporal and spatial visualization. The clock diagrams are analogous to 'rose diagrams' often used to depict wind direction in meteorological maps, or strike direction in geologic applications. The clock diagram method consists of three steps: (1) sample location data, including depth, is compiled along with attribute data such as concentration and time; (2) all data are read into the software application and clock diagram graphics are created; and (3) clock diagram graphics are placed on the map at the location of the well where samples were collected.

Although this method of representing temporal data through the use of clock diagrams is not entirely original, the use of multiple clock diagrams as GIS symbology to emphasize movement of sampled data over time is a new concept. The clock diagram code and method have been proven to be an effective means in identifying temporal and spatial trends when analyzing groundwater data.

12.4 Map series

The discussion thus far has focused on the general principles of map design that apply equally to a single map or a collection of maps for the same area or a common theme being considered over multiple areas. The real power of GIS-based digital cartography is revealed when changes need to be made to maps or when collections of similar maps need to be created. Editing or copying a map composition to create similar maps of any combination of areas and data themes is relatively straightforward with a GIS.

Many organizations use GIS to create collections or series of topographic or thematic maps. Examples include a topographic map series to cover a state or country (e.g., USGS 1:24 000 quad sheets of the United States or Eurogeographics 1:250 000 maps of Europe), an atlas of reference maps (e.g., *National Geographic Eighth Edition Atlas of the World*: Box 12.5), a series of geology maps for a country (Box 12.3), or land parcel (Figure 12.16)

(A)

Figure 12.16 Page from the Town of Brookline Tax Assessor's map book. Each map book page uses a common template (as shown in the legend), but has a different map extent: (A) page 26; (B) legend (Courtesy: ESRI)

(B)

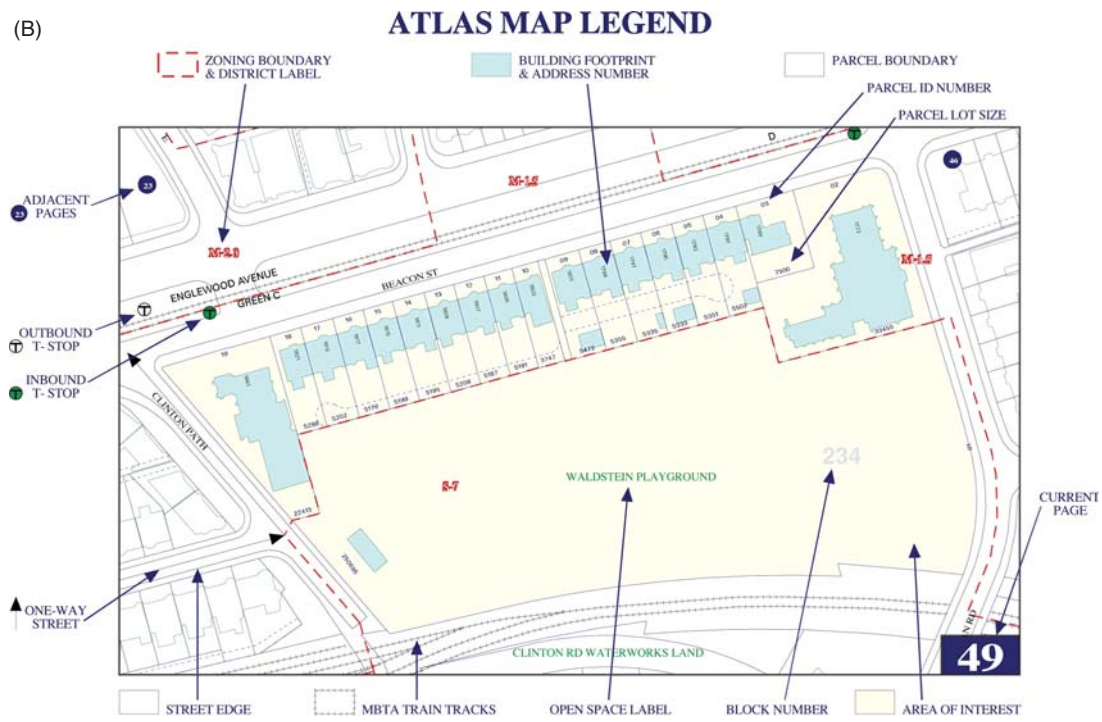


Figure 12.16 (*continued*)

Biographical Box **12.5**

Allen Carroll, Chief Cartographer, National Geographic Society

Figure 12.17 Allen Carroll (left), Chief Cartographer, National Geographic Society

Allen Carroll is chief cartographer and executive vice president of National Geographic Maps. As chief cartographer, Allen presides over the editorial and creative efforts of the Society's map division, including the renowned supplement maps published in *National Geographic* magazine, the *Seventh Edition Atlas of the World*, and the *National Geographic Map Machine*, an innovative world atlas on the Internet. He has been an employee of the National Geographic Society for over twenty years, serving in a variety of positions in the map division and the art department of *National Geographic* magazine.

After graduating from Connecticut College, Allen self-trained in design, illustration, and cartography. He started work as a freelance illustrator and designer, serving clients such as *The Washington Post*, Smithsonian Institute, and *Reader's Digest*. In 1991, he moved to the National Geographic Society, initially as art director of *National Geographic* magazine, producing historical, scientific, and informational artwork, then as managing director of National Geographic Maps where he presided over the shift of the unit from a division of the magazine to the Society's new subsidiary, National Geographic Ventures. During this time NG Maps expanded from its traditional role as service provider to the magazine and book divisions of the Society to become a publisher and distributor of map products. With partnerships and acquisitions, the group extended its product lines to include road atlases, road maps, and outdoor recreation maps.

Speaking about his work Allen Carroll says:

> **My colleagues and I feel extremely fortunate to have inherited *National Geographic*'s ninety-year tradition of superb cartography. I feel doubly blessed because I happen to be chief cartographer at the very moment that**

geographic science and technology is exploding. We in National Geographic Maps are particularly aware of this explosion, having just completed preparation of our *Eighth Edition Atlas of the World*. GIS has vastly improved production of the atlas and maintenance of the data behind it – although building an atlas is still a daunting, labor-intensive task. Our new atlas features breathtaking geographic data of many types and from many sources: satellite imagery, population density, elevation, ocean dynamics, paleogeography, tectonics, protected lands, ecoregions, and more.

I'm constantly surprised and delighted by the contributions of GIS professionals throughout the public and private realm. At the same time I'm concerned about the degree to which the fruits of their labor are sequestered. So much of the GIS data that has been created for specialized uses could benefit the public by being aggregated, interpreted, and distributed to a broader audience. The new atlas, and the Map Machine interactive atlas on the Web, are small steps in the right direction. I'm hoping that we at *National Geographic*, and others throughout the GIS community, can find many other ways to show our data to the world. At the local level, publishing zoning information and other GIS data produced by municipal agencies can help spur citizen activism. Globally, the remarkable efforts of conservation organizations to map biodiversity and protect critical habitats can educate schoolchildren and enlist support for protection of our natural heritage.

In 1888 the National Geographic Society was founded 'to increase and diffuse geographic knowledge'. In the 21st century, as geographic knowledge blossoms in ways unimagined a century ago, we can – and must – diffuse that knowledge to the greatest extent possible.

or utility asset maps that might be bound into a map book. Map series by definition share a number of common elements (for example, projection, general layout, and symbology) and a number of techniques have been developed to automate the map series production process.

GIS makes it much easier to create collections of maps with common characteristics from map templates.

Figure 12.4 showed the main GIS sub-components and workflows used to create maps. In principle, this same process can be extended to build production flow-lines that result in collections of maps such as an atlas, map series, or map book. Such a GIS requires considerable financial and human resource investments, both to create and maintain it, and it will be used by many operatives over a long period of time. In an ideal system, an organization would have a single 'scale-free', continuous database, which is kept up to date by editing staff, and capable of feeding multiple flow-lines at several scales each with different content and symbology. Unfortunately, this vision remains elusive because of several significant scientific and technical issues. Nevertheless, considerable progress has been made in recent years towards this goal.

The heart of map production through GIS is a geographic database covering the area and data layers of interest (see Chapter 10 for background discussion about the basic principles of creating and maintaining geographic databases). This database is built using a data model that represents interesting aspects of the real world in the GIS (see Chapter 8 for background discussion about GIS data models). Such a base cartographic data model is often referred to as a Digital Landscape Model (DLM) because its role is to represent the landscape in the GIS

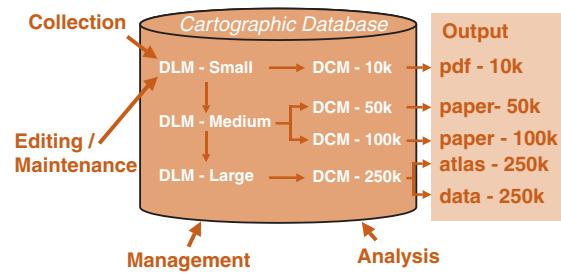


Figure 12.18 Key components and information flows in a GIS map production system

as a collection of features that is independent of any map product representation. Figure 12.18 is an extended version of Figure 12.4 that shows the key components and flows in a map production GIS.

The DLM continuous geographic database is usually stored in a Database Management System (DBMS) that is managed by the GIS. The database is created and maintained by editors that typically use sophisticated GIS desktop software. These must run carefully tailored applications that enforce strict data integrity rules (e.g., geometric connectivity, attribute domains, and topology – see Section 10.7) to maintain database quality.

In an ideal world, it will be possible to create multiple different map products from this DLM database. Example products that might be generated by a civilian or military national mapping agency include: a 1:10 000-scale topographic map in digital format (e.g., Adobe Portable Document Format (pdf)), a 1:50 000 topographic map

sheet in paper format, a 1:100 000 paper map of major streets/highways, a 1:250 000 map suitable for inclusion in a digital or printed atlas, and a 1:250 000 digital database in GML (Geography Markup Language) digital format. In practice, however, for reasons of efficiency and because cartographic database generalization is still not fully automated, a series of intermediate data model layers are employed (see Section 3.8 for a discussion of generalization). The base, or fine-scale, DLM is generalized by automated and manual means to create coarser-scale DLMs (in Figure 12.18 two additional medium- and coarse-scale DLMs are shown). From each of these DLMs one or more cartographic data models (digital cartographic models (DCMs)) can be created that derive cartographic representations from real-world features. For example, for a medium-scale DCM representation a road centerline network can be simplified to remove complex junctions and information about overpasses and bridges because they are not appropriate to maps at this scale.

Once the necessary DCMs have been built the process of map creation can proceed. Each individual map will be created in the manner described in Section 12.3, but since many similar maps will be required for each series or collection, some additional work is necessary to support map creation. Many similar maps can be created efficiently from a common map template that includes any material common to all maps (for example, inset/overview maps, titles, legends, scales, direction indicators, and map metadata; see Section 12.3.1 and Figure 12.8). Once a template has been created it is simple to specify the individual map-sheet content (e.g., map data from DCM data layers, specific title, and metadata information) for each separate sheet. The geographic extent of each map sheet is typically maintained in a separate database layer. This could be a regular grid (Figure 12.7) but commonly there is some overlap in sheets to accommodate irregular-shaped areas and to ensure that important features are not split at sheet edges. An automated batch process can then 'stamp out' each sheet from the database as and when required. Finally, it is convenient to generate an index of names (for example, places or streets), if the sheets are to form a map book or atlas, that shows on which sheet they are located.

12.5 Applications

Obviously there is a huge number and range of applications that use maps extensively. The goal here is to highlight a few examples that raise some interesting cartographic issues.

The relative importance of representing space and attributes will vary within and between different applications, as will the ability to broker improved measures of spatial distributions through integration of ancillary sources. These tensions are not new. In a topographic map, for example, the width of a single-carriageway road may be exaggerated considerably (Figure 12.19). This is done in order to enhance the legibility of features that are

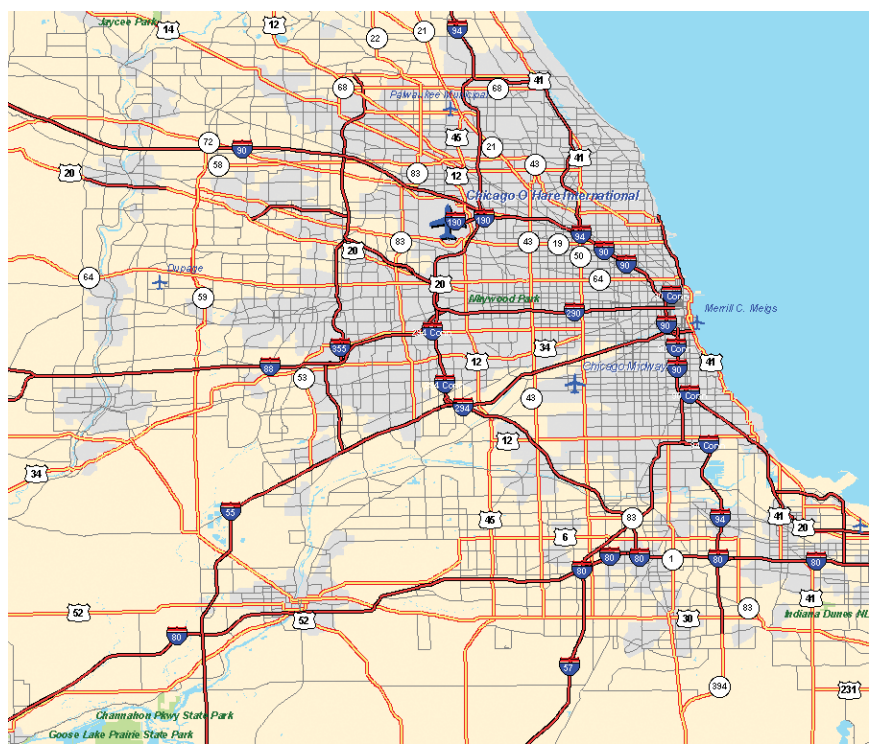


Figure 12.19 Topographic map showing street centerlines from a GDT database for the area around Chicago, USA. The major roads are exaggerated with 'cased' road symbols

central to general-purpose topographic mapping. In some instances, these prevailing conventions will have evolved over long periods of time, while in others the new-found capabilities of GIS entail a distinct break with the past. As a general rule, where accuracy and precision of geo-referencing are important, the standard conventions of topographic mapping will be applied (e.g., Figure 12.1).

Utility applications use GIS that have come to be known as Automated Mapping and Facilities Management (AM/FM) systems. The prime objectives of such systems are asset management, and the production of maps that can be used in work orders that drive asset maintenance projects. Some of the maps produced from an AM/FM system use conventional cartographic representations, but for some maintenance operations schematic representations are preferred (Figure 12.20). A schematic map provides a view of the way in which a system functions, and is used for operational activities such as identifying faults during power outage and routine pipeline maintenance. Hybrid schematic and geographic maps, known as geoschematics, combine the synoptic view of the current state of the network as a schematic superimposed upon a background map with real-world coordinates.

Transportation applications use a procedure known as linear referencing to visualize point (such as street 'furniture', e.g., road signs), linear (such as parking restrictions and road surface quality measures), and continuous events (such as speed limits). In linear referencing, two-dimensional geography is collapsed into one-dimensional linear space. Data are collected as linear measures from the start of a route (a path through a network). The linear measures are added at display time, usually dynamically, and segment the route into smaller sections (hence the term 'dynamic segmentation', which is sometimes used to describe this type of model). Figure 12.21 is a map of Pitt County, USA, that shows road surface (pavement) quality rated on a 100 point scale. The data were collected as linear events by driving the roads with an in-vehicle sensor. For more information on linear referencing see Section 5.4 and 8.2.3.3.

We began this chapter with a discussion of the military uses of mapping, and such applications also have special cartographic conventions – as in the operational overlay maps used to communicate battle plans (Figure 12.22). On these maps, friendly, enemy, and neutral forces are shown in blue, red and green, respectively. The location, size, and capabilities of military units are depicted with special

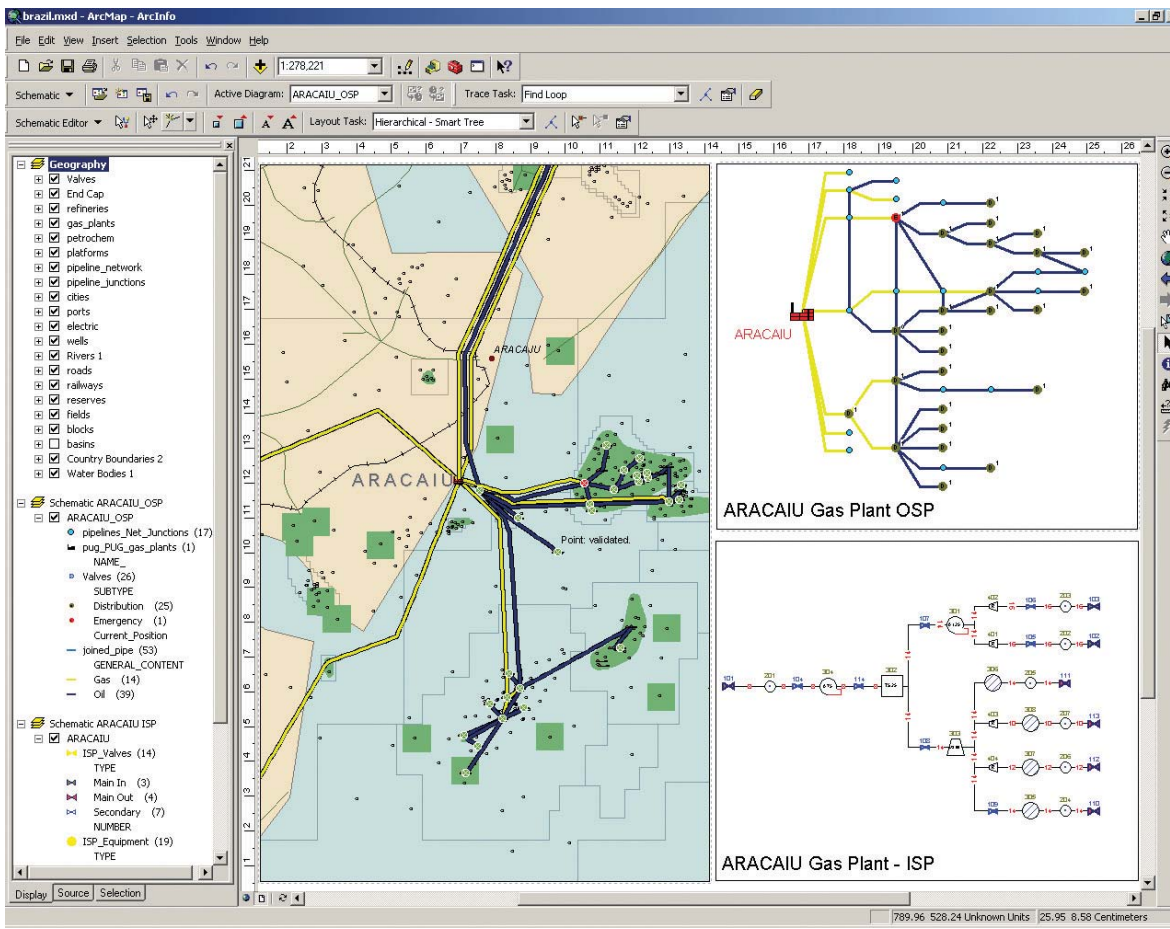


Figure 12.20 Schematic gas pipeline utility maps for Aracaiu, Brazil. Geographic view (left), Outside Plant schematic (top right), Inside Plant schematic (bottom right). Data courtesy of IHS-Energy/Tobin

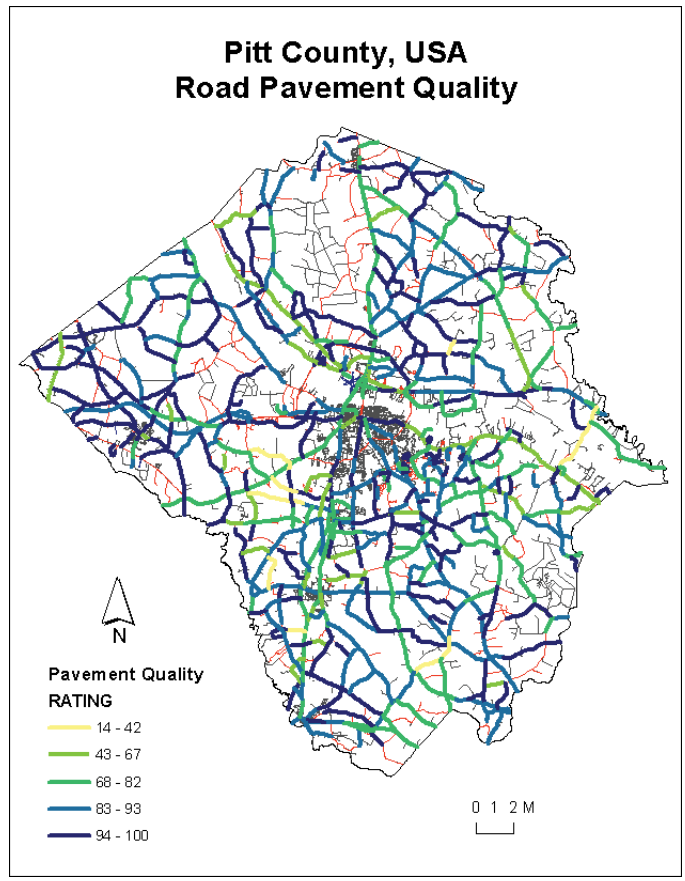


Figure 12.21 Linear referencing transportation data (pavement quality) displayed on top of street centerline files

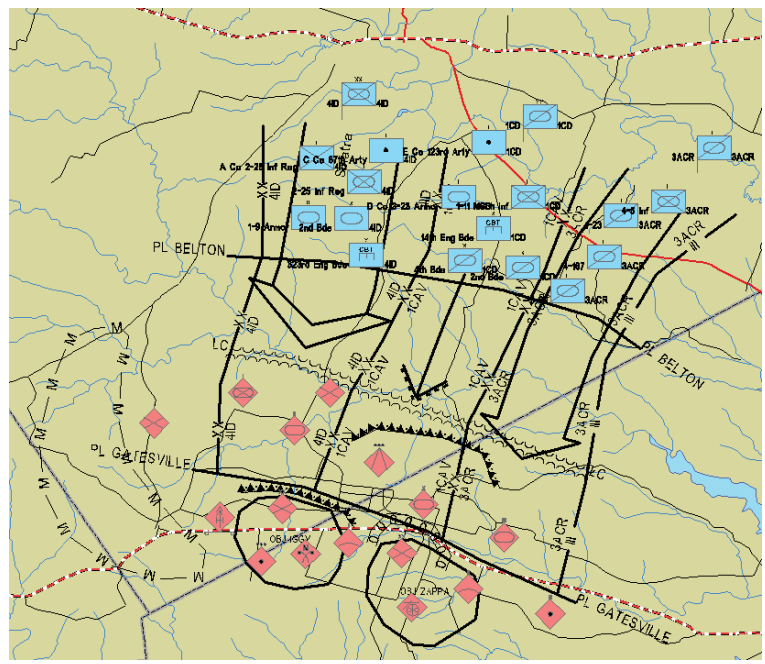


Figure 12.22 Military map showing 'overlay' graphics on top of base map (military vector map level 2) data. Blue rectangles are friendly forces units, red diamonds are enemy force units, black symbols are military operations – maneuvers, obstacles, phase lines, and unit boundaries. These 'tactical graphics' depict the battle space and the tactical environment for the operational units

multivariate symbols that have operational and tactical significance. Other subjects of significance, such as mine fields, impassable vegetation, and direction of movement, also have special symbols. Animations of such maps can be used to show the progression of a battle, including future 'what if?' scenarios.

12.6 Conclusions

Cartography is both an art and a science. The modern cartographer must also be very familiar with the application of computer technology. The very nature of cartography and map making has changed profoundly in the past few decades and will never be the same again. Nevertheless, there remains a need to understand the nature and representational characteristics of what goes into maps if they are to provide robust and defensible aids to decision making, as well as tactical and operational support tools. In cartography, there are few hard and fast rules to drive map composition, but a good map is often obvious once complete. Modern advances in GIS-based cartography make it easier than ever to create large numbers of maps very quickly using automated techniques once databases and map templates have been built. Creating databases and map templates continue to be advanced tasks requiring the services of trained professionals. The type of data that are used on maps is also changing – today's maps often reuse and recycle different datasets, obtained over the Internet, that are rich in detail but may be unsystematic in collection and incompatible in terms of scale. This all underpins the importance of metadata to evaluate datasets in terms of scale, aggregation, and representativeness prior to mapping. Collectively these changes are driving the development of new applications founded on the emerging advances in scientific visualization that will be discussed in the next chapter.

Questions for further study

1. Identify the criteria that you would have used in designing a military mapping system for use during the 2003 Iraq war. How might the system subsequently have been adapted for use in humanitarian relief?
2. Why are there so many classification schemes for area-based choropleth maps? What criteria would you use to select the best scheme?
3. Using Table 12.1 as a guide, create an equivalent table showing example maps for all the map types.
4. Using a GIS of your choice, create a map of your local area containing all the seven elements listed in Section 12.3.1. What makes this a good map?

Further reading

Dent B.D. 1999 *Cartography: Thematic Map Design* (5th edn). Dubuque, Iowa: WCB/McGraw-Hill.

Kraak M.-J. and Ormerling F. 1996 *Cartography: Visualization of Spatial Data*. Harlow: Longman.

Robinson A.H., Morrison J.L., Muehrcke P.C., Kimerling A.J. and Guptill S.C. 1995 *Elements of Cartography* (6th edn). New York: Wiley.

Slocum T.A. 1999 *Thematic Cartography and Visualization*. Upper Saddle River, NJ: Prentice-Hall Inc.

Tufte E.R. 2001 *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

13 *Geovisualization*

This chapter describes a range of novel ways in which information can be presented visually to the user. Using techniques of geovisualization, GIS provides a far richer and more flexible medium for portraying attribute distributions than the paper mapping which is covered in Chapter 12. First, through techniques of spatial query, it allows users to explore, synthesize, present (communicate), and analyze the meaning of any given representation. Second, it facilitates transformation of representations using techniques such as cartograms and dasymetric mapping. Third, GIS-based geovisualization allows the user to interact with the real world from a distance, through interaction with and even immersion in artificial worlds. Together, these functions broaden the user base of GIS, and have implications for public participation in GIS (PPGIS).

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

At the end of this chapter you will understand:

- How GIS affects visual communication;
- The ways in which good user interfaces can help to resolve spatial queries;
- Some of the ways in which GIS-based representations may be transformed;
- How 3-D geovisualization and virtual reality improve our ability to understand the world.

13.1 Introduction: uses, users, messages, and media

Effective decision support through GIS requires that the messages of representations (in the senses defined in Chapter 3) are readily interpretable in the minds of decision makers. Historically, many of the early uses of mapping were to support military operations (Section 12.1). The historic map shown in Figure 13.1, and the case study to which it relates (Box 13.1), illustrates how geovisualization has been used as an indirect tool of warfare. Today's GIS applications environment is immeasurably more data rich than in the past, and a major goal of GIS usage is often to make sense of the huge archive resources that are available, without creating information overload. In some respects at the opposite extreme, historical GIS applications are concerned with assembling together digital shards of evidence that may be scattered across space and time. There are commonalties of interest that link data rich and data poor applications, as Box 13.2 (and also Box 6.1) illustrates.

It is through mapping that the meaning of a spatial representation of the real world is communicated to users. Historically, the paper map was the only available interface between the map-maker and the user: it was permanent, contained a fixed array of attributes, was of predetermined and invariant scale, and rarely provided any quantitative or qualitative indications as to whether it was safe to use (Figure 13.3: see also Section 12.2). These attributes severely limit the usefulness of the paper map in today's applications environment, which can be of seemingly unfathomable complexity, and entails

Applications Box **13.1**

Geovisualization and Geopolitics

In the early 20th century the Oxford, UK, geographer Halford Mackinder proposed a theory that different national groupings organized around their own historic settlement 'heartlands'. He worked at a time when the development of railroads and roads for troop movement was opening up previously inaccessible areas of western Russia, and he envisioned this territory as a new heartland at the center of an increasingly integrated world. In *geopolitical* terms, he saw western Russia as the center of the new joined-up world, and believed that whoever occupied it would come to dominate the emerging 'world island'. Subsequently, the German geographer Friedrich Ratzel developed the concept of the 'heartland' into that of 'Lebensraum', the essence of which was that every different cultural group needed to develop its own geographic living space. In the years immediately prior to World War Two (1939–45), these related ideas of geographic heartland and Lebensraum were developed by General Haushofer, Professor of Geography at the University of Munich. The spatial extent of the Lebensraum of the German national grouping (the German Volk), he argued, had been overly constrained and dissected by the Treaty of Versailles at the end of World War One. Crudely, Haushofer mapped the spatial distribution of 'German-speaking people' (Figure 13.1) to substantiate this thesis, and implied that there was a need to extend the German cultural realm deep into Russia. This argument was subsequently used by the Nazi Party to justify German incursions into Czechoslovakia, the annexation of Austria, and the invasion of Poland that ultimately triggered World War Two. The extensive use of mapping in setting out this geopolitical thesis (termed Geopolitik) provides early and powerful evidence of the ways in which maps can be used as propaganda. Haushofer's maps portray information in order to garner support for a particular worldview and subsequent course of decision making. They illustrate the general point that the map creator is very much the arbiter and architect of what the map portrays. In Haushofer's case, the message of the map ultimately contributed to a pseudo-scientific justification of a worldview that history would rather forget – yet it is periodically rekindled in various guises by extremist groups.

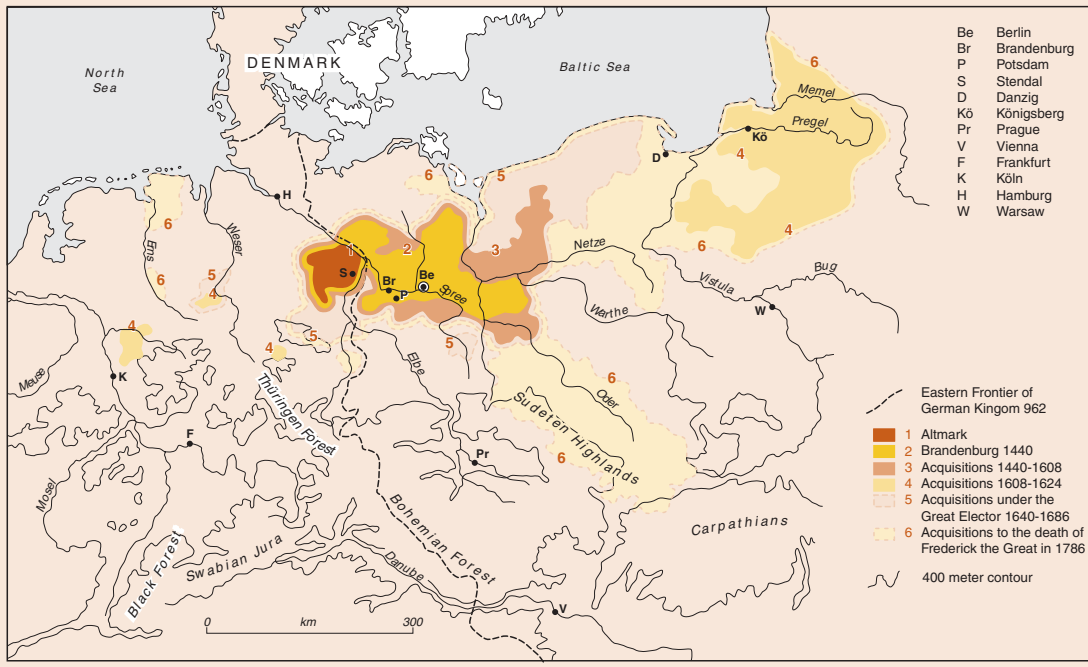


Figure 13.1 The geopolitical center of Germany. (*Source*: Parker G, 1988. Geopolitics: past, present, future. London: Pinter.)

Biographical Box **13.2**

Kevin Schürer, historical demographer and digital data archivist

Kevin Schürer has been fascinated by the past and by maps for as long as he can remember. The young child who was passionate about understanding history, especially through visiting castles and museums, was equally intrigued by atlases and images of far-off places. These dual interests led to his studying both history and geography at London and subsequently as a member of the Cambridge Group for the History of Population and Social Structure at Cambridge University, UK. Kevin recalls: 'In the 1980s the Group provided an exciting mix of historians, geographers, statisticians, computer scientists, with, from time to time, the odd anthropologist or sociologist thrown in for good measure. Each brought their own blend of research skills to problem solving. To me this was ahead of its time as a way of conducting cutting-edge inter-disciplinary team research.'

Today, Kevin is Director of the UK Data Archive (UKDA) at the University of Essex. This is the oldest and largest social science data archive in the UK. It distributes some 18 000 non-census datasets and associated metadata (Section 11.2.1) to some 2000 university-based researchers each year, as well as Census of Population data to over 10 000 registered researchers, teachers, and students. Reflecting on his role, Kevin says:

Figure 13.2 Kevin Schürer, historical demographer and digital data archivist

Providing a service which acts as both provider of and custodian to the major collection of social science research and teaching data in the UK has allowed me to witness first hand the huge impacts that technology, methodology, and geovisualization have had in recent years. Although some might argue that quantitative-focused research is unpopular and no longer in fashion, researchers are year on year using larger amounts of data. In some cases, the scale and complexity of analysis is at a level and requires computing power that could only have been dreamt of even ten years ago. A parallel development has been the increased ease by which data can be viewed via Web-based systems and be migrated between software packages. All of this has made the visualization of complex statistical analyses a greater prospect than previously. Yet, of course, problems and challenges still exist. In particular, bringing data together from several diverse sources or surveys still remains difficult and in some cases impossible. All of which points to the importance of the spatial dimension as a linking mechanism for visualizing diverse datasets from different sources.

In comparing past and present spatial distributions of surnames, Kevin's own research demonstrates why historians and data archivists each need maps (see Box 6.1).

visualization of data which are richer, continuously updated, and scattered across the Internet.

Geovisualization builds on the established tenets of map production and display. Geographer and geovisualization expert Alan MacEachren (Box 13.3) describes it as the creation and use of visual representations to facilitate thinking, understanding, and knowledge construction about human and physical environments, at geographic scales of measurement. Geovisualization is a research-led field that integrates approaches from visualization in scientific computing (ViSC), cartography, image analysis, information visualization, exploratory data analysis (EDA), as well as GIS. Its motivation is to develop theories, methods, and tools for visual exploration, analysis, synthesis, and presentation of geospatial data. The December 2000 research agenda of the International Cartographic Association's Commission on Visualization and Virtual Environments is shown in Table 13.1.

Table 13.1 Research agenda of the International Cartographic Association's Commission on Visualization and Virtual Environments (adapted from **www.geovista.psu.edu/sites/icavis/agenda/index.html**)

1. Cognitive and usability issues in geovisualization
   - Virtual geographic environments and 3-D geovisualization.
   - Dynamic representations (including animated and interactive mapping).
   - Improved interface design.
   - Understanding individual and group interaction using geovisualization.
   - Collaborative visualization.
   - Better evaluation of existing methods.
2. Representation and its relationship with cartographic visualization
   - Existing and potential roles of map displays.
   - Understanding the techniques that map producers use to visualize spatial data.
   - Understanding how technology might improve cartographic practice.
   - The relationships between user cognition, map design, and the nature of spatial data.
3. Uses of geovisualization in knowledge discovery
   - Visual approaches to knowledge discovery.
   - Visual support for knowledge discovery through geocomputation.
   - Use of databases and data models to support knowledge discovery.
4. User interface issues for geovisualization
   - Interface development and usage.
   - Improving usability and navigability of geovisualizations.
   - Improving facilities for user collaboration.

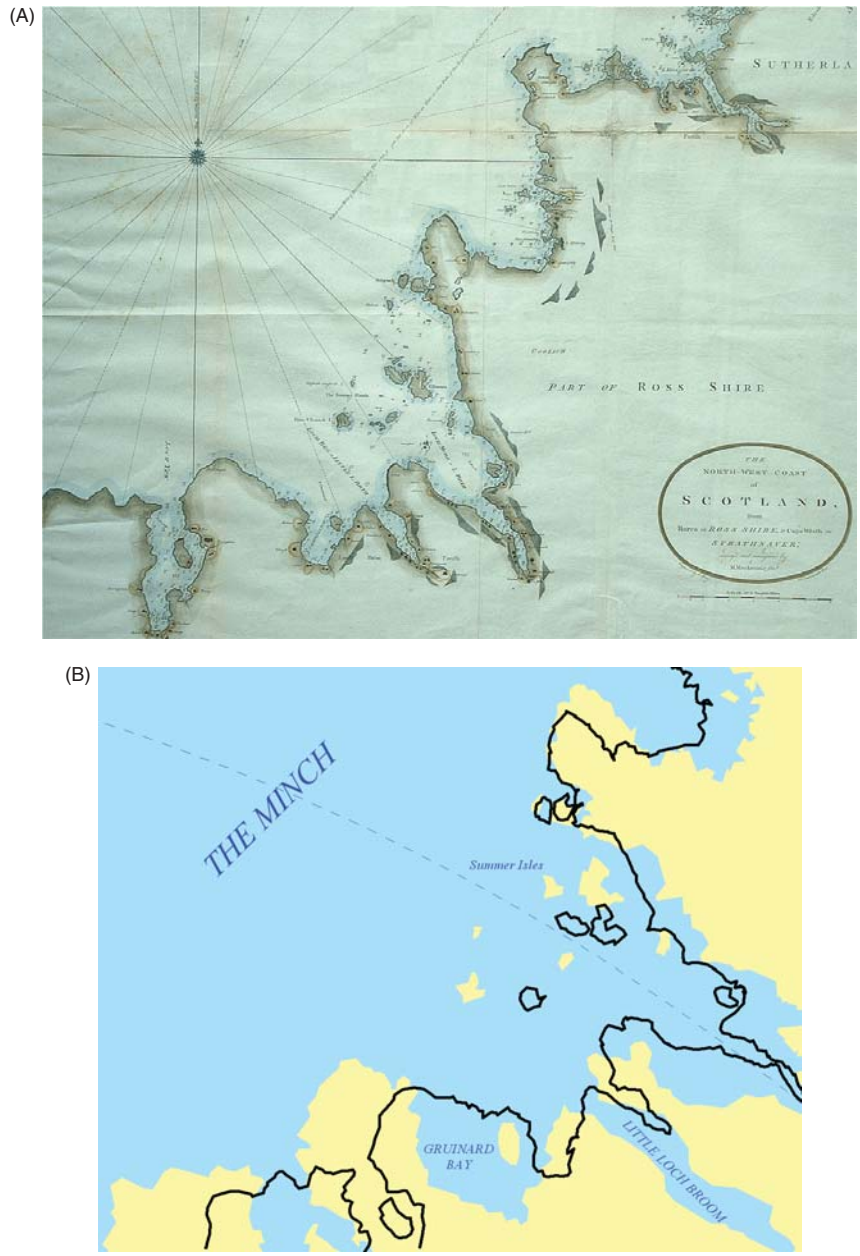


Figure 13.3 Part of the coastline of northwest Scotland (A) according to the hydrographer Murdoc Mackenzie in 1776 (orientation to magnetic north at the time) and (B) Modern day outline of the same coast with trace of Mackenzie chart overlaid in black. (Reproduced by permission of Michael J. de Smith)

As such, today's geovisualization is much more than conventional map design. It has developed into an applied area of activity that leverages geographic data resources to meet a very wide range of scientific and social needs, as well as into a research field that is developing new visual methods and tools to facilitate better user interaction with such data. In this chapter, we discuss how this is achieved through *query* and *transformation* of geographic data, and user *immersion* within them. Query and transformation are discussed in a substantially different context in Chapter 14, where they are encountered as two of six types of spatial analysis.

13.2 Geovisualization and spatial query

13.2.1 Overview

Fundamental to effective geovisualization is an understanding of how human cognition shapes GIS usage,

Biographical Box **13.3**

Alan MacEachren, geographer and geovisualization expert

Alan MacEachren (Figure 13.4) is a Professor of Geography at Pennsylvania State University. He began by studying an eclectic range of disciplines from botany through to experimental psychology at Ohio University before settling on a major in Geography, and next undertook a Masters and Ph.D. in Geography at the University of Kansas. There, he became interested in environmental perception in general and the cognitive aspects of map symbolization in particular. Following his move to Penn State in 1985, he developed interests in the relationships between cognitive science, GIS, and GIScience with an emphasis on visual representation of geographic information. Today he continues to develop a range of geovisualization applications and explores the potential of technology to enable immersive virtual environments and geocollaboration.

He acts as Director of the GeoVISTA Center at the University. Center activities have played a leading role in the development of geovisualization, allied to other developments in GIScience. The Center is avowedly interdisciplinary in addressing two goals: (a) to develop a deep understanding of human interaction with geographic information and technologies; and (b) to use that understanding to develop powerful human-centered scientific methods to solve social, environmental, and other problems. Central to these objectives is the use of computer-supported, visually-enabled exploration and analysis of an ever-expanding range of geographic data.

For Alan, the essence of geovisualization is leveraging the rapid advances in information technology to enable creative thinking. He sees better understanding of how people use dynamic visually enabled representations as key to designing and implementing more effective forms of representation. He says:

> **Most research in geovisualization over the past decade has focused on enabling science. While this remains an important application domain, and there are many unanswered questions, it is time for geovisualization research to take a broader perspective. We need to extend, or reinvent, our geovisualization methods and tools so that they can support real-world decision making. Two particular challenges stand out. First, geovisualization methods and tools must support the work of teams much more directly, with the visual display acting not just as an object of collaboration but as a means to enable dialogue and coordinate joint action. Second, geovisualization must move out of the laboratory to support decision making in the field (e.g., for crisis response).**

Alan's hobbies include birdwatching: identifying birds by jizz (the art of seeing a bird badly but still knowing what it is) involves similar skills to geovisualization, such as the need to identify phenomena that may be difficult to categorize.

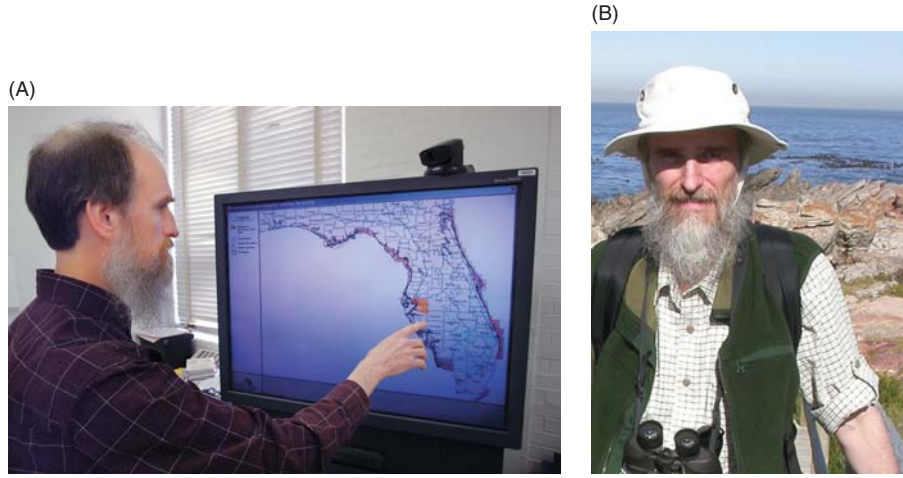


Figure 13.4 Alan MacEachren: (A) geographer and geovisualization expert; and (B) birder

how people think about space and time, and how spatial environments might be better represented using computers and digital data. The conventions of map production, presented in Section 12.3, are central to the use of mapping as a decision support tool. Many of these conventions are common to paper and digital mapping, although GIS allows far greater flexibility and customization of map design. Geovisualization develops and extends these concepts in a number of new and innovative ways.

Viewed in this context, there are four principal purposes of geovisualization:

1. *Exploration*: for example, to establish whether and to what extent the general message of a dataset is sensitive to inclusion or exclusion of particular data elements.
2. *Synthesis*: to present the range, complexity, and detail of one or more datasets in ways that can be readily assimilated by users. Good geovisualization should enable the user to 'see the wood for the trees'.
3. *Presentation*: to communicate the overall message of a representation in a readily intelligible manner, and to enable the user to understand the likely overall quality of the representation.
4. *Analysis*: to provide a medium to support a range of methods and techniques of spatial analysis.

This fourth function that geovisualization supports will be discussed in detail in Chapters 14 and 15.

These objectives cut across the full range of different applications tasks and are pursued by users of varying expertise that desire different degrees of interaction with their data (Figure 13.5). The tasks addressed by geovisualization range from information sharing to knowledge construction (see Tables 13.1 and 1.2); user groups range from lay-public users to GIS specialists (spatially aware professionals, or SAPs: Section 1.4.3.2); and the degree of interaction that users desire ranges from low (passive) to high.

Geovisualization allows users to explore, synthesize, present, and analyze their data more thoroughly than was possible hitherto.

Together these tasks may be considered in relation to the conceptual model of uncertainty that was presented in Figure 6.1, which is presented with additions in Figure 13.6. This encourages us to think of geographic analysis not as an end point, but rather as the start of an iterative process of feedbacks and 'what if?' scenario testing. The eventual best-available reformulated model is used as a decision support tool, and the real world is changed as a consequence. As such, it is appropriate to think of geovisualization as imposing a filter upon the results of analysis and on the nature of the feeds that are generated back to the conception and measurement of geographic phenomena. It can be thought of as an additional 'geovisualization filter' (U4 in Figure 13.6).

The most straightforward way in which reformulation and evaluation of a representation of the real world can take place is through posing *spatial queries* to ask generic spatial and temporal questions such as:

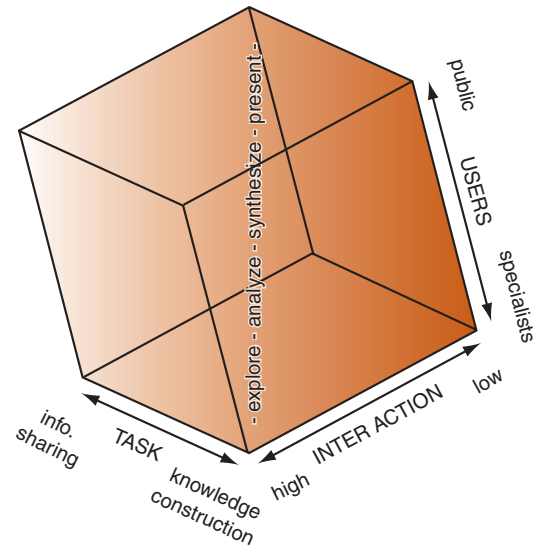


Figure 13.5 Functions of geovisualization (after MacEachren et al 2004) © IEEE 2004. Reproduced by permission.

- Where is . . .?
- What is at location . . .?
- What is the spatial relation between . . .?
- What is similar to . . .?
- Where has . . . occurred?
- What has changed since . . .?
- Is there a general spatial pattern, and what are the anomalies?

These questions are articulated through the graphical user interface (GUI) paradigm called a 'WIMP' interface – based upon **W**indows, **I**cons, **M**enus and **P**ointers (Figure 13.7). The familiar actions of pointing, clicking, and dragging windows and icons are the most common ways of interrogating a geographic database and summarizing results in map and tabular form. By extension, research applications increasingly use multiple displays of maps, bar charts, and scatterplots, which enable a picture of the spatial and other properties of a representation to be built up. These facilitate learning about a representation in a data-led way and are discussed in Section 14.2.

The WIMP interface allows spatial querying through pointing, clicking, and dragging windows and icons.

13.2.2 Spatial query and Internet mapping

Spatial query functions are also central to many Internet GIS applications. For many users, spatial query

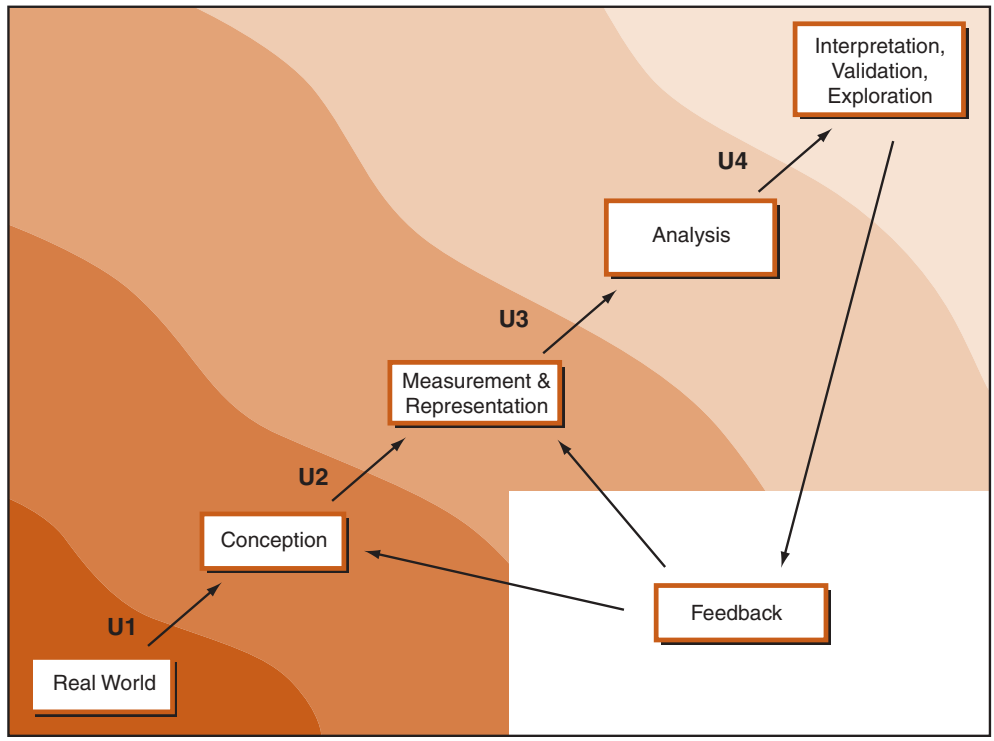


Figure 13.6 Filters U1–U4: conception, measurement, analysis, and visualization. Geographic analysis is not an end point, but rather the start of an iterative process of feedbacks and 'what if?' scenario testing. (See also Figure 6.1.)

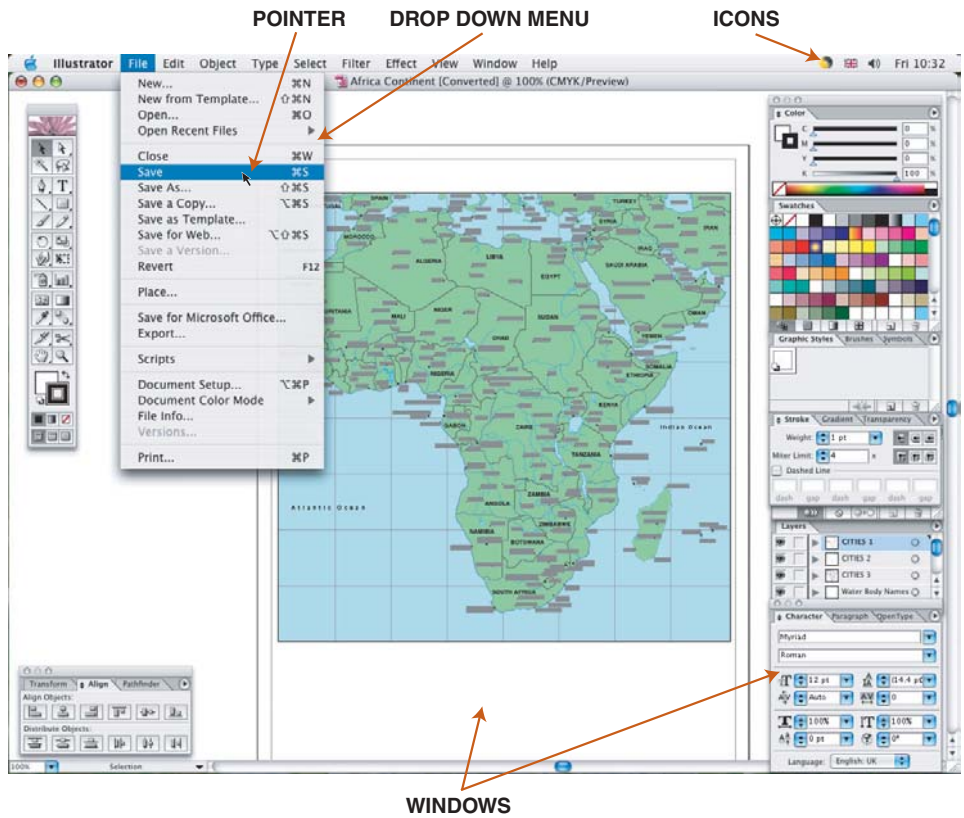


Figure 13.7 The WIMP (Windows, Icons, Menus, Pointers) interface to computing using Macintosh System 10

is the objective of a GIS application, as in queries about the taxable value of properties (Section 2.3.2.2), customer care applications about service availability (Section 2.3.3), or Internet site queries about traffic congestion (Section 2.3.4). In other applications, spatial query is a precursor to any advanced geographic analysis. Spatial query may appear routine, but entails complex operations, particularly when the objects of spatial queries are continuously updated (refreshed) in real time – as illustrated in the traffic application shown in Box 13.4 and on websites concerning weather conditions. Other common spatial queries are framed to identify the location of services, provide routing and direction information (see Figure 1.17), facilitate rapid response in disaster management (see Figure 2.16), and provide information about domestic property and neighborhoods to assist residential search (Figure 1.14).

13.3 Geovisualization and transformation

13.3.1 Overview

Chapter 12 illustrated circumstances in which it was appropriate to adjust the classification intervals of maps, in order to highlight salient characteristics of spatial distributions. We have also seen how transformation of entire scales of attributes can assist in specifying the relationship between spatial phenomena, as in the scaling relations that characterize a fractal coastline (Box 4.6). More generally still, our discussion of the nature of geographic data (Chapter 4) illustrated some of the potentially difficult consequences arising out of the absence of natural units in geography – as when choropleth map counts are not standardized by any numerical or areal base measure. Standard map production and display functions in GIS allow us to standardize according to numerical base categories, yet large but unimportant (in attribute terms) zones continue to accrue greater visual significance in mapping than is warranted. Such problems can be addressed by using GIS to transform the shape and extent of areal units. Our use of the term 'transformation' here is thus in the cartographic sense of taking one vector space (the real world) and transforming it into another (the geovisualization). Transformation in the sense of performing analytical operations upon coordinates to assess explicit spatial properties such as adjacency and contiguity is discussed in Section 14.4.

GIS is a flexible medium for the cartographic transformation of maps.

Some ways in which example real-world phenomena of different dimensions (Box 4.1) may or may not be transformed by GIS are illustrated in Table 13.2. These examples illustrate how the measurement and representation filter of Figure 13.6 has the effect of transforming

Applications Box **13.4**

Use of Internet GIS to manage traffic in Southampton, England

International port cities function as gateways to distant destinations for tourists, freight transporters, and a multitude of other travelers, and so they are often overwhelmed with unique vehicular traffic flow problems. Southampton and its surrounding area have been involved in the development and application of part of a pan-European, GIS-based traffic monitoring and analysis system called the ROad MANagement System for Europe (ROMANSE) for nearly ten years.

ROMANSE is based on Desktop GIS software linked to an Urban Traffic Control (UTC) system. The GIS acts as the graphical front end to the UTC. The GIS software has been customized to create a simplified, specific-purpose display. Basemaps originate from Great Britain's Ordnance Survey, onto which a variety of current traffic information is overlaid. The GIS creates on-the-fly status maps so that engineers at the Traffic and Travel Information Centre (TTIC) in Southampton can analyze existing road hazards and anticipate new ones. They can then issue appropriate advice and warning messages for immediate display on digital road signs in the area or dispatch emergency services.

The information processed at the TTIC comes from a variety of sources, including closed-circuit television, roadside detectors, and satellite tracking systems, all of which are fed into the UTC. It is then processed and integrated by the GIS for instant analysis or display, providing the engineers with a good overview of what is happening on the road network. Information available to the engineers includes traffic flow per minute, traffic speeds, congestion (Figure 13.8A), and car park occupancy (Figure 13.8B).

Internet GIS has been used to provide current traffic information and maps on demand to Internet users visiting the ROMANSE Web site at **www.romanse.org.uk**. From this site, members

of the public can check the current occupancy in car parks before traveling. A city-center map is displayed and by clicking on each car park it is possible to determine both the current occupancy status and what the occupancy is likely to be in the near future, based on historic records. Information for cyclists is also available at the website. Traffic information is also posted digitally on touch-screen displays at main transport interchanges, shopping centers, tourist information centers, and libraries. At bus stops, arrival times are electronically displayed and some stops feature an audio version of the information for visually impaired passengers.

This remains one of the most advanced intelligent transport information systems in the world. A key part of the success was linking GIS software to a real-time traffic control system for data collection, and to messaging systems and the Internet for data display.

Figure 13.8 Monitoring (A) traffic flow and (B) car park occupancy in real time as part of the Web-based ROMANSE traffic management system (Base maps: Crown Copyright **southampton.romanse.org.uk/**) (Reproduced by permission of Southampton City Council)

Table 13.2 Some examples of coordinate and cartographic transformations of spatial objects of different dimensionality (based on D. Martin 1996 *Geographic Information Systems: Socioeconomic Applications (Second edition)*. London, Routledge: 65)

Dimension	**0**	**1**	**2**	**3**
Conception	Population distribution	Coastline (see Box 4.6)	Agricultural field	Land surface
	Coordinate transformation of real world arising in GIS representation/measurement			
Measurement	Imposed areal aggregation, recorded as a sequence of digitized points	Sequence of digitized points	Digitized polygon boundary	Arrangement of spot heights
	Cartographic transformation to aid interpretation of representation			
Visualization	Cartogram or dasymetric map	Generalized line	Integral/natural area	Digital elevation model

some objects, like population distributions, but not others, like agricultural fields. Further transformation may occur in order to present the information to the user in the most intelligible and parsimonious way. Thus, cartographic transformation through the U3 filter in Figure 13.6 may be the consequence of the imposition of artificial units, such as census tracts for population, or selective abstraction of data, as in the generalization of cartographic lines (Section 3.8).

The standard conventions of map production and display are not always sufficient to make the user aware of the transformations that have taken place between conception and measurement – as in choropleth mapping, for example, where mapped attributes are required to take on the proportions of the zones to which they pertain. Geovisualization techniques make it possible to manipulate the shape and form of mapped boundaries using GIS, and there are many circumstances in which it may be appropriate to do so. Indeed, where a standard mapping projection obscures the message of the attribute distribution, map transformation becomes necessary. There is nothing untoward in doing so – remember that one of the messages of Chapter 5 was that all conventional mapping of geographically extensive areas entails transformation, in order to represent the curved surface of the Earth on a flat screen or a sheet of paper. Similarly, in Section 3.8, we described how generalization procedures may be applied to line data in order to maintain their structure and character in small-scale maps. There is nothing sacrosanct about popular or conventional representations of the world, although the widely used transformations and projections do confer advantages in terms of wide user recognition, hence interpretability.

13.3.2 Cartograms

Cartograms are maps that lack planimetric correctness, and distort area or distance in the interests of some specific objective. The usual objective is to reveal patterns that might not be readily apparent from a conventional map or, more generally, to promote legibility. Thus, the integrity of the spatial object, in terms of areal extent, location, contiguity, geometry, and/or topology, is made subservient to an emphasis upon attribute values or particular aspects of spatial relations. One of the best-known cartograms (strictly speaking a *linear cartogram*) is the London Underground map, devised in 1933 by Harry Beck to fulfill the specific purpose of helping travelers to navigate across the network. Figure 13.9A is an early map of the Underground, which shows the locations of stations in Central London in relation to the real-world pattern of streets. The much more recent central-area cartogram shown in Figure 13.9B is a descendent of Beck's 1933 map. It is a widely recognized representation of connectivity in London, and the conventions that it utilizes are well suited to the attributes of spacing, configuration, scale, and linkage of the London Underground system. The attributes of public transit systems differ between cities, as do the cultural conventions of transit users, and thus it is unsurprising that cartograms pertaining to transit systems elsewhere in the world appear quite different.

Cartograms are map transformations that distort area or distance in the interests of some specific objective.

A central tenet of Chapter 3 was that all representations are abstractions of reality. Cartograms depict transformed, hence artificial, realities, using particular exaggerations that are deliberately chosen. Figure 13.10 presents a regional map of the UK, along with an *equal population cartogram* of the same territory. The cartogram is a map projection that ensures that every area is drawn approximately in proportion to its population, and thus that every individual is accorded approximately equal weight. Wherever possible, areas that neighbor each other are kept together and the overall shape and compass orientation of the country is kept roughly correct – although the real-world pattern of zone contiguity and topology is to some extent compromised. When attributes are mapped, the variations within cities are revealed, while the variations in the countryside are reduced in size so as not to dominate the image and divert the eye away

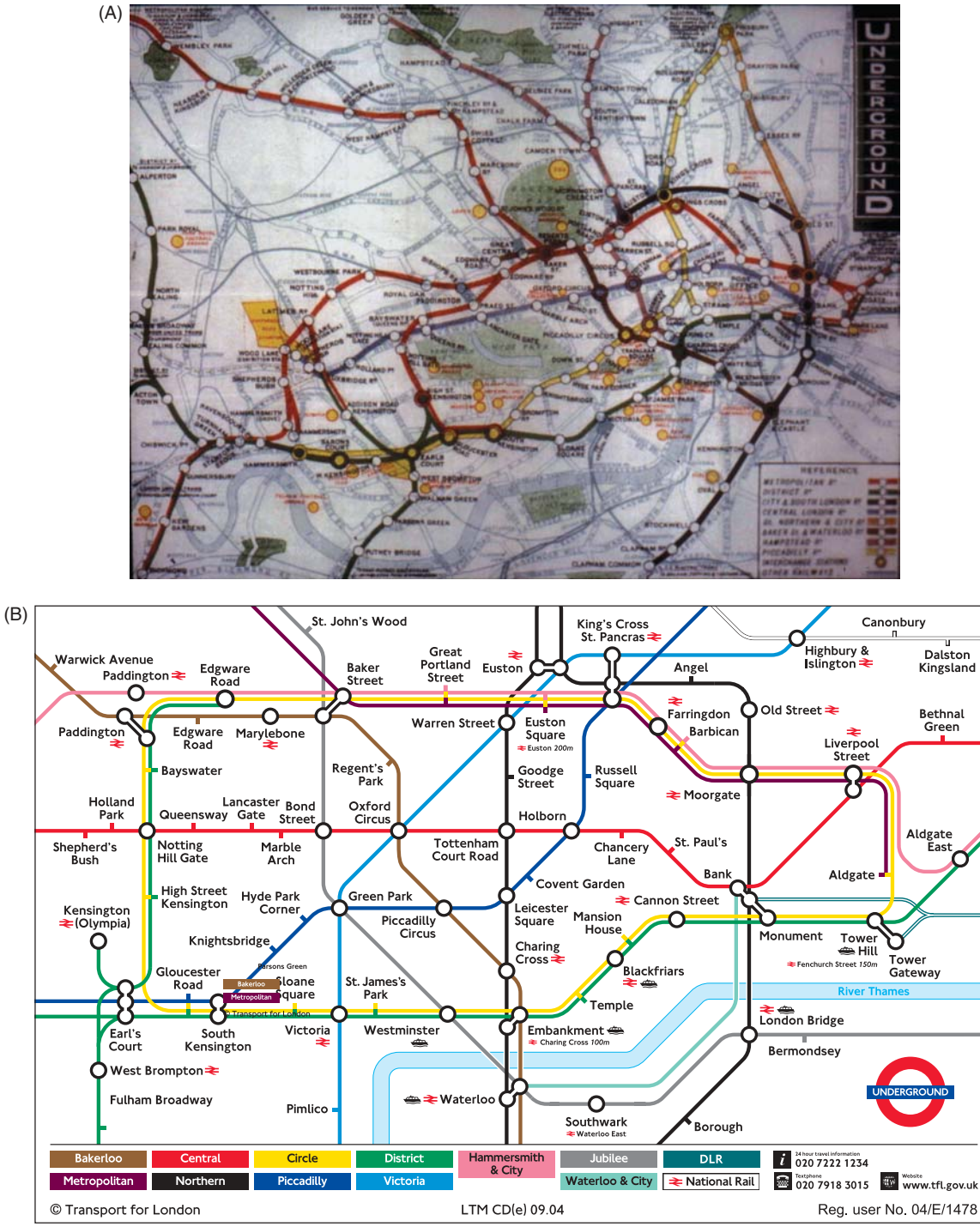


Figure 13.9 (A) A 1912 map of the London Underground and (B) its present day linear cartogram transformation

from what is happening to the majority of the population. Therefore, the transformed map of wealthy areas in the UK, shown in Figure 13.11B, depicts the considerable diversity in wealth in densely populated city areas – a phenomenon that is not apparent in the usual (scaled Transverse Mercator) projection (Figure 13.11A). There are semi-automated ways of producing cartograms, but complicated representations such as those shown in Figures 13.10B and 13.11B almost invariably entail human judgment and design.

Throughout this book, one of the recurrent themes has been the value of GIS as a medium for data sharing and this is most easily achieved if pooled data share common coordinate systems. Yet a range of cartometric

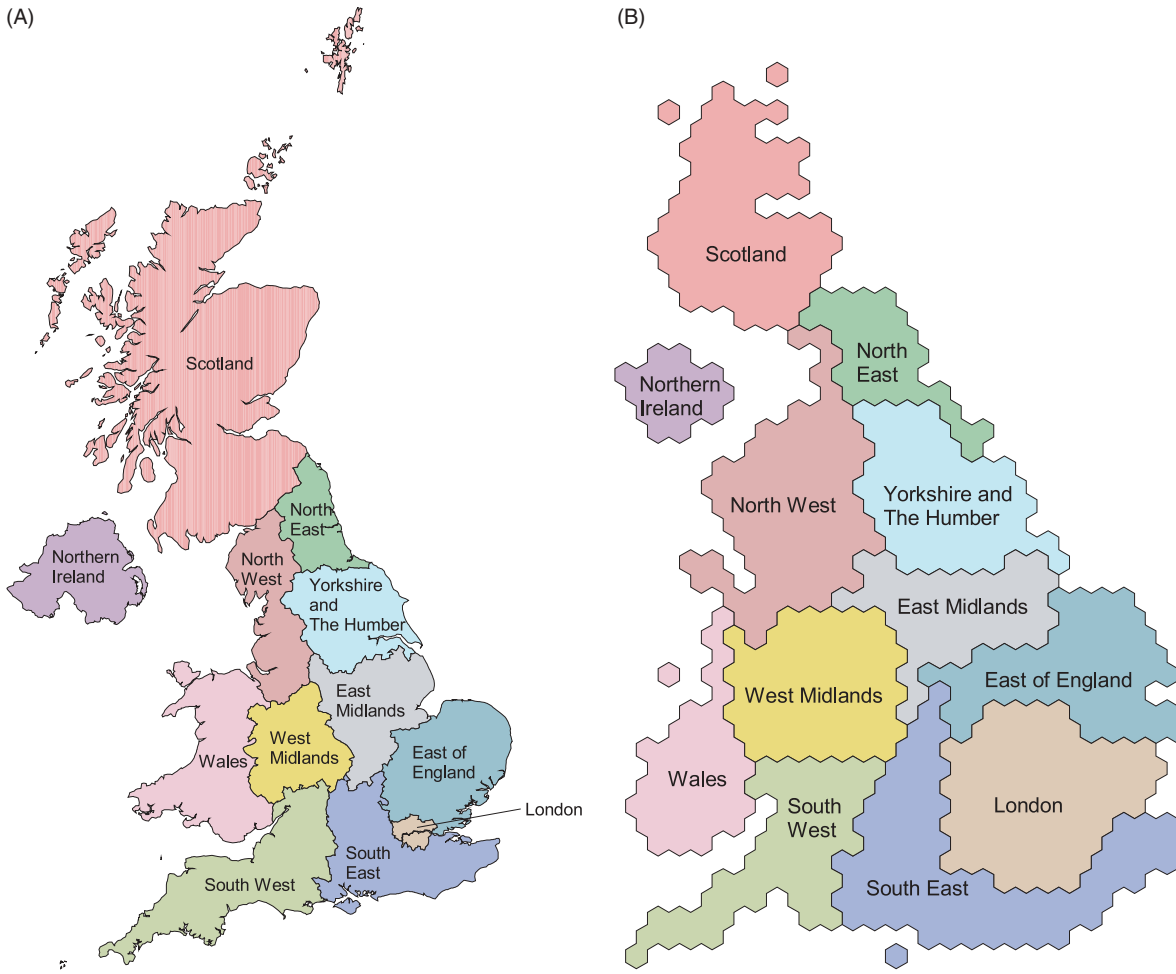


Figure 13.10 (A) A regional map of the UK and (B) its equal population cartogram transformation (Courtesy: Daniel Dorling and Bethan Thomas: from Dorling and Thomas (2004), page (vii)) (Reproduced by permission of The Policy Press)

transformations is useful for accommodating the way that humans think about or experience space, as with the measurement of distance as travel time or travel cost. As such, the visual transformations inherent in cartograms can provide improved means of envisioning spatial structure.

13.3.3 Re-modeling spatial distributions as dasymetric maps

The cartogram offers a more radical means of transforming space, and hence restoring spatial balance, but sacrifices the familiar spatial framework valued by most users. It forces the user to make a stark choice between assigning each mapped element equal weight and being able to relate spatial objects to real locations on the Earth's surface. Yet there are also other ways in which the data-integrative power of GIS can be used to re-model spatial distributions and hence assign spatial attributes to meaningful yet recognizable spatial objects.

One example of the way in which ancillary sources of information may be used to improve the model of a spatial distribution is known as *dasymetric mapping*. Here, the intersection of two datasets is used to obtain more precise estimates of a spatial distribution. Figure 13.12A shows the census tract geography for which small area population totals are known, and Figure 13.12B shows the spatial distribution of built structures in an urban area (which might be obtained from a cadaster or very-high-resolution satellite imagery, for example). A reasonable assumption (in the absence of evidence of mixed land use or very different residential structures such as high-rise apartments and widely spaced bungalows) is that all of the built structures house resident populations at uniform density. Figure 13.12C shows how this assumption, plus an overlay of the areal extent of built structures, allows population figures to be allocated to smaller areas than census tracts, and allows calculation of indicators of residential density. The practical usefulness of dasymetric mapping is illustrated in Figure 13.13. Figures 13.13A and B show the location of an elongated census block (or enumeration district) in the City of

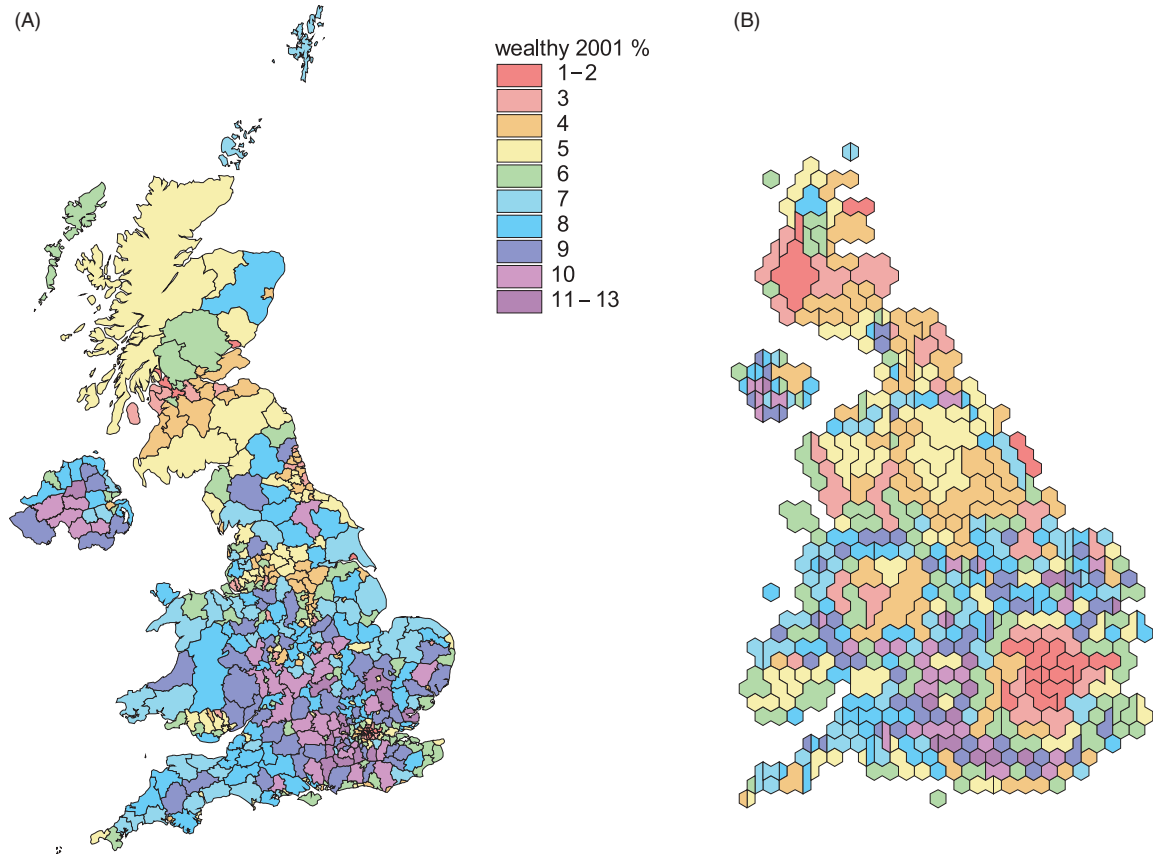


Figure 13.11 Wealthy areas of the UK, as measured by property values, property size, and levels of multiple car ownership: (A) conventional map of local authority districts and (B) its equal population cartogram transformation (Courtesy: Daniel Dorling and Bethan Thomas: from Dorling and Thomas (2004), page 16) (Reproduced by permission of The Policy Press)

Bristol, UK, which appears to have a high unemployment *rate*, but a low absolute *incidence* of unemployment. The resolution of this seeming paradox was examined in Box 4.3 but this would not help us much in deciding whether or not the zone (or part of it) should be included in an inner-city workfare program, for example. Use of GIS to overlay high-resolution aerial photography (Figure 13.13C) reveals the tract to be largely empty of population apart from a small extension to a large housing estate. It would thus appear sensible to assign this zone the same policy status as the zone to its west.

Dasymetric mapping uses the intersection of two datasets (or layers in the same dataset) to obtain more precise estimates of a spatial distribution.

Dasymetric mapping and related techniques present a window on reality that looks more convincing than conventional choropleth mapping. However, it is important to remain aware that the visualization of reality is only as good as the assumptions that are used to create it. The information about population concentration used in Figure 13.13, for example, is defined subjectively and is likely to be error prone (some built forms may be offices and shops, for example), and this will inevitably feed through into inaccuracies in visualization. Inference of land use from land cover, in particular, is an uncertain and error-prone process, and there is a developing literature on best practice for the classification of land use using land cover information and classifying different (e.g., domestic versus non-domestic) land uses.

Inferring land use from land cover is uncertain and error prone.

13.4 Immersive interaction and PPGIS

13.4.1 Overview: scientific visualization (ViSC) and representation

Faster processing hardware and more sophisticated computer graphics, including animation, have led to the development of a new field called 'Visualization in Scientific Computing.' ViSC investigates the use of new technology

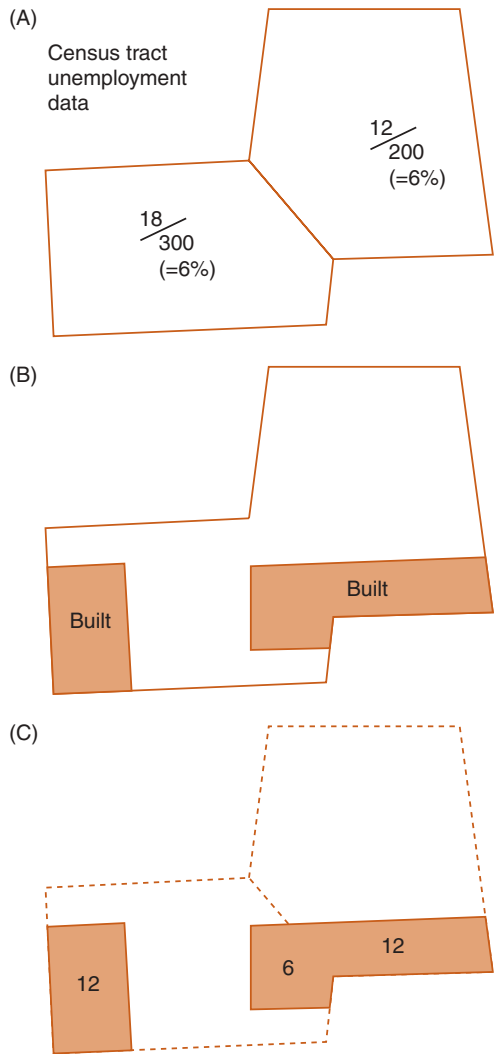


Figure 13.12 Modeling a spatial distribution in an urban area using dasymetric mapping: (A) zonal distribution of census population and unemployment rates; (B) distribution of built structures; and (C) overlay of (A) and (B) to obtain a more contained measure of distribution of unemployed population

and media to convey the multifaceted messages of today's mapping. ViSC provides greater flexibility, sophistication, and interaction in visualizing the world than the paper maps of the past. Some of the advanced techniques are not explicitly spatial, or are considered elsewhere: these include line generalization (Sections 3.8 and 4.8) cartographic smoothing (see the discussion of interpolation in Box 4.3), uncertainty visualization (e.g., Section 6.3.2.2), and parallel coordinate plots (Section 15.1). They may be delivered through a wide range of computer devices, from hand-held computers that help users to navigate the real world to virtual-reality systems that augment user perceptions of the real world or even present users with the artificial worlds of *virtual reality* discussed in Section 11.3.1. These activities pose many technical challenges that are interesting in themselves, but the broader motivation for many of the geovisualization research activities set out in Table 13.1 is to help users better understand geographic data and models, and hence improve understanding of real-world patterns and processes. Related to this is the objective of broadening the base of users that feels confident interacting with GIS. Research suggests that people first desire an overview of the totality of information, and then the ability to narrow down to information that is immediately relevant. Geodemographers (see Section 2.3.3) have investigated the ways in which individuals and different groups interact with new information and communication technologies (NICTs), and have suggested a three-stage model. Initially, people use NICTs to access information; many then graduate to using them to conduct transactions (e.g., through shopping); and user engagement culminates with people using NICTs as a gateway to participation in networked decision making. In a GIS context, this progression depends upon confidence in perceiving, manipulating, and exploring spatial phenomena through geovisualization.

Planner Mike Shiffer (Box 13.5) has described how GIS and a range of other computer media can be used to facilitate the active involvement of many different groups in the discussion and management of urban change through planning, and may act as a bulwark against officialdom or big business. Better human–computer interaction can be used to encourage public participation in (the use of) GIS, and this is often termed PPGIS. PPGIS is fundamentally about how people can perceive, manipulate, and interact with representations of the real world as manifest in GIS: its other concerns include how people evaluate options through multicriteria decision making (Section 16.4), and social issues of how GIS usage remains concentrated within networks of established interests that are often seen to control the use of technology (Section 1.7). A related theme is the use of GIS to create multiple representations – capturing and maintaining the different perspectives of stakeholders, rather than adopting only one authoritative view.

There is a range of potential uses of geovisualization in PPGIS, including:

- Making the growing complexity of land use planning, resource use, and community development intelligible to communities and different government departments.
- Radically transforming the planning profession through use of new tools for community design and decision making.
- Unlocking the potential of the many digital data sources that are collected, but not used, at the local level.
- Helping communities shift land use decisions from regulatory processes to performance-based strategies, and making the community decision making process more proactive and less reactive.
- Improving community education about local environmental and social issues (Figure 13.14).

In each of these applications, there is a strong cognitive component to geovisualization. Users need to feel equipped and empowered to interrogate representations

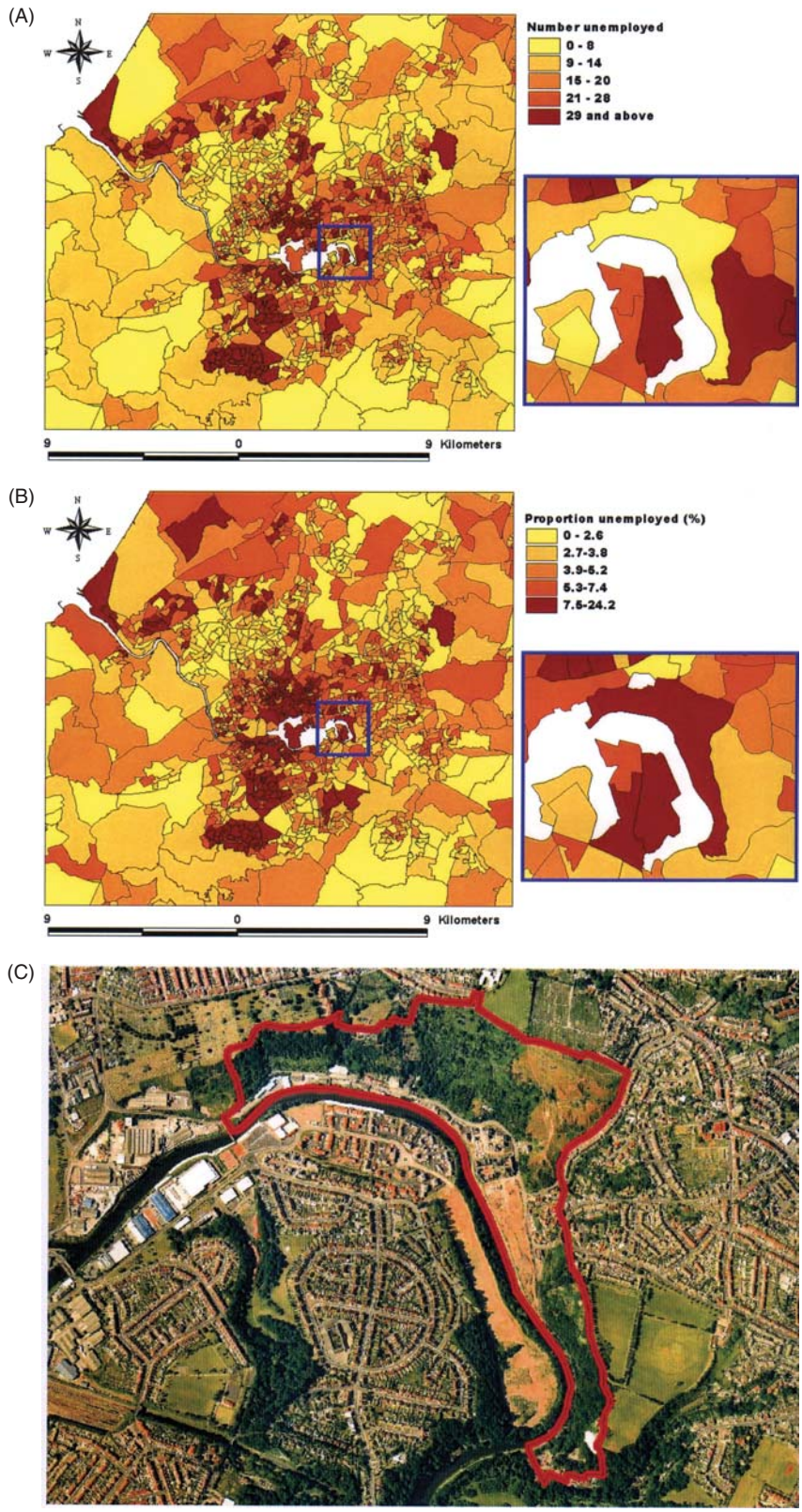


Figure 13.13 Dasymetric mapping in practice, in Bristol, UK: (A) numbers employed by census enumeration district (block); (B) proportions unemployed using the same geography (Reproduced with permission of Richard Harris); and (C) orthorectified aerial photograph of part of the area of interest. (Courtesy: Rich Harris and Cities Revealed: **www.crworld.co.uk**, reproduced by permission of The Geoinformation Group)

Figure 13.14 Field computing encourages education and public participation in GIS (PPGIS) (Courtesy: ESRI)

in order to reveal otherwise-hidden information. This requires dynamic and interactive *software* environments and *people* skills that are key to extracting meaning from a representation. GIS should allow people to use software to manipulate and represent data in multiple ways, in order to create 'what if' scenarios or to pose questions that prompt the discovery of useful relations or patterns. This is a core remit of PPGIS, where the geovisualization environment is used to support a process of knowledge construction and acquisition that is guided by the user's knowledge of the application. PPGIS research entails usability evaluations of structured tasks, using a mixture of computer-based techniques and traditional qualitative research methods, in order to identify cognitive activities and user problems associated with GIS applications.

13.4.2 Geovisualization and virtual-reality systems

Although conventional 2-D representations continue to be used in the overwhelming majority of GIS applications, increasing interaction with GIS now takes place through 3-D representations. This is closely allied with developments in computing, software, and broadband networking, which together allow:

- Users to access virtual environments and select different views of phenomena.
- Incremental changes in these perspectives to permit real time *fly-throughs* (see Figures 13.19, 13.20, and 13.21).
- Repositioning or rearrangement of the objects that make up such scenes.
- Users to be represented graphically as *avatars* – digital representations of themselves.
- Avatars to engage with others connected at different remote locations, creating a networked virtual world.
- The development, using avatars, of new kinds of representation and modeling.
- The linkage of networked virtual worlds with *virtual reality* (VR) systems. Semi-immersive and immersive VR is beginning to make the transition from the realm of computer gaming to GIS application, with new methods of spatial object manipulation and exploration using head-mounted displays and sound (see Section 11.3).

Geovisualization is enabling the creation of 3-D representations of natural and artificial (e.g., cityscapes) phenomena. Box 13.6 presents an illustration of the use of geovisualization in weather forecasting. The 3-D representation of artificial environments, specifically cities, has been aided by the increasing availability of very-high-resolution height data from airborne instruments, such as Light Detection and Ranging (LiDAR), and similar-resolution digital mapping products (Figure 13.17). Such data are achieving increasingly wide application in the extraction of discrete objects and the modeling of continuous elevation of underlying terrain; respective applications include urban planning and flood modeling. A common means of approximating extensive 3-D representations of urban environments is to use local height data to extrude the building footprints detailed on high-resolution mapping (such as GB Ordnance Survey's MasterMap 1:1250 database). Extruded block models provide a basic prismatic representation of buildings and have been augmented using computer-aided design (CAD) software to include details of (in order of complexity) roof structure, building volume, building elevation, and detailed features such as windows. CAD and GIS software also provide the facility of rendering buildings, streetscapes, and natural scenes. Many city visualizations have hitherto been restricted to small areas but, with the wide use of broadband networking, 3-D models can now be effectively 'streamed' over the Internet and make it possible to build together 3-D representations in order to visualize extensive geographic areas (see Box 13.7).

In parallel with the development of fine-scale 3-D models of cityscapes, there have been similarly impressive advances in whole-Earth global visualizations (Figure 13.18). Such software systems allow raster and vector information, including features such as buildings, trees, and automobiles, to be combined in synthetic and photo-realistic global and local displays. In global visualization systems, datasets are projected onto a global TIN-based data structure (see Section 8.2.3.4). Rapid interactive rotation and panning of the globe is enabled by caching data in an efficient in-memory data structure. Multiple levels of detail, implemented as nested TINs and reduced-resolution datasets, allow fast zooming in and out. When the observer is close to the surface of the globe, these systems allow the display angle to be tilted to provide a perspective view of the Earth's surface (Figure 13.19). The overall user experience is one of

Biographical Box **13.5**

Mike Shiffer, public participant in planning

Mike Shiffer (Figure 13.15) is Vice President responsible for planning and development at the Chicago Transit Authority (CTA). He grew up in a car-less family in Chicago where dependence on public transport led to his lifetime interest in improving urban qualities of life and mobility for those without autos. Shiffer is fond of saying that he stumbled into academia because the Chicago Transit Authority was not hiring train drivers every time he graduated (from primary school on). Shiffer received a Ph.D. in Regional Planning and Master's of Urban Planning from the University of Illinois at Urbana-Champaign, before moving in 1991 to work as an academic at the Massachusetts Institute of Technology's (MIT's) Department of Urban Studies and Planning. He returned to work in his native Chicago in 2000. Mike's academic research (he still retains a part-time academic post at the University of Illinois at Chicago) has investigated the ways in which spatial information and multimedia technologies can better inform decision making. He has developed a particular focus on how multimedia representational aids, virtual navigational aids, and broadband networking can better inform deliberation, decision making, and public debate.

At CTA, Shiffer spearheads overall strategic and operations planning efforts of the second-largest public transit system in the US, including service scheduling and facilities development. He is responsible for a staff of planners, architects, engineers, and other transit professionals in five departments. As such, his interests in how GIS technologies shape decision making are much more than academic: he has worked to reshape the CTA's planning process so that it leverages technologies and new models of human capital to create a more agile public transit system for Chicago. As our illustration (Figure 13.15) shows, he has also realized an earlier ambition: Shiffer is now a fully certified train driver.

Figure 13.15 Mike Shiffer, train enthusiast, in 1974 and 2004 (background map courtesy of Chicago Transit Authority)

Biographical Box **13.6**

Penny Tranter, weather forecaster

Figure 13.16 Penny Tranter, weather forecaster (Reproduced with permission of the BBC)

The UK Meteorological (Met) Office has supplied weather presenters, forecasting, and data services to the British Broadcasting Corporation (BBC) since the BBC broadcast the first British TV weather in January 1954. Today, geovisualization of the weather is required for the Corporation's established terrestrial TV channels (BBC1), its more recent all-digital TV services (including BBC4, BBC World, and BBC News 24), its interactive TV service (BBCi) and its real-time website feeds (the weather section is consistently one of the three most used on the BBC website in terms of hits: **www.bbc.co.uk/weather**).

BBC weather forecaster Penny Tranter (Figure 13.16) comes from Ayrshire in Scotland. She became interested in the weather as a teenager and, after completing a degree in Environmental Sciences at the University of East Anglia, joined the Met Office as a trainee forecaster in 1983. She now broadcasts across all BBC channels (BBC1, BBC 4, News 24, BBC World, BFBS, Radio 2, Radio 4, and Radio 5 Live).

BBC weather forecasters are equipped with computer-based weather information displays linked directly to the super-computers at Met Office HQ in Exeter. These powerful computer forecast models take real-time data feeds (Sections 10.8 and 10.9) from around the world and then simulate the world's atmosphere mathematically, from sea-level to the upper atmosphere. Forecasters use their scientific skills and experience as meteorologists to interpret the models and to predict what is going to happen (see Section 1.3). In consultation with the Met Office chief forecasters, they select the details they wish to use to support the weather story for that day. The BBC production team works with the forecasters to make sure that the editorial line 'fits' with the BBC's output for the day, and is consistent across all platforms.

Penny receives continuous data feeds from the Met Office in the form of digital charts showing pressure, temperature, rainfall, and cloud cover. In addition, she receives satellite pictures for every continent and hourly satellite images for the UK providing observational information. Half-hourly radar rainfall charts are also drawn up so that she can see where rain is actually falling and can compare this against the computer forecast model predictions. Over 6000 graphics products are drawn up automatically every day and Penny can also make up her own symbol or text charts using a range of background maps and images.

A state-of-the-art graphics system allows her to put together a sequence of charts and other graphics, including stills, animations, video clips, live weather cameras, and text charts. They can be in or out of vision, and can be customized for different channels or platforms. All the forecasts are ad-libbed and so can be shortened or lengthened as needed. She also prepares a summary chart for use at the end of each bulletin, and can jump to this, using a foot switch, should the bulletin have to be cut short suddenly.

Penny Tranter usually works alone in the studio. She uses an ear-piece to listen to the programme before her broadcast and to any last-minute instructions from the director in the network control room. She stands in front of a translucent screen on which is projected a faint image of the graphics. This gives her an idea of where on the map to point. The back of the screen is flooded with blue light, and an electronic system known as Colour Separation Overlay (CSO) causes any area where the camera detects blue to be replaced by a 'clean feed' of the charts from the computer, called up in sequence by the forecaster at the touch of a button. You will never see Penny dressed in blue clothing as this would merge with the graphics!

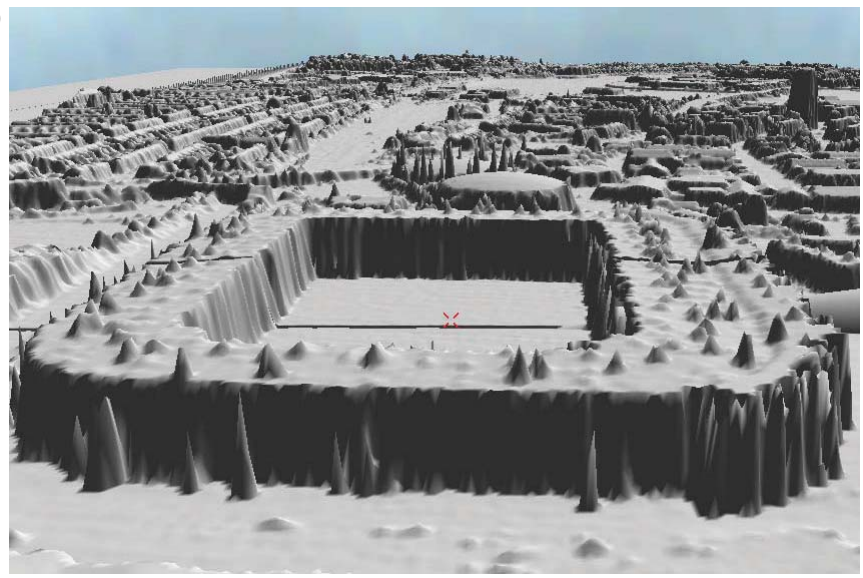
Monitors next to the camera show the output of the graphics computer and the output of the studio. In front of the camera is an autocue screen onto which is projected the service (e.g., BBC1) on which she is about to broadcast. Superimposed on this image is a countdown clock. Penny starts her forecast when she sees herself appear and stops when the countdown clock reaches zero.

A dedicated, BBC-designed weather graphics system has also been installed into each of the national and regional BBC Centres. It was radically revised in 2005, and allows each local weather presenter around the UK to prepare and transmit his or her own weather bulletin using the same information and graphics systems as the forecasters based in the BBC Weather Centre in London.

Bringing together such a range of data sources, building multiple representations of the world into a seamless presentation, and responding to the pressures of real-time delivery of forecasts is not always a straightforward task. Speaking of her work, Penny says:

> **Our state-of-the-art graphics system allows an easy-to-access, easy-to-use, and easy-to-understand approach for both the weather presenter and the viewer. Robust geovisualization conventions make our broadcasts intelligible in any part of the world. Accompanied by the BBC's graphics, they enable us to deliver our weather information – accurately and efficiently – to a wide global audience from Indian farmers to North American retailers.**
>
> (Information courtesy of the British Broadcasting Corporation).

(A)



(B)

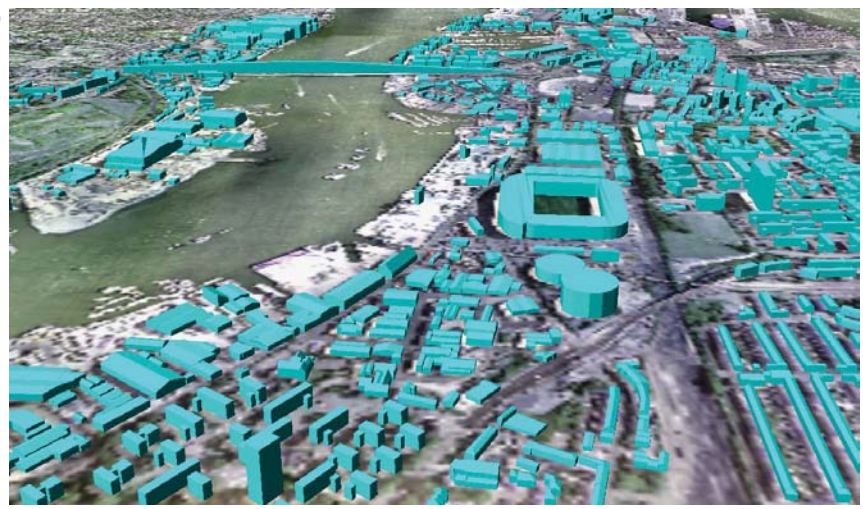


Figure 13.17 (A) Raw LiDAR image of football (soccer) stadium, Southampton, England, and (B) LiDAR-derived bare earth DEM draped with aerial photograph of its wider area, overlaid with 3-D buildings. (Aerial photography reproduced with permission of Ordnance Survey © Ordnance Survey. All rights reserved. (Reproduced with permission from the Environment Agency of England and Wales. Courtesy: Sarah Smith.)

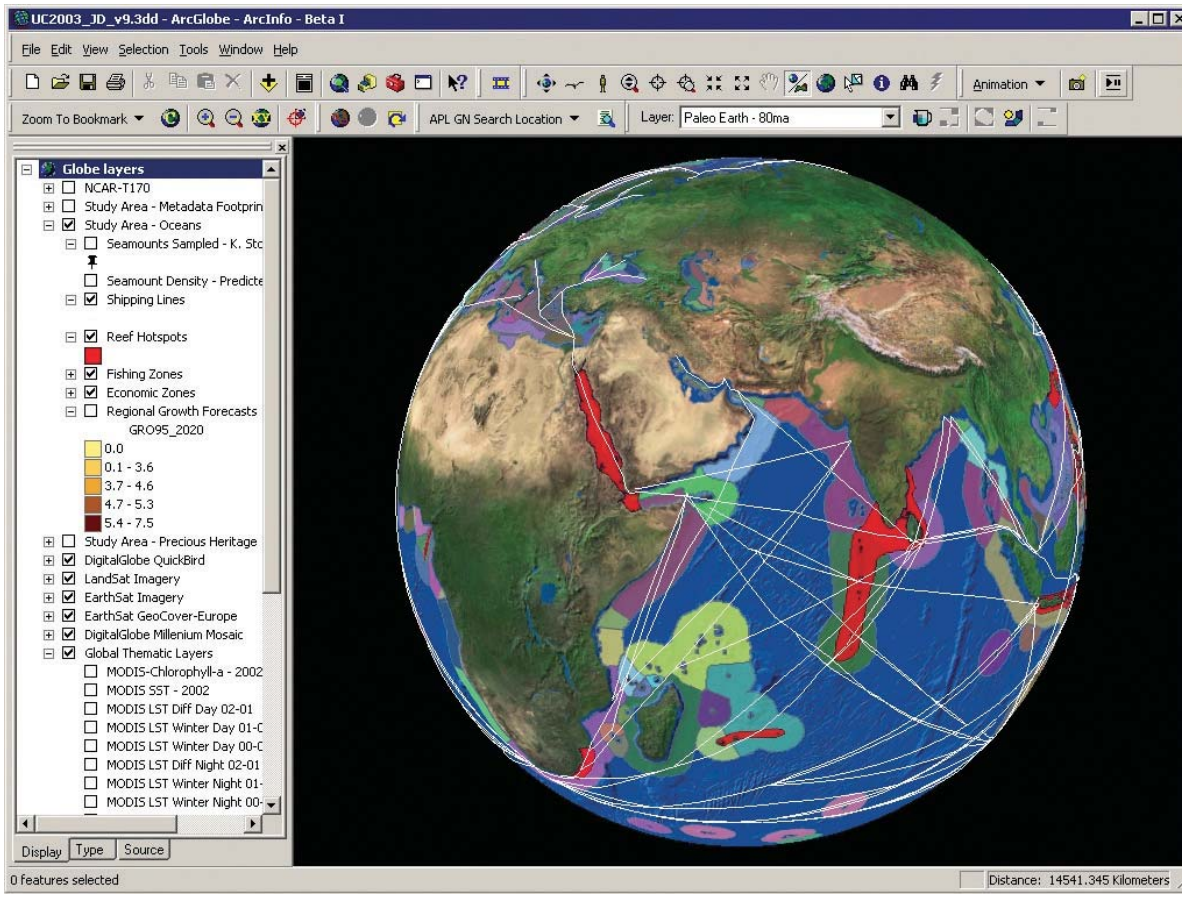


Figure 13.18 A whole-Earth visualization showing vector shipping lanes (white lines), reefs (red), and ocean fishing and economic zones, overlaid on top of raster topography (Courtesy: ESRI)

being able to roam smoothly over large volumes of global geographic information. This enables a better understanding of the distribution and abundance of globally distributed phenomena (e.g., ocean currents, atmospheric temperatures, and shipping lanes), as well as detailed examination of local features in the natural and built environment (e.g., optimum location of cellphone towers, or the impact of tree felling upon the viewscape of tourist areas).

13.4.3 Hand-held computing and geovisualization

Immersion of desk- and studio-bound users in 3-D virtual reality presents one way of promoting better remote interaction with the real world, using computers that are physically amongst the largest of the digital age. At the other end of the computer size spectrum, improved direct interaction with the real world has been made possible by the development of a range of hand-held, in-vehicle, and wearable computer devices. These are discussed, along with the geovisualization conventions they entail, in Sections 7.6.4 and 11.3.

13.5 Consolidation

Geovisualization can make a powerful contribution to decision making and can be used to simulate changes to reality. Yet, although it is governed by scientific principles, the limitations of human cognition mean that geovisualization necessarily presents a further selective filter on the reality that we seek to represent. Mapping is about seeing the detail as well as the big picture, yet the wealth of detail that is available in today's digital environments can sometimes create information overload and threaten to overwhelm the message of the representation. Geovisualization is fostering greater user interaction and participation in the use of GIS as a decision support tool. At its worst, greater sophistication of display may create information overload and obscure the inherent uncertainty in most spatial data: at its best, the media of visualization allow the user to undertake a balanced appraisal of the message of a geographic representation, and to create better description, explanation, prediction, and understanding of GI. The contribution of geovisualization to PPGIS can be viewed as developing and broadening the base of users that not

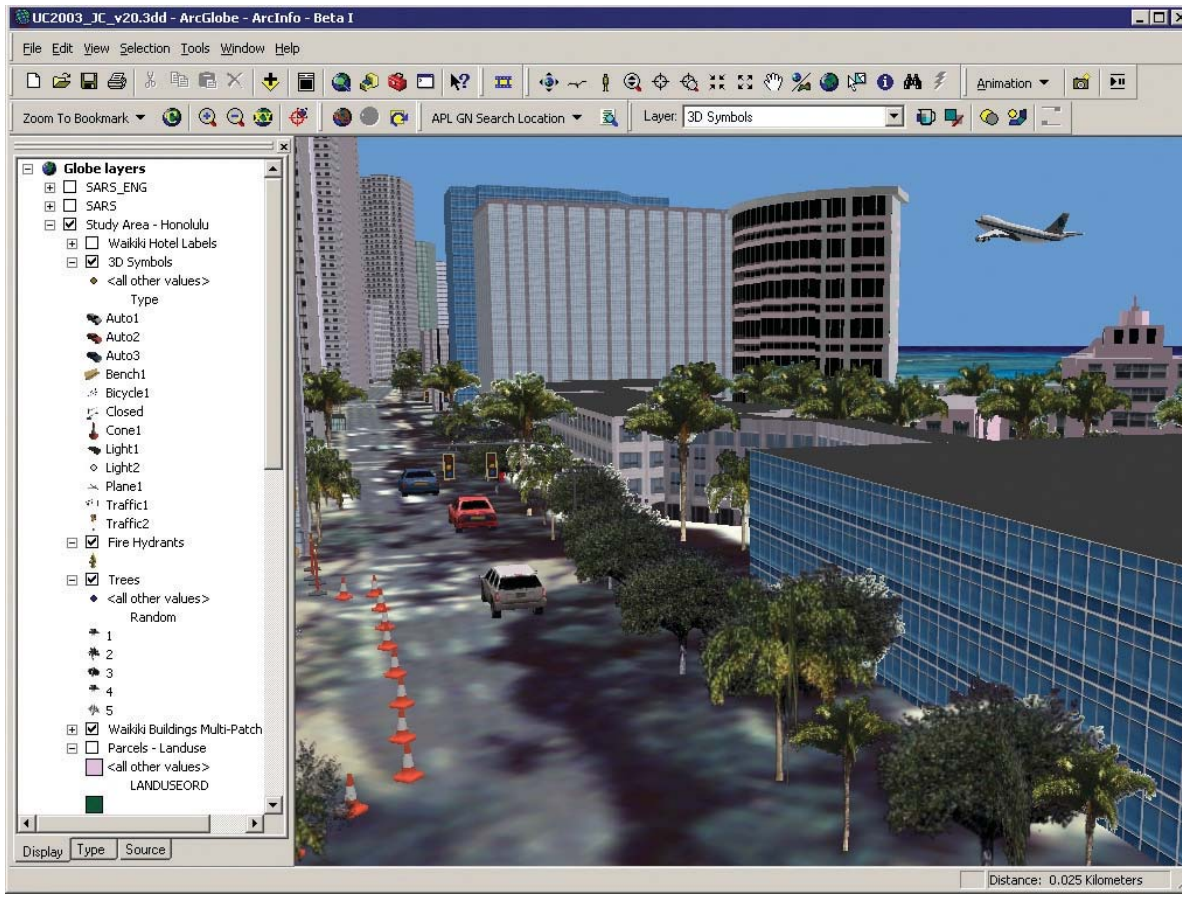


Figure 13.19 A cityscape visualization, comprising extruded buildings and specific vector objects (cars, plane, traffic cones, etc.) to create a photo-realistic simulation (Courtesy: ESRI)

Applications Box **13.7**

Virtual London

In 2002, the UK central government announced its Electronic Government Initiative with the objective of delivering all government procurement services electronically by 2005. This has garnered political support for the vision of a 'Virtual London' and the 3-D GIS research necessary to create it.

Working at the Centre for Advanced Spatial Analysis at University College London, GIS researchers Andy Smith and Steve Evans are creating a 3-D representation of Central London (Figure 13.20). The model is being produced using GIS, CAD, and a variety of new photorealistic imaging techniques and photogrammetric methods of data capture. The core model will be distributed via broadband Internet connections, utilizing techniques to optimize sharing of large datasets and create a multi-user environment. Citizens will be able to roam around a Virtual Gallery as avatars in order to explore issues relating to London. Similar models are available for other cities (Vexcel maintains an 'off-the-shelf' library of US cities, derived principally from remotely sensed data: **www.vexcel.com**), and models of this nature have been used to encourage and enliven public participation in the planning process, particularly with focus upon urban redevelopment proposals (e.g., the Australian community spatial simulation scenarios at **www.c-s3.info**).

The model is designed for four audiences:

- Professionals, such as architects, developers, and planners who are anxious to use a full set of data query and visualization capabilities. For example, architects might place new buildings in the model in order to assess a variety of issues, such as visual impact or likely consequences for traffic and surrounding land use, and law enforcement agencies might use the model for community policing (Figure 13.21).
- Citizens interested in learning about London or in understanding the impacts of development and traffic proposals. This has obvious links to the development of e-democracy and the evaluation of 'what if' scenarios to enable digital planning by citizens.
- Virtual tourists interested in navigating and viewing scenes, or picking up links to websites of interest. This has implications for the way in which city marketing and regional development initiatives are undertaken.
- Classroom users, with diverse interests in geography, civics, and history, and university-based educators in urban planning, architecture, and computer science.

Check out the current state of Virtual London's development at **www.casa.ucl.ac.uk/research/virtuallondon.htm**.

Figure 13.20 Scene from Virtual London, including County Hall and the London Eye (Courtesy: Andy Hudson-Smith, Steve Evans, and ESRI)

Figure 13.21 Virtual London law enforcement analysis: 3-D visualization of closed circuit television viewshed (green) and street illumination (Courtesy: Christian Castle, John Calkins, and ESRI)

only sources information through GIS, but increasingly uses GIS as a medium for information exchange and participation in decision making. The old adage that 'seeing is believing' only holds if visualization and user interaction are subservient to the underlying messages of a representation.

Questions for further study

1. Figure 13.22 is a cartogram redrawn from a newspaper feature on the costs of air travel from London in 1992. Using current advertisements in the press and on the Internet, create a similar cartogram of travel costs, in local currency, from the nearest international air hub to your place of study.
2. How can Web-based multimedia GIS tools be used to improve community participation in decision making?
3. Produce two (computer generated) maps of the Israeli West Bank security fence to illustrate opposing views of its effects. For example, the first might illustrate how it helps to preserve an Israeli 'Lebensraum' while the second might emphasize its disruptive impacts upon access between Palestinian communities. In a separate short annotative commentary, describe the structure and character of the fence at a range of spatial scales.
4. Review the common *sources* of uncertainty in geographic representation and the ways in which they can be *manifest* through geovisualization.

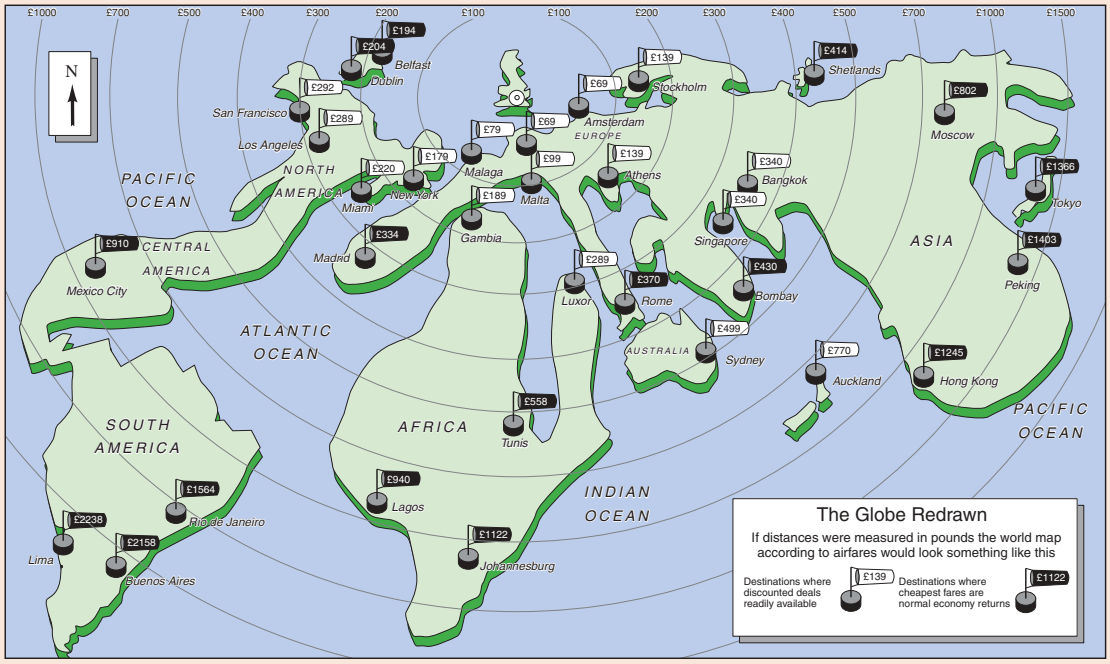


Figure 13.22 The globe redrawn in terms of 1992 travel costs from London (*Source*: redrawn from newspaper article by Frank Barrett in *The Independent* on 9 February 1992, page 6)

Further reading

Craig W.J., Harris T.M. and Weiner D. 2002 *Community Participation and Geographic Information Systems*. Boca Raton, FL: CRC Press.

Dorling D. and Thomas B. 2004 *People and Places: a 2001 Census Atlas of the UK*. Bristol, UK: Policy Press.

Egenhofer M. and Kuhn W. 2005 'Interacting with GIS'. In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley: 401–12.

MacEachren A. 1995 *How Maps Work: Representation, Visualization and Design*. New York: Guilford Press.

MacEachren A.M., Gahegan M., Pike W., Brewer I., Cai G., Lengerich E., and Hardisty F. 2004 'Geovisualization for knowledge construction and decision-support'. *Computer Graphics and Applications* **24**(1), 13–17.

Shiffer M. 2005 'Managing public discourse: towards the augmentation of GIS with multimedia'. In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley: 723–32.

Tufte E.R 2001 *The Visual Display of Quantitative Information* (2nd edn). Cheshire, CT: Graphics Press.

14 *Query, measurement, and transformation*

This chapter is the first in a set of three dealing with geographic analysis and modeling methods. The chapter begins with a review of the relevant terms, and an outline of the eight major topics covered in the three chapters: three in each of Chapters 14 and 15, and two in Chapter 16. Query methods allow users to interact with geographic databases using pointing devices and keyboards, and GIS have been designed to present data for this purpose in a number of standard views. The second area covered in this chapter, measurement, includes algorithms for determining lengths, areas, shapes, slopes, and other properties of objects. Transformations allow new information to be created through simple geometric manipulation.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After working through this chapter you will know:

- Definitions of geographic analysis and modeling, and tests to determine whether a method is geographic;
- The range of queries possible with a GIS;
- Methods for measuring length, area, shape, and other properties, and their caveats;
- Transformations that manipulate objects to create new ones, or to determine geometric relationships between objects.

14.1 Introduction: what is spatial analysis?

The techniques covered in these three chapters are generally termed *spatial* rather than *geographic*, because they can be applied to data arrayed in any space, not only geographic space. Many of the methods might potentially be used in analysis of outer space by astronomers, or in analysis of brain scans by neuroscientists. So the term spatial is used consistently throughout these chapters.

Spatial analysis is in many ways the crux of GIS because it includes all of the transformations, manipulations, and methods that can be applied to geographic data to add value to them, to support decisions, and to reveal patterns and anomalies that are not immediately obvious – in other words, spatial analysis is the process by which we turn raw data into useful information, in pursuit of scientific discovery, or more effective decision making. If GIS is a method of communicating information about the Earth's surface from one person to another, then the transformations of spatial analysis are ways in which the sender tries to inform the receiver, by adding greater informative content and value, and by revealing things that the receiver might not otherwise see.

Some methods of spatial analysis were developed long before the advent of GIS, and carried out by hand, or by the use of measuring devices like the ruler. The term *analytical cartography* is sometimes used to refer to methods of analysis that can be applied to maps to make them more useful and informative, and spatial analysis using GIS is in many ways its logical successor.

Spatial analysis can reveal things that might otherwise be invisible – it can make what is implicit explicit.

In this and the next chapter we will look first at some definitions and basic concepts of spatial analysis. Following introductory material, the two chapters include six sections, which look at spatial analysis grouped into six more-or-less-distinct categories: queries and reasoning, measurements, transformations, descriptive summaries, optimization, and hypothesis testing. Chapter 16 is devoted to *spatial modeling*, a loosely defined term that covers a variety of more advanced and more complex techniques, and includes the use of GIS to analyze and simulate dynamic processes, in addition to analyzing static patterns. Some of the methods discussed in these two chapters were introduced in Chapter 4 as ways of describing the fundamental nature of geographic data, so references will be made to that chapter as appropriate.

Spatial analysis is the crux of GIS, the means of adding value to geographic data, and of turning data into useful information.

Methods of spatial analysis can be very sophisticated, but they can also be very simple. A large body of methods of spatial analysis has been developed over the past century or so, and some methods are highly mathematical – so much so, that it might sometimes seem that mathematical complexity is an indicator of the importance of a technique. But the human eye and brain are also very sophisticated processors of geographic data and excellent detectors of patterns and anomalies in maps and images. So the approach taken here is to regard spatial analysis as spread out along a continuum of sophistication, ranging from the simplest types that occur very quickly and intuitively when the eye and brain look at a map, to the types that require complex software and sophisticated mathematical understanding. Spatial analysis is best seen as a *collaboration* between the computer and the human, in which both play vital roles.

Effective spatial analysis requires an intelligent user, not just a powerful computer.

There is an unfortunate tendency in the GIS community to regard the making of a map using a GIS as somehow less important than the performance of a mathematically sophisticated form of spatial analysis. According to this line of thought, *real* GIS involves number crunching, and users who *just* use GIS to make maps are not serious users. But every cartographer knows that the design of a map can be very sophisticated, and that maps are excellent ways of conveying geographic information and knowledge, by revealing patterns and processes to us. We agree, and believe that map making is potentially just as important as any other application of GIS.

Spatial analysis helps us in situations when our eyes might otherwise deceive us.

There are many possible ways of defining spatial analysis, but all in one way or another express the basic idea that information on locations is essential – that analysis carried out without knowledge of locations is not spatial analysis. One fairly formal statement of this idea is:

Spatial analysis is a set of methods whose results are not invariant under changes in the locations of the objects being analyzed.

The double negative in this statement follows convention in mathematics, but for our purposes we can remove it:

Spatial analysis is a set of methods whose results change when the locations of the objects being analyzed change.

On this test the calculation of an average income for a group of people is not spatial analysis, because it in no way depends on the locations of the people. But the calculation of the center of the US population is spatial analysis because the results depend on knowing where all US residents are located. GIS is an ideal platform for spatial analysis because its data structures accommodate the storage of object locations.

14.1.1 Examples

Spatial analysis can be used to further the aims of science, by revealing patterns that were not previously recognized, and that hint at undiscovered generalities and laws. Patterns in the occurrence of a disease may hint at the mechanisms that cause the disease and some of the most famous examples of spatial analysis are of this nature, including the work of Dr John Snow in unraveling the causes of cholera (Box 14.1).

It is interesting to speculate on what would have happened if early epidemiologists like Snow had had access to a GIS. The rules governing research today

Biographical Box **14.1**

Dr John Snow and the causes of cholera

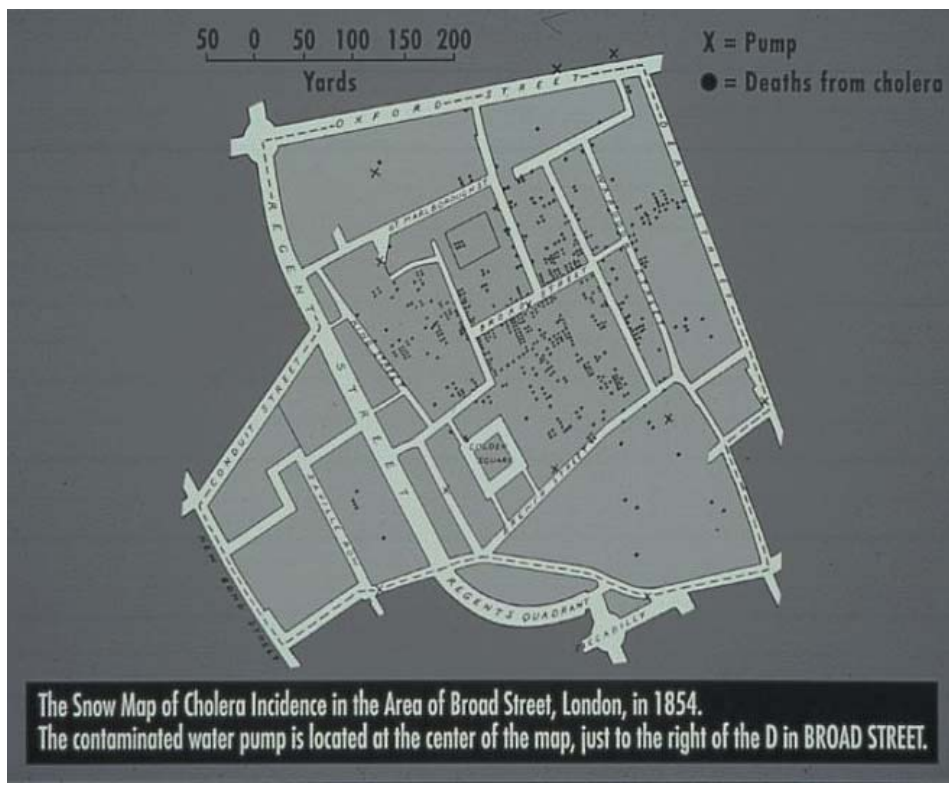


Figure 14.1 A redrafting of the map made by Dr John Snow in 1854 of the deaths that occurred in an outbreak of cholera in the Soho district of London. The existence of a public water pump in the center of the outbreak (the cross in Broad Street) convinced Snow that drinking water was the probable cause of the outbreak. Stronger evidence was obtained in support of this hypothesis when the water supply was cut off, and the outbreak subsided. (*Source*: Gilbert E. W. 1958 'Pioneer maps of health and disease in England.' *Geographical Journal*, **124**: 172–183) (Reproduced by permission of Blackwell Publishing Ltd.)

In the 1850s, cholera was very poorly understood and massive outbreaks were a common occurrence in major industrial cities (today cholera remains a significant health hazard in many parts of the world, despite progress in understanding its causes and advances in treatment). An outbreak in London in 1854 in the Soho district was typical of the time, and the deaths it caused are mapped in Figure 14.1. The map was made by Dr John Snow (Figure 14.2), who had conceived the hypothesis that cholera was transmitted through the drinking of polluted water, rather than through the air, as was commonly believed. He noticed that the outbreak appeared to be centered on a public drinking water pump in Broad Street (Figure 14.3) – and if his hypothesis was correct, the pattern shown on the map would reflect the locations of people who drank the pump's water. There appeared to be anomalies, in the sense that deaths had occurred in households that were located closer to other sources of water, but he was able to confirm that these households also drew their water from the Broad Street pump. Snow had the handle of the pump removed, and the outbreak subsided, providing direct causal evidence in favor of his hypothesis. The full story is much more complicated than this largely apocryphal version, of course; much more information is available at **www.jsi.com**.

Today, Snow is widely regarded as the father of modern epidemiology.

Figure 14.2 Dr John Snow (*Source*: John Snow Inc. **www.jsi.com**)

Figure 14.3 A modern replica of the pump that led Snow to the inference that drinking water transmitted cholera, located in what is now Broadwick Street in Soho, London (*Source*: John Snow Inc. **www.jsi.com**)

would not have allowed Snow to remove the pump handle, except after lengthy review, because the removal constituted an experiment on human subjects. To get approval, he would have had to have shown persuasive evidence in favor of his hypothesis, and it is doubtful that the map would have been sufficient because several other hypotheses might have explained the pattern equally well. First, it is conceivable that the population of Soho was inherently at risk of cholera, perhaps by being comparatively elderly, or because of poor housing conditions. The map would have been more convincing if it had shown the *rate* of incidence, relative to the population at risk. For example, if cholera was highest among the elderly, the map could have shown the number of cases in each small area of Soho as a proportion of the population aged over 50 in each area. Second, it is still conceivable that the hypothesis of transmission through the air between carriers could have produced the same observed pattern, if the first carrier happened to live in the center of the outbreak near the pump. Snow could have eliminated this alternative if he had been able to produce a sequence of maps, showing the locations of cases as the outbreak developed. Both of these options involve simple spatial analysis of the kind that is readily available today in GIS.

GIS provides tools that are far more powerful than the map at suggesting causes of disease.

Today the causal mechanisms of diseases like cholera, which results in short, concentrated outbreaks, have long since been worked out. Much more problematic are the causal mechanisms of diseases that are rare, and not sharply concentrated in space and time. The work of Stan Openshaw at the University of Leeds, using one of his Geographical Analysis Machines, illustrates the kinds of applications that make good use of the power of GIS in this contemporary context.

Figure 14.4 shows an application of one of Openshaw's techniques to a comparatively rare but devastating disease whose causal mechanisms remain largely a mystery – childhood leukemia. The study area is northern England, from the Mersey to the Tyne. The analysis begins with two datasets: one of the locations of cases of the disease, and the other of the numbers of people at risk in standard census reporting zones. Openshaw's technique then generates a large number of circles, of random sizes, and places them (*throws* them) randomly over the map. The computer generates and places the circles, and then analyzes their contents, by dividing the number of cases found in the circle by the size of the population at risk. If the ratio is anomalously high, the circle is drawn. After a large number of circles has been generated, and a small proportion have been drawn, a pattern emerges. Two large concentrations, or clusters of cases, are evident in the figure. The one on the left is located around Sellafield, the location of the British Nuclear Fuels processing plant and a site of various kinds of leaks of radioactive material. The other, in the upper right, is in the Tyneside region, and Openshaw and his colleagues discuss possible local causes.

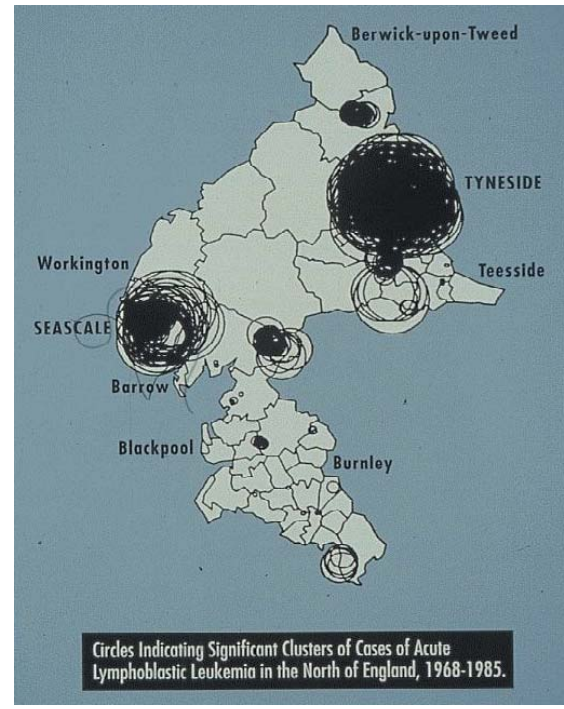


Figure 14.4 The map made by Openshaw and colleagues by applying their Geographical Analysis Machine to the incidence of childhood leukemia in northern England. A very large number of circles of random sizes is randomly placed on the map, and a circle is drawn if the number of cases it encloses substantially exceeds the number expected in that area given the size of its population at risk (*Source*: Openshaw S., Charlton M., Wymer C., and Craft A. 1987 'A Mark I geographical analysis machine for the automated analysis of point datasets.' *International Journal of Geographical Information Systems*, **1**: 335–358. **www.tandf.co.uk/journals**)

Both of these examples are instances of the use of spatial analysis for scientific discovery and decision making. Sometimes spatial analysis is used *inductively*, to examine empirical evidence in the search for patterns that might support new theories or general principles, in this case with regard to disease causation. Other uses of spatial analysis are *deductive*, focusing on the testing of known theories or principles against data (Snow already had a theory of how cholera was transmitted, and used the map as confirmation to convince others). A third type of application is *normative*, using spatial analysis to develop or prescribe new or better designs, for the locations of new retail stores, or new roads, or new manufacturing plant. Examples of this type appear in Section 15.3.

14.1.2 Types of spatial analysis and modeling

The remaining sections of this chapter and the following one discuss methods of spatial analysis using six general headings:

Queries are the most basic of analysis operations, in which the GIS is used to answer simple questions posed by the user. No changes occur in the database, and no new data are produced. The operations vary from simple and well-defined queries like 'how many houses are found within 1 km of this point', to vaguer questions like 'which is the closest city to Los Angeles going north', where the response may depend on the system's ability to understand what the user means by 'going north' (see the discussion of vagueness in Chapter 6). Queries of databases using SQL are discussed in Section 10.4.

Measurements are simple numerical values that describe aspects of geographic data. They include measurement of simple properties of objects, like length, area, or shape, and of the relationships between pairs of objects, like distance or direction.

Transformations are simple methods of spatial analysis that change datasets, combining them or comparing them to obtain new datasets, and eventually new insights. Transformations use simple geometric, arithmetic, or logical rules, and they include operations that convert raster data into vector data, or vice versa. They may also create fields from collections of objects, or detect collections of objects in fields.

Descriptive summaries attempt to capture the essence of a dataset in one or two numbers. They are the spatial equivalent of the descriptive statistics commonly used in statistical analysis, including the mean and standard deviation.

Optimization techniques are normative in nature, designed to select ideal locations for objects given certain well-defined criteria. They are widely used in market research, in the package delivery industry, and in a host of other applications.

Hypothesis testing focuses on the process of reasoning from the results of a limited sample to make generalizations about an entire population. It allows us, for example, to determine whether a pattern of points could have arisen by chance, based on the information from a sample. Hypothesis testing is the basis of inferential statistics and lies at the core of statistical analysis, but its use with spatial data is much more problematic.

In Chapter 16, the attention turns to modeling, a very broad term that is used when the individual steps of analysis are combined into complex sequences. *Static modeling* describes the use of a sequence to attain some defined goal, such as the measurement of the suitability of locations for development, or their sensitivity to pollution or erosion. *Dynamic modeling* uses the GIS to emulate real physical or social processes operating on the geographic landscape. Time is broken up into a series of discrete steps, and the operations of the GIS are *iterated* at each time step. Dynamic modeling can be used to emulate processes of erosion, or the spread of disease, or the movement of individual vehicles in a congested street network.

14.2 Queries

In the ideal GIS it should be possible for the user to interrogate the system about any aspect of its contents, and obtain an immediate answer. Interrogation might involve pointing at a map, or typing a question, or pulling down a menu and clicking on some buttons, or sending a formal SQL request to a database (Section 10.4). The visual aspects of query are discussed in Chapter 13. Today's user interfaces are very versatile, and have very nearly reached the point where it will be possible to interrogate the system by speaking to it – this would be extremely valuable in vehicles, where the use of more conventional ways of interrogating the system through keyboards or pointing devices can be too distracting for the driver.

When a GIS is used in a moving vehicle it is impossible to rely on the normal modes of interaction through keyboards or pointing devices.

The very simplest kinds of queries involve interactions between the user and the various *views* that a GIS is capable of presenting. A *catalog* view shows the contents of a database, in the form of storage devices (hard drives, Internet sites, CDs, ZIP disks, or USB thumb-drives) with their associated folders, and the datasets contained in those folders. The catalog will likely be arranged in a hierarchy, and the user is able to expose or hide various branches of the hierarchy by clicking at appropriate points. Figure 14.5 shows a catalog view, in this case using ESRI's ArcCatalog software (a component of ArcGIS). Note how different types of datasets are symbolized using different icons so the user can tell at a glance which files contain grids, polygons, points, etc.

In contemporary software environments, such as Microsoft's Windows, the Macintosh, or Unix, many kinds of interrogation are available through simple pointing and clicking. For example, in ArcCatalog simply pointing at a dataset icon and clicking the right mouse button exposes basic statistics on the dataset when the Properties option is selected. The metadata option exposes the metadata stored with the dataset, including its projection and datum details, the names of each of its attributes, and its date of creation.

Users query a GIS database by interacting with different views.

The *map* view of a dataset shows its contents in visual form, and opens many more possibilities for querying. When the user points to any location on the screen the GIS should display the pointer's coordinates, using the units appropriate to the dataset's projection and coordinate system. For example, Figure 14.6 shows the most recent location of the pointer in the box below the map window in units of meters east and north, because the map projection in use is UTM (Section 5.7.2). If the dataset is raster the system might display a cell's row and column number, or its coordinate system if the raster is adequately georeferenced (tied to some Earth coordinate system). By

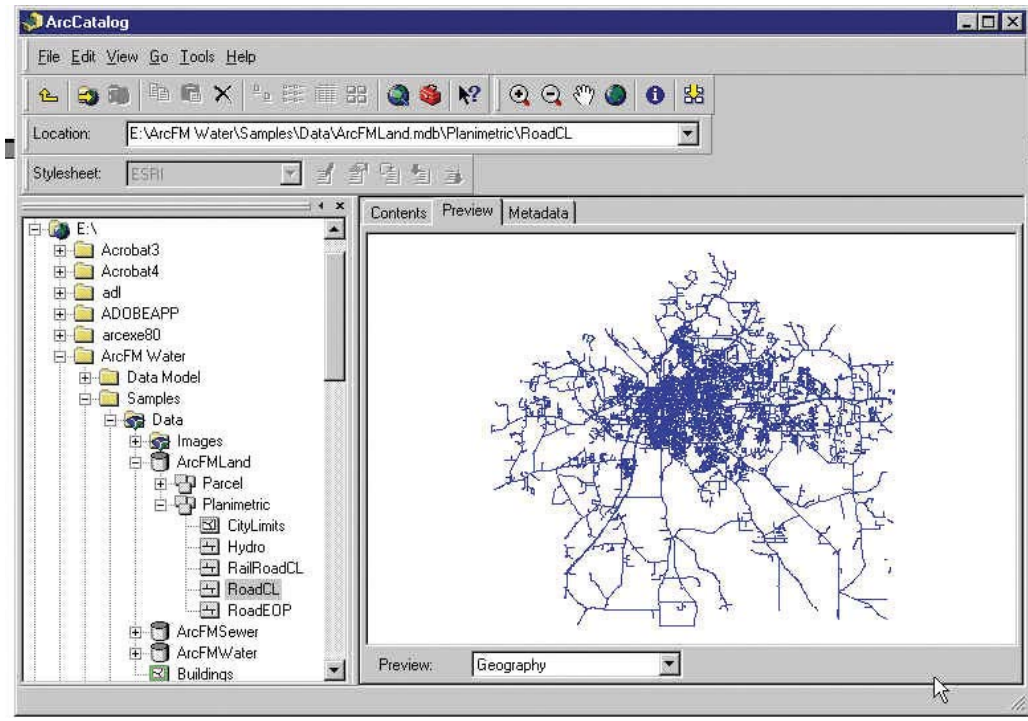


Figure 14.5 A catalog view of a GIS database (ESRI's ArcCatalog). The left window exposes the file structure of the database, and the right window provides a geographic preview of the contents of a selected file (RoadCL). Other options for the right window include a view of the file's metadata (Section 11.2.1), and its attribute table

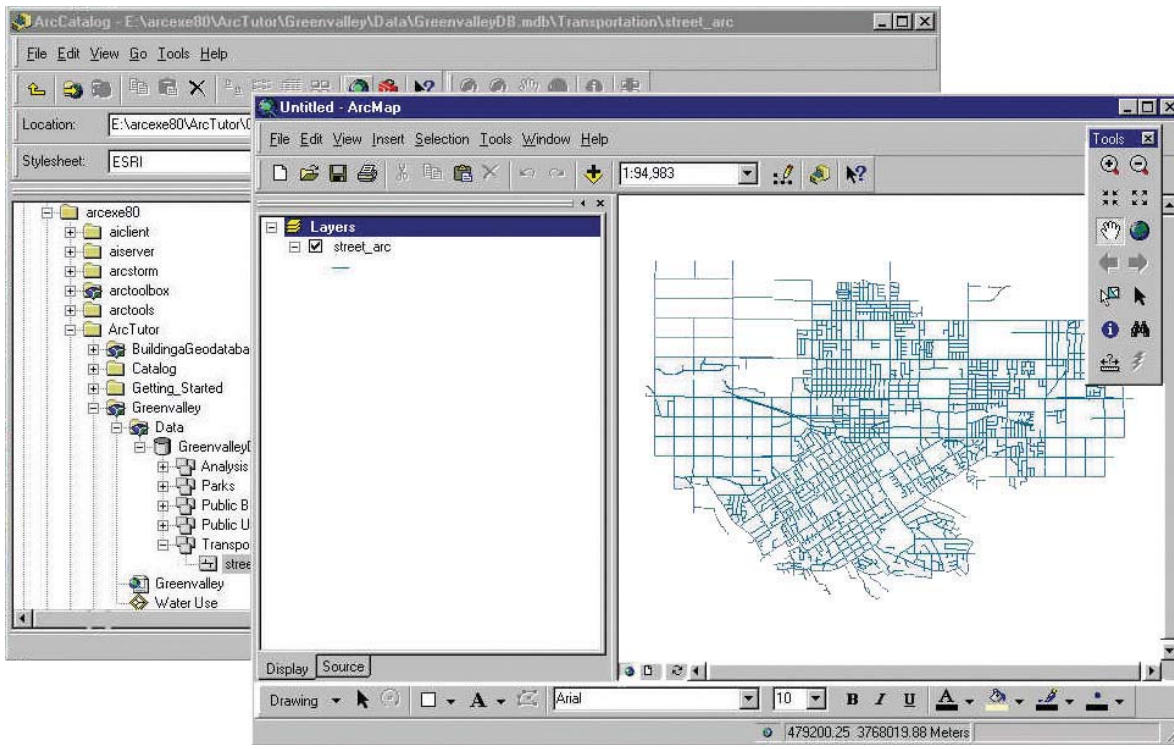


Figure 14.6 ESRI's ArcMap displays the most recent location of the pointer in a box below the map display, allowing the user to query location anywhere on the map. Location is shown in this case in meters east and north (using a UTM projection), since this is the coordinate system of the data being queried

pointing to an object the user should be able to display the values of its attributes, whether the object is a raster cell, a line representing a network link, or a polygon.

Finally, the *table* view of a dataset shows a rectangular array, with the objects organized as the rows and the attributes as the columns (Figure 14.7). This allows the user to see the attributes associated with objects at a glance, in a convenient form. There will usually be a table associated with each type of object – points, lines, areas (see Box 4.1 and Chapter 8) – in a vector database. Some systems support other views as well. In a *histogram* view, the values of a selected attribute are displayed in the form of a bar graph. In a *scatterplot* view, the values of two selected attributes are displayed plotted against each other. Scatterplots allow us to see whether relationships exist between attributes. For example, is there a tendency for the average income in a census tract to increase as the percentage of people with university education increases?

Today's GIS supports much more sophisticated forms of query than these. First, it is common for the various views to be linked, and the visual aspects of this are discussed in Chapter 13. Suppose both the map view and the table view are displayed on the screen simultaneously. Linkage allows the user to select objects in one view, perhaps by pointing and clicking, and to see the selected objects highlighted in both views, as in Figure 14.7. Linkage is often possible between other views, including the histogram and scatterplot views. For example, by linking a scatterplot with a map view, it is possible to select points in the scatterplot and see the corresponding objects highlighted on the map. This kind of linkage is very useful in examining *residuals* (Section 4.7), or cases that deviate substantially from the trend shown by a scatterplot (compare with the idealized scatterplots of Figures 4.14 and 4.15). Figure 14.8 shows four linked views of data on sudden infant death syndrome in the counties of North Carolina. The term *exploratory spatial data analysis* is sometimes used to describe these forms of interrogation, which allow the user to explore data in interesting and potentially insightful ways.

Exploratory spatial data analysis allows its users to gain insight by interacting with dynamically linked views.

Second, many methods are commonly available for interrogating the contents of tables, such as SQL (Section 10.4). Figure 14.7 shows the result of an SQL query on a simple table. The language becomes much more powerful when tables are linked, using common keys, as described in Section 10.3, and much more complex and sophisticated queries, involving multiple tables, are possible with the full language. More complex methods of table interrogation include the ability to average the values of an attribute across selected records, and to create new attributes through arithmetic operations on existing ones (e.g., create a new attribute equal to the ratio of two selected attributes).

SQL is a standard language for querying tables and relational databases.

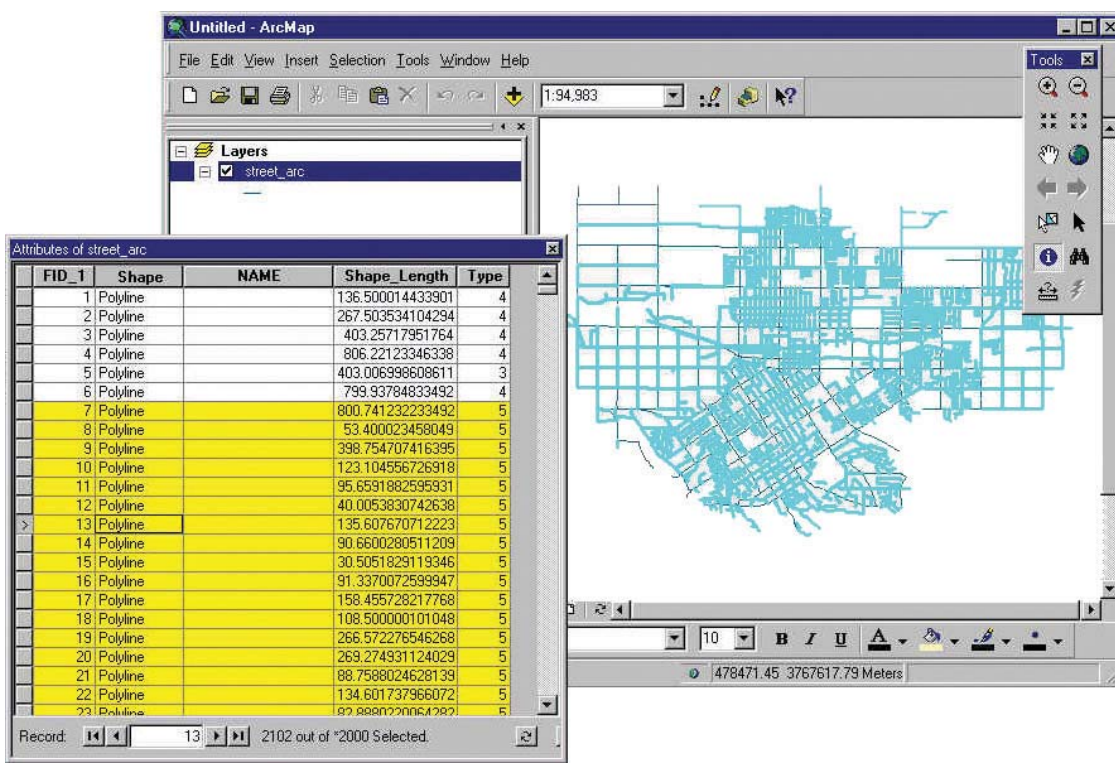


Figure 14.7 The objects shown in table view (left). In this instance all objects with Type equal to 5 have been selected, and the selected objects appear in blue in the map view (ESRI's ArcMap)

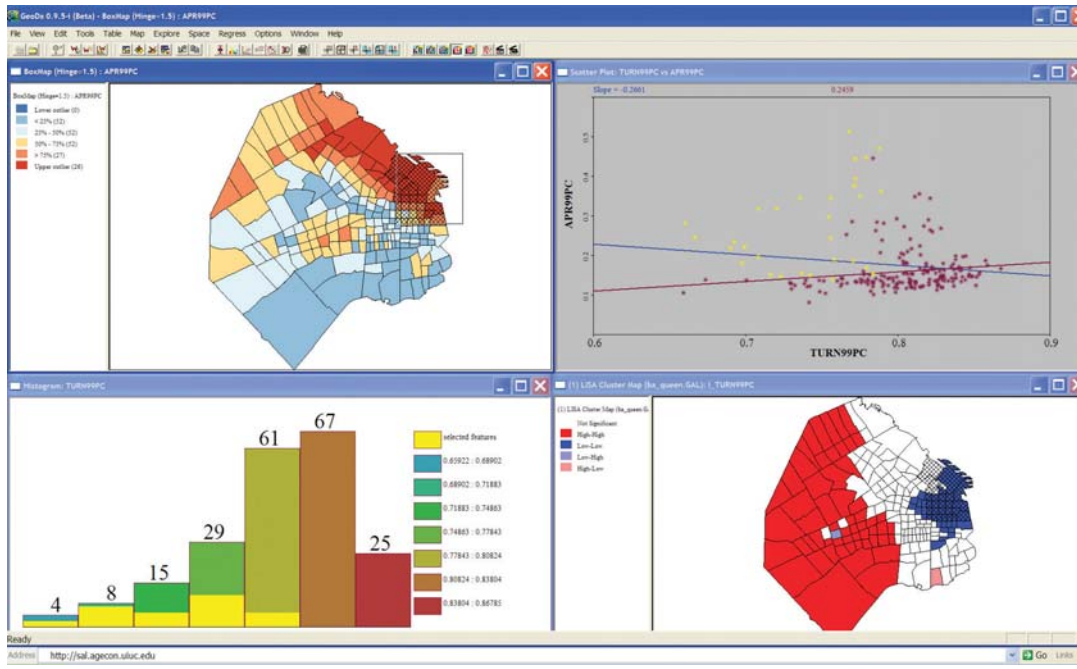


Figure 14.8 Screen shot of GeoDa, a spatial analysis package that integrates easily with a GIS and features the ability to display multiple views (clockwise from the top left a map, a scatterplot, a map showing local indicators of spatial association (LISA; Box 15.3), and a histogram – see Section 14.2 for a discussion of histograms and scatterplots) and to link them dynamically. The tracts selected by the user in the upper left are automatically highlighted in the other windows. GeoDa is available at **www.csiss.org/clearinghouse/GeoDa**. (*Source*: courtesy of Luc Anselin, University of Illinois, Urbana-Champaign)

14.3 Measurements

Many types of interrogation ask for measurements – we might want to know the total area of a parcel of land, or the distance between two points, or the length of a stretch of road – and in principle all of these measurements are obtainable by simple calculations inside a GIS. Comparable measurements by hand from maps can be very tedious and error-prone. In fact it was the ability of the computer to make accurate evaluations of area quickly that led the Canadian government to fund the development of the world's first GIS, the Canada Geographic Information System, in the mid-1960s (see the brief history of GIS in Section 1.4.1), despite the primitive state and high costs of computing at that time. Evaluation of area by hand is a messy and soul-destroying business. The *dot-counting* method uses transparent sheets on which randomly located dots have been printed – an area on the map is estimated by counting the number of dots falling within it. In the *planimeter* method a mechanical device is used to trace the area's boundary, and the required measure accumulates on a dial on the machine.

Humans have never devised good manual tools for making measurements from maps, particularly measurements of area.

By comparison, measurement of the area of a digitally represented polygon is trivial and totally reliable. The common algorithm (see Box 14.2 for a definition of this term) calculates and sums the areas of a series of trapezia, formed by dropping perpendiculars to the x axis as shown in Figure 14.9. By making a simple change to the algorithm it is also possible to use it to compute a polygon's centroid (see Section 15.2.1 for a discussion of centroids).

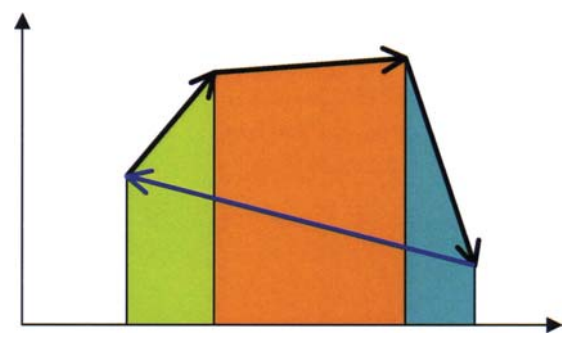


Figure 14.9 The algorithm for calculation of the area of a polygon given the coordinates of the polygon's vertices. The polygon consists of the three black arrows, plus the blue arrow forming the fourth side. Trapezia are dropped from each edge to the x axis and their areas are calculated as (difference in x) times average of y. The trapezia for the first three edges, shown in green, brown, and blue, are summed. When the fourth trapezium is formed from the blue arrow its area is negative because its start point has a larger x than its end point. When this area is subtracted from the total the result is the correct area of the polygon

Technical Box **14.2**

Definition of an algorithm

Algorithm: a procedure that provides a solution to a problem and consists of a set of unambiguous rules which specify a finite sequence of operations. Each step of an algorithm must be precisely defined and the necessary actions must be rigorously specified for each case. An algorithm should always arrive at a problem solution after a finite and reasonable number of steps. An algorithm that satisfies these requirements can be programmed as software for a digital computer.

14.3.1 Distance and length

A *metric* is a rule for the determination of distance between points in a space. Several kinds of metrics are used in GIS, depending on the application. The simplest is the rule for determining the shortest distance between two points in a flat plane, called the Pythagorean or straight-line metric. If the two points are defined by the coordinates (x_1, y_1) and (x_2, y_2), then the distance D between them is the length of the hypotenuse of a right-angled triangle (Figure 14.10), and Pythagoras's theorem tells us that the square of this length is equal to the sum of the squares of the lengths of the other two sides. So a simple formula results:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

A metric is a rule for determining distance between points in space.

The Pythagorean metric gives a simple and straightforward solution for a plane, if the coordinates x and y are comparable, as they are in any coordinate system based on a projection, such as the UTM or State Plane, or National Grid (see Chapter 5). But the metric will not work for latitude and longitude, reflecting a common source of problems in GIS – the temptation to treat latitude and longitude as if they were equivalent to plane coordinates. This issue is discussed in detail in Section 5.7.1.

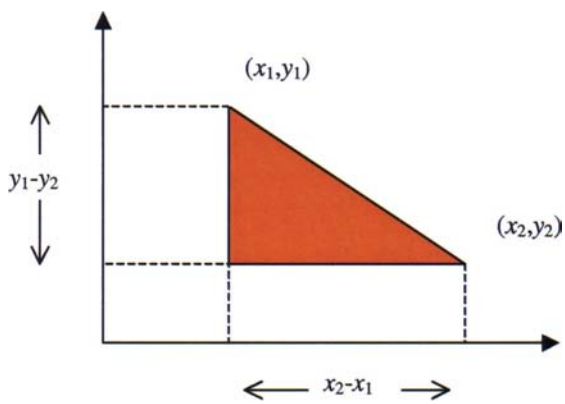


Figure 14.10 Pythagoras's Theorem and the straight-line distance between two points on a plane. The square of the length of the hypotenuse is equal to the sum of the squares of the lengths of the other two sides of the right-angled triangle

The assumption of a flat plane leads to significant distortion for points widely separated on the curved surface of the Earth, and distance must be measured using the metric for a spherical Earth given in Section 5.6 and based on a great circle. For some purposes even this is not sufficiently accurate because of the non-spherical nature of the Earth, and even more complex procedures must be used to estimate distance that take non-sphericity into account. Figure 14.11 shows an example of the differences that the curved surface of the Earth makes when flying long distances.

In many applications the simple rules – the Pythagorean and great circle equations – are not sufficiently accurate estimates of actual travel distance, and we are forced to resort to summing the actual lengths of travel routes. In GIS this normally means summing the lengths of links in a network representation, and many forms of GIS analysis use this approach. If a line is represented as a polyline, or a series of straight segments, then its length is simply the sum of the lengths of each segment, and each segment length can be calculated using the Pythagorean formula and the coordinates of its end points. But it is worth being aware of two problems with this simple approach.

First, a polyline is often only a rough version of the true object's geometry. A river, for example, never makes sudden changes of direction, and Figure 14.12 shows how smoothly curving streets have to be approximated by the sharp corners of a polyline. Box 4.6 discusses the case of a complex shape and the short-cutting that occurs in any generalization. Because there is a general tendency for polylines to short-cut corners, *the length of a polyline tends to be shorter than the length of the object it represents*. There are some exceptions, of course – surveyed boundaries are often truly straight between corner points and streets are often truly straight between intersections. But in general the lengths of linear objects estimated in a GIS, and this includes the lengths of the perimeters of areas represented as polygons, are often substantially shorter than their counterparts on the ground. Note that this is not similarly true of area estimates because short-cutting corners tends to produce both underestimates and overestimates of area, and these tend to cancel out (Figure 14.12; see also Box 4.6).

A GIS will almost always underestimate the true length of a geographic line.

Figure 14.11 The effects of the Earth's curvature on the measurement of distance, and the choice of shortest paths. The map shows the North Atlantic on the Mercator projection. The red line shows the track made by steering a constant course of 79 degrees from Los Angeles, and is 9807 km long. The shortest path from Los Angeles to London is actually the black line, the trace of the great circle connecting them, with a length of roughly 8800 km, and this is typically the route followed by aircraft flying from London to Los Angeles. When flying in the other direction, aircraft may sometimes follow other, longer tracks, such as the red line, if by doing so they can take advantage of jetstream winds

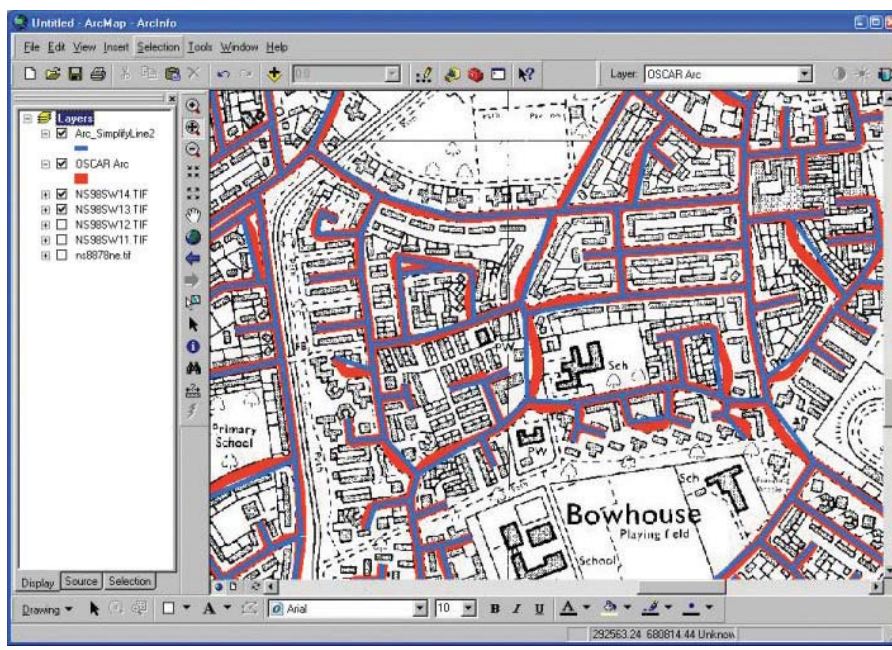


Figure 14.12 The polyline representations of smooth curves tend to be shorter in length, as illustrated by this street map (note how curves are replaced by straight-line segments). But estimates of area tend not to show systematic bias because the effects of overshoots and undershoots tend to cancel out to some extent

Second, the length of a line in a two-dimensional GIS representation will always be the length of the line's planar projection, not its true length in three dimensions, and the difference can be substantial if the line is steep (Figure 14.13). In most jurisdictions the area of a parcel of land is the area of its horizontal projection, not its true surface area. A GIS that stores the third dimension for every point is able to calculate both versions of length and area, but not a GIS that stores only the two horizontal dimensions.

14.3.2 Shape

GIS are also used to characterize the *shapes* of objects, particularly area objects. In many countries the system of political representation is based on the concept of districts or constituencies, which are used to define who will vote for each place in the legislature (Box 14.3). In the USA and the UK, and in many other countries that derived their system of representation from the UK, there is one place in the legislature for each district. It is expected that districts will be compact in shape, and the manipulation of a district's shape to achieve certain overt or covert objectives is termed gerrymandering, after an early governor of Massachusetts, Elbridge Gerry (the shape of one of the state's districts was thought to resemble a salamander, with the implication that it had been manipulated to achieve a certain outcome in the voting; Gerry was a signator both of the Declaration of Independence in 1776, and of the bill that created the offending districts in 1812). The construction of voting districts is an example of the principles of aggregation and zone design discussed in Section 6.2.

Anomalous shape is the primary means of detecting gerrymanders of political districts.

Geometric shape was the aspect that alerted Gerry's political opponents to the manipulation of districts, and today shape is measured whenever GIS is used to aid in the drawing of political district boundaries, as must occur

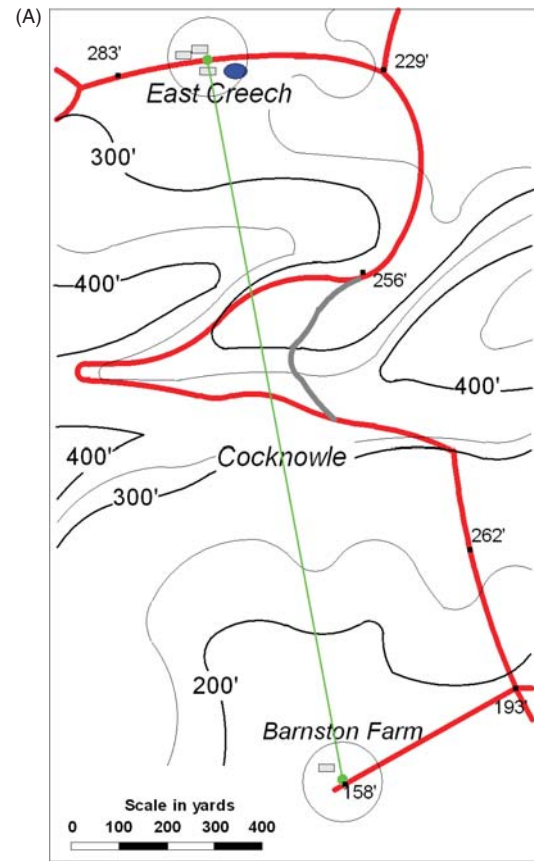


Figure 14.13 The length of a path as traveled on the Earth's surface (red line) may be substantially longer than the length of its horizontal projection as evaluated in a two-dimensional GIS. (A) shows three paths across part of Dorset in the UK. The green path is the straight route, the red path is the modern road system, and the gray path represents the route followed by the road in 1886. (B) shows the vertical profiles of all three routes, with elevation plotted against the distance traveled horizontally in each case. 1 foot = 0.3048 m, 1 yard = 0.9144 m (Courtesy: Michael De Smith)

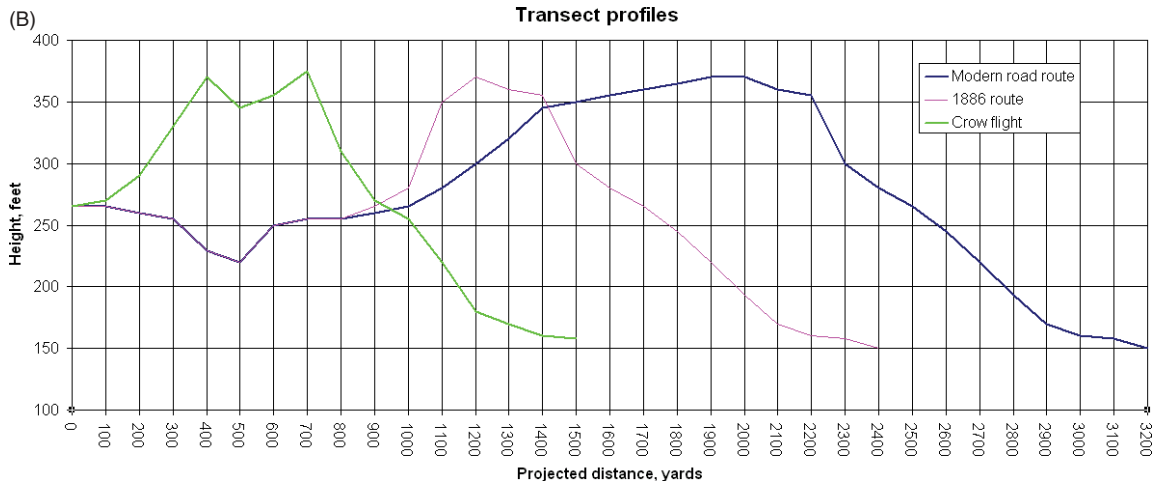


Figure 14.13 (*continued*)

Applications Box **14.3**

Shape and the 12th Congressional District of North Carolina

In 1992, following the release of population data from the 1990 Census, new boundaries were proposed for the voting districts of North Carolina, USA (Figure 14.14). For the first time race was used as an explicit criterion, and districts were drawn that as far as possible grouped minorities (notably African Americans) into districts in which they were in the majority. The intent was to avoid the historic tendency for minorities to be thinly spread in all districts, and thus to be unable to return their own representative to Congress. African Americans were in a majority in the new 12th District, but in order to achieve this the district had to be drawn in a highly contorted shape.

The new district, and the criteria used in the redistricting, were appealed to the US Supreme Court. In striking down the new districting scheme, Chief Justice William Rehnquist wrote for the 5–4 majority that 'A generalized assertion of past discrimination in a particular industry or region is not adequate because it provides no guidance for a legislative body to determine the precise scope of the injury it seeks to remedy. Accordingly, an effort to alleviate the effects of societal discrimination is not a compelling interest.'

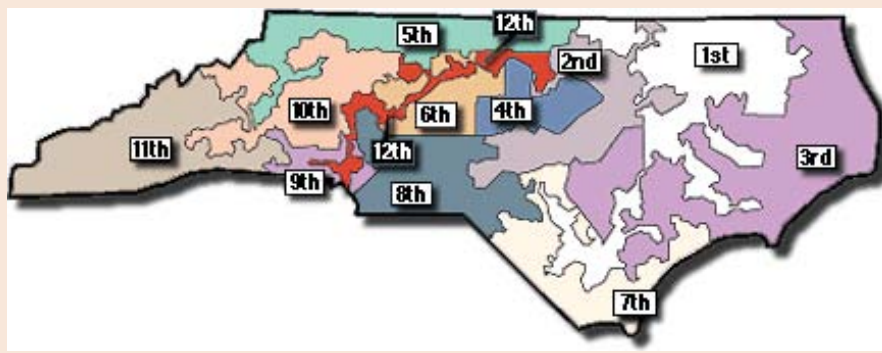


Figure 14.14 The boundaries of the 12th Congressional District of North Carolina drawn in 1992 show a very contorted shape, and were appealed to the US Supreme Court (*Source*: © Durham Herald Company, Inc., **www.herald-sun.com**)

by law in the USA after every decennial census. An easy way to define shape is by comparing the perimeter length of an area to its area measure. Normally the square root of area is used, to ensure that the numerator and denominator are both measured in the same units. A common measure of shape or compactness is:

$$S = P/3.54\sqrt{A}$$

where P is the perimeter length and A is the area. The factor 3.54 (twice the square root of π) ensures that the most compact shape, a circle, returns a shape of 1.0, and the most distended and contorted shapes return much higher values.

14.3.3 Slope and aspect

The most versatile and useful representation of terrain in GIS is the *digital elevation model*, or DEM. This is a raster representation, in which each grid cell records the elevation of the Earth's surface, and reflects a view of terrain as a field of elevation values. The elevation recorded is often the elevation of the cell's central point, but sometimes it is the mean elevation of the cell, and other rules have been used to define the cell's elevation (the rules used to define elevation in each cell of the US Geological Survey's GTOPO30 DEM, which covers the entire Earth's surface, vary depending on the source of data). Because of this variation, it is always advisable to read the available documentation to determine exactly what is meant by the recorded elevation in the cells of any DEM.

The digital elevation model is the most useful representation of terrain in a GIS.

Knowing the exact elevation of a point above sea level is important for some applications, including prediction of the effects of global warming and rising sea levels on coastal cities, but for many applications the value of a DEM lies in its ability to produce derivative measures through transformation, specifically measures of slope and

aspect, both of which are also conceptualized as fields. Imagine taking a large sheet of plywood and laying it on the Earth's surface so that it touches at the point of interest. The magnitude of steepest tilt of the sheet defines the *slope* at that point, and the direction of steepest tilt defines the *aspect* (Box 14.4).

This sounds straightforward but is complicated by a number of issues. First, what if the plywood fails to sit firmly on the surface, but instead pivots, because the point of interest happens to be a peak, or a ridge? In mathematical terms, we say that the surface at this point *lacks a well-defined tangent*, or that the surface at this point is *not differentiable*, meaning that it fails to obey the normal rules of continuous mathematical functions and differential calculus. The surface of the Earth has numerous instances of sharp breaks of slope, rocky outcrops, cliffs, canyons, and deep gullies that defy this simple mathematical approach to slope, and this is one of the issues that led Benoît Mandelbrot to develop his theory of fractals (see Section 4.8, and the additional applications described in Section 15.2.5).

A simple and satisfactory alternative is to take the view that slope must be measured at a particular resolution. To measure slope at a 30 m resolution, for example, we evaluate elevation at points 30 m apart and compute slope by comparing them (equivalent in concept to using a plywood sheet 30 m across). The value this gives is specific to the 30 m spacing, and a different spacing (or different-sized sheet of plywood) would have given a different result. In other words, *slope is a function of resolution*, and it makes no sense to talk about slope without at the same time talking about a specific resolution or level of detail. This is convenient because slope is easily computed in this way from a DEM with the appropriate resolution.

The spatial resolution used to calculate slope and aspect should always be specified.

Second, there are several alternative *measures* of slope, and it is important to know which one is used in a particular software package and application. Slope can be measured as an *angle*, varying from 0 to 90 degrees as the surface ranges from horizontal to vertical. But it can also be measured as a percentage or ratio, defined as *rise over run*, and unfortunately there are two different ways of defining run. Figure 14.16 shows the two options, depending on whether run means the horizontal distance covered between two points, or the diagonal distance (the *adjacent* or the *hypotenuse* of the right-angled triangle respectively). In the first case (opposite over adjacent) slope as a ratio is equal to the tangent of the angle of slope, and ranges from zero (horizontal) through 1 (45 degrees) to infinity (vertical). In the second case (opposite over hypotenuse) slope as a ratio is equal to the sine of the angle of slope, and ranges from zero (horizontal) through 0.707 (45 degrees) to 1 (vertical). To avoid confusion we will use the term slope only to refer to the measurement in degrees, and call the other options tan(slope) and sin(slope) respectively.

When a GIS calculates slope and aspect from a DEM, it does so by estimating slope at each of the data points of the DEM, by comparing the elevation at that point to the elevations of surrounding points. But the number of surrounding points used in the calculation varies, as

Technical Box **14.4**

Calculation of slope based on the elevations of a point and its eight neighbors

The equations used are as follows (refer to Figure 14.15 for point numbering):

$$b = (z_3 + 2z_6 + z_9 - z_1 - 2z_4 - z_7)/8D$$

$$c = (z_1 + 2z_2 + z_3 - z_7 - 2z_8 - z_9)/8D$$

where b and c are tan(slope) in the x and y directions respectively, D is the grid point spacing, and z_i denotes elevation at the ith point, as shown below. These equations give the four diagonal neighbors of Point 5 only half the weight of the other four neighbors in determining slope at Point 5.

$$\tan(slope) = \sqrt{b^2 + c^2}$$

where *slope* is the angle of slope in the steepest direction.

$$\tan(aspect) = b/c$$

where *aspect* is the angle between the y axis and the direction of steepest slope, measured clockwise. Since *aspect* varies from 0 to 360, an additional test is necessary that adds 180 to *aspect* if c is positive.

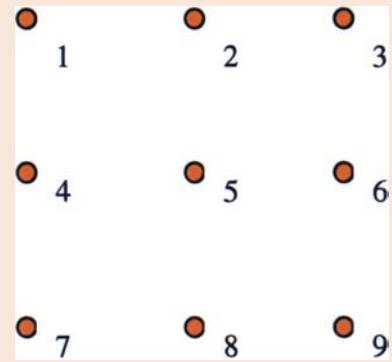


Figure 14.15 Calculation of the slope at Point 5 based on the elevation of it and its eight neighbors

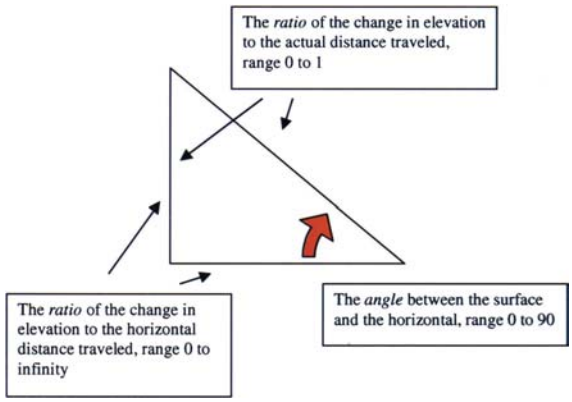


Figure 14.16 Three alternative definitions of slope. To avoid ambiguity we use the angle, which varies between 0 and 90 degrees

do the weights given to each of the surrounding points in the calculation. Box 14.4 shows this idea in practice, using one of the commonest methods, which employs eight surrounding points and gives them different weights depending on how far away they are.

Slope and aspect are the basis for many interesting and useful forms of analysis. Slope is an input to many models of the soil erosion and runoff that result from heavy storms. Slope is also an important input to analyses that find the most suitable routes across terrain for power lines, highways, and military vehicles (see Section 15.3.3).

14.4 Transformations

In this section, we look at methods that transform GIS objects and databases into more useful products, using simple rules. These operations form the basis for many applications, because they are capable of revealing aspects that are not immediately visible or obvious.

14.4.1 Buffering

One of the most important transformations available to the GIS user is the *buffer* operation (see also Section 10.5). Given any set of objects, which may include points, lines, or areas, a buffer operation builds a new object or objects by identifying all areas that are within a certain specified distance of the original objects. Figure 14.17 shows instances of a point, a line, and an area, and the results of buffering. Buffers have many uses, and they are among the most popular of GIS functions:

- The owner of a land parcel has applied for planning permission to rebuild – the local planning authority could build a buffer around the parcel, in order to identify all homeowners who live within the legally mandated distance for notification of proposed redevelopments.
- A logging company wishes to clearcut an area, but is required to avoid cutting in areas within 100 m of streams – the company could build buffers 100 m wide around all streams to identify these protected riparian areas.
- A retailer is considering developing a new store on a site of a type that is able to draw consumers from up to 4 km away from its stores – the retailer could build a buffer around the site to identify the number of consumers living within 4 km of the site, in order to estimate the new store's potential sales (see Section 2.3.3.3).

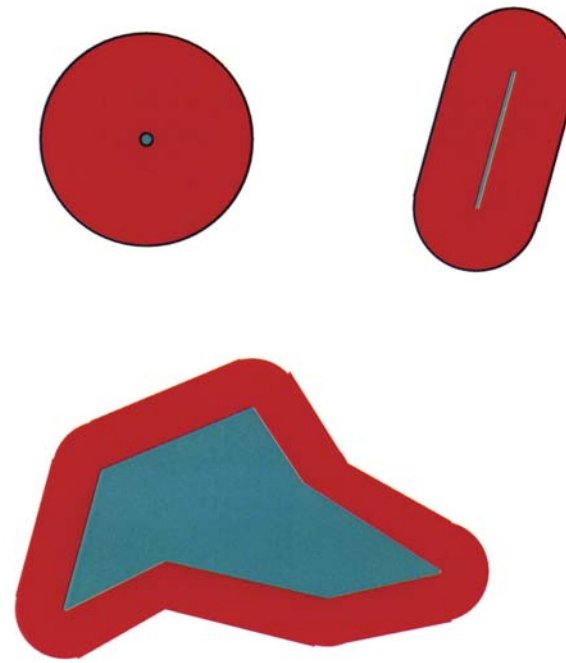


Figure 14.17 Buffers (dilations) of constant width drawn around a point, a polyline, and a polygon

Buffering is possible in both raster and vector GIS – in the raster case, the result is the classification of cells according to whether they lie inside or outside the buffer, while the result in the vector case is a new set of objects (Figure 14.17). But there is an additional possibility in the raster case that makes buffering more useful in some situations. Figure 14.18 shows a city; average travel speeds vary in each cell of the raster outside the city. Rather than buffer according to distance from the city, we can ask a raster GIS to *spread* outwards from the city at rates determined by the travel speed values in each cell. Where travel speeds are high the spread will extend further, so we can compute how far it is possible to go from the city in a given period of time. This idea of spreading over a variable or *friction* surface is easily implemented in raster representations, but very difficult in vector. Another form of analysis that uses a raster surface to control rate of movement is discussed in Section 15.3.3.

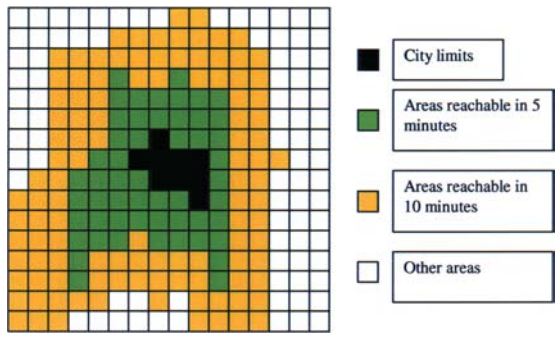


Figure 14.18 A raster generalization of the buffer function, in which spreading occurs at rates controlled by a variable (travel speed, or friction, the inverse of travel speed) whose value is recorded in every raster cell

Buffering is one of the most useful transformations in a GIS and is possible in both raster and vector formats.

14.4.2 Point in polygon

In its simplest form, the point in polygon operation determines whether a given point lies inside or outside a given polygon. In more elaborate forms, there may be many polygons, and many points, and the task is to assign points to polygons. If the polygons overlap, it is possible that a given point lies in one, many, or no polygons, depending on its location. Figure 14.19 illustrates the task. The operation is popular in GIS analysis because it is the basis for answering many simple queries:

- The points represent instances of a disease in a population, and the polygons represent reporting zones such as counties – the task is to determine how many instances of the disease occurred in each zone (in this case the zones should not overlap and each point should fall into exactly one polygon).
- The points represent the locations of transmission-line poles owned by a utility company, and the polygons are parcels of land – the task is to identify the owner of the land on which each pole lies, to verify that the company has the necessary easements and pays the necessary fees.
- The points represent the residential locations of voters, and the polygons represent voting districts – the task is to ensure that each voter receives the correct voting forms in the mail.

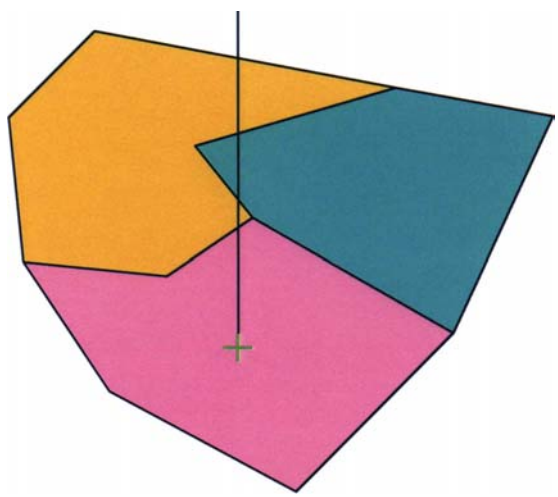


Figure 14.19 The point in polygon problem, shown in the continuous-field case (the point must by definition lie in exactly one polygon, or outside the project area). In only one instance (the pink polygon) is there an odd number of intersections between the polygon boundary and a line drawn vertically upward from the point

The point in polygon operation makes sense from both the discrete-object and the continuous-field perspectives (see Section 3.5 for a discussion of these two perspectives). From a discrete-object perspective, both points and polygons are objects and the task is simply to determine enclosure. From a continuous-field perspective, polygons representing a variable such as land ownership cannot overlap, since each polygon represents the land owned by one owner and overlap would imply that a point is owned simultaneously by two owners. Similarly from a continuous-field perspective there can be no gaps between polygons. Consequently, the result of a point in polygon operation from a continuous-field perspective must assign each point to exactly one polygon.

The point in polygon operation is used to determine whether a point lies inside or outside a polygon.

The standard algorithm for the point in polygon operation is shown in Figure 14.19. In essence, it consists of drawing a line from the point to infinity, in this case parallel to the *y* axis, and determining the number of intersections between the line and the polygon's boundary. If the number is odd the point is inside the polygon and if it is even the point is outside. The algorithm must deal successfully with special cases – for example, if the point lies directly below a vertex (corner point) of the polygon. Some algorithms extend the task to include a third option, when the point lies exactly on the boundary. But others ignore this, on the grounds that it is never possible in practice to determine location with perfect accuracy and so never possible to determine if an infinitely small point lies on or off an infinitely thin boundary line.

14.4.3 Polygon overlay

Polygon overlay is similar to point in polygon transformation in the sense that two sets of objects are involved, but in this case both are polygons. It exists in two forms, depending on whether a continuous-field or discrete-object perspective is taken. The development of effective algorithms for polygon overlay was one of the most significant challenges of early GIS and the task remains one of the most complex and difficult to program.

The complexity of computing a polygon overlay was one of the greatest barriers to the development of vector GIS.

From the discrete-object perspective, the task is to determine whether two area objects overlap, to determine the area of overlap, and to define the area formed by the overlap as one or more new area objects (the overlay of two polygons can produce a large number of distinct area objects, see Figure 14.20). This operation is useful to determine answers to such queries as:

- How much of this proposed clearcut lies in this riparian zone?
- How much of the projected catchment area of this proposed retail store lies in the catchment of this other existing store in the same chain?
- How much of this land parcel is affected by this easement?
- What proportion of the land area of the USA lies in areas managed by the Bureau of Land Management?

From the continuous-field perspective the task is somewhat different. Figure 14.21 shows two datasets, both representations of fields – one differentiates areas according to land ownership and the other differentiates the same region according to land cover class. In the terminology of ESRI's ArcGIS, both datasets are instances of *area coverages*, or fields of nominal variables represented by non-overlapping polygons. The methods discussed earlier in this chapter could be used to interrogate either dataset separately, but there are numerous queries that require simultaneous access to both datasets – for example:

- What is the land cover class and who is the owner of the point indicated by the user?
- What is the total area of land owned by X and with land cover class A?

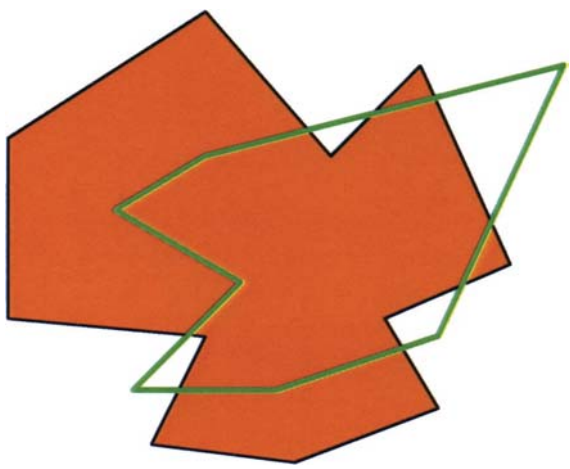


Figure 14.20 Polygon overlay, in the discrete-object case. Here, the overlay of two polygons produces nine distinct polygons. One has the properties of both polygons, four have the properties of the brown polygon but not the green polygon, and four are outside the brown polygon but inside the green polygon

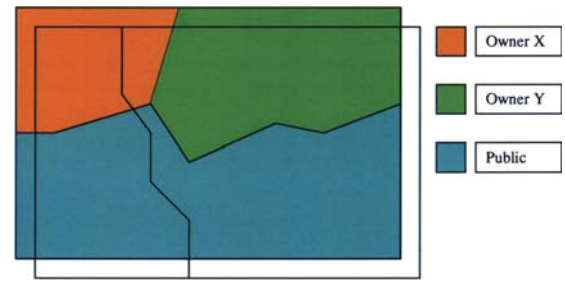


Figure 14.21 Polygon overlay in the continuous-field case. Here, a dataset representing two types of land cover (A on the left, B on the right) is overlaid on a dataset representing three types of ownership (the two datasets have been offset slightly for clarity). The result will be a single dataset in which every point is identified with one land cover type and one ownership type. It will have five polygons, since land cover A intersects with two ownership types and land cover B intersects with three

- Where are the areas that lie on publicly owned land and have land cover class B?

None of these queries can be answered by interrogating one of the datasets alone – the datasets must somehow be combined so that interrogation can be directed simultaneously at both of them.

The continuous-field version of polygon overlay does this by first computing a new dataset in which the region is partitioned into smaller areas that have uniform characteristics on both variables. Each area in the new dataset will have two sets of attributes – those obtained from one of the input datasets and those obtained from the other. All of the boundaries will be retained, but they will be broken into shorter fragments by the intersections that occur between boundaries in one input dataset and boundaries in the other. Note the unusual characteristics of the new dataset shown in Figure 14.21. Unlike the two input datasets, where boundaries meet in junctions of three lines, the new map contains a new junction of four lines, formed by the new intersection discovered during the overlay process. Because the results of overlay are distinct in this way it is almost always possible to discover whether a GIS dataset was formed by overlaying two earlier datasets.

Polygon overlay has different meanings from the continuous-field and discrete-object perspectives.

With a single dataset that combines both inputs, it is an easy matter to answer all of the queries listed above through simple interrogation. It is also easy to reverse the overlay process – if neighboring areas that share the same land cover class are merged, for example, the result is the land ownership map, and vice versa.

Polygon overlay is a computationally complex operation and as noted earlier much work has gone into developing algorithms that function efficiently for large datasets. One of the issues that must be tackled by a practically useful algorithm is known as the *spurious polygon* or *coastline weave* problem. It is almost inevitable that there will be instances in any practical application where the same line on the ground occurs in both datasets. This

happens, for example, when a coastal region is being analyzed, because the coastline is almost certain to appear in every dataset of the region. Rivers and roads often form boundaries in many different datasets – a river may function both as a land cover class boundary and as a land ownership boundary, for example. But although the same line is represented in both datasets, its representations will almost certainly not be the same. They may have been digitized from different maps, digitized using different numbers of points, subjected to different manipulations, obtained from entirely different sources (an air photograph and a topographic map, for example), and subjected to different measurement errors. When overlaid, the result is a series of small slivers. Paradoxically, the more care one takes in digitizing or processing, the worse the problem appears, as the result is simply more slivers, albeit smaller in size.

In two vector datasets of the same area there will almost certainly be instances where lines in each dataset represent the same feature on the ground.

Table 14.1 shows an example of the consequences of slivers and how a GIS can be rapidly overwhelmed if it fails to anticipate and deal with them adequately. Today, a GIS will offer various methods for dealing with the problem, the most common of which is the specification of a *tolerance*. If two lines fall within this distance of each other, the GIS will treat them as a single line and not create slivers. The resulting overlay contains just one version of the line, not two. But at least one of the input lines has been moved, and if the tolerance is set too high the movement can be substantial and lead to problems later.

Overlay in raster is an altogether simpler operation, and this has often been cited as a good reason to adopt raster rather than vector structures. When two raster layers are overlaid, the attributes of each cell are combined according to a set of rules. For example, suppose the task is to find all areas that belong to owner A and have land use class B. Areas with these characteristics would be assigned a value, perhaps 1, and all other areas would be assigned a value of 0. Note the important difference between raster and vector overlay – in vector overlay there is no rule for combination, and instead the result of overlay contains all of the input information, rearranged and combined so that it can be used to respond to queries and can be subjected to analysis. Figure 14.22 shows an example of raster overlay.

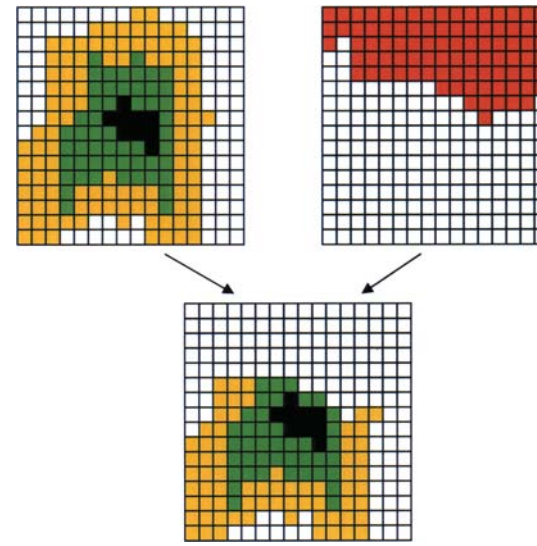


Figure 14.22 The raster overlay case, resulting in a new dataset that applies a set of rules to the input datasets. The two input datasets are maps of (top left) travel time from the urban area (see Figure 14.18) and (top right) county (red indicates County X, white indicates County Y). The output map identifies travel time to areas in County Y and might be used to compute average travel time to points in that county in a subsequent step. This operation is not reversible

Table 14.1 Numbers of polygons resulting from an overlay of five datasets, illustrating the spurious polygon problem. The datasets come from the Canada Geographic Information System discussed in Section 1.4.1, and all are representations of continuous fields. Dataset 1 is a representation of a map of soil capability for agriculture (Figure 15.13 shows a map of this type), datasets 2 through 4 are land-use maps of the same area at different times (the probability of finding the same real boundary in more than one such map is very high), and dataset 5 is a map of land capability for recreation. The final three columns show the numbers of polygons in overlays of three, four, and five of the input datasets

Acres	1	2	3	4	5	1 + 2 + 5	1 + 2 + 3 + 5	1 + 2 + 3 + 4 + 5
0–1	0	0	0	1	2	2640	27566	77346
1–5	0	165	182	131	31	2195	7521	7330
5–10	5	498	515	408	10	1421	2108	2201
10–25	1	784	775	688	38	1590	2106	2129
25–50	4	353	373	382	61	801	853	827
50–100	9	238	249	232	64	462	462	413
100–200	12	155	152	158	72	248	208	197
200–500	21	71	83	89	92	133	105	99
500–1000	9	32	31	33	56	39	34	34
1000–5000	19	25	27	21	50	27	24	22
>5000	8	6	7	6	11	2	1	1
Totals	88	2327	2394	2149	487	9558	39188	90599

Raster overlay is simpler, but it produces a fundamentally different kind of result.

14.4.4 Spatial interpolation

Spatial interpolation is a pervasive operation in GIS. Although it is often used explicitly in analysis, it is also used implicitly, in various operations such as the preparation of a contour map display, where spatial interpolation is invoked without the user's direct involvement. Spatial interpolation is a process of intelligent guesswork, in which the investigator (and the GIS) attempt to make a reasonable estimate of the value of a continuous field at places where the field has not actually been measured. Spatial interpolation is an operation that makes sense only from the continuous-field perspective. The principles of spatial interpolation are discussed in Section 4.5; here the emphasis is on practical applications of the technique and commonly used implementations of the principles.

Spatial interpolation finds applications in many areas:

- In estimating rainfall, temperature, and other attributes at places that are not weather stations and where no direct measurements of these variables are available.
- In estimating the elevation of the surface between the measured locations of a DEM.
- In *resampling* rasters, the operation that must take place whenever raster data must be transformed to another grid.
- In contouring, when it is necessary to guess where to place contours between measured locations.

In all of these instances, spatial interpolation calls for intelligent guesswork, and the one principle that underlies all spatial interpolation is the Tobler Law (Section 3.1) – 'all places are related but nearby places are more related than distant places'. In other words, the best guess as to the value of a field at some point is the value measured at the closest observation points – the rainfall *here* is likely to be more similar to the rainfall recorded at the nearest weather stations than to the rainfall recorded at more distant weather stations. A corollary of this same principle is that in the absence of better information, it is reasonable to assume that any continuous field exhibits relatively smooth variation – fields tend to vary slowly and to exhibit strong positive spatial autocorrelation, a property of geographic data discussed in Section 4.6.

Spatial interpolation is the GIS version of intelligent guesswork.

In this section three methods of spatial interpolation are discussed: Thiessen polygons; inverse-distance weighting (IDW), which is the simplest commonly used method; and Kriging, a popular statistical method that is grounded in the theory of regionalized variables and falls within the field of *geostatistics*.

14.4.4.1 Thiessen polygons

Thiessen polygons were suggested by Thiessen as a way of interpolating rainfall estimates from a few rain gauges to obtain estimates at other locations where rainfall had not been measured. The method is very simple: to estimate rainfall at any point take the rainfall measured at the closest gauge. This leads to a map in which rainfall is constant within polygons surrounding each gauge, and changes sharply as polygon boundaries are crossed. Although many GIS users associate polygons defined in this way with Thiessen, they are also known as Voronoi and Dirichlet polygons (Box 14.5). They have many other uses besides spatial interpolation:

- Thiessen polygons can be used to estimate the trade areas of each of a set of retail stores or shopping centers.
- They are used internally in the GIS as a means of speeding up certain geometric operations, such as search for nearest neighbor.
- They are the basis of some of the more powerful methods for generalizing vector databases.

As a method of spatial interpolation they leave something to be desired, however, because the sharp change in interpolated values at polygon boundaries is often implausible.

Figure 14.23 shows a typical set of Thiessen polygons. If each pair of points that share a Thiessen polygon boundary is connected, the result is a network of irregular triangles. These are named after Delaunay, and frequently used as the basis for the triangles of a TIN representation of terrain (Chapter 8).

14.4.4.2 Inverse-distance weighting

IDW is the workhorse of spatial interpolation, the method that is most often used by GIS analysts. It employs the Tobler Law by estimating unknown measurements as weighted averages over the known measurements

Figure 14.23 Thiessen polygons drawn around each station in part of the Southern California ozone-monitoring network. Note how the polygons, which enclose the area closest to each point, in theory extend off the map to infinity, so must be truncated by the GIS at the edge of the map

Biographical Box **14.5**

Atsu Okabe, expert on Thiessen polygons

Figure 14.24 Atsu Okabe, expert on Thiessen polygons

Atsuyuki Okabe received his Ph.D. from the University of Pennsylvania in 1975 and the degree of Doctor of Engineering from the University of Tokyo in 1977. He is Director of the Center for Spatial Information Science (1998–2005) and Professor of the Department of Urban Engineering, University of Tokyo. His research interests include computational spatial analysis, spatial statistics, spatial optimization, and spatial psychology. He has published many papers in journals, books, and conference proceedings on these topics. He is a co-author (with B. Boots, K. Sugihara, and S. N. Chiu) of *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams* (published by John Wiley). He currently leads the team developing a free GIS-based toolbox for spatial analysis on a network, called SANET.

Asked what had surprised him most about the development of GIS, he wrote: 'It is amazing to consider the remarkable change in spatial analysis brought about by GIS. In the early 1970s when I began to study spatial analysis, it remained conceptual. Digital data were scarce and researchers had to strive to assemble datasets. There were few efficient algorithms for geometrical calculations, and computation often took more than a week. Consequently, methods were difficult to test. I could, for example, consider the concepts of Voronoi diagrams, but it was hard to implement them. Try to construct a Voronoi diagram of five generators by hand, for instance! The first breakthrough was made in the mid 1980s, when 'Computational Geometry' (named by M.I. Shamos in 1975) became established through the book by Preparata and Shamos. The second development occurred in the late 1980s when GIS software became readily available. Surprisingly, almost all modern GIS software packages handle Voronoi diagrams, which we can now construct easily. Furthermore, recent progress in the design of GIS-based toolboxes has changed spatial analysis from an exclusive technique used only in academic research to a routine process used in a wide variety of professional activities. In the same way that Excel has made statistics available for general application, I hope that GIS-based spatial analysis will not only enhance studies in the humanities and social/natural sciences, but also come to be useful in everyday life.'

at nearby points, giving the greatest weight to the nearest points.

More specifically, denote the point of interest as $\mathbf{x}$, and the points where measurements were taken as $\mathbf{x}_i$, where i runs from 1 to n, if there are n data points. Denote the unknown value as $z(\mathbf{x})$ and the known measurements as z_i. Give each of these points a weight d_i, which will be evaluated based on the distance from $\mathbf{x}_i$ to $\mathbf{x}$. Figure 14.25 explains this notation with a diagram. Then the weighted average computed at $\mathbf{x}$ is:

$$z(\mathbf{x}) = \sum_i w_i z_i / \sum_i w_i$$

In other words, the interpolated value is an average over the observed values, weighted by the ws.

There are various ways of defining the weights, but the option most often employed is to compute them as

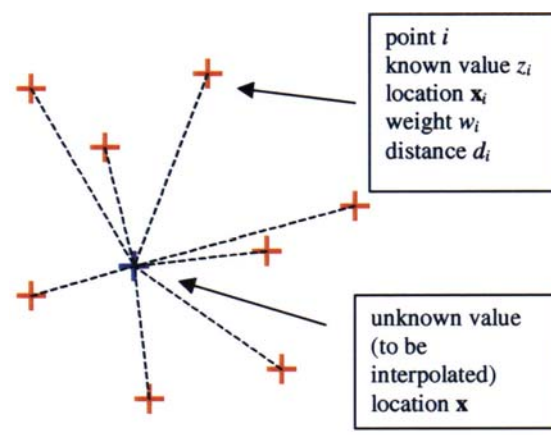


Figure 14.25 Notation used in the equations defining spatial interpolation

the inverse squares of distances, in other words (compare the options discussed in Section 4.5):

$$w_i = 1/d_i^2$$

This means that the weight given to a point drops by a factor of 4 when the distance to the point doubles (or by a factor of 9 when the distance triples). In addition, most software gives the user the option of ignoring altogether points that are further than some specified distance away, or of limiting the average to a specified number of nearest points, or of averaging over the closest points in each of a number of direction sectors. But if these values are not specified the software will assign default values to them.

IDW provides a simple way of guessing the values of a continuous field at locations where no measurement is available.

IDW achieves the desired objective of creating a smooth surface whose value at any point is more like the values at nearby points than the values at distant points. If it is used to determine z at a location where z has already been measured it will return the measured value, because the weight assigned to a point at zero distance is infinite, and for this reason IDW is described as an *exact* method of interpolation because its interpolated results honor the data points exactly. (An *approximate* method is allowed to deviate from the measured values in the interests of greater smoothness, a property which is often useful if deviations are interpreted as indicating possible errors of measurement or local deviations that are to be separated from the general trend of the surface.)

But because IDW is an average it suffers from certain specific characteristics that are generally undesirable. A weighted average that uses weights that are never negative must always return a value that is between the limits of the measured values – no point on the interpolated surface can have an interpolated z that is more than the largest measured z, or less than the smallest measured z. Imagine an elevation surface with some peaks and pits where the peaks and pits have not actually been measured but are merely indicated by the values of the measured points. Figure 14.26 shows a cross-section of such a surface. Instead of interpolating peaks and pits as one might expect, IDW produces the kind of result shown in the figure – small pits where there should be peaks, and small peaks where there should be pits. This behavior is often obvious in GIS output that has been generated using IDW. A related problem concerns extrapolation – if a trend is indicated by the data, as shown in Figure 14.26, IDW will inappropriately indicate a regression to the mean outside the area of the data points.

IDW interpolation may produce counter-intuitive results in areas of peaks and pits, and outside the area covered by the data points.

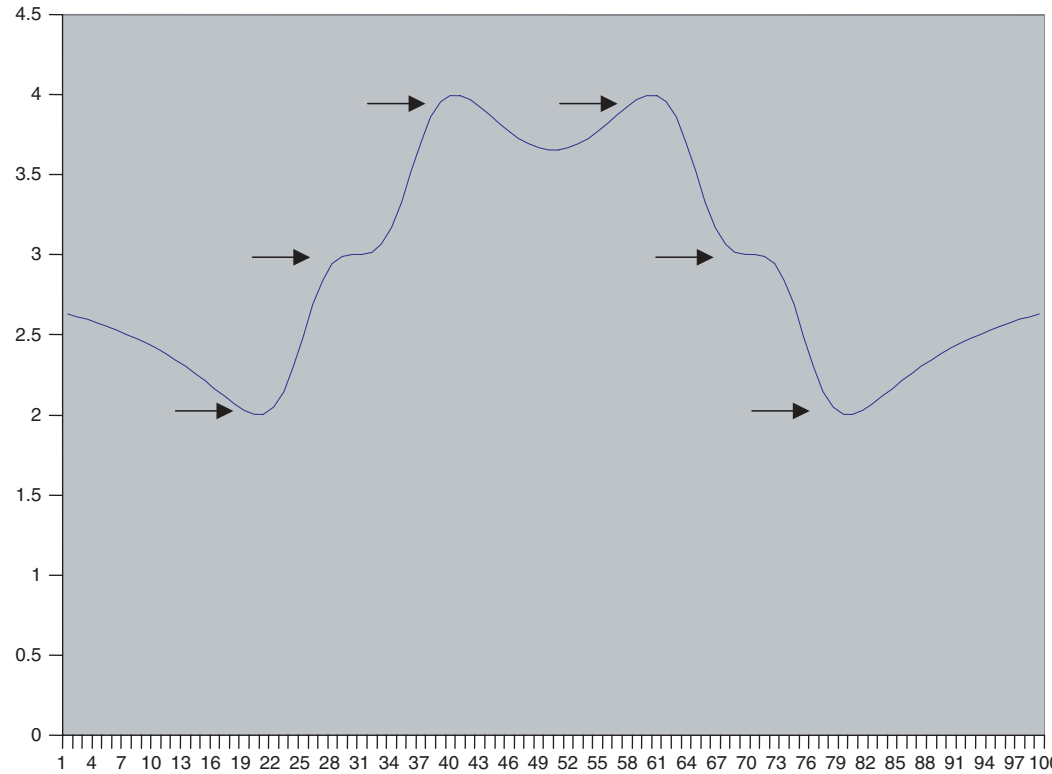


Figure 14.26 Potentially undesirable characteristics of IDW interpolation. Data points located at 20, 30, 40, 60, 70, and 80 have measured values of 2, 3, 4, 4, 3, and 2 respectively. The interpolated profile shows a pit between the two highest values, and regression to the overall mean value of 3 outside the area covered by the data

In short, the results of IDW are not always what one would want. There are many better methods of spatial interpolation that address the problems that were just identified, but the ease of programming of IDW and its conceptual simplicity make it among the most popular. Users should simply beware, and take care to examine the results of interpolation to ensure that they make good sense.

14.4.4.3 Kriging

Of the common methods of spatial interpolation, Kriging makes the most convincing claim to be grounded in good theoretical principles. The basic idea is to discover something about the general properties of the surface, as revealed by the measured values, and then to apply these properties in estimating the missing parts of the surface. Smoothness is the most important property (note the inherent conflict between this and the properties of fractals, Section 4.8), and it is operationalized in Kriging in a statistically meaningful way. There are many forms of Kriging, and the overview provided here is very brief. Further reading is identified at the end of the chapter.

There are many forms of Kriging, but all are firmly grounded in theory.

Suppose we take a point $\mathbf{x}$ as a reference and start comparing the values of the field there with the values at other locations at increasing distances from the reference point. If the field is smooth (if the Tobler Law is true, that is, if there is positive spatial autocorrelation) the values nearby will not be very different – $z(\mathbf{x})$ will not be very different from $z(\mathbf{x}_i)$. To measure the amount, we take the difference and square it, since the sign of the difference is not important: $(z(\mathbf{x}) - z(\mathbf{x}_i))^2$. We could do this with any pair of points in the area.

As distance increases, this measure will likely increase also, and in general a monotonic (consistent) increase in squared difference with distance is observed for most geographic fields (note that z must be measured on a scale that is at least interval, though *indicator Kriging* has been developed to deal with the analysis of nominal fields). In Figure 14.27, each point represents one pair of values drawn from the total set of data points at which measurements have been taken. The vertical axis represents one half of the squared difference (one half is taken for mathematical reasons), and the graph is known as the *semivariogram* (or *variogram* for short – the difference of a factor of two is often overlooked in practice, though it is important mathematically). To express its contents in summary form, the distance axis is divided into a number of ranges or *bins*, as shown, and points within each range are averaged to define the heavy points shown in the figure.

This semivariogram has been drawn without regard to the *directions* between points in a pair. As such it is said to be an *isotropic* variogram. Sometimes there is sharp variation in the behavior in different directions, and *anisotropic* semivariograms are created for different ranges of direction (e.g., for pairs in each 90 degree sector).

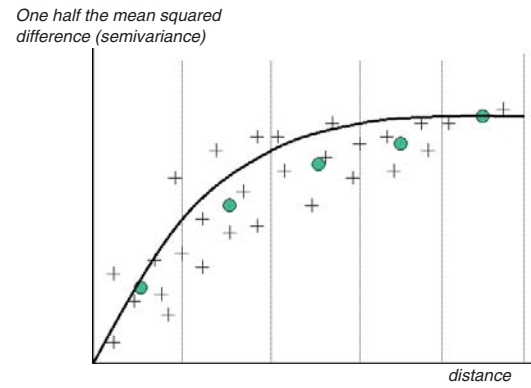


Figure 14.27 A semivariogram. Each cross represents a pair of points. The solid circles are obtained by averaging within the ranges or *buckets* of the distance axis. The solid line is the best fit to these five points, using one of a small number of standard mathematical functions

An anisotropic variogram asks how spatial dependence changes in different directions.

Note how the points of this typical variogram show a steady increase in squared difference up to a certain limit, and how that increase then slackens off and virtually ceases. Again, this pattern is widely observed for fields, and it indicates that difference in value tends to increase up to a certain limit, but then to increase no further. In effect, there is a distance beyond which there are no more geographic surprises. This distance is known as the *range*, and the value of semivariance at this distance as the *sill*.

Note also what happens at the other, lower end of the distance range. As distance shrinks, corresponding to pairs of points that are closer and closer together, the semivariance falls, but there is a suggestion that it never quite falls to zero, even at zero distance. In other words, if two points were sampled a vanishingly small distance apart they would give different values. This is known as the *nugget* of the semivariogram. A non-zero nugget occurs when there is substantial error in the measuring instrument, such that measurements taken a very small distance apart would be different due to error, or when there is some other source of local noise that prevents the surface being truly smooth. Accurate estimation of a nugget depends on whether there are pairs of data points sufficiently close together. In practice the sample points may have been located at some time in the past, outside the user's control, or may have been spread out to capture the overall variation in the surface, so it is often difficult to make a good estimate of the nugget.

The nugget can be interpreted as the variation among repeated measurements at the same point.

To make estimates using Kriging, we need to reduce the semivariogram to a mathematical function, so that semivariance can be evaluated at any distance, not just at the midpoints of buckets as shown in Figure 14.27. In practice, this means selecting one from a set of standard functional forms and fitting that form to the observed data points to get the best possible fit. This is shown in

the figure. The user of a Kriging function in a GIS will have control over the selection of distance ranges and functional forms, and whether a nugget is allowed.

Finally, the fitted semivariogram is used to estimate the values of the field at points of interest. As with IDW, the estimate is obtained as a weighted combination of neighboring values, but the estimate is designed to be the best possible given the evidence of the semivariogram. In general, nearby values are given greater weight, but unlike IDW direction is also important – a point can be *shielded* from influence if it lies behind another point, since the latter's greater proximity suggests greater importance in determining the estimated value, whereas relative direction is unimportant in an IDW estimate. The process of maximizing the quality of the estimate is carried out mathematically, using the precise measures available in the semivariogram.

Kriging responds both to the proximity of sample points and to their directions.

Unlike IDW, Kriging has a solid theoretical foundation, but it also includes a number of options (e.g., the choice of the mathematical function for the semivariogram) that require attention from the user. In that sense it is definitely not a *black box* that can be executed blindly and automatically (Section 2.3.5.5), but instead forces the user to become directly involved in the estimation process. For that reason GIS software designers will likely continue to offer several different methods, depending on whether the user wants something that is quick, despite its obvious faults, or better but more demanding of the user.

14.4.4.4 Density estimation

Density estimation is in many ways the logical twin of spatial interpolation – it begins with points and ends with a surface. But conceptually the two approaches could not be more different, because one seeks to estimate the missing parts of a continuous field from samples of the field taken at data points, while the other creates a continuous field from discrete objects.

Figure 14.28 illustrates this difference. The dataset can be interpreted in two sharply different ways. In the first, it is interpreted as sample measurements from a continuous field, and in the second as a collection of discrete objects. In the discrete-object view there is nothing between the objects but empty space – no missing field to be filled in through spatial interpolation. It would make no sense at all to apply spatial interpolation to a collection of discrete

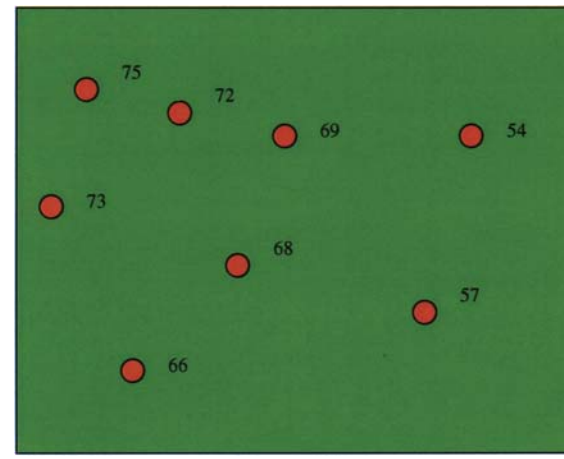


Figure 14.28 A dataset with two possible interpretations: first, a continuous field of atmospheric temperature measured at eight irregularly spaced sample points, and second, eight discrete objects representing cities, with associated populations in thousands. Spatial interpolation makes sense only for the former, and density estimation only for the latter

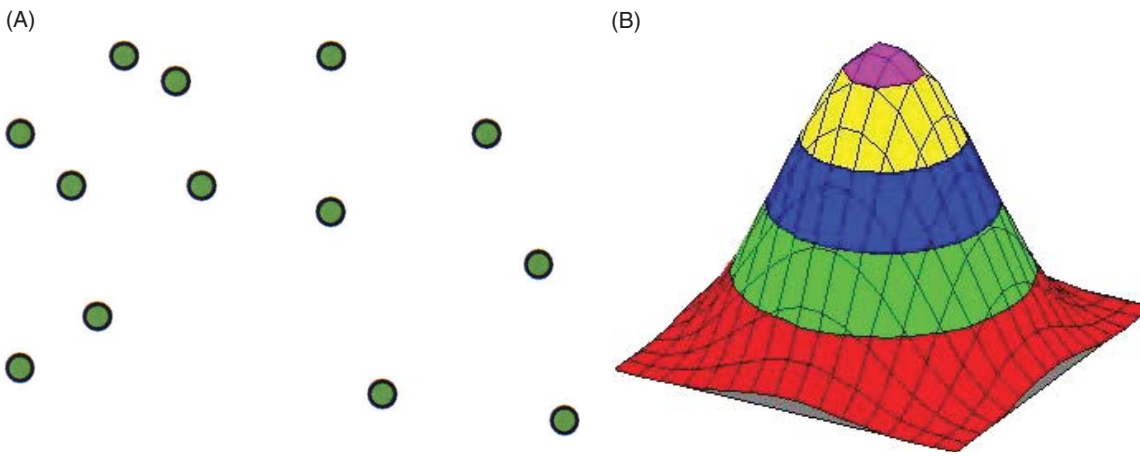


Figure 14.29 (A) A collection of point objects, and (B) a kernel function. The kernel's shape depends on a distance parameter – increasing the value of the parameter results in a broader and lower kernel and reducing it results in a narrower and sharper kernel. When each point is replaced by a kernel and the kernels are added the result is a density surface whose smoothness depends on the value of the distance parameter

objects – and no sense at all to apply density estimation to samples of a field.

Density estimation makes sense only from the discrete-object perspective, and spatial interpolation only from the field perspective.

Although density estimation could be applied to any type of discrete object, it is most often applied to the estimation of point density, and that is the focus here. The most obvious example is the estimation of population density, and that example is used in this discussion, but it could be equally well applied to the density of different kinds of diseases, or animals, or any other set of well-defined points.

Consider the mainland of Australia. One way of defining its population density is to take the entire population and divide by the total area – on this basis the 1996 population density was roughly 2.38 per sq km. But we know that Australia's settlement pattern is very non-uniform, with most of the population concentrated in five coastal cities (Brisbane, Sydney, Melbourne, Adelaide, and Perth). So if we looked at the landscape in smaller pieces, such as circles 10 km in radius, and computed population density by dividing the number of people in each circle by the circle's area, we would get very different results depending on where the circle was centered. So, in general, population density at a location, and at a spatial resolution of d, might be defined by centering a circle at the location and dividing the total population within the circle by its area. Using this definition there are an infinite number of possible population density maps of Australia, depending on the value selected for d. And it follows that there is no such thing as *population density*, only population density *at a spatial resolution of d*. Note the similarity between this idea and the previous discussion of slope – in general, many geographic themes can only be defined rigorously if spatial resolution is made explicit, and much confusion results in GIS because of our willingness to talk about themes without at the same time specifying spatial resolution.

Density estimation with a kernel allows the spatial resolution of a field of population density to be made explicit.

The theory of density estimation formalizes these ideas. Consider a collection of point objects, such as those shown in Figure 14.29. The surface shown in the figure is an example of a *kernel function*, the central idea in density estimation. Any kernel function has an associated length measure, and in the case of the function shown, which is a Gaussian distribution, the length measure is a parameter of the distribution – we can generate Gaussian distributions with any value of this parameter and they become flatter and wider as the value increases. In density estimation, each point is replaced by its kernel function and the various kernel functions are added to obtain an aggregate surface, or continuous field of density. If one thinks of each kernel as a pile of sand, then each pile has the same total weight of one unit. The total weight of all piles of sand is equal to the number of points and the total weight of sand within a given area, such as the area shown in the figure, is an estimate of the total population in that area. Mathematically, if the population density is represented by a field $\rho(x, y)$, then the total population

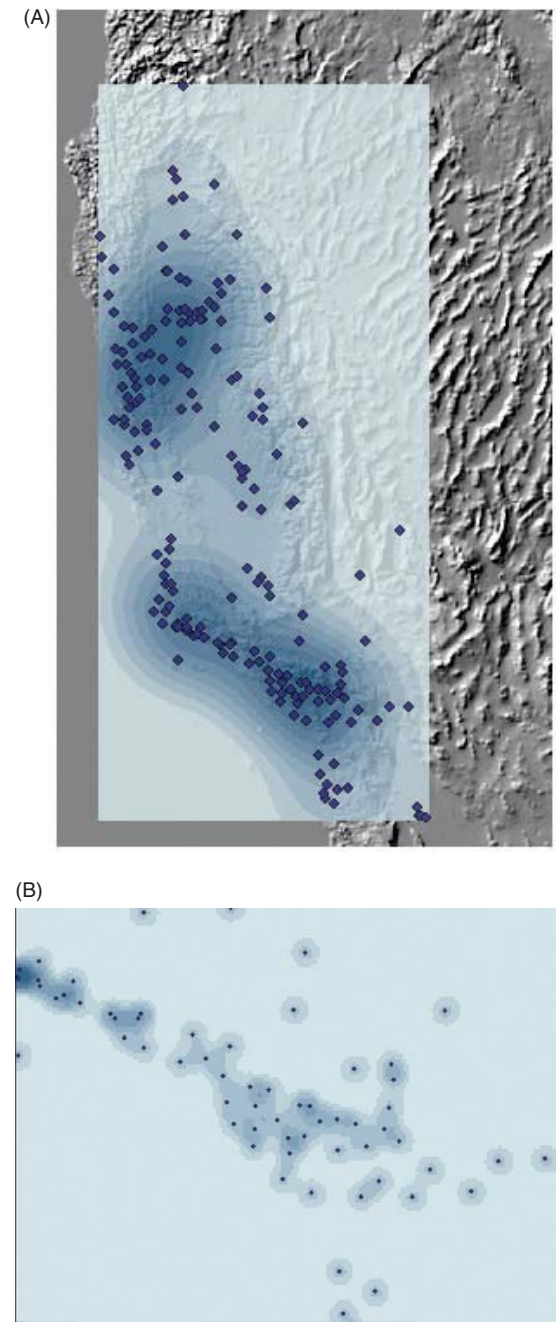


Figure 14.30 Density estimation using two different distance parameters in the respective kernel functions: (A) the surface shows the density of ozone-monitoring stations in California, using a kernel radius of 150 km; (B) zoomed to an area of Southern California, this shows that a kernel radius of 16 km is clearly too small for this dataset, as it leaves each kernel isolated from its neighbors

within area A is the integral of the field function over that area, that is:

$$P = \int_A \rho \mathrm{d}A$$

A variety of kernel functions are used in density estimation, but the form shown in Figure 14.29 is perhaps the commonest. This is the traditional bell curve or Gaussian distribution of statistics and is encountered elsewhere in this book in connection with errors in the measurement of position in two dimensions (Section 6.3.2.2). By adjusting the width of the bell it is possible to produce a range of density surfaces of different amounts of smoothness. Figure 14.30 contrasts two density estimations from the same data, one using a comparatively narrow bell to produce a complex surface, and the other using a broader bell to produce a smoother surface.

14.5 Conclusion

This chapter has discussed the first three types of spatial analysis; Chapter 15 discusses the remaining three, and Chapter 16 examines spatial modeling. Several general issues of scale and resolution, and accuracy and uncertainty, have been raised throughout the discussion: these are discussed in greater detail in Chapter 6.

Questions for further study

1. Did Dr John Snow actually make his inference strictly from looking at his map? What information can you find on the Web on this issue? (try **www.jsi.com**)
2. You are given a map showing the home locations of the customers of an insurance agent and asked to construct a map showing the agent's market area. Would spatial interpolation or density estimation be more appropriate, and why?
3. What is conditional simulation, and how does it differ from Kriging? Under what circumstances might it be useful (a possible source on this question is the chapter by Englund in Goodchild, Steyaert and Parks, *Environmental Modeling with GIS*, Oxford University Press, 1993).
4. What are the most important characteristics of the three methods of spatial interpolation discussed in this chapter? Using a test dataset of your own choosing, compute and describe the major features of the surfaces interpolated by each method.

Further reading

Bailey T.C. and Gatrell A.C. 1995 *Interactive Spatial Data Analysis*. Harlow, UK: Longman Scientific and Technical.

Burrough P.A. and McDonnell R.A. 1998 *Principles of Geographical Information Systems*. New York: Oxford University Press.

Isaaks E.H. and Srivastava R.M. 1989 *Applied Geostatistics*. New York: Oxford University Press.

O'Sullivan D. and Unwin D.J. 2003 *Geographic Information Analysis*. Hoboken, NJ: Wiley.

Silverman B.W. 1986 *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

15 *Descriptive summary, design, and inference*

This is the second of two chapters on spatial analysis and focuses on three areas: descriptive summaries, the use of analysis in design decisions, and statistical inference. These methods are all conceptually more complex than those in Chapter 14, but are equally important in practical applications of GIS.

The chapter begins with a brief discussion of data mining. Descriptive summaries attempt to capture the nature of geographic distributions, patterns, and phenomena in simple statistics that can be compared through time, across themes, and between geographic areas. Optimization techniques apply much of the same thinking, and extend it, to help users who must select the best locations for services, or find the best routes for vehicles, or a host of similar tasks. Hypothesis testing addresses the basic scientific need to be able to generalize results from a small study to a much larger context, perhaps the entire world.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will know about:

- Data mining, a new form of analysis that is enabled by GIS and by vast new supplies of data;
- The concept of summarizing a pattern in a few simple statistics;
- Methods that support decisions by enlisting GIS to search automatically across thousands or millions of options;
- The concept of a hypothesis, and how to make inferences from small samples to larger populations.

15.1 More spatial analysis

The methods introduced in this chapter use more sophisticated conceptual frameworks than those of Chapter 14. That chapter began with simple queries and measurement, but in this chapter the discussion turns to summaries, optimization methods, and inferential methods. Many date from well before the era of GIS, and the advent of geographic databases and cheap computing technology has turned what were once esoteric methods buried in the academic literature into practical ways of solving everyday problems.

The advent of large databases and fast computing has also led to new ways of thinking about spatial analysis and many of these are captured in the term *data mining*. Surely, the thinking goes, there must be interesting patterns, clusters, anomalies, and truths buried in the masses of data being collected and archived every day by humanity's vast new investment in information technology. Cameras now capture images of the traffic on freeways that are precise enough to read license plates. As noted in Section 11.3, every use of a credit card creates a record in a database that includes the location of use, along with information on what was bought, and having the name of the user allows that information to be linked to many other records. If we only had suitable computer methods, the thinking goes, we could program software to scan through these vast databases, looking for things of interest. Many of these might be regarded as invasions of privacy, but others could be very useful in spotting criminal activity, for example. Data mining has become a major area of application in business, where it is used to detect anomalies in spending patterns that might be indicative of a stolen credit card – for example, a card owned by a resident of California, USA is suddenly used for a series of large purchases in a short period of time in the New York area, in types of stores not normally frequented by the card owner (Figure 15.1). The *geographic* aspect of this evidence – the expenditures in a small, unusual area – is one of the most significant.

Data mining is used to detect anomalies and patterns in vast archives of digital data.

Data mining techniques could be used to watch for anomalous patterns in the records of disease diagnosis, or to provide advanced warnings of new outbreaks of virulent forms of influenza. In all of these examples the objective is to find *patterns*, and *anomalies* that stand out from the normal in an area, or stand out at a particular point in time when compared to long-term averages. In recent years surprising patterns in digital data have led to the discovery of the ozone hole over the Antarctic, the recall of millions of tires made by the Firestone-Bridgestone Company, and numerous outbreaks of disease. Figure 15.2 shows an example of a *parallel coordinate plot*, one of many methods developed in recent years to display large multi-dimensional datasets. Note that location plays no part in this plot, which therefore

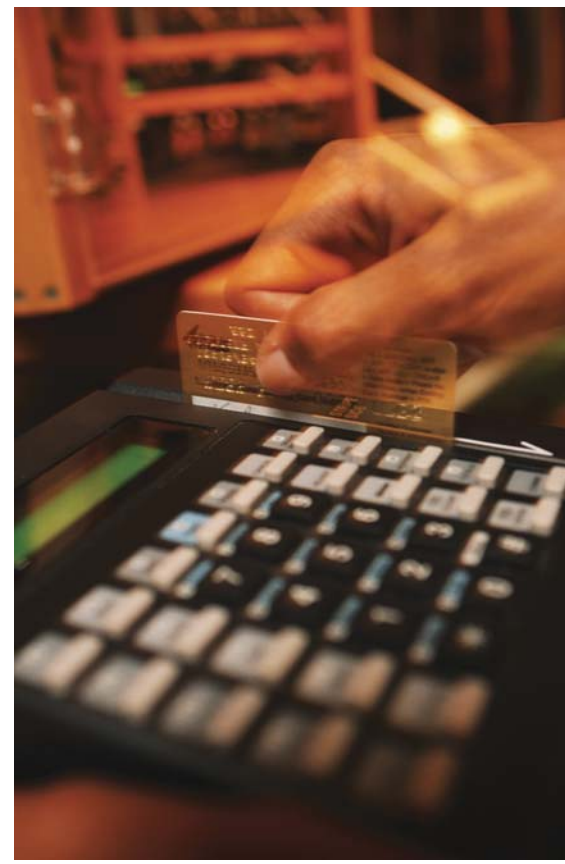


Figure 15.1 Massive geographic databases can be built from records on the use of credit cards

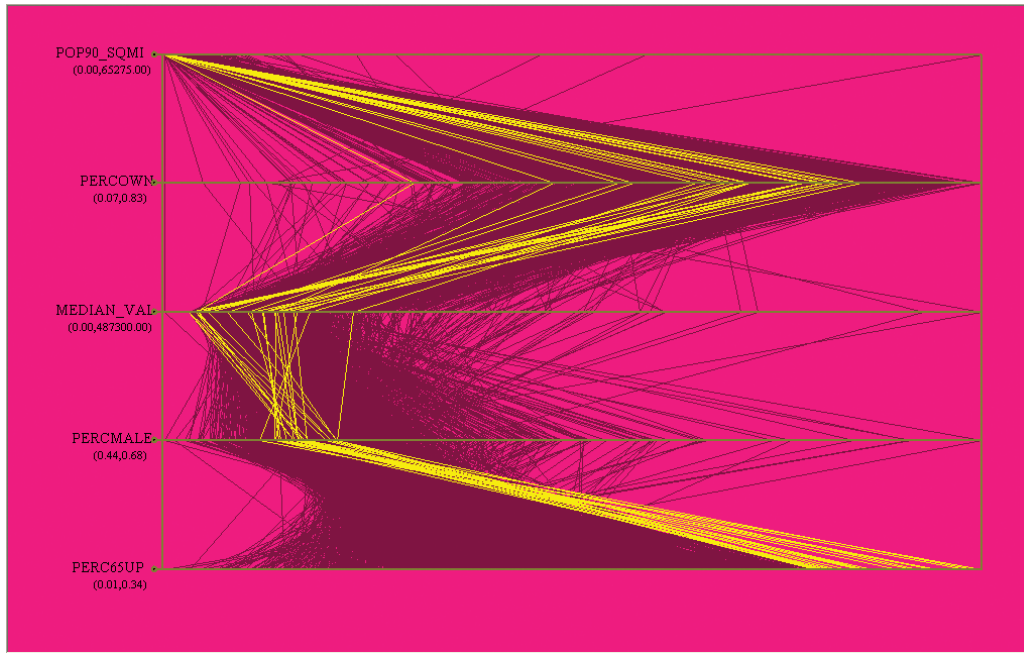


Figure 15.2 The search for patterns in massive datasets is known as data mining. This example of a *parallel coordinate plot* from the GeoDa software (see Figure 14.8) shows data on all 3141 US counties. Each horizontal line corresponds to one variable. Each county is represented by a polyline connecting its scores on each of the variables. In this illustration the user has selected all counties with a very high proportion of people over 65 (the bottom line), and the plot shows that such counties (the yellow polylines) tend also to have low population densities, average rates of owner occupancy, low median values of housing, and average percentages of males

fails to qualify as a form of spatial analysis according to the definitions given in Section 14.1.

15.2 Descriptive summaries

15.2.1 Centers

The topics of generalization and abstraction were discussed earlier in Section 3.8 as ways of reducing the complexity of data. This section reviews a related topic, that of numerical summaries. If we want to describe the nature of summer weather in an area we cite the *average* or *mean*, knowing that there is substantial variation around this value, but that it nevertheless gives a reasonable *expectation* about what the weather will be like on any given day. The mean (the more formal term) is one of a number of measures of *central tendency*, all of which attempt to create a summary description of a series of numbers in the form of a single number. Another is the *median*, the value such that one half of the numbers are larger, and one half are smaller. Although the mean can be computed only for numbers measured on interval or ratio scales, the median can be computed for ordinal data. For nominal data the appropriate measure of central tendency is the *mode*, or the commonest value. For definitions of nominal, ordinal, interval, and ratio see Box 3.3. Special methods must be used to measure central tendency for cyclic data – they are discussed in texts on directional data, for example by Mardia and Jupp.

The spatial equivalent of the mean would be some kind of center, calculated to summarize the positions of a number of points. Early in US history, the Bureau of the Census adopted a practice of calculating a center for the US population. As agricultural settlement advanced across the West in the 19th century, the repositioning of the center every ten years captured the popular imagination. Today, the movement west has slowed and shifted more to the south (Figure 15.3), and by the next census may even have reversed.

Centers are the two-dimensional equivalent of the mean.

The mean of a set of numbers has several properties. First, it is calculated by summing the numbers and dividing by the number of numbers. Second, if we take any value d and sum the squares of the differences between the numbers and d, then when d is set equal to the mean this sum is minimized (Figure 15.4). Third, the mean is the point about which the set of numbers would balance if we made a physical model such as the one shown in Figure 15.5, and suspended it.

These properties extend easily into two dimensions. Figure 15.6 shows a set of points on a flat plane, each

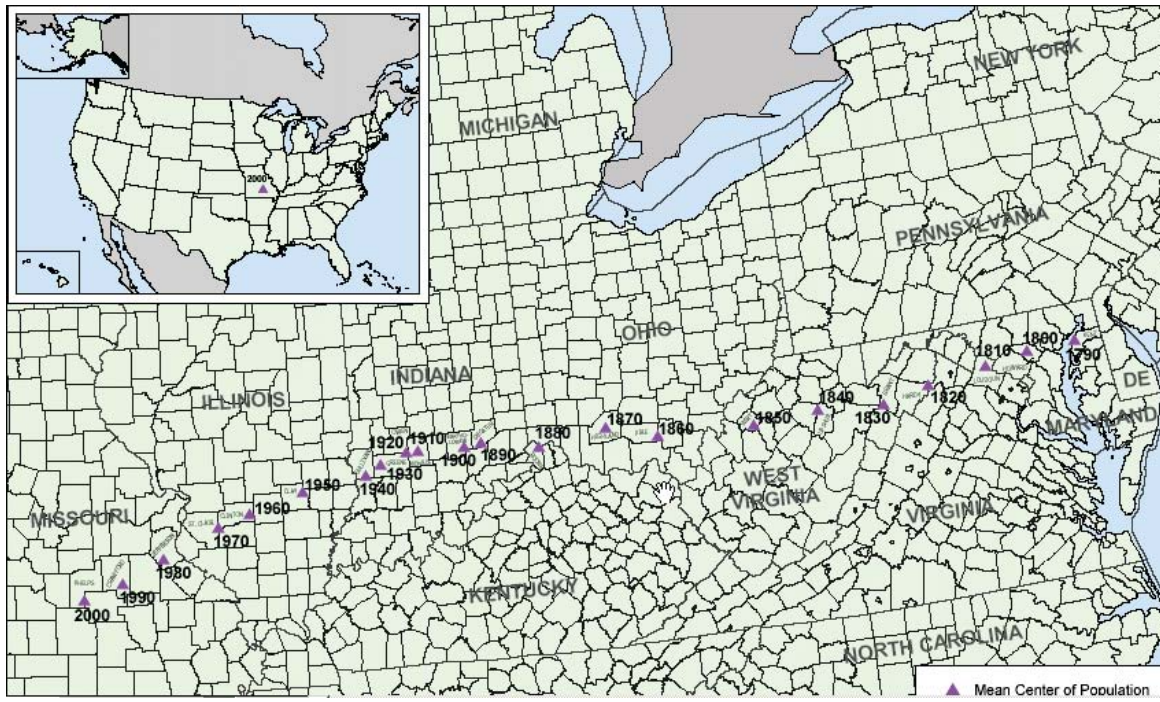


Figure 15.3 The march of the US population westward since the first census in 1790, as summarized in the population centroid (*Source*: US Bureau of the Census)

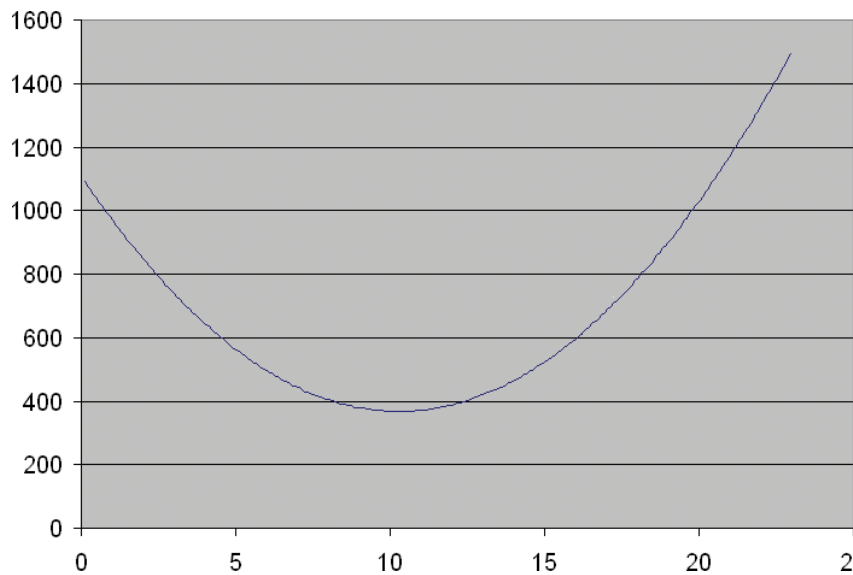


Figure 15.4 Seven points are distributed along a line at coordinates 1, 3, 5, 11, 12, 18, and 22. The curve shows the sum of distances squared from these points and how it is minimized at the mean $[(1+3+5+11+12+18+22)/7 = 10.3]$

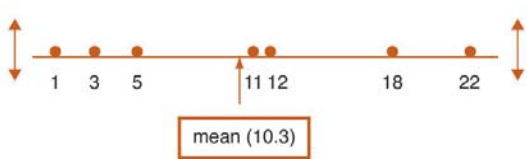


Figure 15.5 The mean is also the balance point, the point about which the distribution would balance if it were modeled as a set of equal weights on a weightless, rigid rod

one located at a point (x_i, y_i) and with weight w_i. The *centroid* or *mean center* is found by taking the weighted average of the x and y coordinates:

$$\overline{x} = \sum_i w_i x_i \Big/ \sum_i w_i$$

$$\overline{y} = \sum_i w_i y_i \Big/ \sum_i w_i$$

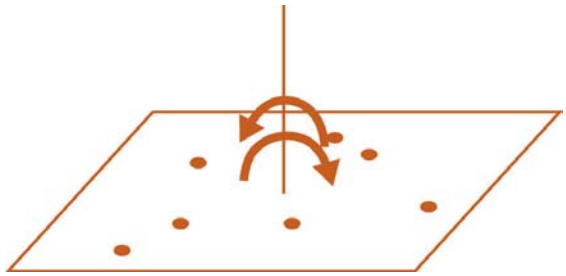


Figure 15.6 The centroid or mean center replicates the balance-point property in two dimensions – the point about which the two-dimensional pattern would balance if it were transferred to a weightless, rigid plane and suspended

It is also the point that minimizes the sum of squared distances, and it is the balance point.

Just like the mean, the centroid is a useful summary of a distribution of points. Although any single centroid may not be very interesting, a comparison of centroids for different sets of points or for different times can provide useful insights.

The centroid is the most convenient way of summarizing the locations of a set of points.

The property of minimizing functions of distance (the square of distance in the case of the centroid) makes centers useful for different reasons. Of particular interest is the location that minimizes the sum of distances, rather than the sum of squared distances, since this could be the most effective location for any service that is intended to serve a dispersed population. The point that minimizes total straight-line distance is known as the *point of minimum aggregate travel* or MAT. Historically there have been numerous instances of confusion between the MAT and the centroid, and the US Bureau of the Census used to claim that the Center of Population had the MAT property, even though their calculations were based on the equations for the centroid.

The centroid is often confused with the point of minimum aggregate travel.

There is no simple mathematical expression for the MAT, and instead its location must be found by a trial-and-error process of *iteration*, in which an initial guess is successively improved using a suitable algorithm. There is also an interesting way of computing the MAT using a physical analog, an experiment known as the *Varignon Frame* (see Box 15.1). There is an intriguing paradox associated with the experiment. Imagine that one of the points must be relocated further away from the MAT but in the same direction. The force pulling on the string remains the same, in the same direction, so the position of the knot does not move. But how could the solution to a problem of minimizing distance not be affected by a change in distance? It is straightforward

Technical Box **15.1**

The Varignon Frame experiment

This experiment provides a way of finding the point of minimum aggregate travel (MAT). Because it uses a physical model to solve the problem rather than a digital representation we can think of it as an *analog* computer (see Section 16.1 for a more extensive discussion of analog models, and see Section 3.7 for a discussion of the paper map as an analog representation). Suppose we want to find the best location for a steel mill, to minimize the total costs of shipping raw materials to the plant (iron ore, limestone, and coal) and the total costs of shipping the product to markets. Each source of raw materials or market is represented by a point, with a weight equal to the cost per km of shipping the required amount of material. Then the total shipping cost is the product of these weights times distances to the plant, summed over all sources and markets.

Take a piece of plywood, and sketch a map on it, locating the sources and markets (see Figure 15.7). At each of these points drill a hole, and hang a weight on a string through the hole, using the appropriate weight (proportional to the shipping cost of the required amount of material per km). Bring all of the strings together into a knot. When the knot is released to find its own position the result is the MAT, though friction in the holes will create some error in the solution.

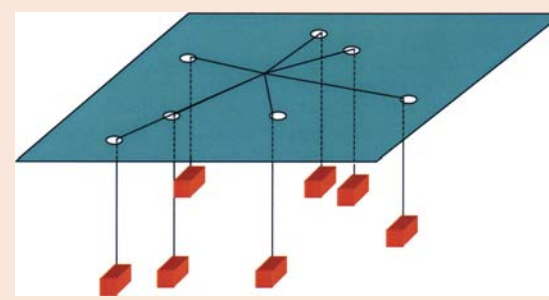


Figure 15.7 The Varignon Frame experiment, an analog computer for the point of minimum aggregate travel (MAT)

to show with a little application of calculus that the experiment is indeed correct, despite what appears to be a counter-intuitive result – that the location of the MAT does indeed depend only on the weights and the directions of pull, not on the distances. The same property occurs in the one-dimensional case: the point that minimizes distance along a line is the median, and the median's location is not affected by moving points as long as no point is relocated to the other side of the current median.

This example serves to make a very general point about spatial analysis and its role in human–computer interaction. In the introduction to Chapter 14 the point was made that all of these methods of analysis work best in the context of a collaboration between human and machine, and that one benefit of the machine is that it sometimes serves to correct the misleading aspects of human intuition. The MAT and more complex problems that are based on it illustrate this principle very well. Intuition would have suggested that moving one of the weights in the Varignon Frame experiment would have moved the knot, even though the movement was in the same direction. But theory says otherwise. Humans can be very poor at guessing the answers to optimization problems in space and the results of machine analysis can often be sharply different from what we expect.

The MAT is one instance of a type of spatial analysis in which the objective is not insight so much as *design* – it is a *normative* or *prescriptive* method because it imposes a norm (minimizing distance) or prescribes a solution. In the 1920s, the Soviet Union supported the Mendeleev Centrographical Laboratory whose objective was to identify the centers of various kinds of economic activity in the USSR as the most suitable sites for the state's central planning functions. For example, the center of piano manufacture would be the best place to locate the state planning office for piano production.

Many methods of spatial analysis are used to make design decisions.

In Section 14.3.1 it was noted that simple metrics of distance, such as the length of a straight line, have the disadvantage that they ignore the effects of networks on travel. The approach to the MAT described here also assumes straight-line travel. Comparable approaches based on networks are often better suited to practical problem solving and some examples of these are discussed in Section 15.3.

All of the methods described in this and the following section are based on plane geometry, and simple x,y coordinate systems. If the curvature of the Earth's surface is taken into account, the centroid of a set of points must be calculated in three dimensions, and will always lie under the surface. More useful perhaps are versions of the MAT and centroid that minimize distance over the curved surface (the minimum total distance and the minimum total of squared distances, respectively, using great circle distances, Sections 5.6 and 14.3.1).

15.2.2 Dispersion

Central tendency is the obvious choice if a set of numbers must be summarized in a single value, but what if there is the opportunity for a second summary value? Here, the measure of choice for numbers with interval or ratio properties is the *standard deviation*, or the square root of the mean squared difference from the mean:

$$s = \sqrt{\sum_i (x_i - \overline{x})^2 \Big/ n}$$

where n is the number of numbers, s is the standard deviation, x_i refers to the ith observation, and $\overline{x}$ is the mean of the observations. In weighted form the equation becomes:

$$s = \sqrt{\sum_i w_i (x_i - \overline{x})^2 \Big/ \sum_i w_i}$$

where w_i is the weight given to the ith observation. The *variance*, or the square of the standard deviation (the mean squared difference from the mean), is often encountered but it is not as convenient a measure for descriptive purposes. Standard deviation and variance are considered more appropriate measures of dispersion than the *range* (difference between the highest and lowest numbers) because as averages they are less sensitive to the specific values of the extremes.

The standard deviation has also been encountered in Section 6.3.2.2 in a different guise, as the root mean squared error (RMSE), a measure of dispersion of observations about a true value. Just as in that instance, the Gaussian distribution provides a basis for generalizing about the contents of a sample of numbers, using the mean and standard deviation as the parameters of a simple bell curve. If data follow a Gaussian distribution, then approximately 68% of values lie within one standard deviation of the mean; and approximately 5% of values lie outside two standard deviations.

These ideas convert very easily to the two-dimensional case. A simple measure of dispersion in two dimensions is the *mean distance from the centroid*. In some applications it may be desirable to give greater weight to more distant points. For example, if a school is being located, then students living at distant locations are comparatively disadvantaged. They can be given greater weight if each distance is squared, such that a student twice as far away receives four times the weight. This property is minimized by locating the school at the centroid.

Mean distance from the centroid is a useful summary of dispersion.

Measures of dispersion can be found in many areas of GIS. The breadth of the kernel function of density estimation (Section 14.4.5) is a measure of how broadly the pile of sand associated with each point is dispersed.

RMSE is a measure of the dispersion inherent in positional errors (Section 6.3.2.2).

15.2.3 Measures of pattern: unlabeled points

A very large number of techniques exist for summarizing the general properties of patterns. Most of these have been developed for points, and their extension to patterns of lines and areas is generally not straightforward, so the focus here will be on points only. The discussion in this section centers on points without distinguishing attributes – such as patterns of incidence of a disease, or sightings of a bird species, or customers of a retail store. The subsequent section looks at methods for describing how attributes are distributed among points, rather than at the locations of the points *per se*.

One of the questions most commonly asked about distributions of points is whether they display a random pattern, in the sense that all locations are equally likely to contain a point, or whether some locations are more likely than others – and particularly, whether the presence of one point makes other points either more or less likely in its immediate neighborhood. This leads to three possibilities:

- the pattern is *random* (points are located independently, and all locations are equally likely);
- the pattern is *clustered* (some locations are more likely than others, and the presence of one point may attract others to its vicinity);
- the pattern is *dispersed* (the presence of one point may make others less likely in its vicinity).

Establishing the existence of clusters is often of great interest, since it may point to possible causal factors, as for example with the case of childhood leukemia studied by Stan Openshaw (Figure 14.4). Dispersed patterns are the typical result of competition for space, as each point establishes its own territory and excludes others – so such patterns are commonly found among organisms that exhibit territorial behavior, as well as among market towns in rural areas, and among retail outlets.

Point patterns can be identified as clustered, dispersed, or random.

It is helpful to distinguish two kinds of processes responsible for point patterns. *First-order* processes involve points being located independently, but may still result in clusters because of varying point density. For example, the drinking-water hypothesis investigated by Dr John Snow and described in Box 14.1 led to a higher density of points around the pump, because of greater access; and similarly the density of organisms of a particular species may vary over an area because of varying suitability of habitat. *Second-order* processes involve interaction between points, and lead to clusters when the interactions are attractive in nature, and dispersion when they are competitive or repulsive. In the cholera case, the contagion hypothesis rejected by Snow is a second-order process and results in clustering even in situations when all other density-controlling factors are perfectly uniform. Unfortunately, as argued in Section 14.1, Snow's evidence did not allow him to resolve with complete confidence between first-order and second-order processes, and in general it is not possible to determine whether a given clustered point pattern was created by varying density factors, or by interactions. On the other hand, dispersed patterns can only be created by second-order processes.

Clustering can be produced by two distinct mechanisms, identified as first-order and second-order.

There are many available descriptive statistics of pattern and some excellent books on the subject. To illustrate the method only one will be discussed in this section. This is the K function, and unlike many such statistics it provides an analysis of clustering and dispersion over a range of scales. In the interests of brevity the technical details will be omitted, but they can be found in the texts listed at the end of this chapter. They include procedures for dealing with the effects of the study area boundary, which is an important distorting factor for many pattern statistics.

$K(d)$ is defined as the expected number of points within a distance d of an arbitrarily chosen point, divided by the density of points per unit area. When the pattern is random this number is πd^2, so the normal practice is to plot the function:

$$\hat{L}(d) = \sqrt{K(d)/\pi}$$

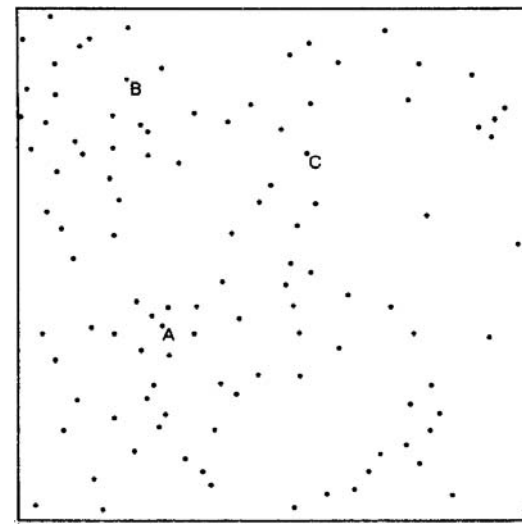


Figure 15.8 Point pattern of individual tree locations. A, B, and C identify the individual trees analyzed in Figure 15.9 (*Source*: Getis A. and Franklin J. 1987 'Second-order neighborhood analysis of mapped point patterns.' *Ecology* **68**(3): 473–477)

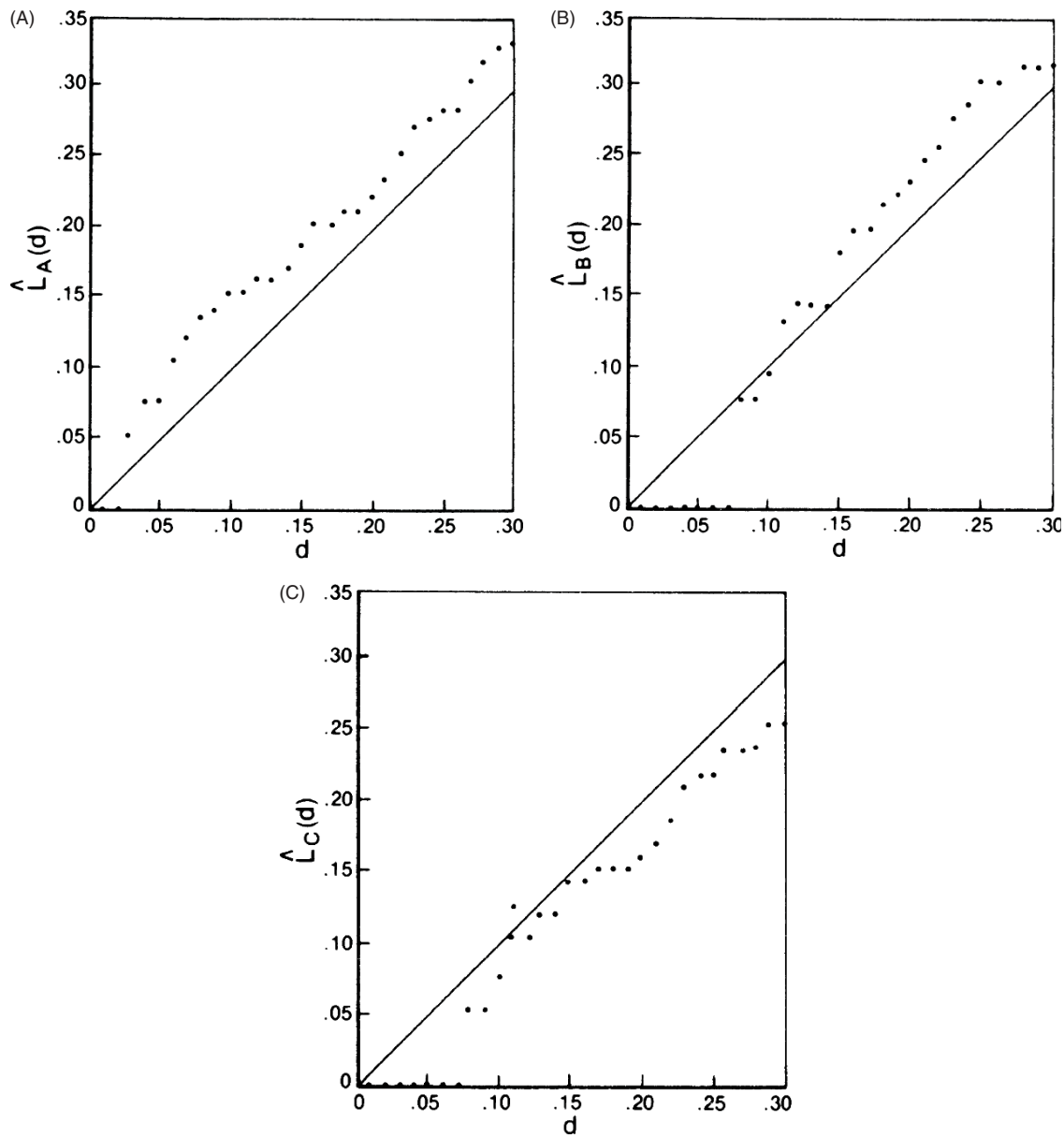


Figure 15.9 Analysis of the local distribution of trees around three reference trees in Figure 15.8 (see text for discussion) (*Source*: Getis A. and Franklin J. 1987 'Second-order neighborhood analysis of mapped point patterns.' *Ecology* **68**(3): 473–477)

since $\hat{L}(d)$ will equal d for all d in a random pattern, and a plot of $\hat{L}(d)$ against d will be a straight line with a slope of 1. Clustering at certain distances is indicated by departures of $\hat{L}$ above the line and dispersion by departures below the line.

Figures 15.8 and 15.9 show a simple example, used to discover how trees are spaced relative to each other in a forest. The locations of the trees are shown in Figure 15.8. In Figure 15.9(A) locations are analyzed in relation to Tree A. At very short distances there are fewer trees than would occur in a random pattern, but for most of the range up to a distance of 30% of the width (and height) of the study area there are *more* trees than would be expected, showing a degree of clustering. Tree B (Figure 15.9(B)) has no nearby neighbors, but shows a degree of clustering at longer distances. Tree C (Figure 15.9(C)) shows fewer trees than expected in a random pattern over most distances, and it is evident in Figure 15.8 that C is in a comparatively widely spaced area of forest.

15.2.4 Measures of pattern: labeled features

When features are differentiated, such as by values of some attribute, the typical question concerns whether the values are randomly distributed over the features, or

whether high extreme values tend to cluster: high values surrounded by high values and low values surrounded by low values. In such investigations the processes that determined the locations of features, and were the major concern in the previous section, tend to be ignored and may have nothing to do with the processes that created the pattern of labels. For example, the concern might be with some attribute of counties – their average house value, or percent married – and hypotheses about the processes that lead to counties having different values on these indicators. The processes that led to the locations and shapes of counties, which were political processes operating perhaps 100 years ago, would be of no interest.

The Moran statistic (Section 4.3) is designed precisely for this purpose, to indicate general properties of the pattern of attributes. It distinguishes between positively autocorrelated patterns, in which high values tend to be surrounded by high values, and low values by low values; random patterns, in which neighboring values are independent of each other; and dispersed patterns, in which high values tend to be surrounded by low, and vice versa.

The Moran statistic looks for patterns among the attributes assigned to features.

In recent years there has been much interest in going beyond these global measures of spatial dependence to identify dependences locally. Is it possible, for example, to identify *hot spots*, areas where high values are surrounded by high values, and *cold spots*, where low values are surrounded by low? Is it possible to identify anomalies, where high values are surrounded by low or vice versa? Local versions of the Moran statistic are among this group and along with several others now form a useful resource that is easily implemented in GIS.

Figure 15.10 shows an example, using the Local Moran statistic to differentiate states according to their roles in the pattern of one variable, the median value of housing. The weights (see Box 4.4) have been defined by adjacency, such that pairs of states that share a common

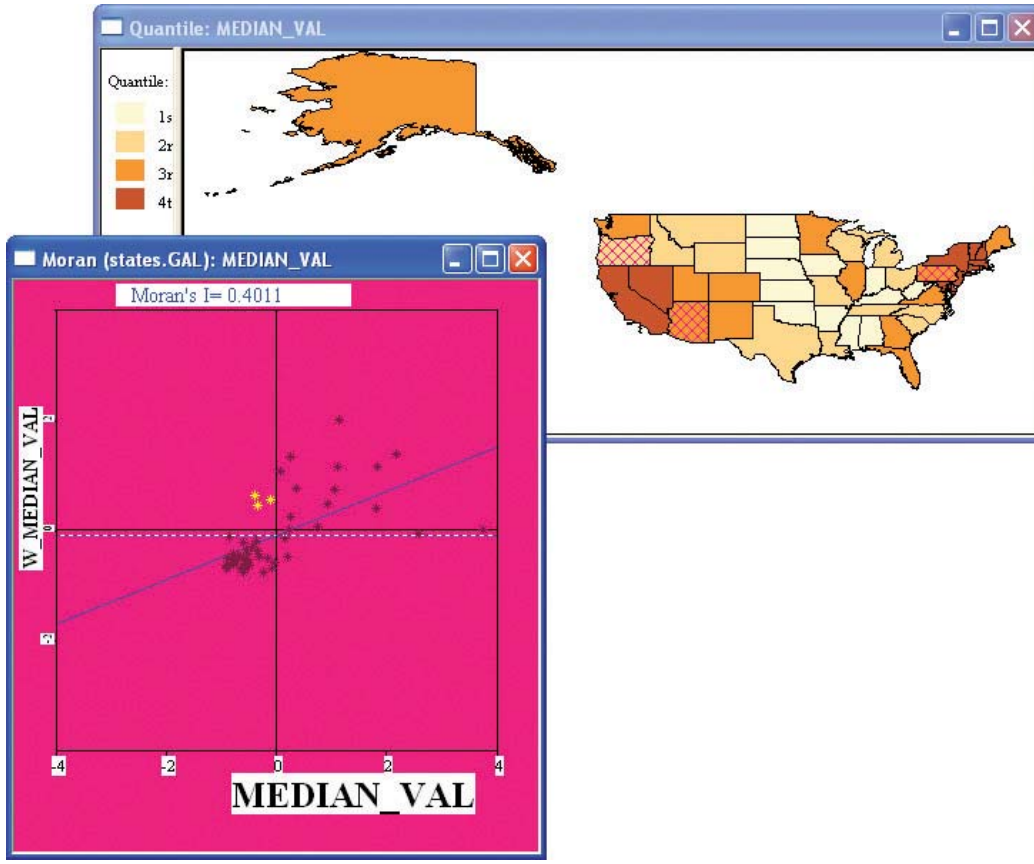


Figure 15.10 The Local Moran statistic, applied to describe local aspects of the pattern of housing value among US states. In the map window the states are colored according to median value, with the darker shades corresponding to more expensive housing. In the scatterplot window the median value appears on the x axis, while on the y axis is the weighted average of neighboring states. The three points colored yellow are instances where a state of below-average housing value is surrounded by states of above-average value. The windows are linked (see Figure 14.8 for details of this GeoDa software), and the three points are identified as Oregon, Arizona, and Pennsylvania. The global Moran statistic is also shown (+0.4011, indicating a general tendency for clustering of similar values)

Applications Box **15.2**

Comparing attributes when objects do not coincide spatially

GIS users often encounter situations where attributes must be compared, but for different sets of objects. Figure 15.11 shows a study area, with two sets of points, where one indicates the locations where levels of ambient sound were measured using recorders mounted on telephone poles and the other shows the locations of interviews conducted with local residents to determine attitudes to noise. We would like to know about the relationship between sound and attitudes, but the locations and numbers of cases (the *spatial support*) are different. A simple expedient is to use spatial interpolation (Section 14.4.4) to map both variables as complete fields. Then it would be possible to conduct the analysis in any of three ways:

- by interpolating the second dataset to the locations at which the first attribute was measured;
- by the reverse – interpolating the first dataset to the locations at which the second attribute was measured;
- by interpolating both datasets to a common geometric base, such as a raster.

In the third case, note that it is possible to create a vast amount of data by using a sufficiently detailed raster. In essence, all of these options involve manufacturing information and the results will depend to some extent on the nature and suitability of the method used to do the spatial interpolation. We normally think of a relationship that is demonstrated with a large number of observations as stronger and more convincing than one based on a small number of observations. But the ability to manufacture data upsets this standard view.

Figure 15.11 shows a possible solution using this third option. The number of grid points has been determined by the smaller number of cases – in this case, approximately 10 (12 are shown) – to minimize concerns about manufacturing information.

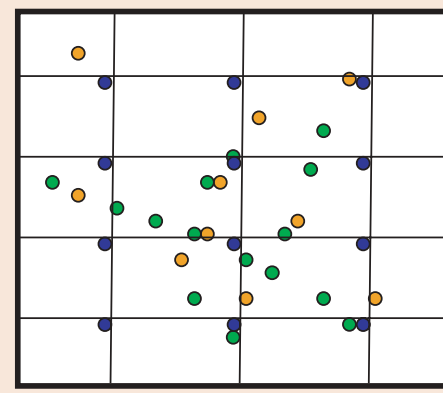


Figure 15.11 Coping with the comparison of two sets of attributes when the respective objects do not coincide. In this instance, attitudes regarding ambient noise have been obtained through a household survey of 15 residents (green dots) and are to be compared to ambient noise levels measured at 10 observation points (brown dots). The solution is to interpolate both sets of data to 12 comparison points (blue dots), using the methods discussed in Section 14.4.4

boundary are given a weight of 1 and all other pairs a weight of zero.

15.2.5 Fragmentation and fractional dimension

Figure 15.13 shows a type of dataset that occurs frequently in GIS. It might represent the classes of land cover in an area, but in fact it is a map of soils, and each patch corresponds to an area of uniform class and is bounded by areas of different class. A landscape ecologist or forester might be interested in the degree to which the landscape is fragmented – is it broken up into small or large patches, are the patches compact or contorted, and how long are the common boundaries between the patches? Fragmented landscapes are less favorable for wildlife because habitat is broken into small patches with potentially dangerous land between. Fragmentation statistics have been designed to provide the numerical basis for these kinds of questions, by measuring relevant properties of datasets.

An interesting application of this kind of technique is to deforestation, in areas like the Amazon Basin. Figure 15.14 shows a sequence of images of part of the basin from space, and the impact of clearing and settlement on the pattern of land use. The impact of land use changes like these depends on the degree to which they fragment existing habitat, perhaps into pieces that are too small to maintain certain species such as major predators.

Biographical Box **15.3**

Luc Anselin, pioneer in spatial analysis

Figure 15.12 Luc Anselin

Luc Anselin (Figure 15.12) is Professor at the University of Illinois, Urbana-Champaign, where he directs a spatial analysis research laboratory and holds appointments in the Departments of Agricultural and Consumer Economics, Economics, Geography, and Urban and Regional Planning. He was originally trained as an economist at the Free University of Brussels (VUB) in Belgium and obtained a Ph.D. in Regional Science from Cornell University. Most of his career has been devoted to the development, implementation, and application of statistical and econometric methods that explicitly take into account spatial effects (spatial dependence and spatial heterogeneity) that may be present in the data. His much-cited *Spatial Econometrics* (1988) was the first comprehensive treatment that outlined the methodology to test for spatial effects in regression models and how to estimate such models. Many of these methods were incorporated into his SpaceStat software package in the early 1990s. Anselin gained an interest in local statistics, particularly how they could be incorporated into software tools for exploratory spatial data analysis (ESDA). He developed the concepts of a Moran scatterplot (Figure 15.10) and LISA test (local indicators of spatial association; Figure 14.8) and implemented them in several software tools. The latest incarnation of these tools is the popular GeoDa package (Figure 14.8). 'I first realized the power of GIS and spatial analysis as a faculty member in the Department of Geography at the University of California, Santa Barbara. I was trained as an econometrician, and even though my research dealt with spatial aspects, I had not fully realized the potential of the interface between GIS and the analytical techniques I had developed. For example, I had carried out all the illustrative regressions and tests for the *Spatial Econometrics* book without ever looking at a map. When GIS tools became available, it was just mind-boggling how much more efficiently analysis could be carried out. Patterns could be readily distinguished that would take hours and hours of tabulations and regression runs in the old days. Especially now, with the geovisualization tools of desktop GIS and ESDA software such as GeoDa, there is a potential for tremendous efficiency gains during the inductive stage of an empirical analysis, in particular when theory is of little guidance or when the object of study is at the interface of multiple disciplines. It is also nothing short of revolutionary to see how in recent years, spatial analysis and the use of GIS methods have become increasingly accepted in mainstream social science.'

An obvious measure of fragmentation is simply the number of patches. If the representation is vector this will be the number of polygons, whereas if it is raster it will be computed by comparing neighboring cells to see if they have the same class. But as argued in Section 6.2.1, landscapes rarely possess well-defined natural units, so it is important that methods for measuring fragmentation take into account the fundamental uncertainty associated with concepts such as *patch*. An area can be subdivided into a given number of patches in a vast number of distinct ways and so measures in addition to the simple count of patches are needed. One simple measure is the average shape of patches, calculated using the shape measures discussed in Section 14.3.2. A somewhat more informative indicator would be a measure of average shape for each class, since we might expect shape to vary by class – the urban class might be more fragmented than the forest class.

The size distribution of patches might also be a useful indicator. Average patch size is uninformative, since it is simply the total area divided by the number of patches, and is independent of their shapes, or of the relative abundances of large and small patches. But a histogram of patch size would show the relative abundances and can be a useful basis for inferences.

The size distribution of patches is a useful indicator of fragmentation.

Summary statistics of the common boundaries of patches are also useful. The concept of fractals was introduced in Section 4.8 as a way of summarizing the relationship between apparent length and level of geographic detail, in the form of the fractional dimension D. Lines that are smooth tend to have fractional dimensions close to one, and more contorted lines have higher values. So fractional dimension has become accepted as

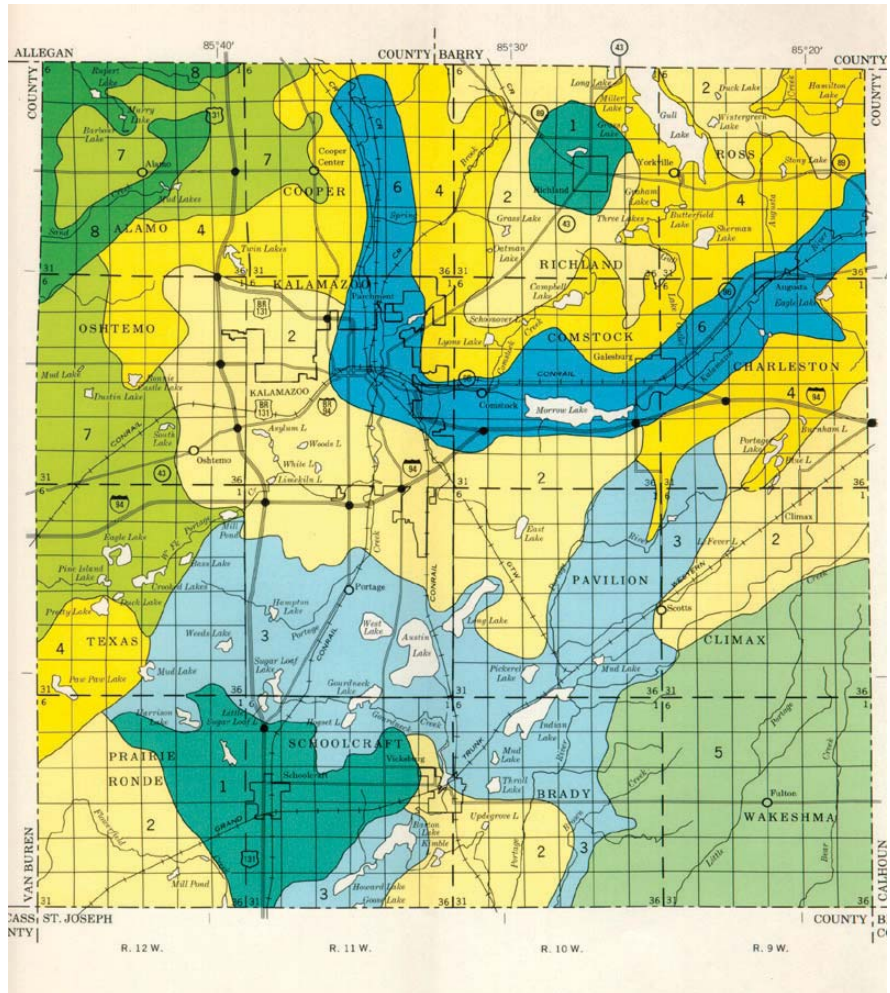


Figure 15.13 A soil map of Kalamazoo County, Michigan, USA. Soil maps such as this one show areas of uniform soil class separated by sharp boundaries (Courtesy: US Department of Agriculture)

a useful descriptive summary of the complexity of geographic objects, and tools for its evaluation have been built into many widely adopted packages. FRAGSTATS is an example – it was developed by the US Department of Agriculture for the purposes of measuring fragmentation using many alternative methods.

15.3 Optimization

The concept of using spatial analysis for design was introduced earlier in Section 15.2.1. The basic idea is to analyze patterns not for the purpose of discovering anomalies or testing hypotheses about process, as in previous sections, but with the objective of creating improved designs. These objectives might include the minimization of travel distance, or the costs of construction of some new development, or the maximization of someone's profit. The three principles of retailing are often said to be *location, location*, and *location*, and over the years many GIS applications have been directed at problems that involve in one way or another the search for optimum designs. Several of the methods described in Section 15.2.1, including centers, were shown to have useful design-oriented properties. This section includes a discussion of a wider selection of these so-called *normative* methods, or methods developed for application to the solution of practical problems of design.

Normative methods apply well-defined objectives to the design of systems.

Design methods are often implemented as components of systems built to support decision making – so-called *spatial-decision support systems*, or SDSS. Complex decisions are often contentious, with many *stakeholders* interested in the outcome and arguing for one position or another. SDSS are specially adapted GIS that can be used during the decision-making process, to provide instant feedback on the implications of various proposals and the evaluation of 'what-if' scenarios. SDSS typically have

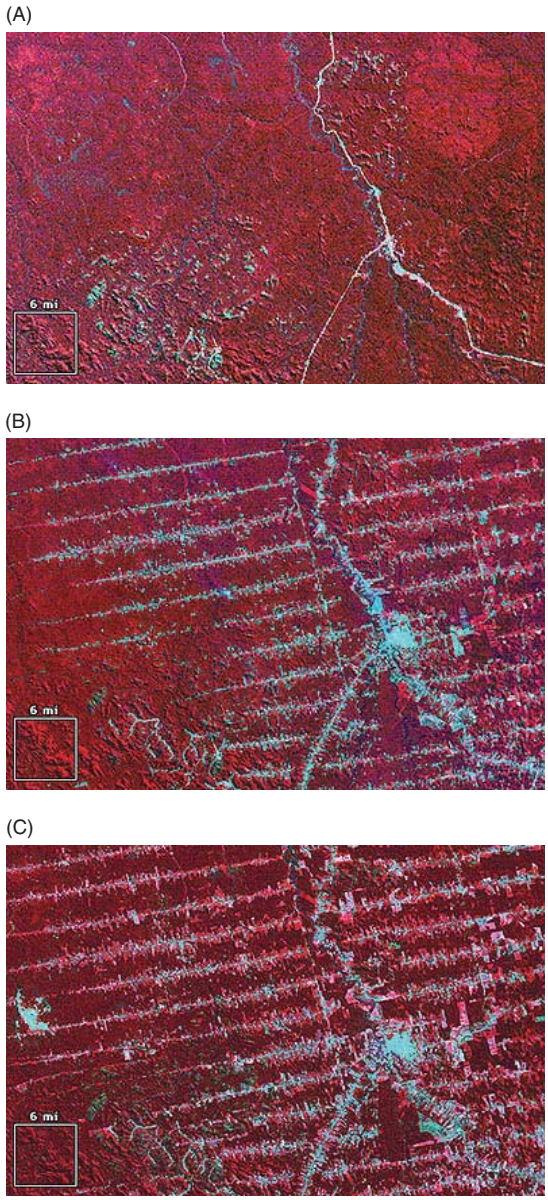


Figure 15.14 Three images of part of the state of Rondonia in Brazil, for 1975, 1986, and 1992, showing the expansion of settlement. Fragmentation statistics can be used to summarize the shapes of areas, their relative sizes, and relative positions (Courtesy: US Geological Survey)

special user interfaces that present only those functions relevant to the application.

The methods discussed in this section fall into several categories. The next section discusses methods for the optimum location of points and extends the method introduced earlier for the MAT. The second section discusses routing on a network and its manifestation in the *traveling-salesman problem* (TSP). The final section examines the selection of optimum paths across continuous space, for locating such facilities as power lines and highways, and for routing military vehicles such as tanks. The methods also divide between those that are designed to locate points and routes on networks and those designed to locate points and routes in continuous space without respect to the existence of roads or other transportation links.

15.3.1 Point location

The MAT problem is an instance of location in continuous space and finds the location that minimizes total distance with respect to a number of points. The analogous problem on a network would involve finding that location on the network that minimizes total distance to a number of points, also located on the network, using routes that are constrained to the network. Figure 15.15 shows the contrast between continuous and network views and Chapter 7 discusses data models for networks.

A very useful theorem first proved by Louis Hakimi reduces the complexity of many location problems on networks. Figure 15.15B shows a typical basis for network location. The links of the network come together in *nodes*. The weighted points are also located on the network, and also form nodes. For example, the task might be to find the location that minimizes total distance to a distribution of customers, with all customers aggregated into these weighted points. In this case the weights would be counts of customers. The Hakimi theorem proves that for this problem of minimizing distance the only locations that have to be considered are the nodes – it is impossible for the optimum location to be anywhere else. It is easy to see why this should be so. Think of a trial point located in the middle of a link, away from any node. Think of moving it slightly in one direction along the link. This moves it towards some weights, and away from others, but every unit of movement results in the same increase or decrease in total weighted distance. In other words, the total distance traveled to the location is a linear function of the location along the link. Since the function is linear it cannot have a minimum mid-link, so the minimum must occur at a node.

Optimum location problems can be solved in either discrete or continuous space, depending largely on scale.

In Section 15.2 the MAT in one dimension was shown to be the location such that half of the weight is to the left and half to the right – in other words, the measure of central tendency for ordinal data discussed in Section 15.2 and known as the *median*. The MAT problem on a network is consequently known as the *1-median* problem. The *p-median* problem seeks optimum locations for any number p of central facilities such that the sum of the distances between each weight and the *nearest* facility is minimized. A typical practical application of this problem is in the location of central public facilities, such as libraries, schools, or agency offices, when the objective is to locate for maximum total accessibility.

Many problems of this nature have been defined for different applications, and implemented in GIS. While the median problems seek to minimize total distance, the *coverage* problems seek to minimize the *furthest* distance

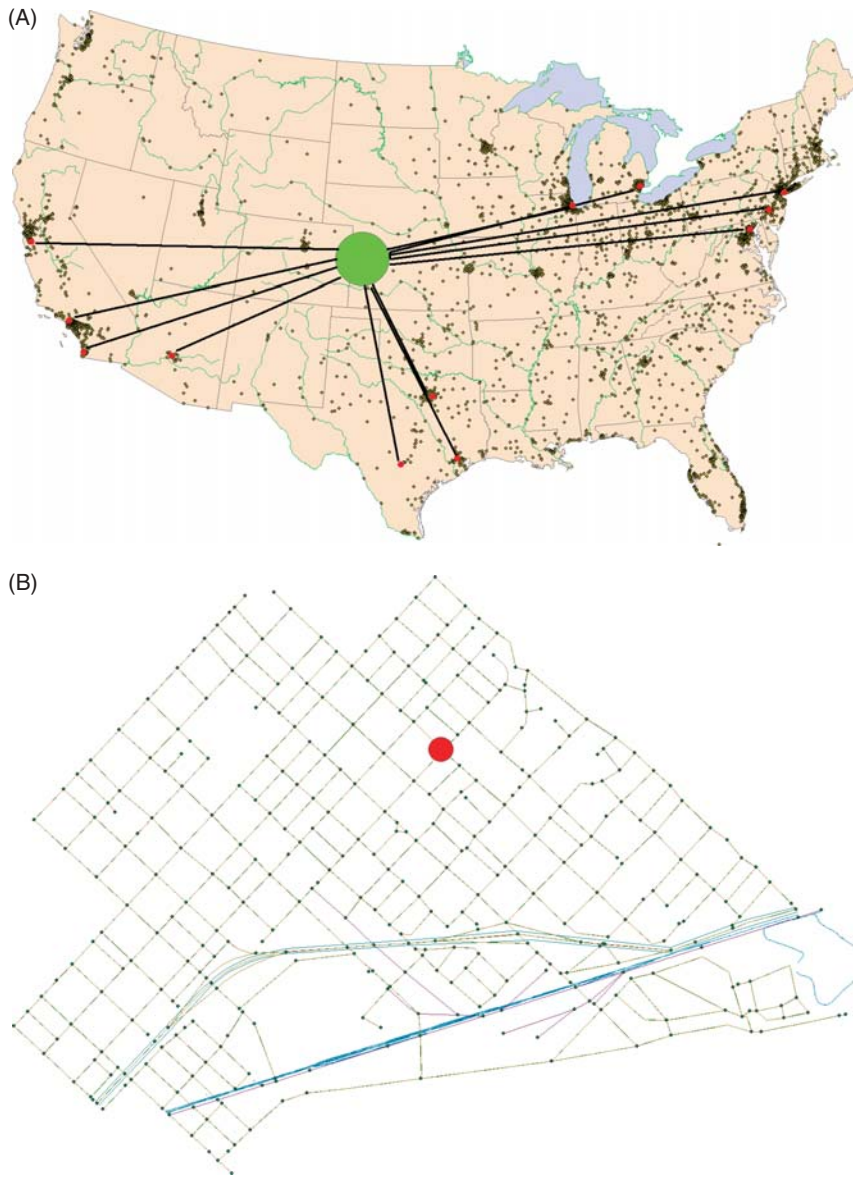


Figure 15.15 Search for the best locations for a central facility to serve dispersed customers. In (A) the problem is solved in continuous space, with straight-line travel, for a warehouse to serve the twelve largest US cities. In continuous space there is an infinite number of possible locations for the site. In (B) a similar problem is solved at the scale of a city neighborhood on a network, where Hakimi's theorem states that only junctions (nodes) in the network and places where there is weight need to be considered, making the problem much simpler, but where travel must follow the street network

traveled, on the grounds that dealing with the worst case of accessibility is often more attractive than dealing with average accessibility. For example, it may make more sense to a city fire department to locate so that a response is possible to *every* property in less than five minutes, than to worry about minimizing the *average* response time. Coverage problems find applications in the location of emergency facilities, such as fire stations (Figure 15.16), where it is desirable that every possible emergency be covered within a fixed number of minutes of response time, or when the objective is to minimize the worst-case response time, to the furthest possible point.

All of these problems are referred to as *location-allocation* problems, because they involve two types of decisions – where to *locate*, and how to *allocate* demand for service to the central facilities. A typical location-allocation problem might involve the selection of sites for supermarkets. In some cases, the allocation of demand to sites is controlled by the designer, as it is in the case of school districts when students have no choice of schools. In other cases, allocation is a matter of choice, and good designs depend on the ability to predict how consumers will choose among the available options. Models that make such predictions are known as *spatial*

Figure 15.16 GIS can be used to find locations for fire stations that result in better response times to emergencies

interaction models, and their use is an important area of GIS application in market research.

Location-allocation involves two types of decisions – where to locate, and how to allocate demand for service.

15.3.2 Routing problems

Point-location problems are concerned with the design of fixed locations. Another area of optimization is in routing and scheduling, or decisions about the optimum tracks followed by vehicles. A commonly encountered example is in the routing of delivery vehicles. In these examples there is a base location, a depot that serves as the origin and final destination of delivery vehicles, and a series of stops that need to be made. There may be restrictions on the times at which stops must be made. For example, a vehicle delivering home appliances may be required to visit certain houses at certain times, when the residents are home. Vehicle routing and scheduling solutions are used by parcel delivery companies, school buses, on-demand public transport vehicles, and many other applications (see Boxes 15.4 and 15.5).

Underlying all routing problems is the concept of the *shortest path*, the path through the network between a defined origin and destination that minimizes distance, or some other measure based on distance, such as travel time. Attributes associated with the network's links, such as length, travel speed, restrictions on travel direction, and level of congestion are often taken into account. Many people are now familiar with the routine solution of the shortest path problem by websites like MapQuest.com (Section 1.5.3), which solve many millions of such problems per day for travelers, and by similar technology used by in-vehicle navigation systems. They use standard algorithms developed decades ago, long before the advent of GIS. The path that is strictly shortest is often not suitable, because it involves too many turns or uses too many narrow streets, and algorithms will often be programmed to find longer routes that use faster highways, particularly freeways. Routes in Los Angeles, USA, for example, can often be caricatured as: 1) shortest route from origin to nearest freeway; 2) follow freeway network; 3) shortest route from nearest freeway to destination, even though this route may be far from the shortest.

The simplest routing problem with multiple destinations is the so-called *traveling-salesman problem* or TSP. In this problem, there are a number of places that must be visited in a tour from the depot and the distances between pairs of places are known. The problem is to select the best tour out of all possible orderings, in order to minimize the total distance traveled. In other words, the optimum is to be selected out of the available tours. If there are n places to be visited, including the depot, then there are $(n-1)!$ possible tours (the symbol ! indicates the product of the integers from 1 up to and including the number, known as the number's *factorial*), but since it is irrelevant whether any given tour is conducted in a forwards or backwards direction the effective number of options is $(n-1)!/2$. Unfortunately this number grows very rapidly with the number of places to be visited, as Table 15.1 shows.

A GIS can be very effective at solving routing problems because it is able to examine vast numbers of possible solutions quickly.

The TSP is an instance of a problem that becomes quickly unsolvable for large n. Instead, designers adopt procedures known as *heuristics*, which are algorithms designed to work quickly, and to come close to providing

Table 15.1 The number of possible tours in a traveling-salesman problem

Number of places to visit	Number of possible tours
3	1
4	3
5	12
6	60
7	360
8	2520
9	20160
10	181440

Applications Box **15.4**

Routing service technicians for Schindler Elevator.

Schindler Elevator Corporation is a US leader in elevator maintenance. It uses a GIS to plan efficient routes for preventative-maintenance technicians. Service work for each technician is scheduled in daily routes, with emphasis on geographically compact areas of operation to allow quick response in case of emergency calls (people stuck in elevators need a quick response). Each building is visited periodically so that the technician can inspect, grease, clean, and perform periodic maintenance tasks. Additionally, visits are constrained by customers who, for example, do not want elevators out of service during the busiest times.

Schindler's GIS, named Planning Assistant for Superintendent Scheduling (PASS), went into operation in mid-1999 (Figure 15.17). It creates optimum routes, based on multiple criteria, for a number of technicians assigned to one of Schindler's 140 offices in the USA. As well as assigning visits for multiple technicians, based on maintenance periods and customer preference, it is able to consider many other factors such as technician skill sets and permissible work hours (including overtime), legally binding break times, visit duration, and contract value. Much of this information is stored in Schindler's central SAP/R3 Enterprise Resource Planning (ERP) system, which must be queried to obtain the data before each routing operation.

The system has had a major impact on Schindler's business, saving them several millions of dollars annually (for an initial outlay of around $1 million). It has also allowed them to improve overall operational efficiency by restructuring offices based on routing requirements.

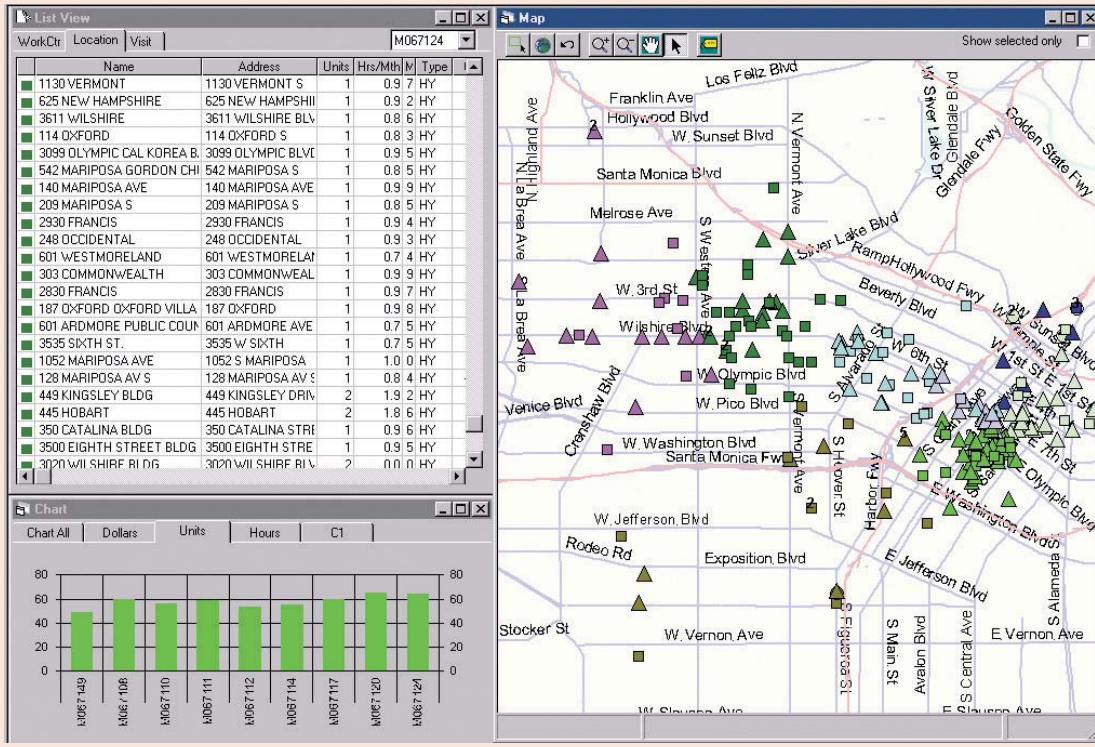


Figure 15.17 A daily routing and scheduling solution for Schindler Elevator. Each symbol on the map identifies a service stop, at the addresses listed in the table on the left

Applications Box **15.5**

Sears Smart Toolbox Mapping software

Sears, Roebuck and Company uses GIS navigational software in 12 000 Sears Home Services vehicles. The mapping application (based on ESRI ArcObjects technology) allows technicians to review daily scheduled routes, modify routes, and receive both audible and visible turn-by-turn directions while driving to a service location. The application runs on a dockable ruggedized touch-screen laptop (Figure 15.18) and communicates to a roof-mounted Mobile Base Station device that provides GPS information as well as satellite and terrestrial wireless data communication capabilities. The real-time GPS location information allows the application to track the vehicle along the optimized route and to reroute the vehicle dynamically when off route. The wireless connection allows the navigation application to download and upload service orders automatically throughout the technician's work day.

The use of GIS in the vehicles provides a tremendous business benefit to Sears. Prior to using this system a technician would need to look up a street name on a paper map and visually determine the best route approximately 8–10 times per day. The GIS eliminates this task from a technician's daily routine and reduces the chance of getting lost on the way to the stop, resulting in significant time and cost savings for Sears. In addition to this, by calculating the most efficient route, the system helps to minimize transportation distances and times, thus realizing additional savings. Along with reduced costs, the increased efficiency has improved customer service, as technicians are now more likely to meet their scheduled customer time windows. However, if a technician is likely to miss a time window the client-side application has the ability to upload a stop and send it to a server where it may be reassigned to another technician. Conversely, if a technician has extra time in his or her day, and a high-priority service call comes in, a new stop may be wirelessly downloaded to the vehicle and added as a stop.

Figure 15.18 Screen shot from the in-vehicle Smart Toolbox Mapping software installed on 12 000 Sears service vehicles, showing the optimum route to the next four stops to be made by this vehicle in Chicago

the best answer, while not guaranteeing that the best answer will be found. One not very good heuristic for the TSP is to proceed always to the closest unvisited destination, and finally to return to the start. Many spatial optimization problems, including location-allocation and routing problems, are solved today by the use of sophisticated heuristics, and a heuristic approach was used to derive the solution shown in Figure 15.17.

There are many ways of generalizing the TSP to match practical circumstances. Often there is more than one vehicle and in these situations the division of stops between vehicles is an important decision variable – which driver should cover which stops? In other situations it is not essential that every stop be visited, and instead the problem is one of maximizing the rewards associated with visiting stops, while at the same time minimizing the total distance traveled. The is known as the *orienteering* problem, and it matches the situation faced by the driver of a mobile snack bar who must decide which of a number of building sites to visit during the course of a day. ESRI's ArcLogistics and Caliper Corp's TransCAD (**www.caliper.com**) are examples of packages built on GIS platforms to solve many of these everyday problems using heuristics.

15.3.3 Optimum paths

The last example in this section concerns the need to find paths across continuous space, for linear facilities like highways, power lines, and pipelines. Locational decisions are often highly contentious, and there have been many instances of strong local opposition to the location of high-voltage transmission lines (which are believed by many people to cause cancer, although the scientific evidence on this issue is mixed). Optimum paths across continuous space are also needed by the military, in routing tanks and other vehicles, and there has been substantial investment by the military in appropriate technology. They are used by shipping companies to minimize travel time and fuel costs in relation to currents. Finally, continuous-space routing problems are routinely solved by the airlines, in finding the best tracks for aircraft across major oceans that minimize time and fuel costs, by taking local winds into account. Aircraft flying the North Atlantic typically follow a more southerly route from west to east, and a more northerly route from east to west (see Figure 14.11).

A GIS can find the optimum path for a power line or highway, or for an aircraft flying over an ocean against prevailing winds.

Practical GIS solutions to this problem are normally solved on a raster. Each cell is assigned a *friction* value, equal to the cost or time associated with moving across the cell in the horizontal or vertical directions (compare Figure 14.18). The user selects a *move set*, or a set of allowable moves between cells, and the solution is expressed in the form of a series of such moves. The simplest possible move set is the *rook's case*, named for the allowed moves of the rook (castle) in the game of chess. But in practice most software uses the *queen's case*, which allows moves in eight directions from each cell rather than four. Figure 15.19 shows the two cases. When a diagonal move is made the associated cost or time must be multiplied by a factor of 1.414 (i.e., $\sqrt{2}$) to account for its extra length.

Friction values can be obtained from a number of sources. Land-cover data might be used to differentiate between forest and open space – a power line might be given a higher cost in forest because of the environmental impact (felling of forest), while its cost in open space might be allocated based on visual impact. Elaborate methods have been devised for estimating costs based on many GIS layers and appropriate weighting factors. Real financial costs might be included, such as the costs of construction or of land acquisition, but most of the costs are likely to be based on intangibles, and consequently hard to measure effectively.

Economic costs of a project may be relatively easy to measure compared to impacts on the environment or on communities.

Given a cost layer, together with a defined origin and destination and a move set, the GIS is then asked to find the least-cost path. If the number of cells in the raster is very large the problem can be computationally time-consuming, so heuristics are often employed. The simplest is to solve the problem hierarchically, first on a coarsely aggregated raster to obtain an approximate corridor, and then on a detailed raster restricted to the corridor.

There is an interesting paradox with this problem that is often overlooked in GIS applications, but is nevertheless significant and provides a useful example of the difference between interval and ratio measurements (Box 3.3). Figure 15.20 shows a typical solution to a routing problem. Since the task is to minimize the sum of costs, it is clear that the costs or friction values in each cell need to be measured on interval or ratio scales. But is interval measurement sufficient, or is ratio measurement required – in other words, does there need to be an absolute zero to the scale? Many studies of optimum routing have been done using scales that clearly did not have an absolute zero. For example, people along the route

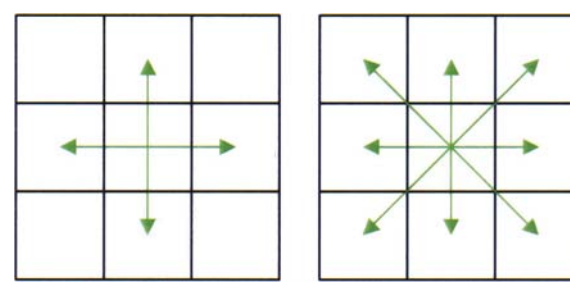


Figure 15.19 The rook's-case (left) and queen's-case (right) move sets, defining the possible moves from one cell to another in solving the problem of optimum routing across a friction surface

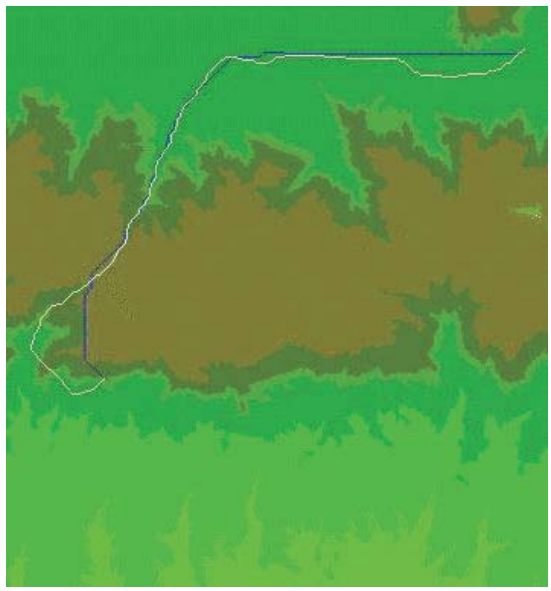


Figure 15.20 Solution of a least-cost path problem. The white line represents the optimum solution, or path of least total cost, across a friction surface represented as a raster, and using the queen's-case move set (eight possible moves). Brown denotes the highest friction. The blue line is the solution when cells are aggregated in blocks of 3 by 3, and the problem is solved on this coarser raster (Reproduced by permission of Ashton Shortridge)

might have been surveyed and asked to rate impact on a scale from 1 to 5. It is questionable whether such as scale has interval properties (are 1 and 2 as different on the scale as 2 and 3, or 4 and 5?), let alone ratio properties (there is no absolute zero and no sense that a score of 4 is twice as much of anything as a score of 2).

But why is this relevant to the problem of route location? Consider the path shown in Figure 15.20. It passes through 456 cells and includes 261 rook's-case moves and 195 diagonal moves. If the entire surface were raised by adding 1.0 to every score, the total cost of the optimum path would rise by 536.7, allowing for the greater length of the diagonal moves. But the cost of a straight path, which requires only 60 rook's-case and 216 diagonal moves, would rise by only 365.4. In other words, changing the zero point of the scale can change the optimum – and in principle the cost surface needs to be measured on a ratio scale. Interval and ordinal scales are not sufficient, despite the fact that they are frequently used.

15.4 Hypothesis testing

This last section reviews a major area of statistics – the testing of hypotheses and the drawing of inferences – and its relationship to GIS and spatial analysis. Much work in statistics is *inferential* – it uses information obtained from samples to make general conclusions about a larger population, on the assumption that the sample came from that population. The concept of inference was introduced in Section 4.4, as a way of reasoning about the properties of a larger group from the properties of a sample. At that point several problems associated with inference from geographic data were raised, and this section revisits and elaborates on that topic and discusses the particularly thorny issue of hypothesis testing.

For example, suppose we were to take a random and independent sample of 1000 people, and ask them how they might vote in the next election. By *random and independent*, we mean that every person of voting age in the general population has an equal chance of being chosen and the choice of one person does not make the choice of any others – parents, neighbors – more or less likely. Suppose that 45% of the sample said they would support George W. Bush. Statistical theory then allows us to give 45% as the best estimate of the proportion who would vote for Bush among the general population, and also allows us to state a *margin of error*, or an estimate of how much the true proportion among the population will differ from the proportion among the sample. A suitable expression of margin of error is given by the 95% confidence limits, or the range within which the true value is expected to lie 19 times out of 20. In other words, if we took 20 different samples, all of size 1000, there would be a scatter of proportions, and 19 out of 20 of them would lie within these 95% confidence limits. In this case, a simple analysis using the *binomial distribution* shows that the 95% confidence limits are 3% – in other words, the true proportion lies between 42% and 48% 19 times out of 20.

This example illustrates the *confidence limits* approach to inference, in which the effects of sampling are expressed in the form of uncertainty about the properties of the population. An alternative, very commonly used in scientific reasoning, is the *hypothesis-testing* approach. In this case, our objective is to test some general statement about the population – for example, that 50% will support Bush in the next election (and 50% will support the other candidate – in other words, there is no real preference in the electorate). We take a sample and then ask whether the evidence from the sample supports the general statement. Because there is uncertainty associated with any sample, unless it includes the entire population, the answer is never absolutely certain. In this example and using our confidence limits approach, we know that if 55% were found to support Bush in the sample, and if the margin of error was 3%, it is highly unlikely that the true proportion in the population is as low as 50%. Alternatively, we could state the 50% proportion in the population as a *null hypothesis* (we use the term *null* to reflect the absence of something, in this case a clear choice), and determine how frequently a sample of 1000 from such a population would yield a proportion as high as 55%. The answer again is very small; in fact the probability is 0.0008. But it is not zero, and its value represents the chance of making an error of inference – of rejecting the hypothesis when in fact it is true.

Methods of inference reason from information about a sample to more general information about a larger population.

These two concepts – confidence limits and inferential tests – are the basis for statistical testing and form the core of introductory statistics texts. There is no point in reproducing those introductions here, and the reader is simply referred to them for discussions of the standard tests – the F, t, χ^2, etc. The focus here is on the problems associated with using these approaches with geographic data, in a GIS context. Several problems were already discussed in Section 4.7 as violations of the assumptions of inference from regression. The next section reviews the inferential tests associated with one popular descriptive statistic for spatial data, the Moran index of spatial dependence that was discussed in Section 4.6. The following section discusses the general issues and points to ways of resolving them.

15.4.1 Hypothesis tests on geographic data

Although inferential tests are standard practice in much of science, they are very problematic for geographic data. The reasons have to do with fundamental properties of geographic data, many of which were introduced in Chapter 4, and others have been encountered at various stages in this book.

First, many inferential tests propose the existence of a population, from which the sample has been obtained by some well-defined process. We saw in Section 4.4 how difficult it is to think of a geographic dataset as a sample of the datasets that might have been. It is equally difficult to think of a dataset as a sample of some larger area of the Earth's surface, for two major reasons.

First, the samples in standard statistical inference are obtained independently (Section 4.4). But a geographic dataset is often *all there is* in a given area – it *is* the population. Perhaps we could regard a dataset as a sample of a larger area. But in this case the sample would not have been obtained randomly – instead, it would have been obtained by systematically selecting all cases within the area of interest. Moreover, the samples would not have been independent. Because of spatial dependence, which we have understood to be a pervasive property of geographic data, it is very likely that there will be similarities between neighboring observations.

A GIS project often analyzes all the data in a given area, rather than a sample.

Figure 15.21 shows a typical instance of sampling topographic elevation. The value obtained at B could have been estimated from the values obtained at the neighboring points A and C – and this ability to estimate is of course the entire basis of spatial interpolation. We cannot have it both ways – if we believe in spatial interpolation, we cannot at the same time believe in independence of geographic samples, despite the fact that this is a basic assumption of statistical tests.

Finally, the issue of spatial *heterogeneity* also gets in the way of inferential testing. The Earth's surface is highly variable and there is no such thing as an average

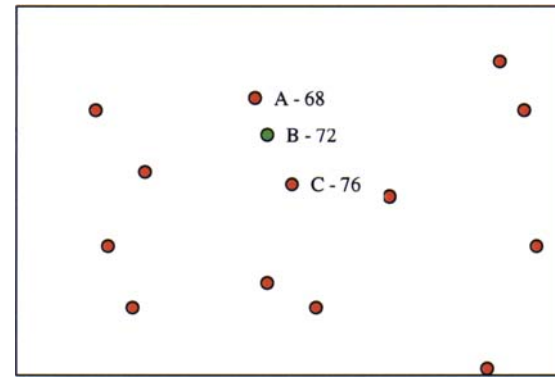


Figure 15.21 Spatial dependence in observations. The value at B, which was measured at 72, could easily have been interpolated from the values at A and C – any spatial interpolation technique would return a value close to 72 at this point

place on it. The characteristics observed on one map sheet are likely substantially different from those on other map sheets, even when the map sheets are neighbors. So the census tracts of a city are certainly not acceptable as a random and independent sample of all census tracts, even the census tracts of an entire nation. They are not independent, and they are not random. Consequently it is very risky to try to infer properties of *all* census tracts from the properties of all of the tracts in any *one* area. The concept of sampling, which is the basis for statistical inference, does not transfer easily to the spatial context.

The Earth's surface is very heterogeneous, making it difficult to take samples that are truly representative of any large region.

Before using inferential tests on geographic data, therefore, it is advisable to ask two fundamental questions:

- Can I conceive of a larger *population* about which I want to make inferences?
- Are my data acceptable as a *random* and *independent* sample of that population?

If the answer to either of these questions is *no*, then inferential tests are not appropriate.

Given these arguments, what options are available? One strategy that is sometimes used is to discard data until the proposition of independence becomes acceptable – until the remaining data points are so far apart that they can be regarded as essentially independent. But no scientist is happy throwing away good data.

Another approach is to abandon inference entirely. In this case the results obtained from the data are descriptive of the study area and no attempt is made to generalize. This approach, which uses local statistics to observe the *differences* in the results of analysis over space, represents an interesting compromise between the nomothetic and idiographic positions outlined in Section 1.3. Generalization is very tempting, but the heterogeneous nature of the Earth's surface makes it very difficult. If generalization is required, then it can be accomplished by appropriate experimental design – by

replicating the study in a sufficient number of distinct areas to warrant confidence in a generalization.

Another, more successful, approach exploits the special nature of spatial analysis and its concern with detecting pattern. Consider the example in Figure 15.10, where the Moran index was computed at +0.4011, an indication that high values tend to be surrounded by high values, and low values by low values – a positive spatial autocorrelation. It is reasonable to ask whether such a value of the Moran index could have arisen by chance, since even a random arrangement of a limited number of values typically will not give the theoretical value of zero corresponding to no spatial dependence (actually the theoretical value is very slightly negative). In this case, 51 values are involved, arranged over the 51 features on the map (the 50 states plus the District of Columbia; Hawaii and Alaska are not shown). If the values were arranged randomly, how far would the resulting values of the Moran index differ from zero? Would they differ by as much as 0.4011? Intuition is not good at providing answers.

In such cases a simple test can be run by simulating random arrangements. The software used to prepare this illustration includes the ability to make such simulations, and Figure 15.22 shows the results of simulating 999 random rearrangements. It is clear that the actual value is well outside the range of what is possible for random arrangements, leading to the conclusion that the apparent spatial dependence is real.

How does this *randomization* test fit within the normal framework of statistics? The null hypothesis being evaluated is that the distribution of values over the 51 features is random, each feature receiving a value that is independent of neighboring values. The population is the set of all possible arrangements, of which the data represent a sample of one. The test then involves comparing the test statistic – the Moran index – for the actual pattern against the distribution of values produced by the null hypothesis. So the test is a perfect example of standard hypothesis testing, adapted to the special nature of spatial data and the common objective of discovering pattern.

Randomization tests are uniquely adapted to testing hypotheses about spatial pattern.

Finally, a large amount of research has been devoted to devising versions of inferential tests that cope effectively with spatial dependence and spatial heterogeneity. Software that implements these tests is now widely available and interested readers are urged to consult the appropriate sources. GeoDa is an excellent comprehensive software environment for such tests (see **www.csiss.org** under Spatial Tools), and many of these methods are available as extensions of standard GIS packages.

15.5 Conclusion

This chapter has covered the conceptual basis of many of the more sophisticated techniques of spatial analysis that are available in GIS. The last section in particular raised some fundamental issues associated with applying methods and theories that were developed for non-spatial data to the spatial case. Spatial analysis is clearly not a simple and straightforward extension of non-spatial analysis, but instead raises many distinct problems, as well as some exciting opportunities. The two chapters on spatial analysis have only scratched the surface of this large and rapidly expanding field.

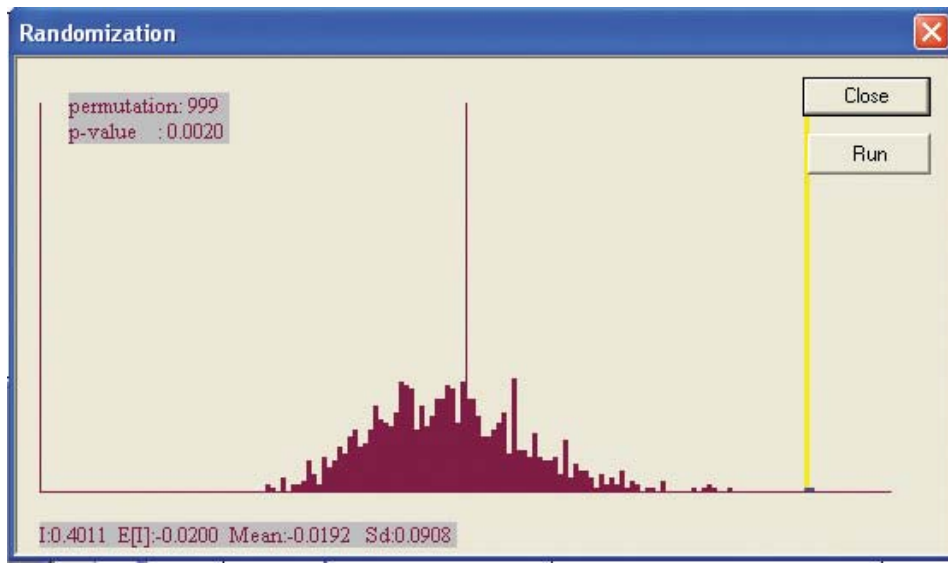


Figure 15.22 Randomization test of the Moran index computed in Figure 15.10. The histogram shows the results of computing the index for 999 rearrangements of the 51 values on the map (Hawaii and Alaska are not shown). The yellow line on the right shows the actual value, which is clearly very unlikely to occur in a random arrangement, reinforcing the conclusion that there is positive spatial dependence in the data

Questions for further study

1. Parks and other conservation areas have geometric shapes that can be measured by comparing park perimeter length to park area, using the methods reviewed in this chapter. Discuss the implications of shape for park management, in the context of a) wildlife ecology and b) neighborhood security.
2. What exactly are *multicriteria* methods? Examine one or more of the methods in the Eastman chapter referenced below, summarizing the issues associated with a) measuring variables to support multiple criteria, b) mixing variables that have been measured on different scales (e.g., dollars and distances), c) finding solutions to problems involving multiple criteria.
3. Besides being the basis for useful measures, fractals also provide interesting ways of simulating geographic phenomena and patterns. Browse the Web for sites that offer fractal simulation software, or investigate one of many commercially available packages. What other uses of fractals in GIS can you imagine?
4. Every point on the Earth's surface has an antipodal point – the point that would be reached by drilling an imaginary hole straight through the Earth's center. Britain, for example, is approximately antipodal to New Zealand. If one third of the Earth's surface is land, you might expect that one third of all of the land area would be antipodal to points that are also on land, but a quick look at an atlas will show that the proportion is actually far less than that. In fact, the only substantial areas of land that have antipodal land are in South America (and their antipodal points in China). How is spatial dependence relevant here, and why does it suggest that the Earth is not so surprising after all?

Further reading

Cliff A.D. and Ord J.K. 1973 *Spatial Autocorrelation*. London: Pion.

Eastman J.R. 1999 'Multicriteria methods.' In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition (2005))*. Hoboken, NJ: Wiley.

Fotheringham A.S., Brunsdon C., and Charlton M. 2000 *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage.

Fotheringham A.S. and O'Kelly M.E. 1989 *Spatial Interaction Models: Formulations and Applications*. Dordrecht: Kluwer.

Ghosh A. and Rushton G. (eds) 1987 *Spatial Analysis and Location-Allocation Models*. New York: Van Nostrand Reinhold.

Mardia K.V. and Jupp P.E. 2000 *Directional Statistics*. New York: Wiley.

16 *Spatial modeling with GIS*

Models are used in many different ways, from simulations of how the world works, to evaluations of planning scenarios, to the creation of indicators of suitability or vulnerability. In all of these cases the GIS is used to carry out a series of transformations or analyses, either at one point in time or at a number of intervals. This chapter begins with the necessary definitions and presents a taxonomy of models, with examples. It addresses the difference between analysis, the subject of Chapters 14 and 15, and this chapter's subject of modeling. The alternative software environments for modeling are reviewed, along with capabilities for cataloging and sharing models, which are developing rapidly. The chapter ends with a look into the future of modeling and associated GIS developments.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

At the end of this chapter you will:

- Know what modeling means in the context of GIS;
- Be familiar with the important types of models and their applications;
- Be familiar with the software environments in which modeling takes place;
- Understand the needs of modeling, and how these are being addressed by current trends in GIS software.

16.1 Introduction

This chapter identifies many of the distinct types of models supported by GIS and gives examples of their applications. After *system* and *object*, *model* is probably one of the most overworked technical terms in the English language, with many distinct meanings in the context of GIS and even of this book. So first it is important to address the meaning of the term as it is used in this chapter.

***Model* is one of the most overworked terms in the English language.**

A clear distinction needs to be made between the *data models* discussed in Chapters 3 and 8 and the spatial models that are the subject of this chapter. A data model is a template for data, a framework into which specific details of relevant aspects of the Earth's surface can be fitted. For example, the raster data model forces all knowledge to be expressed as properties of the cells of a regular grid laid on the Earth. A data model is, in essence, a statement about form or about how the world *looks*, limiting the options open to the data model's user to those allowed by its template. Models in this chapter are expressions of how the world is believed to *work*, in other words they are expressions of process (see Section 1.3 on how these both relate to the science of problem solving). They may include dynamic simulation models of natural processes such as erosion, tectonic uplift, the migration of elephants, or the movement of ocean currents. They may include models of social processes, such as residential segregation or the movements of cars on a congested highway. They may include processes designed by humans to search for optimum alternatives, for example, in finding locations for a new retail store. Finally, they may include simple calculations of indicators or predictors, such as happens when layers of geographic information are combined into measures of groundwater vulnerability or social deprivation.

The common element in all of these examples is the manipulation of geographic information in multiple stages. In some cases these stages will perform a simple transformation or analysis of inputs to create an output, and in other cases the stages will loop to simulate the development of the modeled system through time, in a series of iterations – only when all the loops or iterations are complete will there be a final output. There will be intermediate outputs along the way, and it is often desirable to save some or all of these in case the model needs to be re-run or parts of the model need to be changed.

All of the models discussed in this chapter are *digital* or *computational* models, meaning that the operations occur in a computer and are expressed in a language of 0s and 1s. In Chapter 3, representation was seen as a matter of expressing geographic form in 0s and 1s – this chapter looks for ways of expressing geographic *process* in 0s and 1s. The term *geocomputation* is often used to describe the application of computational models to geographic problems; it is also the name of a successful conference series.

All of the models discussed in this chapter are also *spatial* models. There are two key requirements of such a model:

1. There is variation across the space being manipulated by the model (an essential requirement of all GIS applications, of course); and
2. The results of modeling change when the locations of objects change – location *matters* (this is also a key requirement of spatial analysis as defined in Section 14.1).

Models do not have to be digital, and it is worth spending a few moments considering the other type, known as *analog*. The term was defined briefly in Section 3.7 as describing a representation, such as a paper map, that is a scaled physical replica of reality. The Varignon Frame discussed in Box 15.1 is an analog model of a design process, the search for the point of minimum aggregate travel. Analog models can be very efficient, and they are widely used to test engineering construction projects and proposed airplanes. They have two major disadvantages relative to digital models, however: they can be expensive to construct and operate, and unlike digital models they are difficult to copy, store, or share. Box 16.1 describes the analog experiments conducted in World War II in support of a famous bombing mission.

An analog model is a scaled physical representation of some aspect of reality.

The level of detail of any analog model is measured by its *representative fraction* (Section 3.7, Box 4.2), the ratio of distance on the model to distance in the real world. Like digital data, computational models do not have a well-defined representative fraction and instead the level of detail is measured as *spatial resolution*, defined as the

Applications Box **16.1**

Modeling the effects of bombing on a dam

In World War II, the British conceived a scheme for attacking three dams upstream of the strategically important Ruhr valley, the center of the German iron and steel industry. Dr Barnes Wallis envisioned a roughly spherical bomb that would bounce on the water upstream of the dam, settling next to the dam wall before exploding. Tests were conducted on scaled models (Figure 16.1) to prove the concept. This approach assumes that all aspects of the real problem scale equally, so that conclusions reached using the scale model are also true in reality – and because this assumption may be shaky, it is advisable to test the conclusions of the model later with full-scale prototypes. The successful attacks by the Royal Air Force's 617 Squadron were immortalized in a 1951 book (*The Dam Busters* by Paul Brickhill) and in a 1954 movie (*The Dam Busters* starring Richard Todd as Wing Commander Guy Gibson).

With modern computer power, much of the experimental work could be done quickly and cheaply using computational models instead of scaled analog models. It would be necessary first to build a digital representation of the system (the dam, the lake, and the bomb), and then to build a digital model of the processes that occur when a bomb is dropped at a shallow angle onto a lake surface, bounces on the surface, and then explodes underwater in close proximity to a concrete dam wall.

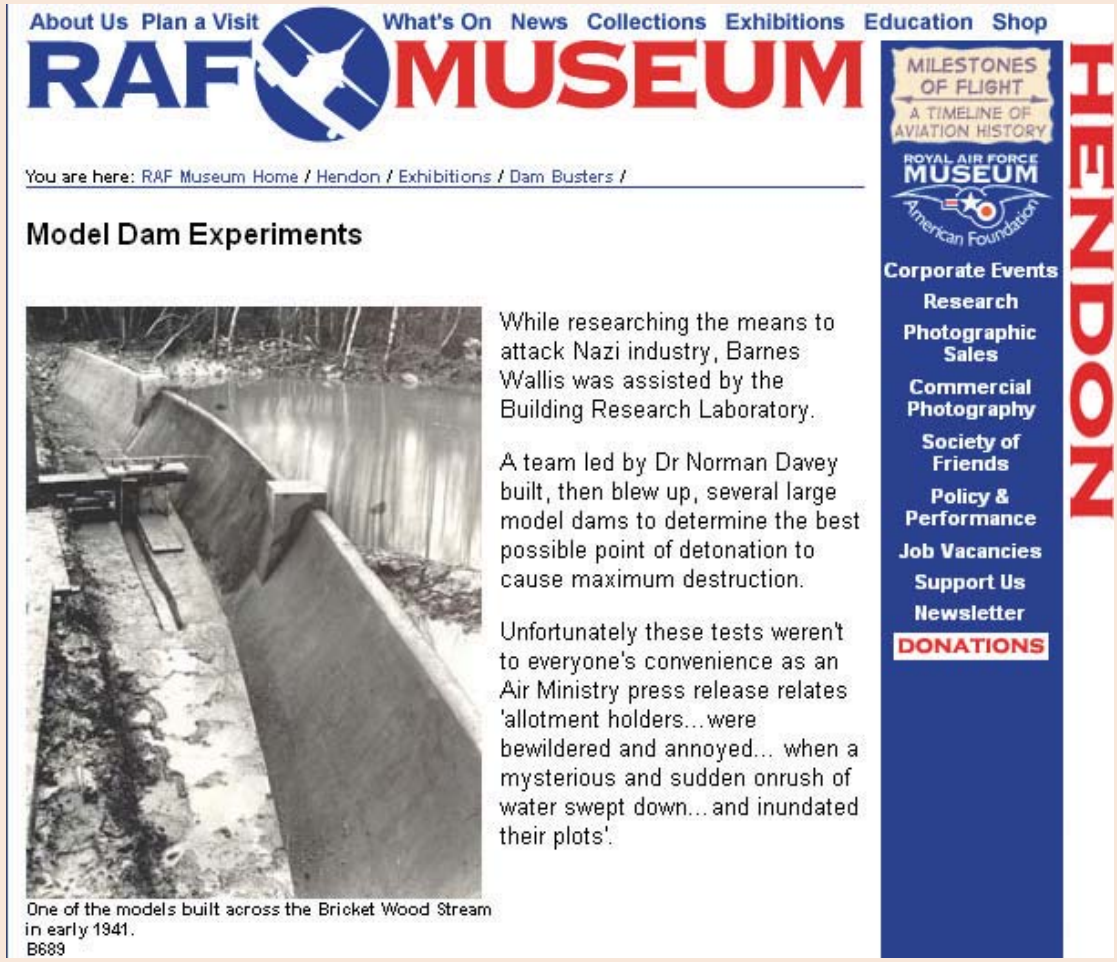


Figure 16.1 An analog model built in 1941 to validate the 'bouncing bomb' ideas of Dr Barnes Wallis in preparation for the successful Royal Air Force assault on German dams in 1943 (Reproduced by permission of Royal Air Force Museum)

shortest distance over which change is recorded. *Temporal resolution* is also important, being defined as the shortest time over which change is recorded and in the case of many dynamic models corresponding to the time interval between iterations.

Spatial and temporal resolution are critical factors in models. They define what is left out of the model, in the form of variation that occurs over distances or times that are less than the appropriate resolution. They also therefore define one major source of uncertainty in the model's outcomes. Uncertainty in this context can be defined best through a comparison between the model's outcomes and the outcomes of the real processes that the model seeks to emulate. Any model leaves its user uncertain to some degree about what the real world will do – a measure of uncertainty attempts to give that degree of uncertainty an explicit magnitude. Uncertainty has been discussed in the context of geographic data in Chapter 6; its meaning and treatment are discussed in the context of spatial modeling later in this chapter.

Any model leaves its user uncertain to some degree about what the real world will do.

Spatial and temporal resolution also determine the cost of acquiring the data because in general it is more costly to collect a high-resolution representation than a low-resolution one. More observations have to be made and more effort is consumed in making them. They also determine the cost of running the model, since execution time expands as more data have to be processed and more iterations have to be made. One benchmark is of critical importance for many dynamic models: they must run faster than the processes they seek to simulate if the results are to be useful for planning. One expects this to be true of computational models, but in practice the amount of computing can be so large that the model simulation slows to unacceptable speed.

Spatial resolution is a major factor in the cost both of acquiring data for modeling, and of actually running the model.

16.1.1 Why model?

Models are built for a number of reasons. First, a model might be built to support a design process, in which the user wishes to find a solution to a spatial problem that optimizes some objective. This concept was discussed in Section 15.3 in the context of locating both point-like facilities such as retail stores or fire stations, and line-like facilities such as transmission corridors or highways. Often the design process will involve multiple criteria, an issue discussed in Section 16.4.

Second, a model might be built to allow the user to experiment on a replica of the world, rather than on the real thing. This is a particularly useful approach when the costs of experimenting with the real thing are prohibitive, or when unacceptable impacts would result, or when results can be obtained much faster with a model. Medical students now routinely learn anatomy and the basics of surgery by working with digital representations of the human body, rather than with expensive and hard-to-get cadavers. Humanity is currently conducting an unprecedented experiment on the global atmosphere by pumping vast amounts of CO_2 into it. How much better it would have been if we could have run the experiment on a digital replica, and understood the consequences of CO_2 emissions well in advance.

Experiments embody the notion of *what-if scenarios*, or policy alternatives that can be plugged into a model in order to evaluate their outcomes. The ability to examine such options quickly and effectively is one of the major reasons for modeling (Figure 16.2).

Models allow planners to experiment with 'what-if' scenarios.

Finally, models allow the user to examine dynamic outcomes, by viewing the modeled system as it evolves and responds to inputs. As discussed in the next section, such dynamic visualizations are far more compelling and convincing when shown to stakeholders and the public than descriptions of outcomes or statistical summaries. Scenarios evaluated with dynamic models are thus a very effective way of motivating and supporting debates over policies and decisions.

16.1.2 To analyze or to model?

Section 2.3.4.2 discussed the work of Tom Cova in analyzing the difficulty of evacuating neighborhoods in Santa Barbara, California, USA, in response to an emergency caused, for example, by a rapidly moving wildfire. The analysis provided useful results that could be employed by planners in developing traffic control strategies, or in persuading residents to use fewer cars when they evacuate their homes. But the results are presented in the form of static maps that must be explained to residents, who might not otherwise understand the role of the street pattern and population density in making evacuation difficult.

By contrast, Box 16.2 describes the dynamic simulations developed by Richard Church. These produce outcomes in the form of dramatic animations that are instantly meaningful to residents and planners. The researcher is able to examine what-if scenarios by varying a range of parameters, including the number of vehicles per household in the evacuation, the distribution of traffic control personnel, the use of private and normally gated exit roads, and many other factors. Simulations like these can galvanize a community into action far more effectively than static analysis.

Models can be used for dynamic simulation, providing decision makers with dramatic visualizations of alternative futures.

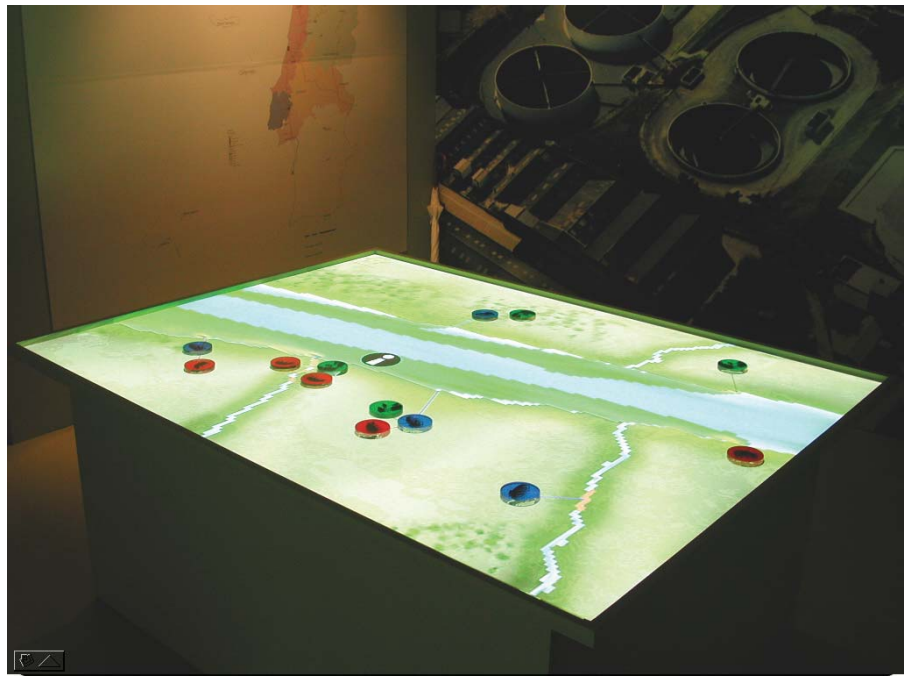


Figure 16.2 A 'live' table constructed by Antonio Câmara and his group (see Box 11.5) as a tool for examining alternative planning scenarios. The image on the table is projected from a computer and users are able to 'move' objects that represent sources of pollution around on the image by interacting directly with the display. Each new position results in the calculation of a hydrologic model of the impacts of the sources, and the display of the resulting pattern of impacts on the table (Reproduced by permission of Antonio Camara/Ydreams)

Biographical Box **16.2**

Richard Church and traffic simulation

Figure 16.3 Richard Church, Professor of Geography, University of California, Santa Barbara

Richard Church (Figure 16.3) received his Bachelor of Science degree from Lewis and Clark College. After finishing two majors, one in Math and one in Chemistry, he continued his graduate education in Geography and Environmental Engineering at The Johns Hopkins University where he specialized in water resource systems engineering. Currently, Rick is Professor of Geography at the University of California, Santa Barbara. He continues to be involved in environmental modeling and has specialized in forest systems modeling and ecological reserve design. He also teaches and conducts research in the areas of logistics, transportation, urban planning, and location modeling.

Rick is often motivated by real-world problems, whether it is operating a reservoir, designing an ambulance posting plan for a city, or finding the best corridor for a right-of-way.

One of his recent projects involves simulating the emergency evacuation of a neighborhood in advance of a rapidly moving fire, such as the one in the Oakland Hills in 1991 that destroyed 2900 structures and resulted in 25 deaths. Figure 16.4 shows the Mission Canyon area of Santa Barbara, a neighborhood that exhibits many of the same characteristics as the Oakland Hills.

Aware of the potential threat, members of the Mission Canyon neighborhood association asked if there was a way in which they could better understand the difficulty they would face in an evacuation. To address

this problem Church utilized a micro-scale traffic simulation program called PARAMICS to model a number of evacuation scenarios. Figure 16.5 shows the simulated pattern some minutes into an evacuation, with cars bumper-to-bumper crowding the available exits, and unable to move should the fire advance over them. Through this work, it was evident that the biggest gain in safety was to encourage residents to use one vehicle per household in making a quick exit. Furthermore, traffic control was found to be very important

Figure 16.4 The Mission Canyon neighborhood in Santa Barbara, California, USA (highlighted in red), a critical area for evacuation in a wildfire emergency. Note the contorted network of streets in this topographically rugged area

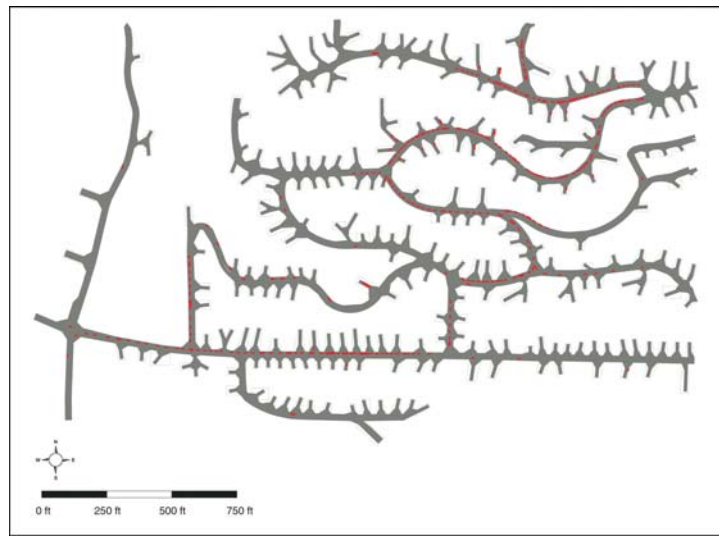


Figure 16.5 Simulation of driver behavior in a Mission Canyon evacuation some minutes after the initial evacuation order. The red dots denote the individual vehicles whose behavior is modeled in this simulation

in decreasing overall clearing time. The County Government has since agreed to pre-position traffic control officers during 'red flag' alerts (times of extreme fire danger) so that the time to respond with traffic control measures is kept as low as possible. The work reported here was sponsored by a grant from the California Department of Transportation (Caltrans).

The simulations using PARAMICS raise an important issue for GIS. The conventional representation of a street in a GIS is as a polyline, a series of points connected by straight lines. However it would be impossible to simulate driver behavior on such a representation because it would be necessary to slow to negotiate the sharp bends at each vertex. Instead PARAMICS uses a representation consisting of a series of circular arcs with no change of direction where arcs join. Integrating PARAMICS with GIS therefore requires transformation software to convert polylines to acceptable representations (Figure 16.6).

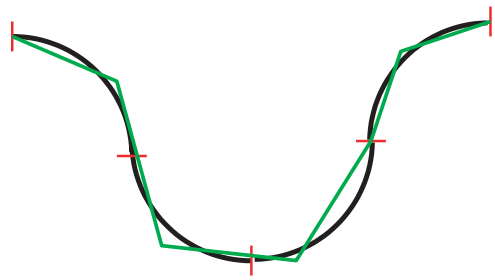


Figure 16.6 The representation problem for driver simulation. The green line shows a typical polyline representation found in a GIS. But a simulated driver would have to slow almost to a halt to negotiate the sharp corners. Traffic simulation packages such as PARAMICS require a representation of circular arcs (the black line) with no change of direction where the arcs join (at the points indicated by red lines)

In summary, analysis as described in Chapters 14 and 15 is characterized by:

- A static approach, at one point in time;
- The search for patterns or anomalies, leading to new ideas and hypotheses; and
- Manipulation of data to reveal what would otherwise be invisible.

By contrast, modeling is characterized by:

- Multiple stages, perhaps representing different points in time;
- Implementing ideas and hypotheses; and
- Experimenting with policy options and scenarios.

16.2 Types of model

16.2.1 Static models and indicators

A static model represents a single point in time and typically combines multiple inputs into a single output. There are no time steps and no loops in a static model, but the results are often of great value as predictors or indicators. For example, the Universal Soil Loss Equation (USLE, first discussed in Section 1.3) falls into this category. It predicts soil loss at a point, based on five input variables, by evaluating the equation:

$$A = R \times K \times LS \times C \times P$$

where A denotes the predicted erosion rate, R is the Rainfall and Runoff Factor, K is the Soil Erodibility Factor, LS is the Slope Length Gradient Factor, C is the Crop/Vegetation and Management Factor, and P is the Support Practice Factor. Full definitions of each of these variables and their measurement or estimation can be found in descriptions of the USLE (see for example **www.co.dane.wi.us/landconservation/uslepg.htm**).

A static model represents a system at a single point in time.

The USLE passes the first test of a spatial model in that many if not all of its inputs will vary spatially when applied to a given area. But it does not pass the second test, since moving the points at which A is evaluated will not affect the results. Why, then, use a GIS to evaluate the USLE? There are three good reasons:

1) because some of the inputs, particularly *LS*, require a GIS for their calculation from readily available data, such as digital elevation models; 2) because the inputs and output are best expressed, visualized, and used in map form, rather than as tables of point observations; and 3) because the inputs and outputs of the USLE are often integrated with other types of data, for further analysis that may require a GIS. Nevertheless, it is possible to evaluate the USLE in a simple spreadsheet application such as Excel, and the website cited in the previous paragraph includes a downloadable Excel macro for this purpose.

Models that combine a variety of inputs to produce an output are widely used, particularly in environmental modeling. The DRASTIC model calculates an index of groundwater vulnerability from input layers by applying appropriate weights to the inputs (Figure 16.7). Box 16.3 describes another application, the calculation of a groundwater protection model, from several inputs in a karst environment (an area underlain by potentially soluble limestone, and therefore having substantial and rapid groundwater flow through cave passages). It uses ESRI's ModelBuilder software which is described later in Section 16.3.1.

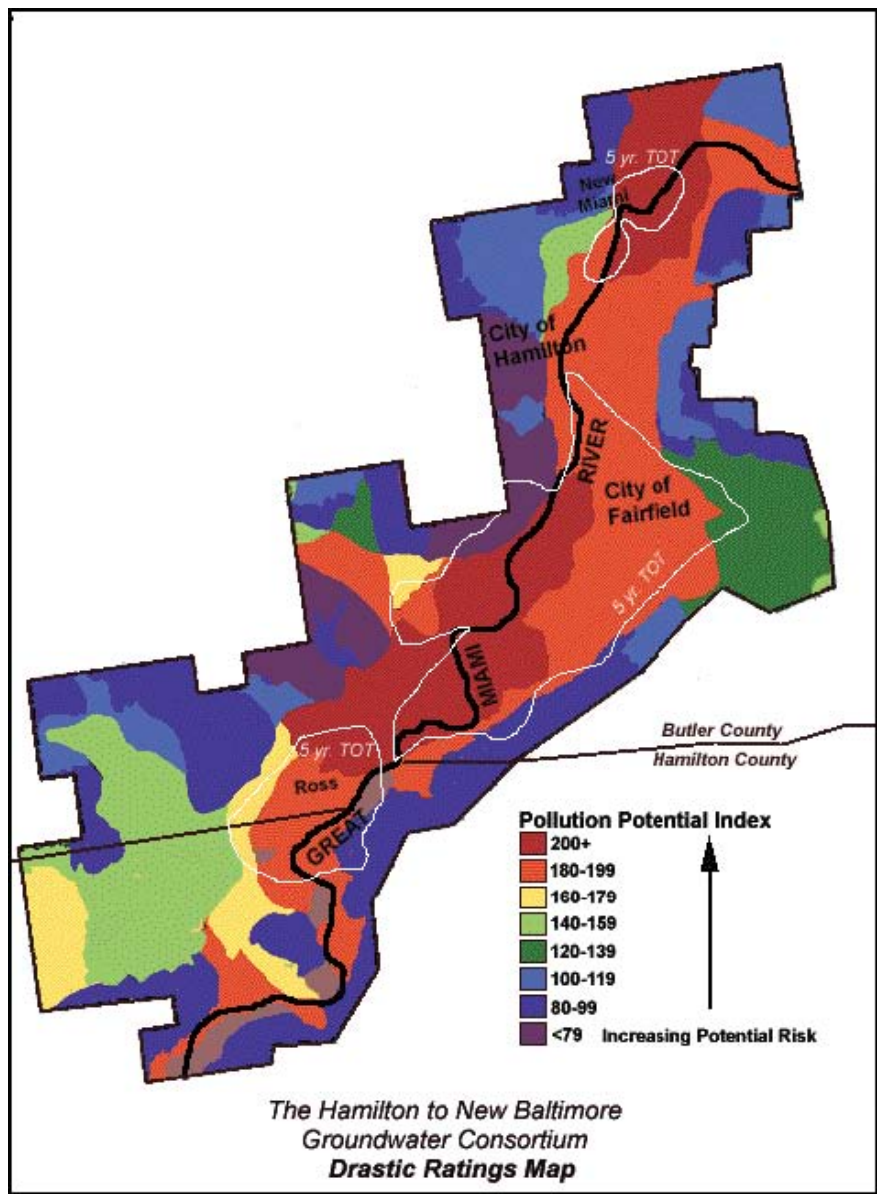


Figure 16.7 The results of using the DRASTIC groundwater vulnerability model in an area of Ohio, USA. The model combines GIS layers representing factors important in determining groundwater vulnerability, and displays the results as a map of vulnerability ratings (Reproduced by permission of Hamilton to New Baltimore Grand Water Consortium Pollution Potential (DRASTIC) map reproduced by permission of Ohio Department of Natural Resources)

16.2.2 Individual and aggregate models

The simulation model described in Box 16.2 works at the individual level, by attempting to forecast the behavior of each driver and vehicle in the study area. By contrast, it would clearly be impossible to model the behavior of every molecule in the Mammoth Cave watershed, and instead any modeling of groundwater movement must be done at an aggregate level, by predicting the movement of water as a continuous fluid. In general, models of physical systems are forced to adopt aggregate approaches because of the enormous number of individual objects involved, whereas it is much more feasible to model individuals in human systems, or in studies of

Applications Box **16.3**

Building a groundwater protection model in a Karst environment

Rhonda Pfaff (ESRI staff) and Alan Glennon (a doctoral candidate at the University of California, Santa Barbara) describe a simple but elegant application of modeling to the determination of groundwater vulnerability in Kentucky's Mammoth Cave watershed. Mammoth Cave is protected as a National Park, containing extensive and unique environments, but is subject to potentially damaging runoff from areas in the watershed outside park boundaries and therefore not subject to the same levels of environmental protection. Figure 16.8 shows a graphic rendering of the model in ESRI's ModelBuilder. Each operation is shown as a rectangle and each dataset as an ellipse. Reading from top left, the model first clips the slope layer to the watershed, then selects slopes greater than or equal to 5 degrees. A landuse layer is analyzed to select fields used for growing crops, and these are then combined with the steep_slopes layer to identify crop fields on steep slopes. A dataset of streams is buffered to 300 m, and finally this is combined to form a layer identifying all areas that are crop fields, on steep slopes, within 300 m of a stream. Such areas are particularly likely to experience soil erosion and to generate runoff contaminated with agricultural chemicals, which will then impact the downstream cave environment with its endangered populations of sightless fish. Figure 16.9 shows the resulting map.

A detailed description of this application is available at **www.esri.com/news/arcuser/0704/files/modelbuilder.pdf**, and the datasets are available at **www.esri.com/news/arcuser/0704/summer2004.html**.

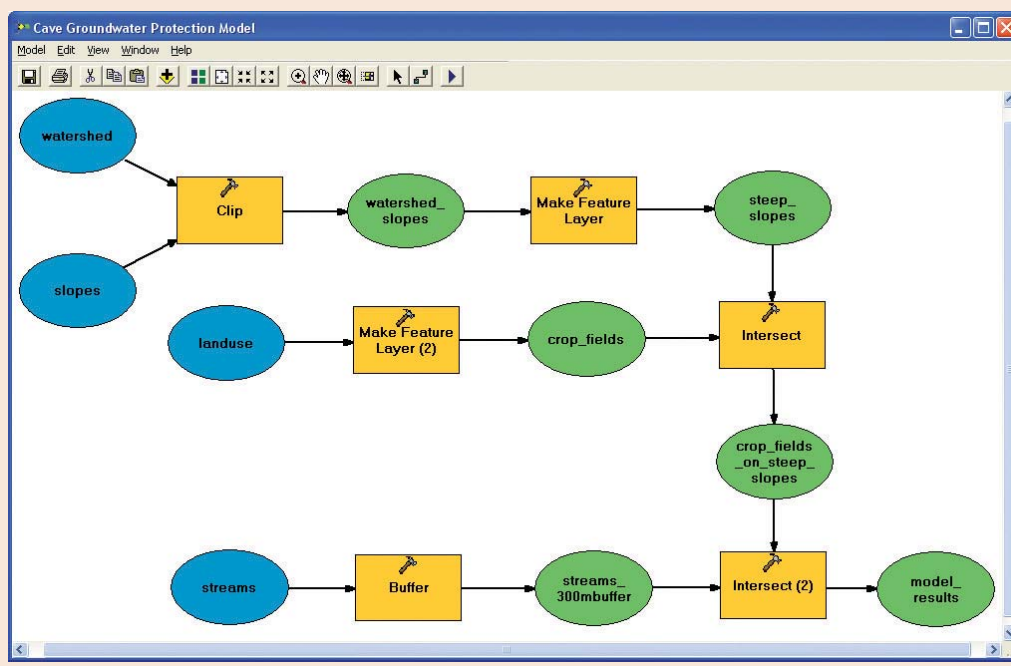


Figure 16.8 Graphic representation of the groundwater protection model developed by Rhonda Pfaff and Alan Glennon for analysis of groundwater vulnerability in the Mammoth Cave watershed, Kentucky, USA

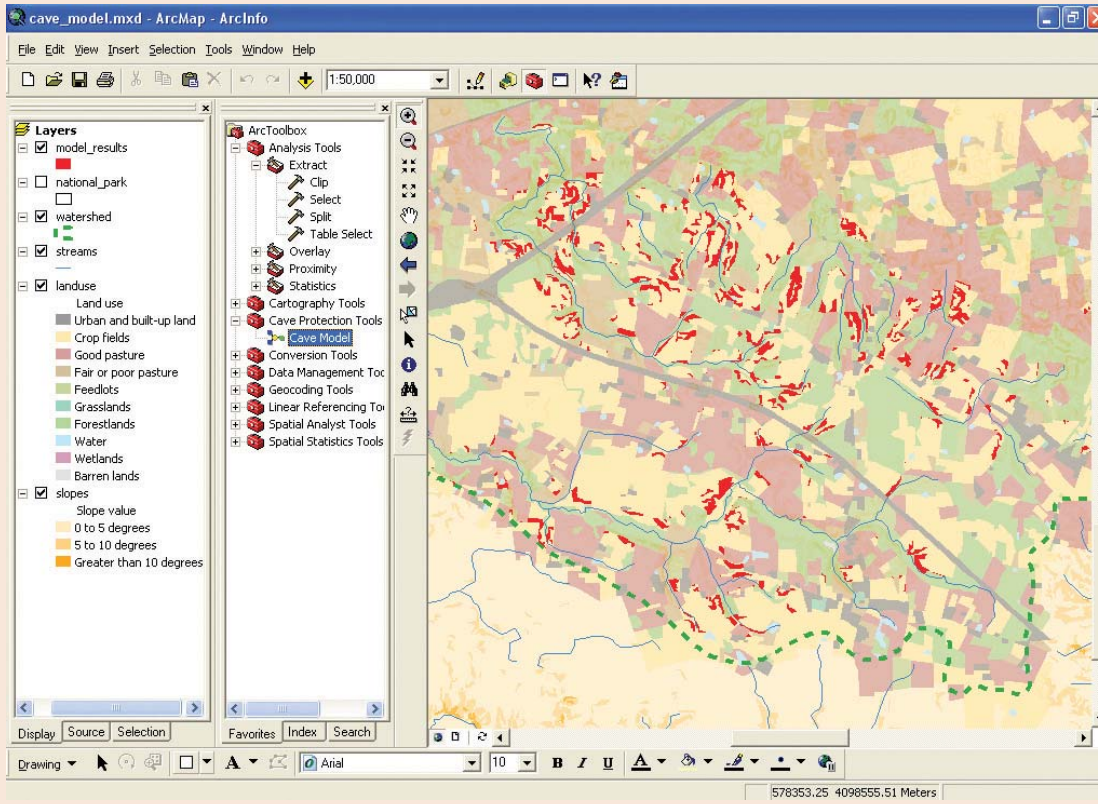


Figure 16.9 Results of the groundwater protection model. Highlighted areas are farmed for crops, on relatively steep slopes and within 300 m of streams. Such areas are particularly likely to generate runoff contaminated by agricultural chemicals and soil erosion and to impact adversely the cave environment into which the area drains

animal behavior. Even when modeling the movement of water as a continuous fluid it is still necessary to break the continuum into discrete pieces, as it is in the representation of continuous fields (see Figure 3.11). Some models adopt a raster approach and are commonly called *cellular* models (see Section 16.2.3). Other models break the world into irregular pieces or polygons, as in the case of the groundwater protection model described in Box 16.3.

Aggregate models are used when it is impossible to model the behavior of every individual element in a system.

Models of individuals are often termed *agent-based* models (ABM) or *autonomous agent* models (Box 16.4). With the massive computing power now available on the desktop, and with techniques of object-oriented programming, it is comparatively easy to build and execute ABMs, even when the number of agents is itself massive. Such models have been used to analyze and simulate many types of animal behavior, as well as the behavior of pedestrians in streets (Box 16.4), shoppers in stores, and drivers on congested roads (Box 16.2). They have also been used to model the behavior of decision makers, not through their movements in space but through the decisions they make regarding spatial features. For example, ABMs have been constructed to model the decisions made over land use in rural areas, by formulating the rules that govern individual decisions by landowners in response to varying market conditions, policies, and regulations. The impact of these factors on the fragmentation of the landscape, with its implications for wildlife habitat, are of key interest in such models.

16.2.3 Cellular models

Cellular models, which were introduced in Section 2.3.5.1, represent the surface of the Earth as a raster. Each cell in the fixed raster has a number of possible states, which change through time as a result of the application of transition rules. Typically the rules are defined over each cell's neighborhood, and determine the outcome of each stage in the simulation based on the cell's state, the states of its neighbors, and the values of cell attributes. The study of cellular models was first popularized by the work of John Conway, Professor of Mathematics at Princeton University, who studied the properties of a model he called the Game of Life (Box 16.5).

Technical Box **16.4**

Agent-based models of movement in crowded spaces

Working at the University College London Centre for Advanced Spatial Analysis, geographer and planner Michael Batty uses GIS to simulate the disasters and emergencies that can occur when large-scale events generate congestion and panic in crowds (Figure 16.10). The events involve large concentrations of people in small spaces that can arise because of accidents, terrorist attacks, or simply by the build-up of congestion through the convergence of large numbers of people into spaces with too little capacity. He has investigated scenarios for a number of major events, such as the movement of very large numbers of people to Mecca to celebrate the Hajj (a holy event in the Muslim calendar; Figure 16.10) and the Notting Hill Carnival (Europe's largest street festival, held annually in West Central London). His work uses agent-based models that take GIScience to a finer level of granularity and incorporate temporal processes as well as spatial structure. Fundamental to agent-based modeling of such situations is the need to understand how individuals in crowds interact with each other and with the geometry of the local environment.

Batty and his colleagues have been concerned with modeling the behavior of individual human agents under a range of crowding scenarios. This style of model is called agent-based because it depends on simulating the aggregate effect of the behaviors of individual objects or persons in an inductive way (bottom-up: Section 4.9). The modeling is based on averages – our objective is modeling an aggregate outcome (such as the onset of panic) although the particular behavior of each agent is a combination of responses which are based on how routine behavior is modified by the presence of other agents and by the particular environment in which the agent finds himself or herself. It is possible to simulate a diverse range of outcomes, each based on different styles of interaction between individual agents and different geometrical configurations of constraints.

Figure 16.11 shows one of Batty's simulations of the interactions between two crowds. Here, a parade is moving around a street intersection – the central portion of the movement (walkers in white) is the parade and the walkers around this in gray/red are the watchers. This model can be used to simulate the build-up of pressure through random motion which then generates a breakthrough of the watchers into the parade, an event that often leads to disasters of the kind experienced in festivals, rock concerts, and football matches as well as ritual situations like the Hajj.

Figure 16.10 Massive crowds congregate in Mecca during the annual Hajj. On February 1, 2004 244 pilgrims were trampled when panic stampeded the crowd. This was a unique but unfortunately not a freak occurrence: 50 pilgrims were killed in 2002, 35 in 2001, 107 in 1998, and 1425 were killed in a pedestrian tunnel in 1990

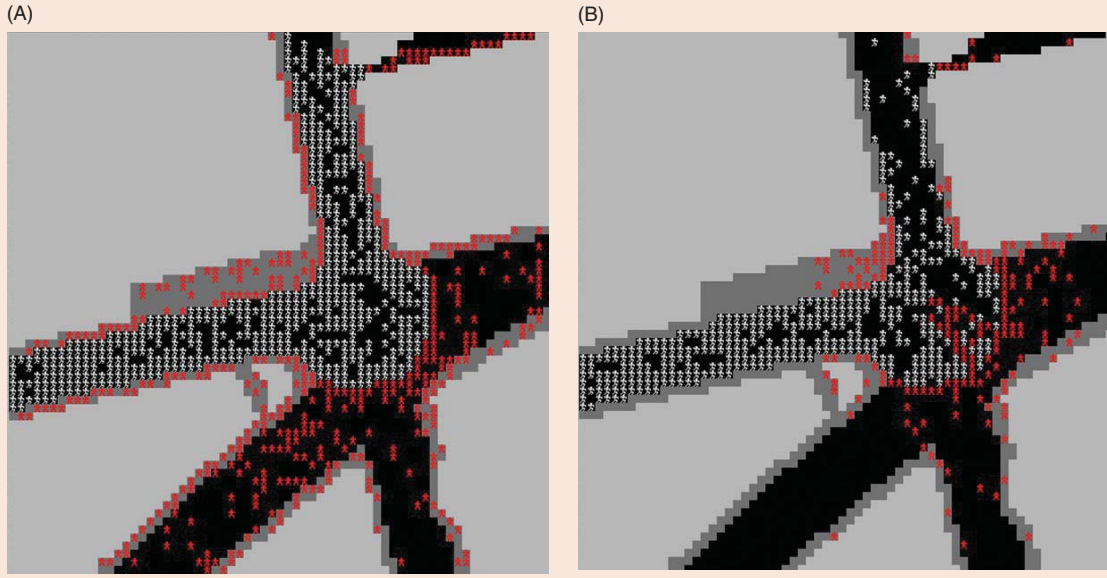


Figure 16.11 Simulation of the movement of individuals during a parade. Parade walkers are in white, watchers in red. The watchers (A) build up pressure on restraining barriers and crowd control personnel, and (B) break through into the parade (Reproduced by permission of Michael Batty)

Cellular models represent the surface of the Earth as a raster, each cell having a number of states that are changed at each iteration by the execution of rules.

Several interesting applications of cellular methods have been identified, and particularly notable are the efforts to apply them to urban growth simulation. The likelihood of a parcel of land developing depends on many factors, including its slope, access to transportation routes, status in zoning or conservation plans, but above all its proximity to other development. These models express this last factor as a simple modification of the rules of the Game of Life – the more developed the state of neighboring cells, the more likely a cell is to make the transition from undeveloped to developed. Figure 16.13 shows an illustration from one such model, that developed by Keith Clarke and his co-workers at the University of California, Santa Barbara, to predict growth patterns in the Santa Barbara area through 2040. The model iterates on an annual basis, and the effects of neighboring states in promoting infilling are clearly evident. The inputs for the model – the transportation network, the topography, and other factors – are readily

Technical Box **16.5**

The Game of Life

The game is played on a raster. Each cell has two states, *live* and *dead*, and there are no additional cell attributes (no variables to differentiate the space), as there would be in GIS applications. Each cell has eight neighbors (the Queen's case, Figure 15.19). There are three transition rules at each time-step in the game:

1. A dead cell with exactly three live neighbors becomes a live cell;
2. A live cell with two or three live neighbors stays alive; and
3. In all other cases a cell dies or remains dead.

With these simple rules it is possible to produce an amazing array of patterns and movements, and some surprisingly simple and elegant patterns emerge out of the chaos (in the field of agent-based modeling these unexpected and simple patterns are known as *emergent properties*). The page **www.math.com/students/wonders/life/life.html** includes extensive details on the game, examples of particularly interesting patterns, and executable Java code. Figure 16.12 shows three stages in an execution of the Game of Life.

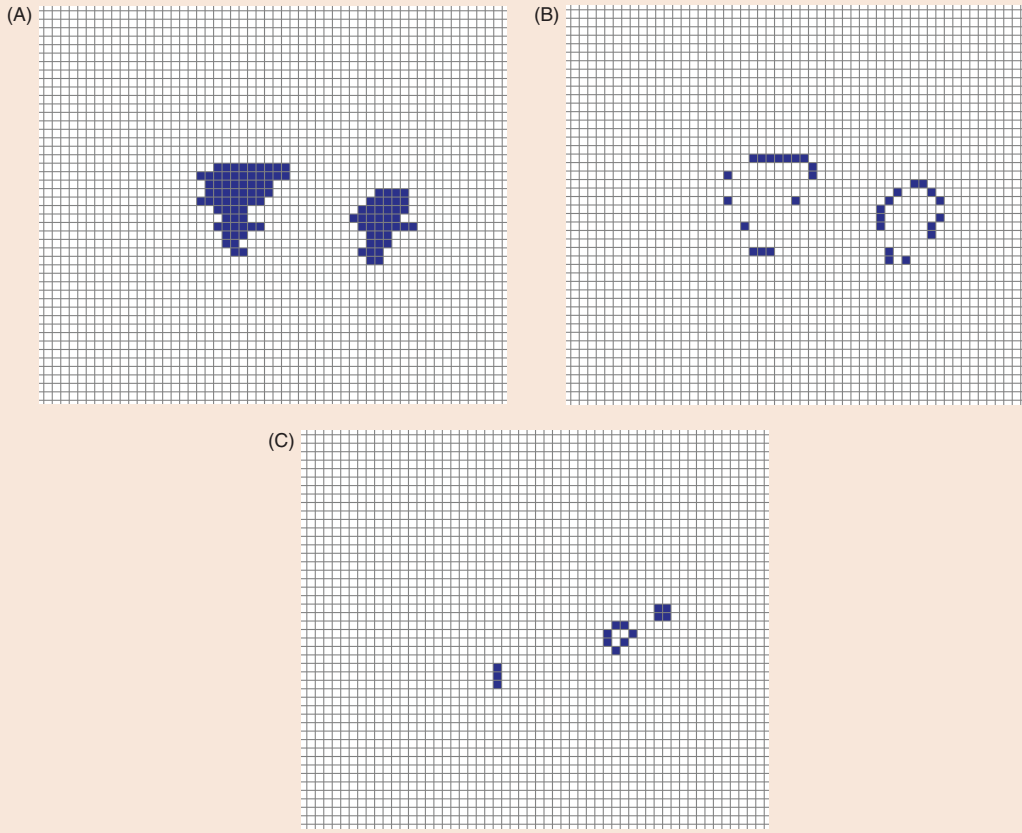


Figure 16.12 Three stages in an execution of the Game of Life: (A) the starting configuration, (B) the pattern after one time-step, and (C) the pattern after 14 time-steps. At this point all features in the pattern remain stable

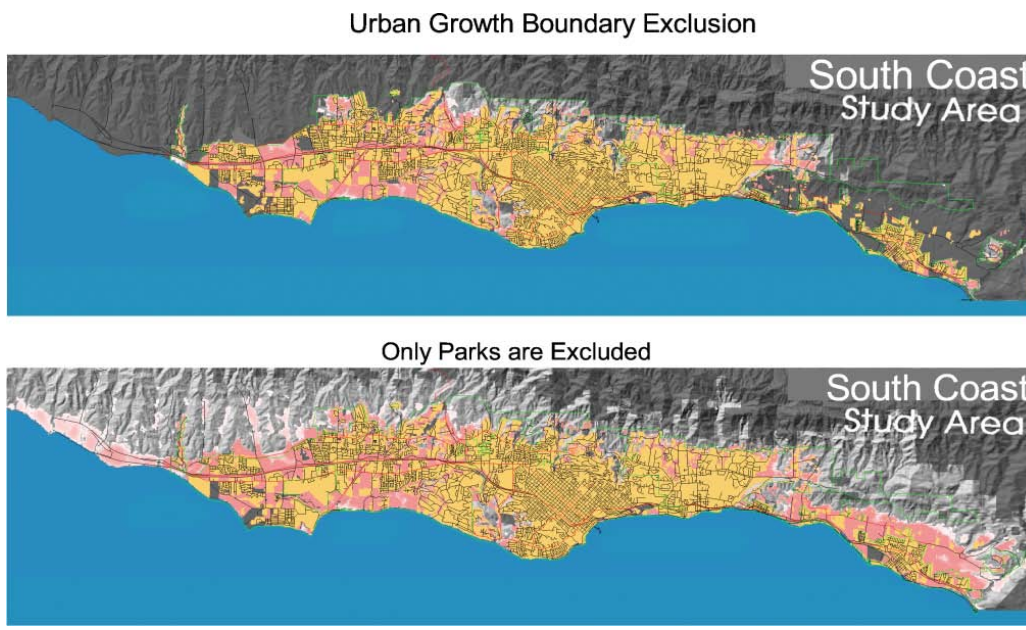


Figure 16.13 Simulation of future urban growth patterns in Santa Barbara, California, USA. (Upper) growth limited by current urban growth boundary; (Lower) growth limited only by existing parks (Courtesy: Keith Clarke)

prepared in GIS, and the GIS is used to output the results of the simulation.

One of the most important issues in such modeling is calibration and validation – how do we know the rules are right, and how do we confirm the accuracy of the results? Clarke's group has calibrated the model by *brute force*, by running the model forwards from some point in the past under different sets of rules and comparing the results to the actual development history from then to the present. This method is extremely time-consuming since the model has to be run under vast numbers of combinations of rules, but it provides at least some degree of confidence in the results. The issue of accuracy is addressed in more detail, and with reference to modeling in general, later in this chapter.

A model must be calibrated against real data to determine appropriate values for its parameters and to ensure that its rules are valid.

16.2.4 Cartographic modeling and map algebra

In essence, modeling consists of combining many stages of transformation and manipulation into a single whole, for a single purpose. In the example in Box 16.3 the number of stages was quite small – only six – whereas in the Clarke urban growth model several stages are executed many times as the model iterates through its annual steps.

The individual stages can consist of a vast number of options, encompassing all of the basic transformations of which GIS is capable. In Section 14.1.2 it was argued that GIS manipulations could be organized into six categories, depending on the broad conceptual objectives of each manipulation. But the six are certainly not the only way of classifying and organizing the numerous methods of spatial analysis into a simple scheme. Perhaps the most successful is the one developed by Dana Tomlin and known as *cartographic modeling* or *map algebra*. It classifies all GIS transformations of rasters into four basic classes, and is used in several raster-centric GISs as the basis for their analysis languages:

- *local* operations examine rasters cell by cell, comparing the value in each cell in one layer with the values in the same cell in other layers;
- *focal* operations compare the value in each cell with the values in its neighboring cells – most often eight neighbors;
- *global* operations produce results that are true of the entire layer, such as its mean value;
- *zonal* operations compute results for blocks of contiguous cells that share the same value, such as the calculation of shape for contiguous areas of the same land use, and attach their results to all of the cells in each contiguous block.

With this simple schema, it is possible to express any model, such as Clarke's urban growth simulation, as a *script* in a well-defined language. The only constraint is that the model inputs and outputs be in raster form.

Map algebra provides a simple language in which to express a model as a script.

A more elaborate map algebra has been devised and implemented in the PCRaster package, developed for spatial modeling at the University of Utrecht (**pcraster.geog.uu.nl**). In this language, a symbol refers to an entire map layer, so the command $A = B + C$ takes the values in each cell of layers B and C, adds them, and stores the result as layer A.

16.3 Technology for modeling

16.3.1 Operationalizing models in GIS

Many of the ideas needed to implement models at a practical level have already been introduced. In this section they are organized more coherently in a review of the technical basis for modeling.

Models can be defined as sequences of operations, and we have already seen how such sequences can be expressed either as graphic flowcharts or as scripts. One of the oldest graphic platforms for modeling is Stella, now very widely distributed and popular as a means of conceptualizing and implementing models. Although it now has good links to GIS, its approach is conceptually non-spatial, so it will not be described in detail here. Instead, the focus will be on similar environments that are directly integrated with GIS. The first of these may have been ERDAS IMAGINE software (**www.gis.leica-geosystems.com/Products/Imagine/**), which allows the user to build complex modeling sequences from primitive operations, with a focus on the manipulations needed in image processing and remote sensing. ESRI introduced a graphic interface to modeling in ArcView 3.x Spatial Modeler, and enhanced it in ArcGIS 9.0 ModelBuilder (see Figure 16.8).

Any model can be expressed as a script, or visually as a flowchart.

Typically in these interfaces, datasets are represented as ellipses, operations as rectangles, and the sequence of the model as arrows. The user is able to modify and control the operation sequence by interacting directly with the graphic display.

Fully equivalent, but less visual and therefore more demanding on the user, are *scripts*, which express models as sequences of commands, allowing the user to execute an entire sequence by simply invoking a script. Initial efforts to introduce scripting to GIS were cumbersome, requiring the user to learn a product-specific language such as ESRI's AML (Arc Macro Language) or Avenue.

Today, it is common for GIS model scripts to be written in such industry-standard languages as Visual Basic for Applications (VBA), Perl, JScript, or Python. Scripts can be used to execute GIS operations, request input from the user, display results, or invoke other software. With interoperability standards such as Microsoft's .Net (Box 7.3), it is possible to call operations from any compliant package from the same script. For example, a script might combine GIS operations with calls to Excel for simple tabular functions, or even calls to another compliant GIS (Section 7.6).

With standards like Microsoft's .Net it is possible for a script to call operations from several distinct software packages.

16.3.2 Model coupling

The previous section described the implementation of models as direct extensions of an underlying GIS, through either graphic model-building or scripts. This approach makes two assumptions: first, that all of the operations needed by the model are available in the GIS (or in another package that can be called by the model); and second, that the GIS provides sufficient performance to handle the execution of the model. In practice, a GIS will often fail to provide adequate performance, especially with very large datasets and large numbers of iterations, because it has not been designed as a modeling engine. Instead, the user is forced to resort to specialized code, written in a source language such as C. Clarke's model, for example, is programmed in C, and the GIS is used only for preparing input and visualizing output. Other models may be spatial only in the sense of having geographically differentiated inputs and outputs, as discussed in the case of the USLE in Section 16.2.1, making the use of a GIS optional rather than essential.

In reality, therefore, much spatial modeling is done by *coupling* GIS with other software. A model is said to be *loosely coupled* to a GIS when it is run as a separate piece of software and data are exchanged to and from the GIS in the form of files. Since many GIS formats are proprietary, it is common for the exchange files to be in export format, and to rely on translators at either end of the transfer. In extreme cases it may be necessary to write a special program to convert the files during transfer, when no common format is available. Clarke's is an example of a model that is loosely coupled to GIS. A model is said to be *closely coupled* to a GIS when both the model and the GIS read and write the same files, obviating any need for translation. When the model is executed as a GIS script or through a graphic GIS interface it is said to be *embedded*.

16.3.3 Cataloging and sharing models

In the digital computer everything – the program, the data, the metadata – must be expressed in the same language of 0s and 1s. The value of each bit or byte (Box 3.1) depends on its meaning, and clearly some bits are more valuable than others. A bit in Microsoft's operating system Windows XP is clearly extremely valuable, even though there are hundreds of millions of them. On the other hand, a bit in a remote sensing image that happens to have been taken on a cloudy day has almost no value at all, except perhaps to someone studying clouds. In general, one would expect bits of programs to be more valuable than bits of data, especially when those programs are useable by large numbers of people.

This line of argument suggests that the GIS scripts, models, and other representations of process are potentially valuable to many users and well worth sharing. But almost all of the investment society has made in the sharing of digital information has been devoted to data and text. In Section 11.2 we saw how geolibraries, data warehouses, and metadata standards have been devised and implemented to facilitate the sharing of geographic data, and parallel efforts by libraries, publishers, and builders of search engines have made it easy nowadays to share text via the Web. But very little effort to date has gone into making it possible to share *process objects*, digital representations of the process of GIS use.

A process object, such as a script, captures the process of GIS use in digital form.

Notable exceptions include ESRI's ArcScripts, a library of GIS scripts contributed by users and maintained by the vendor as a service to its customers. The library is easily searched for scripts to perform specific types of analysis and modeling. At the time of writing, it included 3286 scripts, ranging from one to couple ArcView 3.x to the StormWater Management Model (SWMM) written in Avenue, to a set of basic tools for crime analysis and mapping. ArcGIS 9.0 includes the ability to save, catalog, and share scripts as GIS functions.

The term *GIService* is often used to describe a GIS function being offered by a server for use by any user connected to the Internet (see Sections 1.5.3 and 7.2). In essence, GIServices offer an alternative to locally installed GIS, allowing the user to send a request to a remote site instead of to his or her GIS. GIServices are particularly effective in the following circumstances:

- When it would be too complex, costly, or time-consuming to install and run the service locally;
- When the service relies on up-to-date data that would be too difficult or costly for the user to constantly update;
- When the person offering the service wishes to maintain control over his or her intellectual property. This would be important, for example, when the function is based on a substantial amount of knowledge and expertise.

To date, several standard GIS functions are available as GIServices, satisfying one or more of the above requirements. The Alexandria Digital Library offers a *gazetteer service* (**www.alexandria.ucsb.edu**), allowing a user to send a placename and to receive in return one or more coordinates of places that match the placename. Several companies offer *geocoding services*,

returning coordinates in response to a street address (try **www.geocode.com**). The Geography Network (**www.geographynetwork.com**) provides a searchable catalog to GIServices.

16.4 Multicriteria methods

The model developed by Pfaff and Glennon (depicted in Figure 16.8) rates vulnerability to runoff based on three factors – cropland, slope, and distance from stream – but treats all three as simple binary measures, and produces a simple binary result (land is vulnerable if slope is greater than 5%, land use is cropping, and distance from stream is less than 300 m). In reality, of course, 5% is not a precise cutoff between non-vulnerable and vulnerable slopes, and neither is 300 m a precise cutoff between vulnerable distances and non-vulnerable distances. The issues surrounding rules such as these, and their fuzzy alternatives, were discussed at length in Section 6.2.

A more general and powerful conceptual framework for models like this would be constructed as follows. There are a number of factors that influence vulnerability, denoted by x_1 through x_n. The impact of each factor on vulnerability is determined by a transformation of the factor $f(x)$ – for example, the factor *distance* would be transformed so that its impact *decreases* with increasing distance (as in Section 4.5), whereas the impact of *slope* would be *increasing*. Then the combined impact of all of the factors is obtained by weighting and adding them, each factor i having a weight w_i:

$$I = \sum_{i=1}^{n} w_i f(x_i)$$

In this framework both the functions f and the weights w need to be determined. In the example, the f for slope was resolved by a simple step function, but it seems more likely that the function should be continuously decreasing, as shown in Figure 16.14. The hill-shaped function also shown in the figure would be appropriate in cases where the impact of a factor declines in both directions from some peak value (for example, smoke from a tall smokestack has its greatest impact at some distance downwind).

Many decisions depend on identifying relevant factors and adding their appropriately weighted values.

This approach provides a good conceptual framework both for the indicator models typified by Box 16.3 and for many models of design processes. In both cases, it is possible that multiple views might exist about appropriate functions and weights, particularly when modeling a decision over an important development with impact on the environment. Different *stakeholders* in the design process can be anticipated to have different views

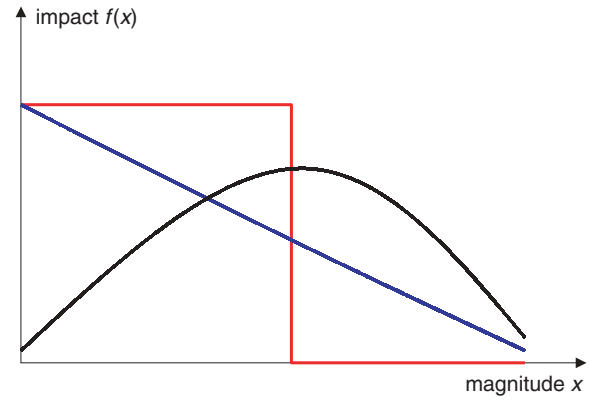


Figure 16.14 Three possible impact functions: (red) the step function used to assess slope in Figure 16.8; (blue) a decreasing linear function; and (black) a function showing impact rising to a maximum and then decreasing

about what is important, how that importance should be measured, and how the various important factors should be combined. Such processes are termed *multicriteria decision making* or MCDM, and are commonly encountered whenever decisions are controversial. Some stakeholders may feel that environmental factors deserve high weight, others that cost factors are the most important, and still others that impact on the social life of communities is all-important.

An important maxim of MCDM is that it is better for stakeholders to argue in principle about the merits of different factors and how their impacts should be measured, than to argue in practice about alternative decisions. For example, arguing about whether slope is a more important factor than distance, and about how each should be measured, is better than arguing about the relative merits of Solution A, which might color half of John Smith's field red, over Solution B, which might color all of David Taylor's field red. Ideally, all of the controversy should be over once the factors, functions, and weights are decided, and the solution they produce should be acceptable to all, since all accepted the inputs. Would it were so!

Each stakeholder in a decision may have his or her own assessment of the importance of each relevant factor.

Although many GIS have implemented various approaches to MCDM, Clark University's IDRISI offers the most extensive functionality as well as detailed tutorials and examples (**www.clarklabs.org**; see Box 16.6). A common framework is the Analytical Hierarchy Process (AHP) devised by Thomas Saaty, which focuses on capturing each stakeholder's view of the appropriate weights to give to each impact factor. First, the impact of each factor is expressed as a function, choosing from options such as those shown in Figure 16.14. Then each stakeholder is asked to compare each pair of factors (with n factors there are $n(n-1)/2$ pairs), and to assess their relative importance in ratio form. A matrix is

Table 16.1 An example of the weights assigned to three factors by one stakeholder. For example, the entry '7' in Row 1 Column 2 (and the 1/7 in Row 2 Column 1) indicates that the stakeholder felt that Factor 1 (slope) is seven times as important as Factor 2 (land use)

	Slope	Land use	Distance from stream
Slope		7	2
Land use	1/7		1/3
Distance from stream	1/2	3	

created for each stakeholder, as in the example shown in Table 16.1. The matrices are then combined and analyzed, and a single set of weights extracted that represent a consensus view (Figure 16.15). These weights would then be inserted as parameters in the spatial model to produce a final result. The mathematical details of the method can be found in books or tutorials on the AHP.

16.5 Accuracy and validity: testing the model

Models are complex structures and their outputs are often forecasts of the future. How, then, can a model be tested, and how does one know whether its results can be trusted? Unfortunately, there is an inherent tendency to trust the results of computer models because they appear in numerical form (and numbers carry innate authority) and because they come from computers (which also appear authoritative).

Results from computers can seem to carry innate authority.

Normally, scientists test their results against reality, but in the case of forecasts reality is not yet available, and by the time future events have played out there is likely little interest in testing. So modelers must resort to other methods to verify and validate their predictions.

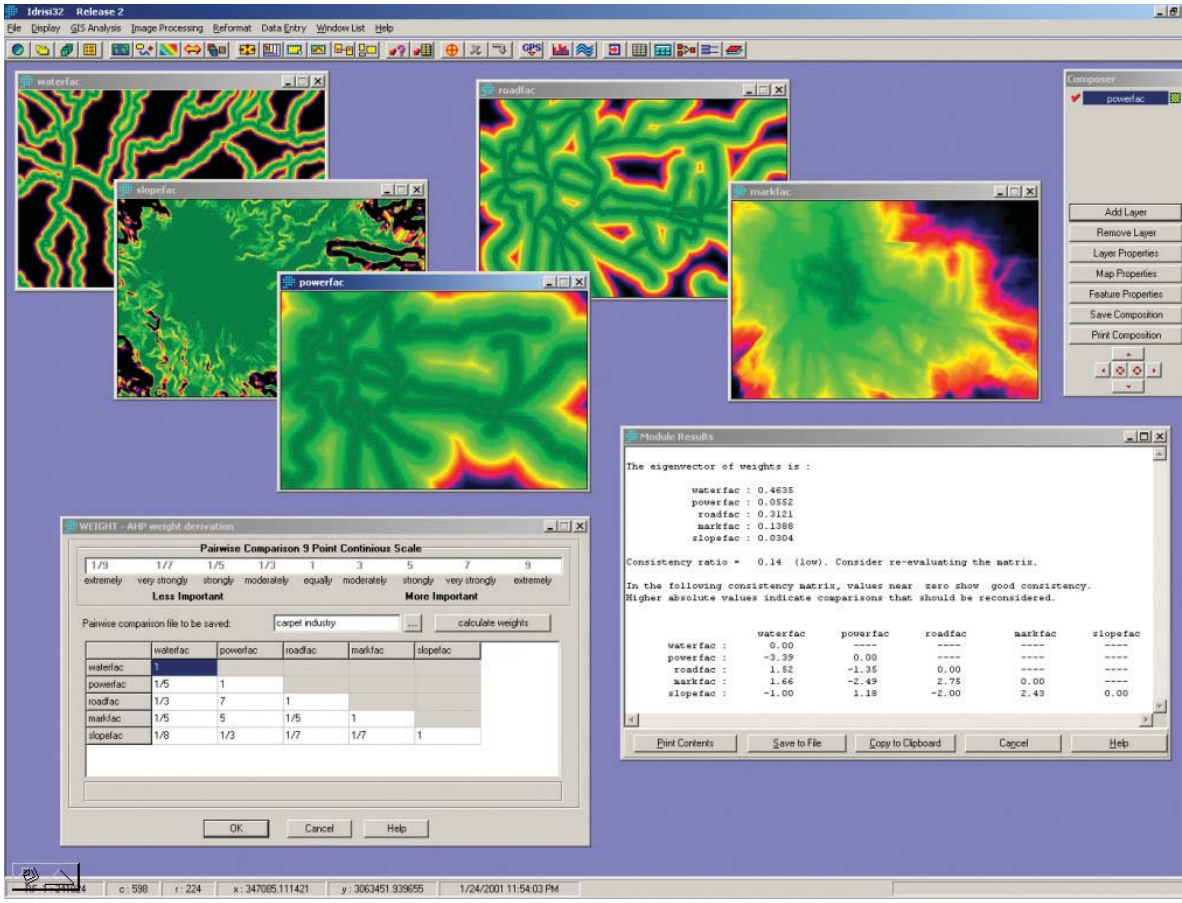


Figure 16.15 Screen shot of an AHP application using IDRISI (**www.clarklabs.org**). The five layers in the upper left part of the screen represent five factors important to the decision. In the lower left, the image shows the table of relative weights compiled by one stakeholder. All of the weights matrices are combined and analyzed to obtain the consensus weights shown in the lower right, together with measures to evaluate consistency among the stakeholders (Reproduced by permission of J. Ronald Eastman)

Biographical Box **16.6**

Ron Eastman, developer of the IDRISI GIS

Figure 16.16 Ron Eastman, developer of IDRISI

J. Ronald Eastman (Figure 16.16) is Director and lead software engineer of Clark Labs, a non-profit center within the George Perkins Marsh Institute of Clark University that produces the IDRISI GIS and image processing system, non-profit software with over 35 000 registered users worldwide. He is also Professor of Geography and former Director of the Graduate School of Geography at Clark University. He gained his Ph.D. in Geography from Boston University in 1982. His research is focused most strongly on the development of analytical procedures for the investigation and modeling of land-cover change and effective resource allocation in environmentally related decision making. He also has extensive experience in the deployment of GIS and remote sensing in the developing world, and is a Senior Special Fellow of the United Nations Institute for Training and Research. He writes: 'With a degree in perceptual psychology and a strong love of travel and the diversity of landscapes and cultures, I became interested in the power of the visual system in facilitating the analysis of geographic phenomena through the medium of maps. My graduate degrees were thus in geography/cartography. However, I had the fortune of witnessing and participating in the birth of GIS and digital image processing. It was immediately clear to me that the analytical potential of this new medium far outstripped what could be achieved by visual analysis alone, particularly in the study of our rapidly changing environment. At the same time, I was also concerned that many in the developing world had only limited access to computing resources and GIS/image processing technology – thus my motivation in developing a system focused on the analytical needs of development and global change.'

Models can often be tested by comparison with past history, by running the model not into the future, but forwards in time from some previous point. But these are often the data used to *calibrate* the model, to determine its parameters and rules, so the same data are not available for testing. Instead, many modelers resort to *cross-validation*, a process in which a subset of data is used for calibration, and the remainder for validating results. Cross-validation can be done by separating the data into two time periods, or into two areas, using one for calibration and one for validation. Both are potentially dangerous if the process being modeled is also changing through time, or across space (it is *non-stationary* in a statistical sense), but forecasting is dangerous in these circumstances as well.

Models of real-world processes can be validated by experiment, by proving that each component in the model correctly reflects reality. For example, the Game of Life would be an accurate model if some real process actually behaved according to its rules; and Clarke's urban growth model could be tested in the same way, by examining the rules that account for each past land use transition. In reality, it is unlikely that real-world processes will be found to behave as simply as model rules, and it is also unlikely that the model will capture every real-world process impacting the system.

If models are no better than approximations to reality, then are they of any value? Certainly human society is so complex that no model will ever fit perfectly. As Ernest Rutherford, the experimental physicist and Nobel Laureate is said to have once remarked, probably in frustration with social-scientist colleagues, 'The only result that can possibly be obtained in the social sciences is, some do, and some don't'. Neither will a model of a physical system ever perfectly replicate reality. Instead, the outputs of models must always be taken advisedly, bearing in mind several important arguments:

- A model may reflect behavior under ideal circumstances and therefore provide a norm against which to compare reality. For example, many economic models assume a perfectly informed, rational decision maker. Although humans rarely behave this way, it is still useful to know what would happen if they did, as a basis for comparison.
- A model should not be measured by how closely its results match reality but by how much it reduces uncertainty about the future. If a model can narrow the options, then it is useful. It follows that any forecast should also be accompanied by a realistic measure of uncertainty.
- A model is a mechanism for assembling knowledge from a range of sources and presenting conclusions based on that knowledge in readily used form. It is often not so much a way of discovering how the world works, as a way of presenting existing knowledge in a form helpful to decision makers.

- Just as the British politician Denis Healey once remarked that capitalism was the worst possible economic system 'apart from all of the others', so modeling often offers the only robust, transparent analytical framework that is likely to garner any respect amongst decision makers with competing objectives and interests.

Any model forecast should be accompanied by a realistic measure of uncertainty.

Several forms of uncertainty are associated with models, and it is important to distinguish between them. First, models are subject to the uncertainty present in their inputs. Uncertainty *propagation* was discussed in Section 6.4.2 – here, it refers to the impacts of uncertainty in the inputs of a model on uncertainty in the outputs. In some cases propagation may be such that an error or uncertainty in an input produces a proportionate error in an output. In other cases, a very small error in an input may produce a massive change in output; and in other cases outputs can be relatively insensitive to errors in inputs. It is important to know which of these cases holds in a given instance, and the normal approach is through repeated numerical simulation (often called Monte Carlo simulation because it mimics the random processes of that famous gambling casino), adding random distortions to inputs and observing their impacts on outputs. In some limited instances it may even be possible to determine the effects of propagation mathematically.

Second, models are subject to uncertainty over their parameters. A model builder will do the best possible job of calibrating the model, but inevitably there will be uncertainty about the correct values, and no model will fit the data available for calibration perfectly. It is important, therefore, that model users conduct some form of *sensitivity analysis*, examining each parameter in turn to see how much influence it has on the results. This can be done by raising and lowering each parameter value, for example by +10% and −10%, re-running the model and comparing the results. If changing a parameter by 10% produces a less-than-10% change in results the model can be said to be relatively insensitive to that parameter. This allows the modeler to focus attention on those parameters that produce the greatest impact on results, making every effort to ensure that their values are correct.

Sensitivity analysis tests a model's response to changes in its parameters and assumptions.

Third, and most importantly, models are subject to uncertainty because of the *labeling* of their results. Consider the model of Box 16.3, which computes an indicator of the need for groundwater protection. In truth, the areas identified by the model have three characteristics – slope greater than 5%, used for crops, and less than 300 m from a stream. Described this way, there is little uncertainty about the results, though there will be some uncertainty due to inaccuracies in the data. But once the areas selected are described as *vulnerable*, and in need of management to ensure that groundwater is *protected*, a substantial leap of logic has occurred. Whether that leap is valid or not will depend on the reputation of the modeler as an expert in groundwater hydrology and biological conservation; on the reputation of the organization sponsoring the work; and on the background science that led to the choice of parameters (5% and 300 m). In essence, this third type of uncertainty, which arises whenever labels are attached to results that may or may not correctly reflect their meaning, is related to how results are described, in other words to their metadata, rather than to any innate characteristic of the data themselves.

16.6 Conclusion

Modeling was defined at the outset of this chapter as a process involving multiple stages, often in emulation of some real physical process. Modeling is often dynamic, and current interest in modeling is stretching the capabilities of GIS software, most of which was designed for the comparatively leisurely process of analysis, rather than the intensive and rapid iterations of a dynamic model. In many ways, then, modeling represents the cutting edge of GIS, and the next few years are likely to see explosive growth both in interest in modeling on the part of GIS users, and software development on the part of GIS vendors. The results are certain to be interesting.

Questions for further study

1. Write down the set of rules that you would use to implement a simple version of the Clarke urban growth model, using the following layers: transportation (cell contains road or not), protected (cell is not available for development), already developed, and slope.
2. Review the steps you would take and the arguments you would use to justify the validity of a GIS-based model.
3. Select a field of environmental or social science, such as species habitat prediction or residential segregation. Search the Web and your library for published models of processes in this field and summarize their conceptual structure, technical implementation, and application history.

4. Compare the relative advantages of the different types of dynamic models discussed in this chapter when applied to a selected area of environmental or social science.

Further reading

Goodchild M.F., Parks B.O., and Steyaert, L.T. (eds) 1993 *Environmental Modeling with GIS*. New York: Oxford University Press.

Heuvelink G.B.M. 1998 *Error Propagation in Environmental Modelling with GIS*. London: Taylor and Francis.

Saaty T.L. 1980 *The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.

Tomlin C.D. 1990 *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs, NJ: Prentice Hall.

V

Management and Policy

17 *Managing GIS*

18 *GIS and management, the Knowledge Economy, and information*

19 *Exploiting GIS assets and navigating constraints*

20 *GIS partnerships*

17 *Managing GIS*

Much of the material in earlier parts of this book assumes that you have a GIS and can use it effectively to meet your organization's goals. This chapter addresses the essential 'front end': the procurement and operational management of such systems. Later chapters deal with how GIS contributes to the business.

This chapter describes how to choose, implement, and manage operational GIS. It involves four key stages: the analysis of needs, the formal specification, the evaluation of alternatives, and the implementation of the chosen system. In particular, implementing GIS requires consideration of issues such as planning, support, communication, resource management, and funding. Successful on-going management of an operational GIS has five key dimensions: support for customers, operations, data management, application development, and project management.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter, you will understand:

- How to go about choosing a GIS to meet your needs;
- Key GIS implementation issues;
- How to manage an operational GIS effectively with limited resources and ambitious goals;
- Why GIS projects fail – some pitfalls to avoid and some useful tips about how to succeed;
- The roles of staff members in a GIS project;
- Where to go for more detailed advice.

17.1 The big picture

This chapter is concerned with the practical aspects of managing an operational GIS. It is embedded deliberately in the part of the book focused on high-level management concepts: success comes from combining strategy and implementation. It is the role of management in GIS projects to ensure that operations are carried out effectively and efficiently, and that a healthy, sustainable GIS can be maintained – one which meets the organization's strategic objectives.

Obtaining and running a GIS seems at first sight to be a routine and apparently 'mechanical' process. It is certainly not 'rocket science'. But neither is it simple. The consequences of failure can be catastrophic, both for the organization and for careers. Success involves constant sharing of experience and knowledge with other people, keeping good records, and making numerous judgments where the answer is not pre-ordained.

Clearly we cannot deal with all the relevant aspects of managing GIS in one chapter. This, then, is a summary and a pointer to more detailed information for those who need it. Perhaps the best 'whole book' general overview of the process of arriving at the running of a successful GIS has been produced by Roger Tomlinson, based on 40 years of experience in building and consulting on GIS for organizations across the world (see Box 17.1). He sees the initial stages of the process as having ten basic steps, encapsulated in Box 17.2.

GIS as we now recognize it began during the spring of 1962.

However, before actually starting on the process of acquiring and implementing a GIS, ask the fundamental question: *do I really need a GIS?* There are many applications where the answer is obvious – as shown in

Biographical Box **17.1**

Roger Tomlinson, GIS pioneer

GIS had many roots, but if one person can be said to be its father, it is Roger Tomlinson (see Figure 17.1: see also Section 1.4.1 and Table 1.4). Roger pioneered the use of GIS worldwide to collect, manage, and manipulate geographic data, changing the face of geography as a discipline.

GIS, as we now recognize it, could be said to have begun during the spring of 1962. While on a plane bound from Ottawa to Toronto, Tomlinson met Lee Pratt, then recently named head of the Canada Land Inventory (CLI). Tomlinson was chief of the computer mapping division at Spartan Air Services, Ottawa. The two men discussed a vast mapping project CLI was about to undertake: a multilayer land-use/planning map of Canada's inhabited and productive land – around 1 million square miles. Tomlinson told Pratt that he used computers for mapping projects and some of his ideas might work for CLI. Pratt went home and did the arithmetic, then called Tomlinson. He said: 'We better talk about this, because we've tested out how to do it manually and it's far too expensive', Tomlinson recalls. Shortly afterwards, he was employed by the Canadian government, heading its GIS development program, where he was instrumental in developing the Canada Geographic Information System or CGIS. Many other developments worldwide drew inspiration from this beginning.

Roger Tomlinson was born in Cambridge, England, in 1933. He flew planes in the Royal Air Force during the early 1950s and twice led expeditions to the Norwegian Ice Cap (1956 and 1957). He holds BAs from Nottingham University, England, and Acadia University, Nova Scotia, Canada; an MA from McGill University, Quebec, Canada; and a Ph.D. from University College, London, England. Tomlinson adopted Canada as his home country in 1957.

Dr Tomlinson established Tomlinson Associates Ltd, a firm of consulting geographers, in 1977. As a geographic consultant, Tomlinson has advised an extraordinary list of clients including the World Bank, United Nations Food and Agriculture Organization, the US departments of Commerce and Agriculture, US Geological Survey, US Forest Service, US Bureau of the Census, the Canadian Forest Service, and numerous US state and Canadian provincial and municipal government agencies – as well as many other bodies worldwide. He has been a keynote speaker at innumerable GIS conferences across the globe.

Dr Tomlinson's contributions have been recognized by a glittering array of high honors including the 2003 Gold Medal of the Royal Canadian Geographical Society, the Murchison Award of the Royal Geographical Society and – most prestigious of all – the Order of Canada.

(A) (B)

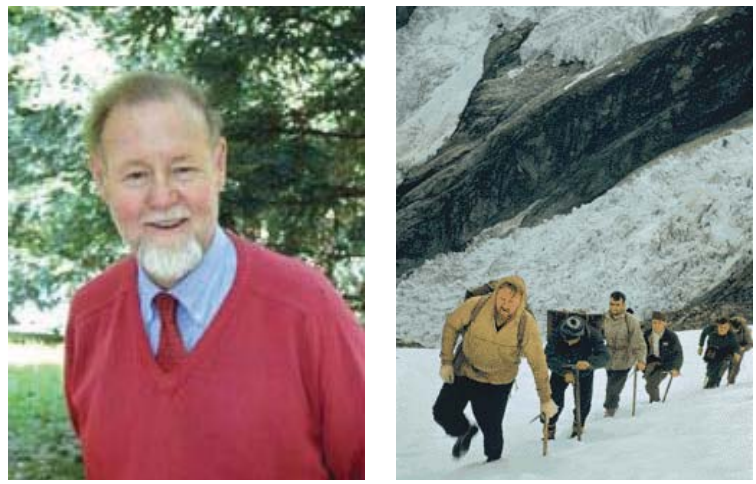


Figure 17.1 Dr Roger Tomlinson, GIS pioneer, in 2003 and as he was in 1957 leading an expedition to the Norwegian Ice Cap

Technical Box **17.2**

The Tomlinson methodology for getting a GIS that meets your needs

Stage	Action	Commentary (after Tomlinson 2003)
1	Consider the strategic purpose	Strategic purpose is the guiding light. The system that gets implemented must be aligned with the purpose of the organization as a whole.
2	Plan for the planning	Since the GIS planning process will take time and resources, you will need to get an approval and commitment at the front end from senior managers in the organization.
3	Conduct a technology seminar	Think of the technology seminar as a sort of 'town-hall meeting' between the GIS planning team, the various staff, and other stakeholders in the organization.
4	Describe the information products	Know what you want to get out of it.
5	Define the system scope	Scoping the system means defining the actual data, hardware, software, and timing.

Stage	Action	Commentary (after Tomlinson 2003)
6	Create a data design	The data landscape has changed dramatically with the advent of the Internet and the proliferation of commercial datasets. Developing a systematic procedure for safely navigating this landscape is critical.
7	Choose a logical data model	The new generation of object-oriented data models is ushering in a host of new GIS capabilities and should be considered for all new implementations. Yet the relational model is still prevalent and the savvy GIS user will be conversant with both (Chapter 8).
8	Determine system requirements	Getting the system requirements right at the outset [and providing the capacity for their evolution] is critical to successful GIS planning.
9	Benefit-cost, migration, and risk analysis	The most critical aspect of doing a realistic benefit-cost analysis is the commitment to include all the costs that will be involved. Too often managers gloss over the real costs, only to regret this later.
10	Make an implementation plan	The implementation plan should illuminate the road to GIS success.

Box 17.3. This response is especially common where other organizations in the same business have shown demonstrable benefits from operating one – and hence where competition decrees the need to mimic or surpass them (see Section 2.3.1). Important aspects of the business case which should be created formally for any GIS acquisition or major enhancement are described in Chapter 18. But at the most strategic level, there are usually two 'demand side' reasons and three general 'supply side' reasons to implement a GIS:

Applications Box **17.3**

Managing land information in Korea through GIS

In South Korea, local government authorities administer the public land through the assessment of land prices, management of land transactions, land use planning and management, and civil services. In many cases, more than one department of a local authority produces and manages the same or similar land and property information; this has led to discrepancies in the information. With the large number of public land administration responsibilities and the control of each given to the local authorities, many problems arose; Korea is a complex and rapidly changing society (see Figure 17.2A). This led to the decision to develop a GIS-based method for sharing the information produced or required for administering land in the public and private sectors (Figure 17.2B).

The Korean Land Management Information System (LMIS) was established in 1998. The purpose of this GIS is to provide land information, increase the productivity of the public land administration, and support the operation of the land planning policies of the Korean Ministry of Construction and Transportation (MOCT). The LMIS database includes much spatial data such as topographic, cadastral, and land use district maps.

(A)

Figure 17.2 (A) Seoul by night. (B) Land information map for part of Seoul (Reproduced by permission of Corbis)

Hyunrai Kim, vice director of the Land Management Division of Seoul Metropolitan City, summarized the advantages this provides: 'By means of the Internet-based Land Information Service System, citizens can get land information easily at home. They don't have to visit the office, which may be located far from their homes.' The system has also resulted in time and cost savings. With the development of the Korean Announced Land Price Management System, it is also possible to compute land prices directly and produce maps of the variations in land price.

Initially, the focus was mainly on the administrative aspects of data management and system development; however, attention then turned to the expansion and development of a decision support system using various data analyses. It is intended that the Land Legal Information Service System will also be able to inform land users of regulations on land use. In essence, LMIS is becoming a crucial element of e-government (see Section 20.3.5). This case study highlights the role that GIS can play beyond the obvious one of information management, analysis, and dissemination. It highlights the value of GIS in enabling organizational integration and the reality of generating benefits through improved staff productivity.

(B)

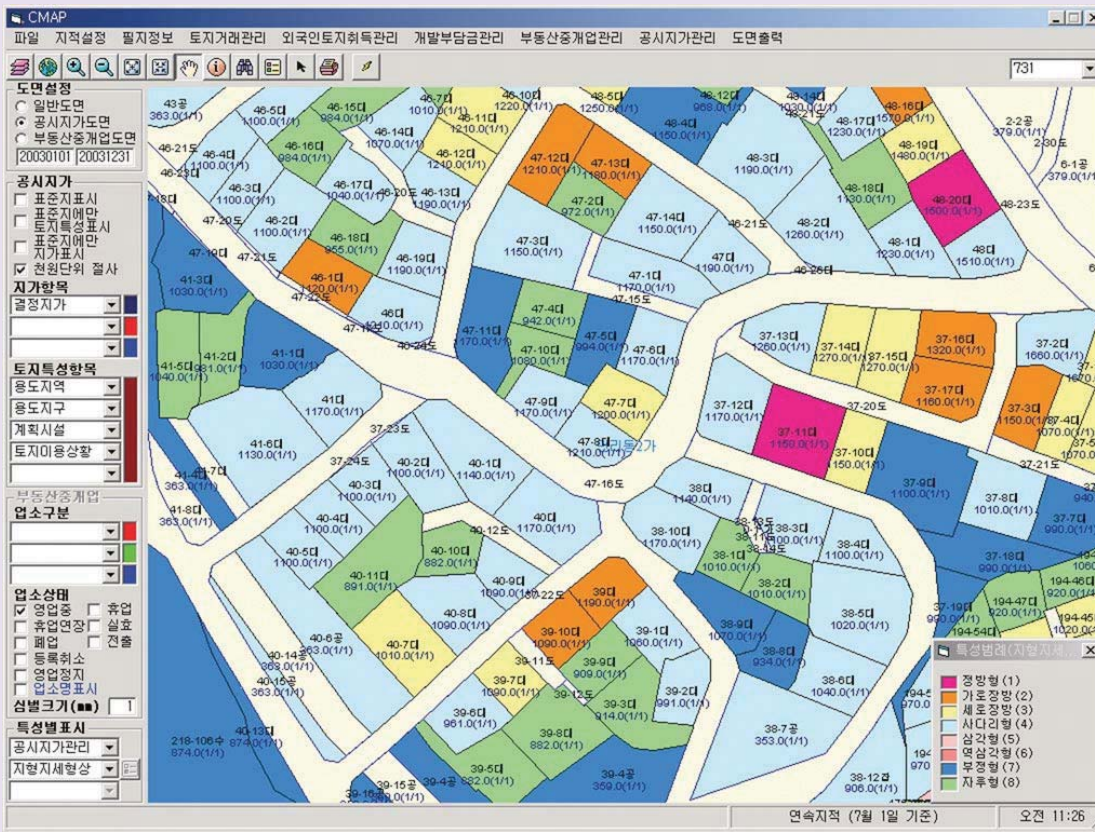


Figure 17.2 (*continued*)

- *Cost reduction*. GIS can replace, in part or completely, many existing operations, e.g., drafting maps, locating and maintaining customers, and managing land acquisition and disposal more efficiently.
- *Cost avoidance*. Examples of GIS use are in locating facilities away from areas at high risk from natural hazards (e.g., tornadoes, floods, or land slides) and by minimizing delivery routes (e.g., letters, refrigerators, and beer).
- *Increased revenue*. Examples include finding and attracting new customers, making maps and data for sale and using GIS as a tool to support consultants (e.g., facility siting, natural resource conservation, and real estate management).

- *Getting wholly new products*. The GIS may enable you to produce products which simply could not have been created previously or which would have been impossibly costly or time-consuming to produce (e.g., satellite imagery of an area draped on a three-dimensional landscape or escape routes from hazards which take into account likely congestion; see Section 2.3.4.2 and Box 16.2).
- *Getting non-tangible (or intangible) benefits*. These benefits are difficult to measure, but they can be very important nonetheless. Examples include making better decisions, providing better service to customers and clients (which can lead to improved public image), the use of consistent information across an organization, the production of reproducible and defensible results, and the ability to document and share the processes and methodology used to solve a problem (see Chapter 19).

17.2 The process of developing a sustainable GIS

GIS projects are similar to many other large IT projects in that they can be broken down into four major life-cycle phases (Figure 17.3). For our simplified purposes, these are:

- business planning (strategic analysis and requirements gathering);
- system acquisition (choosing and purchasing a system);
- system implementation (assembling all the various components and creating a functional solution); and
- operation and maintenance (keeping a system running).

These phases are iterative. Over a decade or more, several iterations may occur, often using different generations of GIS technology and methodologies. Variations on this model include prototyping and rapid application development – but space does not permit much discussion of them here.

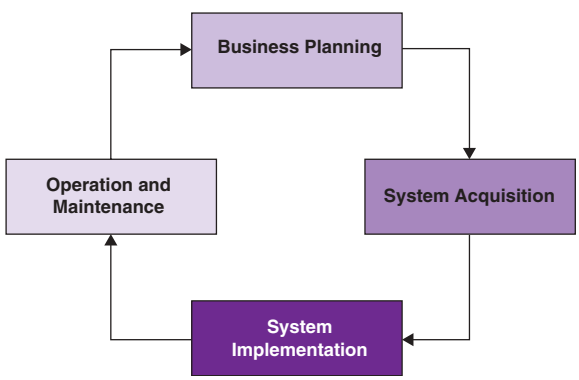


Figure 17.3 GIS project lifecycle stages

GIS projects comprise four major lifecycle phases: business planning; system acquisition; system implementation; and operation and maintenance.

17.2.1 Choosing a GIS

Clarke has proposed a general model of how to specify, evaluate, and choose a GIS, variations of which have been used by organizations over the past 20 years or so. The model we prefer (Figure 17.4) is based on 14 steps grouped into four stages: analysis of requirements; specification of requirements; evaluation of alternatives; and implementation of system; we use it here rather than the 'top level' Tomlinson model because it is more detailed on the implementation phases. Such a process is both time-consuming and expensive. It is really only appropriate for large GIS implementations (contracts over $100 000 in initial value), where it is particularly important to have investment and risk appraisals (see Section 19.4.2 onwards). We describe the model here so that those involved with smaller systems can judiciously select those elements relevant to them. On the basis of painful experience, however, we urge the use of formalized approaches to evaluating the need for and any subsequent acquisition of a system. It is amazing how small projects, carried out quickly because they are small, evolve into big and costly ones!

Choosing a GIS involves four stages: analysis of requirements; specification of requirements; evaluation of alternatives; and implementation of system.

For organizations undertaking acquisition for the first time, huge benefits can be accrued through partnering with other organizations (Chapter 20) that are more advanced, especially if they are in the same field. This is often possible in the public sector, e.g., where local governments have similar tasks to meet. But a surprising number of private sector organizations are also prepared to share their experiences and documents.

Stage 1: Analysis of requirements

The first stage in choosing a GIS is an iterative process for identifying and refining user requirements, and for determining the business case (see Chapter 19) for acquiring a GIS. The deliverable for each step is a report that should be discussed with users and management. It is important to keep records of the discussions and share them with those involved so there can be no argument at a later stage about what was agreed! The results of each report help determine successive stages.

Step 1: Definition of objectives

This is often a major decision for any organization. The rational process of choosing a GIS begins with and spins

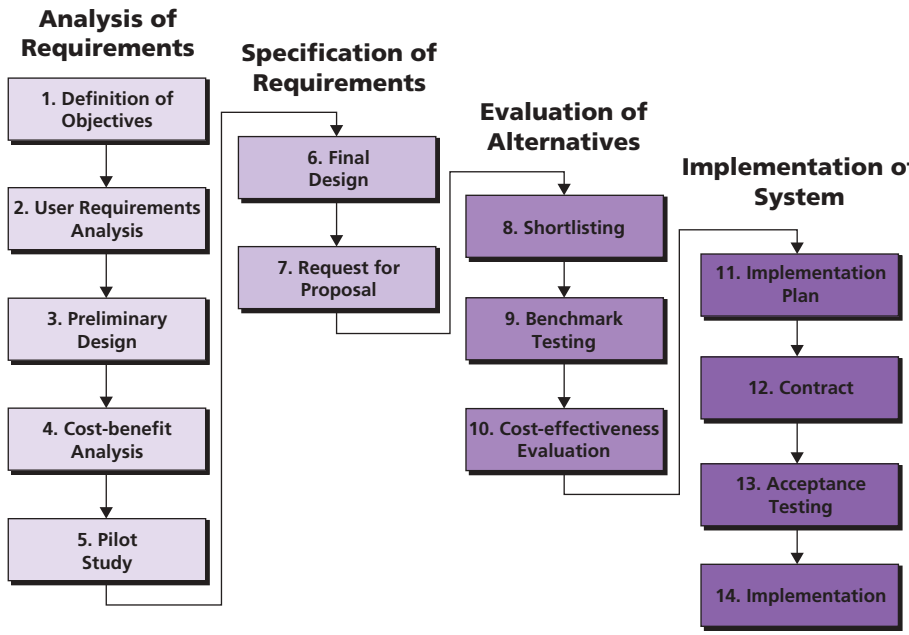


Figure 17.4 General model of the GIS acquisition process

out of the development of the organization's strategic plan (Section 19.4.1) – and an outline decision that GIS can play a role in the implementation of this plan. Strategic and tactical objectives must be stated in a form understandable to managers. The outcome from Step 1 is a document that managers and users can endorse as a plan to proceed with the acquisition, i.e., the relevant managers believe there is sufficient promise to proceed to the next step and commit the initial funding required.

Step 2: User requirements analysis

The analysis will determine how the GIS is designed and evaluated. Analysis should focus on what information is presently being used, who is using it, and how the source is being collected, stored, and maintained. This is a map of existing processes (which may possibly be improved as well as being replicated by the GIS). The necessary information can be obtained through interviews, documentation, reviews, and workshops. The report for this phase should be in the form of workflows, lists of information sources, and current operation costs. The clear definition of likely or possible change (e.g., future applications – see Figure 17.5), new information products (e.g., maps and reports – Figure 17.6) or different utilization of functions and new data requirements is essential to successful GIS implementation.

Step 3: Preliminary design

This stage of the design is based on results from Step 2. The results will be used for subsequent cost-benefit analysis (Step 4 below) and will enable specification of the pilot study. The four key tasks are: develop preliminary database specifications; create preliminary functional specifications; design preliminary system models; and survey the market for potential systems. Database specifications involve estimating the amount and type of data

APPLICATION

Display zoning map information for a user-defined area.

FUNCTIONS USED IN THE APPLICATION:

Review and prepare zoning changes.

DESCRIPTION OF APPLICATION:

This application uses zoning and related parcel-based data from the database to display existing information related to zoning for a specific area that is defined by the user. The application must be available interactively at a workstation when the user invokes a request and identifies the subject land parcel. The application will define a search area based upon the search distance defined and input by the user, and will display all required data for the area within the specified distance from the outer boundary of the subject parcel.

DATA INPUTS:

User defined: Parcel identifier
Search distance

Database: Zoning boundaries
Zoning dimensions
Zoning codes
Parcel boundaries
Parcel dimensions
Parcel numbers
Street names
Addresses

PRODUCTS OUTPUT:

1. Zoning map screen display with subject parcel highlighted, search area boundaries, search distance, all zoning data, parcel data, street names and addresses.
2. Hard copy map of the above.

Figure 17.5 Sample application definition form

(see Chapter 9). Many consultants maintain checklists and vendors frequently publish descriptions of their systems on their websites. The choice of system model involves decisions about raster and vector data models and system

Figure 17.6 A Web-displayed report produced by a local government GIS (*Source*: Mecklenburg County, North Carolina GIS; see **maps2.co.mecklenburg.nc.us/website/realestate/viewer.htm**) (Reproduced by permission of Mecklenburg County, NCIST)

type (see Chapters 7 and 8). Finally, a market survey should be undertaken to assess the capabilities of commercial off-the-shelf (COTS) systems. This might involve a formal Request For Information (RFI) to a wide range of vendors. A balance needs to be struck in writing this between creating a document so open that the vendor has problems identifying what needs are paramount and one so prescriptive and closed that no flexibility or innovation is possible.

Whether to buy or to build a GIS used to be a major decision (Chapter 7). This occurred especially at 'green field' sites – where no GIS technology has hitherto been used – and at sites where a GIS has already been implemented but was in need of modernization. But the situation is now quite different: use of general-purpose COTS solutions are the norm. These have ongoing programs of enhancement and maintenance and can normally be used for multiple projects. Typically they are better documented and more people in the job market have experience of them. As a consequence, risk arising from loss of key personnel is reduced.

There has been a major move in GIS away from building proprietary GIS toward buying COTS solutions.

Step 4: Cost-benefit analysis

Purchase and implementation of a GIS is a non-trivial exercise, expensive in both money and staff resources (typically management time). It is quite common for organizations to undertake a cost-benefit (also called benefit-cost) analysis to justify the effort and expense, and to compare it against the alternative of continuing with the current data, processes, and products – the *status quo*. Box 17.4 summarizes the process.

Cost-benefit cases are normally presented as a spreadsheet, along with a report that summarizes the main findings and suggests whether the project should be continued or halted. Senior managers then need to assess the merits of this project in comparison with any others competing for their resources (see Section 19.4.2).

Technical Box **17.4**

Cost-benefit analysis

Cost-benefit analysis begins with a determination of GIS implementation costs (hardware, software, data, and people) and the projected benefits (improved service, reduced costs, greater use of data, etc.). A monetary value is then assigned to both costs and benefits. When these are summed, the assessed benefits should exceed the projected costs over the full lifecycle if the project is to proceed. This sounds quite simple and straightforward. In practice, however, cost-benefit analysis is more of an art than a science because there are many complicating factors and assumptions to be made. For example, it is difficult to assign a monetary value to many activities, and in many traditional public organizations fixed costs (such as staff time, centralized administration, and computing facilities) are still not taken properly into account. In addition, a positive cost-benefit analysis (benefits exceed costs) may not ensure that the project proceeds – that will also depend on whether the organization's cash flow and management time resources are adequate, whether the risks are acceptable (see Section 19.4.2), and whether another project competing for the same resources has a better cost-benefit out-turn

Table 17.1 Simple examples of GIS costs and benefits (after Obermeyer 2005, with additions)

Category	Costs	Benefits
Economic (tangible)	• Hardware • Software • Training • New staff or skills • Additional space • Data purchase or collection	• Reduced costs (e.g., of staff) • Greater throughput • Increased revenues • New market services or products
Institutional (intangible)	• Interpersonal shifts • Layoffs of low-skilled staff, partially replaced by fewer if individually more expensive skilled staff	• Improved client relationships • Better decisions • Improved morale • Better information flow • Better culture of 'achievers'

The potential costs and benefits of implementing a GIS fall into two categories (Table 17.1): economic (tangible) and institutional (intangible). Unfortunately, experience suggests that the more easily measured economic benefits tend ultimately to be less important in GIS projects than the much more ill-defined institutional benefits. This is true especially of systems that perform decision-support rather than operational tasks (e.g., city planning versus postal delivery routing). But it is an important discipline to identify both in as formal a way as possible: the process helps in decision making and will inevitably be the subject of internal audit at a later stage.

Although cost-benefit analysis is the best-known method of assessing the value of a GIS, a number of alternatives have also been proposed. 'Cost-effectiveness analysis' compares the cost of providing the same service using alternative means. The 'payback period' is derived by dividing the cost of implementing a system by the estimated annual net benefits of using it. The 'value-added approach' emphasizes the new things that a GIS allows an organization to perform. In all assessments, account should be taken of the different value of money at different times.

Surprisingly, there are few readily accessible examples of GIS cost-benefit studies. Many users do not even bother to carry out such a prior evaluation, especially now that GIS is becoming a mainstream science and technology. This is short-sighted because adoption of GIS has long-term consequences – such as staff training and development needs – which go beyond simple initial financial expenditure on equipment and software. Many organizations also gain competitive advantage from GIS and do not want to publish their analyses. Moreover, cost-benefit studies are often carried out by consultants who do not want to make their approach public for fear that they will no longer be able to charge for it.

Step 5: Pilot study

A pilot study is a mini-version of the full GIS implementation that aims to test several facets of the project. The primary objective is to test a possible or likely system design before finalizing the system specification and committing significant resources. Secondary objectives are to develop the understanding and confidence of users and sponsors, to test samples of data if a data capture project is part of the implementation, and to provide a test bed for application development.

A pilot is a mini-version of a full GIS implementation designed to test as many aspects of the final system as possible.

It is normal to use existing hardware or to lease hardware similar to that which is expected to be used in the full implementation. A reasonable cross-section of all the main types of data, applications, and product deliverables should be used during the pilot. But the temptation must be resisted to try to build the whole system at this stage, irrespective of how easy 'the techies' may claim it to be! Users should be prepared to discard *everything* after the pilot if the selected technology or application style does not live up to expectations.

The outcome of a pilot study is a document containing an evaluation of the technology and approach adopted, an assessment of the cost-benefit case, and details of the project risks and impacts. Risk analysis is an important activity, even at this early stage (see Section 19.4.1.2). Assessing what can go wrong can help avoid potentially expensive disasters in the future. The risk analysis should focus on the actual acquisition processes as well as on implementation and operation.

Stage 2: Specification of requirements

The second stage is concerned with developing a formal specification that can be used in the structured process of soliciting and evaluating proposals for the system.

Step 6: Final design

This creates the final design specifications for inclusion in a Request For Proposals (RFP: also called an invitation to tender or ITT) to vendors. Key activities include finalizing the database, defining the functional and performance specifications, and creating a list of possible constraints. From these, requirements are classified as mandatory, desirable, or optional. The deliverable is the final design document. This document should provide a clear description of essential requirements – without being so prescriptive that innovation is stifled, costs escalate, or insufficient vendors feel able to respond.

Step 7: Request for proposals

The RFP document combines the final design document with the contractual requirements of the organization. These will vary from organization to organization but are likely to include legal details of copyright of the design and documentation, intellectual property ownership, payment schedules, procurement timetable, and other draft terms and conditions. Once the RFP is released to vendors by official advertisement and/or personal letter, a minimum period of several weeks is required for vendors to evaluate and respond. For complex systems, it is usual to hold an open meeting to discuss technical and business issues.

Stage 3: Evaluation of alternatives

Step 8: Short-listing

In situations where several vendors are expected to reply, it is customary to have a short-listing process. Submitted proposals must first be evaluated, normally using a weighted scoring system, and the list of potential suppliers narrowed down to between two and four. Good practice is that the scoring must be done by several individuals acting independently and the results – and the differences between the evaluations – compared. This whole process allows both the prospective purchaser and supplier organizations to allocate their resources in a focused way. Short-listed vendors are then invited to attend a benchmark-setting meeting.

Step 9: Benchmarking

The primary purpose of a benchmark is to evaluate the proposal, people, and technology of each selected vendor. Each one is expected to create a prototype of the final system that will be used to perform representative tests. The results of these tests are scored by the prospective purchaser. Scores are also assigned for the original vendor proposal and the vendor presentations about their company. Together, these scores form the basis of the final system selection. Unfortunately, benchmarks are often conducted in a rather secretive and confrontational way, with vendors expected to guess the relative priorities (and the weighting of the scores) of the prospective purchaser. Whilst it is essential to follow a fair and transparent process, maintain a good audit trail and remain completely impartial, a more open co-operative approach usually produces a better evaluation of vendors and their proposals. If vendors know which functions have the greatest value to customers they can tune their systems appropriately.

Step 10: Cost-effectiveness evaluation

Next, surviving proposals are evaluated for their cost-effectiveness. This is again more complex than it might seem. For example, GIS software systems vary quite widely in the type of hardware they use, some need additional database management system (DBMS) licenses, customization costs will vary, and maintenance will often be calculated in different ways. The goal of this stage is to normalize all the proposals to a common format for comparative purposes. The weighting used for different parts must be chosen carefully since this can have a significant impact on the final selection. Good practice involves debate within the user community – for they should have a strong say – on the weighting to be used and some sensitivity testing to check whether very different answers would have been obtained if the weights were slightly different. The deliverable from this stage is a ranking of vendors' offerings.

Stage 4: Implementation of system

The final stage is planning the implementation, contracting with the selected vendor, testing the delivered system, and actual use of the GIS.

Step 11: Implementation plan

A structured, appropriately paced, implementation plan is an essential ingredient of a successful GIS implementation. The plan commences with identification of priorities, definition of an implementation schedule, and creation of a resource budget and management plan. Typical activities that need to be included in the schedule are installation and acceptance testing, staff training, data collection, and customization. Implementation should be coordinated with both users and suppliers.

Step 12: Contract

An award is subject to final contractual negotiation to agree general and specific terms and conditions, what elements of the vendor proposal will be delivered, when they will be delivered and at what price. General conditions include contract period, payment schedule, responsibilities of the parties, insurance, warranty, indemnity, arbitration, and provision of penalties and contract termination arrangements.

Step 13: Acceptance testing

This is to ensure that the delivered GIS matches the specification agreed in the contract. Part of the payment should be withheld until this step is successfully completed. Activities include installation plus tests of functionality, performance, and reliability. It is seldom the case that a system passes all tests first time and so provision should be made to repeat aspects of the testing.

Step 14: Implementation

This is the final step at the end of what can be a long road (see also Section 17.2.1.1 below). The entire GIS acquisition period can stretch over many months or even longer. Activities include training users and support staff, data collection, system maintenance, and performance monitoring. Customers may also need to be 'educated' as well! Once the system is successfully in operation, it may be appropriate to publicize its success for enhancement of the brand image or political purposes.

17.2.1.1 Discussion of the classical acquisition model

The general model outlined above has been widely employed as the primary mechanism for large GIS procurements in public organizations. It is rare, however, that 'one size fits all' and, although it has many advantages, it also has some significant shortcomings:

- The process is expensive and time-consuming for both suppliers and vendors. A supplier can spend as much as 20% of the contract value on winning the business and a purchaser can spend a similar amount in staff time, external consultancy fees, and equipment rental. This ultimately makes systems more expensive – though competition does drive down cost.
- Because it takes a long time and because GIS is a fast-developing field, proposals can become technologically obsolete within several months.

- The short-listing process requires multiple vendors, which can end up lowering the minimum technical selection threshold in order to ensure enough bidders are available.
- In practice, the evaluation process often focuses undue attention on price rather than the long-term organizational and technical merits of the different solutions.
- This type of procurement can be highly adversarial. As a result, it can lay the foundations for an uncomfortable implementation partnership (see Chapter 20) and often does not lead to full development of the best solution. Every implementation is a trade-off between functionality, time, price, and risk. A full and frank discussion between purchaser and vendor on this subject can generate major long-term benefits.
- Many organizations have little idea about what they *really* need. Furthermore, it is very difficult to specify precisely in any contract exactly what a system must perform. As users learn more, their aspirations also rise – resulting in 'feature creep' (the addition of more capabilities) often without any acceptance of an increase in budget. On the other hand, some vendors adopt the strategy of taking a minimalist view of the capabilities of the system featured in their proposal and make all modifications during implementation and maintenance through chargeable change orders. All this makes the entire system acquisition costs far higher than was originally anticipated; the personal consequences for the budget holders concerned can be unfortunate.
- Increasingly, most organizations already have some GIS; the classical model works best in a 'green field site' situation.

As a result of these problems, this type of acquisition model is not used in small or even some larger procurements, especially where the facilities can be augmented rather than totally replaced. A less complex and formal selection method is *prototyping*. Here a vendor or pair of vendors is selected early on using a smaller version of the evaluation process outlined above. The vendor(s) is/are then funded to build a prototype in collaboration with the user organization. This fosters a close partnership to exploit the technical capabilities of systems and developers and helps to maintain system flexibility in the light of changing requirements and technology. This approach works best for those procurements – sometimes even some large ones – where there is some uncertainty about the most appropriate technical solution and where the organizations involved are mature, able to control the process, and not subject to draconian procurement rules.

Prototyping is a useful alternative to classical, linear system acquisition exercises. It is especially useful for smaller procurements where the best approach and outcome are more uncertain.

17.2.2 Implementing a GIS

The goal of this section is to provide a checklist of important management issues when implementing a GIS.

17.2.2.1 Plan effectively

Good planning is essential through the full lifecycle of all GIS projects. Both strategic and operational, or tactical, planning are important to the success of a project. Strategic planning involves reviewing overall organizational goals and setting specific GIS objectives. Operational planning is more concerned with the day-to-day management of resources. There are several general project management productivity tools available that can be used in GIS projects. Figure 17.7 shows one diagramming tool called a Gantt chart. Several other implementation techniques and tools are discussed in Box 17.5.

17.2.2.2 Obtain support

If a GIS project is to prosper, it is essential to garner support from *all* key stakeholders. This often requires establishing executive (director-level) leadership support; developing a public-relations strategy by, for example, exhibiting key information products or distributing free maps; holding an open house to explain the work on the GIS team; and participating in GIS seminars and workshops, locally and sometimes nationally.

17.2.2.3 Communicate with users

Involving users from the very earliest stages of a project will lead to a better system design and will help with user acceptance. Seminars, newsletters, and frequent updates about the status of the project are good ways to educate and involve users. Setting expectations about capabilities, throughput, and turn-around at reasonable levels is crucial to avoid any later misunderstandings with users and managers.

17.2.2.4 Anticipate and avoid obstacles

These may involve staffing, hardware, software, databases, organization/procedures, timeframe, and funding. Be prepared!

17.2.2.5 Avoid false economies

Money saved by not paying staff a reasonable (market value) wage or by insufficient training is often manifested in reduced staff efficiencies. Furthermore, poorly paid or trained staff often leave through frustration. You cannot prevent this by contractual means (see Section 19.1.2) so must do so through paying market rates and/or building a team culture where staff enjoy working for the organization.

Cutting back on hardware and software costs by, for example, obtaining less powerful systems or cancelling maintenance, may save money in the short-term but

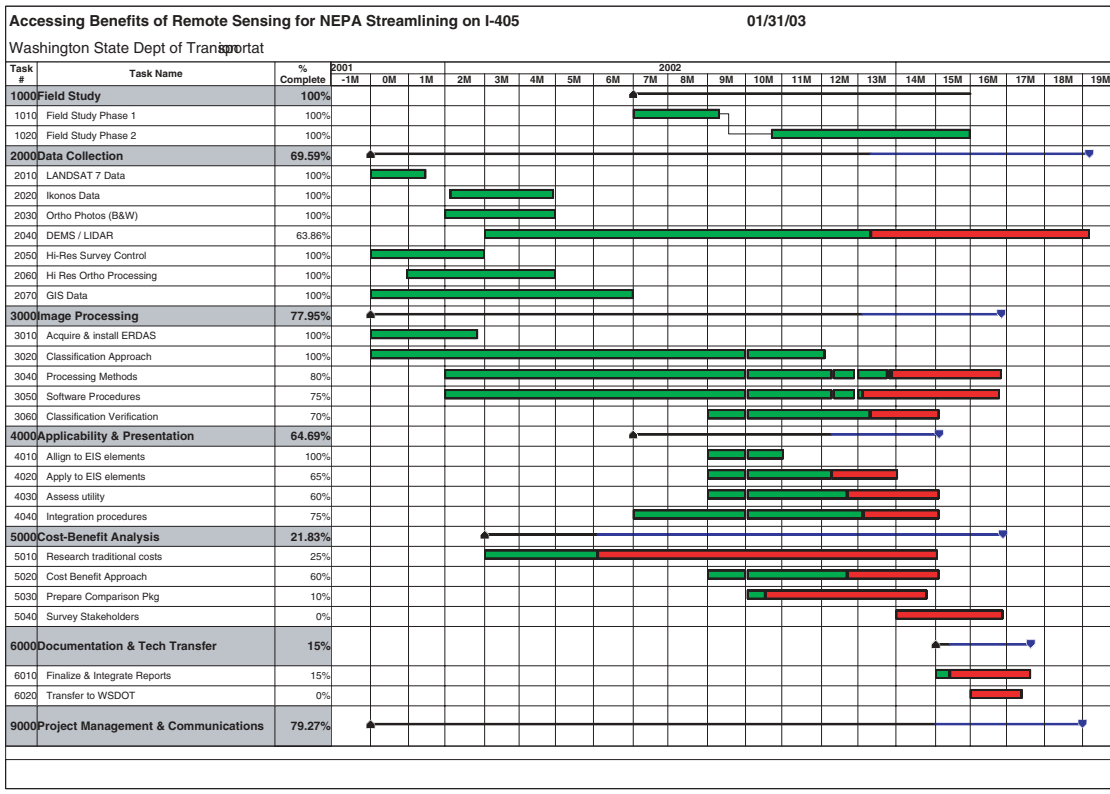


Figure 17.7 Gantt chart of a basic GIS project. This chart shows task resource requirements over time, with task dependencies. This is a relatively simple chart, with a small number of tasks (*Source*: **www.wsdot.wa.gov/environment/envinfo/docs/RSPrj_GanttJan2003.pdf**) (Reproduced by permission of Washington State Department of Transportation)

will likely cause serious problems in the future when workloads increase and the systems get older. Failing to account for depreciation and replacement costs, i.e., by failing to amortize the GIS investment, will store up trouble ahead. The amortization period will vary greatly – hardware may be depreciated to zero value after, say, four years whilst buildings may be amortized over 30 years.

17.2.2.6 Ensure database quality and security

Investing in the database quality is essential at all stages from creation onwards. Catastrophic results may ensue if any of the up-dates or (especially) the database itself is lost in a system crash or corrupted by hacking, etc. This requires not only good precautions but also contingency (disaster recovery and business continuity) plans and periodic serious trials of them.

17.2.2.7 Accommodate GIS within the organization

Building a system to replicate ancient and inefficient ones is not a good idea; nor is it wise to go to the other extreme and expect the whole organization's ways of working to be changed to fit better with what the GIS can do! Too much change at any one time can destroy organizations just as much as too little change can ossify them. In general, the GIS must be managed in a way that fits with the organizational aspirations and culture if it is to be a success (see Section 18.2.3). All this is especially a problem because GIS projects often blaze the trail in terms of introducing new technology, interdepartmental resource sharing, and generating new sources of income.

17.2.2.8 Avoid unreasonable timeframes and expectations

Inexperienced managers often underestimate the time it takes to implement GIS. Good tools, risk analysis, and time allocated for contingencies are important methods of mitigating potential problems. The best guide to how long a project takes is experience in other similar projects – though the differences between the organizations, staffing, tasks, etc., need to be taken into account.

17.2.2.9 Funding

Securing ongoing, stable funding is a major task of a GIS manager. Substantial GIS projects will require core funding from one or more of the stakeholders. None of these will commit to the project without a business case and risk analysis (Section 19.4). Additional funding for special projects, and from information and service sales, is likely to be less certain. It is characteristic of

Technical Box **17.5**

Implementation techniques

There are many tools and techniques available for GIS managers to use in implementation projects. These are summarized in Table 17.2.

Table 17.2 GIS implementation tools and techniques (after Heywood et al 2002, with additions)

Technique	Purpose
SWOT analysis	A management technique used to establish Strengths, Weaknesses, Opportunities, and Threats in a GIS implementation. The output is a list and narrative.
Rich picture analysis	Major participants are asked to create a schematic/picture showing their understanding of a problem using agreed conventions. These are then discussed as part of a consensus-forming process.
Demonstration systems	Many vendors and GIS project teams create prototype demonstrations to stimulate interest and educate users/funding agencies.
Interviews and data audits	These aim to define problems and determine current data holdings. The output is a report and recommendations.
Organization charts, system flow charts, and decision trees	These are all examples of flow charts that show the movement of information, the systems used, and how decisions are currently reached.
Data flow diagrams and dictionaries	These are charts that track the flow of information and computerized lists of data in an organization.
Project management tools	Gantt charts (see Figure 17.7) and PERT (Program, Evaluation, and Review Techniques) are tools for managing time and resources.
Object model diagrams	These show objects to be modeled in a GIS and the relationships between them (see for example Figure 8.16).

many GIS projects that the operational budget will change significantly over time as the system matures. The three main components are staff, goods and services, and capital investments. A commonly experienced distribution of costs between these three elements is shown in Table 17.3.

Table 17.3 Percentage distribution of GIS operational budget elements over three time periods (after Sugarbaker 2005)

Budget item	Year 1–2	Year 3–6	Year 6–12
Staff and benefits	30	46	51
Goods and services	26	30	27
Equipment and software	44	24	22
Total	100	100	100

17.2.2.10 Prevent meltdown

Avoiding the cessation of GIS activities is the ultimate responsibility of the GIS manager. According to Tomlinson, some of the main reasons for the failure of GIS projects are:

- lack of executive-level commitment;
- inadequate oversight of key participants;
- inexperienced managers;
- unsupportive organizational structure;
- political pressures, especially where these change rapidly;
- inability to demonstrate benefits;
- unrealistic deadlines;
- poor planning; and
- lack of core funding.

17.2.3 Managing a sustainable, operational GIS

Larry Sugarbaker has characterized the many operational management issues throughout the lifecycle of a GIS project as: customer support; effective operations; data management; and application development and support. Success in any one – or even all – of these areas does

not guarantee project success, but they certainly help to produce a healthy project. Each is now considered in turn.

Success in operational management of GIS requires customer support, effective operations, data management, and application development and support.

17.2.3.1 Customer support

In progressive organizations *all* users of a system and its products are referred to as customers. A critical function of an operational GIS is a customer support service. This could be a physical desk with support staff or, increasingly, it is a networked electronic mail and telephone service. Since this is likely to be the main interaction with GIS support staff, it is essential that the support service creates a good impression and delivers the type of service users need. The unit will typically perform key tasks including technical support and problem logging plus meeting requests for data, maps, training, and other products. Performing these tasks will require both GIS analyst-level and administrative skills. It is imperative that *all* customer interaction is logged and that procedures are put into place to handle requests and complaints in an organized and structured fashion. This is both to provide an effective service and also to correct systemic problems

Customer support is not always seen as the most glamorous of GIS activities. However, a GIS manager who recognizes the importance of this function and delivers an efficient and effective service will be rewarded with happy customers. Happy customers remain customers. Effective staff management includes finding staff with the right interests and aspirations, rotating GIS analysts through posts and setting the right (high) level of expectation in the performance of all staff. Managers can learn much by taking a turn in the hot seat of a customer support role!

17.2.3.2 Operations support

Operations support includes system administration, maintenance, security, backups, technology acquisitions, and many other support functions. In small projects, everyone is charged with some aspects of system administration and operations support. But as projects grow beyond five or more staff, it is worthwhile to designate someone specifically to fulfill what becomes a core, even crucial, role. As projects become larger, this grows into a full-time function. System administration is a highly technical and mission-critical task requiring a dedicated, properly trained and paid person.

Perhaps more than in any other role, clear written descriptions are required for this function to ensure that a high level of service is maintained. For example, large, expensive databases will require a well-organized security and backup plan to ensure that they are *never* lost or corrupted. Part of this plan should be a disaster recovery strategy. What would happen, for example, if there were a fire in the building housing the database server or some other major problem (e.g., see Box 18.1)?

17.2.3.3 Data management support

The concept that geographic data are an important part of an organization's critical infrastructure is becoming more widely accepted. Large, multi-user geographic databases use database management system (DBMS) software to allocate resources, control access, and ensure long-term usability. DBMS can be sophisticated and complicated, requiring skilled administrators for this critical function.

A database administrator (DBA) is responsible for ensuring that all data meet all of the standards of accuracy, integrity, and compatibility required by the organization. A DBA will also typically be tasked with planning future data resource requirements – derived from continuing interaction with current and potential customers – and the technology necessary to store and manage them. Similar comments to those outlined above for System Administrators also apply to this position.

17.2.3.4 Application development and support

Although a considerable amount of application development is usual at the onset of a project, it is also likely that there will be an on-going requirement for this type of work. Sources of application development work include improvements/enhancements to existing applications, as well as new users and new project areas starting to adopt GIS.

Software development tools and methodologies are constantly in a state of flux and GIS managers must invest appropriately in training and new software tools. The choice of which language to use for GIS application development is often a difficult one. Consistent with the general movement away from proprietary GIS languages, wherever possible GIS managers should try to use mainstream, open languages that appear to have a long lifetime. Ideally, application developers should be assigned full-time to a project and should become permanent members of the GIS group to ensure continuity (but often this does not occur).

17.3 Sustaining a GIS – the people and their competences

Throughout this chapter we have sometimes highlighted and sometimes hinted at the key role of staff as assets in all organizations; their education is discussed in Section 19.2. If they do not function well – individually and as a team – nothing of merit will be achieved.

17.3.1 GIS staff and the teams involved

Several different staff will carry out the operational functions of a GIS. The exact number of staff and their

precise roles will vary from project to project. The same staff member may carry out several roles (e.g., it is quite common for administration and application development to be performed by a GIS technical person), and several staff members may be required for the same task (e.g., there may be many digitizing technicians and application developers). Figure 17.8 shows a generalized view of the main staff roles in medium to large GIS projects.

All significant GIS projects will be overseen by a management board built up of a senior sponsor (usually a Director or Vice-President), members of the user community, and the GIS manager. It is also useful to have one or more independent members to offer disinterested advice. Although this group may seem intimidating and restrictive to some, used in the right way it can be a superb source of funding, advice, support, and encouragement.

Typically, day-to-day GIS work involves three key groups of people: the GIS team itself; the GIS users; and external consultants. The GIS team comprises the dedicated GIS staff at the heart of the project; the GIS manager is the team leader. This individual needs to be skilled in project and staff management (see Section 17.3.2) and have sufficient understanding of GIS technology and the organization's business to handle the liaisons involved. Larger projects will have specialist staff experienced in project management, system administration, and application development.

GIS users are the customers of the system. There are two main types of user (other than the leaders of organizations who may rely upon GIS indirectly to provide information on which they base key decisions). These are professional users and clerical staff/technicians. Professional users include engineers, planners, scientists, conservationists, social workers, and technologists who utilize output from GIS for their professional work. Such users are typically well-educated in their specific field, but may lack advanced computer skills and knowledge of the GIS. They are usually able to learn how to use the system themselves and can tolerate changes to the service. Clerical and technical users are frequently employed as part of the wider GIS project initiative to perform tasks like data collection, map creation, routing, and service call response. Typically, the members of this group have limited training and skills for solving *ad hoc* problems. They need robust, reliable support. They may also include staff and stakeholders in other departments or projects that assist the GIS project on either a full- or part-time basis, e.g., system administrators, clerical assistants, or software engineers provided from a common resource pool or managers of other databases or systems with which the GIS must interface. Finally, many GIS projects utilize the services of external consultants. They could be strategic advisors, project managers, or technical consultants able to supplement the available staffing. Although these may appear expensive at first sight, they are often well-trained and highly focused. They can be a valuable addition to a project, especially if internal knowledge and/or resources are limited and for benchmarking against approaches elsewhere. But the in-house team must not rely too heavily on consultants lest, when they go, all key knowledge and high-level experience goes with them.

The key groups involved in GIS are: the management board; the GIS team (headed by a GIS manager); the users; external consultants; and various customers.

17.3.2 Project managers

A GIS project will almost certainly have several subprojects or project stages and hence require a structured approach to project management. The GIS manager may take on this role personally, although in large projects it is customary to have one or more specialist project managers. The role of the project manager is to establish

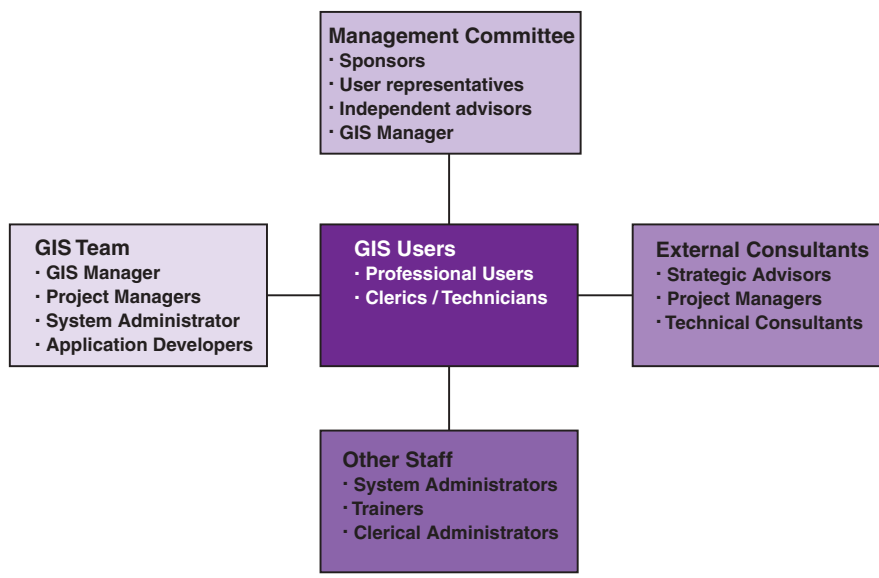


Figure 17.8 The GIS staff roles in a medium to large GIS project

user requirements, to participate in system design, and to ensure that projects are completed on time, within budget, and according to an agreed quality plan. Good project managers are rare creatures and must be nurtured for the good of the organization. One of their characteristics is that, once one project is completed, they like to move on to another so retaining them is only possible in an enterprising environment. Transferring their expertise and knowledge into the heads and files of others is a priority before they leave a project.

17.3.3 Coping with uncertainty

As we have seen earlier, geographic information varies hugely in its characteristics, and rarely is the available information ideal for the task in hand. Staff need a clear understanding of the concepts and implications of uncertainty (Chapter 6) and the related concepts of accuracy, error and sensitivity. Understanding of business risk arising from GIS use and how GIS can help reduce organizational risk is also essential (Chapter 19). This section focuses on the practical aspects relevant to managers of operational GIS – and hence on the skills, attitudes, and other competences they need to bring to their work.

Organizations must determine how much uncertainty they can tolerate before information is deemed useless. This can be difficult because it is application-specific. An error of 10 m in the location of a building is irrelevant for business geodemographic analysis but it could be critical for a water utility maintenance application that requires digging holes to locate underground pipes. Some errors in GIS can be reduced but sometimes at a considerable cost. It is common experience that trying to remove the last 10% of error typically costs 90% of the overall sum. As we concluded in Section 6.5, uncertainty in GIS-based representations is almost always something that we have to live with to a greater or lesser extent. The key issue here is identifying the amount of uncertainty that can be tolerated for a given application, and what can be done at least partially to eliminate it or ameliorate its consequences. Some of this, at least, can only be done by judgement informed by past experience.

A typology of the errors commonly encountered in GIS is also in Chapter 6. Some practical examples of errors in operational GIS are:

- Referential errors in the identity of objects, e.g., a street address could be wrong, resulting in incorrect property identification during an electricity network trace.
- Topological errors, e.g., a highway network could have missing segments or unconnected links, resulting in erroneous routing of service or delivery vehicles.
- Relative positioning errors, e.g., a gas station incorrectly located on the wrong side of a dual carriageway road could have major implications for transportation models.
- Absolute errors in the location of objects in the real world, e.g., tests for whether factories are within a smoke control zone or floodplain could provide erroneous results if the locations are incorrect. This could lead to litigation.
- Attribute errors, e.g., incorrectly entering land use codes would give errors in agricultural production returns to government agencies.

Managing error requires use of quality assurance techniques to identify them and assess their magnitude. A key task is determining the error tolerance that is acceptable for each data layer, information product, and application. It follows that both data creators and users must make analyses of possible errors and their likely effects, based on a form of cost-benefit analysis. And sitting in the midst of all this is the GIS manager who must know enough about uncertainty to ask the right questions; if the system provides what subsequently turns out to be nonsense, he or she is likely to be the first to be blamed! In no sense is the good and long-lasting GIS manager simply someone who ensures that the 'wheels go round'. Box 17.6 presents one highly skilled GIS manager.

17.4 Conclusions

Mostly, any management function in GIS (and indeed elsewhere) is about motivating, organizing or steering, enhancing skills, and monitoring the work of other people.

Managing a GIS project *is* different to using GIS in decision making. Normally, the first requires good GIS expertise and first class project management skills. In contrast, those involved at different levels of the organization's management chain need some awareness of GIS, its capabilities, and its limitations – scientific and practical – alongside their substantial leadership skills. But our experience is that the division is not clear-cut. GIS project managers cannot succeed unless they understand the objectives of the organization, the business drivers, and the culture in which they operate (Chapter 18), plus something of how to value, exploit, and protect their assets (Chapter 19). Equally, decision makers can only make good decisions if they understand more of the scientific and technological background than they may wish to do: running or relying on a good GIS service involves much more than the networking of a few PCs running one piece of software.

So, good management of GIS requires excellent people, technical and business skills, and the capacity to ensure mutual respect and team working between the users and the experts.

Biographical Box **17.6**

Dawn Robbins, GIS manager

Figure 17.9 Dawn Robbins, GIS Manager Ventura County

Dawn Robbins is the Geographic Information Officer (GIO) for Ventura County, California, USA. She has a long history in GIS that includes experience as a consultant. Ms. Robbins holds geography degrees from Dartmouth College (B.A.) and Penn State (M.S.). As GIO, she oversees county-wide geographic data coordination and provision of GIS services such as base map maintenance, system administration, database management, application design and development, and training. Her GIS department has ten staff and an annual budget of over $1m.

Dawn says:

As a GIS manager, my job is to focus on the business viability and day-to-day operations of a GIS Division. My daily duties include project management, review of technical issues, marketing and sales, and staff supervision. I try to get the best use of geographic data throughout the organization (that is, throughout the entire County). Sometimes the solution to a problem does not involve GIS (at least in the traditional sense), but involves a new application that happens to include maps.

For example, two of the questions I recently faced as the GIO were 'Why are the Assessor and the Public Works departments performing seemingly redundant map creation and update procedures – one in GIS and one in CAD?'; and 'how can the County provide access to all land-related information (GIS, CAD, graphics systems, legacy databases, etc.) in a consistent, easy to access manner?' These questions require a 'big-picture' approach and a certain amount of political knowledge and sales ability to find a solution. Challenges that took a great deal of energy over my first year in the job included improving relationships with the other organizations in the region; firming up a budget for the GIS Division; and staff evolution.

A GIO must have the basic technical knowledge of GIS, database, and IT processes. But this knowledge must be coupled with vision, political awareness, and a strong sales ability. A GIO mirrors the eclectic nature of GIS itself, as the individual must be able to work in a variety of locations (agencies, cities, etc.), and at various administrative levels. This is difficult – I don't believe I'm there yet.

Questions for further study

1. How can we assess the potential value of a proposed GIS?
2. Prepare a new sample GIS application definition form using Figure 17.5 as a guide.
3. List ten tasks critical to a GIS project that a GIS manager must perform and the roles of the main members of a GIS project team.
4. Why might a GIS project fail? Draw upon information from various chapters in this book.

Further reading

Harmon J.E. and Anderson S.J. 2003 *The Design and Implementation of Geographic Information Systems*. Hoboken, NJ: Wiley (e-book).

Heywood I., Cornelius S. and Carver S. 2002 *An Introduction to Geographical Information Systems* (2nd edn). Harlow: Prentice-Hall.

Obermeyer N.J. 2005. 'Measuring the benefits and costs of GIS'. In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley: 601–610.

Sugarbaker L.J. 2005. 'Managing an operational GIS'. In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken, NJ: Wiley: 611–620.

Tomlinson R. 2003 *Thinking about GIS: Geographic Information System Planning for Managers*. Redlands, CA: ESRI Press.

Thomas C. and Ospina M. 2004 *Measuring Up: The Business Case for GIS.* Redlands, CA: ESRI Press.

18 *GIS and management, the Knowledge Economy, and information*

It is not enough knowing how to obtain and run a GIS (Chapter 17). Potentially, GIS can now underpin virtually every decision made within and between organizations. Thus sustainable success only comes if the system as used makes a material contribution to the success of the entire enterprise; and any enterprise will only succeed if it is led and managed effectively. Therefore, effective leadership and management are crucial elements for successful GIS in government, business, 'not for profit' organizations, and academia.

The Knowledge Economy (KE – which includes GIS), and management that operates within it, is affected by the growth of networking (such as the Internet) and other technological changes. But many business principles remain unchanged, e.g., understanding the business drivers and how GIS can contribute to those that, locally, are the most important. Technology alone is only one contributor to success. To make a success of GIS, managers need to understand the Knowledge Economy, the role of innovation, what motivates people, and the unique characteristics of geographic information.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter, you will understand:

- Government, academia, 'not for profits', and commerce as 'businesses';
- Management and GIS;
- Business drivers;
- GIS and the growth of the Knowledge Economy and Knowledge Industries;
- The characteristics of geographic information as a commodity;
- The impact of the Internet and the Web upon GI and management;
- GIS as a growth business in its own right.

18.1 Are we all in 'managed businesses' now?

Most readers of this book live in a capitalist society. Our tangible wealth and physical well-being – and hence freedom to do many things – come increasingly from the exploitation of knowledge or skills. Much of this exploitation is underpinned by scientific or technological innovation. Increasingly, we see the breakdown of previously discrete sectors: commercial considerations impinge more widely and there is convergence between the activities and operations of commerce and industry, government, the not-for-profit sector and academia. Increasingly everyone is concerned with explicit goals (many of which share similar traits) and with 'knowledge management'. Moreover, in the GIS world at least, we also see significant overlap in functions – for instance, both government and the private sector are producers of geographic information (GI). The not-for-profit sector increasingly acts as an agent for or supplement to government and usually operates in a very business-like fashion. The result is that previous distinctions are becoming blurred and movement of staff from one sector to another is becoming common.

For these reasons, we use 'business' as a single term to describe the corporate activities in all four sectors. Accordingly, we think of the GIS world as being driven by organizational or individual objectives, using raw material (data or information: see Chapter 1), tools (geographic information systems) and human capital (skills, knowledge, attitudes, and approaches) to create new products or services (e.g., other information or knowledge) which help us to meet the objectives.

This commonality between the different sectors should not be exaggerated. For example, in publicly owned commercial organizations in the USA, pleasing the stockholders every quarter is a commonplace requirement. The lack of profitability and resulting low price of shares in Smallworld – a pioneer in GIS object orientation – led to its purchase by General Electric in fall 2000. Takeovers and mergers are also much more common in the commercial than the government sector – as in the purchase of the major US data firm GDT by the multinational Tele Atlas in 2004. But it is still realistic to regard all good organizations as business-like.

Geography is increasingly recognized as a central concern for all businesses. This book contains many examples of why this is so. In the real world, however, rarely is geography the *only* issue with which managers have to grapple. And rarely do managers have the luxury of being solely in charge of anything or have all the information needed to make truly excellent decisions. Box 18.1 gives a typical example of how geographic and non-geographic issues are intertwined and how many

Applications Box **18.1**

The geography of coping with catastrophe

On the evening of 21 May 2001, The City University in London suffered a disaster. One section of an 1894 building (Figure 18.1A), an icon for generations of students, was engulfed in flames (Figure 18.1B). The London Fire Service was soon there and pouring water on the inferno. As the night rolled on, a helicopter parked overhead sent images of the hottest areas to the firefighters; the desperate battle to contain the flames ebbed and flowed. Soon after midnight, the immediate crisis was over although the fire service did not leave until hours later.

The university was lucky. All 250 people in the building were evacuated safely. There was no wind that evening: had there been any, many more rooms and around a thousand exam scripts stored nearby would have been lost. The fire station was nearby, ensuring the fire tenders arrived within a few minutes of the alarm being

(A)

(B)

(C)

NEWS EXTRA

Evening Standard

LONDON, TUESDAY, 22 MAY 2001 www.thisislondon.co.uk Incorporating THE EVENING NEWS 35p

FIRE DESTROYS CITY UNIVERSITY

40-ft flames send timbers crashing to the ground

By Keith Poole and Jonathan Franks

FLAMES soar 40 feet high above City University in Clerkenwell, destroying a hundred years of history and spreading acrid smoke over a large part of London.

Firefighters battled to quell blazes believed to have started in the university's College Building. Police and students will run intensive evacuation and fire investigators attempt to find the cause of the blaze.

The investigation is being hampered by the fear that a 20-ft-high gable wall could collapse on top of the road. Any more crashing down after supporting timbers disintegrated in the inferno.

Vice chancellor David Rhind said: "This is a disaster for us. Our magnificent listed building is in tatters."

A London Fire Brigade spokesman said: "We are keeping the area sealed off. Crews are standing by but it is unsafe to let them back in as timbers have

Continued on Page 2

M&S profits fall again – but it could have been worse

MARKS & SPENCER today announced profits of £400 million, a seven per cent drop from last year – although this was not as big a fall as City analysts had expected.

The company has been forced to take one-off charges against profits of more than £350 million to cover the costs of its 6,000-plus redundancy programme announced in March.

By Joanne Hart Deputy City Editor

Chairman Luc Vandevelde, who waived an £816,000 bonus after intense internal and external pressure, admitted a recovery would take longer than expected. He said: "Sales of adult clothing during the period were unacceptable and it is this that led us to take many of the tough but necessary actions required to build a platform for recovery. The results will not be seen overnight, but our customers will see a gradual and progressive improvement."

Marks & Spencer has suffered particularly in its core clothing business. Sales fell 5.5 per cent

Continued on Page 2

Figure 18.1 The City University fire of 2001: an example of geographically-related management decisions taken by multiple players (mostly) in concert. (A) College Building before the fire; (B) College Building in flames; (C) the headlines in the next day's London newspaper, replicated by global TV coverage (courtesy Evening Standard)

raised. And there were no other major fires in central London that night so much assistance was on hand.

In the immediate aftermath of the fire – the next morning – both normal operations (including exams) and planning for recovery had to be managed. The geographically-related decisions that were made – some macro-level and some micro-level – by various players and processes included the following:

- The evacuation of students and staff by security staff through safe corridors, necessitating good public records (and staff awareness) of the intricate internal geography of the building.
- The speedy arrival of the fire tenders using address/street network-based routing systems.
- The closing of local streets and re-routing of traffic by the police.
- The mobilization of other suitably located fire tenders on a risk-assessment basis from across London to help fight the fire.
- The telemetering of thermal imaging of 'hot spots' from helicopter to fire fighters.

- The mapping of the areas damaged by fire, flood, or smoke the next morning to quantify the scale of the problem and act as the basis for immediate management action. This included surveys for newly revealed hazards such as exposed asbestos.
- The immediate appropriation of suitable surviving space for examinations and the setting up of posters to guide students to the new venues in good time.
- The search by commercial firms for suitable replacement office space within an acceptable distance, using their GIS and personal knowledge.
- The evaluation of temporary accommodation offered by other universities to test the feasibility of its use given geographical proximity, etc.
- The decisions on which areas of the world to concentrate advertising, given the huge publicity the fire received in the media (including CNN, broadcast worldwide). Their message was that the university had been destroyed (Figure 18.1C). This encouraged overseas students to believe they should switch to another university – which would have led in turn to huge loss of income.
- The progressive (and beneficial) restructuring of space within the university over the next three years, triggered by the fire and carried out partly on space-planning software.

separate groups make decisions, often in parallel, which impact on others.

18.2 Management is central to the successful use of GIS

18.2.1 Some basics

For many people, management implies a dull, routine job ensuring processes are followed and production targets are met. Today's manager, however, is required to anticipate future opportunities and possible disasters and take action appropriately to change the local world. He or she has to keep up-to-date with changes in organizational aims and mores – and help to shape them. Persuading colleagues and staff to give of their best and ensuring targets are achieved is essential. Such pro-active management is often more properly titled 'leadership'; and it takes place at all levels in organizations – from the highest (Box 18.2) to almost the lowest.

Management matters; luck and idiocy on the part of one's competitors usually only plays a small part in what happens. Though necessary, excellent science and technology – however good – are not sufficient conditions for success. Microsoft did not succeed simply because it produced good software. It succeeded because the organization had great management and smart people. Its leaders had a vision of what they wanted to achieve, ideas on how to do it, and superb marketing skills. They also had the ability to reprioritize and reformulate plans and activities as often as necessary. All these management abilities make the difference between success and, at best, mediocrity. It follows that business awareness and flexibility of the staff and flexibility of the GIS tools used are crucial.

There are some things which the manager can rely upon. These include:

- *People cause more problems than technology*.
- *Things change increasingly rapidly* – both the technology and the user expectations.
- *Complexity is normal but some of it is self-imposed* – and all organizations tend to believe they are unique and complicated.
- *Uncertainty is always with us*.
- *Inter-dependence is inescapable*. Creation and implementation of a GIS often have far wider ramifications than originally imagined and hence impacts on many more people, at least indirectly.
- *Customers for a system often have very fixed ideas of what they need* yet these often change by the time a system appears.
- *Detailed specifications of needs are usually seen as necessary* to hold people (contractors, managers, etc.) to account. But these then complicate (and make expensive) changes to a project.
- *There are considerable differences in national or even regional cultures*. These can have significant impacts in running GIS projects in other countries (see Box 18.3).

Management is not 'rocket science' – almost anyone can be competent at it with good training, enough practice, an understanding of 'the big picture', and a modicum of understanding of people. But it is not a trivial pastime – far from it. Few people are good at it all the time. Getting good results normally requires unremitting effort, intelligence, and an ability to welcome criticism and adapt behavior appropriately (see Box 18.3).

Management is not 'rocket science' but neither is it easy. Everyone is a manager at some stage.

Biographical Box **18.2**

Suwit Khunkitti: Deputy Prime Minister of Thailand

Figure 18.2 His Excellency Suwit Khunkitti, pictured at a UN Conference in Geneva in 2004 on 'Hazardous Chemicals and Pesticides in International Trade'

By background a chemist trained at the University of Kentucky, His Excellency Suwit Khunkitti is a strong advocate of GIS. This is derived in part from huge experience in the government of his country. A member of the Thai Parliament since 1983, he has at various times been Minister for Agriculture and Cooperatives, for Justice, for Natural Resources and Environment, and for Science, Technology and Environment. In addition, he chairs the National Research Council of Thailand, the Information and Communication Technology Committee, the National Water Resources Committee, the National Environment Board – and the National Geographic Information System Committee. Amongst many other roles, he has also been Chancellor of the Law Society of Thailand.

Kunkitti says:

As a national executive, I envision the development of information technology as fundamental for providing solutions to national challenges and for our ability to ensure the sustainable growth of the nation. Such development requires databases which are correct and must always be updated. It is vital that government agencies apply contemporary information technology to their day-to-day operations. GIS in particular is an information technology that is essential for *all* aspects of national administration. I have laid the strong foundations for the national GIS development through establishing the national GIS committee and ordering the NSDI study. These actions will allow all organizations in Thailand to develop efficient GIS systems of their own. Already GIS has been used widely in Thailand, including applications to solve problems related to the management of natural resources and the environment. GIS have been used for city planning, public utility services, management of agriculture, flood control, to support the drought protection plan – and much else.

18.2.2 Some GIS-specific challenges for managers

These arise from specific characteristics of GI or of the nature of the industry; these are discussed later. But, as an illustration of the specific challenges, consider:

- There are many different ways of representing the world (Chapter 3) in GI form. How this is done influences what can be achieved and the quality of the results obtained.
- GI is fuzzy in that each and every element of it has associated imprecisions (Chapter 6); mixing data of different accuracies can lead to big problems.
- Our techniques for describing 'data quality' are still primitive. Thus assessing its 'fitness for purpose' is often not simple, especially if the metadata (Section 11.2) are inadequate.

Applications Box **18.3**

Cultural and political consequences of geographical errors

An illustration of the consequences of getting things wrong was given publicly by the senior geopolitical strategist at Microsoft during the four-yearly International Geographical Congress in 2004. This candor is refreshing: all organizations get some things wrong but few admit it so others can learn from their mistakes.

When Microsoft released Windows 95, the time zone map covering India did not include Kashmir, the disputed border area between India and Pakistan. Though only a few pixels in size, this contravened Indian law. It led to the enforced withdrawal of all 200 000 copies of that system and the supply of a replacement, costing the firm millions of dollars. The solution adopted was to degrade the resolution of the map so that individual countries would not be identified.

The map of Turkey in Encarta originally contained the word 'Kurdistan', denoting the cultural homeland of the Kurds. The Turkish government, which has been in armed conflict with Kurdish separatists, arrested Microsoft representatives and interrogated them. Eventually, the name was removed.

Source: as reported by Michael McCarthy in *The Independent* of 19 August 2004

Another embarrassing geographic factual error was the loss of Wales from a map of Europe on the cover of the *Eurostat Statistical Compendium* for 2004 (Figure 18.3). That Europe's statistical agency missed this off caused a political furor. The moral is clearly that a 'heavy-duty' quality control mechanism and highly aware management are essential in encoding geographic data, especially those with political overtones.

Figure 18.3 A bureaucratic blunder left Wales off a map of Europe on the cover of the Eurostat Statistical Compendium for 2004 (Reproduced by permission of European Communities)

- In some cases, combining data gives rise to an 'ecological fallacy' which can produce false correlations between the variables (see Section 6.4.3).
- There remains a general lack of awareness of the value of GIS which means that GIS managers need to be evangelical.

18.2.3 Attitudes towards management

The idea that GIS can be implemented and run simply by following 'cookbook' instructions and that success will inevitably follow is quite wrong. It is contrary to all management experience in projects based on the implementation of high technology. It has been argued there are three different managerial approaches relevant to IT in general and GIS in particular. These can be described as technological determinism, managerial rationalism, and social interactionism. We consider these to be part of a spectrum and consider only the end points, the first and third approaches.

Successful GIS implementation and use require more than just technical solutions.

Technological determinism is a Utopian approach which stresses the inherent technical merits of an innovation. The approach can be recognized in those GIS projects which are defined in terms of equipment and software. The project is usually sold to potential users on the basis that 'There Is No Alternative (TINA); this will do everything you need and has lots of intangible benefits because of the capacity of the technology'. Relatively little emphasis is placed on the human and organizational aspects compared to fine-tuning the software and meeting the detailed technical specification.

Social interactionism involves a formal recognition of uncertainty on how organizations really work and a belief that knowledge (see Section 1.2) and culture within the institution have big impacts on the success of IT projects. In this view, decision making is an interactive, iterative, and often fast-changing process between individuals and groups – both within the organization and without – which are sometimes in conflict and sometimes in collaboration. Success comes from placing stress on the organizational and user acceptance and use of the technology, rather than on the intrinsic merits of the system.

Given the nature of contemporary business life, normally only some version of the social interactionism approach works well (though in war-time other approaches may operate successfully). In practice, of course, most organizations contain people of both persuasions so managers have to ensure they function well together (see Box 18.4).

18.2.4 Business drivers

Overall, we can think of the imperatives that drive *all* businesses (falling within our wide definition) as being:

- Saving money or avoiding new costs (e.g., due to new regulations).
- Saving time (which *may* also involve saving money but sometimes may not, e.g., in emergencies).
- Increasing efficiency, productivity, and/or accuracy, and aiding budgeting.
- Creating new assets, e.g., intellectual property rights, enhanced brands, or trust through new investment.
- Generating additional revenues or other returns from the identification, creation, and marketing of new products and services by exploitation of the assets.

Biographical Box **18.4**

Hugh Neffendorf: a 'can do' manager and expert consultant

Figure 18.4 Hugh Neffendorf: manager and consultant

Hugh's consultancy firm, Katalysis, is so named because of its focus on facilitation, bringing appropriate sets of skills to projects, and getting things done. He has had wide experience in the private, public, and academic sectors and has acted as a consultant on GI for many British public sector organizations and for a number of commercial organizations.

Transplanted in his youth from Texas to Northern England, his involvement in geographic information stretches back over 30 years. It began mainly in the field of transportation planning where he headed a large division of a major multi-national consultancy firm. He then applied his experience to the wider GI arena, leading many strategic and technical projects, and taking responsibility for several software and data product developments, including those creating population census software, map data files, and ADDRESS-POINT™, the standard British data source for address geography.

Hugh's involvement in projects is unusually varied, including areas as diverse as strategy, business management, marketing, product development, information technology, mathematical modeling, and statistics – and dealing with clients. Nowadays, his most usual role is as the leader of technical teams and as facilitator of working groups. Many of his projects relate to official statistics: he is facilitator of the UK Office for National Statistics Geography Advisory Group, an external national 'think tank' that helps to guide statistical geography policy.

Hugh leads the UK Association for Geographic Information's (AGI) Special Interest Group on Address Geography. Fitting with his view (and that of the authors of this book) that work should be fun as well as effective, he also helps to organize the annual AGI Golf Day!

- Identifying risk and reducing it to acceptable levels.
- Supporting better-informed and more effective decision making in the business.
- Ensuring effective communication with key stakeholders.

Every organization now has to listen and respond to customers, clients, or fellow stakeholders. Every organization has to listen to citizens whose power sometimes can be mobilized successfully against even the largest corporations. Every organization has to plan strategically and deliver more for less input, meeting (sometimes public) targets. Everyone is expected to be innovative and deliver successful new products or services much more frequently than in the past. Everyone has to act and be seen to be acting within the laws, regulatory frameworks, and some conventions. Finally, everyone has to be concerned with risk minimization, knowledge management, and protection of the organization's reputation and assets.

All organizations must be responsive to the needs of customers and other stakeholders and must demonstrably be effective. This is achieved through good management, rather than luck or individual genius.

These drivers have different importance in each sector. In addition, the ambitions of and drivers for *individuals* within the organization cannot be ignored, especially in enterprises dependent on clever, highly marketable people (such as some universities and GIS firms). These factors are incorporated and expanded in Table 18.1.

Understand motives and you understand why many things happen.

The relationship between the different sectors also shifts over time. Until the early 1980s, much of the available GIS software was produced by government or by individuals. Only with the arrival of significant commercial enterprises – and hence real competition – has GIS become a global reality, with the number of users at least ten thousand times greater today than it was in the 1970s. Relatively little general-purpose and widely used GIS software is now produced outside the commercial sector, one exception being the IDRISI software from Clark University (see Box 16.6). Also, in part, commercial assemblers and sellers of geographic information have taken over a role traditionally associated with governments.

All organizations and their employees are subject to similar needs and incentives, though what is most important varies locally and over time.

The context in which you manage is rarely constant for long so we now describe the changing world of knowledge creation and exploitation – a world in which GIS is prospering.

Table 18.1 Some business drivers and typical responses. Note that some of the responses are common to different drivers

Sector	Selected business drivers	Possible response	GIS example
Private	Create bottom-line profit and return part of it to shareholders. Build tangible and intangible assets of firm. Build brand awareness.	Get first-mover advantage; create or buy best possible products, hire best (and most aggressive?) staff; take over competitors or promising startups to obtain new assets; invest as much as needed in good time; ensure effective marketing and 'awareness raising' by any means possible; reduce internal cost base.	Purchase and exploitation by Autodesk of MapGuide software and subsequent development – one of earliest Web mapping tools. Engagement of ESRI in collaboration with educational sector since c.1980 – leading to 80% + penetration of that market and most students then becoming ESRI software-literate. Purchase of GDT Technologies by Tele-Atlas to obtain comprehensive, consolidated USA database and to remove major competitor.
	Control risk.	Set up risk management procedures, arrange partnerships of different skills with other firms, establish secret cartel with competitors (illegal)/gain *de facto* monopoly. Establish tracking of technology and of the legal and political environment. Avoid damage to the organization's 'brand image' – a key business asset.	Typically, GIS firms will partner with other information and communication technology organizations (often as the junior partner) to build, install or operate major information technology (IT) systems. Many GIS software suppliers have partners who develop core software, build value-added software to sit 'on top', act as system integrators, resellers, or consultants. Data creation and service companies often establish partnerships with like bodies in different countries to create pan-national seamless coverage. Avoid partnerships with 'fly by night operations'. Know what is coming via network of industry, government, and academic contacts.
	Get more from existing assets.	'Sweat assets', e.g., find new markets which can be met from existing data resources, re-organized if necessary.	Target marketing: use data on existing customers to identify like-minded consumers and then target them using geodemographic information systems (see Section 2.3.3).
	Create new business.	Anticipate future trends and developments – and secure them.	Go to GIS conferences and monitor developments, e.g., via competitors' staff advertisements. Buy start-ups with good ideas (e.g., ESRI and the Spatial Database Engine or SDE). Network with others in GI industry and adjacent ones and in academia to anticipate new opportunities.
Government	Seek to meet the policies and promises of elected representatives or justify actions to politicians so as to get funding from taxes.	Identify why and to what extent proposed actions will impact on policy priorities of government and lobby as necessary for tax appropriations. Obtain political champions for proposed actions: ensure they become heroes if these succeed.	Attempts to create national GIS/GI strategies and NSDIs (see Chapter 20). Impact of President Clinton's Executive Order 12906. Force the pace of progress on interoperability to meet the needs of Homeland Security (see Chapter 20).
	Provide good Value for Money (VfM) to taxpayer.	Demonstrate effectiveness (meeting specification, on time and within budget) and efficiency (via benchmarking against other organizations).	Constant reviews of VfM of British government bodies (including Ordnance Survey, Geological Survey, etc.) over 15 years, including comparison with private sector providers of services. National Performance Reviews of government (including GIS users) in USA and the impact of the 2002 e-government initiative.

Table 18.1 (*continued*)

Sector	Selected business drivers	Possible response	GIS example
	Respond to citizens' needs for information.	Identify these; set up/encourage delivery infrastructure; set laws which make some information availability mandatory.	Setting up of US National Geospatial Clearinghouse under Executive Order 12906 plus equivalent developments elsewhere (Chapter 20). Subsequent development of hundreds of GI portals worldwide.
	Act equitably and with propriety at all times.	Ensure all citizens and organizations (clients, customers, suppliers) are treated identically and that government processes are transparent, publicized, and followed strictly.	Treat all requests for information equally (unless a pricing policy permits dealing with high-value transactions as a priority). Put all suitable material on Web but also ensure material is available in other forms for citizens without access to the Internet.
Academic	Establish high research reputation.	Carry out, publish, and disseminate results of new research, advancing field. Win big research grants from prestigious sources (peer-reviewed academic or science sources normally give more prestige than those from industry or government). Run conferences attended by top-class participants, form informal 'club'. Win prizes.	Focus research on things that business does not do or cannot talk about for fear of losing competitive advantage, i.e., fundamental research (often technical), 'soft, human-focused' research or policy-relevant research. Set up research centers, get funding to attract graduate students and other researchers. Publish – sometimes with overseas colleagues – in top journals in the GIS-related field (Box 1.5), including some (but not too many) in journals outside of field. Talk to people in different disciplines to get new ideas. Attend many conferences; accept keynote speech invitations.
	Establish reputation for teaching excellence and relevance. Build international links to anticipate globalization of learning.	Provide top-class educational experience for students (increasingly important). Alumni are the best form of advertising in the long term.	Recognize that GIS principles and technology are near-identical worldwide so: • Lead/build consortia to create curriculum materials, deliver identical courses worldwide (e.g., UNIGIS model – see Box 19.4) and/or develop state-of-the-art courses which can be run on a residential basis. • Differentiate courses from others in ways suited to the local market needs. Pro-actively seek overseas students to build international base.

18.3 The Knowledge Economy, knowledge management, and GIS

Different people mean different things by the term 'Knowledge Economy'. Central to it, however, is the use of new knowledge to run things more effectively (e.g., local democracy) and/or to create still-newer knowledge and exploit it. The nature of the Knowledge Society is dynamic – knowledge is constantly being added to so keeping up to date is crucial.

18.3.1 Innovation matters

Continuous innovation and creativity are fundamental to the health of the Knowledge Economy. The British Science Minister Lord Sainsbury summarized their importance by saying:

> **We define innovation as 'the successful exploitation of new ideas'. Often it involves new technologies or technological applications. . . For consumers, innovation means higher quality and better value goods, more efficient services (both private and public) and higher standards of living. . .The innovative company or organisation delivers higher profits for its owners and investors. For employees, innovation means new and more interesting work, better skills and higher wages. Equally, an absence of innovation can lead to business stagnation and a loss of jobs.**

For the economy as a whole, innovation is the key to higher productivity and greater prosperity for all. Innovation will also be essential for meeting the environmental challenges of the future – including moving to a low carbon economy and reducing waste. We need to find new ways to break the link between economic growth and resource depletion and environmental degradation.

Source: DTI (2003) *Competing in the Global Economy*, London, Department of Trade and Industry.

Worldwide, there is a gathering pace in commercial developments through such innovation of new products, processes, and services, and in the rapidity of their dissemination and acceptance on a national or worldwide basis (see Figure 18.5). Inventive entrepreneurs in Hong Kong take for granted the truth of the phrase 'If it works, it's obsolete'. The life cycle of new products, services, and even of knowledge itself is becoming ever shorter under competition which is ever more global. This and many other business developments have been facilitated greatly by scientific and technological developments, especially in the information and communication technologies (ICT); but these are not the only factor, and technology has not always done what was anticipated (Box 18.5).

This escalation of innovation is most rapid in the 'Knowledge Industries' where substantial national and business revenues may be at stake through licensing of new processes or products (Figure 18.6). It has long been recognized that these industries will be the key to future national and business success. As an example, the World Bank's *1998 World Development Report* said:

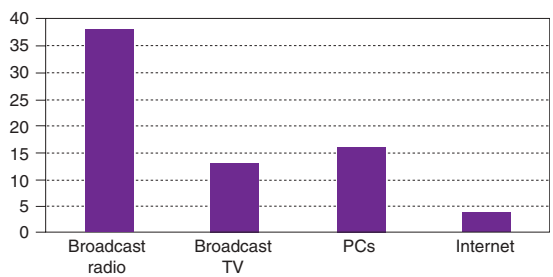


Figure 18.5 Increased speed of up-take of innovations, shown by the years needed for an innovation to be accepted by 50 million Americans (*Source*: US Department of Commerce)

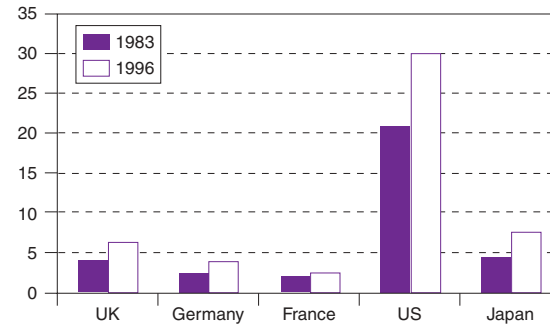


Figure 18.6 Knowledge Economy revenues from overseas earnings on royalties and license fees (US$bn). (*Source*: 'UK Competitiveness White Paper 1998'. © UK Government, Department of Trade and Industry) (Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland)

For countries in the vanguard of the world economy, the balance between knowledge and resources has shifted so far towards the former that knowledge has become perhaps the most important factor determining the standard of living. . . Today's most technologically advanced economies are truly knowledge-based.

GIS is involved inescapably in this innovative gallop and, irrespective of the definitions used, is part of the Knowledge Industries. All the major software vendors have to operate on a continental or global scale: there are enough commonalities of user need – at least at the tool box level – to create huge economies of scale by operating as part of a global business. Each year GIS vendors announce new functionality or new applications for their software; and multi-national mergers and acquisitions in the GI business are no longer unusual.

Those active in GIS are working in the knowledge industries and contributing to the Knowledge Economy; continuing innovation in GIS is strong and crucial to our success.

Applications Box **18.5**

Globalization, technology, and unintended consequences

Globalization is not only a factor of technology: it has been driven as much by worldwide trade liberalization and the removal of restrictions on cross-border capital movements. Moreover, technology effects are not always predictable or benign: some have had unintended consequences. Deaths from hurricanes in the USA have dropped from six thousand a year in 1900 to a few dozen thanks in large measure to radar and other tools to track and predict storms (see Figure 2.16, Box 3.5). Yet insurance payouts have soared – $7.3 billion after Hurricane Andrew – in part because people felt safer in building in hurricane-prone areas.

18.3.2 Putting a value on knowledge and information

Since knowledge – however it is defined – is becoming so crucial, it is no surprise that investors increasingly recognize the growing importance of knowledge assets in the way they value enterprises. Box 18.6 demonstrates one approach to categorizing knowledge in commercial enterprises. However, intangible assets occur at least as frequently in government-based organizations, although valuing the equity in such cases is even more complex. What should be counted as intangible and what is tangible are sometimes of great importance in GIS – as shown by the debate over how to account for the National Geographic Database in the accounts of Ordnance Survey, Britain's national mapping agency. The Auditor General of the UK's National Audit Office made the following comment in his report to Parliament on the OS accounts for 2003/04:

> **Ordnance Survey's turnover of £116 million derives principally from the exploitation of data held in Ordnance Survey's geospatial databases the creation of which has been funded from public monies over many years. . . the Agency has not capitalised the costs of setting up and maintaining the data held in the geospatial databases in its Balance Sheet. In the Agency's view, the data [are] an intangible fixed asset that does not meet the conditions for capitalisation set by Financial Reporting Standard 10. In my opinion, the data held in the geospatial databases [are] a tangible fixed asset that should be capitalised in accordance with Financial Reporting Standard 15. Having taken expert advice about the valuation of the data held, in my view the value to the business is not less than £50 million. Had the data been capitalised at that value, the effect would have been to increase tangible fixed assets included in the Balance Sheet at 31 March 2004 from £36 million to £86 million.**

Also, it is obvious that intangible assets may be negative (e.g., where brand image is destroyed by loss of public acceptance of the quality or safety of data or software products or services provided by the commercial firm or government body). Intangible assets are present in extreme form in academia or small businesses where the poaching of one key member of staff or a team may destroy the reputation of the organization in that field overnight.

The most direct way to put a value on information is to seek to sell or lease it. This is done by many commercial firms and some governments (see Section 19.3.2).

18.4 Information, the currency of the Knowledge Economy

Information, then, is a key ingredient of the Knowledge Economy and Knowledge Industries and is the 'rocket fuel' for GIS. In this section, we discuss several different aspects of the role and characteristics of information in general and geographic information (GI) in particular. We

Applications Box **18.6**

Valuing knowledge

In fast-growing sectors like biotechnology and computer software (including some parts of GIS), a large part of the value of the company resides in the knowledge embodied in its patents and in its staff and, in turn, how the outside world sees these being used and developed to capitalize on new opportunities. A classic example is the huge growth in the difference between Sun Microsystem's stock market value (what people would pay for its stocks) and its book value (based on normal audit of its physical assets and liabilities) in 1995–96 because of the announcement of Java; this represented a major increase in its intangible assets. One way of categorizing the different types of intangible assets is shown below:

Visible equity (book value)	Intangible assets (stock price premium)		
	External structure	**Internal structure**	**Individual competence**
Tangible assets minus visible debt	Brands, customer and supplier relations (e.g., 'trust')	The organization: governance, management, legal structure, manual systems, attitudes, R&D, software	Education, experience, skills

start with how information underpins decision making. Along with projected efficiency improvements and risk moderation (see Section 18.2), this is the basis on which many GIS have been procured in the past.

18.4.1 Information and knowledge – the basis for good decision making?

Until now, we (and many others) have assumed that having more knowledge and information gives more power and the capability to make better decisions. As authors, we hold to the view that better information often *does* permit better decisions to be made. But the reality is usually much more complicated, especially in GIS – in part this is because it is rarely the *only* important input to decision makers (Section 18.1). Many decisions require the balancing of conflicting objectives and judgments about the trade offs, as known in the light of available knowledge at that moment in time. Thus the role of 'raw information' in shaping the running of the world can easily be exaggerated: managers are not information processors, with the 'right decisions' simply falling out of an analytical process. Five commonplace yet erroneous assumptions about management and information are:

- *More information leads to better decisions.* In reality, most managers increasingly suffer from information overload, especially of irrelevant information. Mapping and geovisualization facilities (Chapters 12 and 13) provided by GIS can help to reduce the problem greatly but rarely cure it completely.
- *Managers need the information they want.* Managerial 'wish lists' often do not match to what is actually needed to run the business.
- *Managers are able to model the decision they wish to make.* In the real world it is unusual for a clear model to exist which identifies the relevant variables and their inter-relationships.
- *Managers do not need to understand the information system.* In reality, understanding the strengths and weaknesses of a GIS providing evidence is key to informed management decisions – GIS are not simply 'black boxes' producing 'objective answers'. GIS managers use judgment, operate with certain ethical standards, etc. (see Section 19.2.2)
- *Good information systems lead to enhanced communication.* In practice, the internal structures of the organization may vitiate this potential advantage. Too much information may also clog up responses.

Chapter 2 introduced the application-driven nature of much GIS. But from an institutional perspective, successful use of GIS requires an understanding of the aims, objectives, and ways of operating of the organization and focusing on what is most essential. Given the business drivers set out in Section 18.2.4, the potential benefits of GIS for managers of big projects, or even for corporate bosses, are in creating or clarifying knowledge through:

- Providing 'factual information' about the location of resources, human or natural (*a facilitation process*).
- Computing derived 'facts' (see Box 18.7), e.g., fastest routes through a network, population within five miles of a toxic waste dump, or on what has changed since

Applications Box **18.7**

What are geographic 'facts'?

GIS is seen by many as creating geographic 'facts'. There is, however, a great academic argument about what constitutes 'facts' and to what extent these are universally true or are context-dependent (see Section 1.7). Therefore, managers need to be aware of what can be said to be 'true' in reporting results from GIS to their bosses.

Geographically-referenced statistics are often cited as facts, even if they are based upon measurements of variable quality or where the very concept of a single definitive answer is suspect. For example, the CIA *World Factbook* **(www.cia.gov/cia/publications/factbook/geos/sf.html#Geo)** claims that the South African coastline is 2798 km in length (see Box 4.5 on the measurement of length and the fractal nature of many geographic phenomena).

It is sometimes argued that the outline of a building, the bounds of a land parcel, or a line of constant elevation on a map are facts because any other person or sensor attempting to represent these facts would do it in much the same way. This completely neglects the cartographer's art and, in particular, generalization and different possible geographical representations of the same phenomena as stored in different GIS (see Section 3.8 and Chapter 8).

Fortunately, some sources of 'facts' are more circumspect: the GeoExplorer website **(www.geoexplorer.co.uk/sections/geofacts/geofacts.htm)**, for instance, actually includes the injunction *'remember to question the accuracy of the GeoFact before you use it'!*

last year (*a facilitation process, producing new information*).

- The selection, compression and visualization of information to facilitate decision making (*information filtration*).
- The search for regularity and possible causality: looking for patterns and correlates of geographic distributions, e.g., does spending on certain goods coincide with particular ethnic groups or distance between shops and home? Chapter 2 gives more examples (*information filtration, leading to evidence*), and Section 4.7 examines methods.
- The creation of added value by linkage of information from different sources. In general, such linkage extends the range of applications possible – with two information sets for a given region, we have one combination; with 20 sets, we have over a million combinations in total (*information fermentation, leading to more information, evidence, and possibly knowledge*: see Section 1.2).
- Predicting future events which are geographically distributed (*evidence or knowledge-based operations*). This requires both initial 'state variables' and process models (see Chapter 16).

So, under what circumstances is GI valuable or even vital for decision making? We answer this by citing lots of examples throughout the book. But a better way involves understanding of the characteristics of information, how it is valued, and the sorts of processes which now work well through its use – especially those that rely upon use of the Internet which is now the main mechanism for sharing GI.

18.4.2 The characteristics of information in general

We can think of the very same information being used for two different purposes and thus being valued very differently by the same people in different circumstances. The information may be used:

- *For consumption*. Individuals decide how much value they assign to it based upon their valuation of pleasure, time saving, or some other metric, and their awareness of the uses and the potential benefits of the information.
- *As a factor of production*, where the information is used as part of a good or service. The end-user of that good or service will make his or her decision on the uses to which it can be put and hence its value to him or her. Value will be affected by its availability to others (it will tend to be more valuable if it is not available to others, such as competitor organizations) and the ease of substitutability. Actually measuring these values is rather difficult, so discussion thus far has typically been at a very high conceptual level (but see Box 17.4 on cost/benefit analysis).

Information, from a management perspective, has a number of unusual characteristics as a commodity. In particular, it does not wear out through use (though it may diminish in value as time passes; for example, population census data from ten years ago are less reliable for current policy-making than equivalent data from last year).

It is usually argued that information is in general a 'public good'. A 'pure' public good has very specific characteristics:

- The marginal cost of providing an additional unit is zero. Thus, in effect, copying a small amount of GI adds nothing to the total cost of production.
- Use by one individual does not reduce availability to others ('non-rivalry'). This characteristic is summarized in the famous Thomas Jefferson quotation that: 'He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me'.
- Individuals cannot be excluded from using the good or service ('non-excludability').

A pure public good is a special form of externality. Such externalities arise where production or consumption of a good (e.g., information) by one 'agent' imposes costs on or delivers benefits to other producers or consumers. In this case, provision of a pure public good for one group of users will benefit other potential users because it is not possible to exclude them. Pollution of water or air and by ambient noise form classic examples of external costs (Figure 18.7); refuse collection, education, and public health are classic examples of external benefits.

In practice, information is an optional public good, in that – unlike defense – it is possible to opt to take it or not; you may not feel the need to avail yourself of US Geological Survey data, for example! Moreover, the accessibility and cost of the systems to permit use of information influence whether, in practice, it is a public good in any practical sense. Since maybe only 0.03% of the world's population has ready access to GIS tools, scarcely can it be argued that geographic information in digital form is yet a universal public good in the traditional sense. Finally – to be pedantic – it may also be best to define information as a quasi-public good since it may be non-rival (see above) but its consumption can in certain circumstances be excluded and controlled. The business cases and vast investments of a number of major commercial GI purveyors such as Space Imaging Inc. and DigitalGlobe (Table 1.4) are based on this proposition; if everything they produced could be copied for free and re-distributed at will by anyone, their business would be untenable. Thus the pecuniary value of information may well depend on restricting its availability whilst its social value may be enhanced by precisely the opposite approach.

Traditionally, geographic information has been seen as a type of public good but its characteristics differ in different countries and in different sectors of the economy.

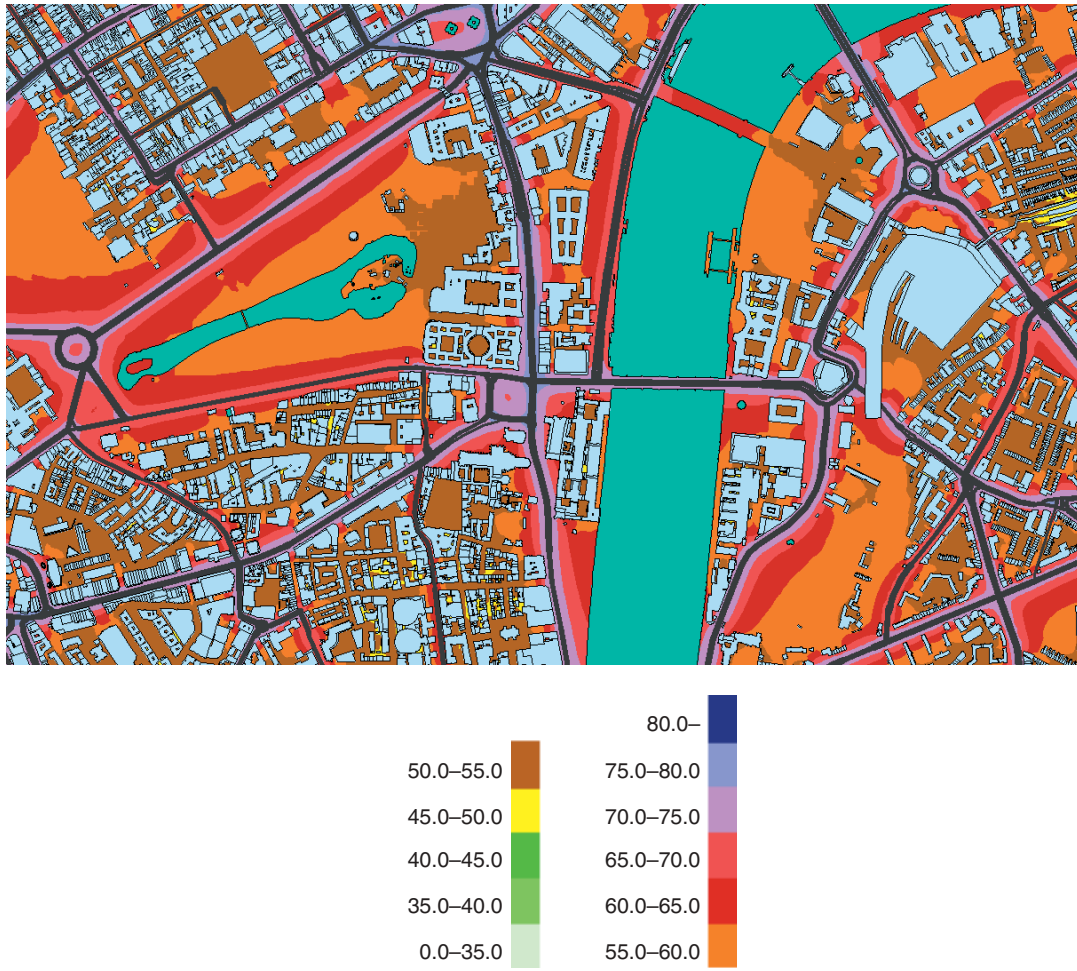


Figure 18.7 The geography of a negative externality: patterns of noise pollution in decibels east from Buckingham Palace (extreme left) in London (*Source*: Department for the Environment, Food and Rural Affairs. Crown copyright reserved)

What is certainly true, however, is that information in digital form – however expensive it is to collect or create in the first instance and even to update – can be copied and distributed via the Internet at near-zero marginal cost. Its widespread availability has many positive (and some negative) externalities which arise far beyond the original creator. In regard to information, the types of externalities can be summarized as:

- Ensuring consistency in the collection of information creates *producer externalities*. Inconsistency can raise the costs of creating and using data and limit the range of applications.
- Providing users access to the same information produces *network externalities*. Network benefits exist for any user of software or data being widely used by others (e.g., Microsoft software). It is, for example, important that all the emergency services use the same geographic framework data.
- Promoting efficiency of decision making generates *consumer externalities*. One example is where access to consistent information allows pressure groups to be more effective in influencing government policy and in monitoring activities in regard to pollution.

Often, a set of information is also an 'experience good' which consumers find hard to value unless they have used it before. Finally, there is no morality about information itself: often, its source (especially GI) can be readily disguised – for example, by changing the representation of roads from single lines to cased ones, changing their color, removing some sinuosity, or computing tables of distance between places. Provenance of data is an important characteristic both because some data producers are trusted more than others and because information theft can undermine businesses.

18.4.3 The characteristics of GI

GI has the standard characteristics of information described above but also some others that mark it out. For instance, central to the perceived value of GIS is that it can be used to integrate information for the same region

derived from many different sources. Such data linkage has at least three potential benefits:

- it permits infilling of missing data, thereby cutting the cost of data collection in some cases;
- it facilitates verifying the quality of individual datasets through a check on their consistency;
- it provides added value, almost for nothing. The number of permutations rises very rapidly as the number of input datasets rises (see Section 18.4.1 and Box 18.8). Hence, when datasets are linked together, many more applications can be tackled and new products and services created than when they are held separately.

Although this presents a colossal advantage, it brings associated problems. Generally, we have to take data 'as they come' from other people. Their classification or level of accuracy may not be what you would like in an ideal world (see Section 6.2.3). There is also a risk in that the results may be provably wrong – and lead to humiliation or being sued (being proved wrong is much easier than being proved right in GIS, just as in most scientific endeavors: see Chapter 6 on uncertainty). Taking information from multiple sources can lead to later conflict over the ownership of products if prior agreements have not been forged.

User attitudes to GI are often that it is, or should be, a free good (provided free or at very low cost and without restriction on its use) as well as a public good. This perception arises from two sources – it has been historically provided free or at low cost by governments to taxpayers and some detailed geographic information has the characteristics of a natural monopoly. The latter effectively means that the first person with the infrastructure wins because normally others can never justify duplication of facilities. Obviously, such monopolies are antipathetic to competition, which demands multiple choices. The obvious parallel for such GI is with telephone lines before the advent of cellular telephones. For instance, assume that only one organization has a complete set of information for every house in the country, that it is unwilling to share it freely, and that the majority of that organization's costs are sunk ones (i.e., they have been spent and cannot be recovered). In such circumstances – not uncommon even now in some parts of the world with government-held information – it is unlikely that the collection and sale of duplicate information can be made competitive.

Some geographic information has the characteristics of a natural monopoly.

Perhaps surprisingly, some competition may not be in the national interest so far as geographic framework information is concerned: it can give rise to negative externalities whilst a monopoly can produce network externalities (Section 18.4.2). Consider, for instance, the situation where information drawn from different sources is brought together and that information was originally compiled on different frameworks or base maps. Even where spatial coincidence occurs in the real world, this would not occur in the GIS because of differences in generalization, resolution, etc., between the two frameworks (see Chapter 5). Putting a value on such externalities is difficult, especially since the costs are visited on those at the end of supply chains and are often only discovered long after analysis of the data has led to mistaken conclusions. This provides great opportunities for lawyers! But it does demonstrate that sometimes there can be a public interest in a natural monopoly in GI in regard to everyone using the same framework or base map – how effective that monopoly is would depend on how well the situation was regulated.

Another factor in making GI special is the relative lack of transience of much traditional geographic information, which is analogous to 'reference books'. Typically, the ground features in topographic maps only change slowly except in very small areas. Satellite imagery of the same area may change much more because of crop change, imperfect conflation of the images sensed at different times, and the application of any classifier used – but any particular business may very well be completely uninterested in this change for its own particular purposes and therefore one purchase per decade may be adequate.

Everyone sharing one 'geographic framework' avoids many problems.

In Chapter 10 we stressed the added value from data linkage; but success in this requires that the results are meaningful and hence the data have been 'fit for purpose'. Measuring the quality of many GI sets and hence of their 'fitness for purpose' is not easy. Moreover, we do not have a good idea of the likely consequences of using 'best available (but probably not good enough)' data for there have been relatively few legal tests as yet of liability involving GI in the courts.

Finally, the role of governments in GI and GIS is, and has been, important. In most parts of the world, the bulk of geographic information has until the last few years been created by governments and these are still major data custodians and providers. Many existing commercial products (e.g., road center-line files) also originated within government, typically with value added by the private sector. Governments also introduce market distortions – varying by country – due to subsidies and legal constraints. Most of this is somewhat different from other Knowledge Industry situations. For example, in the case of pharmaceuticals, government is not directly involved in production; multiple drugs produced by different multi-nationals sometimes compete and government's main role is to regulate safety and ensure that citizens are not being held to ransom by extortionate charging. Quite apart from their own involvement in data creation, GI enterprise is more difficult for governments to regulate than, say, electricity, water, gas, or even telecoms. The same GI may be manifested in many different forms by re-sampling or generalization. There is no single constant unit of GI quantity (c.f. water). Thus the capacity for gross profiteering, creation of monopolies, or even fraud is greater than in some other industries, especially with an immature market.

In summary, GI fits somewhat awkwardly within standard economic classifications. Its market (including

social) value – and hence use – are difficult to predict because of its source and other characteristics. This should be seen as an opportunity as much as a problem!

GI is a tradeable commodity and a strategic resource. Its supply and use is governed by public policy, economics, contractual, financial, and other considerations but regulating the market is difficult.

18.4.4 The Internet, World Wide Web, and GI

The Internet has been hailed as a distribution channel, a communications tool, a marketplace, and an information system – all in one. It is said to have altered almost everything managers do, from finding suppliers to co-ordinating projects to collecting and managing customer and client data. Its use is global and expanding rapidly (see Figure 18.8).

Some people argue that the change is more apparent than real, at least so far as commerce and the economics of information are concerned. They claim that there is no need for a new set of principles to guide business strategy and public policy in the Internet age. Have you read, they ask rhetorically, the standard literature on differential pricing, bundling, signaling, licensing, lock in, or network economics – have you studied the history of the telephone system or the battles between IBM and Microsoft with the US Justice Department? On this argument, technologies change but economic laws do not.

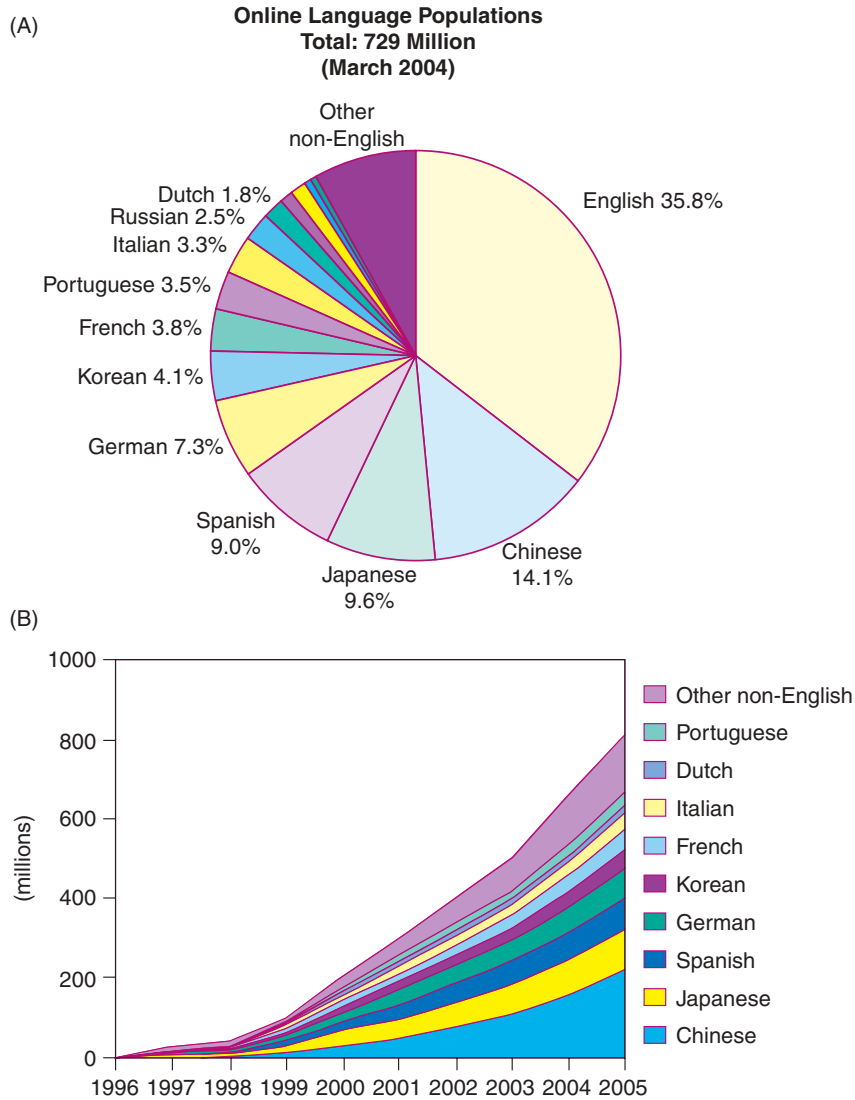


Figure 18.8 Use of the Internet, according to the GlobalReach consultancy in March 2004. (A) the use made by different language groups in 2004; (B) growth of use amongst non-English-speaking groups (*Source*: GlobalReach March 2004 **global-reach.biz/globstats/evol.html**) (Reproduced by permission of GlobalReach)

Applications Box **18.8**

Some Internet hype

Extravagant claims for the Internet have been made by MIT's Media Lab staff, e.g., Nicholas Negroponte said: '[thanks to the Internet, the children of the future] are not going to know what nationalism is'; his colleague Michael Dertouzos has claimed that digital communications will bring 'computer-aided peace. . .which may help to stave off future flare ups of ethnic hatred and national breakups'. All of this is reminiscent of earlier quotes: a telecoms enthusiast said: 'it is impossible that old prejudices and hostilities should longer exist, while such an instrument has been created for the exchange of thought between all the nations of the Earth'. He spoke in 1858 and referred to the advent of the first transatlantic telegraph cable.

Certainly there have been some extravagant claims made for the Internet (see Box 18.8). There remains a substantial 'digital divide' between those who have access to it and those who do not. Whilst there are circumstances in which sales and distribution via the Internet really have changed an industry, this particularly applies to goods in finite supply (e.g., airline seats) and those that are highly 'perishable' (airline seats are valueless after take-off). Much GI is not very perishable.

So far as GIS/GI users and managers are concerned, however, there are a number of very real or apparent advantages of the Internet. These are:

- The ease of setting up and using 'information location tools' (portals and clearinghouses or catalogs for digital libraries: see Chapters 11 and 20).
- The possibility to preview or 'taste' simplified versions of the information to check its suitability (essential in what is for many an 'experience good' – see Section 18.4.2).
- The capacity for mass customization, i.e., meeting the needs of the individual through tailored use of automated tools and standard data – hence segmenting the market to a huge degree and permitting differential pricing (see Figure 18.9).

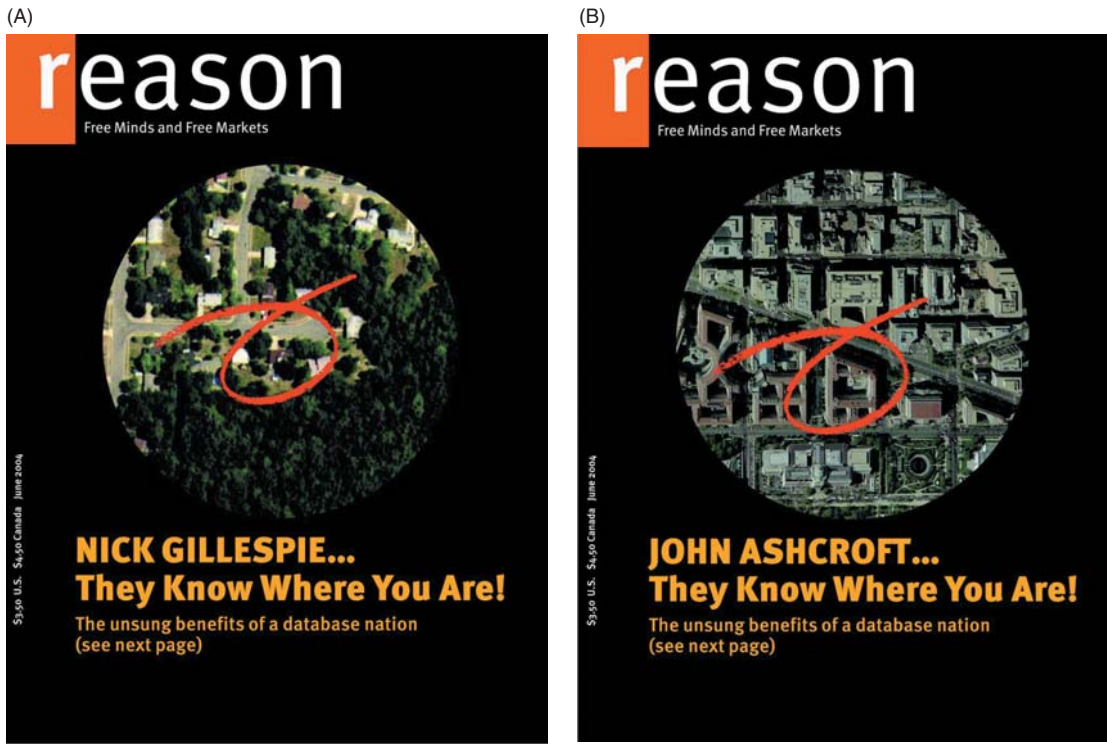


Figure 18.9 Customized covers of the June 2004 issue of *Reason* magazine, with an air photo of the home of each one of the 40 000 individual postal subscribers shown on the copy they received. This was achieved by merging their subscription address through a geo-coded address matching process with suitably geo-referenced aerial photography and generating customized, clipped digital images for the digital printing process. It heralds 'hyper-individualized' publications but also highlights concerns about privacy (see Section 19.1.4). *Source*: **reason.com/june-2004/samples.shtml**

- Its ability to transfer data (generally voluminous in the GI world) at very low cost.
- The ability to transfer costs to the users from the producers. Instead of paying for staff to deal with orders, package and dispatch material, etc., use of the Internet requires that users spend their own time on explicit selection from a menu of options and invest in whatever communication costs are involved.
- The reduction in search time for those who do not live near to good libraries and the immediacy of obtaining the results.
- The seamless way in which information generally may be obtained.
- If charging is implemented, it can be done through simple, low-cost operations like the use of credit cards.
- Initial supply of the information may be tracked and conditions of use enforced on the initial user through his or her inescapable agreement to terms and conditions of use before the information is down-loaded (as for software).
- The astonishing and rapidly growing familiarity with, and use by many people of, the Internet produces network externalities (see Chapter 11, Section 18.4.2, and Figure 18.8).

The Internet facilitates improved marketing, supply, and formatting of GI.

In the GI case, the greatest of all benefits come from customized mass production – where sub-sets of the data are selected to meet particular needs – and from the 'added value' potentially obtained by linking datasets together. Figure 18.9 shows where these two capabilities have been brought together. But the benefit is not because of the Internet *per se* but rather because of the digital form in which that information is held, together with its geographic identifiers. The Internet thus is a 'simplifier' of problems rather than a necessary condition for extracting added value. More importantly, in a world where fast answers are expected, the Internet and World Wide Web may make the delays so short that for the first time they are tolerable for managers and users.

18.5 GIS as a business and as a business stimulant

Thus far, we have treated GIS as a factor in achieving business success in commercial, government, not-for-profit, or academic organizations. But just how big is the GIS business itself – indeed, how can we measure it?

18.5.1 The supply side

First, let us consider the spending on the purchase of software, hardware needed to run it, data, and human resources (people). Truly comparable 'supply side' figures are hard to come by since published accounts of different firms often group activities in different ways reflecting their particular organizational structure. Not surprisingly, firms also portray their figures in the best possible light. The structure of the GIS industry mutates constantly as software vendors provide new services (such as data, publishing or education and training) to strengthen their hold on customers, often through partnership with other players (sometimes also as competitors – see Chapter 20) so a definition of GIS can never be definitive. However, Figure 18.10 shows how global GIS revenues are said to have increased, according to a consultancy specializing in measuring market performance. The figures should be treated with some caution but they show that software sales alone have more than doubled in five years.

An alternative approach is to consider one company and break down its revenues and other indicators of success such as licenses. ESRI – perhaps the largest GIS company worldwide – has no less than 125 000 organizations worldwide with signed license agreements; it has more than a million 'seats' for its different products. It is estimated that this translates to about half a million active users at any one time. ESRI gross incomes have risen between 10% and 20% annually over the last decade, with the financial year 2003 gross revenues being $500 million; if the revenues of the partly owned ESRI 'franchises' in other countries are added in, the total revenue comfortably exceeds $700 million. Another indicator of growth has been the numbers attending

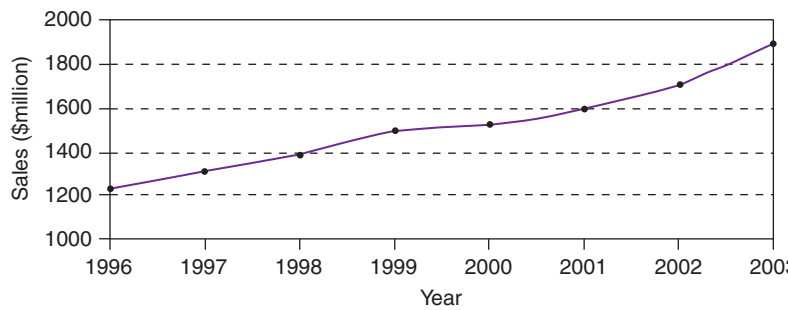


Figure 18.10 The growth in size of the global GIS market, according to Daratech Inc. (© Daratech)

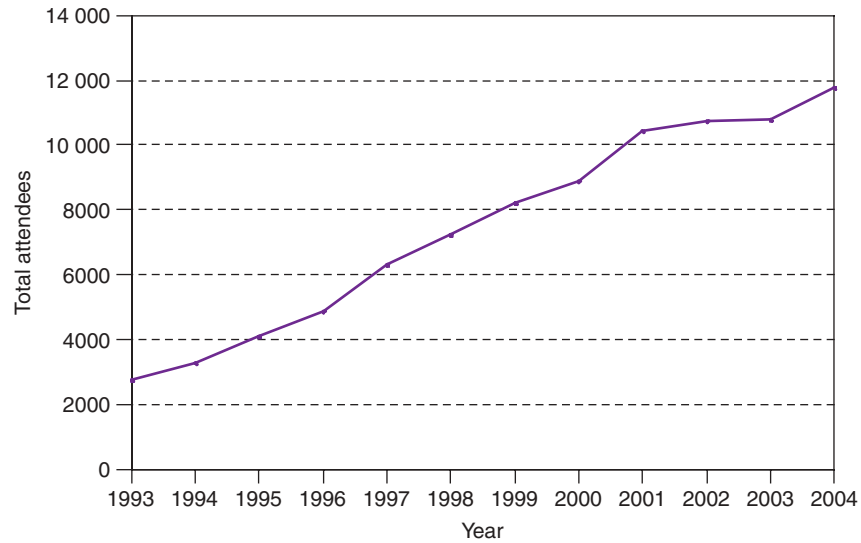


Figure 18.11 Growth in the numbers of attendees at the annual ESRI User Conference, which reflects growth in the GIS community worldwide

the annual ESRI conference: Figure 18.11 shows how this has grown from 23 attendees in 1981 to about 12 000 in 2004 – in many ways an allegory for GIS as a whole. Also the company estimates that their activities leverage at least 15 times the revenues they get in terms of staff, hardware, training, and other expenditures on the part of the users. All of this suggests that ESRI alone generates GI-related expenditures of over $7 billion annually.

If the world population had grown at the same rate as attendance at ESRI conferences since 1981, the global population would now be 2000 billion people, instead of 6.4 billion!

Figure 7.10 shows the proportions of the total global GIS software market achieved by various different players. If we assume each has the same leverage factor as ESRI, the total expenditure on GIS and related activities worldwide cannot be much less than $15 to 20 billion. Depending on the width of the definition of GIS, it could be much higher still (e.g., if the finances of both commercial and military satellites are included). Given the number of ESRI licenses, those of other commercial firms, use of 'free software' provided by governments and other parties plus the illegal, unlicensed copying and use of commercial software which still occurs, together with the contributions of the education sector and other factors, the likely number of active GIS professional users must be well over two million people worldwide. At least double that number of individuals will have had some direct experience of GIS and perhaps an order of magnitude more people (i.e., well over 10 million) will have heard about it (e.g., through such events as GIS Day – see Box 20.1). Yet more will, sometimes unknowingly, have used elementary GIS capabilities in passive Web services such as local mapping.

18.5.2 The demand side

Other ways of quantifying activities are to consider the expenditures of public bodies, all of which – other than security agencies – generally are reported. In the European Union, for example, the civilian national mapping agencies alone spend between $1.5 and $2 billion in total whilst land registration and other map and GI-using public sector bodies spend several times this amount. Worldwide, this figure can be multiplied by perhaps at least ten. The military expenditure in this area is very considerable: the US National Geospatial-Intelligence Agency (NGA) has a publicly stated budget of about $1 billion. How much of all this is on GIS and GI is partly a matter of definition. But, from all this 'user demand side' information, we can get some approximate confirmation of the software supply side estimates in the previous paragraphs. Cary and Associates stated that in 2003 geotechnology spending reported to the Federal Procurement Data System by US federal agencies exceeded $6 billion for the first time. This represents an increase of nearly eight percent over 2002.

The amount of other economic benefits underpinned by this GIS activity is not clear; few studies have been done and disentangling the interactions is difficult. One such study by consultant economists, however, has suggested that the $200 million annual expenditure on keeping up-to-date the national mapping of Britain supports and distributes products about one thousand times that sum in the Gross Value Added to the economy. But in no sense is the connection absolute: other sources of information – or guesswork – would be used if the national Ordnance Survey mapping were not available. Nevertheless, even if wrong (either way) by a factor of three this does demonstrate the scale of the GI/GIS business when considered on a wide definition.

18.6 Discussion

In this chapter, we have seen that GIS is now nothing less than an increasingly useful part of management's weaponry across almost all sectors of human activity. GIS is used in commerce to obtain a comparative advantage for one group over others; it is used as a means of evaluating policy and practice options in both the public and private sectors. Also, it is used as a highly effective information sharing device where that is deemed desirable.

To be useful to an individual manager, such generalizations need to be refined to fit the business environments faced by different organizations. Manifestly, these differ on many detailed counts: by virtue of the level of operations (e.g., international, national, or local); sector (private/public and government of different levels/not-for-profit' bodies); by the level of competition (formal and informal, as occurs in many governments); by business culture (e.g., willingness to collaborate with other bodies); and other factors. Some environments are much more conducive to success than others: for instance, the constraints of working with poorly skilled staff, on out-of-date equipment, and to impossible deadlines are non-trivial. As always, it is the interaction of the various elements of the business environment and the incentive structures and operating constraints that pose the greatest challenges for management in general and GIS in particular.

Questions for further study

1. Write down – without looking it up on your Intranet or in paper documentation – the five or six key strategic aims of your organization. Are the most important drivers in the organization in accord with these aims?
2. What are the special characteristics of geographic information as a commodity?
3. What potential benefits does GIS bring to businesses of different types (as defined in Section 18.1)?
4. Imagine that tomorrow you are to take over responsibility for five demoralized and under-funded staff in a GIS group in academia, business, or government. What would be your first actions? How do you decide whether someone is doing a good job and what do you do if they are not?

Further reading

Brown J.S. and Duguid P. 2000 *The Social Life of Information*. Cambridge, MA: Harvard Business School Press.

Shapiro C. and Varian H.R. 1999 *Information Rules: A Strategic Guide to the Network Economy*. Cambridge, MA: Harvard Business School Press.

Thomas C. and Ospina M. 2004 *Measuring Up: The Business Case for GIS*. Redlands, CA: ESRI Press.

19 *Exploiting GIS assets and navigating constraints*

Recent years have seen a big increase in the value of 'g-business' in commerce, government, not-for-profit organizations and academia. This has been built on use of both tangible and intangible business assets such as legal ownership or rights to physical and intellectual property, the caliber and skills of staff, the customer or client base, brand image and organizational reputation, and organizational compliance with accepted standards. Sometimes however the same factors can act as constraints. This chapter considers the importance of three major factors both as assets and constraints – the law relating to GIS, staff competences and the availability and accessibility of 'core' datasets. GIS is inevitably disruptive to most organizations in the early stages of its adoption even if it is crucial to their long term success. And events (such as 9/11) or new technologies (such as the Web) can reshape our worlds remarkably rapidly. From all this, we demonstrate the need for an organization-wide GIS strategy and implementation plan and the need to identify and reduce risk to acceptable levels.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter, you will understand:

- How the law impacts on GIS assets, plans, and operations;
- The need for particular staff skills and attitudes plus the reasons why professional certification is advocated by some experts;
- How availability, accessibility, and quality of core GI act as business constraints;
- The issues about making public sector GI freely available to all users;
- How a GIS strategy helps us to exploit assets, cope with constraints, and reduce risk.

Figure 19.1 The scales of justice: the essence of the law. Actions of individuals or organizations are assessed against the laws of the land and judgments handed down (Reproduced by permission of Richard Bailey)

19.1 GIS and the law

The law touches everything (Figure 19.1). During a career in GIS, we may have to deal with several manifestations of the law. These could include copyright and other intellectual property rights (IPR), data protection laws, public access issues enabled, for example, through Freedom of Information Acts (FOIA), and legal liability issues. But since laws of various sorts have several roles – to regulate and incentivize the behavior of citizens and to help resolve disputes and protect the individual citizen (Figure 19.2) – almost all aspects of the operations of organizations and individuals are steered or constrained by them. One complication in areas such as GIS is that the law is always doomed to trail behind the development of new technology; laws only get enacted after (sometimes long after) a technology appears.

There is also a geography of the law. The legal framework varies from country to country. For example, the Swedish tradition of open records on land ownership and many personal records dates back to the 17th century, but much less open frameworks exist even in other countries of Europe. The creation, maintenance, and dissemination of 'official' (government-produced) geographic information (see Section 19.3) is strongly influenced by national laws and practice. Box 19.1

Figure 19.2 Corporations – or governments – can often act in ways which infringe the rights of individuals, such as building highly intrusive structures adjacent to homes. The law can give protection against certain acts of this kind (Courtesy: Landmark Information Group)

summarizes the situation and the way in which commerce has to adjust to the legal framework of each country in

which it operates – though supra-national bodies like the World Trade Organization and the European Commission increasingly force greater homogeneity in certain laws and regulations.

Given this heterogeneity and on-going change, none of what follows should be taken as other than a general discussion of the issues. It is vital that everyone with management responsibilities working in GIS reads more detailed treatments of GIS and the law (see the guide to further reading at the end of this chapter) *and* seeks local, professional legal advice at an early stage wherever the law might be involved in a problem.

The law touches everything. Be prepared.

19.1.1 Protecting information assets

Innovation is key to GIS success (Section 18.3.1) Through it, individuals or organizations may create something better than has been done before. It is a normal reaction to consider how to achieve the greatest benefit from that investment of time and other resources. One response is to make it generally and freely available – either because that is 'good science', because wider availability will benefit the owners more, or because laws (e.g., those under which the US Federal Government operates) dictate that should happen. In many commercial or artistic operations, however, protection of one's own creative works (e.g., through licensing of the intellectual property rights) is the very basis of their operations. Governments have long recognized the need for such protection of investment – witness the Berne Convention of 1886, subsequently modified on several occasions. This treaty has been signed by most countries.

At the other end of the spectrum, there is clearly a societal interest in ensuring that innovations diffuse widely, rather than society being 'held to ransom' indefinitely by an unscrupulous monopolist. The law attempts to balance this right to exploit results of creativity and of investment against the need to foster society-wide gains. Most typically this is done by limiting the time of protection enjoyed by an innovation, such as for patents for new drugs or copyright on published material. A variety of means of legal protection exist – through legislation on copyright, designs, patents, trademarks, and trade secrets. Because of its nature, most of the protected material in GIS has been safeguarded by copyright legislation. The GIS world is becoming increasingly litigious: lawyers will continue to benefit.

Copyright is a universally agreed form of protecting investment – but copyright owners can opt not to enforce it.

Patenting is also a process impinging increasingly on the IT industry generally and GI in particular. Traditionally given for a new manufactured good, some patents have been awarded for software but a recent development has been their award for processes. There is a geography to this – what can be patented in the USA sometimes cannot be so protected in Britain or Australia. The reality, however, is that patenting is much more widely applicable than is commonly supposed and it is possible for anything which is useful, novel, and non-obvious – including business methods. In mid-2000, Etak – a pioneer in car guidance systems – was awarded four patents relevant to GIS. These included a method for storing map data in a database using space-filling curves and a system for path-finding computation. An example of a GIS-related patent is shown in Box 19.2. Firms owning patented processes commonly seek royalties from others who use arguably identical approaches. In such cases, the choices are either to give in – avoiding legal costs but paying for something which may have been developed independently – or fight the case in court arguing independent development and prior use. The latter course ties up senior management, incurs legal costs, and may end in failure and awards of punitive damages. A number of enterprises have judged it better to license rather than challenge claimed infringement of patents.

Applications Box **19.1**

Globalization of commerce and the law

Commerce is global. Law, for the most part, is not. . .never has that truth caused greater uncertainty than now, as global business clashes with local and national law in the borderless new world of electronic commerce. . . all the uncertainty is good for lawyers but not for the growth of e-commerce. . .The threshold question is: whose laws apply? . . . the question of jurisdiction remains unresolved, both internationally and within the US: what consumer laws, contract law, privacy laws and other laws apply to e-commerce applications? Where does a transaction take place? How will conflicts in law be determined?

Financial Times, 23 December 1999

Applications Box **19.2**

Interactive automated mapping patent

Patent Number: 5,214,757

Name: Interactive Automated Mapping

Issuing office: United States Patent Office

Date of Patent: May 25,1993

Inventors: Thad Mauney, Aglaia C. F. Kong, Douglas B. Richardson, all of Yellowstone County, Montana

Abstract: An automated, fully transportable mapping system utilizes position information gathered from a Global Positioning Satellite (GPS) capture program to create new maps or annotate existing maps contained in a Geographic Information System (GIS) database in real time. In addition, the present invention displays position information captured by GPS in real time, enabling users to track the path on which they are traveling. Attributes related to the position information may also be entered in real time, and are stored in a file for subsequent inclusion in a GIS database.

The rest of the patent document describes the specific inventions claimed, the technical field, background art (context), and detailed description of the preferred embodiment.

Society has ordained that there must be a balance between the right to benefit from one's innovations and protection against long-term monopolies.

19.1.2 Some big information ownership questions

In this section, we show how some frequently asked questions about GIS assets and the law can be answered.

Can geographic data, information, evidence, and knowledge be regarded as property?

The answer to this is normally yes, at least under certain legal systems. The situation is most obviously true in regard to data or information – under either copyright or database protection (see below). Who owns the data is sometimes difficult to define unequivocally, notably in aggregated personal data. An example is bills at a supermarket which have been paid by credit card and subsequently geocoded to generate aggregate purchasing characteristics for people in a zipcode (or postcode) area.

Can geographic information always be legally protected?

There is little argument about copyright where it is manifestly based on great originality and creativity, e.g., in relation to a painting. Rarely, however, is the situation as clear-cut, especially in the GIS world where some information is widely held to be in the public domain, some is regarded as 'facts' (see Section 18.4.1), and some underpins either the commercial viability or the public rationale of government organizations.

In the USA, the Supreme Court's ruling in the famous *Feist* case of 1991 has been widely taken to mean that factual information gathered by 'the sweat of the brow' – as opposed to original, creative activities – is not protectable by copyright law. Names and addresses are regarded widely as 'geographic facts'. Despite this, several jurisdictions in the USA and elsewhere have found ways to protect compilations of facts, provided that these demonstrate creativity and originality. The real argument is about which compilations are sufficiently original to merit copyright protection. A number of post-*Feist* instances have occurred where maps have been recognized by courts not as facts but as creative works, involving originality in selection and arrangement of information and the reconciliation of conflicting alternatives. Thus uncertainty still exists in US law about what GI can be protected.

In Europe, different arrangements pertain – both copyright and database protection exists. For copyright protection to apply to a database, it must have originality in the selection or arrangement of the contents and for a database right to apply, the database must be the result of substantial investment. It is, of course, entirely possible that a database will satisfy both these requirements so that both copyright and database rights may apply.

A database right is in many ways very similar to copyright so that, for example, there is no need for registration – it is an automatic right like copyright and commences as soon as the material that can be protected exists in a recorded form. Such a right can apply to both paper and electronic databases. It lasts for 15 years from making but, if published during this time, then the term is 15 years from publication. Protection may be extended if the database is refreshed substantially within the period. As stressed earlier, geographic differences in the law have potential geographic consequences for commercial GI activity. Protection equivalent to the database right will not necessarily exist in the rest of the world, although all members of the World Trade Organization (WTO) do

have an obligation to provide copyright protection for some databases.

Can information collected directly by machine, such as a satellite sensor, be legally protected?

Given the need for originality, it might be assumed that automated sensing should not be protectable by copyright because it contains no originality or creativity. Clearly this interpretation of copyright law is not the view taken by the major players, such as Space Imaging Inc., OrbImage, and DigitalGlobe (see Section 20.5.1.2 and Table 1.4) since they include copyright claims for their material and demand contractual acceptance of this before selling products to firms or individuals. It helps that there are at present only a few suppliers of satellite imagery. The cost of getting into the business is high; thus any re-publication or re-sale of an image could be tracked back to source relatively easily. In addition, there is some evidence that some of the key customers (notably the military) are interested primarily in near-real-time results. In any event, imagery up to 15 years old would be protected under the EC's Copyright Directive, at least within the European Union area itself.

How can tacit geographic and process knowledge – such as that held in the heads of employees and gained by experience – be legally protected?

'Know how' is a foot-loose element. Suppose that the lead designer of a new software system is seduced away by a large software vendor. Though no code is actually transported, the essence of the software and the lessons learned in creating the first version would facilitate greatly the creation of a competitor. This situation has happened in GIS/GI organizations from the earliest times. The only formal way to protect tacit knowledge is to write some appropriate obligations into the contract of all members of staff – and be prepared to sue the individual if he or she abrogates that agreement. Human rights legislation may complicate winning such action. Keeping your staff busy and happy is a better solution!

You can rely upon someone copying your work. If you don't want that to happen without your permission, fingerprint your data and be ready to sue.

How can you prove theft of your data or information?

It is now quite common to take other parties to court for alleged theft of your information. One of the authors previously led an organization which successfully litigated against the Automobile Association in the UK for illegal copying, amending, and reproducing (in millions of copies) portions of Ordnance Survey (OS) maps. This was despite the terms, conditions of use, and charges for such mapping being published and well-known: the OS received over $30 million in back payment.

The legal process in such circumstances is an adversarial one. Therefore, it is crucial to have good evidence to substantiate one's case and to have good advice on the laws as they exist and as they have been previously interpreted by courts. How then do you prove that some GI is really yours and that any differences which actually exist do not simply demonstrate different provenances – especially since digital technology makes it easy to disguise the 'look and feel' of your GI, and reproduce products to very different specifications, perhaps generalized?

The solution is both proactive and reactive. In the first instance, the data may be watermarked in both obvious and non-visible ways. Though watermarks can be removed by duplicitous individuals, they serve as a warning. A series of groups of small numbers of colored pixels scattered apparently randomly throughout a raster dataset ('salting') can be effective. Proof acceptable to courts requires a good audit trail within one's own organization to demonstrate that these particular pixels were established by management action for that purpose. In addition to watermarking, finger-printing can also be used: at least one major US commercial mapping organization is said to add occasional fictitious roads to its road maps for this purpose. In areas subject to temporal change (like tides or vegetation change), it is very unlikely that any two surveys or uses of different aerial photographs could ever have produced the same result in detail. Such evidence should, in the right hands, be enough to demonstrate a good case though rather different techniques may be called for in regard to imagery-based datasets of slow-changing areas.

Who owns information derived by adding new material to source information produced by another party?

Suppose you find that you have obtained a dataset that is useful but which, by the addition of an extra element, becomes immensely more valuable or useful. This is common: the 'First Law of GIS Management' says that you get something for nothing by adding geographic information together. The number of data combinations and potential uses rise very rapidly as you add a few more 'data layers' – some of which may have been derived from pre-existing maps (Section 18.4.1 and Figure 19.3).

The 'First Law of GIS Management' says you get something for nothing by bringing together GI from different sources and using it in combination.

So who owns and can exploit the results in derivative works? The answer is both you and the originator of the first dataset. Without his or her input, yours would not have been possible. This is true even if the original data are only implicit – if, for instance, you trace off only a few key features and dump all the rest of the original. Moreover, you may be in deep trouble if you combine your data with the original data and omit to ask the originator for permission to do so. Where 'moral rights'

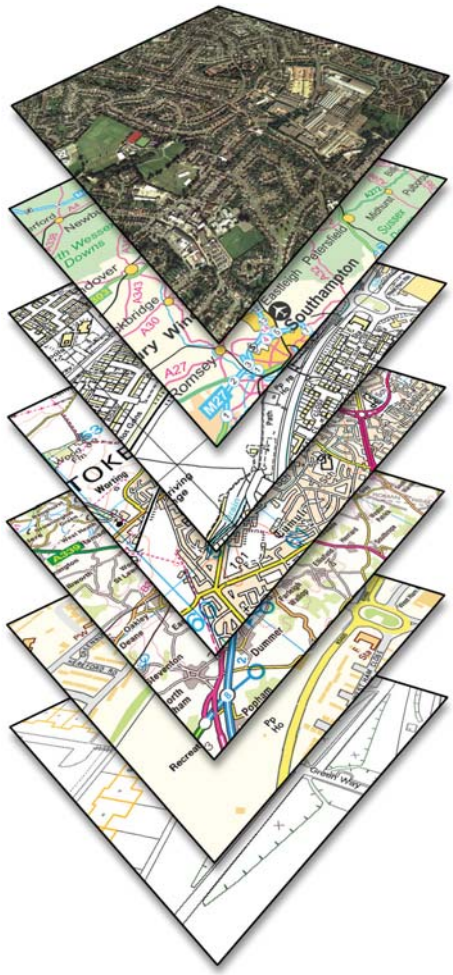


Figure 19.3 The concept of adding layers of geographic information (e.g., that stored in different maps) to get added value (Courtesy: Landmark Information Group)

exist, e.g., in European countries, the author has the right to integrity (i.e., to object to any distortion or modification of his or her work which might be prejudicial to his or her honor). He or she also has the right to attribution as the author, to decide when a work will be published, and to withdraw a work from publication under defined circumstances.

19.1.3 Legal liability and GIS

Liability is a creation of the law to support a range of important social goals, such as avoidance of injurious behavior, encouraging the fulfillment of obligations established by contracts, and the distribution of losses to those responsible for them. It is a huge and complicated issue. Cho identifies Alaska as the only US state that has specific provisions for limiting liability for GIS. In many cases liability in data, products, and services related to GIS will be determined by resort to contract law and warranty issues. This assumes that gross negligence has not occurred, when the situation can be far worse and individuals as well as corporations can be liable for damages. Other liability burdens may also arise under legislation relating to specific substantive topics such as intellectual property rights, privacy rights, anti-trust laws (or non-competition principles in a European context), and open records laws.

Clearly all this is a serious issue for all GIS practitioners. Consider, for instance, the following realistic scenario. You create some software and assemble some data from numerous sources in your spare time and for your own interest. A friend admires the result and asks for a copy: you are glad to oblige, knowing he is doing some important work on behalf of the local community. Decisions are made using these tools which subsequently are shown to be founded on results that are in error because of poor programming and inadequate, error-prone data. You get ostracized and sued by the community, by your friend – and by data suppliers who are annoyed that you have ripped off their data and, by implication, also brought them into disrepute.

Minimizing losses for users of geographic software and data products and reducing liability exposure for creators and distributors of such products is achieved primarily through performing competent work and keeping all parties informed of their obligations. If you have a contract to provide software, data, services, or consultancy to a client, that contract should make the limits of your responsibility clear. Disclaimers rarely count for much.

19.1.4 GIS and privacy

Societies capable of constant and pervasive surveillance are being built rapidly around us, sometimes with our co-operation, more often without our knowledge. GIS is part of all this activity. For example, the use of postal addresses of individuals provides some form of geographic identification for many purposes. Adding value (Section 18.4.1) via the bringing together of geographic information and personal information can lead to unverifiable conjectures about the characteristics of human individuals or households which are then subsequently used in marketing and elsewhere. In some eyes, this is unacceptable and immoral. Typically, civil rights campaigners favor an 'opt in' approach. In general terms, industry favors 'opt out' provisions – the data subject must say he or she wants to be excluded from having their records used by firms.

Some of the reasons for this move are that technically it is now possible; but others arise from security considerations. *The Economist* of 23 January 2003 summarized the situation as follows:

> **. . .since the terrorist attacks of September 11th 2001. . .governments everywhere, not just in the United States, have rushed to expand their powers of surveillance (see Box 20.6). Even the European Union, which has long been at loggerheads with America about data protection, performed a sharp U-turn, relaxing legal restraints on the collection**

and use of information by governments. In June 2002 the EU adopted a new directive to allow member states to require firms to retain data on everyone using mobile phones, landlines, e-mails, chatrooms, the Internet or any other electronic device. In America, the USA Patriot Act, passed in the wake of September 11th, made it much easier for the authorities to obtain court permission to monitor Internet traffic and obtain wiretaps. The technology to do this is being built into the Internet's infrastructure.

There is no escape from the uncomfortable reality: GI and GIS are in the front line so far as both security and sales and marketing are concerned and their effective use can erode privacy. The only practical (though modest) amelioration is through the use of agreed codes of professional conduct and, where appropriate, laws (where the situation continues to mutate).

19.2 GIS people and their skills

In many businesses, especially those forming part of the Knowledge Economy (Section 18.3), staff are usually seen as the prime business asset and the factor differentiating one organization from another. The supply of GIS experts and GIS-literate people is also a crucial factor in determining the rapidity of up-take of the existing facilities and their successful use.

Actually, we know very little about the skills and other competences and characteristics of the millions of people working in the GIS industry worldwide at present. But anecdotal evidence suggests that most see themselves and their careers as being technology-focused. Of the 134 jobs advertised on the GIS Jobs Clearinghouse in a 2 month period in summer/fall 2004, some 40% were for GIS analysts, 25% were for technicians, 22% for managers, and 10% for programmers. It seems that the historical origins of GIS have led to many GIS-trained employees being highly skilled technically but being somewhat ill at ease and unfamiliar with business issues – or even with the user community outside of that one in which they were first employed. This situation calls for improvement; appropriate education, training, and personal development is essential.

Most GIS industry people think of themselves as technical experts. An over-strong focus in this area is not in the long-term interests of GIS as a whole.

19.2.1 Existing GIS education and training

Despite very different cultures and ways of operating worldwide, there are at least three reasons why we would expect global similarities in GIS curricula and educational approach. The first is the pervasive influence of the software vendors whose systems have harmonized some GIS concepts and terminology. The second is the effect of the GIS Core Curriculum created by a team led from the University of California Santa Barbara in 1990. This curriculum was made available, cheaply, worldwide and has been used in over 1000 universities. Finally, there have been a number of subsequent attempts to create international consortia delivering the GIS education material in different countries, most notably the UNIGIS consortium, with nodes in 14 countries (see Box 19.4). Web-based material is enabling many universities to deliver courses to individuals irrespective of where they are physically located. The different types of courses operating are shown in Table 19.1.

Typically, however, much learning in GIS is achieved in substantial measure by 'doing', rather than by copying and reciting 'facts' given in lectures. That 'doing' is best guided or supported by a tutor or mentor. Such tutoring can be provided in a rote-learning situation by a software wizard. More complex educational issues, where judgment is concerned, require a human being to act in this capacity.

19.2.2 Professional accreditation in GIS

From the public's point of view, there is an attraction in having properly certified or accredited learning or training – especially where safety-critical activities are involved, for example, in engineering. Globalization is leading to globally acceptable certification in some fields. At present – and despite early moves (often by land surveyors) – there is no known professional certification of GIS/GI courses analogous to those in other professions. But there are areas of GIS where professional skills are being exercised, where its exploitation is not simply a matter of pressing buttons, and where the results may impact strongly on people's lives. In such circumstances, the whole question of judgment, quality of work, and liability (see Section 19.1.3) comes into play. Two approaches exist – to do it through existing professional or statutory bodies (as for accountants, surveyors, or those in the insurance community) or to set up a scheme administered centrally by individuals or bodies experienced in GIS and GI.

Within the USA, the Urban and Regional Information Systems Association (URISA) launched a certification program in 2004 whose purpose is to provide those professionals who work in the field of geographic information systems with a formal process that will:

- allow them to be recognized by their colleagues and peers for having demonstrated exemplary professional practice and integrity in the field;
- establish and maintain high standards of both professional practice and ethical conduct;
- encourage aspiring GIS professionals to work towards certification for the purpose of professional development and advancement;

Table 19.1 A simple summary of GIS/GI-related training and education courses

	Main delivery mode	
	Face-to-face courses	'Distance learning'
Software training courses	Commonplace – all vendors provide this themselves or via licensed partners. Can be very remunerative.	Becoming ever more important, e.g., ESRI Virtual Campus
Software development and customization	Varies greatly. Fundamental training is provided by university computer science courses. Vendors may provide specific courses, e.g., on their macro language.	As above
School-level education	GIS was once part of the UK government's Core Curriculum for schools. Elsewhere the school curriculum is typically defined locally, e.g., see Box 19.3.	
Undergraduate-level education	GIS is taught in many hundreds of undergraduate programs around the world but in many instances as modules in degree courses in geography and non-geography courses (e.g., forestry or environmental sciences). Only a few universities run undergraduate courses entirely in GIS.	The UNIGIS consortium (see Box 19.4) has set up a distance-learning course with a common 'pick and mix' curriculum now used by universities around the world. This benefits from some face-to-face teaching.
Postgraduate-level university education	Numerous Masters programs exist	Some of these programs can be taken in a distance mode but relatively few have been designed from the outset for that.
Short courses for professionals (i.e., continuing professional development or CPD) – the systematic maintenance, improvement, and broadening of knowledge and skill, and the development of personal qualities necessary for the execution of professional and technical duties throughout one's working life.	This is a matter of rapidly growing importance. Taking CPD is often mandatory to remain a member of a professional body.	An increasing number of these courses are given by distance-based means.

Applications Box **19.3**

GIS in schools: a Texas success story

From being a failing school a few years ago to winning one of the top 20 Blue Ribbon schools awards in summer 2004 – against entries from 1200 schools across the USA – is a stunning achievement. This demonstrates leadership of a high order. But GIS helped as well. The Bishop Dunne Catholic K–12 School, Dallas, won an INTEL Scholastic twenty-first century schools of distinction award for their innovative use of GIS technology (**www.bdhs.org/gis/**) – along with a slew of other awards (Figure 19.4A).

The GeoTech Research Lab at Bishop Dunne Catholic School is a unique place where students become the researchers, active community stakeholders, and the creators of geographic and environmental solutions. For example, the Bishop Dunne students partnered with the Dallas Police Department for several projects. In the early stages of the partnership, students produced maps for neighborhood crime watch organizations. The maps illustrated where crimes were and what types of crimes were occurring on a monthly basis. Later projects

included producing maps to help determine where the Dallas Police Department should deploy its Robbery Task Force.

Fire Ants are a growing environmental problem in the North Texas area. Students of Bishop Dunne Catholic School's GIS program studied fire ant colonization in the areas surrounding the school, beginning with field investigations on campus, and expanding later to nearby Kiest Park. Figure 19.4B shows some of the results. In this case, the students' community partners were City of Dallas Parks & Recreation, the Department of Agriculture – at Texas A&M University, and the USDA–Dallas County Extension Agents.

The use of GIS to achieve results of real value for the community, to enhance student skills, and as a means of inter-working with various organizations is an example that with value could be cloned worldwide.

(A) (B)

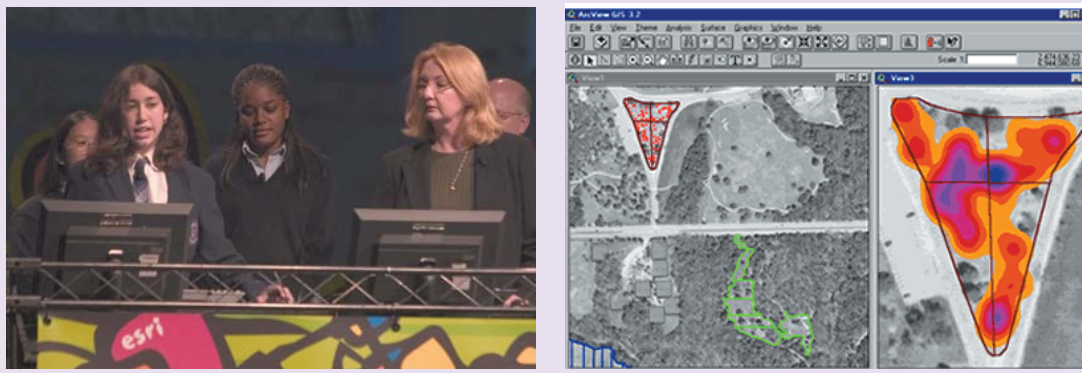


Figure 19.4 (A) Some of the students (and teachers) at Bishop Dunne Catholic School, Texas, who have won a host of national awards for their GIS work with local communities. They are seen here receiving a major award at the 2004 ESRI User Conference. (B) The clustering of fire ants around the school, as based upon student surveys and aggregated using a GIS

- encourage established GIS professionals to continue to hone their professional skills and ethical performance even as GIS technology changes.

The URISA aim is to provide a formal system to evaluate the competency of GIS professionals. This is a noble aspiration. In practice, it was felt essential to: be voluntary and open to all; be flexible; use existing GIS educational bodies; be collaborative; and include a code of ethics. As implemented through the GIS Certification Institute (**www.gisci.org/**), it is not achieved through any test but by self-certification based on points calculated from educational attainment, professional experience, and contributions to the profession. Benchmarks in each of these three areas are, respectively, a Bachelor's degree with some GIS courses (or equivalent), four years in GIS application or data development (or equivalent), and annual membership and modest participation in a GIS professional association (or equivalent). A committee adjudicates upon each application for certification based on the fit of the application to the benchmarks and each successful applicant must sign a code of ethics.

Such a process can be scorned easily: it is based largely on self-certification, measures inputs (e.g., hours in classes) not outcomes, and the process for coping with unprofessional behavior is not yet clear. Moreover, it implies the acceptability of some 'authorized' course contents rather than other material; this has led in other professional certification programs to ossification of what is acceptable and much-reduced innovation. All that said, this is the first major, national attempt to put GIS certification in place.

Certification of GIS practitioners could have the benefit that it may enhance professional standards – but it may also stunt innovation.

19.2.3 Fixing the shortcomings of current GIS education and learning

The 'survey' of GIS staff characteristics quoted earlier plus other evidence suggests that a technical fixation remains current in GIS. This creates a self-perpetuating role for GIS as a pursuit for those 'with clever fingers'. It is no surprise, therefore, that relatively few GIS people have yet made it to the highest levels in major organizations. If we wish to have GIS skills and approaches more deeply involved in high-level decision making, we need to modify the nature of our education and learning.

The reality – and one both espoused and noted throughout this book – is that GIS is now a much 'broader church' than simply a collection of technical expertise. As a result, it seems unwise to have a substantial GIS/GI

Biographical Box **19.4**

Josef Strobl and UNIGIS

Figure 19.5 Josef Strobl, GIS educator

Josef Strobl (Figure 19.5) is an associate professor of Geography at the University of Salzburg in Austria and Director of the Centre for GeoInformatics (**www.zgis.at**). His interests are focused very much internationally, working with colleagues in many other countries.

But our particular interest here is in his role as co-founder, and his leadership of, the UNIGIS Consortium (**www.unigis.net**). UNIGIS is a worldwide network, now of 15 universities, co-operating since 1990 in the design and delivery of distance learning in GIS and GIScience. UNIGIS host universities currently operate in Austria, Canada, the Czech Republic, Ecuador, Hungary, India, the Netherlands, Poland, Portugal, Russia, South Africa, Spain, the UK, and the USA. The network offers Postgraduate Certificate, Diploma, and Masters courses. The courses are supported by an international network of qualified academics and other professionals and use industry-standard software to give appropriate practical experience. UNIGIS courses are modular and flexible study programs. The content varies to meet the needs of local students. Optional courses provide the opportunity to tailor the program to meet individual student's needs, and residential workshops are offered to support communication and networking. Local centers provide administrative support and interaction with experienced tutors. Tutoring and assessment is by electronic communication. Dedicated websites provide virtual e–Learning environments and support at all stages of the learning process.

UNIGIS is a unique entity in GIS and has enrolled about 6200 students since it began. Josef and his colleagues have shared generously their experience in running the network, including the development of online teaching material for use in diverse educational and national cultures.

course which does not include at least some treatment of the following:

- entrepreneurial skill development and leadership;
- the principles of geographic science;
- understanding of and familiarity with GIS technologies;
- understanding organizations;
- finance, investment criteria, and risk management;
- human resources policies and practice;
- legal constraints to local operations;
- cultural differences between disciplines;
- awareness of international differences in culture, legal practice, and policy priorities;
- formal management training, including staff development and presentational and analytic skills.

Clearly not all courses and learning needs to include all of this material. Some will be particularly relevant to those engaged (as all professionals should be) in continuing professional development short courses. The introduction of locally related legal, cultural, and application-related elements – as well as buttressing the global technical, business, and management issues – will be to the benefit of GIS practitioners, the discipline, and business (used in our wide sense) alike.

19.3 Availability of 'core' geographic information

If GI is needed but not available, someone will be faced with a substantial cost in creating or buying it. Without such 'rocket fuel', vendors have much more difficulty in selling software. And politicians who require speedy guidance on what to do after natural disasters or quick answers in regard to a policy issue simply will not be able to use GIS. It follows, therefore, that data existence, availability, and ease and cost of access have major impacts on the entire GIS enterprise and on its future potential. Central to all this is the availability of the 'core GI' – the geographic framework on which all other information is assembled (see Boxes 19.6 and 19.7). Typically, geographic information is collected by many different agencies, usually within the boundaries of national territories; and, historically, most of these agencies have collected data primarily to suit their own local purposes. Thus, if the benefits of added value are to be obtained from data integration within a GIS (see Section 18.4.1 and Box 19.5), there must be a mechanism (or spatial 'key') to register the different sets of information together in their correct

Applications Box **19.5**

A value-adding GI organization

Landmark Information Group is a leading supplier of digital mapping, property, and environmental risk information. The organization has forged data supply agreements and partnerships with a number of government bodies. Its business is founded upon two principles. The first is that all property-related investment decisions are affected by potential environmental risks and liabilities. The second is that combining and consolidating environmental data from a wide variety of sources not only improves the availability of data and makes them more convenient to access, it also creates new types of information services which can provide new solutions to old problems. In other words, it creates added

(A)

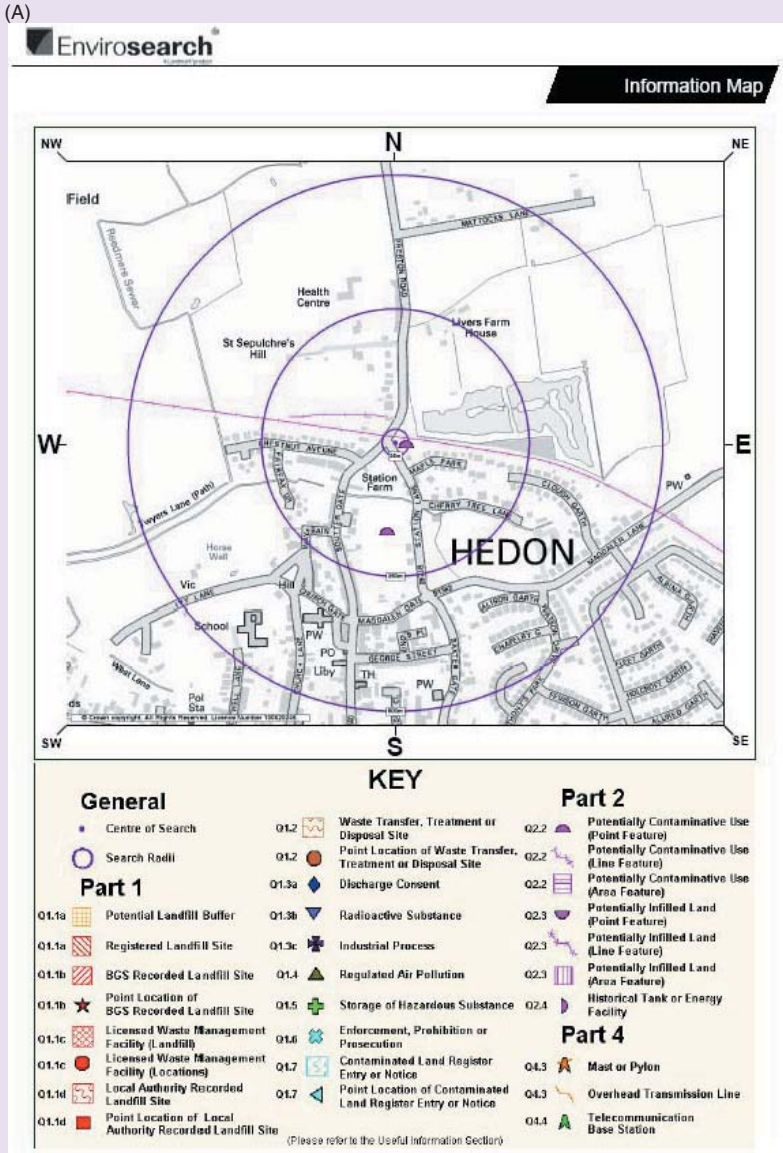


Figure 19.6 (A) Hazards in an area within specified distances of a property which is being purchased. The results are based on bringing together environmental data from many sources to give added value. (B) Flood hazards in the same area as shown in (A). The results are again derived by bringing together environmental data from many sources (Courtesy: Landmark Information Group – see **www.landmarkinfo.co.uk**)

value. On this basis, a wide range of nationally complete digital GI has been assembled from different sources and geo-referenced so it fits together with the Ordnance Survey framework data (see Box 19.6). These include:

- One million historical Ordnance Survey maps dating back to the 1850s, together with a record including address and coordinates for every address in the country.
- A century-long set of Trade Directory Entries and Business Census data indicating sites with potentially contaminative uses, plus other such sites identified by analysis of the OS maps.
- Areas at risk of coastal or riverine flooding.
- Natural hazard areas (e.g., those liable to coal mining subsidence, radon-affected areas, and groundwater vulnerability areas).
- Planning consents and site information (e.g., discharge consents into streams and rivers, hazardous substance consents, registered landfill sites).
- Air Pollution Control records, i.e., authorizations.

From all this and other information integrated within their GIS, Landmark offers a variety of services tailored to different markets – to environmental professionals, real estate professionals, property lawyers, local governments, utility companies, and individual house owners. The business has grown from its origins in 1995 to a $80 m turnover business in 2004. A sample of two parts of the standard report for house owners on the potential environmental hazards surrounding their home is shown in Figure 19.6A and B; the full report is produced for $50.

(B)

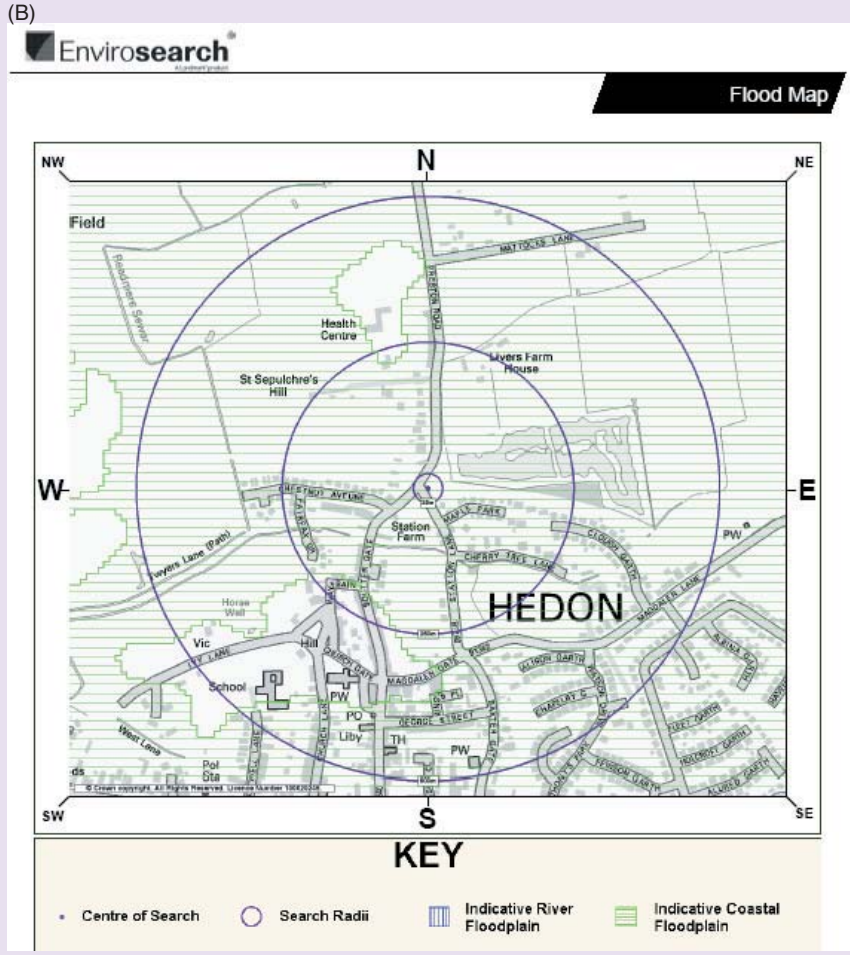


Figure 19.6 (*continued*)

relative positions. The geographic framework provides this ubiquitous data linkage key through the use of co-ordinates or (sometimes) via implicit geographic coding (e.g., zip or postcodes).

Geographic framework information underpins all use of GIS.

19.3.1 The nature of the framework

What constitutes the framework varies between different countries. In the USA, the Federal Geographic Data Committee has defined the information content of the US framework as consisting of:

- geodetic control (control points, datums, etc.);
- elevation data;
- hydrography;
- public land cadastral information;
- digital orthoimagery (see Section 9.3.2.4);
- transportation;
- the geography of governmental/administrative units.

Similar definitions – including placenames – would in practice be found in many countries. Most typically, the framework is manifested in user terms through topographic mapping – the 'topographic template'. Normally, such mapping will contain all this information but partly in implicit form (e.g., geodetic control). It has long been produced by national mapping agencies; as arms of government, these have provided continuity of provision over decades or even centuries in many countries.

Generally, the framework's quality is publicly defined and well known and its governmental source is likely to stay in operation to update and guarantee it.

Important differences do exist between national frameworks. These are tailored to different major types of application, were created in response to different legal and governmental environments, and have persisted in a historic form for different periods. Typically, only one or two public-sector organizations are responsible for providing the overall national framework (as in Britain, described in Box 19.6; and also France, the Nordic countries, India, and Japan). But, in countries with federal governments, the normal situation is that the national government produces relatively low-resolution materials (see Box 19.7) and local and more detailed versions of the framework are produced by other (e.g., state or local) governments (as in Australia, Germany, and the USA). Detailed parts of the framework may also be unavailable in some countries for reasons of state security or lack of investment. Moreover, at present, much framework information is still only available in paper map or coordinate list form in some countries.

Primarily, to date, the private sector has been involved in marketing some of this framework data or as contractors in its collection. However, as the availability and use of digital imagery from commercial satellites becomes more common, the situation is changing quite rapidly. The world of national mapping agencies is also changing rapidly: the new US approach to creating The National Map is extraordinarily different to the traditional one, involving sewing together material from many different sources and the use of volunteers to update it using GPS (see Box 19.7).

19.3.2 Making the framework available

The quality, currency, ready accessibility, and price of a framework are also important constraints for users. Commercial imperatives in selling framework data are readily understood. But a number of governments are also

Applications Box **19.6**

The British national framework

Britain's physical and governance characteristics and its history ensured that its national mapping differs significantly from countries having federal government structures, such as the USA: Britain is much smaller (one-fortieth of the USA in area) but the level of geographic resolution provided is over 100 times greater than the basic map information made available by USGS and is continuously updated. Almost all of the geographic framework is crafted by one central government body, Ordnance Survey (OS). Unlike US Federal Government bodies, this charges for its products and services (see Section 19.3.2); numerous partnerships exist with commercial organizations which are licensed to add value and distribute OS data (e.g., see Figure 19.6).

The core product from the national topographic database is MasterMap which has been produced by re-engineering the original OS large-scale (highly detailed) database, completed – as a world first – in 1995, into an object-oriented form. It comprises detailed topographic, cartographic, administrative boundary, postal address, and network features positioned on the National Grid plus an Imagery Layer. Every OS MasterMap feature has a unique identifier called a TOpographic IDentifier (TOID®), which is used to refer to the feature. Key features are complete

up-to-date coverage, the seamless data (i.e., no map tiles), orthorectified aerial imagery, topographic area features, and a topologically structured transport network.

All 440 million features are codified on a new classification basis with its unique identifier, a version number, and a defined life cycle linked to the 'real world object'. Data are available by theme as well as by geographic area; the customer can be supplied with 'change only updates'. From MasterMap data, other products can be spun off at lower levels of resolution. Figure 19.7 shows an example of some of the data layers.

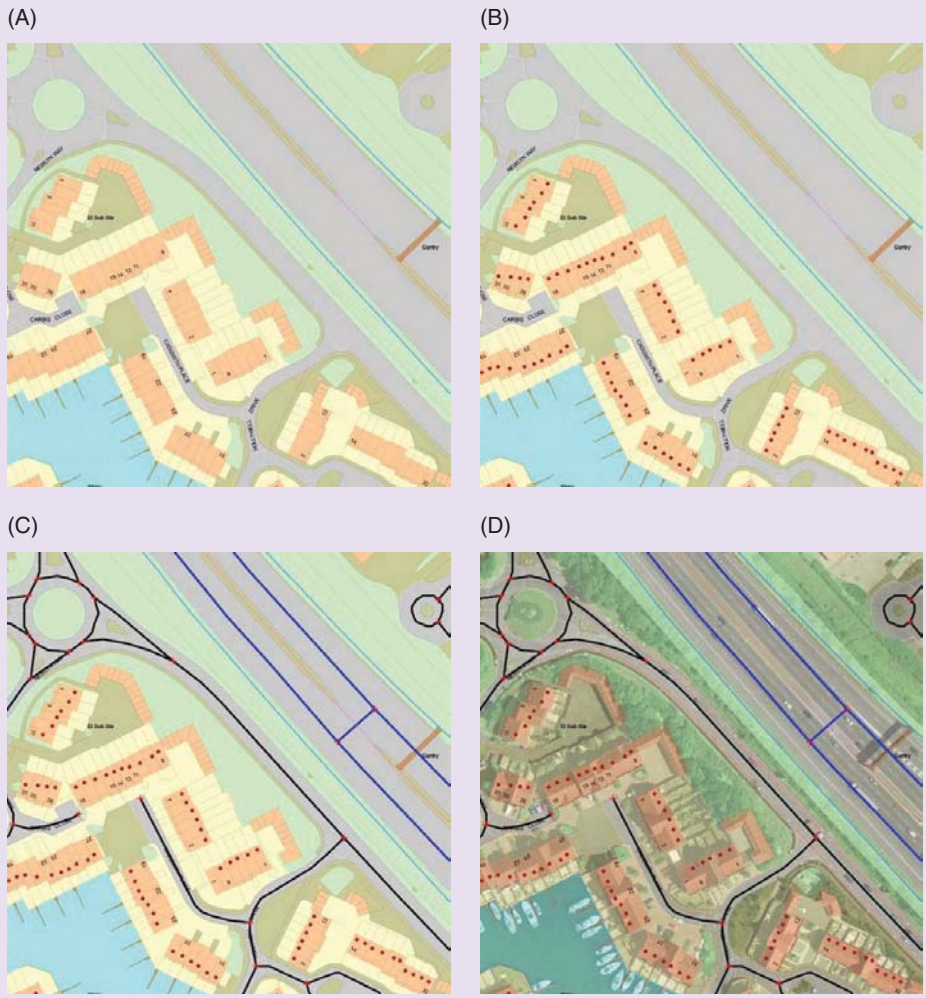


Figure 19.7 Sample of part of the Ordnance Survey MasterMap, showing various layers built up from the topography (A), with addresses added (B), with integrated transport links added (C), and with the orthophoto layer (D) (Courtesy: Ordnance Survey; Crown copyright reserved)

seeking to recover some of the heavy up-front costs of information (especially GI) collection through charging for its use.

Some of this has been brought about through governments worldwide reviewing their roles, responsibilities, and taxation policies and reforming their public services as a consequence. In some cases this has led to privatizations of government functions or at least out-sourcing. The imperative for some reviews has been financial, with a search for greater efficiency and wiping out of duplication (e.g., the US e-government initiative). In others, the imperative was ideological – that the state should do less and the private sector should do more.

The reform of government has led to two approaches in regard to GI. The first has been the outsourcing of production: thus large parts of the imagery created by many national mapping organizations (e.g., US Geological Survey) are now produced under contract by the private

Applications Box **19.7**

The US national framework

Essentially, the US situation has two parts – the traditional approach and a new one. The traditional manifestation is what is available out of the US National Mapping Program (NMP). The NMP incorporates the documented needs of 40 Federal agencies and the 50 States, which are solicited and analyzed as part of a continuous requirements assessment process. Products available include:

- Topographic quadrangle maps, derived originally from digitizing of 1:24 000-scale paper maps and available in vector and raster forms. On average, the contents of these maps were last substantially up-dated around 25 years ago. Less-detailed national coverage is also produced.
- digital elevation models, orthophotoquads, satellite images derived from a variety of sources, and the Geographic Names Information System (GNIS). The latter forms the official national record of domestic geographic names information.

Given its difficulty in addressing all the myriad needs for up-to-date framework data and the increasing role of other players in providing such data across the nation, USGS announced in 2001 the concept of The National Map (TNM). This plan involves a novel partnership approach drawing data from multiple (other federal agencies, states, counties, academic, and private sector) sources to create, make consistent, and distribute much-more-current framework data. A highly innovative USGS initiative was to seek to estimate the creation costs and the application benefits of building a US national map in this fashion; though some of the estimates made are (necessarily) heroic and the projected costs seem modest, the simulation study was ground-breaking and suggested a break-even time for a successful TNM of 14 years and a net benefit of between $1 and $3bn over a 30 year lifetime. At the time of writing, the ambitious TNM vision was fairly clear but the reality is embryonic (see **nationalmap.usgs.gov/**). The relationship between this and various other initiatives is described in Chapter 20.

sector. Another approach has been simultaneously to force economies within government and get the user to pay for the bulk of GI. In some European countries, for instance, land records and the cadaster are paid for totally through user fees; since many people do not own land or property (around 30% in the UK), it is deemed unfair for them to contribute to the costs via general taxation. Also, it is worth stressing that this charging is nothing new in principle: historically, almost all government organizations have made *some* charge for data, including Federal Government bodies in the USA. The only issue is for what elements users should be charged, e.g., the whole product or only the costs of documentation and distribution. The use of the Internet has dramatically changed this equation (see Section 18.4.4).

The advantages and disadvantages of the 'charging for everything' and 'free access' approaches are summarized in Box 19.8, though these are subject to much argument. In Britain, the on-going costs of running the national mapping organization are met totally from revenues raised from users. Perhaps as a consequence, the main elements of the database are guaranteed never to be more than six months out-of-date. Figure 19.8 shows how the proportion of funding arising from the users rose over many years. That said, the policy also has some downsides, as suggested in Box 19.8, and policy shifts have happened in various countries – sometimes several times (see Box 19.9).

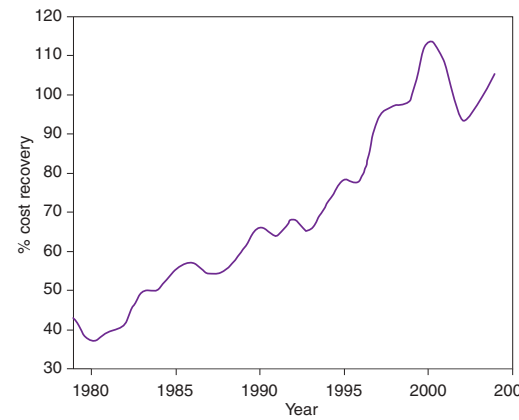


Figure 19.8 Cost recovery by Ordnance Survey which operates a 'user pays' model for digital geographic information and paper maps. Prior to 1999, Ordnance Survey was set a percentage cost recovery target by government that rose steadily towards 100% from a start of 25% in the mid 1970s. Since then, Ordnance Survey has been tasked with a target of achieving a surplus on operating activities. In the graph, post-1999 cost recovery has been calculated as trading revenue as a percentage of operating costs. Exceptional costs (provision for substantial early retirement programs), exceptional revenues (settlement with the Automobile Association's new owners – see Section 19.1.2), and net interest movements have been excluded from the calculation. The dip below 100% since 1999 was the result of a planned investment program

Applications Box **19.8**

The advantages and disadvantages of the 'user pays' and 'free access' philosophies (*and some contrary views*)

Arguments for cost recovery from users:

- Charging which reflects the cost of collecting, checking, and packaging data actually measures 'real need', reduces market distortions, and forces organizations to establish their real priorities (*but not all organizations have equivalent purchasing power, e.g., utility companies can pay more than charities; differential pricing invites legal challenge*).
- Users exert more pressure where they are paying for data and, as a consequence, data quality is usually higher and the products are more 'fit for purpose'.
- Charging minimizes the problem of subsidy of users at the expense of taxpayers – about a 30 000 to 1 ratio globally (*but some users are acting on behalf of the whole populace, such as local governments*).
- Empirical evidence shows that governments are more prepared to part-fund data collection where users are prepared to contribute meaningful parts of the cost. Hence full data coverage and update is achieved more rapidly: the average age of the traditional US basic scale mapping is over 30 times that in the UK (*but once the principle is conceded, government usually seeks to raise the proportion recovered inexorably*).
- It minimizes frivolous or trivial requests, each consuming part of fixed resources. (*This rationale is now less relevant given the user-driven access to the Internet in many countries*).
- It enables government to reduce taxes.

Arguments for dissemination of data at zero or copying cost:

- The data are already paid for hence any new charge is a second charge on the taxpayer for the same goods (*but taxpayers are not the same as users – see above*).
- The cost of collecting revenue may be large in relation to the total gains (*not true in the private sector – why should it be so in the public sector?*).
- Maximum value to the citizenry comes from widespread use of the data through intangible benefits or through taxes paid by private sector added-value organizations (*this contention is neither provable or disprovable because of the many factors involved*).
- The citizen should have unfettered access to any information held by his/her government, other than that which is sensitive for security or environmental protection reasons (e.g., the location of nests of endangered species).

19.4 Navigating the constraints

We have seen that various legal, staff skills, information availability, and other constraints can complicate the full exploitation of a GIS severely. The introduction of a GIS will introduce risk to an organization – though in the longer term it may well reduce it. Each constraint or problem can be tackled individually but a much better approach is to have a GIS strategy and carry out regular reviews of risk. By being proactive, many constraints can be avoided or minimized.

19.4.1 The role of the GIS strategy

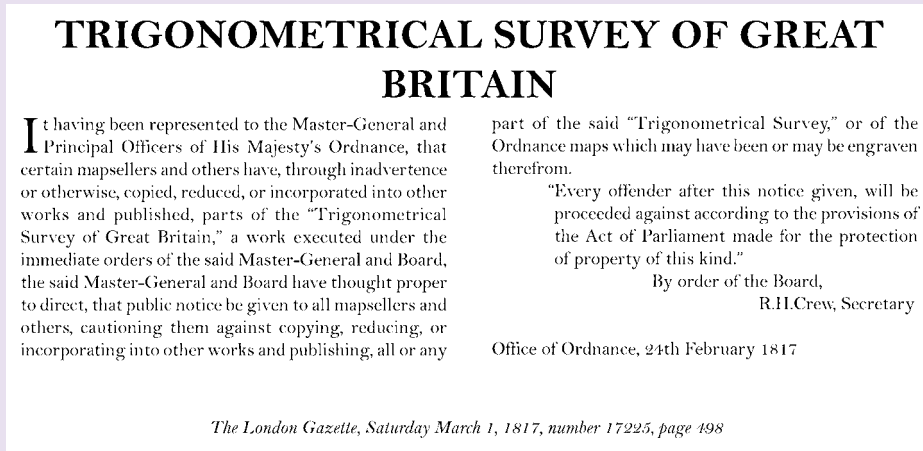
Strategies set direction, based upon the organization's aims, its resources, and current position. It will be rare for any organization to have a wholly free-standing GIS strategy: normally, it will have to fit within and support the organization's overall strategy and business plan. Moreover, it will in turn also need to fit and be supported by other strategies with which it will be associated – such as the finance, human resources, information systems, and services and estates strategies, and possibly others such as external relations/publicity and international ones. Sometimes the GIS strategy may simply be embedded, even buried, within the information technology, systems, or services strategy. We think that this is often a mistake because of the cross-cutting nature of GIS and (especially) GI across many levels and (potentially) most of the organization's activities – part of why we stated that 'spatial is special' in Section 1.1.1.

Here we treat the GIS strategy as a free-standing entity for illustrative purposes. The particular form of that strategy will vary depending on how GIS is seen within the organization (see Chapter 7.3.1). Finally, strategies are useless without an implementation plan, senior management commitments to proceed, and milestones at which progress is monitored. Most of these issues are dealt with in more detail in Chapter 18.

Applications Box **19.9**

Swings and roundabouts of public policy on GI

Following the policies of the British government of the day, Ordnance Survey took action to prevent 'free riding' or illegal use of the government's maps at least as early as 1817. They warned expressly anyone infringing their copyright that this would lead to legal action (see Figure 19.9). In the 1930s world of mapping, however, public policy changed to a 'charge only for ink and paper' basis; in the 1960s it changed back again to charging users a fraction of the total costs of operations. In 1999, the move to a Trading Fund status effectively led to OS having to become profitable and thus meet full costs and interest on capital (but see Section 18.3.2 on whether all the assets were capitalized).



TRIGONOMETRICAL SURVEY OF GREAT BRITAIN

It having been represented to the Master-General and Principal Officers of His Majesty's Ordnance, that certain mapsellers and others have, through inadvertence or otherwise, copied, reduced, or incorporated into other works and published, parts of the "Trigonometrical Survey of Great Britain," a work executed under the immediate orders of the said Master-General and Board, the said Master-General and Board have thought proper to direct, that public notice be given to all mapsellers and others, cautioning them against copying, reducing, or incorporating into other works and publishing, all or any part of the said "Trigonometrical Survey," or of the Ordnance maps which may have been or may be engraven therefrom.

"Every offender after this notice given, will be proceeded against according to the provisions of the Act of Parliament made for the protection of property of this kind."

By order of the Board,
R.H.Crew, Secretary

Office of Ordnance, 24th February 1817

The London Gazette, Saturday March 1, 1817, number 17225, page 498

Figure 19.9 The earliest known warning against infringing GI copyright: the warning printed by the Board of Ordnance in the *London Gazette* of 1 March 1817

19.4.1.1 Getting to the GIS strategy and the implementation plan

Organizations achieve things routinely through standard processes. An illustrative top-down process view of how to get to a GIS strategy is shown in Figure 19.10. To aid realism, we use a hypothetical example of a mid-size local government, with perhaps a thousand or so employees. Sometimes things happen in such a top-down way but often, in practice, GIS strategies are grown partly from the 'grass-roots'. And the strategy would be different in some respects in a commercial enterprise or a government one which sells data; a pricing sub-strategy would contribute to the overall GIS and organizational one (see Box 19.10 for some elements of it).

In essence, getting to a GIS (and corporate) strategy requires a statement of the current situation, agreed by all the key players as an accurate summary. This needs to cover the reason why the organization exists, its strengths, and how well it is achieving its current aims. Also, it needs to describe any major shortcomings and the causes of them (based largely on customer, client, or 'owner's' views). From this must come some vision of where the organization should go. The strategy must include summaries of the threats from others and the new opportunities available, plus the constraints that will need to be circumnavigated. In addition, a new strategy needs to be underpinned by a coherent summary of the organization's culture and values – or new ones if these need to be changed. To enable implementation of the strategy, an explicit description of the present structure and operational management and a process map of the organization are required. All of this needs to be as dispassionate as possible – which is why consultants are often called in to help.

Any strategy will fail if it does not contain a statement of all the major changes necessary and a comparison of these with the financial, staffing, and other constraints plus an assessment of their ramifications. These may involve changing the numbers, disposition, and skills of staff, changes of software, the need to acquire or create additional data/information or technology, and much else. Finally, it is crucial to get 'buy in' from senior managers and the staff concerned.

19.4.1.2 Identifying risk

Risk identification and management is fundamental to the success of any organization – and to its strategies. Risk can arise internally from new initiatives, simply from on-going operations, or because of external changes. It is important to realize that, inherently, some risks are

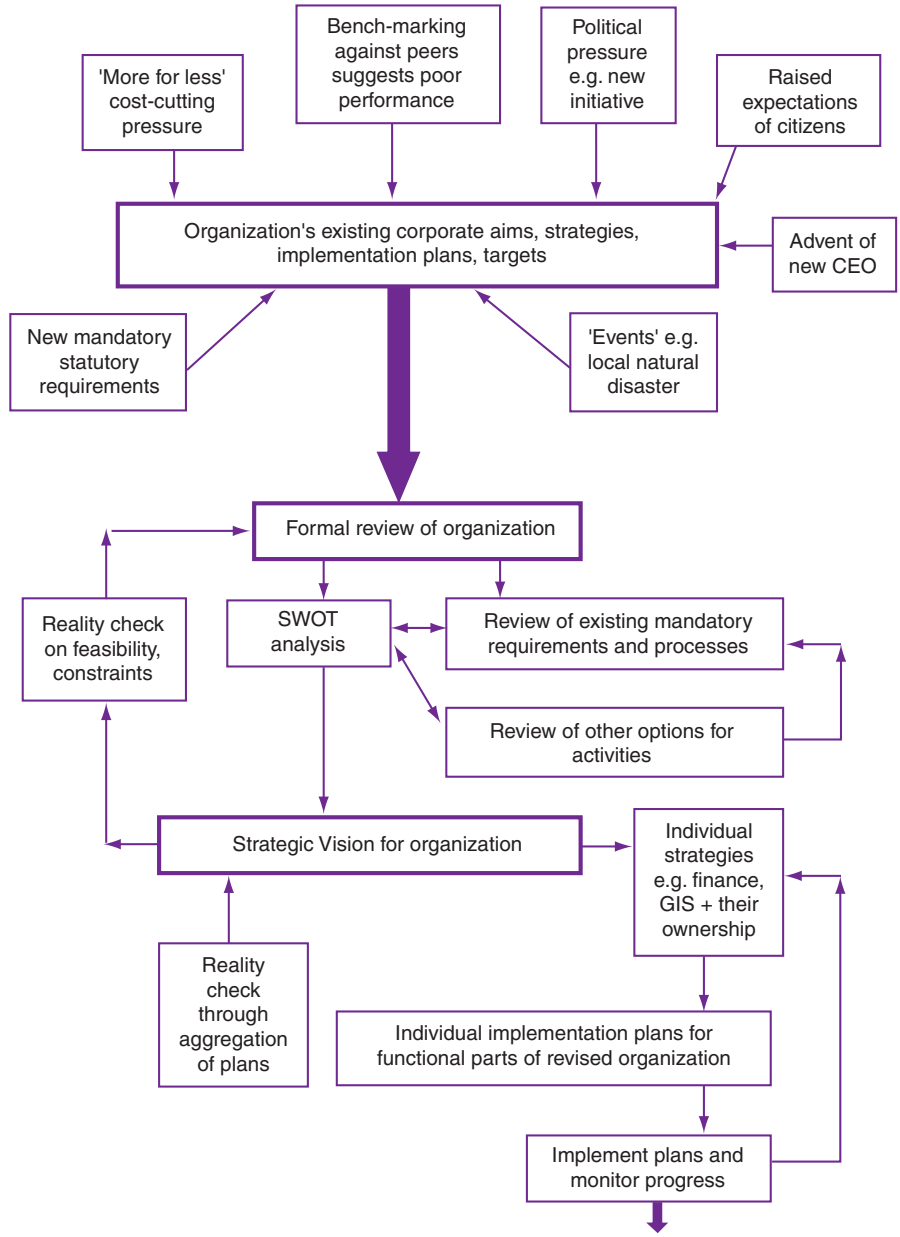


Figure 19.10 External and internal pressures on a local government organization and consequential moves to a corporate and a GIS strategy and implementation plans

unpredictable in detail though their aggregate likelihood can sometimes be insured against (e.g., via professional indemnity insurance). Failure to control risk can lead to financial disaster, to public humiliation for the organization, and the loss of jobs, especially of senior personnel. Central to the notion of risk is accountability. It is essential on a practical basis to 'lock in' your superiors in the project, obtaining their agreement to the strategy and key operational decisions.

Different types of risk need to be considered separately (Table 19.2). In general, risks are greater when GIS are being used for wholly new activities within organizations, rather than when such systems are being used to do existing things better or cheaper, especially if they are introduced incrementally (see Section 17.1).

On the other side, GIS is now a major contributor to the estimation and control of risk in the world's largest insurers – in tasks ranging from estimates of creditworthiness of individuals to monitoring the distribution and frequency of natural hazard events, especially major natural disasters (see Section 21.4). Indeed, technology is changing the very nature of how risk is assessed: 'pay-as-you-go' insurance, with charges varying by the speed at which you are driving and the levels of traffic and road conditions, is now possible thanks to the integration of GPS and GIS (Figure 19.11).

Applications Box **19.10**

Building sub-strategies – an information pricing example

Getting pricing wrong is usually disastrous. In general, the information owner should set a strategy which:

- involves setting prices by perceived value, rather than cost of production;
- achieves cost leadership through economies of scale and scope; and/or
- differentiates the products, i.e., creates multiple versions of the same product differing in accuracy, appearance, and currency and creates different products by linking different datasets together and extracting 'new' information goods (see Box 19.5). Avoid letting the information become a commodity, i.e., where everyone is selling the same good. In that case, fierce price competition will decide who wins.
- personalizes the pricing as well as the product. One way to achieve this is to group prices (e.g., for all local governments) via some kind of sector-wide deal or Service Level Agreement, thereby permitting bulk discounts quite legally. Ordnance Survey Great Britain now has one deal – the Pan-Government Agreement – with 280 departments, agencies, and other central government bodies in the UK and has parallel collective deals with all the utility companies and all the local governments in Britain.
- seizes 'first mover' advantages (i.e., gets into the market first and establishes brand image, customer base, etc.) but continues to innovate. Avoid greed if you are the market leader. For example, employ 'limit pricing', which are prices set just low enough to discourage competitors from investing to join the market since they will never make a suitable return on their employed capital. And 'play tough' – a new potential competitor should know that you are willing to lower your prices aggressively if necessary and sue for any infringement of intellectual property rights.
- is built upon understanding of the markets and their current profitability and use of promotions to measure demand.

Table 19.2 A classification of risk faced by organizations

Major risk category	Sub-category	GIS-related example
Business risk	Product/service risk	Implementation of the system over-runs financial or time budgets. Analyses, outputs, outcomes found flawed or unconvincing due to software inadequacy or staff incompetence
	Economic risk (via change of markets)	No-one wants 'line maps' when up-to-date image maps available?
	Technology risk (out-of-date c.f. competitors)	Competitors can achieve the same results (e.g., analyses, maps) faster, better, more cheaply
Event risk	Reputation risk	Erroneous or delayed results lead to collapse in confidence in organization, fanned by the media (e.g., 'State denounces GIS firm for unprofessional work.')
	Legal risks	As above, leading to legal action to recover damages and/or costs
	Disaster risks	Software incapable of doing what was promised; fire destroys all facilities including data and results to date
	Regulatory or policy shift risk	Government decrees that information policy has changed or that all materials contributing to end results have to be in public domain
	Political risk	Elected official supports competitor
Financial risk	Credit, market, liquidity, operational, price, delivery, etc., risks and fraud	Other 'normal' business risks – all must be guarded against

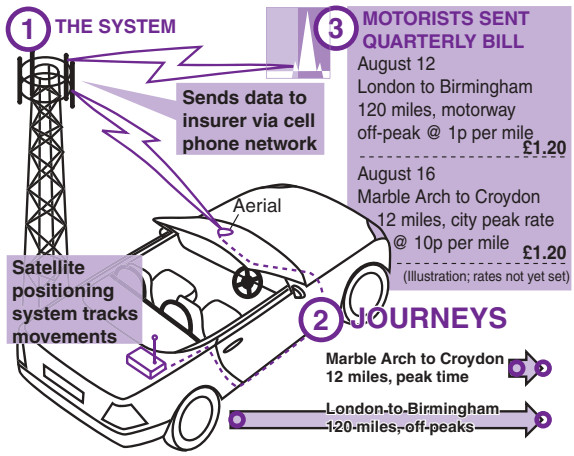


Figure 19.11 The pay-as-you-go car insurance scheme being piloted by Norwich Union in the UK – an enterprising combination of new technology and digital maps to differentiate charging for different customers with different driving habits (based on a risk assessment)

19.4.2 Managing risks in GIS projects

Managing risk within significant GIS projects requires at least two appraisals, carried out at different stages. These may overlap with – and be part of – the business case for a GIS. The appraisals are:

- a sound investment appraisal before commitment is made to proceed;
- a risk appraisal to identify the significant factors which may go wrong in implementation of the project.

The investment appraisal is needed because someone above you in the management chain needs to analyze your logic and test it to destruction. The opportunity costs of going ahead also need to be exposed. In many organizations, there is only a finite amount of financial resource available for new developments. This limitation may arise from an unwillingness to add to the tax burden (in the public sector) or from restrictions on bank or market support (in the private sector). Such investment appraisals also need to be done when money is not a major problem: the most precious commodity in many businesses is senior management skills and it may be necessary to use these to take forward or to rescue another project. In short, the GIS business case is a necessary but not sufficient case for proceeding – a more wide-ranging investment appraisal is needed and rarely can this be done by GIS staff alone. It needs at least to be sanctioned by the finance director or his/her equivalent.

Managing risk requires both an investment appraisal and a risk appraisal.

An investment appraisal needs to cover at least:

- The predicted costs of people, materials, hardware and software, training, and other organizational costs insofar as they can be quantified, all broken down by time.
- The predicted costs and benefits (see Box 17.4) and a comparison with those of the *status quo*, i.e., the 'do nothing option'.
- Sensitivity analyses of the impacts of variations on the plan on the business case. Such variations could include cost over-runs, failure to achieve the predicted benefits in the predicted time, etc.
- Conclusions on the basis of all the above, preferably agreed individually with as many of the key players as possible before the formal consideration of the appraisal.

Anticipation of risk requires a formal process and people charged to identify and manage it. No project should start without a table of potential risks, classified by probability and anticipated magnitude of impact (which may be very different) and by steps to be taken to reduce the probability and/or minimize the impact. Moreover, the risk assessment is iterative: after every major event or even periodically, the process shown in Figure 19.12 must be repeated.

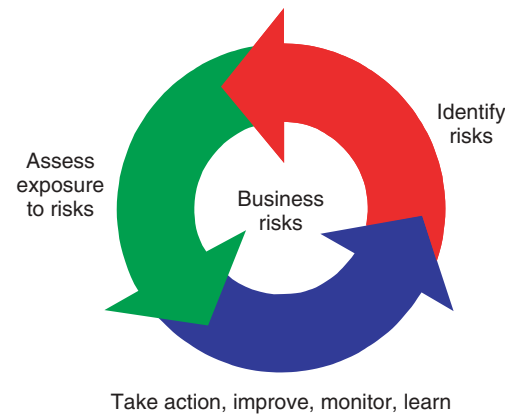


Figure 19.12 The iterative nature of risk identification, assessment, and management

19.5 Conclusions

Throughout this book we have argued that GIS can – and increasingly does – underpin many of the activities of businesses of all types, whether academic, governmental, not-for-profit, or commercial. To succeed, these businesses need to exploit their assets, including all the capabilities that GIS can provide – and avoid disaster through following a strategy, anticipating risk, and navigating past the constraints, i.e., being business-like is essential. Such plans and actions need to be reviewed frequently: the impacts of technology change, the advent of new competitors, and changing expectations on the part of the business's stakeholders ensure that tomorrow will not be the same as today.

Questions for further study

1. Who owns the geographic information you use? What rights do you have in it? How would you protect your organization's GI against misuse by competitors?
2. If you were being educated for your own job, what would be the ideal components of a GIS course? Is distance-based learning for GIS a satisfactory approach? For what types of personal development is it best suited?
3. Suppose tomorrow morning you receive a writ for damages from a commercial firm for whom you have carried out consultancy. It alleges your results – on which they acted – are now demonstrably false and you did not take due care and attention. What do you and your lawyer do first? How do you 'cover your back'?
4. How would you go about creating a new GIS strategy for your organization and what are the main risks that you have to guard against?

Further reading

Cho G. 2005 *Geographic Information Science: Mastering the Legal Issues*. John Wiley & Sons Ltd., Chichester, UK.

Photogrammetric Engineering and Remote Sensing. Special Issue on The National Map. 2003 (69).

Shapiro C. and Varian H.R. 1999 *Information Rules*. Cambridge, MA: Harvard Business School Press.

20 *GIS partnerships*

'Business' (as we define it) may operate within commerce, government, the not-for-profit sector, and in academia – or between them. Many aspects of it are often highly competitive, even (though it might be denied) in government. But competition is not always the right way: frequently collaboration pays. This chapter examines how collaboration has occurred in GIS and GI, the extent to which it has worked, and the factors that sometimes make it fail. It covers collaborations at the local, national, and the global levels – and seeks to separate the hype from the reality. The US National Spatial Data Infrastructure – one of the truly 'big ideas' in the GIS world – is reviewed and lessons are drawn. Finally, the impact of overwhelming external events as business drivers is considered, based on the events of 11 September 2001 and its aftermath. These have raised crucial issues about the balance between collaboration and coercion in GIS – and beyond.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter, you should know about the value of GIS partnerships through familiarity with:

- The characteristics, benefits, and some of the potential pitfalls of collaboration;
- National spatial data infrastructures;
- The need for globally consistent GI;
- The importance of political support for GIS;
- The role of extreme events as business drivers.

20.1 Introduction

Chapters 18 and 19 have emphasized the competitive nature of our world. But there are many occasions in GIS on which we find it best to collaborate or partner with other institutions or people. The need for partnerships applies at the personal, institutional, local area, national, and global levels – so there is great scope for different approaches. The form of the partnerships can range from the highly formal, based on contract, to the informal where participation is entirely voluntary. A good example of the latter is GIS Day (see Box 20.1). Some partnerships result from events which span national boundaries (like Chernobyl), and some are both multi-national and created and driven by short-term events. Figure 20.1 shows a locust infestation which demanded collaboration in data collection and pest eradication between FAO and various African countries in August 2004, with financial support by other members of the United Nations (**www.fao.org**). In this chapter, we look at some major enterprises based on collaboration – including one that failed and another that faded away – to attempt to draw some general conclusions; and we consider how global terrorism seems to be changing some aspects of GIS partnerships.

Partnerships can add vital staff skills, technology, marketing know-how, and brand image to your capabilities, leading to fresh insights on the real user needs, new products, and good publicity. It can also lead to cost- and risk-sharing.

20.2 Collaborations at the local level

There are around 80 000 different public-sector governing bodies in the USA (e.g., counties, school boards) – perhaps the most complex structure anywhere in the world, with different jurisdictions overlapping in geography. The size of these organizations and the resources available to them ensure that, for some administrations, investment in GIS is a heavy (or sometimes impossible) burden: Figure 20.3 illustrates the town hall in the county seat in one small county, which is open by appointment. The scope for collaboration or partnerships at this level – typically involving operational matters or sharing of knowledge – is considerable.

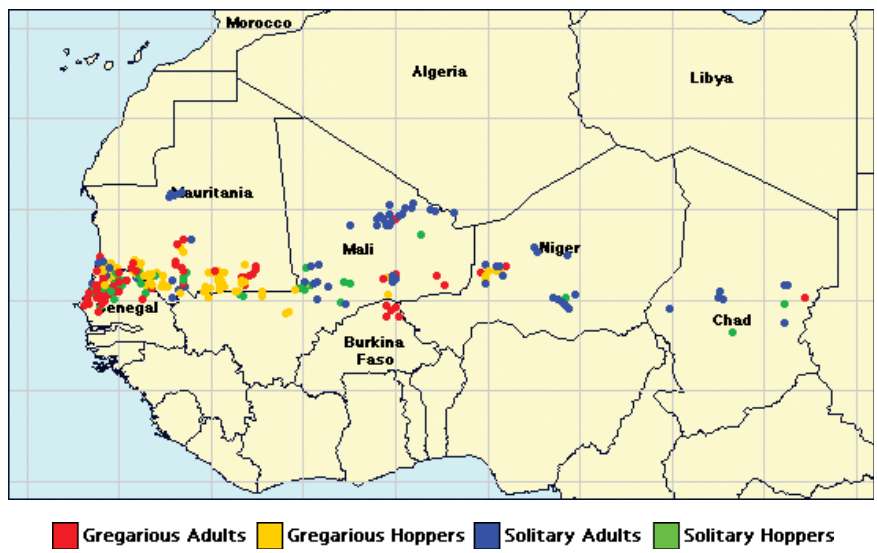


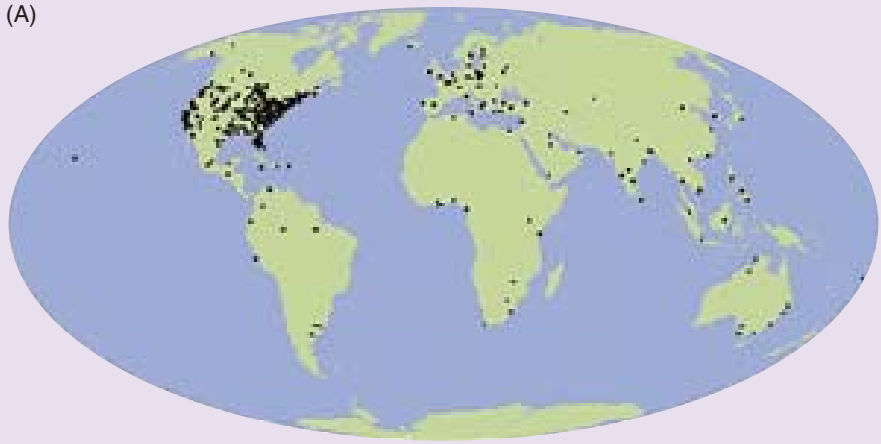
Figure 20.1 Locust 'hot spots' in West African countries in August 2004 (*Source*: FAO)

Applications Box **20.1**

GIS Day: a global but informal partnership

GIS Day is an annual event which began in 1999. It is part of the US National Geographic Society's Geography Awareness Week, designed to promote geographic literacy in schools, communities, and organizations. The aim of GIS Day is for GIS users and vendors to open their doors to schools, businesses, and the general public to showcase real-world applications of the technology. News of the event is spread by use of the Internet and by advertising. Any organization can host such an event. Potential visitors can search for the nearest site to themselves. No less than 2400 organizations hosted GIS Day events in more than 76 different countries in 2003 (see Figure 20.2(A) and 20.2(B)). It was estimated that over two million children and adults were educated on GIS technology through geography on that day.

(A)



(B)

Figure 20.2 (A) The distribution of worldwide GIS Day events in 2003. (B) A GIS Day event underway at Kyrgyz State University of Construction, Transportation and Architecture (KSUCTA), Bishkek, Kyrgyzstan

Figure 20.3 The town hall in Franconia, Minnesota where meetings with the part-time administration are held by appointment when it is open (Reproduced by permission of Francis Harvey)

Nowhere is this more significant than in Public Participation GIS (PPGIS). These have been variously defined (see Section 13.4), including:

- being an interdisciplinary research, community development and environmental stewardship tool grounded in value and ethical frameworks that promote social justice, ecological sustainability, improvement of quality of life, redistributive justice, nurturing of civil society, etc.;
- endeavoring to involve youth, elders, women, First Nations, and other segments of society that are traditionally marginalized from decision-making processes.

Box 20.2 provides an example of such community-based GIS projects. Central to successful work in this area is good communication: once data and analysis are credible (not necessarily highly accurate or complex), having a sound strategy to introduce that information to targeted audiences is essential. The power of high-quality, well-designed maps to persuade decision makers is a crucial asset.

20.3 Working together at the national level

In some cases, national associations of local bodies have been set up to share knowledge and information and even to act as a co-ordinating mechanism. The 400 or so local governments in the UK, for instance, formed a Geographic Information House to trade in information and have forged a partnership with the Land Registry and other central government partners to create a National Land Information Service (**www.nlis.org.uk/docs/index.htm**).

Beyond such associations of like bodies, we now see truly national partnership initiatives in a number of countries. Most are based around some form of national spatial data infrastructure (or NSDI) – one of the 'big ideas' in GIS. It is believed that all of the countries in Table 20.1 have or are implementing partnerships through forms of NSDI. Web addresses of these NSDIs are given in the Global Spatial Data Infrastructure (GSDI) website (**www.gsdi.org/SDILinks.asp**). The nature of these initiatives, their title, who is involved (e.g., see Box 20.3), and much else differs from country to country but there are also a number of striking similarities – reflecting global convergence in the nature of some of the biggest problems faced in many countries.

We use the US example as a case study since this was probably the earliest coherent scheme, it has influenced many others, and it was initially underpinned by the leader of the nation state.

Many countries are implementing GI partnerships through varied forms of national spatial data infrastructure.

20.3.1 The US National Spatial Data Infrastructure

The US 'Strategy for NSDI' document was initially published by the Federal Geographic Data Committee (FGDC) in 1994 and revised in April 1997. Box 20.4 summarizes FGDC's view of the problem NSDI was intended to fix. As ever, a few individuals were important to what actually happened: Box 20.5 describes how one key player contributed to the formative stages of NSDI.

20.3.2 Getting towards a solution

The origins of the NSDI in the early 1990s involved the Mapping Science Committee (MSC) of the United States National Research Council and the US Geological Survey. From the outset, NSDI was seen as a comprehensive and co-ordinated environment for the production, management, dissemination, and use of geospatial data, involving the totality of the relevant policies, technology, institutions, data, and individuals. A 1994 MSC report urged specifically the use of partnerships in creating the NSDI.

Table 20.1 Countries with NSDIs in August 2004 according to the Global Spatial Data Infrastructure website (**www.gsdi.org/SDILinks.asp**)

Australia	France	Cambodia	South Africa
Canada	Germany	Malaysia	Spain
Chile	Hungary	Nicaragua	Sweden
China	India	Norway	Switzerland
Colombia	Indonesia	Philippines	The Netherlands
Cuba	Ireland	Poland	United Kingdom
Czech Republic	Iceland	Portugal	United States
Denmark	Israel	Russia	Uruguay
Dominican Republic	Italy	Salvador	Venezuela
Finland	Japan	Slovenia	

Applications Box **20.2**

A successful PPGIS

Larry Orman summarized his experience in PPGIS at the 2003 conference on that topic, together with his work in GreenInfo Network (**www.greeninfo.org/**). Figure 20.4(A) is an example of a GreenInfo map which helped conservation advocates to show what could happen if suburban sprawl were to continue and forced answers to the question: 'should

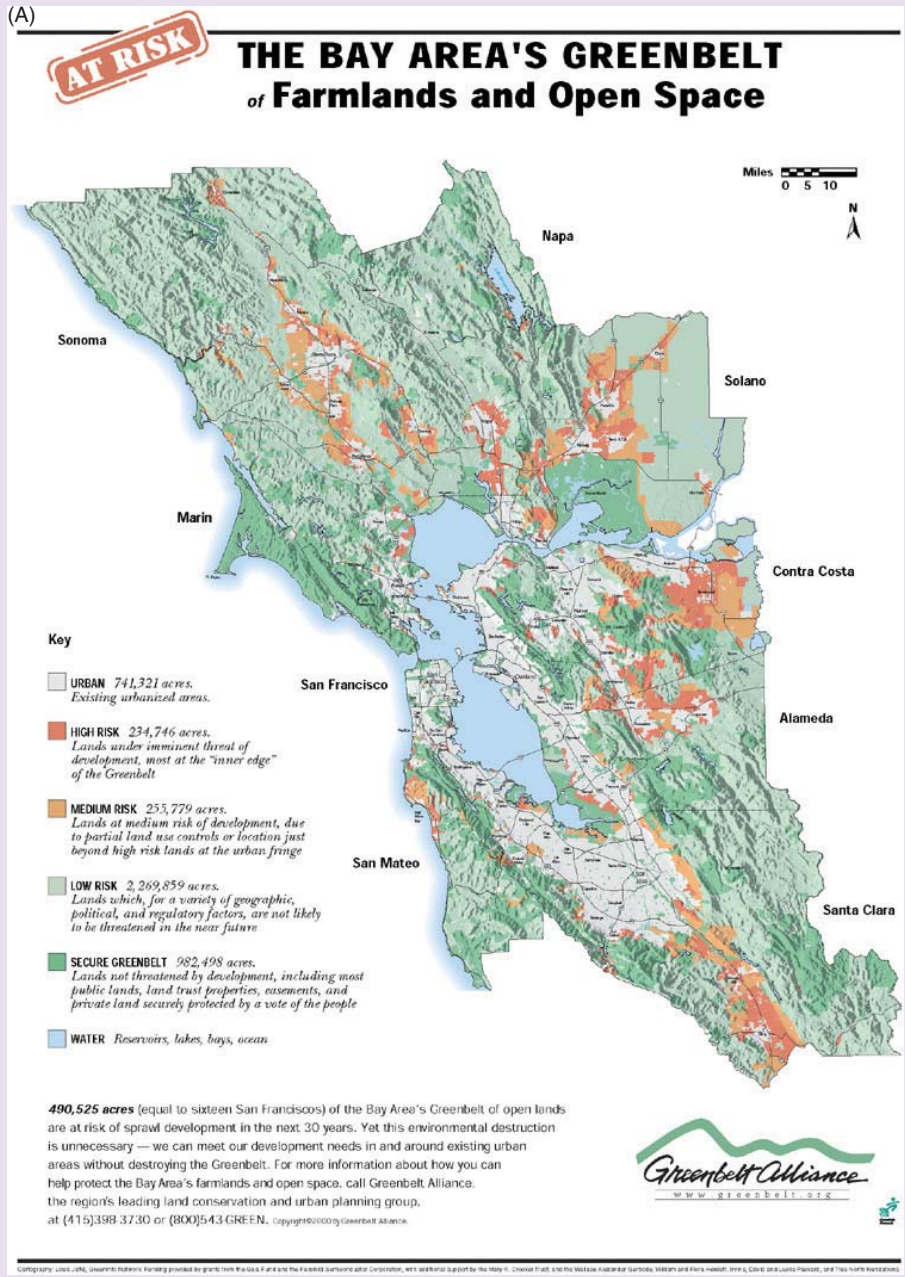


Figure 20.4 (A) Map showing threats from advancing urbanization in the Bay Area produced through a public participation GIS project carried out by the GreenInfo Network, the source of the map. (B) A simulation of wind turbines on a Scottish hillside to enable the local population to visualize a development proposal (*Source*: Macaulay Institute, Aberdeen)

(B)

Figure 20.4 (*continued*)

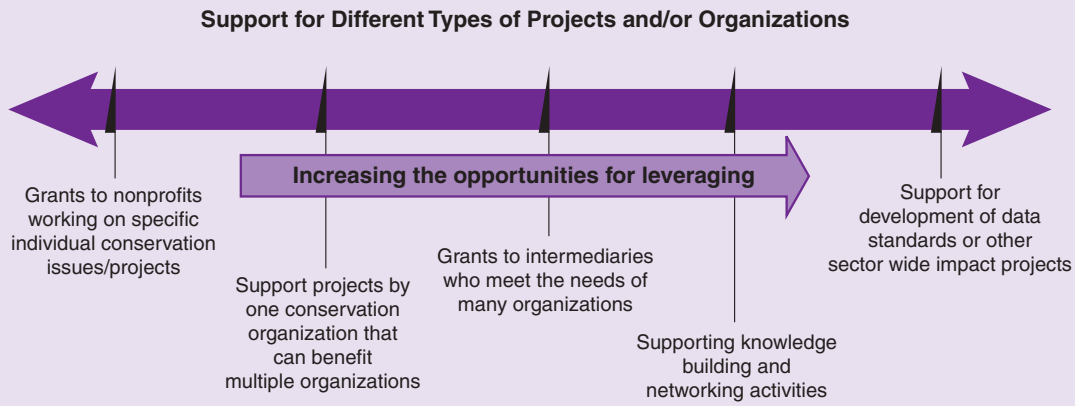


Figure 20.5 The differential leverage of GIS public participation projects. Greater long-term benefit tends to come from working towards the right end of the scale (Reproduced by permission of GreenInfo Network)

the future be this way?'. Highly compelling visual design helped the project to obtain good coverage in the news media.

For many public interest groups, GIS software and data are still too complicated, expensive, and frustrating. To circumvent this problem, GreenInfo Network was set up to be a service center for non-governmental organizations (NGOs), especially in California. The Network reached a critical mass of staff and began producing highly influential work approximately three years later. Financial stability is crucial for such organizations: getting grants or gifts for hardware, etc., is relatively easy initially but these soon dry up so project funding must be obtained. But 'one off' projects often have only a very local effect: Orman argues for choosing to do projects which leverage the effort in various ways (Figure 20.5). In short, community groups need to be business-like in how they operate if they are to continue to prosper (see Section 18.1)

Typically, GreenInfo Network works in the areas of: conservation for land trusts, public agencies, advocacy organizations, and various foundations; for a range of advocacy groups in environmental matters; and for research organizations, service providers, and others in social well-being matters. It helps to link up client groups with projects and information from other clients, building broader education about geographic literacy and engagement. The organization's staff of eight GIS specialists usually works with 30 to 40 groups at any one time and collaborated with over 300 in the period 1999–2004. Most projects are in the $1000–10 000 range, with a few in the $25 000–75 000 range.

At the time of writing, there were two other independent non-profit GIS service centers besides GreenInfo (New York – **www.nonprofitmaps.org**, and Seattle – **www.commenspace.org**), both founded in the late 1990s. There are, of course, also many thousands of community groups working in the USA and elsewhere globally. Therefore, use of GIS tools to visualize proposed developments and communicate these to the public (e.g., see Figure 20.4(B)) is becoming common, helping to sustain local democracy. But, typically, such success is only achieved through successful partnering and investment in people.

Biographical Box **20.3**

Ian Masser: European SDI pioneer

Figure 20.6 Ian Masser

Ian Masser's early work was as a computer-literate urban planner with a human geography background. He became involved in matters relating to GI policy as the national coordinator of the UK Regional Research Laboratory (RRL) initiative in the late 1980s. This coincided with the British Government's publication of the pioneering 'Chorley report' on *Handling Geographic Information* in 1987. This report was accepted immediately by the UK government and was an important milestone in the development of thinking about GI policy.

In 1993 he became co-director of the European Science Foundation's GISDATA scientific program. This was funded by 15 National Science Councils and also had strong links with the US National Center for Geographic Information and Analysis. An important spin-off from the program was the establishment of the Association of Geographic Information Laboratories in Europe (AGILE) in 1998. He was involved in the creation of an independent European Umbrella Organisation for Geographic Information (EUROGI) in 1993, the debates regarding a European Geographic Information Infrastructure between 1995 and 1999, and the first GSDI Conference in Bonn in 1996. He published a comparative evaluation of NSDIs in Australia, Britain, the Netherlands, and the US in 1998. Based in the Netherlands from 1998, Ian undertook missions to places as far afield as China, Colombia, India, Peru, and Vietnam.

EUROGI (see above) has 22 member organizations which collectively represent more than 6500 organizations in total; it has been actively involved in various European Commission-funded SDI-type projects. Ian Masser was its President from 1999 to 2003. He became President of the Global Spatial Data Infrastructure Association in 2002 with a remit to make it operational rather than simply a 'nice idea'.

A lever for realizing the vision was the Clinton Administration's wish to 'reinvent' the federal government in early 1993. Staff of the US Federal Geographic Data Committee (FGDC) outlined the concept of the NSDI as a means to foster better intergovernmental relations, to empower state and local governments in the development of geospatial datasets, and to improve the performance of the federal government. In September 1993, the NSDI was listed as one of the National Performance Review (NPR) initiatives to reinvent that government.

Applications Box **20.4**

FGDC's statement of the problem

In the United States, geographic data collection is a multibillion-dollar business. In many cases, however, data are duplicated. For a given piece of geography, such as a state or a watershed, there may be many organizations and individuals collecting the same data. Networked telecommunications technologies, in theory, permit data to be shared, but sharing data is difficult. Data created for one application may not be easily translated into another application. The problems are not just technical – institutions are not accustomed to working together. The best data may be collected on the local level, but they are unavailable to state and federal government planners. State governments and federal agencies may not be willing to share data with one another or with local governments. If sharing data among organizations were easier, millions could be saved annually, and governments and businesses could become more efficient and effective.

Public access to data is also a concern. Many government agencies have public access mandates. Private companies and some state and local governments see public access as a way to generate a revenue stream or to recover the costs of data collection. While geographic data have been successfully provided to the public through the Internet, current approaches suffer from invisibility. In an ocean of unrelated and poorly organized digital flotsam, the occasional site offering valuable geographic data to the public cannot easily be found.

Once found, digital data may be incomplete or incompatible, but the user may not know this because many datasets are poorly documented. The lack of metadata or information on the 'who, what, when, where, why, and how' of databases inhibits one's ability to find and use data, and consequently, makes data sharing among organizations harder. If finding and sharing geographic data were easier and more widespread, the economic benefits to the nation could be enormous.

Biographical Box **20.5**

Nancy Tosta: NSDI pioneer

Figure 20.7 Nancy Tosta

Nancy Tosta is a Vice President at Ross & Associates Environmental Consulting, Ltd in Seattle, Washington, USA, where she seeks to assist federal government agencies with their environmental, public health, and information challenges. From 1992–1994, however, she was centrally involved at the birth of the US National Spatial Data Infrastructure (NSDI). As the first Staff Director to the Federal Geographic Data Committee (FGDC), housed within the US Department of the Interior, she was a major source of thought and energy behind the initial conceptualization and implementation of the US NSDI.

She was the lead drafter of the April 1994 Presidential Executive Order 12906 that established the components of the NSDI as a combination of partnerships, standards, framework data, and data clearinghouses. She oversaw the FGDC development and approval of the first geospatial metadata standard. She was a firm advocate of the 'bottom-up, patchwork quilt' approach to spatial data infrastructures, believing strongly in the need for spatial data efforts to be rooted in serving community needs – be they in cities, counties, watersheds, parks, states, regional, or national jurisdictions. She recognized that the value of the NSDI was in helping to promote a culture of data sharing and that the NSDI was not a tangible 'thing' that would ever be constructed or finished, but more a continually evolving 'state of mind'.

Nancy has since given talks worldwide about the value of sharing geospatial information and conceptualizing a common and coordinated approach for doing so – continuing to give impetus to the NSDI movement.

In April 1994, Executive Order 12906: 'Co-ordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure' was signed by President Clinton. This directed that federal agencies, coordinated by FGDC, carry out specified tasks to implement the NSDI; it created an environment within which new partnerships were not only encouraged, but required. Also, it raised the national and international political visibility of geographic information collection, management, and use. Given the nature of the US constitution and the many parties involved from very different sectors, no one body was ever seen as being 'in charge'.

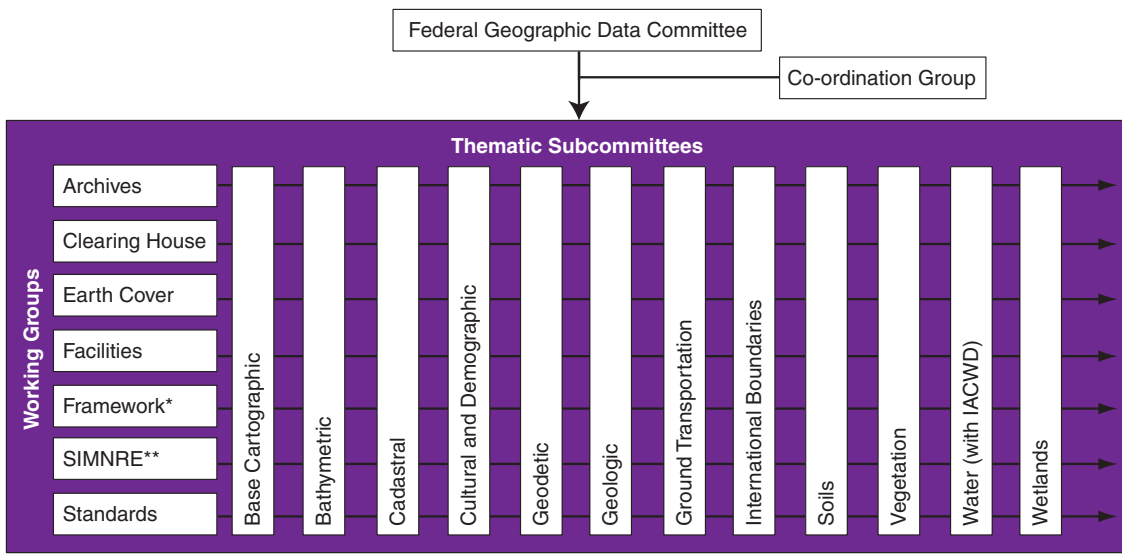
20.3.3 What has been the reality of the NSDI?

The Presidential Executive Order identified three primary areas to promote development of the NSDI: (i) the development of standards; (ii) improvement of access to, and sharing of, data by developing a National Geospatial Data Clearinghouse; and (iii) the development of the National Digital Geospatial Data Framework.

The Federal Geographic Data Committee, the stipulated coordinating mechanism, has operated as a bureaucracy through a series of subcommittees based on different themes of geospatial data (e.g., soils, transportation, cadastral), each chaired by a different federal agency. Figure 20.8 illustrates the original sub-committees and who in the federal government led each one.

Many of these groups developed standards for data collection and content, classifications, data presentation, and data management to facilitate data sharing. For example, a Standards Working Group developed the metadata standard, which was adopted formally by the FGDC in mid-1994 after an extensive public review process. It has since been adapted in the light of international and other national developments (see Section 11.2). The NSDI Executive Order mandated that all federal agencies must use all FGDC-adopted standards.

The second activity area was intended to facilitate access to data, minimize duplication, and assist partnerships for data production where common needs exist. This has been done both by advertising the availability of existing data and institutional willingness to create new data through the National Geospatial Data Clearinghouse. Agencies producing geographic information describe its existence with metadata and serve those metadata on the Internet, to be accessed by commonly used commercial Internet search and query tools (see Section 11.2). As a partial result of all this, nearly all federal agencies, as well as most states and numerous local jurisdictions, have become active users of the Internet for disseminating geographic information. This model does not assume



FGDC Subcommittees and Lead Departments

Base Cartographic - Interior
Bathymetric - Commerce
Cadastral - Interior
Cultural and Demographic - Commerce
Geodetic - Commerce
Geologic - Interior
Ground Transportation - Transportation
International Boundaries - State
Soils - Agriculture
Vegetation - Agriculture
Water - co-sponsored with the Interagency Advisory Committee on Water Data
Wetlands - Interior

Figure 20.8 The original structure of the Federal Geographic Data Committee sub-committees and lead responsibilities for them (Reproduced by permission of the FGDC)

necessarily that GI will be distributed for free. Obtaining some of the datasets requires the payment of a fee, while others are free.

The third NSDI activity area is the conceptualization and development of a US digital geospatial framework dataset. This has proved somewhat more difficult than the other two activities (see discussion of The US National Map in Box 19.7).

The NSDI is not a concrete 'thing' but is more of a vision, a state of mind, a campaign, and an enabler for better use of scarce resources.

20.3.4 The NSDI stakeholders

The stakeholders are now defined to include public sector bodies, within federal, state, and local government, and the private sector. In general terms, the US NSDI is steered by representatives of the following organizations: the National States Geographic Information Council, the National Association of Counties, the Open Geospatial Consortium, the University Consortium for Geographic Information Science, the National League of Cities, the FGDC, cooperating state councils, the International City/County Management Association (ICMA), and the Intertribal GIS Council. Engagement with the private sector was slow to occur. This is now channeled through the Open Geospatial Consortium, itself a partnership (see Section 20.5.2)

20.3.5 Has the US NSDI been a success?

Since the creation of America's NSDI, many more organizations – from different levels of government and occasionally from the private sector – have formed consortia in their geographic area to build and maintain digital geospatial datasets. Examples include various cities in the US where regional efforts have developed among major cities and surrounding jurisdictions (e.g., Dallas, Texas), between city and county governments (e.g., San Diego, California), and between state and federal agencies (e.g., in Utah). The characteristics of these partnerships vary depending on the level of technology development within the partner jurisdictions, on institutional relations, on the funding, and on the type of problems being addressed.

Inevitably, many problems have arisen in such an ambitious project involving huge numbers of people and organizations. Incentivizing different organizations to work together and produce benefits for all those organizations incurring costs is not a simple matter – especially when there is no agreed measure of success or even progress and 'no-one is in charge'. As indicated above, initially the inputs of the private sector were modest except in delivering metadata software – which they needed to do in any case to meet the needs of their government customers. The NSDI concept of 'bottom up' aggregation of data from many (perhaps thousands of) sources to form quality national datasets is intrinsically complex, even daunting. Also it is surprising that NSDI was not more concerned with education and training in its early stages – these have been key influences on the development and use of GIS and GI (see Section 1.5.5).

Typically, government initiatives wane after their authors have moved on. NSDI's profile and prestige diminished somewhat around 2000 when it was no longer new. Internal dissension within the federal government about who led on what, some shortcomings in agreements between federal, state, and local government and with the private sector on how to progress, budget constraints, a host of other new initiatives demanding staff and resources, all contributed to a slowdown of progress of the NSDI in its heartland. A new 'second phase' was triggered, however, by the US e-government initiative launched in 2002, an initiative designed to improve the value for money of government. Driven by the Office of Management and Budget, it led directly to the Geospatial One-Stop geoportal (see Box 11.2). As part of this, the NSDI has been re-examined in Congress by the Committee on Government Reform (Box 20.6). The greatly increased focus on Homeland Security after 9/11 is also forcing a re-think of some aspects of NSDI (see Section 20.6).

The impetus for greater government efficiency and effectiveness and the advent of Homeland Security have re-ignited the cause of NSDI.

Confusion about what all the different initiatives were and how they related to each other began to become common in 2003/04. A paper published by senior USGS and FGDC staff in April 2004 sought to define the relationships: according to this, the US National Map program (Box 19.7) enables the provision of content via integration of data from different partners, the FGDC (Section 20.3.2) provides tested standards, and Geospatial OneStop (see Box 11.2) provides the access tool to a range of federal and some state GI. The FGDC also produced a thoughtful paper in mid-2004 entitled 'Towards a National Geospatial Strategy and Implementation Plan'. This argued for 'partnerships with purpose' between commerce, academia, and all levels of government, led by FGDC; the standardization of geographic framework themes; and a need to communicate better the message about NSDI through a business case, a strategic communications plan, and training programs. All these were accompanied by proposed targets to be achieved by different dates in the period up to 2007.

Highly partisan views exist on what NSDI has achieved and what should be done next. This is not surprising given its ambitious scope, its nature, and the lack of simple measures of success.

Shortly afterwards, however, the USGS Director reorganized all of the components described above to be within a new National Geospatial Programs Office within his own organization, under the direction of the former Geographic Information Officer (now an Associate Director for Geospatial Information and Chief Information

Applications Box **20.6**

The Putnam NSDI hearings in Congress

At the time of writing, Congressman Putnam's Sub-Committee on technology, information policy, intergovernmental relations, and the Census (which neatly covers many of the concerns of this book) had held two hearings with witnesses. The first (in June 2003) is summarized at **reform.house.gov/TIPRC/Hearings/EventSingle.aspx?EventID=510.** The second, a year later, was entitled 'Geospatial Information: Are we headed in the right direction or are we lost?' and is summarized at **reform.house.gov/TIPRC/Hearings/Event-Single.aspx?EventID=1150.**

Essentially, the first hearing was setting the scene. In it, a senior official of the Office for Management and Budget (OMB) conjectured that up to half of all spending on geospatial information across all levels of government was duplicative – a contention later disputed. Amongst the conflicting views expressed at the second meeting (by representatives from the bodies in parentheses) were:

- A complete and up-to-date strategic plan is missing, federal agencies are not complying with General Accounting Office (GAO) direction on coordination and OMB's oversight methods to prevent duplication have proved ineffective (GAO).
- Considerable progress has been made but more needs to be done (OMB).
- The Geospatial OneStop (GOS) portal in mid-2004 held information from 155 federal and state sources and received visits from 6600 different individuals per month. It will continue to be developed (USGS).
- The National Geospatial-Intelligence Agency of the Department of Defense works to coordinate Defense involvement in GOS and spends about $0.5 m per annum directly on it. Because of its statutes and regulation, NGA cannot provide geospatial intelligence support to state, county, tribal, and other political administrations (NGA).
- Data sharing between local governments, states, and federal government has to meet the most exacting requirements – for local detail – and hence is more expensive than data created for federal agencies alone but is more up-to-date. . . A single model does not fit all states in regard to NSDI. . . The US needs a single federal agency, with cross-cutting authority, that can direct and speak for all federal agencies on geospatial development and coordination. Legislation constrains collaboration, e.g., the Federal Advisory Committee Act prevents federal agencies from putting local, state, or tribal organizations on their advisory bodies (National States Geographic Information Council).
- Federal government geospatial programs would benefit from the private sector being a full partner in developing plans and policies. . . A well-funded business plan is required for the NSDI (Spatial Technologies Industry Association).
- We are not proceeding with a good road map. Numerous studies detail the lack of coordination of federal mapping and geospatial activities and government's duplication with the private sector (Management Association for Private Surveyors).
- We are basically on the right track but some mid-course corrections are needed. . . policy makers have overlooked the importance of the Open Geospatial Consortium (OGC)'s interoperability standards effort . . . the FGDC and other federal agencies need to continue to participate in the OGC to ensure that unfinished standards reflect the needs of the public and the requirements of the government agencies entrusted to serve the public interest (OGC).
- The FGDC has done a good job, especially in metadata standards. . . [but] voluntary partnerships are not working and the federal government must find new carrots and sticks to realize the potential of the NSDI (Mapping Science Committee).

It is obvious that partisan, and strongly held, views were expressed. In so far as this is representative of wider views amongst the stakeholders, little consistency of view exists – other than agreement on the fact that an NSDI is (actually or potentially) a good thing and the federal government has an important role (though there is no clear consensus on what the latter should be). It is obvious that the NSDI is controlled by statutory, political, legal, commercial/financial, and other factors as well as technical ones.

Officer). The clear aim was to reduce any confusion and foster inter-working, at least within key components of the most central federal government agency (USGS) so far as GI is concerned.

20.3.6 Conclusions on the progress of the US NSDI

It is no surprise that NSDI matters have not gone smoothly. Anything so broadly defined and forming such an intangible asset (Box 18.6), whose creation involves huge numbers of different organizations and individuals – each with their own views, values, and objectives – and having 'no-one in charge' was always going to be difficult to achieve. At a high level, there are few who dispute the merits of achieving data sharing, reduction of duplication, and risk minimization through the better use of good-quality GI. But how to make it happen for real is a different matter.

Something as diffuse as NSDI will never be seen as a success by everyone, but it has been a catalyst for many positive developments.

Notwithstanding all these complications, in its short lifetime NSDI has generated high levels of interest in the USA and beyond. Some real successes have been achieved, as indicated above. Perhaps its greatest success, however, has been as a catalyst, acting as a policy focus, publicizing the importance of geographic information, and focusing attention on the benefits of collaboration. This is especially important in a country as large and governmentally complex as the USA.

20.4 Multi-national collaborations

There are many multi-national partnership arrangements in GIS, especially trans-border and bi-lateral ones. Here we concentrate on two such examples – the GIS and GI developments in the European Union and in Southeast Asia. The similarities and the differences between them and with the USA are obvious.

20.4.1 The European dimension

After the accession of 10 new member countries in 2004 – taking the total to 25 countries – the European Union (EU) contains some 456 million people. The European Commission (EC), the public service of the EU, has been considering a variety of GIS and GI issues since the 1980s. Progress has been slow, complicated by the decentralized, multi-cultural, and multi-lingual nature of decision making amongst the countries involved and the relatively low priority accorded to GI matters in EC work. A serious attempt to define a GI policy framework for Europe through the GI2000 initiative failed. There have been some achievements in GIS research terms via the coordination and funding provided by the EC's Joint Research Centre in Ispra, Italy, including work by EUROGI and AGILE (see Box 20.3). More productive thus far has been the incorporation of GI within broader programs. Examples of this include the eEurope and eContent initiatives to foster European industry and the use of new technologies and content by citizens. The EC's ideas on public sector information, which highlighted the need for much easier access to information by citizens, may have long-running consequences since much information of this type is geographic. Once initiatives to take forward policies are agreed and financed by the European Union, they then acquire considerable momentum: an EC Directive approved by the European Parliament and the Council of Ministers must be implemented in national law within a defined period (normally two years).

A major initiative of the EC is the INSPIRE program (**www.ec-gis.org/inspire/**). This is a Directive proposed formally by the European Commission to establish a legal framework for the establishment and operation of an Infrastructure for Spatial Information in Europe. Based on much work by the Directorate of the Environment, Eurostat (Europe's statistical agency), and the Joint Research Centre at Ispra, this is designed to meet both community and local needs. The value of what is proposed is anticipated to be greatest to policy-makers such as legislators and public authorities but citizens, academics, and business are also expected to benefit if INSPIRE is adopted. The legitimation for INSPIRE is similar to that of the US NSDI (see Section 20.3.1) – even if the process for taking it forward is very different. The proposal was produced after making impact and risk assessments, having a consultation period and a public hearing, and taking account of the EU principles of subsidiarity (i.e., that nothing should be done at Community level which can be done more efficiently at national or lower level).

The proposed directive covers spatial datasets held by public authorities or their agents and spatial data services which employ these datasets. It mandates the use of a Community geoportal (Section 11.2) by all member states to supply their data to all other member states plus national and Europe-wide coordination structures and the creation of metadata. Procedures are laid down for establishing harmonized data specifications, so as to facilitate safe added-value operations, including unique identifiers for spatial objects across the continent. Most specifically, member states of the EU will be mandated (after acceptance of the Directive) to provide certain classes of services. These are discovery services permitting search for spatial datasets and related services, viewing services to display and navigate data, download services, data transformation services, and the capacity for invoking spatial data services. The first two of these must be provided free of charge though public access to the others may be curtailed for a number of reasons (e.g., national security and statistical or commercial confidentiality). One public authority in each country is responsible for liaising with the EC on the directive.

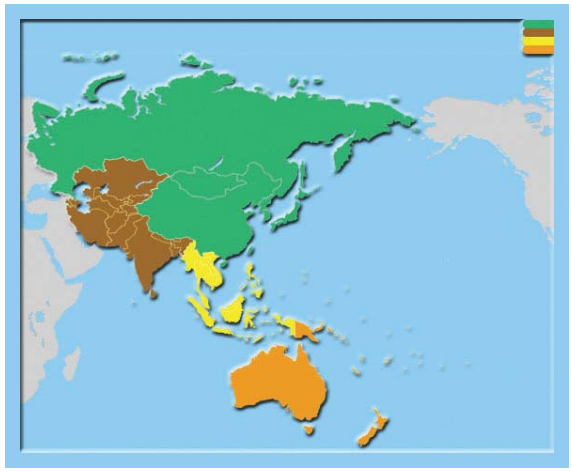


Figure 20.9 The member countries of the Permanent Committee on GIS Infrastructure for Asia and the Pacific (*Source*: PCGIAP)

INSPIRE was created following many political, scientific, and administrative battles over previous attempts to create an NSDI for an enterprise almost with twice the size of population of the USA and one in which Presidential Executive Orders (see Section 20.3.2) do not exist. The progress of the directive through the approval processes will not be easy but could make a material difference to how European GIS and GI operate. In microcosm, the attempts to build US and European NSDIs illustrate how differently the USA and the European Union operate and the importance of political leadership.

20.4.2 The Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP)

This is the largest multi-national GIS body known at present. It was set up under UN auspices in 1994 and involves 55 countries (Figure 20.9). Its aims include maximizing the economic, social, and environmental benefits of geographic information in accordance with Agenda 21 (Section 20.5.1.3) by providing a forum for nations from Asia and the Pacific to cooperate in the development of a regional geographic information infrastructure and share experiences. 'GIS infrastructure' is taken as including the institutional framework (which defines the policy, legislative, and administrative arrangement for building, maintaining, accessing, and applying standards and fundamental datasets), technical standards, fundamental datasets, and a framework which enables users to access them. For the members of PCGIAP, key applications of GIS are national or regional land administration, land rights and tenure, resource management and conservation, and economic development. Considerable GIS activity is underway across the member nations, including the development of numerous NSDIs, and this information is shared through numerous meetings and newsletters (**www.gsi.go.jp/PCGIAP/**).

Continent-wide NSDIs are much more difficult to launch in Europe than in the USA because of the power structures.

20.5 Nationalism, globalization, politics, and GIS

We have shown above how the USA and groupings of other nation states have attempted to enhance the quality, utility, accessibility, and awareness of geographic information – and its use. This has been carried out through institutional structures and 'umbrella bodies', holding together disparate partners. But much GI has hitherto been created to suit national needs (at best). Its wider use is constrained by its historical legacy and the need for continuity through time for comparative purposes.

Other than that collected via satellite remote sensing, truly global coverage GI is currently little more (and sometimes less) than the sum of the national parts and is not always readily available. Yet the need for globally or regionally available, easily accessible, and consistent GI – especially the 'geographic framework' variety (see Section 19.3) – is increasingly evident. Even though few people currently have management responsibilities which are global, the present shortcomings in global GI have severe consequences for many actual and potential users who operate beyond national frameworks. The United Nations Regional Director of Development for the Pacific argued forcefully for the need for greater partnering and coordination of data collection and supply on a regional and global basis. He emphasized that development of such datasets is labor-intensive and costly and, although these datasets could support a wide variety of applications, generally no single use can justify the full cost of development.

Other than for imagery, global GI is currently little more than the sum of the highly varied national parts and detailed information is not readily available.

20.5.1 Global databases

There are two obvious approaches to creating global databases: to use existing mapping or to create something wholly new from imagery or other sources (though typically the products are often later used in combination). The latter has developed hugely in recent years through imagery produced by NASA, ESA, and similar bodies (e.g., see Figure 20.10), and commercial organizations.

20.5.1.1 Databases and maps

Global mapping has progressed rather more slowly than satellite imagery as a data source. And there have been false steps along the way: the International Map of

Figure 20.10 One section of a 'night time map' of the world produced as a composite of hundreds of images made by the Defense Meteorological Satellite Program which can detect low levels of visible/near-infrared radiance at night including lights from cities, towns, industrial sites, gas flares, and ephemeral events such as fires. Clouds have been removed. It was compiled from the October 1994–March 1995 data collected when moonlight was low (Courtesy NASA and US National Oceanic and Atmospheric Administration (NOAA))

the World was assembled over 60 years to provide a common geographic framework worldwide. Some 750 map sheets were published eventually but IMW is now simply history. The primary reasons for its ultimate failure were organizational and political. There was a lack of commitment of finance by those who agreed to participate – national mapping agencies – and conflicts in priority with national objectives. There was a lack of clearly articulated needs which the IMW was designed to meet, and few demonstrations of success from its use, and a lack of clear management responsibility for overall success. Finally, there was much duplication of work because of the technologies then available: much of the same mapping had to be created separately on both national and global map projections (see Section 5.7).

In the 1990s, in particular, a number of developments occurred in global topographic mapping – the traditional form of a geographic framework and still crucial. The US and allied military organizations (see below) created and distributed at low cost the Digital Chart of the World (DCW), subsequently known as VMap Level 0. Derived from the paper Operational Navigation Chart (ONC) 1:1 million scale map of the world created by NATO partners, this has been widely used in a variety of applications – including in low-price hand-held GPS receivers. Its popularity is despite well-known shortcomings in the quality of the original cartographic source material. This project was brought to reality because one organization led it strongly, there was a clear operational need, and financing of it could be justified on that basis. The release of DCW in 1993 led to robust discussions between a number of national mapping organizations (NMOs) and the then US Defense Mapping Agency (now part of the US National Geospatial-Intelligence Agency or NGA) about infringements of national Intellectual Property Rights (see Section 19.1). Since then, an agreement has been forged for subsequent VMap products. This ensures that, where NMO-sourced material is used in NGA products, the providing nation has the option of restricting availability to NATO military use and making other arrangements (if desired) for commercial and public dissemination of its material.

Finally, there have been many other plans to build global databases which include mapping – all based on some form of partnership. The most advanced map-based project is that originally proposed by the Japanese NMO, initially for a new 1:1 million scale global digital map. An International Steering Committee for Global Mapping (ISCGM) has produced a full specification. The committee's initial proposals involve using a combination of VMap level 0, 30 arc second (i.e., approximately 1 km resolution) digital elevation model data and the Global Land Characteristics Database. Thus ISCGM proposed using vector and raster data already in the public domain, though not hitherto linked together, to provide the default dataset. The plan was that this would be replaced on a country-by-country basis wherever NMOs agreed to provide higher resolution data – though inevitably the final result would be somewhat inconsistent, especially at national boundaries in areas of territorial disputes. Recognizing such inherent difficulties, the ISCGM's specification ensures that – where any difference occurs – both sets of international boundaries will be stored. Encouragement to include data deriving from more detailed 1:250 000 scale mapping has now occurred. Figure 20.11 shows the countries that had committed to an involvement in this project by 2004. Although a sizeable number of these countries is not yet able or willing to supply

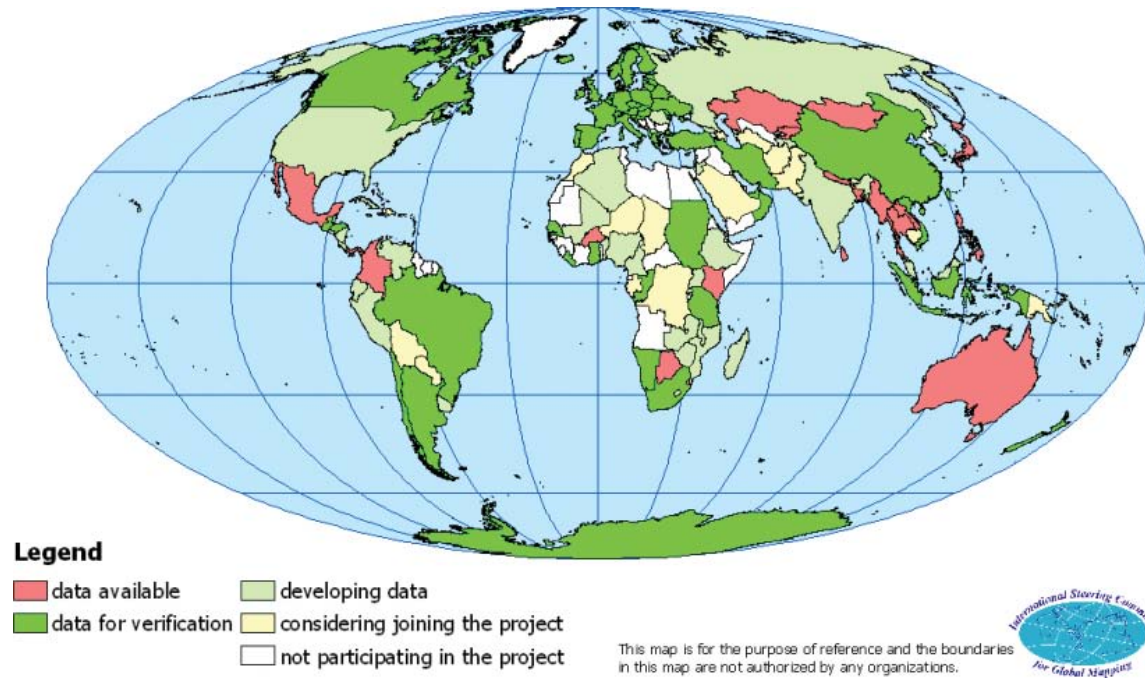


Figure 20.11 The countries that had committed themselves to the Global Mapping Project by August 2004

data in practice, materially this project has improved the availability of strategic-level GI for some areas of the world.

In some instances, certain aspects of maps have been abstracted and are being enhanced and up-dated by GPS and other means. The prime example is roads databases. Here the key players have almost all been private-sector bodies – even though many obtained raw data from national mapping organizations (NMOs). Navigation Technologies (NAVTEQ), a company operating in North America and Europe, has been building databases for well over a decade for the intelligent transportation industry and for in-vehicle navigation. These are built to one worldwide specification. Every road segment in the NAVTEQ database can have up to 150 attributes (pieces of information) attached to it, including information such as street names, address ranges, and turn restrictions (see Section 8.2.3.3). In addition, the database contains Points of Interest information in more than 40 categories. The information has multiple sources and several NMOs have relationships with Navigation Technologies. For example, the Institut Géographique National in France bought shares in NAVTEQ. A competitor of NAVTEQ is Tele Atlas which has coverage in Singapore, Hong Kong, and large parts of Australia as well as North America and Europe; the industrial scale of the applications of such data can be seen from the client lists. For Tele Atlas this includes Blaupunkt, BMW, Daimler Chrysler, Ericsson, ESRI, Michelin, Microsoft, and Siemens. All of these firms have business partners who may include surveying firms, national mapping organizations, value-added resellers, and much else.

Public and private sector partnerships are creating new GI which, in time, will probably be global in coverage.

20.5.1.2 Databases from satellite imaging

The variety of global databases now available from satellite-based sensing is extraordinary. One of the most relevant for many GIS users is that arising from NASA's $220 million Shuttle Radar Topography Mission launched in February 2000. The mission's aim was to obtain the most complete 'high resolution' (by the standards of what previously existed) digital topographic database of the Earth. The primary intended use of this data is for scientific research. It used C-band and X-band interferometric synthetic aperture radars to acquire topographic data between latitudes 60°N and 56°S during the 11-day mission. The resulting grid of heights is available on CD. The data released into the public domain are 3 arc seconds resolution (i.e., about 90 meters at the Equator) for the land areas of the Earth; however, 1 arc second data are available for the continental USA.

The first high-resolution civilian satellite was launched by Space Imaging Inc. in September 1999. This produced various imagery sets, the most detailed being 1 meter resolution. Space Imaging also collects and distributes Earth imagery from a range of partner organizations in Canada, Europe, India, and the USA. In essence, it is setting out to be a global geo-information, products, and services organization, working with a variety of regional partners across the world. Other competitors such as DigitalGlobe have joined Space Imaging and data resolutions as detailed as 60 cm are now available from these commercial providers for many areas of the world

(see Figure 20.12), with 42 cm resolution planned by 2006/07. A variety of public and private sector channels exist for obtaining access to these data (see, for instance, **www.terraserver.com**)

20.5.1.3 Why do we not yet have good global GI?

What is evident from the last 20 years or more of experience is that everyone thinks that digital databases of the world are 'a good thing', especially if they are produced to a standard specification, sufficiently detailed, up-to-date, and readily available.

There are, for instance, many examples where GI and GIS have been identified by politicians as central to improving some situation. Perhaps the most striking example is that of Agenda 21. This agreement was designed to facilitate sustainable development. The manifesto was agreed in Rio in 1992 at the 'Earth Summit' and was supported by some 175 of the governments of the world. It is noteworthy that eight chapters of the Agenda 21 plan dealt (in part) with the need to provide geographic information. In particular, Chapter 40 aimed at decreasing the gap in availability, quality, standardization, and accessibility of data between nations. This was reinforced by the Special Session of the United Nations General Assembly on the Implementation of Agenda 21 held in June 1997. The report of this session includes specific mention of the need for global mapping, stressing the importance of public access to information and international cooperation in making it available. The subsequent UN World Summit on Sustainable Development in 2002 included many GIS activities and demonstrations to the 40 000 participants, including leading politicians. Again there were public calls for spatial data to be regarded as a public good (see Section 18.4.2), accessible to potential users in a common format for the common good.

Unsurprisingly, therefore, many partnerships have been set up to plan or produce such global databases. In practice, however, commerce is only interested if the real prospect of profit occurs; usually this necessitates military expenditure. Government leadership of such schemes tends to falter with change of key personnel (e.g., see Box 20.7).

Past failures to deliver promises or meet aspirations, however, do not undermine the large potential value of such databases; and political support is essential (if often ephemeral) for investment if that can only come from governments. Rarely does this political support come from a simple interest in technical matters or even a fascination with geography on the part of the politicians. To obtain support requires both fitting the GIS agenda to the existing political agendas and highly professional raising of awareness and lobbying.

20.5.2 Global partnerships for standards

There are already a number of well-resourced and long-term groupings which bring together countries for particular purposes and which have direct or indirect impacts on GIS – such as the International Standards Organization (ISO), the World Intellectual Property Organization

Figure 20.12 Part of a Quickbird 62 cm resolution image of Buckingham Palace and the Mall, London (© Quickbird)

Applications Box **20.7**

The fading of Digital Earth

'Digital Earth' was first proposed by US Vice-President Gore in his 1992 book *Earth in the Balance*. Re-launching the concept in 1998, he said a pressing problem was how to turn the huge amounts of geospatial data that were being collected into information that people could use. He stated that:

- there was no effective connection (interoperability) between the many datasets required to accomplish better management of the Earth's resources and minimization of the problems of natural disasters;
- there was no common vision to activate cooperation among the many elements of the technology base and the national and international community;
- there was low awareness and little accessibility by the 'common person' to the vast information resources about our planet and its people.

Gore's analysis was a damning indictment of what had been achieved to that date. His solution was a mechanism for tying together data from multiple, networked sources and making that integrated resource available visually as a 'browsable', 3-D version of the planet via an enhanced Internet. In the longer term, he envisaged having a complete and consistent digital map of the Earth at 1 meter resolution.

Following Gore's defeat in the 2000 US Presidential elections, Digital Earth rapidly faded from view as a project. However, some Chinese activity in this sphere continues, many relevant (but re-titled) developments continue in NASA, and commercial developments like *ArcGlobe* and *Earthviewer* (**www.earthviewer.com**) are practical demonstrations of some of Gore's aspirations.

(WIPO), and the World Trade Organization (WTO). These are voluntary groupings bound through signature of treaties or through the workings of a consensus process. But other groupings of bodies exist which have specific importance for GIS. Perhaps the most important of these is the Open Geospatial Consortium (OGC), formerly the Open GIS Consortium (see also Section 11.2).

OGC is a voluntary consensus standards organization operating in the GIS arena. The OGC is a not-for-profit, global industry association founded in 1994 to address geospatial information-sharing challenges. Its membership is primarily US-based, reflecting its origins and also the US dominance of the GIS marketplace, but the geographic spread of membership has continually increased. OGC's worldwide membership, which totals 260 entities, includes geospatial software vendors, government integrators, information technology platform providers, US federal agencies, agencies of other national and local governments, and universities. The OGC's President has claimed that the network of public/private partnerships embodied by the organization has accomplished for geospatial information what the US railroad companies had accomplished by 1886, when they achieved consensus on the adoption of a common rail gauge.

20.5.3 The Global Spatial Data Infrastructure

The other approach to building worldwide consistent facilities has been to seek to create a Global Spatial Data Infrastructure. GSDI is the result of a voluntary coming together of national mapping agencies and a variety of other individuals (see Box 20.3) and bodies. The Steering Committee defined the GSDI as 'the broad policy, organizational, technical, and financial arrangements necessary to support global access to geographic information'. Initially, the organization sought to define and facilitate creation of a GSDI by learning from national experiences. An early scoping study showed that the reality of the benefits was difficult to prove, largely because it is very difficult to define the product that governments and other agencies might be asked to fund. Since then, perhaps wisely, the GSDI Association (**www.gsdi.org/**) has concentrated on

- serving as a point of contact and effective voice for those in the global community involved in developing, implementing, and advancing spatial data infrastructure concepts;
- fostering spatial data infrastructures that support sustainable social, economic, and environmental systems integrated from local to global scales; and
- promoting the informed and responsible use of geographic information and spatial technologies for the benefit of society.

All this has been done through a series of partnerships with many other bodies (such as OGC and regional GIS organizations). Certainly, the result has broadened the knowledge of GIS use and GI challenges. In addition, such GSDI publications as the GSDI 'cookbook' on how to create a SDI have helped many countries (**www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf**).

Building a Global Spatial Data Infrastructure would be even more difficult than a national one, but discussions on GSDI have proved especially helpful to those outside the wealthier countries.

20.6 Extreme events can change everything

Much of what has been written above has arisen from a situation where it is essentially 'business as usual'. But the events of 11 September 2001, seared into the memory of millions of people across the globe (Figure 20.13), have also begun to change many aspects of GIS and GI.

The death and disruption caused by the Al Queda hijacking of four aircraft and their use as guided missiles ruined lives and caused billions of dollars of harm to the US and other Western economies. The immediate aftermath was marked by heroic efforts, not least amongst the GIS teams who recovered information with which to support the rescue and reconstruction efforts in New York (Box 1.1). But the event itself also led to an intense questioning of how we can best protect ourselves from further acts of terrorism – without turning ourselves into a police state. One manifestation of this debate was the creation of the US Department of Homeland Security (see Box 20.8) but the reactions were not restricted to the USA: countries as far apart as India, Russia, and Colombia have also taken action as they too have been

Figure 20.13 Newspapers from around the world on 12 September 2001 reporting the terrorist attacks on the World Trade Center. (*Source*: **www.september11news.com/index.html**. Copyright in the images subsists in the newspapers shown; *Herald Sun* page reproduced by permission of Herald & Weekly Times Pty Ltd; *Dagens Naeringsliv* page reproduced by permission of Dagens Naeringsliv)

Applications Box **20.8**

The creation of the US Department of Homeland Security

The Office of Homeland Security was set up by the US President on 8 October 2001 and the Department of the same name was formed by an Act passed by the Congress some three months later. The Department of Homeland Security (DHS) has three primary missions: to prevent terrorist attacks within the United States, reduce America's vulnerability to terrorism, and minimize the damage from potential attacks and natural disasters. It was formed by merging all or parts of 22 pre-existing agencies under new integrated management. It was also set up from the outset to have a Directorate of Information Analysis and Infrastructure Protection. The Secretary of the department has sweeping powers of access to virtually any information which he or she may consider relevant to the department's tasks, whether held by the federal government or elsewhere (see **www.dhs.gov**).

affected by terrorist action. All this highlights the need for flows of intelligence across institutional and national boundaries – requiring new forms of trust and partnership which must operate speedily and without failure.

Since everything happens somewhere and since targets of terrorism are picked to cause maximum effect, GI and systems to deal with it are at the forefront of governmental efforts to carry out their first duty – to protect their citizens. To aid this mission, we must first know how GIS can potentially help terrorists so we can guard against such threats. We must also know how best to use such systems to minimize the consequences of any terrorist actions. And we need to work out the best form of partnerships in these new circumstances.

20.6.1 How GIS can potentially help attackers

In principle, GIS can help a potential terrorist in at least four ways:

- finding locations of a particular type (e.g., major 'choke points' in roads, population centers, or highly critical places such as nuclear plants);
- adding data layers to enrich the understanding of that location's 'uniqueness' or 'connectedness';
- modeling of the likely effects of a disruption (in three dimensions and time, if desired) to the facility and of escape routes for the attackers;
- providing descriptive material or GPS coordinates to lead the attackers to the spot and to exit after the attack.

20.6.2 Is public access to GI adding to the problem?

The nature of Western society is that much information is made readily available: that is a fundamental tenet of democracy. The situation is perhaps most extreme in the USA where the huge bulk of federal government information is made available without constraints on who can get it or on its further dissemination (and commercial exploitation – see Section 19.3). Table 20.2 summarizes the critical physical infrastructure of the USA as recognized in the national strategy for its protection.

Immediately after 9/11, a number of US government websites withdrew access to various datasets (see Box 20.9). Much was restored after quick reviews. The National Geospatial-Intelligence Agency, with US Geological Survey support, sponsored a major study by the Rand Corporation to assess the homeland security implications of publicly available geospatial information. This reviewed the types of data which might be helpful to terrorists in different ways – the 'demand side' – and studied a large sample of publicly available websites – the 'supply side' (see Table 20.2). Their conclusions can be summarized as follows:

- Publicly accessible geospatial information has the potential to be useful to terrorists for selecting a target and determining its location but is probably not the first choice to fulfill their needs – there are many alternative sources.
- Less than 6% of the publicly available GI from federal government agencies appeared as if it would be useful to potential attackers and none appeared of critical importance to them; fewer than 1% appeared both useful and unique.
- Much of the information needs of potential attackers can be met from non-governmental websites – such as industry and business, academia, NGOs, state and local governments, international sources, and even private individuals.
- The societal costs of restricting access to geospatial information are difficult to quantify but are large.
- The federal government should develop a comprehensive model for addressing the security of geospatial information, focusing on what level of protection should be afforded to data about a particular location. In addition, it should raise public awareness of the potential sensitivity of GI.

Table 20.2 US critical infrastructure sectors and accessibility to information about individual instances. About 85% of the US critical infrastructure is owned by the private sector

Sector	Example assets	Degree of public accessibility
Agriculture	Grain storage elevators	Medium to substantial
Food	Meat processing plants	Medium to substantial
Water	Drinking water facilities	Substantial
	Dams	Limited to substantial
Public health	Hospitals	Substantial
	National pharmaceutical stockpiles and supplies	Limited
Emergency services	Emergency operations centers	Medium to substantial
Government	Government agency headquarters	Limited to medium
	Regional offices	Medium to substantial
Defense industrial base	Military equipment manufacturing plants	Medium to substantial
Information and telecommunications	Transmission sites	Medium
	Internet backbone facilities	Medium to substantial
Energy	Nuclear power plants	Limited to medium
	Oil refineries	Limited to medium
Transportation	Bridges	Substantial
	Tunnels	Substantial
	Pipelines	Substantial
Banking and finance	Major financial exchanges	Substantial
	Financial utilities	Medium
Chemical industry and hazardous materials	Chemical processing plants	Limited to medium
	Hazmat material transportation	Medium to substantial
Postal and shipping	Mail processing centers	Medium to substantial

Source: Table D.1 in Baker J., Lachman B., Frelinger D., O'Connell K., Hou A., Tseng M., Orletsky D. and Yost C. 2004 *Mapping the Risks: Assessing the Homeland Security Implications of Publicly Available Geospatial Information*. Rand Corporation Monograph (see **www.rand.org/publications/MG/MG142**)

Applications Box **20.9**

The withdrawal of public geospatial information after 9/11

Before September 11, 2001, most federal groups viewed the Internet as a place to store their vast library of public documents and as a way to network with community leaders in a timelier manner. But in the days that followed 9/11, authorities realized that terrorists also used the Web. Government agencies suddenly scrambled to assess what they released into cyberspace, vetting it for any sign that it could be used to exploit structural or security vulnerabilities.

Beth Hayden, spokeswoman for the Nuclear Regulatory Commission, one of more than a dozen federal agencies that altered their sites. . . said the agency's entire site initially was stripped from the Internet on October 11 but was restored October 17. The Environmental Protection Agency acted within 48 hours of the 9/11 attacks, opting to remove risk-management plans that offer specific outlines of the nation's chemical plants, including how they're handled and scenarios for possible releases. The plans had been posted on the Web so state or city officials and the public could view them in the event of an emergency.

Source: **CNN.com**, 10 September 2002

One state took its clearinghouse offline for two months while it removed 'tiles' that contained sensitive information. There were 500 sensitive sites on 170 quads and 1,600 orthophoto tiles. . . As of last November all the quads were back online. Ortho tiles are in the process of being 'blurred'. The state called upon a local university to develop an algorithm to blur just the footprints of the sensitive areas, again, attempting not to draw attention to them. All of the data was [sic] widely available elsewhere, but both states felt it politically correct to alter at least some of it.

Source: GISmonitor 1 April 2004

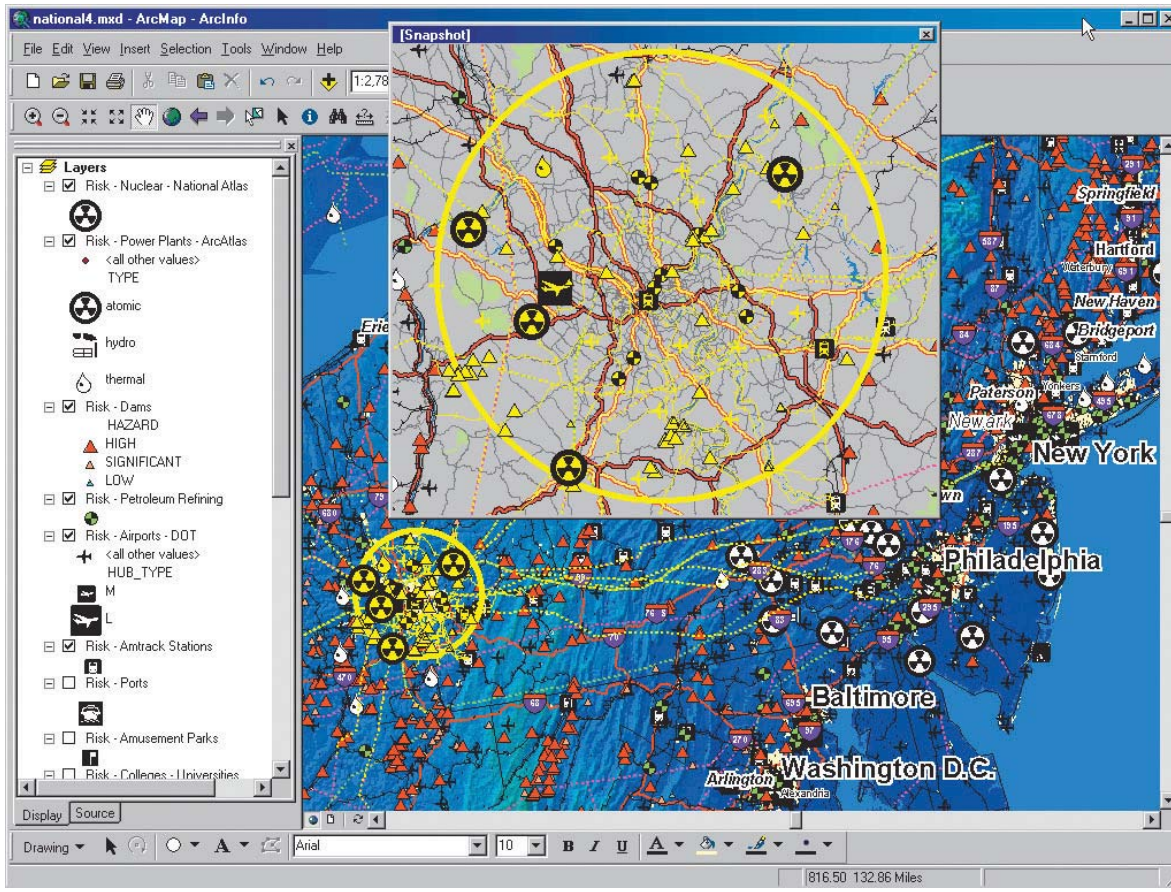


Figure 20.14 Key facilities in the North East USA, as assessed in the joint Homeland Security project of USGS and the Department of Homeland Security (Courtesy US Geological Survey)

The study's overall conclusion was that, whilst it might be useful, geographic information was unlikely to be of *critical* value to terrorists and is, in any case, so widely available that nothing much could be done to minimize the risk except in a few cases. This is comforting – but seems somewhat surprising given the final recommendation above.

20.6.3 Coping with terrorism and the GIS contribution

We can think of dealing with terrorism or some major national disasters as having five phases, each of which we review below. In each case GIS and GI can play a valuable role and partnerships between (often many) players are required.

20.6.3.1 Risk assessment

In principle, this is identical to all good business planning (see Section 19.4.1). It involves assessing the likelihood and possible impacts (on life, property, and other assets and the environment) of terrorist activity, an emergency, or disaster – and then communicating this to the appropriate authorities. It involves identifying risks and their potential impacts (Figure 20.14), which organizations should be involved, and the necessary mitigation, response, and recovery procedures – and testing the procedures.

20.6.3.2 Preparedness

This covers those activities necessary because terrorist attack or other hazards cannot be confidently anticipated and obviated. It involves creating a set of operational plans to deal with a set of defined scenarios, including who is to be in charge (e.g., in different zones) for the different scenarios. Inventories of governmental and other resources which may be mobilized need to be created and shared on a 'need to know' basis. Early warning systems need to be installed. Training exercises will be carried out. Emergency supplies of food and medical supplies need to be stockpiled and their delivery planned. All of these involve use of geographic information and GIS (see Figures 20.15 and 20.16).

20.6.3.3 Mitigation

This involves identifying activities that will reduce or obviate the probability and/or impact of an event, e.g.,

Figure 20.15 A scene from a training exercise simulating a chemical attack on the financial district of London (Reproduced by permission of PA photos)

maintaining an inventory of the contents and locations of hazardous materials and mapping the geography and sensitivity of environmental dissemination of contaminants (e.g., reservoirs and water distribution facilities – see Figure 20.14).

20.6.3.4 Response

These are the activities immediately following a terrorist attack, or other emergency, and are designed to provide emergency assistance for victims, stabilize the situation, reduce the probability of secondary damage, and speed up recovery activities. They include providing search and rescue, emergency shelter, medical care, and mass feeding; controlling transport links so as to prevent further injury, looting, or other problems (see Figure 20.17); shutting off contaminated utilities; and/or carrying out a damage assessment (see Figure 20.18).

Organizing evacuations may be essential (Sections 2.3.4.2 and 16.1.2) – as in areas affected by hurricanes (e.g., see Figure 2.14) – but it is also planned and periodically tested in most major cities as a prudent safeguard. In London, for instance, plans for evacuation of areas containing several hundred thousand people exist and tests of readiness against various attacks have been made; the transport system GIS plays a significant role.

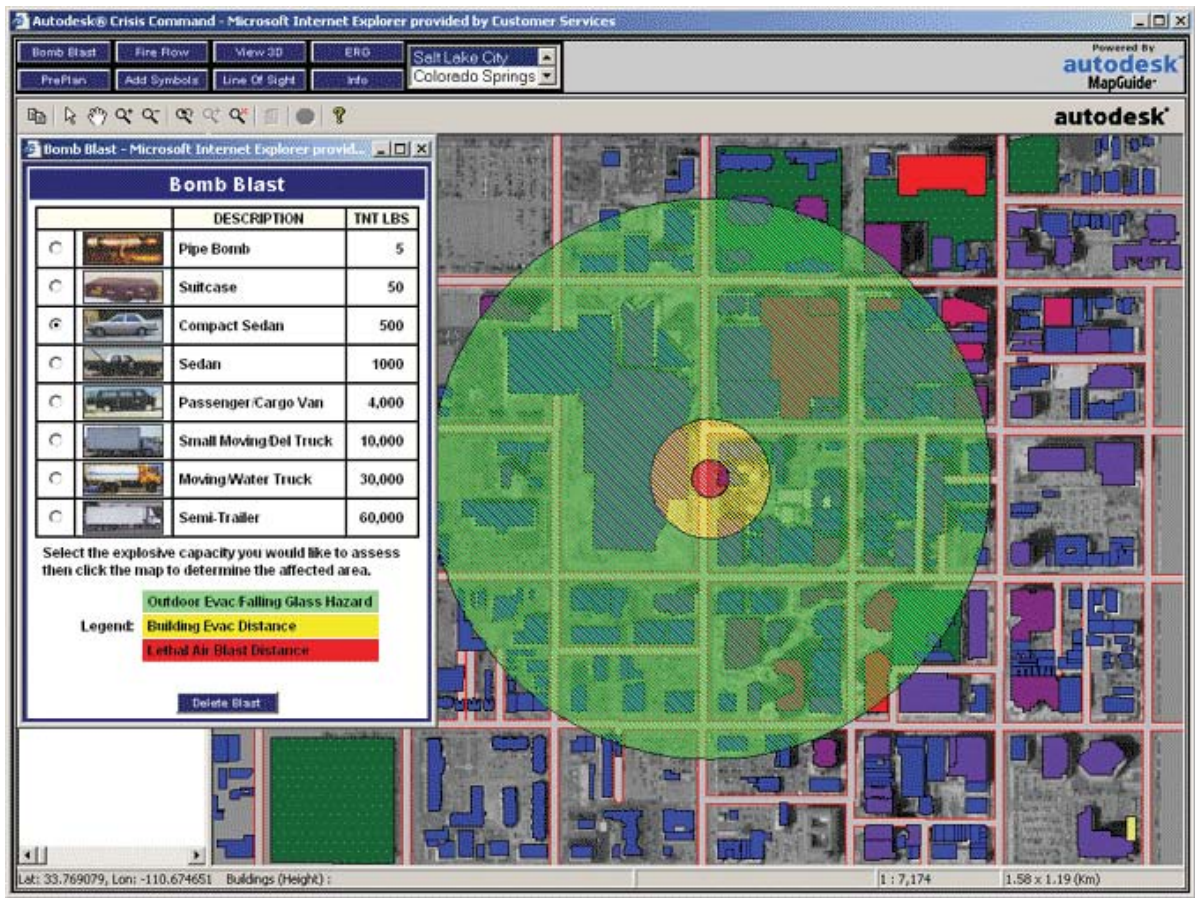


Figure 20.16 The potential impact of explosions within Colorado Springs, based on the size of an explosion (that selected is a simulation of a sedan packed with 500 pounds of TNT). The red area is the lethal air blast distance, yellow and green are the indoor and outdoor evacuation areas (Reproduced by permission of Autodesk, Inc. © 2004. All rights reserved.)

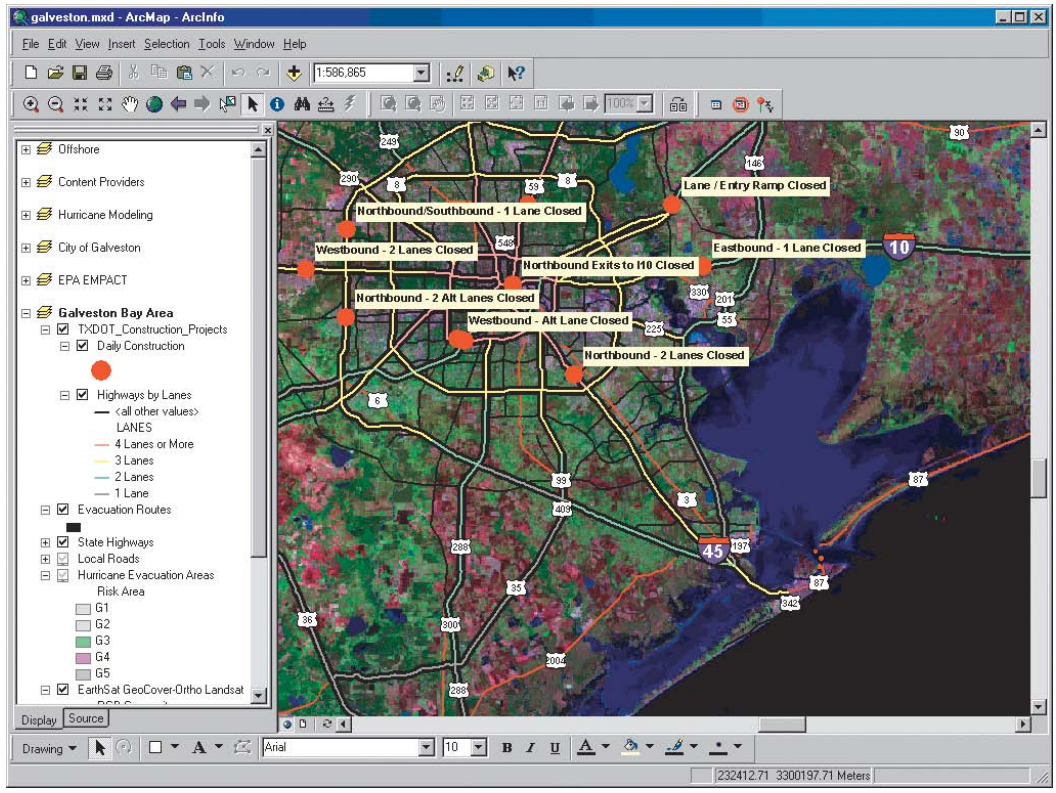


Figure 20.17 Evacuation routes and roads temporarily closed due to construction in the Galveston Bay area after an incident simulated in the joint Homeland Security project of USGS and the Department of Homeland Security (Courtesy: US Geological Survey)

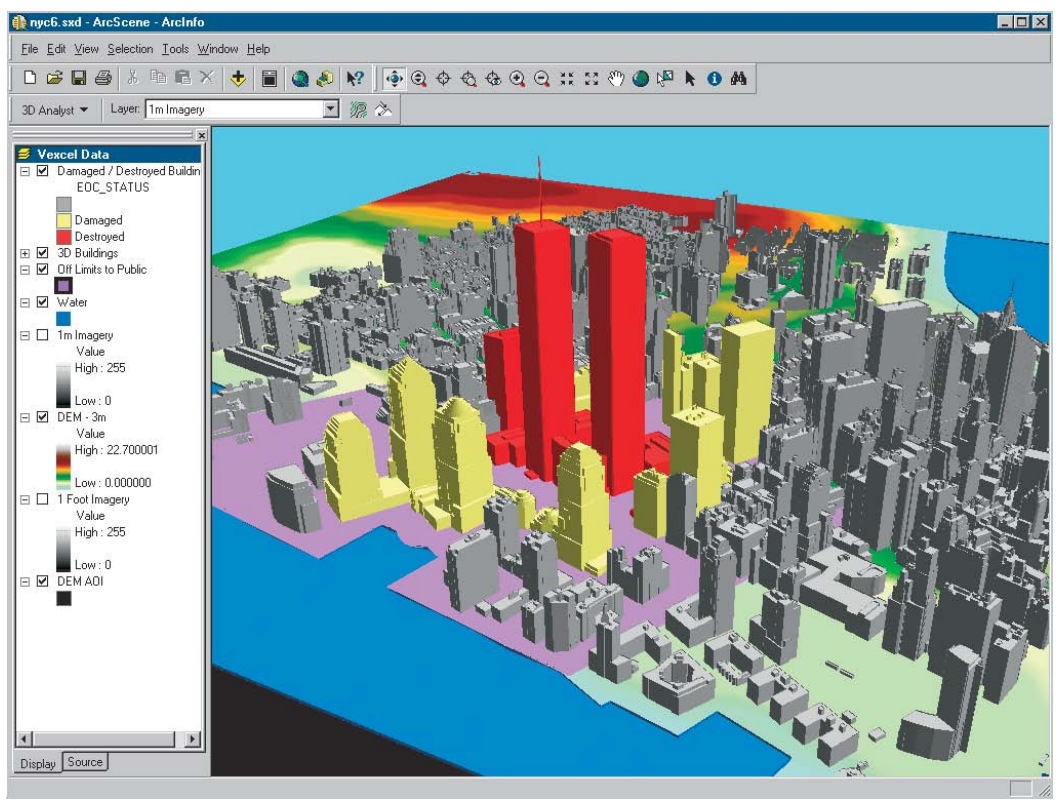


Figure 20.18 The damage caused at Ground Zero. Map created in the Homeland Security project of USGS and the Department of Homeland Security

20.6.3.5 Recovery

These are the actions necessary to return all systems to the pre-recovery state or better. They include both short- and long-term activities such as creating an agreed and shared 'status map'. 'Cleaning up', provision of temporary housing, return of power and water supplies and allocation of government assistance are essential parts of aiding recovery.

20.6.4 Does GIS help or hinder?

There is some doubt about the utility of GI and GIS to those planning terrorist attacks. But there is no doubt that informed and expert use of such assets is enormously beneficial to the populace, both in anticipating and planning for possible attacks and in recovering from these and from natural disasters. Such extreme events place great reliance on system and information inter-operability (see Section 11.2), on frequent practices, and good inter-institutional relationships and partnerships.

Geographic Information – and the ability to assemble, analyze, depict, and communicate it – is key to dealing with terrorism attacks or most other emergencies.

20.7 Conclusions

Despite all the problems of partnerships due to the sometimes different agendas of different players, often there is no alternative to being part of them, e.g., where public and private sectors each provide some crucial element. It can bring vital components – staff skills, technology, marketing skills, and brand image – to add to yours and make a new product or service. It can bring fresh insights on the real user needs and can create superb support and publicity from fellow citizens. It can lead to cost- and risk-sharing.

The probability of success in partnering is raised by keeping the objectives clear, ensuring overall responsibility is agreed and that adequate resourcing exists, and keeping the numbers of partners to the minimum possible, consistent with achieving the desired aims. Remember that ultimately the interactions and trust between the people involved will determine how well the partnership succeeds. But the changes in the world post-9/11 may well ensure that our previous ways of partnership-working may well be modified in the need to safeguard our very way of life: someone has to be in overall charge and procedures enforced. In that sense, the partnerships are more like military ones than the voluntary ones common until now in the GIS community. If so, this could strongly affect NSDI and much else in GIS.

Questions for further study

1. Imagine that you are trying to get three local parties – a local government, a utility company, and a commercial supplier of GI – to work together to produce a Web-accessible GI database for free use. What do you imagine would be the problems? How would these change as you move up the spatial scale to the nation, to the region (e.g., the Americas or Europe), and to the world as a whole?
2. Do a one hour search by whatever means you have available locally for a global database in one thematic area (e.g., land use). How many sources of supply have you found? How many organizations were involved, in what sort of partnerships? What is the relationship between the datasets? What is the internal coherence and quality of what you have found?
3. Find out what preparations have been made in your area to guard against terrorism or major disasters and what role GIS is playing in them
4. How would you 'sell' the benefits of GIS and GI to a politician?

Further reading

Masser I. 2005 *Spatial Data Infrastructure: An Introduction*. Redlands, CA: ESRI Press.

NRC (2004) *Licensing geographic data and services*. National Research Council of the National Academies, Division of Earth and Life Sciences. National Academies Press, Washington DC.

Rhind D. 2005 'National and international geospatial data policies'. In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W. (eds.) *Geographical information Systems: Principles, Techniques, Management and Applications (abridged edition)*. 767–787, Hoboken NJ: Wiley.

21 *Epilog*

This final chapter revisits and summarizes the key ideas introduced in this book and offers some predictions about the future of GIS. We assess the implications of GIS as a technology, as a framework for applications, as an important management tool, and as a fast-developing science. Simply considering any one of these in isolation understates the crucial integrating capacity of GIS that makes it hugely valuable in the real world. We develop this theme of enduring issues and practices in a brief critique of current preoccupations with 'geospatial' systems and technologies. In focusing upon explicitly *geographic* information systems and *geographic* science we also identify some grand challenges that will likely shape the near future of the field, and assess the remit of GIS in the developing global Knowledge Economy.

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.
© 2005 John Wiley & Sons, Ltd. ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

Learning Objectives

After reading this chapter you will:

- Understand some of the main themes that link GISystems, GIScience, and GIStudies, including
  - The impact of the Internet on GIS;
  - How GIS and management interact;
  - How GIS can be applied in real-world practical contexts;
  - The interdependencies between GISystems and GIScience;
- Understand the core organizing themes that link the principles, techniques, analysis, and practice of GIS;
- Be familiar with how the current usage of the term 'geospatial' relates to GIS;
- Be aware of some of the grand challenges that face GIS, and understand some of the ways in which GIS is likely to develop in the next few years.

21.1 Introduction

In *Geographic Information Systems and Science*, we have tried to define the content and context of GIS. For the many that are already experienced in the field we have introduced some different ideas and new ways of thinking about GIS as systems, science, technology, and as a management tool. For those new to GIS, we have tried to write a comprehensive primer that introduces prevailing thinking about the subject. There is compelling evidence that GIS in the 21st century is an immensely important field of endeavor and that rapid progress is taking place on several fronts. With heightened awareness of GIS comes increased uptake and wider participation, informed critical analysis, clearer scientific thinking, and greater realization that management is central to successfully achieving business and societal aims. In this epilog, we close the book with a brief review of some recurring themes, and some personal views on what the future might hold for GIS.

21.2 A consolidation of some recurring themes

21.2.1 Systems, science, and studies

In Chapter 1, we identified geographic information science (GIScience) as the set of fundamental organizing principles and methods of analysis that arise from the use of geographic information systems (GISystems). This definition is helpful, because it suggests that people using GIS must inevitably ask questions about such topics as scale, accuracy, inference, and the ways in which humans interact with computers. Answers to questions like these require systematic scientific inquiry, and access to canons of knowledge such as spatial statistics and cognitive science. Expertise at the cutting edge of GIScience is allied to an understanding of state-of-the-art technology, and we have tried to illustrate how GISystems provide efficient and effective tools to assemble, manage, and display geographic data. For almost all applications, appropriate technology and analysis are necessary but not sufficient preconditions for success in GIS. Therefore, we have emphasized the tactical and strategic importance of management in governing the way in which GIS applications are conceived, resourced, and executed. This is the way in which geographic information *systems support and drive science* in unique applications contexts (Figure 21.1).

Geographic information systems support and drive science.

There are few new principles amongst the many we have enunciated in this book – indeed, many of the issues raised by GIS have been recognized for centuries. Humans have accumulated vast stores of relevant theory and technique, and only the most precocious of GIScientists would suggest that today's GIScience is an activity that has become cut free of its diverse disciplinary roots. Core to this activity is the geographer's ability to relate unique applications to generalized understanding of the ways in which the world works (Section 1.3). Cartographic design principles, so important for effective communication using GIS (Chapter 12), originated long before the advent of either digital computers or GIS. The unique

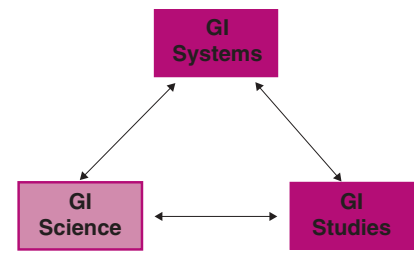


Figure 21.1 The interrelatedness of GI Systems, Science, and Studies

properties of spatial data present daunting challenges for the use of classical statistics, but the established framework of statistics nevertheless provides a starting point for understanding the nature of spatial data and the specification, estimation, and testing of geographic relationships (Chapter 14). The goal of broadening the user base of GIS requires our understanding of the established cognitive principles that guide user interaction. Taken together, this suggests a symbiotic relationship between science and systems: successful applications drive science, as noted above, but geographic information *systems also implement the storehouse of knowledge known as geographic information science*.

Geographic information science grounds successful applications in established scientific practices.

We should, of course, accept that science itself and the exploitation of technology are changing. Most of the exciting developments are now at the boundaries between what we used to think of as different branches of science – notably in the bio-sciences and social sciences. Thus inter- and multi-disciplinary science is at the center of crucial developments (see Box 21.1). The traditional linear model of technology exploitation – where curiosity-driven research spawned new fundamental findings, which were then taken up by engineers and made into diverse products which marketers sought to sell to the public – is now seen as increasingly irrelevant. It is much closer involvement of all parties at all stages and much greater iterative experimentation that works now. The outward-looking but integrative nature of geography as an academic subject, allied to the technical and scientific foundations of GIS, ensure that people educated in both have the right sets of skills to prosper in this changing world.

While GISystems originated in the 1960s, the term *geographic information science* was first coined in the 1990s. Its remit is to make old science more relevant to new systems, and to motivate new research in old areas of science – in order to solve problems that wider changes in data supply and access have brought to the fore. From this perspective GISystems mine GIScience,

Biographical Box **21.1**

Dr Rita Colwell, scientist

Dr Rita Colwell (Figure 21.2) is one of the world's most distinguished scientists. She has authored or co-authored 16 books and more than 600 scientific publications, and produced the award-winning film *Invisible Seas*. She has also received 26 honorary degrees from institutions of higher education, and a geological site in Antarctica, Colwell Massif, was so named in recognition of her work in the Polar Regions. A marine microbiologist by training, she became the first woman to be Director of the US National Science Foundation (NSF) in 1998, a post she held for six years serving both Presidents Clinton and Bush. The NSF initiated its first interdisciplinary research program in the 1970s, but Dr Colwell increased the organization's focus on interdisciplinary research as a result of highly compelling evidence that this was essential in contemporary science and society.

Figure 21.2 Rita Colwell, scientist

Her own research has demonstrated this commitment to working with scientists and practitioners from many disciplines. Dr Colwell's work in Bangladesh spans a quarter of a century and has focused upon solutions to the continuing cholera outbreaks that have plagued the local people. This is a different setting in time, scale, and circumstance to the pioneering work of Dr John Snow (Box 14.1), and today's solutions rely on an interdisciplinary approach that integrates and shares information from medical, environmental, and socioeconomic resources, among others. The language that enables effective communication among the various disciplines is geography, which Dr Colwell has described as the 'ultimate, original multi-disciplinary language', through which data lead to wisdom (see Section 1.2). It is by bringing together and analyzing data from many sources that geography articulated through GIS 'contributes to science and technology and improves the human condition'.

Dr Colwell's own applications have examined the relationships between weather, sea temperature, the population of copepods (shrimp-like microscopic organisms in which cholera bacteria naturally occur), and the ways in which people in Bangladesh live and interact with their environment. These enabled her to devise simple interventions that have helped to reduce seasonal cholera outbreaks by nearly 50 percent – such as filtering water with a folded piece of cloth, typically used in making women's clothing, to reduce the number of copepods and other particulates.

and the future of GISystems depends on a constant flow of new ideas channeled through GIScience. While many of these ideas are not new, the environment of GIS is at last realizing many of the objectives of the 1960s spatial-science paradigm – in an environment that is data rich, offers excellent processing and visualization capabilities, and provides increasingly robust spatial-analysis tools. As such, GIScience extends well beyond the immediate set of issues motivated by today's GISystems, into areas such as cognitive science and information science, and tomorrow's discoveries may make possible a new generation of GISystems that bear little relationship to today's designs. The core organizing principles of today's GIScience may predate GISystems, but together they nevertheless bring huge resources of accumulated knowledge to the operational setting of GIS.

Geographic information science drives the development of systems.

21.2.2 What's in a name? GIS and geospatial

We have already noted (Sections 1.4.2 and 1.6) that the interdisciplinary lexicon includes a number of different terms to describe the activity that we call 'GIS'. The term 'geospatial' (see Section 1.1.1) was first coined in the mid-1990s to describe the data activities of a number of US federal government agencies, and has since entered common parlance, at least among spatially aware professionals. Today, it is also used by some private-sector companies and by the Open GIS Consortium (now the Open Geospatial Consortium: an organization dedicated to interoperability between GISystems – see Section 11.1). The popularity of the term appears to reside in its convenient contraction of geographic ('relating to the Earth') and spatial (relating to any coordinate system at any scale, as in medical imaging or plotting multi-dimensional statistical relationships). This is appealing to some for a number of reasons – the connotation of ownership of GIS by the academic geography community (which has been responsible for rather few technical developments in GIS) is avoided, and this loosening of ties may be taken to suggest that applications need no longer be restricted to geographic scales of measurement (usually taken as ranging from the architectural to the global). It is also a convenient label to use for instruments and technologies (such as GPS and remote sensing) that have been associated with academic disciplines other than geography, and also conveys to some a greater sense of 'hard science'. Last, and not least, it breaks commercial links with the past and allows new companies to compete more effectively.

Our own interim view is that while this terminology succeeds in conveying what is 'special about spatial' (Section 1.1), it fails to convey the achievement of bringing together the power of computer-based generalization with sensitivity to the uniqueness of the real places that are the focus of GIS applications. Much of the thrust of research in GIScience is concerned with making GIS-based representations sensitive to context, and representing an exciting world that is vivid, unique, and real. 'Geospatial' has been used to imply commonalties of approach with GIS in describing the very small minority of applications that are neither at geographic scales of measurement nor are adequately described by other discipline-specific labels (such as computer-aided design and photogrammetry). Our own view is that the richness and diversity of GIScience may appear compromised by a term that implies that space is sterile and amorphous rather than variegated and rich in texture. We think that this may also be the view of academics working in the range of geographically enabled disciplines, generations of whom have dedicated their careers to measuring and understanding the variable characteristics of the Earth's surface. In short, we have concerns that a focus upon the 'geospatial' implies a preoccupation with technology and low-order data concepts, rather than the higher levels of the chain of understanding (evidence, knowledge, and wisdom) described in Section 1.2. Still others would contend that, in a different sense, it is graphicacy (visual interpretation of explicitly geographic phenomena) that is most immediate, exciting, and central to GIS-based analysis.

The term 'geospatial' acknowledges what is 'special about spatial' but neglects what is 'magic [and unique] about geographic'.

This point assumes more than semantic importance when we consider its implications for GI education (Section 1.5.5). Given the long heritage of the term 'GIS', its popularity in universities, and formal place in the academic curriculum, this would seem to be the preferred term for an important scientific, technical, and business field. It is also a term that is recognized increasingly in high schools, not least through outreach activities such as GIS Day. Geography does not always make the best of its opportunities to promote either GIScience or real-world relevance, but considerable intellectual capital is tied up in a discipline that enjoys a privileged status on the (high school and university) curriculum in many parts of the world, and has considerable momentum behind it in others. Promulgation of a newer term, which apparently fails to confer any unequivocal advantages, is likely to confuse the uninitiated and, worse still, present applications as devoid of real-world scientific context.

21.2.3 Geography's revenge – and GIStudies

The Internet evolved out of projects funded by the US Defense Advanced Research Projects Agency (DARPA), beginning in the late 1960s, to tie computers together with high-speed communications links. It was designed to be highly redundant, so that several possible paths could be followed between any pair of computers, ensuring that the network would survive limited attack. Thus, an email sent from one computer to another will be broken into packets by the Internet's protocols, and individual packets may follow several different routes, being reassembled on arrival at their destination. It was also designed to

overcome geography, to make it possible to communicate without respect to distance. Today, few Internet protocol (IP) addresses give any indication of the user's location (Section 11.3.2). This ability of the Internet to overcome distance inspired much extravagant prose in the early years of the network's popularization, culminating in 1997 with the publication of the first edition of Frances Cairncross's book *The Death of Distance: How the Communications Revolution Will Change Our Lives*.

However, it soon became obvious that while electronic communication had major benefits in its ability to overcome distance – allowing people to send email to friends on other continents or to access foreign websites at speeds that bore no relationship to distance – there were still major reasons why location and distance remained important. Most people who work at home still need to visit the office occasionally; most people are more likely to send email to nearby addresses than distant ones; and information on nearby restaurants is usually more useful than information on restaurants in other countries. Businesses were quick to recognize this, and to find ways to determine the location of a customer's computer or cellphone, using the methods discussed in Section 11.3.2. An article titled 'The Revenge of Geography' appeared in *The Economist* in March 2003, and laid out the reasons for a resurgence of interest in distance and location in Internet applications: 'It was naive to imagine that the global reach of the Internet would make geography irrelevant. Wireline and wireless technologies have bound the virtual and physical worlds closer than ever'.

Thus far our discussion has not addressed the important need to reflect on GISystems, their impact on society, and their appropriate place within it. We noted in Section 1.7 that any technology inevitably reflects in part the nature of the society that invents and uses it, and in turn influences the way that society evolves. GIStudies is not about the technology itself, or about the science that makes that technology possible. Rather it concerns how that technology is embedded in a societal context, and how it affects such issues as the importance of distance and location to that society. Careful reflection is a vital part of the role of the academy, and GIStudies is a vital part of the academy's relationship to GISystems. As its title suggests, this book has tended to focus more on the technical and scientific aspects of GISystems and GIScience than on the reflective aspects of GIStudies, or the fast-developing field of public participation in GIS. There is an expanding literature expressing a critical perspective on GIS and its role in science and society. We welcome this as long as there remains recognition that applied science creates value, and that there is a net worth to activities that nevertheless inevitably benefit different groups in society to differing extents.

GIStudies is concerned with the ways in which systems and science interact with society.

However important it may seem, information technology is inevitably limited in its ability to represent the complexity of the real world, or to address and solve its problems. Understanding its limitations has been one of the themes of this book, notably in the chapters on representation, uncertainty, and management. Sometimes even the most skeptical of GISystems users may fail to 'see the wood for the trees', and it is important that GIStudies continue to attract people who are less immersed in the technology and can think deeply about the meaning of GIS and its role in society, and about GIS weaknesses and limitations. Yet we believe that all users need a good grounding in principles and techniques in order to clarify rather than obscure geographic issues.

21.2.4 GIS and technology

One of our main motivations for writing this book is to rebalance the preoccupation of many in the GIS field with technology. Technology is important to GIS and one of the drivers for change in a rapidly evolving field. Appropriate technology is clearly necessary, but alone it is not sufficient for success in GIS. For some, technology is a constraint on their ability to participate in GIS – adequate and appropriate hardware, software, data, and network connections are not universally accessible – and we have commented elsewhere on digital differentiation between GIS haves and have-nots. For others, GIS technology provides opportunities – most obviously the business opportunities of the GIS vendor and consultant community, but also for those interested in adapting technology to problem solving.

Technology continues to be a major GIS driver and constraint on GIS access.

In the four years since the publication of the first edition of this book, there have been significant changes in technology. Computers and networks are now approximately five times faster, there have been multiple major software releases from each of the main GIS vendors, and the problems of data availability have lessened. As we look forward, some of the key technology trends that will shape GIS in the rest of this decade include:

- GIS servers – a move from the desktop to centrally managed GIS servers possessing similar functionality will provide new ways of delivering GIS content and capabilities to a much enlarged user community. GIS servers make possible the vision of a 'GIS Appliance' – everything in a single package to 'do GIS' (hardware, software, data, and applications). Users will only need to plug their GIS Appliance into the power supply and network to be up and running.
- Services-oriented architectures – GIS enabled Web services will make it much easier to integrate GIS into the information systems strategies of still more organizations, and will drive the acceptance and adoption of GIS by the IT mainstream.
- Spatial data infrastructures (SDI) and geoportals – access to data is still a constraint on GIS, and top-down programs that organize data access via Web geoportals will alleviate these problems.
- Sensor webs – new hardware and network developments now allow low-cost unmanned sensors to be placed in many more remote locations (e.g., on

flying drones, on shipping containers, in vehicles, or on people). These can track and monitor objects that move or change state, and can relay this information to central GIS logging and analysis systems. These open up many new applications for GIS, especially in the military and business logistics.

- Modeling and analysis – our ability to perform GIScience is being enhanced by the availability of improved software for spatial analysis and modeling. Key developments in GIS-based exploratory spatial data analysis (ESDA), geostatistics, and process flow modeling are all stimulating interest in GIS in the traditional sciences, as well as improving our inferential abilities.
- Data management – GIS now have the capability to manage some of the world's largest databases (multi-terabyte: Section 1.2 and Table 1.1), to provide multi-user read-write transactional access, and to coordinate distributed (federated) geographic databases. This will be one of the pillars of future 'massive GIS'.
- Visualization – the ability to visualize geographic data in conventional 2-D maps, 3-D whole Earth views, and a variety of graphical and cartographic visualizations is already providing new scientific insights, as well as helping to popularize GIS.

21.2.5 GIS and management

In a more applications-specific sense, it is clear that a critical understanding of management context is important to ascertain the most efficient and effective means of deploying GIS in organizations. There are two important dimensions to GIS and management: the management of GIS and the use of GIS for management.

Poor management remains the Number One reason why GIS projects fail.

GIS provides an increasingly useful part of management's weaponry in all 'businesses' – whether in commerce, government, not-for-profit organizations, or even academia. It can provide competitive advantage (Section 2.3) and is a useful means of evaluating policy and practice options in both the public and the private sectors. GIS also has an important role to play in communication and information sharing, and in public-policy formulation and implementation.

Although business leaders often do not realize it, most businesses are organized geographically and geography plays an important role in success or failure. Government services departments such as road and park maintenance, retail businesses such as banks and high-street stores, and transportation distribution companies are all critically dependent on geography to organize their business activities. For many of these, overcoming the constraints of geography is central to their strategic and operational decision making, and often gives them a competitive edge. Through GIS, businesses can save money and time, increase efficiency and productivity, improve communication and collaboration, and generate revenue. It can be important to think of GIS as much as a source of revenue as an outflow of expenditure. Generally applicable and accessible return on investment and cost-benefit analysis templates, as well as case studies, will be of considerable value in propelling GIS use forward.

21.2.6 GIS applications

Over the past 10 years, many have begun to realize that a significant number of the world's problems are inherently geographic, both in terms of their causes and their impacts. At the same time, investigation has revealed that cross-cutting interdisciplinary science and social science are required in order to seek and implement solutions. Global warming, the hole in the ozone layer, the AIDS epidemic, desertification, homeland security, social unrest, and other massive problems can all benefit from using GIS to manage data and support investigation. GIS can help integrate data from multiple sources, record how events and processes change over space and time, support the evaluation of alternative explanatory models, and communicate results in an effective and understandable way. Some representative and successful applications include the following.

- *Military geointelligence.* Military organizations throughout the world are changing rapidly and there is a new emphasis on gathering, storing, analyzing, and visualizing intelligence data about territory both overseas and at home. Important sources of intelligence include satellite images, surveillance from aircraft and pilotless drones, images collected from ground-based video cameras, and records of the locations of credit-card and bank transactions (Section 11.3.2). A key aspect of virtually all of this intelligence is its geographic location. Indeed, geography is commonly used to organize and index intelligence information. During the Iraq war, for example, General Tommy Franks – the commanding officer of the USA-led forces – was briefed daily using geointelligence in a GIS. During the recent civil war in the Sudan–Chad border region, humanitarian relief was coordinated using intelligence entered into mobile GIS (see Figure 13.14). An important tool for intelligence gathering is *geoparsing*, a term used to describe tools that scan massive amounts of text, searching for references to places, and using those references to link the text to maps and to other information about the referenced place. MetaCarta (**www.metacarta.com**) is an advanced geoparsing technology, developed in collaboration with US intelligence agencies, that can geoparse extensive collections of Web documents and other material.
- *Megan's Laws.* In many jurisdictions, particularly in the US, laws have been passed requiring local authorities to provide public access to information about the residential locations of sex offenders. They are known as Megan's Laws, named for the young victim of a repeat sex offender who had been resettled in a community without the community's knowledge,

and who reverted to previous behavior. Communities find GIS, and Web-based GIS maps, to be invaluable in meeting the requirements of Megan's Laws, since they give citizens ready access to information in an easily understood form. Figure 21.3 shows an example site maintained by the Sheriff's Department of Pierce County, Washington, which covers Metropolitan Tacoma.

- *West Nile Virus.* West Nile Virus was first recorded in Uganda, but has since spread into many other countries. In North America, it is spread by a species of mosquito, and infects both birds (typically crows and jays) and humans. Research has yielded detailed and accurate models of the relationships between mosquito prevalence, bird infection, and human infection, with the result that effective early warning systems are now available. They require GIS-based mapping of bird habitat and mosquito habitat, records of dead and infected birds, and records of human infections. Figure 21.4 shows an example map of the status of West Nile virus in Pennsylvania, USA. Sean Ahearn of Hunter College of the City University of New York has developed the DYCAST system, which uses GIS functions to monitor the progress of West Nile outbreaks and to provide early warning to health agencies.
- *Agricultural policy.* The European Union Common Agricultural Policy (CAP) has played a very

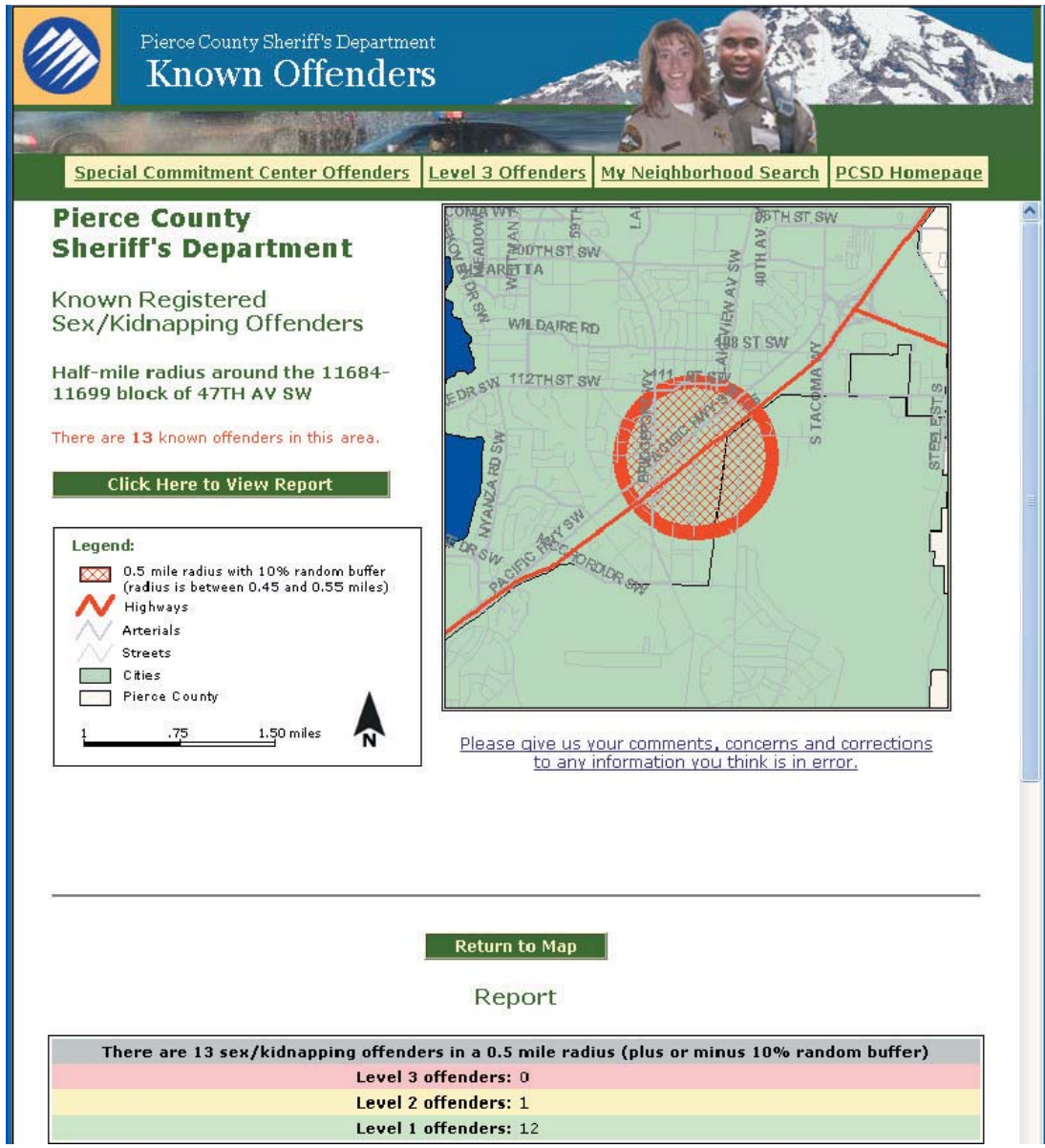


Figure 21.3 Sheriff's Department of Pierce County, Washington, Known registered sex/kidnapping offenders map (**www.co.pierce.wa.us/pc/abtus/ourorg/sheriff/sexoff/sormain.htm**)

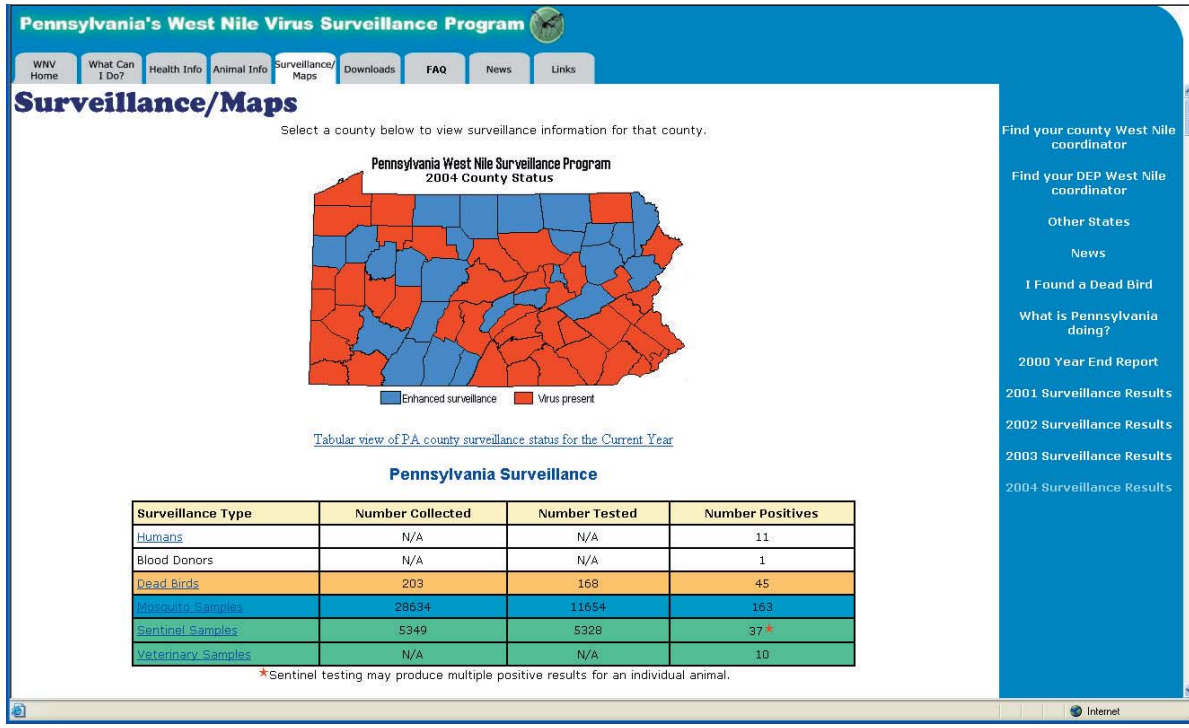


Figure 21.4 Pennsylvania West Nile Virus Surveillance Program website showing 2004 County Status Statistics

significant role in the decision-making process of farmers through enforcement of production quotas, and market subsidies. It has caused a restructuring of farm systems in countries from Ireland in the west to the new succession countries in the east. GIS has been used to spot fraudulent claims from EU farmers that declare more land than they have in order to increase the subsidies they receive. Satellite imagery combined with land parcel boundaries has been used to assess the crop area under production, and is then compared with the census returns from the farmers. This type of land resource assessment is a classic use of GIS that was employed in the first system that created a land inventory for Canada in the 1960s (Section 1.4.1). GIS automates the process of assessing and measuring, standardizes the operations, and builds a database for long-term use. In Italy alone, the system is understood to have narrowed the difference between the area declared for subsidies and the actual area from 9 to 2 per cent since 1999, saving the taxpayers millions of euros. In the enlarged EU, it is expected that over six million farmers will declare fifty million fields every year (see **www.euractiv.com**). GIS is also equally applicable to other agricultural policies – such as systems that monitor whether farmers are safeguarding their land, or the extent of land opened up for recreational activities.

- The United Nations Office on Drugs and Crime and the Myanmar government are using GIS to monitor the poppy harvest in Myanmar, the main opium producer in Southeast Asia. Using a combination of satellite imagery and ground survey, they demonstrated that there was a 29% reduction in crop area from 2003 to 2004. Only by effective monitoring can the size of the problem and success of eradication programs be measured. Figure 21.5 shows one of the maps produced from this survey that are used to communicate the status of the program to key decision makers and stakeholders.

GIS can play a significant role in helping to resolve many types of global problems – but only if it works 'with the grain' inside other initiatives.

These examples are by no means unique – indeed they are repeated over and over again in many application areas. Each illustrates visibly how GIS is literally being used to save and change people's lives, as well as to deliver significant economic, social, environmental, and security benefits.

21.3 Ten 'grand challenges' for GIS

Clearly, GIS has come a very long way in its four decades of existence. Nevertheless, there is still much to be done

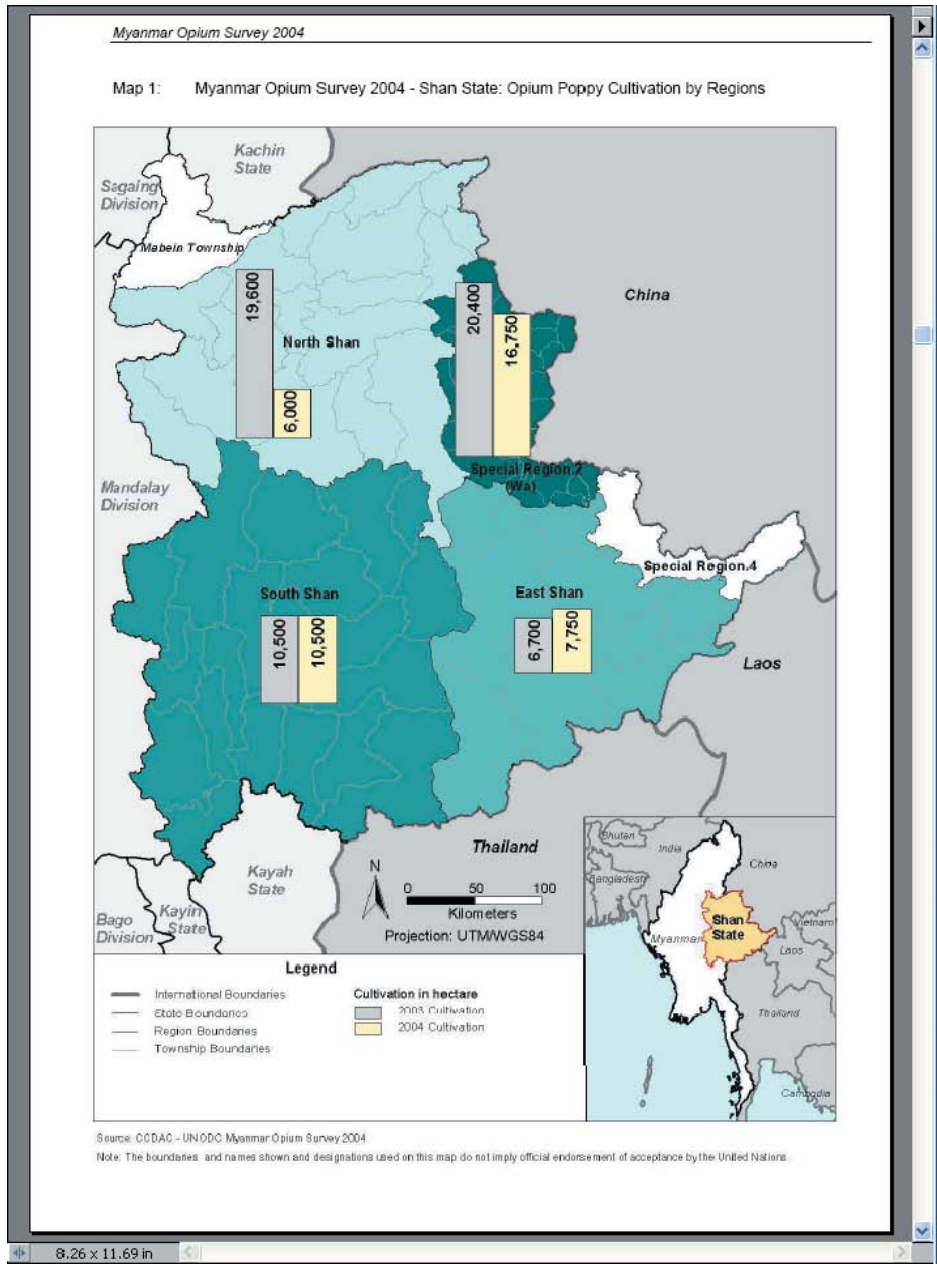


Figure 21.5 Myanmar Opium Survey 2004 – Shan State: Opium Poppy cultivation by regions (Reproduced by permission of CCDAC, Union of Myanmar)

if GIS is to realize its full potential in the vast majority of public and private organizations. Here we outline a series of 'grand challenges', in no particular order, which we believe represent the most significant and urgent action items. Each of these grand challenges can be achieved within the next five years.

21.3.1 Global data layers

Without data, GIS can make little or no contribution to the suite of problems which face us. Over the past 100 years, there have been many attempts to assemble global data layers. In Chapter 20, we described the painfully slow progress to date. The fact that higher resolution data is available for the USA as compared with much of the rest of the world reflects not a technology issue but rather a political challenge – centered on the question of whether one nation should produce data for another without the latter's agreement, the scale at which data might be created and commercially exploited, and the data standards that should be adhered to.

There is no doubt in our minds that GIS use is being held back by a lack of basic data layers, and

there remains much wasteful duplication of effort and expense. Many of the outstanding problems are not technical, but (as indicated above) relate to commercial issues, coordination, agreement on standards, and politics. We feel that wider use of GIS would be facilitated if global layers were widely available for: street centerlines (at a positional accuracy of 200 m for 1:100 000 topographic mapping); surface images at a similar or better resolution (this would be around 1 petabyte of data for the land area alone); a terrain dataset; a comprehensive gazetteer (placename layer); hydrography data; data on cultural features to an agreed global specification; and geodetic control data. It is easy to say what is needed, especially from the perspective of academics without exposure to risk from getting conjectures wrong. The reality is that things happen when the incentive mechanisms are strong enough (Chapter 18). As we showed in Chapter 20, such incentives are usually commercial, related to defense and homeland security, or are political grand gestures. Altruism usually plays little part and the last driver is often ephemeral (but see Section 21.4). Fulfillment of this agenda thus requires that those making the decisions understand that there are advantages to their missions in creating data and subsequently sharing them. This will only occur through carefully planned and targeted strategies and persistent lobbying.

21.3.2 The GIS profession

Historically, GIS has grown bottom up as a loose assemblage of university academics, government employees, commercial workers, and other interested individuals. The academic discipline of geography has contributed many of the intellectual underpinnings of GIS, but other academic and non-academic societies have also contributed a great deal. GIS has now reached a critical mass that has prompted the creation of a code of ethics, and a certification scheme. These could be considered the first stages on the road to establishing a GIS profession, following much the same pattern established in the fields of, for example, geology, chemistry, land surveying, and law. Internationally, this community would embrace the places in the world where GIS has a long history, together with those that are relative newcomers to GIS as an area of activity (see Box 21.2)

For this to happen, the presently disparate GIS-related societies would need to set aside their competitive instincts towards each other and pool resources to create a charter that is endorsed by public bodies of some significance. The purpose of this professional body would be to establish standards for craftsmanship, best practices, and basic qualifications, to disseminate information, and to promote the field. Whether this can be done outside a national context is dubious: hitherto, national laws and conventions have strongly shaped professional practice in

Biographical Box **21.2**

Dr Young-Pyo Kim, Korean Pioneer of NGIS

After a successful career in the Korea Marine Corps, Young-Pyo Kim (Figure 21.6) joined the Korea Research Institute for Human Settlements (KRIHS) in 1979. Over the next decade, he defined the concept and method of Land Assessment in Korea and focused upon a range of land planning and policy issues. In 1986, he became director of the Computing Center in KRIHS, and a champion of using GIS to devise more efficient and effective measures for land planning and policy. Implementing this in a country that, at the time, had little GIS expertise was far from straightforward, and strenuous efforts were necessary to introduce the technology and build capacity. In 1993, Dr Kim organized an official government meeting to establish policy directions for GIS development in Korea. This meeting brought the National GIS project in Korea into being, and he subsequently played a key role in making it happen.

Figure 21.6 Young-Pyo Kim, Korean Pioneer of NGIS

Today Young-Pyo Kim is director of the GIS Center in KRIHS, where he also acts as a coordinator of the Korean National GIS (NGIS) Steering Committee. His leading role in the National GIS project is fundamentally pragmatic, but he believes that GIS is much more than a toolbox. Speaking of the social and philosophical ramifications of GIS he says: 'All things in the universe are restricted by the axes of time and space, and these are recognized by humans. In Eastern classical philosophy, space, time, and humans have been called the three fundamental elements of the universe. These elements cannot be completely integrated in the real world because of the constraints of time and space. However, human construction of three-dimensional cyber-geospace can overcome the separation of time and space and create movement towards a more ubiquitous world. GIS will be a foundation technology in this project.'

different countries (for instance, teaching qualifications and professional accreditation achieved in one country are still not recognized in others). Since GIS impacts on many different disciplines, some of which already have codes of practice and conduct (e.g., engineering), we envisage that the GIS practitioner will be formally assessed upon a wide range of criteria. Establishing a formal profession of GIS is not essential within the field, but will create credibility outside of it. It will draw attention to the science, systems, and studies of GIS in ways that are not currently possible.

21.3.3 The GIS curriculum

One of the major road-blocks to growth of GIS is the paucity of suitably qualified GIS people: technicians, application analysts, and managers. Even though the tertiary education sector supply-side capability has increased several-fold in the past decade and now exceeds 150 000 per annum at over 10 000 institutions, there is considerable variability in content and quality. Moreover, we need GIS 'plugged in' more systematically to a wide range of different discipline-led courses. A global curriculum for GIS will help institutions to create courses in many disciplines, and will also lead to better core GIS courses. It will establish and implement programs reflective of the modern GIS era. In advocating this we recognize the importance of plurality of approach and diversity of outcome; education must be sensitive to different cultural contexts. We believe that much can be common, however, and that economies of scale can be achieved without overly inhibiting further technical and intellectual development.

Thus, our vision is of a select core of spatial principles and techniques that will be used in outreach from geography to GIS-enable the widest range of relevant disciplines in the social and environmental sciences. The skills that qualify professionals to work with GIS are in many respects similar to those needed by professionals who work with any kind of visual information – in short they are the skills of spatial reasoning. Such skills are too rarely taught, but instead must often be acquired informally when students encounter spatial data and visual display, whether in physics, mathematics, or graphic arts. In our vision GIS will form a core component of a systematic approach to the development of spatial skills. This approach – of broadening the outreach of GIS to new disciplinary markets and enhancing the core courses (see Section 19.2.1) – nevertheless only satisfies part of the educational needs. Even more important is the progressive introduction of GIS into schools across the globe. There are already good examples where this has happened, not just in affluent areas of the world (see, for example, Box 21.3).

Although much material taught in schools across the world is of national importance (such as in the teaching of History), an equivalent local focus is essential to

Applications Box **21.3**

GIS in Indian schools

On 5 October 2004, *The Financial Express*, an Indian newspaper, described how the Indian government has empowered around 250 students of 20 schools in the Almora district in Uttaranchal to create digital maps and landscapes of their neighborhood (Figure 21.7). Equipped with hand-held computing devices supported with satellite imagery, the children have been compiling content for a GIS. The pilot program, called 'Mapping the Neighbourhood,' was initiated jointly by the Department of Science and Technology (DST), the Department of Space, and the Centre for Spatial Database Management and Solutions (CSDMS), a non-government organization. The children gathered data on resources, community demographies, livelihood information, land use, infrastructure, and much else, primarily through mapping and questionnaire surveys in villages; they produced – with help – maps and analyses of the data. The aim was to foster greater linkage between young people and their local communities and, hopefully, to solve some of the local problems of water supply and distribution.

In the words of Amitabha Pande, Joint Secretary, Department of Science and Technology, Government of India: 'As a dream, the programme has the potential of transforming the design and structure of education, making it into a creative, interactive, participatory activity which is at the same time socially relevant and useful.'

Figure 21.7 School children in India working on the 'Mapping the Neighbourhood' GIS project (Reproduced by permission of CSDMS)

encourage engagement in the community and to foster a civil society and better inter-generational harmony. As it happens, GIS and geography together have much to offer school-level education now that the PC is ubiquitous and inexpensive software is widely available. Students can gain understanding of current IT through 'doing GIS'. At the same time, they can tackle projects of manifest importance and interest to their parents. The integrative nature of both GIS and geography ensures that many different factors can be brought to bear; understanding the trade offs between multiple factors is excellent training for many problem solving needs in later life. Because of all this, we are convinced that there is a real prospect of introducing more project-led GIS in schools across the globe. A key requirement for success is suitable case studies in different national and regional contexts for teachers, and suitable education and training of the teachers by their peers.

21.3.4 Near-universal empowerment of GIS users

At a routine level, much has been achieved in developed countries through the wide use of Web-enabled technologies, but there is much to be done. At one level the problem centers upon exposure to geographic concepts through education and familiarity with technology through experience – appropriately educated users will typically use a technology first to solicit information, before likely using it to conduct transactions and eventually feeling confident to use it to participate in collaborative decision making, or so one line of thought goes. In practice, however, there is some evidence that the networks that govern access to and control of decision making fail to empower wider user communities, and there is a need to research the way that such characteristics reinforce other sources of structural disadvantage in society. These aggregate perspectives need to be complemented by extending work in public participation in GIS (PPGIS) to GIScience (PPGISc). Such work includes investigation of how user interfaces and exploratory spatial data analysis techniques might foster greater intelligibility and wider understanding. There is a need to develop software that is both easy and safe to use by the overwhelming majority users in any setting – from the street, to the classroom, to the public appeal arising out of development proposals. In the longer term, there is a need to develop our understanding of social and collective cognition, and thus develop our understanding of how GIS can accommodate the widest range of institutions and practices. Collectively, people need to be able to understand both the power of generalization and the mitigating effects of uncertainty.

21.3.5 Global population statistics

Our global knowledge about the most basic of things – how many people there are in different parts of the world and how they live, their state of health, their vulnerability to disasters – is surprisingly poor. Even in Westernized countries, the collection of demographic data by censuses is now under challenge as individuals become less ready to fill in forms. Tallying the population through linking together administrative records is technically demanding and only works well in certain cultures (e.g., in Scandinavia). In other parts of the world, knowledge of populations and (sometimes massive) migrations is often years out of date and of highly questionable accuracy; in most cases, the data published only pertain to large areas, complicating vastly the tackling of very real problems even if helpful in drawing attention to 'big picture' problems.

There have also been many attempts to estimate population numbers from aerial photographs or satellite imagery – using good local knowledge and plausible modeling methods (e.g., see Section 13.3.3). GIS is already making a significant contribution to creating global population data but the source data are highly varied in their quality. Heroic assumptions are necessarily made in the modeling, often derived from data with a spatially variable correlation with population numbers. We must not denigrate the efforts made thus far, but neither should we condone data hoarding beyond the public gaze. If a plethora of new data sources can be brought into the public domain, it should be possible to improve data infrastructures through statistical and GIS skills, honed to particular cultural contexts (e.g., through the national statistical agencies). A global partnership, led by statisticians with GIS skills, could and should benefit everyone. We think this is not just desirable but possible, that progress in the global information economy can make the underpinnings to statistical estimation more secure, and that United Nations agencies may have an important role to play in this regard.

21.3.6 Development of richer geodemographic data infrastructures

We know too little about the functioning of contemporary societies. If robust and fair rules controlling data supply and access can be developed and widely accepted, there is no reason why the tremendous potential of rich government and private-sector datasets should not be unlocked for the greater good. Uninteresting, zonally coarse, infrequently collected surrogate indicators of society are simply no longer good enough for tactical and strategic decision making in consumer-led markets. Nevertheless, huge resource expenditure (especially in the public sector) continues to be authorized using such data. Organizations with international management responsibilities also need more data that traverse national boundaries, but at fine levels of granularity. There is also a need to extend the success of business applications to the challenges of fair and equitable access to public services. We need to unlock the potential of geographically referenced data, many of which are collected at the local level, to represent living conditions, in terms of health, law enforcement, and education. Too much policy is still top down and we need to empower real communities to develop and negotiate their desires and preferences in

order to establish priorities in the provision of public goods. The speed bumps on the information highway manifest an unwillingness to share information, very often because of unfounded fears of disclosure of individual circumstances. We need to lighten up in our views of data supply and access. One domain in which there is evident need concerns the application of geodemographics to the take-up of educational opportunity (Box 21.4).

21.3.7 The transition from geo-centered to ego-centered mapping

The innovation of location based services is providing new challenges to the cartography and mapping communities, driven by restrictions in the viewable areas of

Applications Box **21.4**

The geodemography of geography

Many of tomorrow's GIS experts will gain their first exposure to GIS through undergraduate university programs, in geography itself or in disciplines that are at least partly spatially enabled. In the UK, the issue of access to all higher-education courses is presently a hot political topic, as increasing attention is focused upon the supplementation of educational attainment with other access criteria designed to broaden the social base of those that benefit from an education after high school.

As we have seen, 'doing GIScience' is in part a socially constructed activity. What then are the social backgrounds of its future practitioners? Working at the UK Universities Central Admissions Service (UCAS), researcher Alex Singleton has investigated some of the ways in which geography is a socially divided discipline. Figure 21.8 presents a geodemographic profile (see Section 2.3.3) created using geographic and small-area data of the profiles of parents of UK geography undergraduates admitted in 2003.

Those whose parents are within the wealthiest group (dark blue) are much more likely to be amongst those accepted onto geography courses, and other well-to-do groups (light blue and gray) are similarly over-represented. A final group that is over-represented has parents that are rural as well as affluent (dark green). To what extent is this because aspiring geographers are motivated by the physical environment (e.g., people growing up in rural areas are more likely to become interested in Earth surface processes and landforms) or because they come from affluent social environments? A social engineer might wish to target the high-status groups to broaden the social base of geography, but this would also likely penalize candidates whose choices arise out of subject concerns with the environment, or who have been motivated by the travel opportunities that affluence brings. These are the sorts of issues that are central to understanding fairness, equity, and access, but which are usually either overlooked or crassly simplified when discussing participation in higher education. Geography is pivotal to the creation of modern geodemographic classifications, and the geography of access to different universities within nations and (increasingly) internationally is an increasingly important concern. These topics will be hotly debated in the years to come. Perhaps you have views of your own, based on the characteristics of your own contemporaries at university! Geography is key to understanding geodemography, and vice versa.

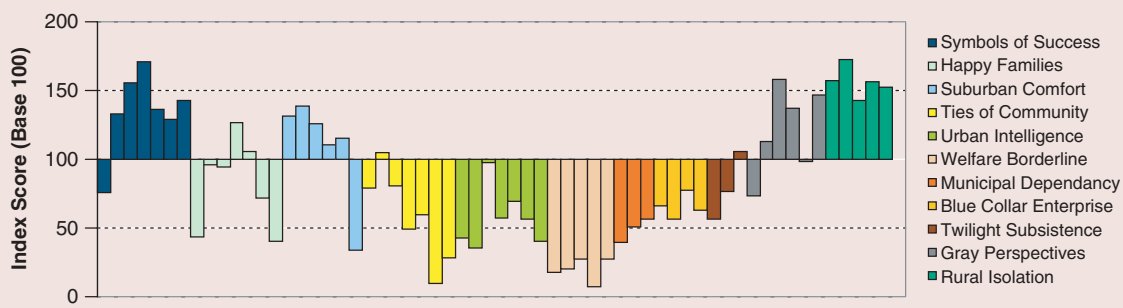


Figure 21.8 Deviations away from a base score of 100 for acceptances to study Geography in 2003, according to Type and Group of the 2003 Mosaic geodemographic system

devices. This technical constraint is associated with a change in the goal of mapping from supplying *products* to providing *services* that are *tailored* to the precise needs of users, and a shift in the application of the cartographer's skill set from developer- to user-centered visualization. Instead of the conventional *bird's eye* view oriented to North and framed by arbitrary lines of latitude and longitude, the user is likely to prefer views that are centered on his or her current position, oriented in the direction of travel (*heads-up*), and with a ground-based perspective (such views are now common in in-car navigation systems). Different classes of users will have different needs, but common characteristics will center upon the need for greater immediacy and ready intelligibility in field situations.

21.3.8 Support data models for a complete range of geographic phenomena

Throughout the book, and particularly in Chapters 3 and 8, we have encountered limitations to the ability of current GIS to represent real geographic phenomena. These range from the emphasis on polylines, which limits the ability of GIS to represent the continuous curves of rivers and roads, to a comparative dearth of methods for representing time and dynamics. In the best of all possible worlds, GIS would include data models for all types of geographic phenomena, including those that are well beyond the range of currently supported data models. It is impossible, for example, to build effective representations today of phenomena such as storms that simultaneously move, change shape, and change internal structure. Similarly, there are no methods for representing watersheds, since a watershed exists for every point on a landscape. This latter example is an interesting hybrid between fields and objects, since every point on a terrain (a field) maps to a unique area (a discrete object) – the term *object field* has been suggested for this kind of phenomenon. Many more examples exist, leading us to suggest that the development of data models for complex geographic phenomena should be given high priority by the GIScience research community.

21.3.9 Combating terrorism yet preserving culture

We treasure our freedom of speech and the other characteristics of democracy. But these are under threat in many parts of the world from terrorism. In Chapter 20, we showed how GIS and GI can make real contributions to anticipating threats and preparing to recover from attacks; it may be – though it is unproven – that GIS and GI can also actually aid the attackers. Put another way, the approach we have espoused – of being open and publishing strategies and plans, working in partnership through consensus between various (sometimes many) parties, sharing and publishing data – might actually benefit attackers. The agents of the state charged with protecting life and limb (e.g., in homeland security, in the intelligence services, and the military) broadly take the view that more directed action to prepare for such eventualities and 'sensible safeguarding of information' is required. An incidental benefit of this concern is that huge resources are being invested in new (e.g., surveillance) technologies, especially by the military, and these are likely to find their way into civilian use at a later stage – just as high-resolution satellites became civilianized. Yet the price of being a more closed and secretive society is a heavy one and such a move may not even be possible: the Internet and Web have empowered people to such an extent that geographic (and other) information is readily available from multiple sources. As individuals, we have little ability to influence the bigger picture and events. As individuals working in GIS, however, we have an obligation to seek ways of reconciling the need to protect human life and assets whilst at the same time maintaining our way of life. Furthermore, as GIS professionals, we have an obligation to act ethically. Formal GIS accreditation (see Section 19.2.2) should be extended to encompass these dilemmas.

21.3.10 Support a wide range of types of geographic simulation

If humans are to be successful in their efforts to improve life on Earth, then an understanding of how the Earth system *works* is essential – any amount of mapping and description of how the Earth *looks* cannot substitute. How much better it would have been, for example, to have understood how the atmosphere's concentration of CO_2 and other greenhouse gases affected climate *before* we embarked on the wholesale burning of hydrocarbons, or to have understood the relationship between forest cover and climate *before* we embarked on wholesale clearing, whether it was of the Amazon basin in the late 20th century, the eastern United States in the early 19th century, or western Europe in the 17th and 18th centuries? Ideas of process are best tested in computational environments such as GIS before they are the subject of human-induced experiment, and in Chapter 16 we described several examples of process simulations in GIS. Alan Turing, the early computer scientist, once proposed an acid test of a computing system – that its outputs would be indistinguishable from those of a real person – and we would like to propose a variant of the Turing test applicable to process simulation. A GIS would pass our version of the Turing test if its predictions about future landscapes and future conditions on the Earth's surface were indistinguishable from actual outcomes – in essence, if the predictions of the computational system exactly matched those of reality. Simulations of many processes, including those discussed in Chapter 16, already come close to this standard, and we anticipate substantial progress in this direction in the next few years as tools for simulation become more and more prevalent in GIS.

The recent advances in GIS modeling and analysis discussed in Chapters 14–16 are already beginning to help us model the world better and test hypotheses about the future in a digital simulation setting. The requirement looking forward is not so much technical as scientific. We need scientific understanding to create and interpret models and scientific data to parameterize and calibrate them.

21.4 Conclusions

The journey that this book has taken us on is an excursion into the work of Geographic Information Systems and Science. We have tried to define the terms GISystems and GIScience and show how they interact closely. For us, both are critically juxtaposed and closely interwoven. One relies intimately on the other and once separated both are less useful and more prone to misinterpretation. We have also discussed the relationship to GIStudies and GIServices. The former connotes the embedding of GIS in society, the latter a mechanism to deliver GIS more effectively to a wider group of users than has heretofore been technically and commercially feasible.

It will be obvious, both from the first 20 chapters but especially from the Epilog, that we are convinced of the intrinsic merits of a geographic perspective on understanding the functioning of the Earth and the societies living on it. Repeatedly, we have demonstrated many examples where GIS has proven valuable to science and society. The potential value is much greater still, however, as horrific natural disasters like the devastation wrought by the tsunami in South East Asia of December 2004 has demonstrated: the use of formal and informal geographic information in the search for survivors and delivering aid to them was crucial. Figure 21.9 shows the scale of the devastation and the need for international contributions and partnerships in the face of such disasters.

Only through GIS can we bring together efficiently information from many different sources, integrate it, and assess the trade offs between different actions. In this sense, GIS is the language of geography since it facilitates communication between the constituent parts of the field.

Although we are confident of a bright future for GIS, we are far from complacent; GIS is still used by only a tiny fraction of human beings. The data and models necessary for evaluating problems and possible solutions rarely exist in the right form. Furthermore, political, economic, and other factors often constrain or deny what is possible given GIS technology and human skills. We see the next decade as one in which technology will inevitably improve still further, but one in which human factors in GIS will become ever more important.

The future of GIS lies in the hands of GIS – not systems, science, studies, or services, but geographic information *students* – the next generation of software developers, application analysts, managers, researchers, and users. We hope that this book will be a contribution to their future and, through their use of GIS, to all our futures. It is our firm view that without GIS the future of the third rock from the sun will be more uncertain and less well managed.

(A)

Figure 21.9 (A) Banda Aceh, Sumatra, Indonesia on 23 June 2004, six months before the tsunami struck. (B) The same area on 28 December 2004, two days after the earthquake nearby and the consequent tsunami. High resolution satellite images such as these are of value in directing disaster relief. They were also reproduced by many newspapers around the world and helped to stimulate understanding of the scale of the problem and financial and other support for the survivors. (Images courtesy of DigitalGlobe, **www.digitalglobe.com**.)

(B)

Figure 21.9 (*continued*)

Questions for further study

1. What range of interests do your student contemporaries bring to the study of GIS? Do you think that a subject of study with diverse roots attracts students with diverse interests and motivations? Give reasons for your answer.
2. What do you think is the most significant impediment to the future success of GIS?
3. With reference to the definitions of GIS in Section 1.4, discuss the view that GIS and GIScience is a socially constructed subject.
4. Choose an application domain and give specific examples of what is?, where is?, why is?, and what if? questions. What datasets would you need to answer these questions successfully?

Further reading

Cairncross F. 2001 *The Death of Distance 2.0: How the Communications Revolution Will Change Our Lives*. Harvard: Harvard Business School Press.

Johnston R.J. 2005 'Geography and GIS'. In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W. (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications (abridged edition)*. Hoboken: Wiley: 39–47.

Nyerges T. 2001 *Geographic Information Systems for Group Decision Making*. London: Taylor and Francis.

Index

Page numbers in *italics* refer to figures; page numbers in **bold** refer to tables; ***bold, italicized*** page numbers refer to boxes.

AA (Automobile Association) (UK), 429, *439*
Abingdon (UK), *17*
ABMs (agent-based models), 372
absolute errors, 401
absolute position, 209
abstraction, 343
  levels, 178–9
Acadia University (Canada), ***386***
acceptance testing, 395
Access (Microsoft), 168, 225, 229
accident rates, spatial autocorrelation studies, 95–101
accuracy
  definition, 128
  in geographic representation, 68
  per-pixel, 139
  per-point, 139
  per-polygon, 139
  and precision compared, 140
  spatial models, 379–81
  see also inaccuracy
acquisition models, 390–5
  limitations, 395–6
Active Server Pages (ASP), ***172***
ADDRESS-POINT, ***411***
adjacency, 54, 349–50
  polygons, 185, 186
ADL *see* Alexandria Digital Library (ADL)
Adobe, 283–4
Advanced Research Projects Agency Network (ARPANET), 18
AED Sicad, 165
aerial photography, 180, 200
  and dasymetric maps, 302, *304*
  geo-referenced, *421*
  photogrammetry, 209
  raster data capture, 202–4
  scanning, 205–6
African Americans, ***327***
Age of Discovery, 67, ***68***
Agenda 21, 462
agent-based models (ABMs), use of term, 372
aggregate models, 371–2
aggregation, 80, *81*, 137, 326
  internal validation, 146–8
  and scale, 147–8
AGI (Association for Geographic Information), 212
AGILE (Association of Geographic Information Laboratories in Europe), ***453***
agricultural policies, and GIS, 477–8
Ahearn, Sean, 477
AHP (Analytical Hierarchy Process), 378–9
Air Ministry (UK), ***106***
Airborne Visible InfraRed Imaging Spectrometer (AVIRIS), 75
AirPhoto, 218
Al Qaeda, 464
Alabama (USA), 165
Alaska (USA), **72**, 430
Alexandria Digital Library (ADL)
  gazetteer, 125, 377
  geolibrary, ***248***
Alexandria Library, 67
algorithms, ***166***, 276, 323, 345
  definition, ***324***
  Douglas–Poiker, 82
  point in polygon analysis, 330
  polygon overlays, 331–2
  for routing problems, 355
  for traveling-salesman problem, 355–8
Allaire Corporation, ***172***
Almora (India), ***481***
AM/FM (automated mapping and facilities management) systems, 285
amalgamation, 80, *81*
Amazon Basin, 484
  deforestation, 55, 350, *353*
ambiguity, 130–2
  use of term, 128
American Standard Code for Information Interchange (ASCII), ***66***, 243–4
AmericanFactFinder, 170
AML (Arc Macro Language), ***170***, 376
analog computers, ***345***
analog models, 364, ***365***
analytical behavioral geography, ***30***
analytical cartography
  use of term, 316
  see also spatial analysis
Analytical Hierarchy Process (AHP), 378–9
angles, 328
Annals of the Association of American Geographers, 27
Announced Land Price Management System (South Korea), ***389***
anomalies, and data mining, 342
Anselin, Luc
  biographical notes, ***351***
  Spatial Econometrics (1988), ***351***
Antarctica, 67, 342, ***473***

Geographic Information Systems and Science, 2nd edition Paul Longley, Michael Goodchild, David Maguire, and David Rhind.

© 2005 John Wiley & Sons, Ltd.

ISBNs: 0-470-87000-1 (HB); 0-470-87001-X (PB)

APIs (application programming interfaces), 159, 163, 213, 219, 230, 231
Apple Macintosh, *296*, 320
application programming interfaces (APIs), 159, 163, 213, 219, 230, 231
Aracaiu (Brazil), *285*
Arc Macro Language (AML), ***170***, 376
ArcCatalog, ***170***, 320, *321*
ArcGIS, 18, 162, *163*, ***169–70***, 320, 331
  development, ***166***
  features, 377
  object-level metadata, 245
  product family, 165
  Spatial Analyst, 173
ArcGIS Desktop, 165
ArcGIS Engine, ***166***, 172
ArcGIS Server, ***166***, 170
ArcGlobe, ***463***
ArcIMS, 18
ArcInfo, *163*, ***166***, 167, ***169–70***, 186
ArcLogistics, 358
ArcMap, *163*, ***170***, *321*
ArcObjects, ***357***
ArcPad, 173
ArcReader, 167
ArcScripts, 377
ArcSDE, 18, 173, 221
ArcToolbox, ***170***
ArcTools, ***170***
ArcView, 18, ***166***, 167, 377
  Spatial Modeler, 376
ArcWeb Services, 170, 174
area coverages, 331
areas, ***88***
Argo 1 (deepsea vehicle), ***101***
Argus, ***171***
Arizona (USA), *349*
Arno, River, 67
ARPANET (Advanced Research Projects Agency Network), 18
arrays, ***183***
ASCII (American Standard Code for Information Interchange), ***66***, 243–4
Asia, 458, 459
Asian bird 'flu, *275*, 279
ASP (Active Server Pages), ***172***
aspect, 327–9
  definition, 328
assets
  intangible vs. tangible, 415
  see also GIS assets
Association for Geographic Information (AGI) (UK), 212
  Special Interest Group on Address Geography, ***411***
Association of Geographic Information Laboratories in Europe (AGILE), ***453***
AT&T, 254
Atlas of Cybergeography, 18
atlases, Internet-based, ***282***
atomistic fallacy, definition, 147
attenuation functions, 94
attribute errors, 401
attributes, 181
  breaks, 277–9
  characteristics, 69–70
  comparing, ***350***
  data capture, 215
  mapping, 273–9
  measurement, 110
  ordinal, ***69***
  representation, 273–9
  transformations, 273–9
  types of, ***69***
  use of term, 68–9
  see also interval attributes; nominal attributes; ratio attributes
augmented reality, 251–2
  use of term, 251
Austerlitz, Battle of (1805), ***266***
Australia, 251, ***310***, 338, 461
  immigration, ***11***
  marine charts, *265*
Austria, ***434***
  annexation, ***290***
AutoCAD, 166, 213
auto-closure, 185
autocorrelation, 104
  temporal, 86–7
  see also spatial autocorrelation
Autodesk GIS Design Server, 173
Autodesk Inc., 24, 165, 170
  GIS software, 166–7
Autodesk Map 3-D, 165, 166, 167, 173
Automated Fluid Minerals Support System, ***195***
automated mapping and facilities management (AM/FM) systems, 285
Automobile Association (AA) (UK), 429, *439*
autonomous agent models, 372
avatars, 305
Avenue, 376, 377
averages, 343
AVIRIS (Airborne Visible InfraRed Imaging Spectrometer), 75
Ayrshire (Scotland), ***307***
azimuthal projections, 118

balanced tree indexes, 231
  limitations, 232
Bangladesh, ***473***
batteries, and location-based services, 256
battles, maps, 285–7
Batty, Michael, ***373***
Bay of Plenty (New Zealand), *275*
BBC (British Broadcasting Corporation), ***307–8***
bears, 71, **72**
Beck, Harry, 299
behavior
  consumer, 46, 50–1
  modeling, ***373–4***
behavioral geography, ***30***
bell curve *see* Gaussian distribution
benchmarks, 204, 395
Bentham, Jeremy, ***124***
Bentley GeoGraphics, 173
Berne Convention (1886), 427
Berners-Lee, Tim, 18
Bertin, Jacques, 273, *274*, 275
Bethlehem (Israel), *276*
Bézier curves, 82, 183
binary counting system, 65–6, 181
binomial distribution, 359
bins, 336
bird flu, 4

Birkin, Mark, 50
Bishop Dunne Catholic School (USA), ***432–3***
bits (data), ***66***
black box approach, 59, 337, 416
BLM (Bureau of Land Management) (USA), ***195***, 331
block adjustment, 210
block encoding, ***182***, ***183***
Bloomingdale Insane Asylum (USA), 252, *253*
Blue Marble Geographics, 172
Bluetooth, ***256***
Bojador, Cape, ***68***
bombs, bouncing, ***365***
Boolean operators, 226
Bosnia, partitioning, 265
Boston University (USA), ***380***
boundary definition, zones, 130
Bournemouth (UK), 48, 50
Brazil, *285*
  settlement expansion, *353*
Brickhill, Paul, *The Dam Busters* (1951), ***365***
Bristol (UK), 301–2, *304*
Britain *see* United Kingdom (UK)
British Broadcasting Corporation (BBC), ***307–8***
British Columbia (Canada), *264*
British Geological Survey, 17
British Nuclear Fuels, 319
broadband communications, 250, 257, 305, ***310***
Brookline, Massachusetts (USA), *281–2*
Brown, James, ***147***
Brown University (USA), ***243***
browsers, 21, 23
B-tree indexes *see* balanced tree indexes
buckets, 336–7
Budapest (Hungary), *11*
buffers, 329–30
  functions, 329
Bureau of the Census (USA), 16–17, 53, 149, 170, ***213***, ***387***
  field work, 242
  population centers, 343
Bureau of Land Management (BLM) (USA), ***195***, 331
Burrough, Peter A., 139, ***147***
Bush, George W., 359, ***473***
business
  direct mail, 26
  geographic issues, 406–8
  GIS applications, 46–51
  and GIS Day, ***449***
  managed, 406–8
  use of term, 406
  see also g-business; GIS business
Business Analyst, 25
business drivers, 410–11, **412–13**, 422–3, 472
Business Objects, 165
bytes, ***66***

C (programming language), 162
C (programming language), 162, ***168***
CAD *see* computer-aided design (CAD)
cadasters, 114–15, ***388***
  definition, 114
Cairncross, Frances, *The Death of Distance: How the Communications Revolution Will Change Our Lives* (1997), 475
calendar systems, 110
calibration, 380
  cellular models, 376
California (USA), 52, *73*, *149*, 165, ***166***, *189*
  geographic information officers, ***402***
  geolibraries, ***248–9***
  GIS initiatives, 456
  IP addresses, *257*
  non-governmental organizations, ***452***
  ozone monitoring, *333*, *338*
  satellite images, *179*
  urbanization, ***451***
California Department of Transportation (Caltrans) (USA), ***369***
Caliper Corporation, 358
CAMA systems *see* Computer Assisted Mass Appraisal (CAMA) systems
Câmara, Antonio, *367*
  biographical notes, ***254–6***
Cambridge (UK), 167, ***386***
Cambridge Group for the History of Population and Social Structure (UK), ***291***
Cambridge University (UK), ***106***, ***291***
Camden Town (UK), ***150–1***, ***152***
Canada, *264*
  land classification, ***69***
  land inventories, 478
  mapping projects, ***386***
  postcodes, 113
  projections, 121, 122
Canada Geographic Information System (CGIS), 16, 17
  datasets, **332**
  development, 42, 55, 323, ***386***
Canada Land Inventory (CLI), 16
  mapping projects, ***386***
Canadian Forest Service, ***387***
cannibalizing, 50
CAP (Common Agricultural Policy), 477–8
capitalism, 381, 406
car insurance, pay-as-you-go, 442, *444*
car park occupancy, ***297–8***
carbon dioxide emissions, *13*
Carrizo Plain (California, USA), *73*
Carroll, Allen, biographical notes, ***282–3***
Cartesian coordinates, 118, 119
cartograms, 299–301, *312*
  definition, 299
  equal population, 299, *301*, *302*
  linear, 299
cartographic models, 376
cartography, ***124***, 263–87, 472
  analytical, 316
  automated, 17
  computer, 180
  digital, 269
  geo-centric vs. ego-centric, 483–4
  goals, 484
  historical background, 265–7
  and maps, 267–70
  overview, 264–7
  see also maps
Cary and Associates, 423
CASE *see* computer-aided software engineering (CASE)
catastrophes, coping strategies, ***406–8***
CBDs (central business districts), 94, 104
cellphones *see* mobile telephones

cellular models, 372–6
  applications, 374–6
  calibration, 376
  validation, 376
Celsius scale, ***69***
CEN (Comité Européen de Normalisation), 214
Census for England and Wales (1881), ***131***
Census of Population (1970) (USA), 17
censuses, 25, 54, ***213***, 327
  data, 170
    reliability, 417
  datasets, ***291***
  future trends, 482
  software, ***411***
  see also Bureau of the Census (USA)
Center for Spatial Information Science (Japan), ***334***
centers, 343–6
central business districts (CBDs), 94, 104
Central Intelligence Agency (CIA) (USA), ***416***
Central Park, Manhattan (USA), 70
central place theory, 104
central tendency, 346
  measures of, 343, 353
Centre for Advanced Spatial Analysis (CASA) (UK), ***150–2***, ***310–11***, ***373***
Centre for GeoInformatics, ***434***
Centre for Spatial Database Management and Solutions (CSDMS) (India), ***481***
centroid, ***88***, 344–5, 346
  mean distance from, 346
CERN (Conseil Européen pour la Recherche Nucléaire), 18
CGIS *see* Canada Geographic Information System (CGIS)
CGS (Czech Geological Survey), ***271***
Chad, *266–7*, 476
Chernobyl (Russia), 448
chess, moves, 358
Chicago (USA), *284*
Chicago Transit Authority (CTA), ***306***
China, 189, *190*, ***463***
Chittenden County, Vermont (USA), ***133***
Cho, G., 430
cholera, ***473***
  Snow's studies, 317–19, 347
Chorley Report (1987) (UK), ***453***
choropleth maps, ***96–8***, 299, 302
  classification schemes, 277–9
Chrisman, N. R., 31
Christaller, Walter, 104
Church, Richard, 366
  biographical notes, ***367–9***
CIA (Central Intelligence Agency) (USA), ***416***
CIO Magazine, ***244***
circles, proportional, 277
Citrix, 161
City University, London (UK), fire, ***406–8***
City University of New York (USA), Hunter College, 477
cityscapes, 305
  visualization, *310*
civil rights, 430
Clark Labs (USA), 173
  George Perkins Marsh Institute, ***380***
Clark University (USA), 378, 411
Clarke, Keith, 374, 376, 377, 380, 390
classification, 14
  fuzziness in, ***133***
clearing houses, 421
CLI *see* Canada Land Inventory (CLI)
client–server GIS, costs, 201
clients
  thick, 23
  thin, 23
climate maps, 279
climatology, ***77***
Clinton, Bill, 453, 455, ***473***
CLM *see* collection-level metadata (CLM)
clock diagrams, 279, ***280–1***
coastline weave problem, 331–2
coastlines
  analysis, 332
  representations, 136–7
Codd, E., 222–3
coding standards, 66
cognition
  and geovisualization, 294–5
  spatial, ***30***
cognitive mapping, ***30***
cognitive science, ***30***
COGO *see* coordinate geometry (COGO)
cold spots, 349
Cold War, and GIS development, 18
ColdFusion, ***172***
collapse, 80, *81*
collection-level metadata (CLM), 247–50
  standards, 247
  use of term, 247
color, in maps, 274, *276*, 277
color separation overlay (CSO), ***307***
Colorado Springs (USA), *468*
Colorado State University (USA), ***195***
Columbia University (USA), 252, *253*
Columbus, Christopher, 122
Colwell, Rita, biographical notes, ***473***
Colwell Massif (Antarctica), ***473***
COM standard, ***170***, 377
Comité Européen Normalisation (CEN), 214
command line interfaces, 159
commerce
  globalization of, ***427***
  see also e-commerce
commercial off-the-shelf (COTS) software, 39, 158, 162
  databases, 173
  evaluation, 393
Committee on Government Reform (USA), 456
Common Agricultural Policy (CAP), 477–8
competition, 50, 419
computational geometry, ***334***
computational models, 364–6
Computers, Environment and Urban Systems, 27
Computer Assisted Mass Appraisal (CAMA) systems
  GIS links, 45
  tasks, 44
computer cartography, 180
computer-aided design (CAD), 166–7
  in geovisualization, 305
  in GIS, 173
  and GIS data models, 180–1
computer-aided software engineering (CASE), 193, *194*
  tools, 219
computers, 22–3
  analog, ***345***

and binary digits, 65–6
developments, 250, 475
early, 242
portable, 250
reduced costs, 18, 242
wearable, 251–2
see also personal computers (PCs)
concatenation, use of term, 148
conceptual models, 178
Concurrent Computer, ***243***
Cone, Leslie, biographical notes, ***195***
confidence limits, 359, 360
confidentiality issues, 429
conflation, use of term, 148
conformal property, 118
confusion matrices, 138, 139, 142–3
Congress (USA), 456, ***465***
NSDI hearings, ***457***
conic projections, 118, *119*
see also Lambert Conformal Conic projection
Connecticut College (USA), ***282***
connectivity
concept of, 54
rules, 192, 229
Conseil Européen pour la Recherche Nucléaire (CERN), 18
conservation, ***451–3***
consolidation, spatial, 47
constant terms, 102
consultancy, ***411***
consumer behavior, 46, 50–1
consumer externalities, 418
consumer's perspective, 138
consumption, 417
contagious diffusion, 47, *48*
containment, 185
Content Standards for Digital Geospatial Metadata (CSDGM) (USA), 245–7
features, ***246***
context sensitivity, 50
continuous fields, 70, 72–4, 136–7, 372
polygon overlays, 331
polygons, 330
representation, 78, *77*
variables, 72
contours, 79
contracts, 395
convenience shopping, GIS applications, 48–51
Conway, John, 372
Cooke, Donald, biographical notes, ***213***
Cooperating State Councils (USA), 456
coordinate geometry (COGO)
object construction tools, 210, *211*
in vector data capture, 210
coordinates, 324
Cartesian, 118, 119
geographic, 115–17
and projections, 117–22
and surveying, 204
Universal Transverse Mercator, 111
copepods, ***473***
copyright, 426, 427
legal actions, 429, ***441***
Cornell University (USA), ***351***
Corps of Engineers (USA), ***243–4***
correlation, 59, 148
see also spatial autocorrelation
Corvallis, Oregon (USA), ***101***
cost recovery issues, ***437***, 439, ***440***
cost-benefit analysis, 391, ***393–4***, 417
cost-effectiveness evaluation, 395
COTS (commercial off-the-shelf) software, 39
Council of Ministers (EU), 458
coupling, models, 377
courses, in GIS, 27–8, **432**, ***434***
Cova, Tom, 52–3, 54, 366
covariance, 86
coverage problems, 353–4
credit cards, 342
critical infrastructure sectors, **466**
cross-validation, 380
crowds, movement modeling, ***373–4***
Crystal Reports, 165
CSDGM (Content Standards for Digital Geospatial Metadata) (USA), 245–7
CSDMS (Centre for Spatial Database Management and Solutions) (India), ***481***
CSO (color separation overlay), ***307***
CTA (Chicago Transit Authority), ***306***
culture, vs. terrorism, 484
customer support, 399
cyberinfrastructure, use of term, 244
cyclic attributes, ***69***
cyclic data, 343
cylindrical equidistant projection, 119–20
cylindrical projections, 118, *119*
Czech Geological Survey (CGS), ***271***
Czech Republic, ***271***
Czechoslovakia (former), ***290***

Dallas, Texas (USA), ***432–3***, 456
Dallas Police Department (USA), ***432–3***
Dam Busters, The (film) (1954), ***365***
dams, bombing, ***365***
Dangermond, Jack, 165, ***166***
Dangermond, Laura, 165
Daratech Inc., 165, 166, *422*
DARPA (Defense Advanced Research Projects Agency) (USA), 474
Dartmouth College (USA), ***402***
dasymetric maps
and aerial photography, 302, *304*
definition, 301
spatial distributions as, 301–2, *303*, *304*
data
analog, 200
cyclic, 343
distributing, 244–50
provenance, 418
theft, 429
use of term, 11–12
value-added resellers, 39
see also digital data; geographic data; GIS data; lifestyles data; metadata; raster data; vector data
data access
direct read, 213, *214*
translation, 213, *214*
data capture
attributes, 215
costs, 201

data capture (*continued*)
use of term, 200
see also geographic data capture; raster data capture; vector data capture
data collection
costs, 201
external sources, 211–15
field work, 242
marine environments, ***102***
projects
evaluation, 201
management, 215
tradeoffs, 215
surveys, *243*
workflow, 201
see also data capture; data transfer; GIS data collection
data control language (DCL), 225, 226
data definition language (DDL), 225
data industry, 25
data maintenance, costs, 201
data management
future trends, 476
support, 399
data manipulation language (DML), 225
data mining, 342, *343*
data models, 64, 162, 219
geographic phenomena, 484
in GIS software, 162–3, 180
overview, 178
and spatial models compared, 364
use of term, 178
see also geographic data models; GIS data models; network data models; object data models; raster data models; vector data models
Data Store, 212
data switchyards, 215
data transfer, 211–15
use of term, 200
data validation, 184–5
database administrators (DBAs), 219, 238
responsibilities, 399
database design, 227–9
models
conceptual, 229
logical, 229
physical, 229
database management systems (DBMSs), 44, 158, *161*, *162*, 192, 218–22
administration tools, 219
advantages, 218
applications, 219
backups, 219
capabilities, 219
commercial off-the-shelf, 173
data loaders, 219
definition, 218–19
disadvantages, 218
extensions, 220–1
GIS links, 45
in GIS software, 165
indexes, 219, 220
licenses, 395
in map production, 283
operations support, 399
recovery, 219
roles, *221*
security, 219
software, 168
spatial, 238
tables, data storage, 222–5
types of, 219–20
updates, 219
see also object database management system (ODBMS); object-relational database management system (ORDBMS); relational database management system (RDBMS)
databases, 12
and data mining, 342, *343*
legal protection, 428–9
overlays, 149
placenames, 125
quality issues, 397
replication, 220
security issues, 397, 465–7
transactional, 7–8
see also geographic databases; GIS databases; global databases
datasets
applicability, 148–9
censuses, ***291***
geographic, 12
histogram views, 322
map views, 320
scatterplot views, 322
spatial, 458
table views, 322
Davey Jones Locker, ***101***
Davidson, John, ***243***
Daylight Saving Time, 110
DB2, ***170***, ***221***
extensions, 220
DBAs *see* database administrators (DBAs)
DBMSs *see* database management systems (DBMSs)
DCL (data control language), 225, 226
DCMs (digital cartographic models), 284
DCW (Digital Chart of the World), 460
DDL (data definition language), 225
Dead Sea, *73*
Death Valley, California (USA), *189*
decimal numbers, ***66***
decision-making
and geovisualization, 309
government, 42
and information, 416–17
and knowledge, 416–17
support infrastructure, **13**
see also multicriteria decision making (MCDM); spatial decision support systems (SDSS)
Declaration of Independence (1776) (USA), 326
deduction, 95, 106
internal validation, 149–52
deductivism, 42
Deepsea Dawn, 265
deepsea vehicles, ***101***
Defense Advanced Research Projects Agency (DARPA) (USA), 474
Defense Mapping Agency (USA), 17, 460
Defense Meteorological Satellite Program (USA), *460*
deforestation, 350, *353*
GIS studies, 55–9

degrees of freedom, 144
Delaunay triangulation, 189, 333
Delft University of Technology (Netherlands), ***270***
DEM *see* digital elevation model (DEM)
density estimation, 337–9
  applications, 338
  kernel, ***152***, 338
Department of Agriculture (USA), 242, 352, ***387***
Department of Commerce (USA), ***387***
Department of Defense (USA), 18, ***457***
Department of Homeland Security (DHS) (USA), *469*
  concept of, 456
  establishment, 464, ***465***
  key facilities, *467*
Department of the Interior (USA), ***248***
  see also Geospatial One-Stop (USA)
Department of Science and Technology (DST) (India), ***481***
Department of Space (India), ***481***
Dertouzos, Michael, ***421***
descriptive summaries, 320, 343–52
  patches, 351–2
deserts, 13, *14*
design
  optimization, 352–3
  and spatial analysis, 346, 352
  see also computer-aided design (CAD); database design; map design
detail, level of, 86
determinism, technological, 410
DGPS (Differential Global Positioning System), 123
DHS *see* Department of Homeland Security (DHS) (USA)
diagrams
  clock, 279, ***280–1***
  rose, 281
  three-dimensional, 64
Differential Global Positioning System (DGPS), principles, 123
diffusion
  contagious, 47, *48*
  hierarchical, 47, 48–51
  innovation, 41
digital, use of term, 65–6
digital cartographic models (DCMs), 284
Digital Chart of the World (DCW), 460
digital data, 200
  storage, 67
  transmission, 67
digital differentiation, 39
digital divide, 39
Digital Earth, 247
  concept of, ***463***
digital elevation model (DEM), 141, *142*, ***182***, ***439***
  advantages, 327–8
  data point estimation, 328–9
  errors, 143–4
  and photogrammetry, 210
  uncertainty effects, 146
digital landscape models (DLMs), 283–4
digital line graphs (DLGs), 79
digital maps, 79, 267
  development, 17
  vs. paper maps, 269–70
digital models, 364
digital raster graphics (DRGs), 79, *79*
digital representation, 65–7, 79
  advantages, 66–7
digital terrain models (DTMs), ***88***
DigitalGlobe, 417, 429, 461–2
digitizing
  heads-up, 207
  human errors, 208
  manual, 206–7
    processes, ***207***
  on-screen, 59
  stream-mode, 207
DIME (Dual Independent Map Encoding), 17, ***213***
dimensions, ***72***
  2.5-D, ***72***, 189, 209
  3-D, 189, 209
  4th, ***88***
  fractal, 57, ***88***
  zero, ***88***
direct mail businesses, 26
direction indicators, 273
Directorate of the Environment (EU), 458
Directorate of Information Analysis and Infrastructure Protection (USA), ***465***
Dirichlet polygons, 333, ***334***
disaster management, 4
discrete objects, 130, 136–7
  geographic representation, 70–2, 74
diseases, 342
  cholera studies, 317–19, 347, ***473***
  global diffusion, 4
  leukemia studies, 319, 347
  West Nile virus, 477, *478*
dispersion, 346–7
displacement, 80, *81*
distance, 324–6
distance decay, 93–5
  effects, 49
  exponential, 94–5
  isotropic, 94
  linear, 94
distance effects, and spatial autocorrelation, 95–101
distance measuring instruments (DMIs), ***105***
distributed GIS, 159, 164, 241–59
  advantages, 243
  data distribution, 244–50
  future trends, 259
  locations, 242–3
  mobile users, 250–7
  networks, 242
  overview, 242–4
  software distribution, 257–8
  standards, 243–4
distribution, of goods, 47
distribution (statistics), 141
  binomial, 359
  see also Gaussian distribution
DLGs (digital line graphs), 79
DLMs (digital landscape models), 283–4
DMIs (distance measuring instruments), ***105***
DML (data manipulation language), 225
DNA, analysis, 8
Dodge, Martin, 18
domains, in georeferencing, 110–11
Dorset (UK), *326*
dot-counting, 323
dot-density maps, 277
Douglas, David, 82

Douglas–Poiker algorithm, 82
Downs, Roger, 251
downtowns, ***150–2***
Dr. Foster (consultancy), ***47***
DRASTIC model, 370
DRGs (digital raster graphics), 79, *79*
DST (Department of Science and Technology) (India), ***481***
DTMs (digital terrain models), ***88***
Dual Independent Map Encoding (DIME), 17, ***213***
Dublin Core metadata standard, 245–7
  elements, ***246***
DWG files, 213
DXF files, 213
DYCAST system, 477
dynamic modeling, 320, 366
dynamic segmentation, 188–9
Dynamic Server, 220
dynamic simulation models, 55

Earth
  curvature effects, *325*
  knowledge of, 64
  measuring, 115–17, 122, 324
  night time map, *460*
  projections, 117–22
  remote sensing, 74–5
  representations, 116
  systems, 484
  topographic databases, 461
  see also Digital Earth
Earth Observing System Data and Information System (EOSDIS), 245
Earth summits, 462
earthquakes, 52
Earth's surface, 64, 69, 70, 79, 80, 82
  domains, 110–11
  generalization, 360–1
  geovisualization, 305–9
  heterogeneity, 360
  measurement, 209
  modeling, 364
  as raster, 372, 374
EarthSat, 218
Earthviewer, ***463***
East Bay Hills, California (USA), 52
East River, *70*
eastings, 117, 121
Eastman, J. Ronald, biographical notes, ***380***
eclipses, 65
ecological fallacy, 147, *148*, 410
  use of term, 49–50
ecological processes, 58, 59
e-commerce
  growth, ***427***
  use of term, 22
  see also g-business
econometrics, ***351***
Economist, The, 430–1, 475
eContent initiative, 458
ECU (Experimental Cartography Unit) (UK), 17
EDA (exploratory data analysis), 292
education and training
  GIS, 27–8, 431, **432**
  web-based, ***28***
eEurope initiative, 458
Einstein, Albert, 15
  theory of relativity, ***77***
electromagnetic spectrum, 75
elevation
  measurement, 122–3
  sampling, 360
elevator maintenance, routing problems, ***356***
ellipsoid of rotation, 116
emergency evacuations, 52–5, 366
  modeling, ***367–9***
  routes, 468, *469*
emergency incidents, *355*
  transportation, 51
emergency management, ***5–7***
  GIS applications, 52–5
  linear referencing, 114
  and location-based services, 254
emergent properties, ***374***
emigration, ***9–11***
encapsulation, ***191***
Encarta (Microsoft), ***410***
Engenuity JLOOX, 172
England, parishes, ***131***
enhancement, 80, *81*
Enterprise Resource Planning (ERP) system, ***356***
environment, GIS applications, 55–60
environmental impact, minimizing, 15
Environmental Protection Agency (USA), ***466***
Environmental Systems Research Institute (ESRI), 23–4, 245, 320, 331, ***357***
  conferences, 422–3, ***433***
  GIS databases, *321*
  GIS software, 165, 167, ***169–70***, 174, 186
    customization, 162, *163*
    developer, 172
    hand-held, 173
  growth, 422–3
  logistic software, 358
  middleware, 221
  modeling software, 370, ***371***, 376
  products, 18, 25
  script libraries, 377
  software development, ***166***
  see also ArcGIS
EOSDIS (Earth Observing System Data and Information System), 245
equal area property, 118
equal interval breaks, 278
Equator, 115, 116, 117, 119, 120
  and absolute position, 209
ERDAS IMAGINE, 173, 376
Erie County, Ohio (USA), *15*
EROS Data Center, 247
ERP (Enterprise Resource Planning) system, ***356***
error function *see* Gaussian distribution
errors, 104, 359
  absolute, 401
  attribute, 401
  distribution, 141
  interval attributes, 139–41
  in nominal attributes, 138–9
  positional, 141–2
  ratio attributes, 139–41
  realization, 143, *145*
  referential, 401

relative positioning, 401
spatial structure of, 142–4
topological, 401
typology, 401
see also measurement error; root mean square error (RMSE)
ESDA (exploratory spatial data analysis), ***351***
ESRI *see* Environmental Systems Research Institute (ESRI)
e-tailing
development, 51
use of term, 47
Eurogeographics, map series, 281
EUROGI (European Umbrella Organisation for Geographic Information), ***453***, 458
Europe
GIS partnerships, 458–9
urbanization, *56*
European Commission (EC), 427, ***453***
Copyright Directive, 429
Directives, 458
Joint Research Centre, 458
European Parliament, 458
European Science Foundation, GISDATA program, ***453***
European Umbrella Organisation for Geographic Information (EUROGI), ***453***, 458
European Union (EU), 423, 429, 458
GIS applications, 477–8
Eurostat, 458
Eurostat Statistical Compendium (2004), ***410***
evacuations *see* emergency evacuations
Evans, Steve, ***310***
events, extreme, 464–70
Everest, Mount, 69
evidence, use of term, 13
exaggeration, 80, *81*
Excel (Microsoft), ***334***, 370, 377
Exeter (Devon), ***307***
expansion, spatial, 47
Expedia (Microsoft), 170
Experian, ***47***
Mosaic, 25, 46, 48
Experimental Cartography Unit (ECU) (UK), 17
exploration, 67
exploratory data analysis (EDA), 292
exploratory spatial data analysis (ESDA), ***351***
developments, 476
Explorer (Microsoft), 23
extensible markup language (XML), 163, 244, 245
external validation
data integration, 148–9
definition, 144
and uncertainty, 144, 148–52
see also internal validation
externalities, 417
types of, 418
valuation issues, 419
extreme events, 464–70

factorials, 355
FAO (Food and Agriculture Organization), ***387***, 448
farmers, 478
Farr, Marc, ***47***
Feature Manipulation Engine, 165, 214
features, labeled, 348–50
Federal Advisory Committee Act (1994) (USA), ***457***
Federal Geographic Data Committee (FGDC) (USA), 128, 450, 453, ***454***
framework definitions, 437
roles, 455
standards, 245
structure, *455*
'Towards a National Geospatial Strategy and Implementation Plan', 456
Federal Information Processing Standard (FIPS), 222
Federal Procurement Data System (USA), 423
Feiner, Steven, 252, *253*
FGDC *see* Federal Geographic Data Committee (FGDC) (USA)
file formats, 213–14
film, coding standards, 66
Financial Express, The (India), ***481***
Find Friends, 254
finger-printing, 429
FIPS (Federal Information Processing Standard), 222
fire ants, ***433***
fires, 67, 366, ***367***, ***406–8***
from space, *460*
Firestone-Bridgestone Company, 342
Fisher, Peter, 129
flattening, definition, 116
flight simulators, 68
floods, 67
hazard mapping, ***435–6***
Florence (Italy), floods, 67
Florida (USA), 254
hurricanes, 52
fly-throughs, 305
FOIA (Freedom of Information Act), 426
Food and Agriculture Organization (FAO), ***387***, 448
footpaths, in national parks, *8*
forestry management, 37, *38*
see also deforestation
forests
and land use, 58
see also deforestation
forms, 13
Formula 1 games, ***255***, ***256***
Fortran (programming language), ***213***
Forward Sortation Areas (FSAs), 113
Fotheringham, Stewart, 31
biographical notes, ***32***
Geographically Weighted Regression: The Analysis of Spatially Varying Relationships (2002), ***32***
Quantitative Geography: Perspectives on Spatial Data Analysis (2000), ***32***
fractal dimensions, 57, ***88***
fractal geometry, 105
fractals, 90, 106, ***106***, ***183***, 351
fractional dimension, 350–2
fractions, ***66***
fragmentation, 350–2
fragmentation statistics, 57
FRAGSTATS, 352
France, 80
Franconia, Minnesota (USA), *450*
Frank, A. U., 139
Franks, Tommy, 476
Free University of Brussels (Belgium), ***351***
Freedom of Information Act (FOIA), 426
freeware, 158, 173–4
frequentism, 133

friction values, 358
FSAs (Forward Sortation Areas), 113
fuzziness, 132–5
  use of term, 128–9
fuzzy logic, 133
fuzzy membership, 133
fuzzy sets, 133

Galileo system, 122
Galveston Bay, Texas (USA), *469*
Gama, Vasco da, ***68***
Game of Life, 372, ***374–5***, 380
games, location-based, ***254***
Gantt charts, 396, *397*
Gates, Bill, 253
Gaussian distribution, 145, 338, 339
  bivariate, 141–2
  characteristic curve, *141*
  definition, 141
  properties, 346
gazetteers, 125, 377
g-business, 25
  use of term, 21
  see also GIServices; location-based services (LBS)
g-commerce, use of term, 22
GBF-DIME *see* Geographic Base Files – Dual Independent Map Encoding (GBF-DIME)
GE *see* General Electric (GE)
GenaMap, ***243***, ***244***
Genasys, ***243***
genealogical research, ***8–10***
General Electric (GE), 406
  GE Energy, 165, 167
  Power Systems, 167
  Smallworld, 48, 167, 406
  Spatial Application Server, 170
  Spatial Intelligence, 167
generalization, ***88***, 103, 269, 343
  classification, 80
  Earth's surface, 360–1
  in geographic representation, 80–2
  methods, 80, *81*
  rules, 80
  symbolization, 80
  weeding, 80–2
geocoding, 123
  postal addresses, ***125***
  services, 377–8
  use of term, 110
  see also georeferencing
GeoCommunicator, ***195***
geocomputation, 152
  use of term, ***32***, 364
geocomputational approach, 149
GeoDa, 173, *323*, *343*, *349*, ***351***, 361
geodemographics, 42, ***47***, 303
  classification schemes, 46
  of geography, ***483***
  infrastructure development, 482–3
  profiling, 49–50
  systems, 25
GeoExplorer, ***416***
geographic, use of term, 8
geographic areas, issues, ***409–10***
Geographic Base Files – Dual Independent Map Encoding (GBF-DIME), 19
geographic data, 67–8
  classification, **200**
  formats, 212–15
  hypothesis testing, 360–1
  Internet resources, 211–12, **214**, 244–5
  nature of, 85–106
  non-spatial, 110
  primary, 200, 211
  as property, 428
  secondary, 200, 211
  sources, **212**
  structuring, 229–35
  types of, 200
  see also GIS data
geographic data capture
  primary, 201–4
  secondary, 205–10
geographic data models, 177–97
  abstraction levels, 178–9
  applications, 186, 192–4
  practice issues, 195
  use of term, 178
  see also GIS data models
geographic databases, 200, 217–39
  design, 227–9
  editing, 235
    multi-user, 235–7
  functions, 226–7
  indexing, 231–5
  limitations, 195
  locking strategies, 236
  maintenance, 235
  management, 45
  object creation, ***191***
  overview, 218
  queries, 44
  query access, 235
  storage, 180, 218
  topology creation, 229–31
  transactional, 7–8
  transactions, 235–6
  types of, 226–7
  versioning, 236–7
  see also database management systems (DBMSs); GIS databases
geographic datasets, 12
geographic detail, 4
geographic facts, ***416***
geographic frameworks, 434–9
  availability, 437–9
  cost recovery issues, ***437***, 439, ***440***
  definitions, 437
  and government, 438–9
geographic information (GI)
  availability, 434–9
  characteristics, ***12***, 415–16, 418–20
  framework, 437–9
  global, 462
  and Internet, 67, 420–2
  legal protection, 428–9
  markets, 419–20
  and monopolies, 419
  organizations, ***435–6***

pricing strategies, ***443***
production, 406
as property, 428
public access, ***454***, 455–6, 465–7
public policy, ***453***
issues, ***441***
security issues, 465–7
updates, ownership issues, 429–30
withdrawal, ***466***
see also geographic(al) information system (GIS); GIScience
Geographic Information House (UK), 450
geographic information officers (GIOs), ***402***
geographic(al) information science (GISc) *see* GIScience
geographic(al) information system (GIS)
acceptance testing, 395
access, 39
acquisition processes, 390, *391*, 394
applications, development, 399
applications-led, 40
bias, 31
business of, 24–8
as business stimulant, 422–3
CAD-based, 173
challenges, 478–85
chronology, **19–21**
components, 18–24, 158, 159
constraints, 440–4
contracts, 395
cost-benefit analysis, 391, ***393–4***
cost-effectiveness evaluation, 395
costs, 422
coupling, 377
customer support, 399
data management support, 399
definitions, 16, 31
evaluation, 395
funding, 397–8
future trends, 478–85
and geography, 31–3
and globalization, 459–64
hardware, 22–3, 158
historical background, 16–18, **19–21**
impacts, 39
implementation, 386, 388–90, 395, 396–8
innovation diffusion, 41
and knowledge economy, 413–15
and knowledge management, 413–15
legal issues, 426–31
legal liability, 430
limitations, 484
linkages, 418–19
management, 24, 385–403, 406–11, 476
management issues, 409–10
markets, 422–3
and nationalism, 459–64
networks, 18–22, 158
object-oriented concepts, ***191***
operations support, 399
personnel issues, 399–401
perspectives, 18
pilot studies, 391, 394
and politics, 459–64
power issues, 32
privacy issues, 430–1
and publishing industry, 27
relevance, 4–11
representation, 63–83
roles, 31, *221*
scholarly journals, ***27***
scope, 4
selection criteria, ***387–8***, 390–6
social contexts, 31
spatially aware professionals, 24
specifications, 394
standards, 159
sustainable, 390–9
and technology, 475–6
themes, 472–8
usage motivations, 41
users, 396
empowerment, 482
requirements, 391
see also distributed GIS; GIS applications; GIS business; GIS data models; GIS databases; GIS software; mobile GIS; public participation in GIS (PPGIS)
Geographic Names Information System (GNIS) (USA), 125, ***439***
geographic networks, themed, 22
geographic objects, ***191***
classification, ***88***, 190
collections, 190–1
connectivity rules, 192, 229
lengths of, ***105***
properties, ***191***
relationships, ***191***, 192
three-dimensional, 71
two-dimensional, 71
see also object data models
geographic phenomena
data models, 484
discrete-object view, 130
measurement, 136–7
representation, 136–7
and scale, 135–6
and uncertainty, 129–52
geographic problems
classification, 4–8
practical, 5, 7
scientific, 5, 7
solving, 4, 16
see also problem-solving
geographic representation
accuracy issues, 68
continuous fields, 70, 72–4, 76, *77*
definition, 67
discrete objects, 70–2, 74
fundamentals, 68–70, 86–7
generalization, 80–2
historical background, 67–8
principles, 86
quality, 128
selection, 229
simplification, 70
tables, 71, **72**
three-dimensional, 71, 305
use of term, 178
see also GIS data models
Geographic Resources Analysis Support System (GRASS) (USA), 174, ***243***

geographic simulation, future trends, 484–5
geographic variables, field representations, 88–9
Geographical Analysis Machines, 319
geographically weighted regression (GWR), ***32***
geography
  behavioral, ***30***
  geodemographics of, ***483***
  human, ***453***
  idiographic, 14, 59
  nomothetic, 14, 59
  Quantitative Revolution, ***32***
Geography Markup Language (GML), 244, 284
Geography Network, 170, 258, 378
geoinformatics, use of term, 28
geoinformation science, use of term, 28
geolibraries, 247, ***248–9***, 258
geolocating
  use of term, 110
  see also georeferencing
geological maps, *17*, 264, ***271***
  series, 283
geomatics, use of term, 28
GeoMedia, 165, 167, ***168–9***
  Grid, 173
  Image, 173
GeoMedia Pro, ***169***
GeoMedia Transaction Server, 222
GeoMedia Viewer, ***169***
GeoMedia WebMap, ***169***, 170
geometries
  classes, 226–7
  computational, ***334***
  fractal, 105
  spatial analysis, 227, *228*
  see also coordinate geometry (COGO)
geomorphology, ***77***
GeoObjects, 172
geoparsing, use of term, 476
geopolitics
  and geovisualization, ***290***
  Germany, ***290***, *291*
Geopolitik, ***290***
geoportals, 21, 22, ***248***
georeferences
  converting, 123–5
  metric, 111
  persistence, 110
  properties, 110
  and spatial resolution, 110
  systems, **112**
georeferencing, 109–26
  accuracy issues, 125
  domains, 110–11
  use of term, 110
  see also linear referencing
georelational models, 186, *187*
geoschematics, 285
geospatial, use of term, 8, 474
Geospatial Asset Management Solution, 167
Geospatial One-Stop (USA), 8, 22, 247, 258
  geolibrary, ***248***, ***249***
  portal, ***195***, 456, ***457***
geostatistical interpolation, ***101***
geostatistics, ***147***, 333
GeoTIFF, 250
geovisualization, 289–313
  applications, 303–5
  and cognition, 294–5
  consolidation, 309–12
  and decision-making, 309
  field of study, 292–3
  and geopolitics, ***290***
  global, 305
  goals, 295
  and hand-held computing, 309
  overview, 290–3
  research, ***294***
  and spatial queries, 293–7
  and transformations, 297–302
  and virtual reality, 305–9
  whole-Earth, 305, *309*
Germany, ***365***
  geopolitics, ***290***, *291*
Gerry, Elbridge, 326
gerrymandering, 148, 326
GI *see* geographic information (GI)
GI2000 initiative, 458
Gibson, Guy, ***365***
GIOs (geographic information officers), ***402***
GIS *see* geographic(al) information system (GIS)
GIS appliances, 475
GIS applications, 4–8, 35–60, 290, 476–8
  business, 46–51
  classification, 41
  daily usage, 36–9
  environmental, 55–60
  genealogical research, ***8–10***
  government, 42–6
  issues, 31–3
  logistics, 51–5
  normative, 7
  in organizations, 41
  positive, 7
  public service, 42–6
  research, 32–3
  sciences, 39–40
  service planning, 46–51
  and spatial queries, 295–7
  transportation, 51–5
GIS assets, 425–45
  legal protection, 427–30
  ownership issues, 428–30
GIS business
  demand, 423
  future trends, 476
  growth indicators, 422–3
  supply, 422–3
GIS Certification Institute, 433
GIS community, use of term, 16
GIS curriculum, 431, 481–2
  global, 481
GIS data, 158
  fitness for purpose, 419
  industry, 25
  management, 25
  use of term, 16, 30
GIS data collection, 199–216
GIS data models, 179–92
  applications, 180
  and computer-aided design, 180–1

development, 179–80
graphical, 180–1
image, 180–1
roles, 179
see also geographic representation
GIS databases, 14, 17, 24, 178, *321*
volumes, **12**, 24
GIS Day, 423, 448
events, ***449***
GIS education, 27–8, 431
courses, **432**, ***434***
curriculum, 481–2
improvement strategies, 433–4
in schools, ***432–3***, 481–2
India, ***481***
GIS Management, First Law of, 429
GIS partnerships, 447–70
advantages, 448
global, for standards, 462–3
local, 448–50
multi-national, 458–9
national, 450–8
GIS profession, 480–1
see also spatially aware professionals (SAPs)
GIS projects
community-based, 450, ***451–3***
failure, 398, 476
Gantt charts, 396, *397*
life-cycle phases, 390
personnel roles, 399–400
GIS servers, 475
GIS software, 14, 16, 23–4, 157–75
applications, 174
architecture, 159–64
three-tier, 159–62
availability, 39
components, 158, 165, ***168***
construction, 165
costs, 24
customization, 162–3, 165
data models, 162–3, 180
database management systems, 165
desktop, 160–2, 163–4, 167, ***168–70***
developer, 172
developments, 18, 159
distributing, 257–8
distribution models, 158
evolution, 158–9
future trends, 174
hand-held, 172–3
industry, 24–5
Internet, 20, 159, 163–4, 170
market share, *166*
markets, 18, 159, 174
and organizations, 159
overlays, 159
professional, 167
server, 164, 167–72
types of, 167–74
vendors, 158, 165–7
see also commercial off-the-shelf (COTS) software
GIS strategy
and implementation plans, 441, *442*
roles, 440–2
GIS teams, 399–400
GIScience, ***77***, ***270***
applications, 40, 59
conferences, 28
development, 28–9
fundamental laws, 86
goals, 31
representation, 70
themes, 472–4
use of term, 28, 473, 474, 485
GISDATA program, ***453***
GIServices, 21–2, 243, 257–8
definition, 257
development, 51
features, 377–8
industry, 25–6
use of term, 377
GIStudies, 30–1, 33
themes, 474–5
GISystems, 29–30
themes, 472–4
use of term, 485
Glacier National Park, Montana (USA), *134*
Glennon, Alan, ***371***, 378
global data layers, 479–80
global databases, 459–62
from satellite imaging, 461–2
Global Land Characteristics Database, 460
Global Mapping Project, *461*
Global Positioning System (GPS), 39, 92, ***428***
accuracies, 123
applications, 37, 123, ***357***
development, 18
differential, 123
hand-held, 172–3
and location-based services, 253–4
principles, 122–3
receivers, ***213***, 460
global spatial data infrastructure (GSDI) Association, *450*, 463–4
activities, 463
conferences, ***453***
definition, 463
globalization, ***414***, 431
of commerce, ***427***
and GIS, 459–64
GlobalReach, *420*
GlobeXplorer, 247
GLONASS system, 122
GML (Geography Markup Language), 244, 284
GNIS (Geographic Names Information System) (USA), 125, ***439***
God's eye view, 251
Goleta, California (USA), *149*, ***249***
Golledge, Reginald, ***30***, 252
Goodchild, M. F., 86, 148
goods, distribution, 47
Google, 213, 256
Gore, Al, *Earth in the Balance* (1992), ***463***
government
decision-making, 42
expenditure on GIS, 423
and geographic frameworks, 438–9
GIS applications, 42–6
and GIS assets, 427

government (*continued*)
GIS initiatives, 456
roles, 419
see also local government
GPRS/UMTS, ***256***
GPS *see* Global Positioning System (GPS)
graphic primitives, *274*
graphic variables, 275–7
graphical user interface (GUI), 159, 160, 161, 165, 295
embedded SQL, 225
Windows-based, ***168***
see also windows, icons, menus, and pointers (WIMP) interface
GRASS (Geographic Resources Analysis Support System) (USA), 174, ***243***
graticules, 118
gravitation, universal, ***77***
Great Britain *see* United Kingdom (UK)
great circles, 117, 324, 346
Great Sandy Strait, Queensland (Australia), *265*
GreenInfo Network, ***451–3***
Greenwich (UK), *115*, 116, 117
Greenwich Meridian, 209
grid computing, use of term, 244
grid indexes, 232
Ground Truth: The Social Implications of Geographic Information Systems (1993), 31
groundwater analyte, multivariate mapping, ***280–1***
groundwater protection models, 370, ***371–2***, 378, 381
GSDI Association *see* global spatial data infrastructure (GSDI) Association
GUI *see* graphical user interface (GUI)
GWR (geographically weighted regression), ***32***

Hajj, ***373***
Hakimi, Louis, theorem, 353, *354*
hand-held computing, and geovisualization, 309
hand-held GIS, 172–3, 242
hardware, 22–3
Harte-Hanks, ***213***
Harvard University (USA), Laboratory for Computer Graphics and Spatial Analysis, 17, ***166***
Haushofer, Karl, ***290***
Hayden, Beth, ***466***
hazard mapping, ***435–6***
Healey, Denis, 381
heartlands, ***290***
Heisenberg's uncertainty principle, 70, ***77***
Henry the Navigator, 67, ***68***
heterogeneity, 87
see also spatial heterogeneity
heuristics, 355–8
Heuvelink, Gerard, 128, 145
biographical notes, ***147***
Error Propagation in Environmental Modelling with GIS (1998), ***147***
hierarchical diffusion, 47, 48–51
highways, 51
linear referencing, 114
routing problems, 355
histograms, 322
historical demography, ***291–2***
history, and uncertainty, ***131–2***
HLS (hue, lightness, and saturation), 277
homeland security, 456, 464, *469*
key facilities, *467*
studies, 465
Hong Kong, 414, 461
hot spots, 349
house hunting, via Internet, 37–8
households, and market area analysis, 48–9
housing value, patterns, *349*
HTML (hypertext markup language), ***172***
HTTP protocol, ***171***
Hudson River, *70*
hue, 274, *276*, 277
hue, lightness, and saturation (HLS), 277
human capital, 406
human–computer interactions, 346
human genome, analysis, 8
human geography, ***453***
humanitarian relief, 255, *266–7*
Huntsville, Alabama (USA), 165
hurricanes
Andrew (1992), ***414***
Floyd (1999), 52
hypertext, 18
hypertext markup language (HTML), ***172***
hypotenuse, 328
hypothesis testing, 320, 359–60
geographic data, 360–1

IBM, 165, 173, 220, 235, ***243***
litigation, 420
ICMA (International City/County Management Association) (USA), 456
ICT *see* information technology (IT)
IDEO, ***255***
IDEs (integrated development environments), 162–3
idiographic geography, 14, 59
Idrisi, 24, 173, 378, *379*, 411
development, ***380***
IDW *see* inverse-distance weighting (IDW)
IGN *see* Institut Géographique National (IGN) (France)
IKONOS (satellite), 25, 200
imagery, classification, 59
images
coding standards, 66
processing, 180
scanning, 205
simplification, 70
stereo, 202
immersive interactions, and public participation in GIS, 302–9
immigration, Australia, ***11***
implementation, 386, 388–90, 395, 396–8
and GIS strategy, 441, *442*
techniques, ***398***
imprecision, use of term, 128–9
IMW (International Map of the World), 459–60
inaccuracy, *140*
use of term, 128–9
indexes
database management systems, 219, 220
geographic databases, 231–5
grid, 232
maintenance, 231–2
R-tree, ***221***, 234–5
see also balanced tree indexes; quadtree indexes

India, ***409***
GIS education, ***481***
Indiana University (USA), ***32***
indicators, ***150–2***
direct, 130
historical, ***131***
indirect, 130
individual models, 371–2
induction, 95, 104
internal validation, 149–52
inductivism, 42
inference, 342
statistical, 91, 359
inferential tests, 359–60
information
applications, 417
characteristics, 417–18
and decision-making, 416–17
filtration, 417
interpretation, 12
and knowledge compared, 12–13
and knowledge economy, 415–22
and management, 416–17
theft, 429
use of term, 12
valuation, 415
see also geographic information (GI); knowledge
information assets, legal protection, 427–8
Information and Communication Technology Committee (Thailand), ***409***
information systems, use of term, 11
information technology (IT), 4, 39, ***409***
developments, ***294***
and GIS education, 482
innovation, 414
investment, 342
and knowledge, 410
Informix, ***170***, 173, 220, ***221***, 235
InfoSplit, 256, *257*
Infrastructure for Spatial Information in Europe (INSPIRE), 458
establishment, 459
inheritance, ***191***
innovation, and knowledge economy, 413–14
innovation adopters, classification, 41
innovation diffusion, 41
inset maps, 272
INSPIRE *see* Infrastructure for Spatial Information in Europe (INSPIRE)
Institut Géographique National (IGN) (France), 17, 80
maps, 200
and NAVTEQ, 461
integers, ***66***, ***88***
integrated development environments (IDEs), 162–3
intellectual property rights (IPR), 426
national, 460
see also copyright
Intelliwhere, 173
interactive automated mapping, patents, ***428***
intercept terms, 102
Intergraph Corporation, 23, 24
GIS software, 165, 167, ***168–9***, 170
hand-held, 173
middleware, 221–2
internal validation
aggregation, 146–8
deduction, 149–52
definition, 144
error propagation, 144–6
induction, 149–52
and uncertainty, 144–52
see also external validation
International Cartographic Association, Commission on Visualization and Virtual Environments, 292
International City/County Management Association (ICMA) (USA), 456
International Geographical Congress, ***409***
International Institute for Geo-Information Science and Earth Observation, ***270***
International Map of the World (IMW), 459–60
International Standards Organization (ISO), 214, 226
and GIS partnerships, 462–3
International Steering Committee for Global Mapping (ISCGM), 460
International Training Centre for Aerial Survey (ITC), ***270***
Internet
access, 421
advantages, 421–2
atlases, ***282***
development, 18–22, 474–5
distribution maps, *22*
gazetteers, 125
geographic data, 211–12, **214**, 244–5
geographic information, 67, 420–2
and GIS, 18, 244–5
GIS data management, 25
GIS software, 159, 163–4, 170
GIServices, 25
house hunting, 37–8
hype, ***421***
impacts, 18
information packets, 66
mapping, and spatial queries, 295–7
protocols, 475
roles, 18
software transmission, 158
usage, *420*
see also World Wide Web (www)
intersection, 54, 185
representation, 71
Intertribal GIS Council (USA), 456
interval attributes, ***69***, 346
classification, 277–9
errors, 139–41
intranets, 18
Inverness Street Market, Camden Town (UK), ***150–1***
inverse-distance weighting (IDW), 333–6
disadvantages, 335
goals, 335
Invisible Seas (film), ***473***
IP addresses, 256, *257*, 475
IPR (intellectual property rights), 426
Iraq War (2003), 476
Ireland, 478
ISCGM (International Steering Committee for Global Mapping), 460
ISO *see* International Standards Organization (ISO)
isochrones, ***96***
isodapanes, ***96***
isohyets, ***96***
isolines, ***96***

isopleth maps, 94, ***96–8***, 137
Ispra (Italy), 458
Israel, *197*, *276*
IT *see* information technology (IT)
Italy, 189, *190*, 458, 478
ITC (International Training Centre for Aerial Survey), ***270***
iteration, 320, 345, 364

Jacqueline Kennedy Onassis Reservoir (USA), *70*
jagged irregularity, 104
Japan, 460
Java, 159, 162, 163, 164, 165, ***374***
  toolkits, 172
Java Script, ***172***
Jefferson, Thomas, 68, 417
Jenks breaks, 277
Johns Hopkins University (USA), ***367***
JPEG 2000 standard, ***183***
JScript, ***172***
JSP (Java Server Pages), ***172***
Jupp, P. E., 343

K function, 347–8
Kalamazoo County, Michigan (USA), *352*
kappa index, 138–9
karst environments, 370, ***371–2***
Kashmir, ***409***
Katalysis, ***411***
Kenai Peninsula (Alaska), 71, **72**
Kendall, Sir Maurice George, 135, **136**
Kentucky (USA), 64, *65*, ***371***
Kenya, 251
kernel density estimation, ***152***, 338
kernel functions, *337*, 338, 346
keys, 215
Kfar-Saba (Israel), *197*
Khunkitti, Suwit, biographical notes, ***409***
Kim, Hyunrai, ***389***
Kim, Young-Pyo, biographical notes, ***480***
knowledge
  codified, 12
  and decision-making, 416–17
  of Earth, 64
  and information compared, 12–13
  as property, 428
  tacit, 12
  types of, 14–15
  use of term, 12–13
  valuation, 415
  see also information
knowledge economy, 431
  and GIS, 413–15
  and information, 415–22
  and innovation, 413–14
  use of term, 413
knowledge industries, 415
  innovation, 414
knowledge management, and GIS, 413–15
Korea Marine Corps, ***480***
Korea Research Institute for Human Settlements (KRIHS), ***480***
Kraak, Menno-Jan, 268
  biographical notes, ***270***
Kraak, Menno-Jan, and Ormeling, Ferjan, *Cartography: Visualization of Spatial Data* (1996), ***270***
Kriging, 333, 336–7
  indicator, 336
  see also semivariograms
KRIHS (Korea Research Institute for Human Settlements), ***480***
KrKonose–Jizera mountains (Czech Republic), ***271***
KSUCTA (Kyrgyz State University of Construction, Transportation and Architecture), ***449***
Kurdistan, ***410***
Kurds, ***410***
Kyrgyz State University of Construction, Transportation and Architecture (KSUCTA), ***449***
Kyrgyzstan, ***449***

lakes, 70–1, 72–4
Lambert Conformal Conic projection, *119*, *120*
  applications, 121, 122
Lancaster University, ***47***
land assessment, ***480***
land inventories, 478
land management, 55, 331
  South Korea, ***388–9***
  systems, ***195***
Land Management Information System (LMIS), ***388–9***
Land Registry (UK), 450
land resource management, 115, 478
Land and Resources Project Office (USA), ***195***, ***196***
land taxes, 114, 223
land use, 55, 58–9, ***371***
  change, 56–7
  classification, ***69***
  classification issues, 302
  fragmentation, 350
  maps, 137
  patterns, 59
  uncertainty issues, 138–9
landbases, 201
  use of term, 193
Landmark Information Group, ***435–6***
Landsat, 17, 218
  images, *181*
  Thematic Mapper, 74
landscapes
  heterogeneity, 87
  settled, 13, *14*
landslides, *14*
  risk, 189, *190*
language, and placenames, 112
LANs (local area networks), 161, 164
latitude, 115–17, ***125***, 324
  definition, 116
  measurement, 122–3
Law Society of Thailand, ***409***
laws
  generalized, 86
  use of term, 15
laws of motion, 15
layers, 180
  global, 479–80
LBS *see* location-based services (LBS)
leadership, 408
least-cost paths, 358–9
Lebensraum, ***290***
legal issues, 426–31
legal liability, 430
legal protection, GIS assets, 427–30

legends, 273
Leica Geosystems, 24, 165, 173
  Total Station, *205*
Leicester (UK), 238
Leicestershire (UK), 92, *93*
length, 324–6
leukemia, studies, 319, 347
Lewis, Meriwether, 68
Lewis and Clark College (USA), ***367***
Lex Ferrum (game), ***255***, ***256***
Lexington, Kentucky (USA), 64, *65*
liberalization, ***414***
libraries, 67
  digital, 421
  geolibraries, 247, ***248–9***, 258
LiDAR (light detection and ranging), 305, *308*
lifestyles data, 46
  applications, 49, 50
  generation, 48
light detection and ranging (LiDAR), 305, *308*
light metadata, concept of, 247
Limerick soils (USA), ***133***
lineage, shared, 148–9
linear referencing, 188
  applications, 114
  systems, 114
lines, 183, *184*, ***220***
  duplicate, 185
  properties, ***88***
  straight, 324
  see also polylines
lines of best fit, 102
Linux, 164
LISA *see* local indicators of spatial association (LISA)
liteware, 158
LizardTech, ***183***
Lloyd, Daryl, ***131–2***
LMIS (Land Management Information System), ***388–9***
local area networks (LANs), 161, 164
local government
  and GIS, 450
  GIS applications, 42, **43–4**, *392*
local indicators of spatial association (LISA), *323*
  development, ***351***
location, 275
  georeferencing, 109–26
  measuring systems, 17, 110
  network, 353, *354*
  point, 353–5
location-allocation problems, 354, 355
location-based services (LBS), 22, 25, 253–6
  applications, ***124***
  definition, 253
  games, ***254–6***
  limitations, 256–7
locator services, use of term, 125
locust infestations, 448
logical positivism, 33
logistics, GIS applications, 51–5
London (UK), *325*
  BBC Weather Centre, ***307***
  Buckingham Palace, *462*
  carnivals, ***373***
  cholera, 317–19, 347
  County Hall, ***311***
  evacuation plans, 468
  fires, 67
  London Eye, ***311***
  Mall, *462*
  noise pollution, *418*
  simulated attacks, *468*
  travel costs, *312*
  Underground maps, 299, *300*
  virtual reality, 251, ***310–11***
London Fire Service, ***406–8***
longitude, 115–17, ***125***, 324
  definition, 115–16
  measurement, 122–3
Longley, P. A., 148
Los Angeles (USA), *325*, 355
loxodromes, 118
loyalty card programs, 48, 50
Lucas County, Ohio (USA), *45*

MacEachren, Alan, 264, 273, *274*, 275, 292
  biographical notes, ***294***
McGill University (Canada), ***386***
Mackenzie, Murdoc, *293*
Mackinder, Halford, ***290***
McMaster R. B., 80, *81*
McMaster University (Canada), ***32***
macros, 24
magazines, ***26***
Maguire, D. J., 165
Maine (USA), 69–70, ***238***
  coasts, ***105***, 106
malaria, and rainfall risk, *267*, 279
Mammoth Cave, Kentucky (USA), ***371***
management
  attitudes toward, 410
  geographic information systems, 24, 385–403, 406–11, 476
  and information, 416–17
  issues, 408
  use of term, 408
  see also data management; emergency management; land management
managers
  assumptions, 416
  GIS issues, 409–10
  project, 400–1
Mandelbrot, Benoit, ***105***, 105, ***106***, 328
Manhattan (USA), 70
map algebra, 376
map composition
  components, 272–3
  process, 272
map design, 28
  controls, 272
  goals, 270–1
  principles, 270–9
map production, 263–87
  key components, 283
  vs. spatial analysis, 316
  workflows, *268*
MapGuide, 166, 167, 170
  components, ***171–2***
MapInfo Corporation, 24, 165, 167, 170
  middleware, 222
MapInfo MapX, 172
MapInfo MapXtreme, 170

MapInfo Professional, 165, 167
MapInfo ProViewer, 167
MapInfo SpatialWare, 173, 222
Mapping Science Committee (MSC) (USA), 450
MapPoint (Microsoft), 40, 170, 174
MapQuest (USA), 25, 170, ***213***, 257, 258
website, 355
maps
applications, 284–7, 290
and cartography, 267–70
as communication mechanisms, 268
definitions, 267
dot-density, 277
early, 67
editing, 17
elements, 283
formal, 264
functions, 268
global, 459–61
inset, 272
isopleth, 94, ***96–8***, 137
limitations, 268
and media, 269–70
multivariate, 279
paper, 76–9, 267
limitations, 290–2
vs. digital, 269–70
reference, 272
scale, 76–8, 80, ***90***, 269, 272, 273
scanning, 205
series, 281–4
soil, 132, ***133***, 279, 350, *352*
specifications, 80
as storage mechanisms, 268
symbolization, 273–9
templates, 283
transitory, 264
types of, 264–5
use of term, 265
on World Wide Web, 21
see also choropleth maps; dasymetric maps; digital maps; geological maps; military maps; thematic maps; topographic maps
Mardia, K. V., 343
margins of error, 359
marine charts, 265
marine environments, and data collection, ***102***
market area, defining, 50
market area analysis, 46
strategies, 48–9
mass communication, 421
Massachusetts (USA), ***166***, ***243***, *281–2*, 326
Massachusetts Institute of Technology (MIT) (USA), ***306***, ***421***
Masscomp, ***243***
Masser, Ian, biographical notes, ***453***
MasterMap, ***437–8***
MAT *see* point of minimum aggregate travel (MAT)
mathematics, ***238***
MAUP *see* modifiable areal unit problem (MAUP)
Maxfield, Bill, ***213***
Maxwell School of Citizenship and Public Affairs (USA), ***124***
MBR *see* minimum bounding rectangle (MBR)
MCDM *see* multicriteria decision making (MCDM)
mean, 343, *344*, 346
mean distance from the centroid, 346
measurement error, 50, 128
in vector data capture, 208–9
measurements, 320, 323–9
reporting, ***140***
Mecca (Saudi Arabia), ***373***
media, and maps, 269–70
median, ***69***, 343, 346, 349, 353
1-median problem, 353
Megan's Laws, 476–7
Mendeleev Centrographical Laboratory (USSR), 346
Mercator projection, 118, *119*, *325*
Oblique, 121–2
see also Universal Transverse Mercator (UTM) system
merge rules, 192
mergers, 406
merging, 80, *81*
meridians, 116
Mersey, River (UK), 319
meta-analyses, 13
MetaCarta, 476
meta-categories, 132
metadata, 59, 152, 215, 273, ***291***, 320
definition, 245
light, 247
standards, 455
see also collection-level metadata (CLM); Dublin Core metadata standard; object-level metadata (OLM)
Meteorological Office (UK), ***107***, ***307***
meteorology, ***77–8***
metrics, 324
Mexico, soil maps, 279
Michigan (USA), *352*
Microsoft, 406, 418
geographic issues, ***409–10***
litigation, 420
Microsoft Access, 168, 225, 229
Microsoft COM, ***170***, 377
Microsoft Encarta, ***410***
Microsoft Excel, ***334***, 370, 377
Microsoft Expedia, 170
Microsoft Explorer, 23
Microsoft MapPoint, 40, 170, 174
Microsoft .Net, 162, 163, 164, 172
Microsoft Terraserver, 247
Microsoft Windows, 164, 167, ***168***, 320
Windows 95, ***409***
Windows XP, 377
Microstation DGN, 213
middleware, 173
extensions, 221–2
Midlands Regional Research Laboratory (UK), 238
migration, patterns, ***9–10***
military geointelligence, 476
military maps, 264, 265
applications, 285–7
in history, ***266***
minimum bounding rectangle (MBR), ***234***
applications, 235
Ministry of Construction and Transportation (MOCT) (South Korea), ***388–9***
Minnesota (USA), 70–1, 72–4, *450*
Minnesota Map Server, 173
minorities, ***327***
Missouri (USA), 68
MIT (Massachusetts Institute of Technology) (USA), ***306***, ***421***

mixels, use of term, 137
mobile GIS, 250–7
  issues, 256–7
mobile telephones, 22, 23, ***357***, 475
  games, ***254–6***
  GPS in, 253–4
MOCT (Ministry of Construction and Transportation) (South Korea), ***388–9***
mode, 343
ModelBuilder, 370, ***371***, 376
modeling
  dynamic, 320, 366
  static, 320
models
  agent-based, 372
  aggregate, 371–2
  cartographic, 376
  conceptual, 178
  digital, 364
  digital terrain, ***88***
  dynamic simulation, 55
  georelational models, 186, *187*
  groundwater protection, 370, ***371–2***, 378, 381
  individual, 371–2
  physical, 178–9, 195
  process, 178
  spatial interaction, 15, 50, 354–5
  static, 369–70
  use of term, 64, 364
  see also acquisition models; agent-based models (ABMs); cellular models; data models; digital elevation model (DEM); spatial models
modifiable areal unit problem (MAUP), 148
  effects, ***152***
MODIS, *70*
Monmonier, Mark, 110, 123
  biographical notes, ***124***
  How to Lie with Maps (1991), ***124***
  Rhumb Lines and Map Wars: A Social History of the Mercator Projection (2004), ***124***
monopolies, and geographic information, 419
Monte Carlo simulations, 145, 381
Moran index, ***100***, 349, 360, 361
  local, 349–50
Moran scatterplots, ***351***
Morehouse, Scott, biographical notes, ***166***
Mosaic, 25, 46, 48, ***483***
mosquitoes, 477
mountains, 70–1
move sets, 358
MrSID (Multiresolution Seamless Image Database), ***183***
MSC (Mapping Science Committee) (USA), 450
Muir Woods National Monument, California (USA), ***248***
multicollinearity, 104
multicriteria decision making (MCDM), 15
  implementation, 378–9
  use of term, 378
multicriteria techniques, 59
Multiresolution Seamless Image Database (MrSID), ***183***
multivariate mapping, 279
  groundwater analyte, ***280–1***
multivariate statistics, error in, 128
Munro, Sir Hugh, 71
Munros (Scotland), 71
Murray Basin (Australia), 251
Muslims, ***373***
Myanmar, 478, *479*

NAD27 (North American Datum of 1927), 116
NAD83 (North American Datum of 1983), 116, 121
Nairobi (Kenya), 251
NASA (National Aeronautics and Space Administration) (US), 245, 459, 461, ***463***
National Academy of Sciences Mapping Science Committee (USA), ***213***
National Aeronautics and Space Administration (NASA) (US), 245, 459, 461, ***463***
National Alliance of Business (USA), ***213***
National Association of Counties (USA), 456
National Association of Realtors (USA), 170
National Audit Office (UK), 415
National Center for Geographic Information and Analysis (USA), ***238***, ***453***
National Cooperative Soil Survey (USA), ***133***
National Digital Geospatial Data Framework (USA), 455, 456
National Environment Board (Thailand), ***409***
National Geographic, ***282***
National Geographic Eighth Edition Atlas of the World, 281, ***282–3***
National Geographic Information System Committee (Thailand), ***409***
National Geographic Map Machine, ***282***
National Geographic Maps, ***283***
National Geographic Society, ***282–3***
  Geography Awareness Week, ***449***
National Geographic Ventures, ***282***
National Geospatial Data Clearinghouse (USA), 455
National Geospatial Programs Office (USA), 456
National Geospatial-Intelligence Agency (NGA) (USA), 8, 17, 423, ***457***, 460
  homeland security studies, 465
National GIS (NGIS) (Korea), ***480***
National Grid (UK), ***111***, 121–2, 320, ***437***
National Image Mosaic (USA), 218
National Integrated Land System (NILS), ***195***
National Intelligence and Mapping Agency (USA), 8
National Land Information Service (UK), 450
National League of Cities (USA), 456
National Map Accuracy Standards (USA), 142
national mapping organization (NMO), 17, 460
  raw data, 461
National Mapping Program (NMP) (USA), ***439***, 456
national parks, footpaths, *8*
National Performance Review (NPR) (USA), 453
National Research Council (NRC) (USA), 247, 450
National Research Council of Thailand, ***409***
national science councils, ***453***
National Science Foundation (NSF) (USA), ***473***
National Spatial Data Infrastructure (NSDI) (USA)
  development, 455–6
  establishment, 450–5
  evaluation, 456–8
  hearings, ***457***
  implementation, ***454***
  stakeholders, 456
  Standards Working Group, 455
national spatial data infrastructures (NSDIs)
  countries, *450*
  establishment, 450, 459

National States Geographic Information Council (USA), 456, ***457***
National Taiwan University, ***77***
National University of Ireland, National Centre for Geocomputation, ***32***
National Water Resources Committee (Thailand), ***409***
nationalism, and GIS, 459–64
NATO (North Atlantic Treaty Organization), 460
natural breaks, 277, *278*
Navigation Technologies, 53, 461
NAVTEQ, 53, ***213***, 461
Nazi Party, ***290***
Neffendorf, Hugh, biographical notes, ***411***
Negroponte, Nicholas, ***421***
neighbors
  similarity, ***100***
  use of term, 89
Netherlands, The, ***147***, ***270***, ***453***
  land management, 55
Netscape, 23, ***248***
network data models, 186–9
  applications, 188
network externalities, 418
network location, 353, *354*
networks, 18–22
  connectivity, 54, 185
  and distributed GIS, 242
  local area, 161, 164
  looped, 186–7
  radial, 186–7
  shortest path analysis, 54
  topology, 192
  tracing, 185
  wide area, 161, 164
  see also triangulated irregular network (TIN)
New England (USA), ***133***
New Haven Census Use Study (USA), ***213***
new information and communication technologies (NICTs), 303
New South Wales (Australia), 14
  soil erosion, 24
New University of Lisbon (Portugal), ***254***
New York (USA), 477
  GIS service centers, ***453***
  Ground Zero, *469*
  terrorist attack, 4, ***5–7***, *70*, 430–1
New York Stock Exchange (USA), ***6***
New Zealand, tidal movements, *275*
Newton, Sir Isaac, 15
  universal gravitation theory, ***77***
NGA *see* National Geospatial-Intelligence Agency (NGA) (USA)
NGIS (National GIS) (Korea), ***480***
NGOs (non-governmental organizations), ***452***
NICTs (new information and communication technologies), 303
NILS (National Integrated Land System), ***195***
NMO *see* national mapping organization (NMO)
NMP (National Mapping Program) (USA), ***439***
nodes, 184, 353
noise
  pollution, *416*
  studies, ***350***
NOKIA N-Gage, ***256***
nominal attributes, ***69***, 89
  errors, 138–9
nomothetic geography, 14, 59
non-governmental organizations (NGOs), ***452***
non-linear effects, 49
normal distribution *see* Gaussian distribution
North America, digital elevation model, ***182***
North American Datum of 1927 (NAD27), 116
North American Datum of 1983 (NAD83), 116, 121
North Atlantic, *325*
North Atlantic Treaty Organization (NATO), 460
North Carolina (USA), 322, *323*
  12th Congressional District, ***327***
northings, 117, 121
Norwegian Ice Cap, ***386***, ***387***
Norwich Union (UK), *444*
Notting Hill Carnival (UK), ***373***
Nottingham University (UK), ***386***
Nova Scotia (Canada), ***386***
NPR (National Performance Review) (USA), 453
NRC (National Research Council) (USA), 247, 450
NSDI *see* National Spatial Data Infrastructure (NSDI) (USA)
NSDIs *see* national spatial data infrastructures (NSDIs)
NSF (National Science Foundation), ***473***
Nuclear Regulatory Commission (USA), ***466***
null hypothesis, 359, 361

Oakland Hills, California (USA), 52, 53, ***367***
object classes, 191, 192, 193
  defining, 229
object data models, 190–2, 196
  applications, 192–4
object database management system (ODBMS)
  definition, 219
  limitations, 219–20
object fields, use of term, 484
ObjectFX, 172
object-level metadata (OLM), 245–7, 250
  advantages, 245
  definition, 245
  disadvantages, 245
  standards, 245
object-relational database management system (ORDBMS), 219
  capabilities, 220
  definition, 220
  extensions, 220
  replication, 220
  storage management, 220
  transactions, 220
Oblique Mercator projection, 121–2
oceanography, ***101***
OCR (optical character recognition), 200, 215
ODBMS *see* object database management system (ODBMS)
ODYSSEY GIS, 17, ***166***
Office of the Deputy Prime Minister (UK), ***150***
Office of Management and Budget (OMB) (USA), 456, ***457***
Office for National Statistics (UK), Geography Advisory Group, ***411***
OGC *see* Open Geospatial Consortium (OGC); Open GIS Consortium (OGC)
Ohio (USA), *370*
  tax assessment, 44, *45*
Ohio University (USA), ***294***
Okabe, Atsuyuki, biographical notes, ***334***
Okabe, Atsuyuki, *et al, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams,* ***334***
Oklahoma (USA), tornadoes, *77*

old age pensioners, spatial autocorrelation studies, *98–9*
OLM *see* object-level metadata (OLM)
OLS (ordinary least squares), 103
Olympic Peninsula, Washington State (USA), *181*
OMB (Office of Management and Budget) (USA), 456, ***457***
ONC (Operational Navigation Chart), 460
OnSite, 166, 173
OnStar, 51, *52*, 254
  applications, 114
Ontario (Canada), accident rate studies, 95–101
ontologies, 64, 82
Open Geospatial Consortium (OGC), 214, 456, 463
  activities, 463, 474
Open GIS Consortium (OGC), ***244***, 463, 474
  standards, 226, ***457***
Open GIS Project, ***244***
open source software, 158, 173–4
Openshaw, Stan, ***32***, 147–8, 149, 319, 347
operational decisions, 7, 55
operational management, issues, 398–9
Operational Navigation Chart (ONC), 460
operations support, 399
opium poppies, 478, *479*
optical character recognition (OCR), 200, 215
optimistic versioning, 236
optimization, 320, 342, 352–9
optimum paths, 358–9
Oracle, 168, ***170***, 173, 220, 229
Oracle Spatial, ***220–1***, 231, 235
OrbImage, 429
ORDBMS *see* object-relational database management system (ORDBMS)
ordinal attributes, ***69***
ordinary least squares (OLS), 103
Ordnance Survey (OS) (UK), 17, 152, *293*, 423, ***436***
  copyright actions, 429, ***441***
  cost recovery, ***439***, ***441***
  MasterMap database, 218, 305
  national framework, ***437–8***
  National Geographic Database, 415
  pricing strategies, ***443***
  and traffic management systems, ***297***
  turnover, 25
Oregon (USA), *349*
Oregon State University (USA), ***101***
organizations
  consolidation, 47
  functions, 46
  GIS applications, 41
  and GIS software, 159
  rationalization, 47
  restructuring, 47
  see also national mapping organization (NMO)
orientation, 210
orienteering problem, 358
Orman, Larry, ***451–3***
Ormeling, Ferjan, ***270***
OS *see* Ordnance Survey (OS) (UK)
Ottawa (Canada), ***386***
overview maps, 272
Oxford (UK), ***290***
ozone monitoring, *333*, *338*, 342

Pacific Ocean, 68, 459
PAF (postcode address file), 49
Pakistan, ***409***
paleoclimatology, ***78***
Pande, Amitabha, ***481***
Pan-Government Agreement (UK), ***443***
panic, ***373***
panoptic gaze, use of term, ***124***
Panopticon, ***124***
parallel coordinate plots, 342, *343*
parallels, 116
PARAMICS, ***368–9***
parcels (property), 45, 114–15, 138–9, 192, 223
  and agricultural policy, 478
  buffering, 329
  map series, 281
parishes, civil, ***131***
PASS (Planning Assistant for Superintendent Scheduling), ***356***
patches
  concept of, 351
  descriptive summaries, 351–2
patents, 427, ***428***
path analysis, 54, *326*
  optimum paths, 358–9
Patriot Act (USA), 431
patterns
  clustered, 347
  and data mining, 342
  dispersed, 347
  measures, 347–50
  random, 347
  see also point patterns
pavement quality, 285, *286*
PCC (percent correctly classified), 138–9
PCGIAP (Permanent Committee on GIS Infrastructure for Asia and the Pacific), 459
PCI Geomatics, 24
PCRaster, 376
PCs *see* personal computers (PCs)
PDA *see* personal digital assistant (PDA)
pdf (Portable Document Format), 283–4
pedestrians, behavior modeling, 372, ***373–4***
Pennsylvania (USA), *142*, *349*, 477, *478*
Pennsylvania State University (USA), 251, ***402***
  GeoVISTA Center, ***294***
percent correctly classified (PCC), 138–9
Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP), 459
personal computers (PCs)
  components, 164
  GIS software, 160–2, 163–4, 167
personal digital assistant (PDA), 22, 23, 60, 253, 264
  in field work, 242
personnel
  confidentiality issues, 429
  issues, 399–401
  professional accreditation, 431–3
  roles, 399–400
  skills, 431–4
  see also managers; spatially aware professionals (SAPs)
pessimistic locking, 236
Pfaff, Rhonda, ***371***, 378
pharmaceutical industry, 419
Philippines, 55–9
photogrammetry
  in vector data capture, 209–10
  workflow, 210, *211*

physical models, 178–9, 195
Pickles, John, 31
picture elements *see* pixels
Pierce County, Washington (USA), 477
pilot projects, 215
pilot studies, 391, 394
Pisa (Italy), 189, *190*
Pitt, William, the Younger, ***266***
Pitt County (USA), 285, *286*
pixels, 70, 137, 181
  accuracy, 139
  in remote sensing, 202
  use of term, 74
placenames, 112
  databases, 125
  duplicated, 110–11
planar enforcement, 186
Planck, Max, quantum theory, ***77***
planimeters, 323
planning
  operational, 396
  service, 46–51
  strategic, 396
Planning Assistant for Superintendent Scheduling (PASS), ***356***
planning consents, ***436***
Plate Carrée projection, 119–20
PLSS *see* Public Land Survey System (PLSS) (USA)
p-median problem, 353
Poiker, Tom, 82
point location, optimization, 353–5
point of minimum aggregate travel (MAT)
  definition, 345
  determination, 345–6
  optimization, 353
point patterns
  analysis, ***88***
  first-order processes, 347
  measures, 347–8
  second-order processes, 347
point in polygon analysis, 50
  procedures, 330
point-of-sale systems, 254–6
points, 183, *184*, ***220***
  accuracy, 139
  properties, ***88***
  shielded, 337
  unlabeled, 347–8
Poland, ***290***
Poles, 120, 121
political districts, 326–7
politics
  and GIS, 459–64
  see also geopolitics
polygon overlays, 148, 330–3
  algorithms, 331–2
  computational issues, 331
  continuous fields, 331
polygon in polygon tests, ***234***
polygons, 71, 75–6, *77*, 180, 183, *184*
  accuracy, 139
  adjacency, 185, 186
  area, *323*
  centroids, 323
  complex, ***220***, ***234***
  errors, 208
  overlapping, 185
  Thiessen, 333
  topology, 185–6, *187*
  in triangulated irregular network, 189
  see also point in polygon analysis
polylines, 76, *77*, 79, 82, 183, 184
  approximations, 324, *325*
  duplicate, 185
  intersections, 185
  in parallel coordinate plots, *343*
polymorphism, ***191***
poppy cultivation, 478, *479*
Popular Science, 254
population, 103, 130, 360
  cartograms, 299, *301*, *302*
  centers, 343
  global statistics, 482
  shifts, 343, *344*
  studies, ***291–2***
population density, *269*, 338
Portable Document Format (pdf), 283–4
Portugal, 67, ***106***
positional control, 17–18
positional error, 141–2
positivism, 33
postal addresses, 110, 113
  geocoding, ***125***
  legal issues, 428
postal systems, 110, 111–12
postcode address file (PAF), 49
postcodes, 17, 48–9, 113
PostGIS, 173–4
PostgreSQL, 174
potatoes, yields, 135, **136**
PPGIS *see* public participation in GIS (PPGIS)
PPGISc (public participation in GIScience), 482
Pratt, Lee, ***386***
precision
  and accuracy compared, 140
  definitions, 140
predictor variables, 50, 101
Preparata, Franco P., ***334***
Presidential Executive Orders (USA), 459
  12906, ***454***, 455
pricing strategies, ***443***
prime meridian, 115
Princeton University (USA), 372
privacy issues, 31–2, 430–1
probability
  conceptions
    frequentist, 133
    subjectivist, 133
problem-solving
  GIS applications, 39–40
  objectives, 39
    multiple, 15
    tangible vs. intangible, 15
  science of, 13–16
  technology of, 16–24
process models, 178
process objects, 377
processes, 13–14
  ecological, 58, 59
  geographic, 364
  social, *13*

producer externalities, 418
producer's perspective, 138
professional accreditation, 431–3
programming languages, 165, ***168***, 225
  high-level, 159
  visual, 162
project evaluation, 201
project managers, 400–1
projections, 275
  azimuthal, 118
  classification, 118
  conic, 118, *119*
  and coordinates, 117–22
  cylindrical, 118, *119*
  planar, 118
  Plate Carrée, 119–20
  properties
    conformal, 118
    equal area, 118
  secant, 118, *119*
  tangent, 118
  unprojected, 119–20
  see also Lambert Conformal Conic projection; Mercator projection; Universal Transverse Mercator (UTM) system
properties
  inventories, 46
  tax assessment, 42–6, 297
  topological, 54
  valuations, 46
property databases, applications, 46
prototyping, 396
proximity analysis, 45
proximity effects, and spatial variation, 86
psychological distance, 50
public access, geographic information, ***454***, 465–7
public domain software, 158, 173–4
public good, 417, 482–3
Public Land Survey System (PLSS) (USA), 91, 114–15
  development, 115
  procedures, 115
public participation in GIS (PPGIS), ***150***, 309–12, ***451–3***, 482
  goals, 450
  and immersive interactions, 302–9
  research, 305
  use of term, 303
public participation in GIScience (PPGISc), 482
public policy, and geographic information, ***441***, ***453***
public sector organizations, and GIS, 448–50
public service, GIS applications, 42–6
publishing industry, 27
Putnam, Adam, NSDI hearings, ***457***
Pythagoras's theorem, 324
Python, 162

Qatar, 159
quadtree indexes, ***182***, ***183***, ***221***, 232–4, 235
  classification, 233
  linear, 233–4
  point, 233
  region, 233
quality, of geographic representation, 128
quality control, 409, ***410***
quality of life, 15
quantile breaks, 277–8
Quantitative Revolution, ***32***
quantum theory, ***77***
Quebec (Canada), ***386***
queen's case, 358, ***374***
Queensland (Australia), *265*
queries, 293, 320–2
  linkages, 322
  see also spatial queries
query access, 235
query languages, 219, 220
  see also Structured/Standard Query Language (SQL)
query optimizers, 220
query parsers, 220
query performance, 231
Quickbird, images, *462*

rainfall, Thiessen polygons, 333
rainfall risk, and malaria, *267*, 279
Rand Corporation, 465
randomization tests, 361
range, 336, 346
ranges (land unit), 115
raster cells, 139, 372, ***374–5***
  focal operations, 376
  global operations, 376
  local operations, 376
  zonal operations, 376
raster compression, 181, ***182–3***
raster data, 59, 74–5, 200
  buffering, 329
  storage, 181
  tiling, 75
  transformations, 376
  vs. vector data, 76, 136–7
  watermarks, 429
  see also pixels
raster data capture, 202–4
  via scanners, 205–6
raster data models, 136, 181–3, 196, 364, 391–3
  and vector data models compared, 179
raster overlays, 332–3
raster-based GIS, 173
rasterization, 207
ratio attributes, ***69***, 346
  classification, 277–9
  errors, 139–41
rationalization, 47
Ratzel, Friedrich, ***290***
RDBMS *see* relational database management system (RDBMS)
Reader's Digest, ***282***
reality
  and GIS data models, 178–9
  and representation, 128
  see also augmented reality; virtual reality
Reason, 421
Redlands, California (USA), 165, ***166***
Reed, Carl, ***243***
referential errors, 401
refinement, 80, *81*
Regional Research Laboratory (RRL) (UK), ***453***
regression analysis, 59, 101–3, 106, 148, ***351***
regression parameters, 102
Rehnquist, William Hubbs, ***327***

relational database management system (RDBMS), 186, *187*
definition, 219
extensions, 219–20
limitations, 219
relationships
geographic objects, ***191***, 192
rules, 192
spatial, 226–7
relative positioning errors, 401
relativity, theory of, 15, ***77***
remote sensing, 74–5, 200, 459
characteristics, *203*
development, 17–18
raster data capture, 202
resolution, 202, 204
sensors, 202
use of term, 202
see also aerial photography; satellites
representation
attributes, 273–9
floating-point, ***66***, 181
geographic phenomena, 136–7
in GIS, 63–83
laws, 65
occurrence, 64–5
real, ***66***
and reality, 128
spatial variation, 86
and uncertainty, 128
use of term, 64
and visualization in scientific computing, 302–5
see also digital representation; geographic representation
representative fraction, ***90***, 273, 364
definition, 76–8
see also scale
Request for Proposals (RFP), 394
residuals, 103, 322
resolution
remote sensing, 202, 204
and slope, 328
spectral, 202
see also spatial resolution; temporal resolution
response variables, 101
restructuring, 47
retailing
catchment areas, 94–5, ***150***
GIS applications, 48–51
strategies, 15
see also e-tailing
RFP (Request for Proposals), 394
Richardson, Lewis Fry, 106
biographical notes, ***106***
right-angled triangles, 328
Rio de Janeiro (Brazil), 462
rise over run, 328
risk
analysis, 394
classification, **443**
identification, 441–2
management, 444
terrorism, 467
RMSE *see* root mean square error (RMSE)
Road Management System for Europe (ROMANSE), ***297–8***
Robbins, Dawn, biographical notes, ***402***
Rogers, Everett, 41
ROMANSE (Road Management System for Europe), ***297–8***
Rome (Italy), 251, *252*
Rondonia (Brazil), *353*
rook's case, 358, 359
root mean square error (RMSE), 141–2, 143, 144, 145, 346, 347
definition, 140–1
rose diagrams, 281
Ross & Associates Environmental Consulting (USA), ***454***
routing problems, 355–8
algorithms, 355
Royal Air Force (UK), ***365***, ***386***
Royal Canadian Geographic Society, ***387***
Royal Geographical Society, ***387***
Royal Observatory, Greenwich (UK), 116, 117
RRL (Regional Research Laboratory) (UK), ***453***
R-tree indexes, ***221***, 234–5
rubberbanding, 185
rubber-sheeting, 209
Ruhr Valley (Germany), ***365***
rule sets, 14–15
run-length encoding, ***182***, ***183***
Russia, ***290***, 346
regional geography, *135*
Rutherford, Ernest, 380

Saaty, Thomas, 378–9
Safe Software, 165, 214
Sainsbury, Lord David, 413–14
samples, random and independent, 359, 360
sampling, 87, 360
random, 91
stratified, 139
see also spatial sampling
sampling designs, 103
sampling frames
definition, 90–1
periodicity, 91
sampling intervals, 89
San Andreas Fault, 104
San Diego, California (USA), *179*, 456
San Francisco (USA)
fires, 67
urbanization, ***451***
San Rafael, California (USA), 166, ***248***
SANET, ***334***
Santa Barbara, California (USA), 53, *257*, 366
Mission Canyon, ***367–9***
urban growth models, 374–6
Santa Maria Maggiore, Rome (Italy), 251, *252*
SAPs *see* spatially aware professionals (SAPs)
SARS (severe acute respiratory syndrome), 4
satellite images, 180, 200, 429, ***439***, 459
advantages, 202–4
and agricultural policy, 478
changes, 419
and global databases, 461–2
legal protection, 429
photogrammetry, 209
raster, *179*, *181*
raster data capture, 202
see also aerial photography
satellites, 25, 70, 74–5, 200
civilian, 17

Earth-orbiting, 202
geostationary, 75, 202
in Global Positioning System, 122–3
high-resolution, 461–2, 484
military, 17
remote sensing, 202
and traffic management, ***297***
see also Landsat; remote sensing
satisfaction, maximizing, 15
scalar fields, 72
scale, 4, 86, 364
and aggregation, 147–8
concept of, 87
definition, 76–8
effects, ***106***
and geographic phenomena, 135–6
maps, 76–8, 80, ***90***, 269, 272, 273
and spatial autocorrelation, 87–90
use of term, ***90***
see also representative fraction
scale dependence, 89
scanners
operation, 205
in raster data capture, 205–6
spatial resolution, 205
scatterplots, 102, 106, 322
Moran, ***351***
scenarios, 56–7
planning, *367*
what-if, 366
Schell, David, biographical notes, ***243–4***
Schindler Elevator Corporation, ***356***
scholarly journals, ***27***, 28
schools
and GIS Day, ***449***
GIS in, ***432–3***, 481–2
Schürer, Kevin, ***131–2***
biographical notes, ***291–2***
sciences
GIS applications, 39–40
see also GIScience
Scotland, 70, 71, ***307***
maps, *293*
wind turbines, ***451–2***
scripts, 24, 376–7
libraries, 377
SDI *see* spatial data infrastructure (SDI)
SDSS *see* spatial decision support systems (SDSS)
SDTS files, 213
search engines, 250
Search for Extraterrestrial Intelligence (SETI), 244
Sears, Roebuck and Company, ***357***
Seattle (USA), *98*, *99*, ***454***
GIS service centers, ***453***
secant projections, 118, *119*
security issues, 219, 387
geographic information, 465–7
see also homeland security
self-similarity, 90, 104
Sellafield (UK), 319
semivariograms, 336, 337
anisotropic, 336
nuggets, 336
sensitivity analysis, 381
sensor webs, 475–6
sensors, 202
characteristics, *203*
Seoul (South Korea), ***388***
Seoul Metropolitan City (South Korea), Land Information Service System, ***389***
September 11 2001 (terrorist attack), 4, ***5–7***, *70*, 430–1, 464–5, 470
and geospatial information withdrawal, ***466***
see also terrorism
servers, 23
GIS, 475
GIS software, 164, 167–72
service planning, GIS applications, 46–51
services-oriented architectures, 475
SETI (Search for Extraterrestrial Intelligence), 244
settlement expansion, *353*
severe acute respiratory syndrome (SARS), 4
sex offenders, 476–7
Shamos, M. I., ***334***
Shanghai (China), *275*
Shapefile, 213
shapes, 326–7
proportional, 277
symbols, 274–5
shared editing, 185
shares, 406
shareware, 158
Shea, K. S., 80, *81*
Shiffer, Mike, 303
biographical notes, ***306***
shortest path analysis, 54, 355
Shuttle Radar Topography Mission (NASA), *73*, 461
Sibuyan Island (Philippines), deforestation, 55–9
Sierpinski carpets, *90*
SIG (Special Interest Group), ***213***
sills, 336
similarity, between neighbors, ***100***
simplification, 80, *81*
Singapore, 461
Singleton, Alex, ***483***
Sioux Falls, South Dakota (USA), 247
skills, personnel, 431–4
slope, 327–9
calculation, ***328***
definitions, 328, *329*
and resolution, 328
small circles, 117
small office/home office (SOHO), 41
Smallworld, 48, 167, 406
Smart Toolbox Mapping software, ***357***
smartphones, 172–3
Smith, Andy, ***310***
Smithsonian Institute (USA), ***282***
smoothing, 80, *81*
snapping, 185
SNIG, ***254***
Snow, John, cholera studies, 317–19, 347, ***473***
Sobel, D., 122
social interactionism, 410
software, 14, 16
distribution models, 158
public domain, 158, 173–4
reduced costs, 18
source code, 158

software (*continued*)
see also commercial off-the-shelf (COTS) software; GIS software
software components, 163
software data models, 162–3
software industry, 24–5
Soho (London), 317–19
SOHO (small office/home office), 41
soil erosion, 15, 24
soil maps, 132, ***133***, 279, 350, *352*
soils, ***371***
classification, ***133***, *134*
Solar System, 65
sound, coding standards, 66
source code, 158
South Dakota (USA), 247
South Korea, land management, ***388–9***
Southampton (UK), ***297–8***
LiDAR images, *308*
Southern United States
population density, *269*
topographic maps, *268*
Soviet Union (former), 346
see also Russia
Space Imaging Inc., 25, 247, 417, 429
activities, 461–2
SpaceStat, ***351***
Spain, ***106***
Spartan Air Services, ***386***
spatial, use of term, 8–11, 218, 316
Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, ***147***
spatial analysis, 31, 33, 144, 342–3, ***351***
applications, 317–19
deductive, 319
definitions, 316–17
and design, 346, 352
distortions, 144
future trends, 476
geometries, 227, *228*
inductive, 319
methods, 316
modeling, 319–20
multivariate, 148
normative, 319, 346, 352
overview, 316–17
prescriptive, 346
types of, 319–20
vs. map production, 316
vs. spatial modeling, 366–9
spatial autocorrelation, 129
concept of, 87
and distance effects, 95–100
measurement, 95
negative, 88
positive, 88, 143–4
and scale, 87–90
zero, 88
spatial cognition, ***30***
spatial data infrastructure (SDI), 212, 475
projects, ***453***
Spatial Datablade, 220, ***221***, 235
spatial decision support systems (SDSS), 16
user interfaces, 352–3
spatial distributions, as dasymetric maps, 301–2, *303*
spatial econometrics, ***351***
Spatial Extender, 220, ***221***
spatial heterogeneity, 87, 361
and inferential testing, 360
spatial information science, use of term, 28
spatial interaction models, 15, 50, 354–5
spatial interactions, 50, ***96***
spatial interpolation, 54, 93–4, 333–9
applications, 333, ***350***
processes, 333
spatial invariance, ***32***
spatial modeling, 363–82
future trends, 476
motivation, 366
multicriteria methods, 378–9
overview, 364–9
technology for, 376–8
use of term, 316
vs. spatial analysis, 366–9
spatial models
accuracy, 379–81
aggregate, 371–2
calibration, 380
cartographic, 376
cataloging, 377–8
cellular, 372–6
coupling, 377
and data models compared, 364
implementation, 376–7
individual, 371–2
requirements, 364
sharing, 377–8
static, 369–70
testing, 379–81
types of, 369–76
and uncertainty, 366, 381
validity, 379–81
spatial objects
artificial, ***88***
mapping, 275, **277**
natural, ***88***
types of, ***88***
spatial patterning, 57
spatial queries
and geovisualization, 293–7
and Internet mapping, 295–7
spatial relationships, testing, 226–7
spatial resolution, 70, 202, 364–6
and georeferences, 110
scanners, 205
spatial sampling, 90–3
designs, 91–2
spatial structures, 50
spatial support, ***350***
spatial variation
controlled vs. uncontrolled, 86, 87
issues, 105–6
and proximity effects, 86
representation, 86
SpatialFX, 172
spatially aware professionals (SAPs), 47, 121, 167, 250, 295
roles, 51
use of term, 24
see also GIS profession
spatially extensive variables, ***96***

spatially intensive variables, ***96***
SPCs (State Plane Coordinates), 121–2, 324
Special Interest Group (SIG), ***213***
spectral resolution, 202
spectral signatures, 202, *204*
spheroid, use of term, 116
splines, 183
split rules, 192
SPOT *see* Système Probatoire d'Observation de la Terre (SPOT)
spot heights, ***88***
spurious polygon problem, 331–2
SQL *see* Structured/Standard Query Language (SQL)
SQL Server, 168, ***170***, 220
staff *see* personnel
stakeholders, 352, 378, 396, 456
standard deviation, 145, 278–9
  determination, 346
standards, 142, 214, 226, 245, 455
  coding, 66
  distributed GIS, 243–4
  GIS, 159
  global GIS partnerships for, 462–3
  metadata, 455
    collection-level, 247
    object-level, 245
State College, Pennsylvania (USA), *142*
State Plane Coordinates (SPCs), 121–2, 324
State University of New York, Buffalo (USA), ***32***, ***77***
state variables, 417
static modeling, 320
static models, 369–70
statistical inference, 91, 359
Stella, 376
Storm Water Management Model (SWMM), 377
straight lines, 324
strategic decisions, 7, 55
street centerlines, 284–5, 480
  databases, 53, ***125***
street layouts, 54
street names, 111–12
streets
  addresses, ***125***, 377–8
  networks, 187–8
  shortest path analysis, 54
  see also postal addresses
Strobl, Josef, biographical notes, ***434***
Structured/Standard Query Language (SQL), 227, 229, 230, 231, 320
  applications, 219
  principles, 225–6
  query parsers, 220
  statements, 225
  tables, 322
studies, qualitative vs. quantitative, 56–7
subjectivism, 133
Sudan, *266–7*, 476
sudden infant death syndrome, 322, *323*
Sugarbaker, Larry, 398–9
Summer Time, 110
Sun Microsystems, 162, 163, 244
  valuation, ***415***
  see also Java
surnames, distribution, ***8–10***, 39, ***131–2***, ***292***
surveillance, ***124***, 430
  issues, 32
survey monuments, 122
surveying, 204, 235
  field data, 242
surveys, data collection, *243*
Sweeney, Jack, ***213***
SWMM (Storm Water Management Model), 377
Sybase, ***170***
symbolization, maps, 273–9
symbols
  multivariate, 287
  shape of, 274–5
Syracuse University (USA), ***124***
system administrators, 399
Système Probatoire d'Observation de la Terre (SPOT), 200
  images, *179*
  overview, 202

tables, 322
  in database management systems, 222–5
  geographic information as, 71, **72**
Tacoma, Washington (USA), 477
tactical decisions, 7, 55
tagging, use of term, 110
takeovers, 406
tangent projections, 118
tangents, 328
Tapestry, 25
Tauranga Harbour (New Zealand), *275*
tax assessment, 223, 297
  GIS applications, 42–6
  maps, *281–2*
tax assessors, roles, 44
Tax Assessor's Office (USA), 44
technological determinism, 410
technology
  and GIS, 475–6
  see also information technology (IT)
Tele Atlas, 25, 53, ***213***, 406, 461
telegraph cables, transatlantic, ***421***
telemetry, ***407***
temperature, scales, ***69***
temporal autocorrelation, 86–7
temporal resolution, 202
  definition, 366
TERRA satellite, 70
terrain nominal, 80
Terraserver (Microsoft), 247
terrorism, 4, ***5–7***, *70*, 430–1, 464–5
  coping strategies, 467–70
  and GIS, 465, 470
  global, 448
  mitigation, 467–8
  preparedness, 467
  recovery, 470
  responses, 468
  risk assessment, 467
  simulated attacks, *468*, *469*
  vs. culture, 484
  see also September 11 2001 (terrorist attack)
Tesco
  Express store, *48–9*
  GIS applications, 48–51
  growth, 48

tessellation, 75
Texas (USA), 75, 121, ***411***, ***432–3***, 456
  evacuation routes, *469*
  projections, *122*
Texas A&M University (USA), ***433***
Thai Parliament, ***409***
Thailand, ***409***
The National Map (TNM) (USA), 437, ***439***
thematic maps, 272
  series, 281
theodolites, 204
there is no alternative (TINA), 410
thermal imaging, ***407***
Thiessen, Alfred H., 333
Thiessen polygons, 333, ***334***
tidal movements, *275*
TIGER *see* Topologically Integrated Geographic Encoding and Referencing (TIGER)
time, ***88***
  measurement, 110
time series, 86–7
time zones, 110
TINA (there is no alternative), 410
TIN *see* triangulated irregular network (TIN)
Titanic, HMS, discovery, ***101***
TMS Partnership, ***47***
TNM (The National Map) (USA), 437, ***439***
Tobler, Waldo, 277
Tobler's First Law of Geography, 92, 104, 106, 129, 333, 336
  applications, 45
  definition, 65
  and distance decay effects, 49
  interpretation, 89, 93
  violation, 87
Todd, Richard, ***365***
TOID (Topographic Identifier), ***437***
Tokyo (Japan), fires, 67
tolerance, 332
Tomlin, Dana, 376
Tomlinson, Roger, 398
  biographical notes, ***386–7***
Tomlinson Associates Limited, ***387***
Tomlinson model, ***387–8***, 390
Topographic Identifier (TOID), ***437***
topographic maps, *264*, *268*, 273, ***439***
  changes, 419
  production, 283–4
  series, 281
  street centerlines, 284–5
topologic features, in vector data models, 184–6, *187*
topological dimension, ***88***
topological errors, 401
topological properties, 54
Topologically Integrated Geographic Encoding and Referencing (TIGER), 53, 54
  database, 149, ***213***
topology, 184
  creation, 229–31
  networks, 192
  normalized model, 229–30, 231
  physical model, 229, 230–1
tornadoes, *78*
Toronto (Canada), ***386***
Torrens, Paul, 55
Tosta, Nancy, biographical notes, ***454***
total stations, 204, *205*, 235
town centeredness, degree of, ***152***
town centers, definition, and uncertainty, ***150–2***
townships, 115
toxic chemicals, 52
tracing, 185
trade directories, ***436***
trade marks, 427
traffic management, geovisualization, ***297–8***
traffic simulation, ***367–9***
Traffic and Travel Information Centre (TTIC) (UK), ***297***
transactional databases, 7–8
transactions, 220, 235–6
  long, 236
  short, 235, 236
TransCAD, 358
transformations, 50, 293, 320, 329–39, ***369***
  attributes, 273–9
  and geovisualization, 297–302
  raster data, 376
transit authorities, 51
transits (surveying), 204
translation
  data access, 213, *214*
  semantic, 214
  syntactic, 214
transportation
  cartograms, 299, *312*
  emergency incidents, 51
  geovisualization in, ***306***
  GIS applications, 37, *38*, 51–5
  linear referencing, 114
  map applications, 285
  routing problems, 355
  traffic management, ***297–8***
Tranter, Penny, biographical notes, ***307–8***
travel speed analysis, 329, *330*
traveling-salesman problem (TSP), 353
  algorithms, 355–8
Treaty of Versailles, ***290***
trees, point patterns, *347*, 348
triangulated irregular network (TIN), 76, *77*, 305, 333
  data models, 189–90
  topology, 189, *190*
triangulation, 210
trip planners, 254
tropical rainforests, deforestation, 55
TSP *see* traveling-salesman problem (TSP)
tsunami, 485, 486
TTIC (Traffic and Travel Information Centre) (UK), ***297***
Turing, Alan, 484
Turkey, ***410***
Tyneside (UK), 319

UCAS (Universities Central Admissions Service), ***47***, ***483***
UCLA (University of California, Los Angeles) (USA), 251, *252*
UCSB *see* University of California, Santa Barbara (UCSB) (USA)
UDDI (Universal Description, Discovery, and Integration), 258
UDP (Urban Data Processing), ***213***
Uganda, 477
UML (Unified Modeling Language), 193, *194*
uncertainty, 50, 70, 86, 127–53
  concept of, 129–36, 295, *296*
  coping strategies, 401

external validation, 144, 148–52
and history, ***131–2***
internal validation, 144–52
measurement, 129
and representation, 128
rules, 152–3
sources, 129
and spatial models, 366, 381
statistical models, 137–41
and town center definition, ***150–2***
units, 129–30
use of term, 128–9
Undercover (game), ***254–6***
unemployment rates, dasymetric mapping, 301–2, *304*
Unified Modeling Language (UML), 193, *194*
UNIGIS, 431, ***434***
United Kingdom (UK)
consultancies, ***411***
functional zones, *136*
local government, 450
MasterMap database, 218
national framework, ***437–8***
old age pensioners, *98–9*
projections, 121–2
Wales, ***410***
wealthy areas, 300, *302*
see also Scotland
United Kingdom Data Archive (UKDA), ***291–2***
United Nations (UN), ***387***, 448, 459
roles, 482
United Nations General Assembly, 462
United Nations Institute for Training and Research, ***380***
United Nations Office on Drugs and Crime, 478
United Nations World Summit on Sustainable Development (2002), 462
United States of America (USA)
business issues, 406
critical infrastructure sectors, **466**
homeland security, 465
housing value patterns, *349*
hurricanes, 52, ***414***
key facilities, *467*
land resource management, 115
national framework, ***439***
projections, *120*, 121
public sector organizations, 448–50
zip codes, 113, *114*
see also Southern United States
United States Forest Service, 14–15, ***195***
United States Geological Survey (USGS), 17, 53, 152, ***387***, 417, 450
contracts, 438
digital elevation model, 141, *142*
geographic data, 247
and government initiatives, 456–8
GTOPO30 DEM, 327
homeland security studies, 465
income, 25
map scales, 80
maps, 200, ***213***
series, 281
topographic, 78, 80
national framework, ***439***
United States Justice Department, 420
United States Supreme Court, ***327***, 428
units
natural, 129–30, 135
usable, 135
Universal Description, Discovery, and Integration (UDDI), 258
Universal Polar Stereographic (UPS) system, 121
Universal Soil Loss Equation (USLE), 15, 369–70, 377
Universal Transverse Mercator (UTM) system, 111, 225, 300, 320, *321*, 324
applications, 120
projections, 120–1
Universities Central Admissions Service (UCAS), ***47***, ***483***
University of Aberdeen (UK), ***32***
University of Amsterdam, ***147***
University of California, Los Angeles (UCLA) (USA), Cultural Virtual Reality Laboratory, 251, *252*
University of California, Santa Barbara (UCSB) (USA), ***30***, ***101***, 252, ***351***, ***367***
GIS curriculum, 431
Map and Imagery Laboratory, ***248***
urban growth studies, 374–6
University College London (UK), ***9***, ***131***, ***150***, ***310–11***, ***373***, ***386***
University Consortium for Geographic Information Science (USA), 28, 456
research agenda, ***29***, 31
University of East Anglia (UK), ***307***
University of Essex (UK), ***291***
University of Florida (USA), ***32***
University of Illinois (USA), ***306***, ***351***
University of Kansas (USA), ***294***
University of Kentucky (USA), ***409***
University of Leeds (UK), 319
University of Munich (Germany), ***290***
University of New Mexico (USA), ***195***
University of Newcastle (UK), ***32***
University of North Carolina (USA), ***243***
University of Oklahoma (USA), ***77***, ***78***
University of Pennsylvania (USA), ***334***
University of Salzburg (Austria), ***434***
University of Tokyo (Japan), ***334***
University of Wageningen (Netherlands), 56
University of Wisconsin-Madison (USA), 134
UNIX, 164, ***170***, 320
development, ***243***
Upper Lake McDonald basin, Montana (USA), *134*
UPS (Universal Polar Stereographic) system, 121
Urban Data Processing (UDP), ***213***
urban environments
geovisualization, 305
three-dimensional representations, 305
Urban and Regional Information Systems Association (URISA) (USA), ***213***
certification program, 431–2
urban traffic control (UTC) systems, ***297***
Urbana-Champaign, Illinois (USA), ***306***, ***351***
urbanization, 55, ***451***
American Mid-West, *57*
Europe, *56*
growth models, 374–6, 380
URISA *see* Urban and Regional Information Systems Association (URISA) (USA)
USGS *see* United States Geological Survey (USGS)
USLE (Universal Soil Loss Equation), 15, 369–70, 377
USSR (former), 346
see also Russia

Utah (USA), 456
UTC (urban traffic control) systems, ***297***
utilities, 16
  geographic objects, 192
  GIS applications, 36–7, 41, *181*
  mains records, *36*
  map series, 283
  network data models, 188
  pipeline schematics, *285*
  see also water distribution systems
UTM system *see* Universal Transverse Mercator (UTM) system
utopianism, 410
Utrecht University (Netherlands), ***147***, ***270***, 376
Uttaranchal (India), ***481***

vagueness, 130
  use of term, 128–9
validity, spatial models, 379–81
Valuation Office Agency (UK), 44
value-added approach, ***394***
value-added resellers (VARs), 39
Varenius Project, 28, *29*
variables
  and continuous fields, 72
  dependent, 101
  geographic, 88–9
  graphic, 275–7
  independent, 101
  omission, 128
  predictor, 50, 101
  response, 101
  spatially extensive, ***96***
  spatially intensive, ***96***
  state, 417
variance, 346
Varignon Frame, ***345***, 346
variograms, 336
  isotropic, 336
  see also semivariograms
VARs (value-added resellers), 39
VBA (Visual Basic for Applications), ***170***, 377
vector data, 74, 75–6, 200
  buffering, 329
  vs. raster data, 76, 136–7
vector data capture, 204
  coordinate geometry data entry, 210
  manual digitizing, 206–7
  measurement error, 208–9
  photogrammetry, 209–10
  secondary, 206–10
vector data models, 183–90, 391–3
  data validation, 184–5
  editing, 185
  features, 183–6
    modeling, 185
    topological, 184–6, *187*
  query optimization, 185–6
  and raster models compared, 179
vector fields, 72
vector overlays, 332
vector polygon overlay algorithms, ***166***
vector-based GIS, 173
vectorization, 229
  batch, 207, *208*
  heads-up, 207–8
vegetation, 204
  maps, *139*
  representation, 80
vehicles
  deepsea, ***101***
  tracking, 51–2
Veldkamp, Tom, 56
Ventura County, California (USA), ***402***
Verburg, Peter, 56, 59
Vermeer, Johannes, *The Geographer,* 251, *252*
Vermont (USA), ***133***
versioning, 236–7
  optimistic, 236
vertices, 75
Vexcel, ***310***
Virtual Field Course, 251
Virtual London, 251, ***310–11***
virtual reality, 64, 251–2, 303
  and geovisualization, 305–9
  immersive, 305
  use of term, 251
ViSC *see* visualization in scientific computing (ViSC)
Visual Basic, 159, 162, *163*, 165, ***168***
Visual Basic for Applications (VBA), ***170***, 377
Visual C++, 165
visualization
  future trends, 476
  see also geovisualization
visualization in scientific computing (ViSC), 292
  field of study, 302–3
  and representation, 302–5
VMap, 460
Vodafone, ***255***
volume, ***88***
Voronoi polygons, 333, ***334***
VPF files, 213

Wageningen University (Netherlands), ***147***
Wales, ***410***
walk-throughs, three-dimensional, 273
Wallis, Barnes, ***365***
WANs (wide area networks), 161, 164
Washington Post, The, ***282***
Washington State (USA), *181*, 477
water, modeling, 371–2
water distribution systems
  geographic database versioning, 236–7
  GIS applications, *181*
  networks, 187
  object data models, 192–4
water supplies, and cholera, 317–19, 347
water-facility objects, 192–3, *194*
watermarks, 429
wavelet compression, ***182***, ***183***
wayfinding, websites, 257–8
wealthy, distribution maps, 300, *302*
weather forecasting, ***307–8***
Web *see* World Wide Web (WWW)
Web Coverage Services, 244
Web Feature Service, 244
Web Map Service, 244
Web mapping, ***243***
Web Services Definition Language (WSDL), 258
Webber, Richard, 46
websites, ***26***

weeding, 80–2
weighting, 87, 93, ***100***, 349–50, 378–9
  inverse-distance, 333–6
  see also inverse-distance weighting (IDW)
weights matrix, ***100***
West Africa, locust infestations, 448
West Nile virus, 477, *478*
West Parley (UK), 50
wetlands, 14, *15*, 132
WGS84 (World Geodetic System of 1984), 116
what-if scenarios, 366
wheat, yields, 135, **136**
Whistler, British Columbia (Canada), *264*
wide area networks (WANs), 161, 164
WiFi, 250, 254, 257
WIMP interface *see* windows, icons, menus, and pointers (WIMP) interface
wind turbines, ***451–2***
windows, icons, menus, and pointers (WIMP) interface, 295
  Macintosh, *296*
Windows (Microsoft) *see* Microsoft Windows
Windows NT, ***170***
Windows Terminal Server, 161
WIPO (World Intellectual Property Organization), 462–3
wireless communications, 250, ***357***, 475
  limitations, 257
wisdom, use of term, 13
woodlands, 130
Worboys, Mike
  biographical notes, ***238***
  GIS: A Computing Perspective (2004), ***238***
word processors, 40
World Bank, ***387***
World Factbook, ***416***
World Geodetic System of 1984 (WGS84), 116
World Intellectual Property Organization (WIPO), 462–3
World Trade Center, *70*, *464*
World Trade Organization (WTO), 427, 428–9, 463
World War I (1914–18), ***290***
World War II (1939–45), ***290***
  bombing missions, ***365***
World Wide Web (WWW), ***244***
  browsers, 21, 23
  development, 18
  geographic information, 420–2
  map access, 21
  search engines, 250
  user empowerment, 482
  see also Internet
Wright, Dawn, biographical notes, ***101***
WSDL (Web Services Definition Language), 258
WTO (World Trade Organization), 427, 428–9, 463
WYSIWYG environments, 235

XML (extensible markup language), 163, 244, 245
X-Windows, ***170***

Yale University (USA), ***213***
Yangtse River (China), 189, *190*
YDreams, ***254–6***
Yell, 257, 258
Yellow Pages (UK), 25, *26*
Yuan, May, biographical notes, ***77–8***
Yule, George Udny, 135, **136**

Zhu, A-Xing, 133–4
zip codes, 17, 48, 113, *114*
zones
  aggregation, 147–8, 326
  boundary definition, 130
  defining, 135
  functional, 135, *136*
  homogeneous, 135
  uniform, 135, 137
zoning, effects, 149–52

Applied GIS and Spatial Analysis

Editors

JOHN STILLWELL
GRAHAM CLARKE

School of Geography, University of Leeds

Copyright © 2004 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com
or www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-84409-4

Typeset in 10/12pt Times by SNP Best-set Typesetter Ltd., Hong Kong
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

List of Contributors ix

Acknowledgements xiii

1 Introduction 1
Graham Clarke and John Stillwell

PART 1: GEOBUSINESS 9

2 Retail Applications of Spatial Modelling 11
Ken Jones and Tony Hernandez

3 Using Spatial Models to Solve Difficult Retail Location Problems 35
Mark Birkin, Graham Clarke, Martin Clarke and Richard Culf

4 Location-based Services for WAP Phone Users in a Shopping Centre 55
António Câmara and António Eduardo Dias

5 Mass Appraisal and Noise: the use of Lifestyle Segmentation Profiles to Define Neighbourhoods for Hedonic Housing Price Mass Appraisal Models 71
Steve Laposa and Grant Thrall

PART 2: SOCIAL DEPRIVATION 87

6 Targeting Clusters of Deprivation within Cities 89
Richard Harris and Paul Longley

7 Assessing Deprivation in English Inner City Areas: Making the Case for EC Funding for Leeds City 111
Paul Boyle and Seraphim Alvanides

8 GIS for Joined-up Government: the Case Study of the Sheffield Children Service Plan 137
Massimo Craglia and Paola Signoretta

9 The Application of New Spatial Statistical Methods to the Detection of Geographical Patterns of Crime 151
Peter Rogerson

PART 3: TRANSPORT AND LOCATION 169

10 Modelling and Assessment of Demand-Responsive Passenger Transport Services 171
Mark E.T. Horn

11 The South and West Yorkshire Strategic Land-use/Transportation Model 195
David Simmonds and Andy Skinner

12 The Relocation of Ambulance Facilities in Central Rotterdam 215
Stan Geertman, Tom de Jong, Coen Wessels and Jan Bleeker

13 A Probability-based GIS Model for Identifying Focal Species Linkage Zones across Highways in the Canadian Rocky Mountains 233
Shelley M. Alexander, Nigel M. Waters and Paul C. Paquet

PART 4: NATIONAL SPATIAL PLANNING 257

14 Modelling Migration for Policy Analysis 259
Phil Rees, A.Stewart Fotheringham and Tony Champion

15 Modelling Regional Economic Growth by Means of Carrying Capacity 297
Leo van Wissen

16 Planning a Network of Sites for the Delivery of a New Public Service in England and Wales 315
Mike Coombes and Simon Raybould

17 New Methods for Assessing Service Provision in Rural England 335
Martin Frost and John Shepherd

18 Forecasting River Stage with Artificial Neural Networks 353
Pauline Kneale and Linda See

CONCLUSION 375

19 Undertaking Applied GIS and Spatial Analysis Research in an Academic Context 377
Robin Flowerdew and John Stillwell

Author Index 397

Subject Index 403

List of Contributors

Shelley Alexander, Department of Geography, Faculty of Social Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada

Seraphim Alvanides, Department of Geography, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne NE1 7RU, United Kingdom

Mark Birkin, School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

Jan Bleeker, GGD Rotterdam e.o., PO Box 70032, 3000 LP Rotterdam, The Netherlands

Paul Boyle, School of Geography and Geosciences, University of St Andrews, St Andrews, Fife KY16 9AL, United Kingdom

António Câmara, New University of Lisbon, 2825 Monte de Caparica, Lisbon, Portugal (and YDreams, SA, Madan Park, Caparica, Lisbon, Portugal)

Tony Champion, Department of Geography, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne NE1 7RU, United Kingdom

Graham Clarke, School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

Martin Clarke, GMAP Limited, 1 Park Lane, Leeds LS3 1EP, United Kingdom

Mike Coombes, Centre for Urban and Regional Development Studies (CURDS), University of Newcastle-upon-Tyne, Newcastle NE1 7RU, United Kingdom

Massimo Craglia, Sheffield Centre for Geographic Information and Spatial Analysis, University of Sheffield, Sheffield S10 2TN, United Kingdom

Richard Culf, GMAP Limited, 1 Park Lane, Leeds LS3 1EP, United Kingdom

António Eduardo Dias, University of Evora, Evora, Portugal (and YDreams, SA, Madan Park, Caparica, Portugal)

Robin Flowerdew, School of Geography and Geosciences, University of St Andrews, St Andrews, Fife KY16 9AL, United Kingdom

A. Stewart Fotheringham, Department of Geography, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne NE1 7RU, United Kingdom

Martin Frost, South East Regional Research Laboratory (SERRL), School of Geography, Birkbeck College, University of London, 7–15 Gresse Street, London W1T 1LL, United Kingdom

Stan Geertman, URU and Nexpri, Faculty of Geographical Sciences, Utrecht University, PO Box 80.115, 3508 TC Utrecht, The Netherlands

Richard Harris, School of Geography and South East Regional Research Laboratory (SERRL), Birkbeck College, University of London, 7–15 Gresse Street, London W1T 1LL, United Kingdom

Tony Hernandez, Centre for the Study of Commercial Activity (CSCA), 350 Victoria Street, Ryerson University, Toronto, Ontario M5B 2K3, Canada

Mark E.T. Horn, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Mathematical and Information Sciences, GPO Box 664, Canberra A.C.T. 2001, Australia

Ken Jones, Centre for the Study of Commercial Activity (CSCA), 350 Victoria Street, Ryerson University, Toronto, Ontario M5B 2K3, Canada

Tom de Jong, URU, Faculty of Geographical Sciences, Utrecht University, PO Box 80.115, 3508 TC Utrecht, The Netherlands

Pauline Kneale, School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

Steven Laposa, PricewaterhouseCoopers (PwC), 1670 Broadway, Suite 1000, Denver, CO 80202, USA

Paul Longley, Department of Geography and Centre for Advanced Spatial Analysis (CASA), University College London, 1–19 Torrington Place, London WC1E 6BT, United Kingdom

Paul Paquet, Faculty of Environmental Design, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada

Simon Raybould, Centre for Urban and Regional Development Studies (CURDS), University of Newcastle-upon-Tyne, Newcastle NE1 7RU, United Kingdom

Philip Rees, School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

Peter Rogerson, Department of Geography, University of Buffalo, Buffalo NY 14261, USA

Linda See, School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

John Shepherd, South East Regional Research Laboratory (SERRL), School of Geography, Birkbeck College, University of London, 7–15 Gresse Street, London W1T 1LL, United Kingdom

Paola Signoretta, Sheffield Centre for Geographic Information and Spatial Analysis, University of Sheffield, Sheffield S10 2TN, United Kingdom

David Simmonds, David Simmonds Consultancy, Suite 23, Miller's Yard, Mill Lane, Cambridge CB2 1RQ, United Kingdom

Andy Skinner, MVA, 26th Floor, Sunley Tower, Manchester M1 4BT, United Kingdom

John Stillwell, School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom

Grant Thrall, PricewaterhouseCoopers (PwC), 1670 Broadway, Suite 1000, Denver, CO 80202, USA (and University of Florida, Gainesville, FL 32611, USA)

Nigel Waters, Department of Geography, Faculty of Social Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada

Coen Wessels, Nexpri, Utrecht University, PO Box 80.115, 3508 TC Utrecht, The Netherlands

Leo van Wissen, Faculty of Spatial Sciences, University Groningen, PO Box 800, NL-9700 AV Groningen, The Netherlands, and Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, The Netherlands

Acknowledgements

Various people have helped in the preparation of this book, not least Alison Manson in the Graphics Unit in the School of Geography at the University of Leeds who has drawn and improved many of the figures, and Lyn Roberts at Wiley who provided support and encouragement to us as editors throughout the duration of the project. However, no collection like this would be possible without the contributions of the authors themselves and we are very grateful for the efforts that all our collaborators have made in producing their chapters in the first instance and in responding to our editorial suggestions.

1
Introduction

Graham Clarke and John Stillwell

Abstract

Applied work in geographical information systems (GIS) and spatial analysis has been a persistent component of research activity in geography and regional science for several decades. This introductory chapter of this book establishes the background to the contemporary flurry of applied activity, explains the structure of the book and introduces the contents of the chapters that constitute the various parts into which the book has been divided.

1.1 Background

The history of applied spatial modelling has been a chequered one. If we ignore the elements of simple numerical reporting in the early commercial geographies of the nineteenth century, then we can probably trace the origins of applied quantitative spatial analysis to the fields of transportation modelling, especially at the University of Pennsylvania, in the early 1960s (Herbert and Stevens, 1960; Harris, 1962). Mathematical geography became prevalent in planning circles both in the USA and in Europe from this time onwards. The early 1960s also witnessed the emergence of statistical geography: the search for patterns or similarities in spatial data sets and the testing for significance, with the aim of providing universal spatial theories and laws (or at least theories that worked well in the real world). Although this was often very empirical in nature, it seldom extended beyond the academic geography community. There are excellent summaries in Haggett (1965) and Haggett *et al*. (1977).

Foot (1981) has reviewed a number of the more well-known applications of large-scale models in the United Kingdom (UK) during the 1960s. The Portbury Dock

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

study in 1964 was one example of an application of mathematical modelling to estimate the future flow of exports through UK ports and hence the evaluation of the need for a new major dock facility at Portbury, in South West England. The Haydock shopping study (1966) was another classic example, incorporating a spatial interaction model of retail expenditure in North West England to estimate the potential revenues for a new regional shopping centre located roughly half-way between Liverpool and Manchester (a concept ahead of its time perhaps!). Another major study was the analysis carried out to quantify the impact of a third London airport at various locations in the South East of England. Paralleling these studies, were a large number of comprehensive land-use/transportation models built as part of local authority strategic planning. Batty (1989, 1994) has reviewed these applications in detail.

The fierce critiques of urban modelling in the 1970s (Lee, 1973; Sayer, 1976) brought the first era of applied modelling to an end. Lee believed the various large-scale models to be generally too complex, mechanical and ultimately unworkable. Sayer's critique was based on the role of agents in modelling – that is, that urban dynamics were as much the result of key decisions made by urban gatekeepers as they were due to land-use/transportation changes. In 1976, models were banned from being used in many UK retail planning enquiries because of the conflicting evidence they supposedly produced. More generally in land-use modelling, May (1991) argues that the decline in planning applications using models resulted from model-builders being too keen to present blueprints rather than alternative options. Wilson (1997) queries whether it was simply easier for planners to face the very difficult political decisions associated with cities 'in a haze rather than with maximum clarity' (Wilson, 1997, p. 18). Academics, meanwhile, retreated back to theory in an attempt to address the concerns expressed by humanist and Marxist geographers. For example, the 1970s and the early 1980s witnessed a research agenda dominated by the desire to incorporate dynamics into urban models (Allen and Sanglier, 1979; Wilson, 1981; Dendrinos, 1985). As Wilson (1997) reflects: 'While it can be argued that the full breadth of the attack was misguided – Marxian economists, for example, had no hesitation in using mathematical models – the analysis does point to the need for an understanding of deeper underlying structures that might be at the core of urban evolution' (Wilson, 1997, p. 10).

For much of mainstream academic geography, however, quantitative methods had become irrelevant and indeed counterproductive as the focus switched from serving the interests of planners to looking for grand theories of social change. The cause of quantitative geography was not helped by the desertion of some of its key pioneers (especially in the UK) such as David Harvey and Doreen Massey. The UK story is mirrored in many other parts of Europe and the USA. The full story of the retreat from quantitative geography is well told by Johnston (1996).

However, some remained loyal to the cause. A few geography departments around the world remained strong quantitatively, whilst the growth of the discipline of regional science helped the surviving quantitative geographers find allies amongst regional economists and planners (Plane, 1994). Slowly, through the 1980s and into the 1990s, quantitative geography began to claw its way back into the discipline (although it would not, to date, seriously challenge the stranglehold of 'critical human geography'). The reasons for this are numerous and a fuller account is

provided by Clarke and Wilson (1987). However, it is clear that there were two fundamental drivers of change. The first was the focus anew on applications. Statistical and mathematical geography became much more focused on how it could solve real location problems in fields such as health, education, retailing, transport and deprivation analysis. The pioneers here were undoubtedly the quantitative geographers at the University of Leeds who began to find commercial applications for their models in both the public and private sector, leading to the formation of GMAP in the late 1980s (see Chapter 3). The second driver of change was the developments occurring in the fields of geographical information systems (GIS) and geodemographics. Interestingly, both of these originated outside the discipline of geography. However, as they became well known in planning and business circles, the quantitative geographers were quick to realise that they provided a new platform on which to sell their wares. In the UK, a number of geography departments benefited from the Economic and Social Research Council (ESRC) initiative to create a series of research laboratories in GIS (see the final chapter in this collection by Flowerdew and Stillwell for a review of this initiative).

Alongside these two major developments came a series of other enabling factors which helped to promote applied quantitative analysis. First, the personal computer (PC) allowed the planner (both public and private sector) to have immediate desktop access to powerful software for spatial mapping and analysis. Second, spatial data was also becoming more routinely available in most application areas, including planning, as evidenced in Stillwell *et al.* (1999). Third, pressure was mounting on academics to interact more with the outside world in order to attract new income streams. Finally, we might acknowledge a genuine sea-change in attitudes of many academics – an increase in the desire to see their models usefully applied to social and economic problems – as well as a parallel change of attitude by private and public sector clients who began to recognise more seriously the benefits available from the exploitation of large data sets by new geotechnologies. The variety of planning support systems now being used in practice (Geertman and Stillwell, 2002) is one reflection of this new interest. The more generic set of issues also allowed a new emphasis on applied research in regional science (Clarke and Madden, 2001), and in non-quantitative aspects of human and physical geography (Pacione, 1999).

1.2 Aims and Contents

Against this background, the aim of this book is to illustrate the applied nature of contemporary quantitative geography through a series of case studies. It is clearly only a selection but there are still only relatively few examples of commissioned research in the literature (for further examples, see Clarke and Madden, 2001). The paucity of applied work in academic journals is perhaps not surprising since the results of many such projects are confidential, especially for private sector applications. Here, however, we have collected together a suite of case studies of funded work which has been applied to particular real-world problems confronted by either private or public sector organisations. Between them, these case studies involve the application of GIS, statistical models, location-allocation models and network or

flow models. However, rather than presenting these case studies on the basis of the methodologies they adopt, the book is arranged thematically in four parts around important subject areas: business applications; social deprivation studies; transport and location problems; and national spatial planning applications.

The initial quarter of the book (Part One) contains a collection of chapters focused on applications in the private sector, predominantly drawn from retailing. The first two chapters draw upon experiences with various clients. This shows the strength of applied work in retail analysis. Both Jones and Hernandez and Birkin *et al*. draw upon their experiences with many global organisations involved in retail location analysis. Jones and Hernandez give a very useful US/Canadian overview of commercial applications of GIS and spatial analysis. Their main conclusion is that the market is moving (slowly!) from one of simple mapping to the use of more sophisticated data mining and visualisation techniques. This argument is taken up in Birkin *et al.* where the authors examine a number of difficult applied location problems that have been faced by a number of clients. They begin to show how these problems may be addressed from an analytical point of view. The study by Câmara and Dias is a very novel use of network analysis for guiding shoppers around a major shopping centre complex in Lisbon, Portugal, undertaken in cooperation with SONAE, the company responsible for running the shopping centre. As the use of mobile phones for location-based services becomes more sophisticated, it is apparent that their use for pinpointing the geographical location of various types of services will increase. There is an important role for spatial analysis in optimising the best ways to maximise access to geographical information of this sort. Finally, in this part of the book, Laposa and Thrall investigate the use of GIS and models for the study of house price variations for PricewaterhouseCoopers. Thrall has been involved with applications of GIS in the business sector for many years, specialising in the real estate sector (Thrall, 2002). The main contention of the chapter is that the use of GIS, three-dimensional modelling and visualisation of house sale prices provides greater explanatory power than typical hedonic residential price models for estimating zonal or regional average house prices.

The set of chapters that are presented in Part Two have been commissioned by various local authorities helping to expose different aspects of social problems, especially deprivation and crime. The second chapter by Boyle and Alvanides examines a problem that many urban areas face in the UK. This is the so-called 'two-speed' growth problem. On the one hand, there are areas of cities experiencing rapid growth in the number of jobs, levels of service provision, house prices and/or residents. On the other hand, other parts of cities get left behind and remain as pockets of acute deprivation. In Europe, there are European Union (EU) monies available for those cities in most need of support. However, often the pockets of deprivation are not large enough to allow credible and integrated programmes to be mounted. This was the case in point with the city of Leeds, UK. The authors of this chapter, commissioned by the City Council, demonstrate how it is possible to identify a contiguous set of enumeration wards within the Leeds area which, when aggregated together, represent a relatively large concentration of households with a high deprivation score. The initial chapter in this part of the book, by Harris and Longley, takes a more reflective look at existing deprivation indicators used in socio-economic planning. In

particular, they evaluate the combination of data obtained from lifestyle databases with that available from high-resolution satellite imagery. They show how the latter can help to evaluate the degree of homogeneity in census-based deprivation scores (or income estimates) and offer a potential for representing deprivation or income clusters at more precise spatial scales. The third chapter relates deprivation in the city to the problems of child care. Sponsored by the local social service department, Craglia and Signoretta explore the use of GIS for the construction of a 'Children's Services Plan' in Sheffield, UK. In a sense, this is a children's census; data is put together on all aspects of children's health, status of care and behaviour (crime, attendance at school, etc.). Each of these variables is then ranked by small geographical area. The final product is a child deprivation score for the entire city. The final chapter in this part of the book is by Rogerson who looks at one dimension of social deprivation in more depth: namely, crime within Metropolitan Buffalo, sponsored by various city agencies. He uses GIS linked to the very latest spatial statistical analysis tools to find regions with significantly higher crime than could be expected under a random distribution of crime activity. He also uses various statistical techniques to examine the dynamics of crime patterns.

The set of chapters comprising Part Three relate to transport networks and location problems. The first chapter by Horn is a good example of using transport models to investigate the potential impact of a number of future transport plans. The case study is based on the Gold Coast, a rapidly expanding area of South East Queensland in Australia. The model is set up to help assess the viability of several road-based 'demand-responsive' transport modes (taxis, multi-taxis and so-called roving buses) designed to supplement existing bus, rail and taxi services. The analysis tool is a simulation model driven by demand simulators that replicate likely or future patterns of consumer demand. Network simulators help to service that demand by tracking the movements of individual vehicles and allocating the closest vehicle. The benefits and costs of operating such a system have been articulated and now the transport planners in Queensland have to make the ultimate decision of whether to introduce the scheme. The second chapter by Simmonds and Skinner uses a well-known land-use/transport model (DELTA) to examine the future transport plans of both West and South Yorkshire in the UK in the context of the preparation of Regional Planning Guidance for Yorkshire and the Humber (Government Office for Yorkshire and the Humber, 2001). As was noted in Section 1.1, land-use/transport models have a long history of applied success in planning. The model here helps to address two key questions: how can the authorities design an integrated and sustainable transport policy for the future, and how can the most urgent problems be specifically addressed? The following chapter by Geertman *et al*. reports the response to a classic geographical problem – where to site ambulance stations in a city context in order to minimise the time taken to access all parts of that city. The methodology adopted combines traditional shortest-path network analysis from a major proprietary GIS package with an in-house accessibility indicator package based on spatial interactions of flows. The work was commissioned by Rotterdam Municipal Health Authority. The final chapter by Alexander *et al.* reports on a very common environmental consequence of road construction, especially in areas well inhabited by wildlife. The construction of roads in such areas can cut-off access to pathways or

routeways historically used by various types of wildlife. The authors present a case study of the use of GIS to identify optimal placement sites for structures that facilitate wildlife passage across, or under, highways in the Canadian Rocky Mountains.

Part Four of the book contains a set of chapters that look at large-scale national social and economic problems. These are largely funded by government departments. The first chapter by Rees *et al*. was commissioned by the UK Department of Environment, Transport and the Regions (DTLR now the Office of the Deputy Prime Minister, ODPM). A key concern of DTLR was to balance population growth across the UK. In order to achieve this, they need to equalise future regional population growth in situ, but also to reduce migration losses from northern regions (flowing to the south). The chapter describes a two-stage migration model built to enable policy makers to investigate the first-round quantitative impacts of alternative economic and policy scenarios on gross flows of population between regions. The model used is a very disaggregated spatial interaction model, calibrated using statistical regression techniques. The second chapter continues the theme of policies for regional balanced population and economic growth. Van Wissen explores the concept of regional 'carrying capacities' as a framework for establishing policies on regional economic growth in the Netherlands. The concept of 'carrying capacity' is borrowed from ecology – the maximum use of land that can be sustained over time. This idea is translated into an economic context by a model of inter-industry linkages. It is argued that the growth potential of an individual firm in a locality relates to the size and composition of the population of firms in that locality. The final model is a mixture of a spatial interaction and an input–output model.

The chapter by Coombes and Raybould reports on the process of finding sites for a potential new UK government information service (funded by the Lord Chancellor's Department). It is another classic location problem and, not surprisingly, it fits into a framework of many public sector location–allocation problems. In this case, the problem is to locate a new information service for those couples facing the prospect of filing for divorce from a possible 647 candidate sites. The chapter by Frost and Shepherd looks at another issue of increasing concern to the UK DTLR (now ODPM) – rural accessibility. As the level of service provision declines in rural areas, there is an urgent need to measure rural accessibilities. Frost and Shepherd use GIS to build a parish-based survey of access to services. They also evaluate the changing role of the local market town in rural areas, in order to be able to identify market towns with strong/weak service centre functions. The final chapter by Kneale and See addresses a key problem faced by the Environmental Agency in the UK and elsewhere – how can we improve our flood forecasting methodologies. This comes at a time when the UK has faced a number of very wet periods and the amount of flooding has been severe. Yet, despite advances in most aspects of computer technology, the ability to predict the consequences of these floods has not been very successful. The approach by Kneale and See uses neural networks to improve the forecasting process and to give more time for operators to send out alarms.

The concluding chapter of the book addresses some of the issues and concerns that confront those working in an academic environment when attempting to undertake applied research. Flowerdew and Stillwell reflect on the advantages and limitations of applied research by drawing on two different university experiences during the 1990s: the ESRC-funded Regional Research Laboratory (RRL) initiative (with

comments based largely on the activities of the North West RRL at the University of Lancaster) and the Yorkshire and Humberside Regional Research Observatory established at the University of Leeds. They conclude that despite various difficulties associated with applied research (frequently done through consultancy arrangements), much valuable work has been undertaken already and the opportunities for making use of new GIS, analysis methods and modelling techniques in the future are very exciting.

We hope that the contributions assembled in this collection provide a useful representation of what has been achieved in recent years.

References

Allen, P. and Sanglier, M. (1997) A dynamic model of growth in a central plane system, *Geographical Analysis*, **11**: 256–73.

Batty, M. (1989) Urban modelling and planning: reflections, retrodictions and prescriptions, in MacMillan, B. (ed.) *Remodelling Geography*, Blackwell, Oxford.

Batty, M. (1994) A chronicle of scientific planning: the Anglo-American modelling experience, *Journal of the American Institute of Planners*, **60**: 7–16.

Clarke, G.P. and Madden, M. (2001) *Regional Science in Business*, Springer, Berlin.

Clarke, M. and Wilson, A.G. (1987) Towards an applicable human geography, *Environment and Planning A*, **19**: 1525–41.

Dendrinos, D.S. (1985) *Urban Evolution: Studies in the Mathematical Ecology of Cities*, Oxford University Press, New York.

Foot, D. (1981) *Operational Urban Models*, Methuen, London.

Geertman, S. and Stillwell, J.C.H. (eds) (2002) *Planning Support Systems in Practice*, Springer Verlag, Heidelberg.

Government Office for Yorkshire and the Humber (2001) *Regional Planning Guidance for Yorkshire and the Humber (RPG 12)*, Department of Transport, Local Government and the Regions, The Stationery Office, London.

Haggett, P. (1965) *Locational Analysis in Human Geography*, Edward Arnold, London.

Haggett, P., Cliff, A.D. and Frey, A. (1977) *Locational Analysis in Human Geography*, 2nd edition, Edward Arnold, London.

Harris, B. (1962) Linear programming and the projection of land uses, *Paper 20*, Penn-Jersey Transportation Study, Philadelphia.

Herbert, D.J. and Stevens, B.H. (1960) A model for the distribution of residential activity in an urban area, *Journal of Regional Science*, **2**: 21–36.

Johnston, R.J. (1996) *Geography and Geographers*, 5th edition, Edward Arnold, London.

Lee, D.B. (1973) Requiem for large-scale models, *Journal of the American Institute of Planners*, **39**: 163–78.

May A.D. (1991) Integrated transport strategies: a new approach to urban transport policy formulation in the UK, *Transport Reviews*, **11**: 223–47.

Pacione, M. (ed.) (1999) *Applied Geography: Principles and Practice*, Routledge, London.

Plane, D. (1994) Comment: on discipline and disciplines in regional science, *Papers in Regional Science*, **7**: 19–23.

Sayer, R.A. (1976) A critique of urban modelling, *Progress in Planning*, **6**: 187–254.

Stillwell, J.C.H., Geertman, S. and Openshaw, S. (eds) *Geographical Information and Planning*, Springer Verlag, Heidelberg.

Thrall, G. (2002) *Business Geography and New Real Estate Market Analysis*, Oxford University Press, New York and London.

Wilson, A.G. (1981) *Geography and the Environment: Systems Analytical Methods*, Wiley, New York.

Wilson, A.G. (1997) Land-use/transport interaction models: past and future, *Journal of Transport Economics and Policy*, **32**(1): 3–26.

PART 1
GEOBUSINESS

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

2

Retail Applications of Spatial Modelling

Ken Jones and Tony Hernandez

Abstract

This chapter examines the use of geographical information systems (GIS) and associated forms of spatial analyses currently used by North American retailers, property managers and urban planners in assessing various strategic issues (see Goodchild, 2000, for a more general assessment of GIS and spatial analysis). Initially, the chapter evaluates the value and adoption of spatial analysis and associated technologies by retail sector analysts. This discussion will be followed by five recent applications. These examples, drawn from client-based and academic research, will serve to illustrate the use of spatial analysis and GIS at various scales and for a variety of users groups. These cases will examine: (i) an assessment of regional variations in retail sales performance; (ii) the impact of 'big boxes' on the competitive structure of a retail environment; (iii) the spatial variability of customer penetration associated with a destination retailer; (iv) the internal dynamics and performance of a regional shopping centre; and (v) the spatial/temporal variation of sales volumes by product category for a network of retail outlets. The chapter will conclude with an evaluation of data mining and visualisation technologies as a potential means of enhancing the value of the spatial approach to retail corporate decision making. Due to commercial sensitivity, the names of the retail organisations involved have been omitted to maintain confidentiality.

2.1 Introduction

Over the last 25 years the awareness and use of GIS technologies for retail planning in North America has increased significantly (Sheerwood, 1995; Hernandez, 1999;

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

Hernandez and Biasiotto, 2001). Initially, GIS technology was considered by most retail organisations as having limited application, being viewed by many users as nothing more than an elegant form of digital mapping. GIS-based applications were typically housed in either the real estate or market research departments and normally were used to provide visual support for internal decision making (Reid, 1993). The early adopters, often associated with the grocery and department store chains, were inhibited by a number of constraints. These typically related to the limitations imposed by existing computer capacity, the general lack of spatially referenced databases, the elementary nature of the GIS software and the lack of trained analysts. As a consequence, early applications were associated with the automation of various traditional trade area methodologies such as customer spotting, boundary definition and market share estimates. Often retailers used simplistic trade area demographics to examine issues that related to store cannibalisation and market penetration. The next major stage in the development of spatial modelling is related to the development and widespread adoption of spatial demographic cluster systems in North America and Europe (see Flowerdew and Goldstein, 1989, for example).

In North America, two firms dominated the commercialisation of these advances – Claritas Corporation (US) and Compusearch Micromarketing (Canada). In both cases, these research firms took existing census data and developed spatially defined socioeconomic clusters for both the American and Canadian markets, respectively. Typically, these cluster profiles were developed for specific geographic areas (i.e., the enumeration area, block face or various postal geographies ZIP+4 or FSALDU). Many retail organisations subscribed to these databases and used them to assess existing store performance, evaluate future market opportunities, and/or to plan and execute marketing or direct mail campaigns. Shopping centre developers also used these systems to market various properties to prospective tenants, to examine mall performance, and to evaluate potential expansion plans. More recently, the use of GIS and associated spatial technologies has been extended into a number of new areas. This has been facilitated by the development of new databases that relate to the supply side of the retail economy, the acceptance of GIS as a mainstream technology by a growing number of retail analysts and decision makers, and the increased availability of more sophisticated spatial modelling and data visualisation software.

This chapter will illustrate the use of various spatial modelling applications in somewhat non-traditional, but burgeoning, retail applications. By adopting this approach, we hope to demonstrate the increased potential of spatial data, spatial analysis, and a spatial perspective to various levels and types of decision-making activities associated with the retail and service economy. In North America, the term *business geomatics* has been coined recently to reflect the emergence of a new interdisciplinary area that is beginning to link spatial information, spatial technologies, the spatial sciences and applications (Yeates, 2001). In the long run, this development will increase the awareness and credibility of spatial modelling and GIS applications to a much wider community in both the academic and private and public sectors.

2.2 Retail Sales Performance: A Macro Approach

The first application focuses on the analysis of the major concentrations of retail sales activity at the national or regional level using small area retail trade estimates data (SARTRE). SARTRE reports total annual retail sales for approximately 142 000 retail locations across Canada aggregated to Forward Sortation Area (FSA) level. FSAs are a unit of postal geography comprising 1452 areas covering all of Canada. SARTRE is based on a combination of data from the Retail Chain Location file of Statistics Canada (that tracks the locations of all retailers operating in Canada with four or more locations) and corporate tax returns from Revenue Canada. The increasing availability of macro data permits the retail analyst to rank major retail areas in Canada by retail sales performance. In addition, given the longitudinal nature of these data, the fastest growing retail markets can be readily identified. In this example, these spatial databases are used to link retail trade sales to store locations at a fine level of geographic detail – typically, postal areas or census areas. The example combines data from survey and taxation information for all incorporated retailers in Canada and groups these businesses according to retail category (i.e., grocery retailing, fashion, department stores, automotive). This level of supply-side information allows for a detailed assessment of retail performance by major retail category for relatively small geographical areas.

The analysis of data of this type can yield some interesting general trends at the national scale. Often retail chains seek out locations in fast-growing areas. The urban market areas identified in Table 2.1 are confined to areas with retail sales in excess of $100 million and sales growth greater than 200%. Interestingly, the two largest regional markets in Canada, Ontario and Quebec, combine for only four of the top 10 fastest growing retail concentrations (defined by FSA). A detailed analysis suggests that these high growth locations tend to exhibit a suburban character, although two of the three Edmonton locations, along with the Montreal FSA, are certainly central within those cities. The locational dynamics of retailing for the period are also uncovered when we compare the sales growth to the change in the number of

Table 2.1 *Fastest growing retail areas in Canada, 1990–95*

FSA	Market	% Retail sales growth[1]	% Retail location growth[1]
V9V	Nanaimo, BC	3283.29	133.3
L5R	Mississauga, Ont.	1817.46	94.7
V2Y	Langley, BC	1628.72	285.7
T5V	Edmonton, Alta.	1451.68	157.1
H3K	Montreal, Que.	672.28	–4.8
T5A	Edmonton, Alta.	373.92	–9.3
V3N	New Westminster, BC	276.28	2.1
G1X	Ste-Foy, Que.	246.57	8.6
M8X	Etobicoke, Ont.	223.79	–3.2
T0C	Edmonton, Alta.	213.76	–17.9

Note: [1]Percentage growth calculated as follows: (1995 value minus 1990 value) divided by the 1990 value ×100

Table 2.2 *Top 20 Canadian retail hot spots*

FSA	Major shopping node	Market	Sales score[1]	Sales/Location ($000)[2]
M5B/C	Eaton Centre/The Bay	Toronto (CBD)	8.02	1626.204
T2H	Chinook SC	MacLeod Trail, Calgary	6.76	2051.122
L3R	Markville SC	Markham, Ont.	6.02	1326.133
H3A/B	Place Ville Marie/ Eaton Ctr.	Montreal (CBD)	5.89	1469.971
V6X	Lansdowne, SC	Richmond, BC	5.86	1727.706
M6A	Yorkdale SC	Toronto	5.39	1965.889
V5H	Metrotown	Burnaby, BC	5.34	2182.542
H9R	Fairview Ctr.	Pointe Claire, Que.	5.09	2053.887
M1P	Scarborough Town Ctr.	Toronto, Ont.	5.05	1981.640
L4M	Bayfield St./ Georgian Mall	Barrie, Ont.	4.74	1860.050
V1Y	Orchard Park SC	Kelowna, BC	4.64	1452.908
T1Y	Sundridge SC	Calgary, Alta	4.59	2349.063
T5T	West Edmonton Mall	West Edmonton, Alta	4.56	1408.541
T2J	MacLeod Trail	Calgary, Alta	4.48	1547.741
S7K	Midtown SC	Saskatoon (CBD)	4.48	1198.968
N8X	Devonshire Mall	Windsor, Ont.	4.35	2074.595
L5B	Square One	Mississauga City Ctr., Ont.	4.25	1847.830
L3Y	Upper Canada Mall	Newmarket, Ont.	4.16	1594.634
N6E	White Oaks SC	S. London, Ont.	4.14	2025.862
G1V	Place Laurier	Quebec City (CBD)	4.08	1255.477

Notes:
[1] Based on the 20 FSAs with the greatest big-box square footage in 1995. The sales index values for each FSA were normalised to assure confidentiality. These values show the relationship between sales in each of these FSAs and average sales in all FSAs in the GTA. An FSA with sales equal to the average for all FSAs, would have a sales index value equal to 1.00
[2] Sales/Location calculated as retail sales in FSA divided by the number of retail locations in FSA

locations. Over the five-year period, four markets reported an increase in sales greater than 200% while the actual number of retail locations declined. This may suggest the growing importance of large, highly productive 'big box' retailers in these areas. Growth of 3000% in Nanaimo highlights the vigour of the market in western Canada.

When the retail areas in the database are ranked by their reported retail sales figures (Table 2.2), it is possible to identify major shopping nodes that exist within the top 20 markets. A number of high profile locations appear in the *Top 20 Retail Hot Spots*, along with what may be some unexpected additions. The major Toronto area regional shopping malls like the Eaton Centre, Yorkdale, Scarborough Town Centre and Square One, are joined by Markham's Markville Shopping Centre. In addition, just to the north of the Greater Toronto Area (GTA), Barrie's Bayfield St. and Georgian Mall area makes the list with the tenth highest reported sales in the country. Of note

is the strong presence of Calgary's MacLeod Trail area and the Chinook Shopping Centre. It actually tops the list if the two retail areas that make up the top position are separated. Just under half of the top 20 locations are within Ontario, reaffirming the province's strong retailing position and demonstrating continued market concentration and relative affluence. The list also appears fairly compact, showing about the same range in sales score and sales per location, with the first place position doing only double that of the twentieth location.

The use of macro retail sales databases is particularly powerful for assessing general location strategies and does provide a much needed overview of the performance of the national retail system. For the trained analyst, a diversity of locations and patterns appears as the data are explored in terms of the overall strength of retail, fashion retail and general merchandising. Such a high-level database provides benchmarks for the retail organisation and provides valuable insights for retail managers and property owners into the operation of the retail spatial economy. In particular, one can assess the changing role of downtown areas, identify healthy shopping centre locations and speculate on the impact of new retail formats such as power centres or major retailers such as Wal-Mart. The data can also be input into various data visualisation programs and dynamic three-dimensional (3D) views of the retail economy can be generated (Plates 1 and 2). MineSet software produced by Silicon Graphics Inc. was used to generate the 3D maps. The software allows users to view temporal animations of spatial data.

2.3 Impact of Big-Box Retailing

The second application illustrates the influence of big-box retailing (e.g., Home Depot, Wal-Mart) on the health of street-front (strip) retailing. Increasingly, the impact of big-box/power centre development has become a major concern for urban planners, retail organisations with street locations, and financial institutions who invest heavily in retail property. One way of measuring the influence of big-box activity on the operation of the urban retail system is to examine changes in the functional composition of those retail strips that are in direct competition with these new large format retailers. This type of analysis requires an extensive spatial database of retail activity that is collected and maintained on a yearly basis. The particular case examines retail change in the City of Toronto. In this urban area, as elsewhere, strip retail areas are dominated by independent retailers, and often provide a social, cultural and economic focus for their surrounding neighbourhoods. Indeed, one of the major distinguishing features of the Toronto area from a North American perspective is the health, diversity and vibrancy of its retail streets. In total, the urban area is served by a network of over 200 retail strips that provide the citizens of the city with access to over 18 000 retail shops, restaurants and personal and business services. However, this system has been threatened by the recent arrival of big-box retailing and this change has created concerns for local planners, neighbourhood groups and local business organisations. The analysis of spatial data can help to provide empirical evidence that is needed to assess and monitor the magnitude of the impact of large format retailers on the local retail environment.

One fundamental measure of retail restructuring can be derived through an examination of changes in the economic role of retail areas and the relative growth or decline in vacancy rates. For the 1994–97 period, data on the commercial structure of Toronto's major retail streets, including vacancy rates, were collected and tabulated by the Centre for the Study of Commercial Activity (CSCA) at Ryerson University. This information is presented in Table 2.3, which provides an aggregate view of the changing structure of street-front retailing in Toronto over the 1994–97 period for selected retail categories. These categories reflect the areas of retailing that are in direct competition with the dominant big-box retailers currently in operation (for example, hardware – Home Depot; books – Chapters; office products – Office Place/Business Depot; supermarkets – Price Costco; and various other superstores – PETsMART). The data suggest that, during this period of significant big-box and power centre development in the GTA, the relative importance of retailing on the retail streets of Toronto, as measured by the proportion of all occupied stores, declined from 53.7% in 1994 to 49.5% in 1997. In addition, certain key retail categories that compete directly with a major big-box format either declined (hardware –10.4%, General Merchandise –3.9%) or remained relatively stable (food –0.8%, Pharmacy +0.4%, and furniture/housewares +1.7%). The significant decline in men's and women's fashion corresponds with the growth of discount department store chains (e.g., Wal-Mart) and the re-mixing of the traditional department stores. Complementing these changes was the trend toward a general increase in the importance of restaurants and personal and business services along Toronto's retail streets. These non-retail functions experienced a 24.1% growth in the number of stores over the study period, and all services now account for more occupied street-front units than retail activities. It is also worth observing the steady decline in the financial category within Toronto's retail strips during this period; a change that would, without question, accelerate if proposed mergers between some major Canadian banks were to be approved by the Federal Government. The inclusion of automated banking machines in many big-box outlets further emphasises the potential of new format retailing to alter street-front activities in major parts of the tertiary sector. Finally, the data suggest that the overall health of street retailing in Toronto has declined somewhat over this four-year period, as the number of vacancies increased by 10.6%. These figures, however, may well mask real changes to the overall 'quality' of the shopping experience provided along particular retail streets. Canadian business leaders have suggested that the shift from traditional retail activities, such as clothing stores and hardware shops, to cut-price stores and doughnut shops, while better than vacancies, is not always a positive change.

Table 2.4 examines the change in the relative share of the eight retail categories that are most vulnerable to big-box retailing. Over this period, the total number of stores within Toronto's retail strips grew from 13 026 to 14 448, an increase of 10.9%. However, for the eight selected categories (those most impacted by the big boxes), the number of stores experienced only a 2.9% increase, from 982 to 1010. Further, the total share of all stores captured by the eight categories fell from 7.3% to 6.8%. What is evident from a closer analysis of these figures is that in all eight categories studied, the percentage share of these store types declined by an average of 7%, with

Table 2.3 *Changes in the economic role of Toronto's major retail strips, 1994–97*

Activity	Number of establishments				% Growth rate 1994–97
	1994	1995	1996	1997	
Food	1 694	1 692	1 681	1 681	–0.8
Liquor	67	69	73	79	17.9
Pharmacy	238	244	245	245	0.4
Shoes	199	207	203	210	5.5
Men's clothes	184	173	167	168	–9.5
Women's clothes	565	551	539	515	–8.7
Other clothes	414	441	429	445	7.5
Furniture & housewares	720	723	712	732	1.7
Automotive	410	426	428	428	4.4
General merchandise	254	253	255	244	–3.9
Books & stationery	179	184	180	184	2.8
Florists	180	185	187	181	0.5
Hardware	222	216	209	199	–10.4
Sports	156	152	151	158	1.3
Music	264	272	275	279	5.7
Jewellery	171	178	179	195	14.0
Other retail	1 076	1 143	1 155	1 206	12.1
Total retail	*6 992*	*7 172*	*7 095*	*7 149*	*2.2*
% Retail	*53.7*	*52.3*	*51.0*	*49.5*	*–7.8*
Financial	*615*	*597*	*582*	*575*	*–6.5*
% Financial	*4.7*	*4.4*	*4.2*	*4.0*	*–14.9*
Restaurants	2 531	2 676	2 823	2 953	16.7
Recreation	179	227	247	259	44.7
Medical	405	475	502	515	27.2
Hair/Beauty	798	900	950	1 002	25.6
Cleaning	367	380	382	397	8.2
Other personal	615	680	749	836	35.9
Business services	524	613	678	762	45.4
Total services	*5 419*	*5 951*	*6 331*	*6 724*	*24.1*
% Services	*41.6*	*43.4*	*45.2*	*46.5*	*11.8*
Total activity	*13 026*	*13 720*	*14 008*	*14 448*	*10.9*
Vacancies	*1 259*	*1 440*	*1 490*	*1 392*	*10.6*
%Vacancies	*9.7*	*10.5*	*10.6*	*9.6*	*–1.0*
Total stores	*14 285*	*15 160*	*15 498*	*15 940*	*11.6*

the largest declines found in office products (–23%) and hardware (–16.5%), respectively. Even categories that have only recently been affected by big-box competition experienced a decline in their street-front locations (e.g., books down 2.2% and pet stores down 2.1%) in the years in question. Given the relatively brief time period for which data are available, these changes to the retail structure of Toronto's shopping streets are striking.

Table 2.4 *Change in the composition of retail strips in the categories most impacted by big boxes: the metropolitan Toronto experience*

Category	1994		1997		Change in relative share by category 1997–94 %
	No. of stores	% share	No. of stores	% share	
Supermarkets	253	1.94	261	1.81	–6.7
Sporting goods	104	0.8	111	0.77	–3.8
Books	120	0.92	130	0.9	–2.2
Office products	35	0.03	33	0.023	–23.3
Hardware	142	1.09	131	0.91	–16.5
Toys	79	0.61	83	0.57	–6.6
Electronics	188	1.44	195	1.35	–6.3
Pet stores	61	0.47	66	0.46	–2.1
All eight categories	*982*	*7.3*	*1 010*	*6.79*	*–7.0*
Total stores	*13 026*		*14 448*		

Table 2.5 *Changes in the number of independents on Toronto retail streets, 1994–97*

Category	Number of stores		Change	% Change
	1994	1997		
Supermarkets	225	220	–5	–2.3
Sporting goods	91	100	9	9.9
Books	99	118	19	19.2
Office products	32	31	–1	–3.1
Hardware	107	96	–11	–10.3
Toys	73	82	9	12.3
Electronics	171	163	–8	–4.4
Pets	39	55	16	41.0
All eight	*837*	*865*	*28*	*3.3*
All independents	*11 096*	*12 646*	*1550*	*13.9*

Table 2.5 examines the same database in order to trace the change in the number of independent retailers in the same eight categories. In this case, two categories of retailer were more affected – hardware and electronics. These categories suffered losses of 11 and eight stores, respectively, between 1994 and 1997. There was also a slight reduction in the number of independent office products retailers and supermarkets during this interval. Independent retailers in the remaining categories experienced some growth between the years in question, most notably book sellers, sporting goods, toys and pet stores, with net gains of 19, nine, nine, and 16 establishments, respectively; a reflection of growth trends in these retail sectors. Overall, independent stores on Toronto's retail strips grew by 13.9% during this period, a

Table 2.6 *Major concentrations of big-box retailing in the GTA*[1]

Name of node	FSA	Sales index	% increase in sales 1989–95	Big-box growth 1989–95 (000s of sq. ft.)	No. of boxes	
					1995	1998
Orion Gate	L6W	2.46	37	353	11	14
Woodbine & Hwy. 7	L6G	n.a.	n.a.	387	10	14
Crossroads	M9N	1.82	7	32	10	11
Thickson	L1N	2.15	19	229	9	16
Pickering	L1V	3.29	16	71	9	10
Warden & Ellesmere	M1R	1.58	20	314	3	3
Heartland	L5R	1.47	3940	238	7	9
400 & Hwy. 7	L4L	3.31	214	295	6	18
Hyde Park	L6H	1.31	22	272	8	14
Burlington	L7N	0.95	–3	36	5	5
SW Etobicoke	M9B	2.89	74	200	3	3
Central Mississauga	L4Z	0.64	23	194	3	5
Oakville	L6M	0.58	343	50	8	9
Unionville/Markham	L3R	5.19	30	60	6	7
Yorkdale	M6A	4.65	10	102	6	6
Rosart Centre	L7T	0.64	79	200	1	4
Ikea	M2K	2.09	–16	0	1	1
Warden Woods	M1L	2.07	9	101	6	8
Erin Mills	L5L	2.17	13	21	5	10
Younge/ Richmond Hill	L4C	2.39	24	180	2	4
Big-box nodes		*$4702.3*[2]	*18.60*	*3335*	*119*	*171*
Total GTA		*$21454.8*[2]	*9.20*	*6237*	*251*	*320*
Residual		*$16752.5*[2]	*6.80*			

Notes:
[1] Based on the 20 FSAs with the greatest big-box square footage in 1995. The sales index values for each FSA were normalised to assure confidentiality. These values show the relationship between sales in each of these FSAs and average sales in all FSAs in the GTA. An FSA with sales equal to the average for all FSAs, would have a sales index value equal to 1.00
[2] Expressed ($millions) includes all retail sales minus automotive sales (SIC 60, 61, 62, 64, 65)

marked contrast to the trends for the selected categories that, together, experienced only a 3.3% increase. As indicated in Table 2.3, the service sector has been experiencing substantial growth over the 1994–97 period, accounting for the vast majority of new stores within Toronto's retail strips.

In order to evaluate the impact of the big-box phenomenon on the distribution of retail sales, the 20 areas in the GTA with the greatest concentration of big-box retailers were identified (Table 2.6). These areas captured 47.4% of the big-box outlets in the GTA in 1995, and collectively accounted for over 3.3 million square feet

Table 2.7 *Number and probability of retail closures and distance from the nearest competing big box, 1994–97*

Category	Distance to the nearest big-box competitor			Probability of closure by category
	0–2 km	2.01–4 km	>4 km	
Supermarkets	43/329 (.13)	61/375 (.16)	15/69 (.22)	.154
Electronics	60/257 (.23)	39/177 (.22)	20/136 (.15)	.208
Office products	5/40 (.13)	7/40 (.18)	2/20 (.10)	.14
Sporting goods	6/44 (.14)	9/52 (.17)	35/213 (.16)	.162
Toys	12/73 (.16)	16/93 (.17)	12/74 (.16)	.167
Hardware	8/32 (.25)	10/82 (.12)	19/170 (.11)	.13
Books	7/42 (.17)	5/39 (.13)	4/47 (.09)	.125
Pet stores	1/2 (.5)	0/2 (.00)	2/59 (.03)	.048
Probability of closure by distance band	.173	.171	.138	.161

(53.5%) of big-box expansion during the 1989–95 period. What is particularly noteworthy is that these 20 areas experienced an 18.6% increase in retail sales between 1989 and 1995. This compares with a growth of just 9.2% for the entire urban region for the same interval. In addition, some of the most impressive sales growth occurred in the region's emerging power nodes: Highways 400 and 7 (214%), Orion Gate (37%), Thickson (19%), Hyde Park (22%), and Heartland (3940%). Each of these nodes experienced a growth of at least 200 000 square feet of big-box space between 1989 and 1995 and, collectively, these power nodes accounted for 1.4 million square feet of new, big-box retail space, or 22.2% of the total growth. In comparison, the areas associated with Toronto's traditional retailing – street fronts and shopping centres experienced more limited sales growth of just 6.8%.

This section has examined the competitive pressure that big-box retailing has placed on the retail strips in Toronto over the 1994–97 period. By using the annual inventories of street retailing collected by the Centre for the Study of Commercial Activity, the *number* and *probability* of store closures for each of the retail categories in direct competition with big-box retailers were calculated for the three-year period (Table 2.7). Over the three-year period, the number of retail categories in competition with the street retailers increased from five (supermarkets, electronics, office products, sporting goods, and toys) to eight with the introduction of Home Depot (hardware), Chapters (books), and PetSmart (pet stores) into the competitive mix in the Toronto market. As Table 2.7 shows there was a 16.1% chance that retailers

operating on street-front locations in Toronto in these eight categories would close. The sectors with the highest store closure rates were electronics, toys and sporting goods. Those locations within 4 kilometres of a competing big box experienced the greatest impact with a closure rate of approximately 17%. The probability of failure exhibited a slight distance decay effect declining to 14% when the nearest big-box competitor was located more than 4 kilometres away. However, what was striking was the broad areal extent of big-box retailers over space. In aggregate, the competitive effect of big-box retailers, with their large trade areas, is distance insensitive. One location can affect the entire urban system. Among the activities that did exhibit some distance sensitivity were hardware and books, while the toys category exhibited the most homogeneous spatial impact. It is of interest to note that big-box developments in the GTA occurred primarily in the outer suburbs in the early 1990s and have subsequently 'infilled' to the edge of the downtown core. During the last few years, big-box developments, and the grouping of these to form 'power centres' have significantly impacted traditional strip retailers and shopping mall tenants alike.

Over the three years for which data were available, street retailers in the eight categories in direct competition with the selected big-box retailers experienced a closure rate of 16.1%. This figure compared to a 14.8% closure rate for all other retail categories. When the data is analysed on a yearly basis there appears to be a competitive adjustment to the big-box phenomenon. In the initial years of competition, the closure rate in a category is high as the less competitive retailers are forced out of business. This adjustment phase appears to have taken place in a variety of categories – especially food, office products and toys. Furthermore, the competitive pressures in some categories (e.g., electronics) appears to be always high – a function of the dynamic nature of the product line in these retail categories. A number of forces tend to relate to the impact of big-box retailers on local retail communities – time, distance, product category and the health of the retail economy.

2.4 Market Share of a Destination Retailer

A more typical application of spatial analysis in retailing has been the assessment of market areas. Various indicators are normally used including: market share, market penetration and market opportunity. Many of these applications have been incorporated in various software packages (e.g., Huff, 2000). In this case, the market area of a major destination is examined using GIS technology (normally through 'buffer' and 'overlay' features). The ability to visualise areas or market concentration, the spatial variability of the expenditure surface and the distance decay effects provides a wealth of information to the retailer.

In Figure 2.1, the typical map of market penetration around the destination retailer is presented. Here, a highly variable market coverage is presented, with obviously the highest concentration of retail expenditures coming from areas directly north and north-west of the site. However, pockets of high expenditures are found throughout the trade area, located in both areas that are associated with high income and/or adjacent to an expressway, thus providing greater accessibility to the retail location. In assessing this map, typically one attempts to seek relationships between the levels

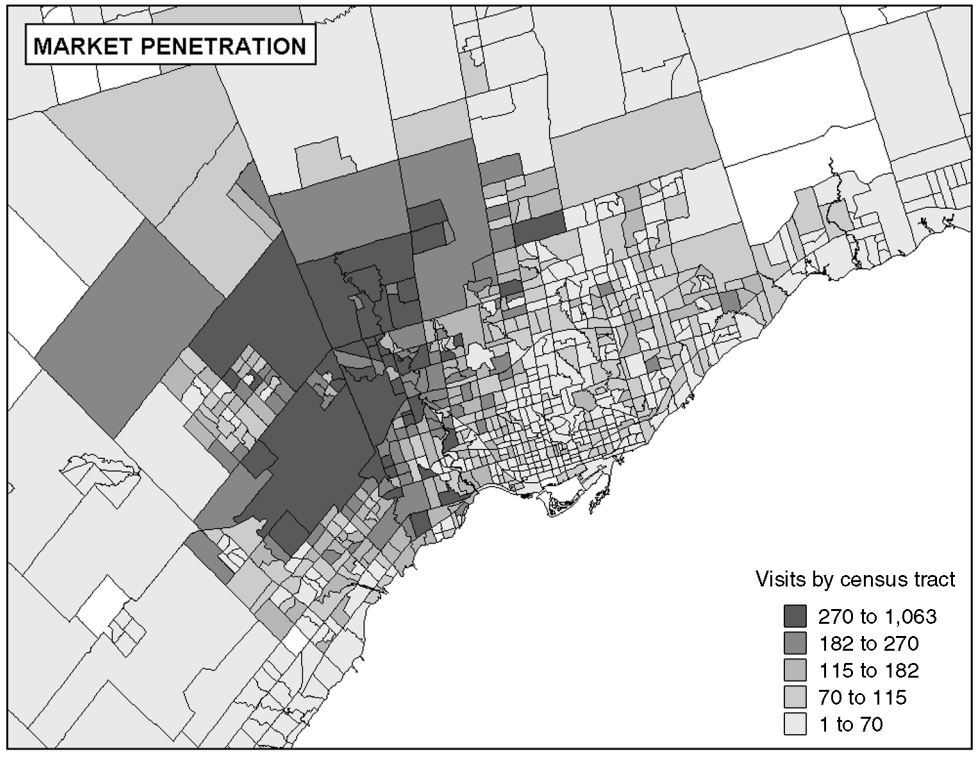


Figure 2.1 *Market penetration of a major destination retailer*

of market penetration, retail expenditure and the socioeconomic characteristics of the population. If strong associations are found, this information can be used in various marketing campaigns and can be used to assess the market potential of future locations.

Figure 2.2 uses the same data to develop a three-dimensional map of the trade area. This technology is being increasingly used by North American retailers to assess the spatial extent of their markets. When compared with the two-dimensional version, this representation provides much greater insights into the nature of the trade area. Clear distance decay effects become obvious and the importance of the immediate trade area becomes apparent. The value of this information is increased significantly when data are presented on a monthly or annual basis. Spatial trends become readily observed and these can be linked to a variety of corporate marketing strategies or can help assess various exogenous changes (e.g., changes in the economy or new competitive entries). The ability to incorporate simple data visualisation technologies into GIS has provided a major advance into the use of spatial mapping for many organisations and once again places more pressure on the need to collect and maintain increasingly detailed and current spatial databases (see also the case studies in Chapter 3).

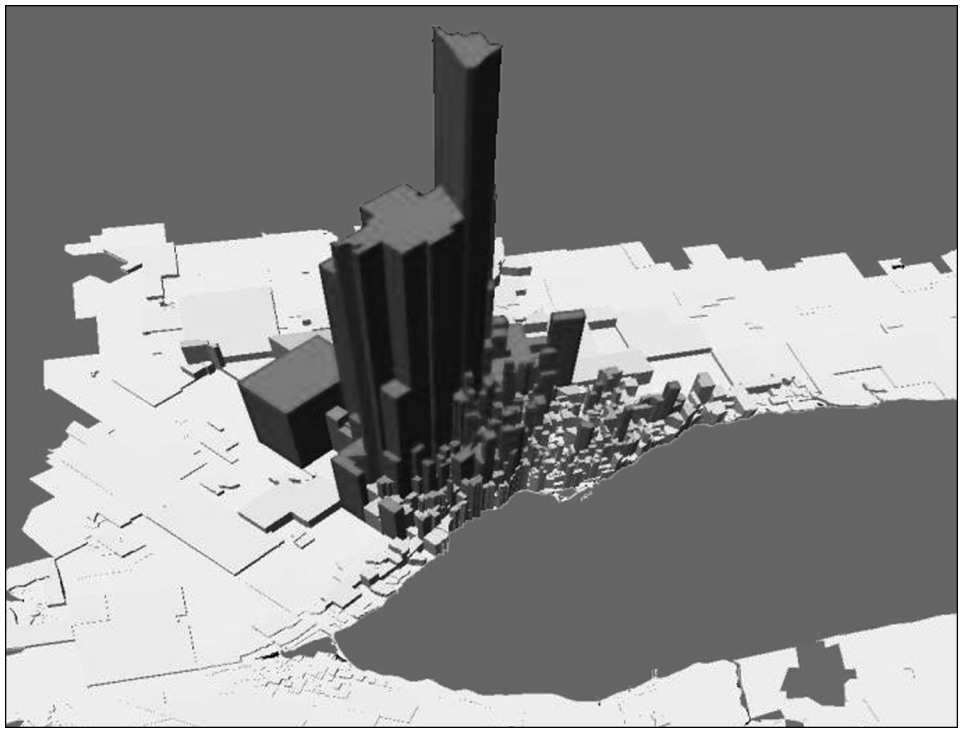


Figure 2.2 *Three-dimensional trade area representation*

2.5 Evaluating Mall Dynamics

One area that has experienced relatively little attention in the literature has been the modelling of tenant performance within a shopping centre (Yeates *et al.*, 2001). This is somewhat suprising given the level of investment in these properties and the size and nature of the databases that are collected by shopping centre developers and ownership groups. Mall management is a complex process. It requires that successful shopping centre developers are able to respond to and anticipate change effectively. This reality requires the use of sophisticated management systems and the ability to mine and effectively analyse a large array of information. One means of addressing these needs is to incorporate spatial support systems into their day-to-day operations.

This application examines the development of a spatially referenced management system for a regional shopping centre. This prototype system was developed in the mid-1990s (Jones *et al.*, 1995). It linked the shopping centre developer's operational databases, market area demographics, the distribution of competing facilities and the road network. The objective of the system was to provide the shopping centre management group with the ability to examine in 'real time' the performance of the shopping centre at various spatial scales. It provided the management with a means of

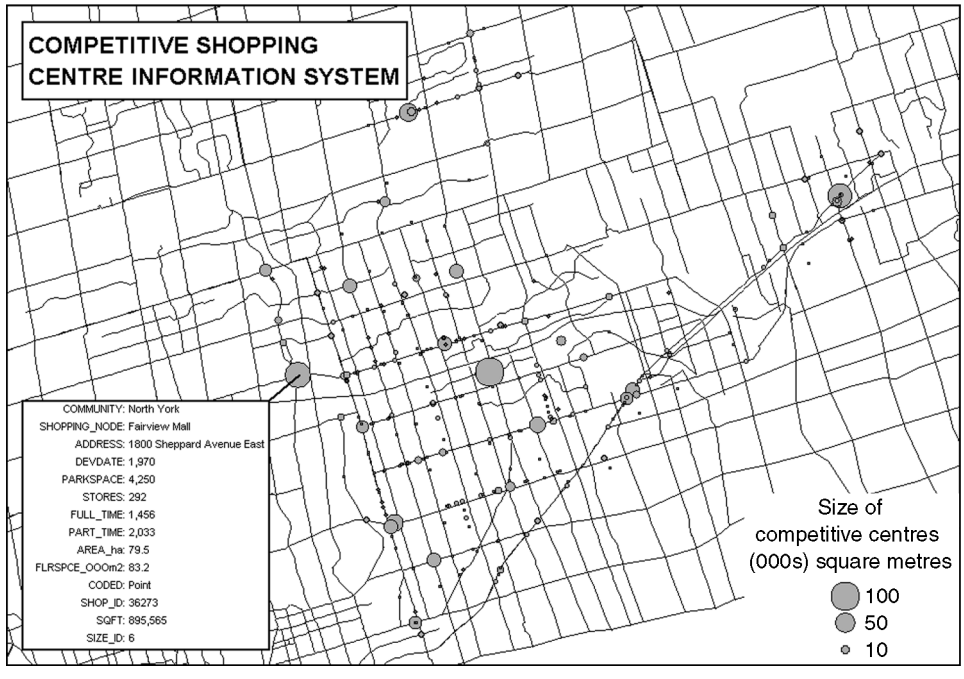


Figure 2.3 *The competitive retail system*

generating quick answers to a number of fundamental strategic management concerns. These included:

- the characteristics and changing composition of the trade area;
- areas of over and under-penetration;
- the impact of mall advertising and mall promotions;
- the location and impact of the competition;
- changes in tenant performance on a monthly and year over basis;
- the dynamics of the tenant mix;
- the effect of new tenants on mall performance;
- identification and measurement of 'mall dead spots' and vacancy rates; and
- evaluation of tenant turnover rates.

The following maps provide an illustration of the system. In Figure 2.3, the competitive retail environment is shown. With this database, the management could quickly link to supporting data that would provide a series of measures of each competitive facility in the primary trade area. These measures included such variables as date of development, number of stores, gross leasable area, number of parking spaces, employment vacancy rates, turnover rates, tenant mix and total retail sales generated in competitive retail nodes. This supply-side database was supported by a demand-side database that captured various demographic data at the enumeration

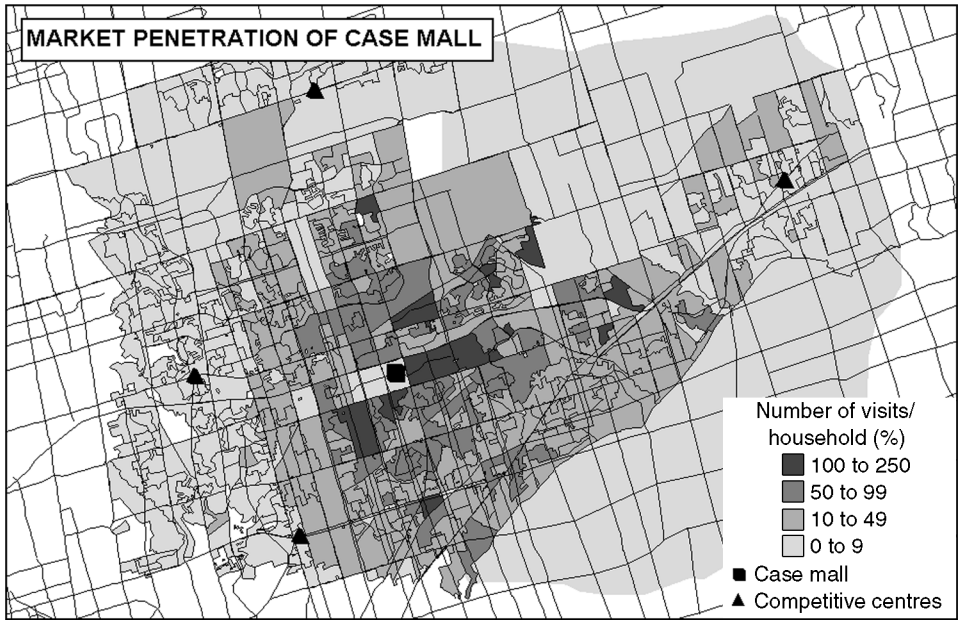


Figure 2.4 *Shopping centre market penetration by EA*

level. This data could be used when combined with customer-spotting data to (i) evaluate mall penetration by demographic segment and (ii) identify market segments that were over or under-served by the mall (Figure 2.4).

The principal advantage of the system was to develop a mechanism to monitor the performance of tenants within the mall. A GIS was developed that integrated the mall and tenant floor space with the historic performance data 'spatial record keeping' and provided mall management with a powerful, quick response analytical tool. It addressed a long-standing problem that too often shopping centre developers had no cost-effective means of systematically tracing the historical evolution of the tenant mix, structural changes or performance histories of their various properties at any spatial scale. By incorporating a GIS-based management system into their operations, shopping centre managers have a spatially based decision support system that can be applied to track and model the dynamics of tenant and mall performance. Figures 2.5, 2.6 and 2.7 illustrate the use of this system. In Figure 2.5 variations in tenant performance on an annual sales per square foot basis is depicted for a two-level mall. The map clearly shows areas of high and low performance and identifies the potential locations that should be performing at a higher level. In Figure 2.6, changes in annual sales performance are shown. Once again this system, when applied and tracked on a monthly basis, can provide mall managers with a diagnostic tool with respect to isolating those tenants that are unproductive, and more generally, identifying areas of the mall that are either gaining or losing sales. What typically emerge are well-defined spatial patterns that may reflect the importance of mall anchors, the decline or growth of certain chains, or changes in consumer

Figure 2.5 *Sales per square foot analysis*

Figure 2.6 *Tenant percentage sales change, 1994–97*

Figure 2.7 *Standardised tenant performance*

behaviours. In Figure 2.7, a more refined view of tenant performance was developed. Here, each tenant's sales performance is evaluated relative to its product category. This approach normalises the data, otherwise food court tenants would always outperform fashion tenants simply on their ability to generate much higher sales per square foot. The use of this tenant performance index provides a standardised view of mall performance and clearly identifies the best performing tenants and the hot and cold spots within the shopping centre.

2.6 Longitudinal Sales Volumes for a Retail Chain

The final application focuses on the development of data visualisation systems to analyse and track sales volume change for a retailer operating a large network of stores across Ontario. The company in this example was faced with an ever-increasing volume of transactional and operational data, in particular, weekly sales by store by product category (amounting to a substantial spatial–temporal retail performance database). The challenge for the company was in turning vast amounts of data into valuable insight and knowledge. The solution was found in the use of a geovisualisation approach using Silicon Graphics Inc.'s MineSet software and hardware.

Geovisualisation (also referred to as visual data mining) is a process of selecting, exploring and modelling large amounts of spatial data to uncover previously unknown patterns of data for competitive advantage. As MacEachren and Kraak (2001, p. 3) define, 'geovisualization integrates approaches from visualization in scientific

computing, cartography, image analysis, information visualization, exploratory data analysis and geographic information systems to provide theory, methods and tools for visual exploration, analysis, synthesis and presentation of geospatial data'. There is an extensive literature within the field of data visualisation and mining (Slocum, 1999; Lloyd, 1997; Brown *et al.*, 1995; MacEachren, 1995). The traditional focus of data mining and visualisation research has been in fields such as biotechnology, engineering and medical sciences. Geovisualisation, however, is an emerging research area (MacEachren and Kraak, 2001; Slocum *et al.*, 2001).

By applying visualisations and data-mining techniques, retailers can fully exploit their data warehouses and associated large-scale relational databases, to gain a greater understanding of the markets in which they operate. By reducing complexity, encouraging model interpretation, and easily depicting multidimensional data, the visual paradigm empowers retail decision makers and potentially reduces the time and effort required to gain valuable insight from reams of data. With large amounts of data collected by retailers it is now possible to shift away from *a priori* hypothetico-deductive forms of modelling toward inductive (or querying) approaches in which models are deciphered from the data themselves.

Visualisation facilitates the development of dynamic interactive decision support tools that are a means for data exploration and provide immediate feedback to the decision maker (user). As Slocum *et al.* (2001, p. 17) note, 'developments in hardware and software have led to (and will continue to stimulate) novel methods for visualizing geospatial data'. For example, data animation can be used to depict trends and patterns by expressing how critical attributes change over key variables such as time and space. Geovisualisation tools can also depict business data using three-dimensional visualisations, enabling users to explore data interactively and discover meaningful new patterns quickly. Moreover, animated three-dimensional landscapes take advantage of a human's ability to navigate in three-dimensional space, recognise patterns, track movement and compare objects of different sizes and colours.

The use of GIS mapping is now commonplace in retail organisations in North America. GIS-derived maps are typically depicted as abstract two-dimensional plan views, such as a choropleth map of market penetration, that is viewed from directly overhead and represents data values through colour or shading, with vision the primary means of acquiring spatial knowledge (Slocum and Egbert, 1993). Three-dimensional mapping potentially assists in identifying spatial trends, intensity and variation in data. There have been a number of studies that have focused on the utilisation and potential for three-dimensional mapping (Ledbetter, 1999; Haeberling, 1999; Kraak, 1994). The market share application presented earlier in the chapter clearly illustrated the potential benefits from viewing a map in three dimensions, with spatial patterns accentuated when compared with traditional two-dimensional choropleth representation (for example, compare the visual impact of Figures 2.1 and 2.2).

Often, data sets are just too complex for representation in two or even three dimensions. By animating the display across user-defined independent variables, users can easily observe trends in extremely complex data sets. With animation, the user has the ability to discover trends, patterns and anomalies in data. The term 'dynamic representations' refers to displays that change continuously, either with or without user

control. Dynamic representation has changed the way users obtain and interact with information across the full range of display technologies (Andrienko and Andrienko, 1999; Koussoulakou and Stylianidis, 1999; Bishop *et al*., 1999; Blok *et al*., 1999; Acevedo and Masuoka, 1997; DiBiase *et al*., 1992; Dorling, 1992). One form of dynamic representation is the animated map, in which a display changes continuously without the user necessarily having control over that change. An argument for utilising animation is that it is natural for depicting temporal data because changes in real-world time can be reflected by changes in display time. In addition, dynamic representations also permit users to explore geospatial data by interacting with mapped displays, a process sometimes referred to as direct manipulation (Slocum, 1999).

Plate 3 provides a sequence of temporal snapshots from an animation of weekly sales data for a five-year period (1996 to 2001) that was developed by the CSCA for an Ontario-based retailer. The geovisualisation software allowed the analyst to define variables to be mapped, the time period, speed of animation, and to navigate the map area interactively whilst viewing the data animation (i.e., utilising a dynamic map representation). This provided the opportunity to search for spatial patterns and trends over a range of 'critical' operating periods during the five-year animation sequence. For example, analysing the weeks (year-over) in the run up to and after the peak December–January sales period. Plate 3 illustrates the significant effect of sales in December 2000 – the retail millennium factor – with sales dramatically increasing in December 2000 (far more than would be expected in a typical December trading period). Spatial–temporal variations in total sales provided a base level of information. The geovisualisation approach was also used to analyse spatial variation by product category over time, providing insight for merchandise-mix decisions and local area promotional campaigns. Studies to date on the merits of three-dimensional mapping and animation vary significantly, some advocate such techniques, while others highlight their limitations (Morrison *et al.*, 2000; Robertson *et al*., 1999; Openshaw *et al*., 1993; Slocum *et al*., 1990). In this example, the adoption of a geovisualisation approach provided additional insight into total and category sales performance across a network of stores in Ontario, revealing patterns that would otherwise have remained hidden within the data.

2.7 Conclusions

The five examples presented in this chapter provide insights into the current use of GIS for retail applications within a North American context. They have detailed the use of spatial data and associated technologies at a variety of levels of sophistication, and reflect the broad spectrum of GIS application. These include, at the simplest level, traditional market mapping and spatial inventory reporting. Increasingly more organisations are applying a spatial approach to augment their decision support activities, for example, assessing the impact of new retail formats, evaluating retail dynamics and developing spatial management systems. More recently, the potential use of data mining and geovisualisation are being evaluated by some major retail corporations. However, what is apparent is the increasing adoption and awareness of the value of GIS by more retail organisations, but the potential benefits to be gained

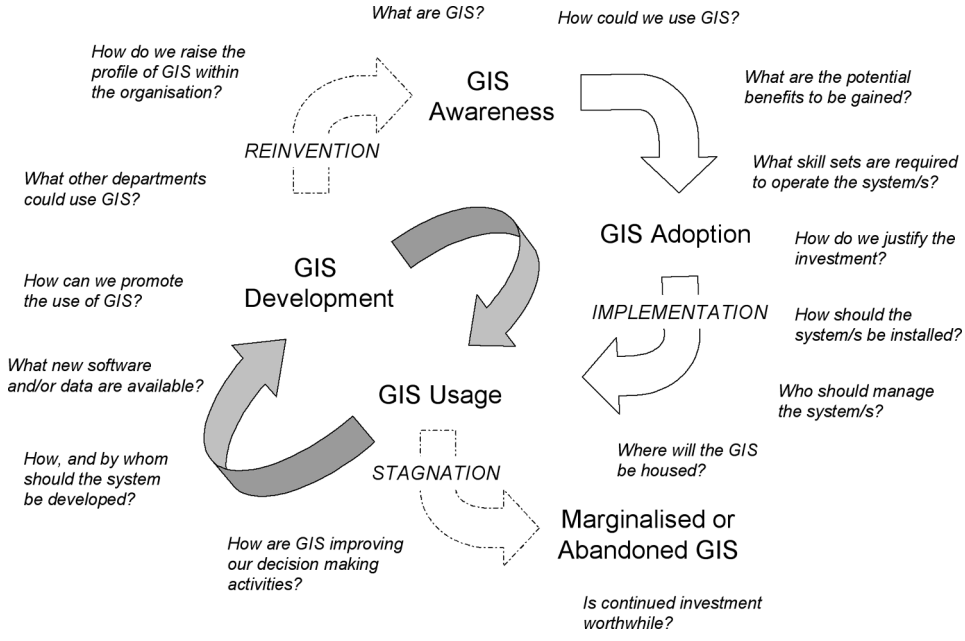


Figure 2.8 *GIS adoption history in retail organisations*

from using GIS technologies and spatial modelling are not always easily recognised. Advocates of these spatial approaches within organisations are still required if various applications of spatial models are to become common fixtures in the retail sector (Hernandez, 1999). Our assessment of the use of GIS and associated technologies is presented in Figure 2.8. In order for GIS to be successfully introduced, operated and embedded within a retailer's decision support systems, the technology requires careful planning, monitoring and management. The decision-making reality for many businesses remains, and March (1991) has summarised that they: (a) gather information but do not use it; (b) ask for more and ignore it; and, (c) gather and process a great deal of information that has little or no relevance to decisions. Traditional methods of spatial data analysis have not facilitated interactive exploration of data, and served to isolate the decision maker as the recipient of static information as opposed to an analytical actor gathering information for knowledge creation. Technological developments have enabled the development of new 'visual' spatial approaches to decision support that aim to harness the intuitive cognitive powers of decision makers and, more generally, promote the adoption and use of GIS.

If the spatial approach is to become a central element in the decision systems of major retail organisations a number of issues must be addressed. Foremost, GIS technology and spatial analysts must have access to and utilise 'real-time' data. Retail organisations must react with increasing speed to changes in the retail economy. As a consequence, the ability to integrate current data (whether it relates to sales levels, competitive effects or reliable census estimates) into the decision making of the

organisation, will have a major impact on the future adoption of GIS. In addition, there must be a conscious effort to link current GIS software with more sophisticated statistical models. Without this ability the value of GIS to any organisation will be limited, as indicated further in Chapter 3 of this volume. There are obvious extensions of spatial analysis and GIS that can be developed in areas that related to data visualisation and data mining. Here, the opportunity exists for spatial analysts to distil and summarise the large databases that relate to customer sales that reside in most retail organisations but are typically not analysed in any systematic manner. Finally, the incorporation of satellite imagery and the potential of GPS-based data to improve the precision of our spatial data is another area that offers potential for the future. However, what offers the greatest source of optimism is the growing awareness of GIS in the retail industry. The term GIS is now recognised by most retail organisations. Courses in business geographics are being incorporated into many university and college business courses and business geomatics has become a recognised field of enquiry.

Glossary of Key Terms

Big box – a large-format retailer operating with depth and breadth in a well-defined merchandise category, e.g. books, sporting goods and home improvement.

Power centre – two or more big boxes sharing the same parking facilities and tenants of the same development.

Power node – a grouping in proximity to a major road/highway intersection of at least two power centres.

Market share – a measure that captures the relative share of total sales allocated to a given retailer or retail destination.

Market penetration – a measure that captures the proportion of actual against potential customers for a given area, typically based on household or population counts.

References

Acevedo, W. and Masuoka, P. (1997) Time-series animation techniques for visualizing urban growth, *Computers & Geosciences*, **23**(4): 423–35.

Andrienko, G.L. and Andrienko, N.V. (1999) Interactive maps for visual data exploration, *International Journal of Geographical Information Science*, **13**(4): 355–74.

Bishop, I.D., Ramasamy, S.M., Stephens, P. and Joyce, E.B. (1999) Visualization of 8000 years of geological history in Southern India, *International Journal of Geographical Information Science*, **13**(4): 417–27.

Blok, C., Köbben, B., Cheng, T. and Kuterema, A.A. (1999) Visualization of relationships between spatial patterns in time by cartographic animation, *Cartography and Geographic Information Science*, **26**(2): 139–51.

Brown, J., Earnshaw, R., Jern, M. and Vince, J. (1995) *Visualization: Using Computer Graphics to Explore Data and Present Information*, Wiley, New York.

Business Geographics (1993–2001) Fort Collins, Colorado: GIS World.

DiBiase, D., MacEachren, A.M., Krygier, J.B. and Reeves, C. (1992) Animation and the role of map design in scientific visualization, *Cartography and Geographic Information Systems*, **19**(4): 201–14, 265–6.

Dorling, D. (1992) Stretching space and splicing time: from cartographic animation to interactive visualization, *Cartography and Geographic Information Systems*, **19**(4): 215–27, 267–70.

Flowerdew, R. and Goldstein, W. (1989) Geodemographics in practice: developments in North America, *Environment and Planning A*, **21**: 605–16.

Goodchild, M.F. (2000) The current state of GIS and spatial analysis, *Journal of Geographical Systems*, **2**: 5–10.

Haeberling, C. (1999) Symbolization in topographic 3D maps: conceptual aspects for user-oriented design, in *Proceedings of the 19th International Cartographic Conference*, Ottawa, Canada, Section **7**: 62–9.

Hernandez, T. (1999) GIS in retailing in Canada, the UK and the Netherlands, *Research Report 1999-10*, Centre for the Study of Commercial Activity, Ryerson University, Toronto.

Hernandez, T. and Biasiotto, M. (2001) The status of corporate retail planning in Canada, *Research Report 2001-3*, Centre for the Study of Commercial Activity, Ryerson University, Toronto.

Huff, D. (2000) *Huff's Market Area Planner*, DataMetrix Inc., Tulsa, Oklahoma.

Jones, K., Pearce, M. and Biasiotto, M. (1995) The management and evaluation of shopping centre mall dynamics and competitive positioning using a GIS technology, *Journal of Shopping Center Research*, **2**: 49–82.

Koussoulakou, A. and Stylianidis, E. (1999) The use of GIS for the visual exploration of archaeological spatio-temporal data, *Cartography and Geographic Information Science*, **26**(2): 152–60.

Kraak, M. (1994) Interactive modelling environment for three-dimensional maps: functionality and interface issues, in MacEachren, A.M. and Taylor, D.R.F. (eds) *Visualization in Modern Cartography*, Pergamon, Oxford, pp. 269–85.

Ledbetter, M. (1999) 3-D visualization helps solve real-world problems, *GEOWorld*, **12**(9): 52–4, 56.

Lloyd, R. (1997) *Spatial Cognition, Geographic Environments*, Kluwer Academic Publishers, Dordrecht.

MacEachren, A.M. (1995) *How Maps Work: Representation, Visualization, and Design*, The Guildford Press. New York.

MacEachren, A. and Kraak, M. (2001) Research challenges in geovisualisation, *Cartography and Geographic Information Science*, **28**(1): 1–11.

March, J.G. (1991) How decisions happen in organisations, *Human Computer Interaction*, **6**(2): 97–117.

Morrison, J.B., Tversky, B. and Betrancourt, M. (2000) Animation: does it facilitate learning?, in *AAAI Spring Symposium on Smart Graphics*, Stanford, California, pp. 53–9.

Openshaw, S., Waugh, D. and Cross, A. (1993) Some ideas about the use of map animation as a spatial analysis tool, in Hearnshaw, H.M. and Unwin, D.J. (eds) *Visualization in Geographical Information*, Wiley, Chichester, pp. 131–8.

Reid, H.G. (1993) Retail trade, in Castle, G.H. (ed.) *Profiting from a Geographic Information System*, GIS World Inc., Fort Collins, Colorado.

Robertson, G.G., Card, S.K. and Mackinlay, J.D. (1999) Information visualization using 3D interactive animation, in Card, S.K., Mackinlay, J.D. and Shneiderman, B. (eds) *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann, San Francisco, pp. 515–29.

Slocum, T.A. (1999) *Thematic Cartography and Visualization*, Prentice Hall, Upper Saddle River, New Jersey.

Slocum, T.A. and Egbert, S.L. (1993) Knowledge acquisition from choropleth maps, *Cartography and Geographic Information Systems*, **20**(2): 83–95.

Slocum, T.A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D., Fuhrmann, S. and Hedley, N. (2001) Cognitive and usability issues in geovisualisation, *Cartography and Geographic Information Science*, **28**(1): 61–75.

Slocum, T.A., Robeson, S.H. and Egbert, S.L. (1990) Traditional versus sequenced choropleth maps: an experimental investigation, *Cartographica*, **27**(1): 67–88.
Yeates, M., Charles, A. and Jones, K. (2001) Anchors and externalities, *Canadian Journal of Regional Science*, **24**(3): 456–84.
Yeates, M. (2001) Business/retail geomatics: a developing field, *Canadian Journal of Regional Science*, **24**(3): 375–86.

3

Using Spatial Models to Solve Difficult Retail Location Problems

Mark Birkin, Graham Clarke, Martin Clarke and Richard Culf

Abstract

This chapter looks at the experience at the University of Leeds of applying spatial interaction models in a number of retail business scenarios. The argument is that retail conditions in the real world present a number of important challenges for our models. In some cases, the existing models are sufficiently robust to deal with these challenges. In other cases, we need to refine the models so that they capture more complex types of consumer behaviour. Three case study clients are presented here. The business concerns of each vary considerably. The chapter thus moves from the classic use of the models for new store openings, through to refinements associated with modelling flows and impacts of branches within shopping centres, dealing with elastic demand, incorporating shopping from the workplace and finally, through to an alternative spatial modelling framework to handle very complicated consumer markets such as petrol retailing. Hence, we conclude that applied modelling is as likely to enrich our theoretical development work as it is simply to demonstrate the usefulness of our existing methods. The link between applied and theoretical work is thus much stronger than many researchers appreciate.

3.1 Introduction and Client Needs

Spatial interaction models have a rich tradition in the geographical literature on retail market analysis. In recent years, there has been increasing commercial interest in the

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

application of these techniques within real business environments. Such was the commercial interest in models developed by the University of Leeds that a unique university company (GMAP) was formed to professionalise the links between academia and the outside world. The range of applications and models used has been described elsewhere (Clarke and Clarke, 2001; Birkin and Culf, 2001; Birkin *et al*., 1996, 2002a). The objective of this chapter is to reflect on the experience of developing and transferring models from an academic into a commercial environment. We will describe some of the technical developments which have been found necessary to make the models work in the real world, and to review, with examples, the degrees of success which have been achieved. It will be shown that in most cases, the models require significant extension of the basic concepts. However, in certain cases, we have concluded (reluctantly) that no amount of re-engineering can tailor the modelling approach to the requirements of difficult retail markets. In these cases, we believe that a multivariate approach to rating sites is the most sensible strategy. An alternative modelling framework of this type is also described and evaluated.

To explore these issues, we introduce three clients from the range of many that have worked with the School of Geography, University of Leeds and with GMAP (see Section 3.2). Each client had/has very different requirements in terms of store location research. Two of these are drawn from the financial services industry. The first client is the State Bank of South Australia. This is an example of the most straightforward use of spatial interaction modelling: the desire to use the models to evaluate a number of what-if proposals concerning new branch openings. This application is described in Section 3.3. The second client is a leading financial organisation based in the north of England whose spatial analysis requirements were built around two major issues. The first issue was the impact of new branch openings on its existing client base. This involved a much greater consideration of the success of the models for impact assessment. Second, was the desire to locate a network of automatic telling machines (ATMs) in good or 'optimal' locations. There was also a third concern regarding the evaluation of potential mergers and acquisitions from the perspective of the combined branch network that might be inherited: however, this is dealt with in detail elsewhere (Birkin *et al.*, 2002b). These objectives require a much more critical look at the traditional spatial interaction models typically used for such analyses. The issues these applied problems raised are addressed in Section 3.4. Finally, in Section 3.5, we review an area deemed to be extremely difficult to model – namely the petrol market. Here, the client is Exxon/Mobil. The switch from a traditional modelling approach to a multivariate approach is discussed with reference to the very difficult issues that the petrol market brings to spatial interaction modelling.

3.2 Background to Commercial Modelling

For 20 years, GMAP Limited has successfully applied geographical modelling tools within a business planning context. Clients of GMAP include retail organisations in the broadest sense, from supermarkets (Asda-Walmart, Iceland), high street multiples (Dixons, W.H. Smith) and out-of-town retailers (IKEA, MFI), to travel

agents (Thomas Cook), petrol companies (BP, Exxon/Mobil), banks (Halifax, Alliance & Leicester) and automotive manufacturers and distributors (Ford, Toyota, Volvo, Mazda). The intellectual foundations of this applied research programme reside in the 1960s and 1970s, when gravity and spatial interaction models found widespread application in transportation and land-use planning, and in planning assessments of retail impact analysis (e.g., Boyce *et al.*, 1970, for examples of the former; Foot, 1981, for a review of the latter: see also Clarke and Stillwell, this volume). At this time, however, there was little or no commercial awareness or interest in these models.

Interest in commercial applications of spatial interaction models was more obvious in the 1980s and early 1990s for a number of reasons. From a retail point of view, markets were increasingly competitive, and tending towards saturation, making profitable new sites increasingly difficult to find. From a methodological perspective, the increased availability of data (about customers, markets and competitors), of PCs to process and distribute the data, and of geographical information systems (GIS) for the analysis of data, were all factors providing an impetus to market analysis. These factors were supplemented by an economic boom in the 1980s which led many UK companies to want to expand their networks. This increased interest in modelling was manifest in a number of ways:

- provision of spatial interaction modelling capabilities as standard within GIS systems (e.g., ARC/Info);
- development of specialist market analysis software (e.g., InSite, Environ, MicroVision) by business consultants; indeed, some of these business consultants were new to the scene, specialising in spatial modelling (e.g., GMAP, Geobusiness Solutions); and
- creation of dedicated site analysis and modelling departments within major retail organisations (e.g., Tesco, Marks and Spencer, J. Sainsbury, Whitbread).

There is a reasonable amount of evidence to suggest, however, that increased commercial interest in modelling has not been associated with increased technical sophistication. Despite the increasing availability of spatial interaction models, it is clear that proprietary GIS (normally standard buffer and overlay techniques) remain the most common methodology for site location (if organisations use anything over and above gut feeling – see Hernandez and Bennison, 2000). Reynolds (1991) and Elliot (1991) describe the application of such simple catchment area analysis techniques within GIS. There does not seem much evidence to support more sophisticated use of GIS since then (see Jones and Hernandez, this volume). This is somewhat surprising to the extent that our own experience demonstrates unequivocally that substantial development work has been required to transform spatial interaction models from interesting pedagogic and academic tools into valuable commercial weapons. Similarly, it seemed to be the case that, whilst the number of packages including models was increasing, the concept of the gravity model was much misunderstood and mistrusted in the commercial world. In part this was because the models were synonymous with very aggregate versions of the spatial interaction model, and in many instances simply did not work well enough. Similarly, users often did not have

the experience or expertise to be able to build models from scratch and attempt the detailed disaggregation needed to make the models operational. In the next section of this chapter, we set out the classic model before discussing some of the developments and extensions required to answer the applied problems set by the clients.

3.3 Classic Model

The classic retail spatial interaction model has the following structure:

$$S_{ij} = A_i \times O_i \times W_j \times f(c_{ij}) \tag{3.1}$$

and,

S_{ij} is the flow of people or money from residential area i to shopping centre j;
O_i is a measure of demand in area i;
W_j is a measure of the attractiveness of centre j;
c_{ij} is a measure of the cost of travel or distance between i and j; and
A_i is a balancing factor which takes account of the competition and ensures that all demand is allocated to centres in the region. Formally it is written as:

$$A_i = \frac{1}{\sum_j W_j \times f(c_{ij})} \tag{3.2}$$

In summary, the model assumes that in order to forecast the flow of expenditure from an origin zone i to a destination (retail outlet or centre) j, then one begins with a fixed demand pool at i. This demand is shared between retailers in proportion to their relative size (or other measures of attractiveness), and in proportion to their geographical proximity. Individual parameters are added to these variables so that the model can be fitted (or calibrated) to real-world data (see equations below). At the very heart of what makes spatial interaction modelling so appealing as an analysis tool is that predictions are based on the replication of customer trips between individual origin and destination pairs rather than by using statistical or analogous evidence (see Birkin *et al.*, 2002a for more details on alternative methodologies for store location analysis).

For all real-world applications this model is likely to require disaggregation. This might be by demand, i.e., different parameters of the model are appropriate for different income groups, or age groups – or supply based (a new version of the attractiveness term – see below). As we shall argue below, new trends in retail geography have caused other types of refinement to be introduced.

Before we examine these new model developments, it is useful to present an application used in a classic what-if fashion for the first of our clients to be reported here – the State Bank of South Australia (SBSA). The model was of the type shown in equations (3.1) and (3.2) but disaggregated (in standard GMAP fashion) as follows:

$$T_{ij}^{lk} = \sum_{k \in j} A_i^{lk} O_i^l W_j^{lk} \exp^{-\beta_i^{lk} d_{ij}} \tag{3.3}$$

where:

$$A_i^{lk} = \frac{1}{\sum_{jk} W_j^{lk} \exp^{-\beta_i^{lk} d_{ij}}} \tag{3.4}$$

to ensure that:

$$\sum_{jk} T_{ij}^{lk} = O_i^l \tag{3.5}$$

and:

$$W_j^{lk} = R_j^{lk} Z_j^{lk} \gamma_{l(j)}^{lk(j)} \mu_j^{lk} \chi_j^{lk} \tag{3.6}$$

where:

T_{ij}^{lk} = estimated sales of product l, for organisation k in origin zone i at destination j;
O_i^{kl} = demand for product l in origin zone i;
W_j^{kl} = attractiveness of outlet j selling product l, organisation k;
β_i^{kl} = distance decay parameter disaggregated by origin i, product l, and organisation k;
d_{ij} = drive-time from zone i to outlet j;
R_j^{lk} = attractiveness of outlet j, belonging to organisation k, selling product l;
Z_j^{lk} = size of outlet j (i.e. square feet) belonging to organisation k, for product l;
μ_j^{kl} = parameter reflecting proximity of outlet j belonging to organisation k to surrounding outlets for product l;
χ_j^{kl} = parameter reflecting operating period (i.e. months open/12) for outlet j belonging to organisation k, selling product l; and
$\gamma_{l(j)}^{lk(j)}$ = brand attractiveness of organisation k, selling product l, to customers resident in i.

Figure 3.1 shows the predicted market share of SBSA across the Adelaide Metropolitan Region. It shows that generally SBSA did have a very high market share, but as with most organisations that deliver services from fixed points, that market share varied significantly across the region. The areas of low market share clearly provide an immediate snapshot of market potential. The models were then used in a systematic fashion to explore the impacts of adding a new SBSA branch in all of these market gaps. Figure 3.2 shows the market share before and after the opening of a new branch at St Agnes. It can be immediately seen that market share improved significantly. However, as important is the total revenue predicted for the new branch, and where those sales are likely to come from (including cannibalisation from the

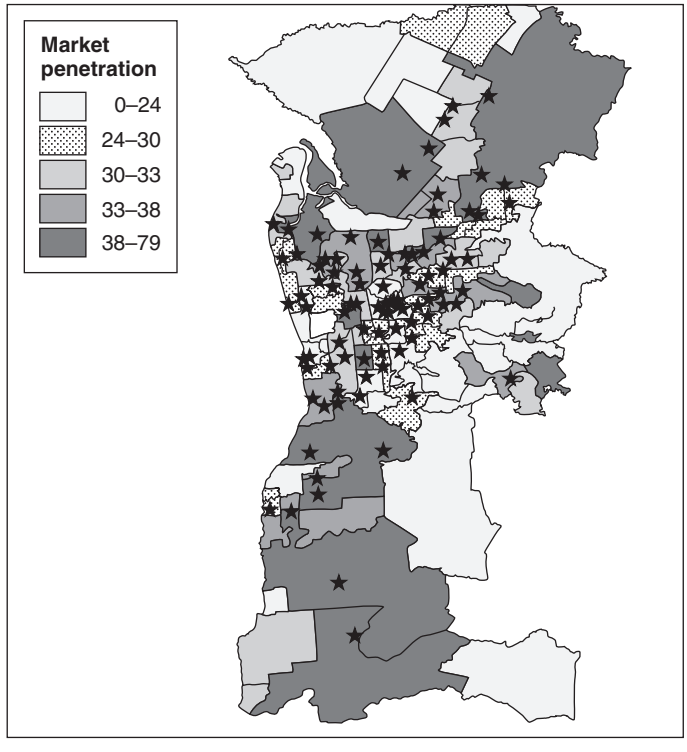


Figure 3.1 *SBSA market share in Adelaide metropolitan market in mid-1990s*

SBSA existing branches). Table 3.1 shows the net change in sales. It can be seen that the impacts on existing SBSA branches are not high and hence this looks like a good new location.

As we noted above, this kind of application was typical of GMAP applications in the 1980s and early 1990s. However, a more sophisticated approach to the art of store location research was required from the mid-1990s onwards. We shall explore the implications of this change in the next section.

3.4 Model Extensions

3.4.1 Introduction

The aim in this section is to explore new challenges in spatial interaction modelling, and store location research more generally, brought about by new types of application and client requirements. First, there has been an increasing concern to explore the predictive power of the models in terms of their ability to represent real deflection rates accurately. Second, there are those applications that have challenged the nature of demand: where does this come from and how can this be incorporated into

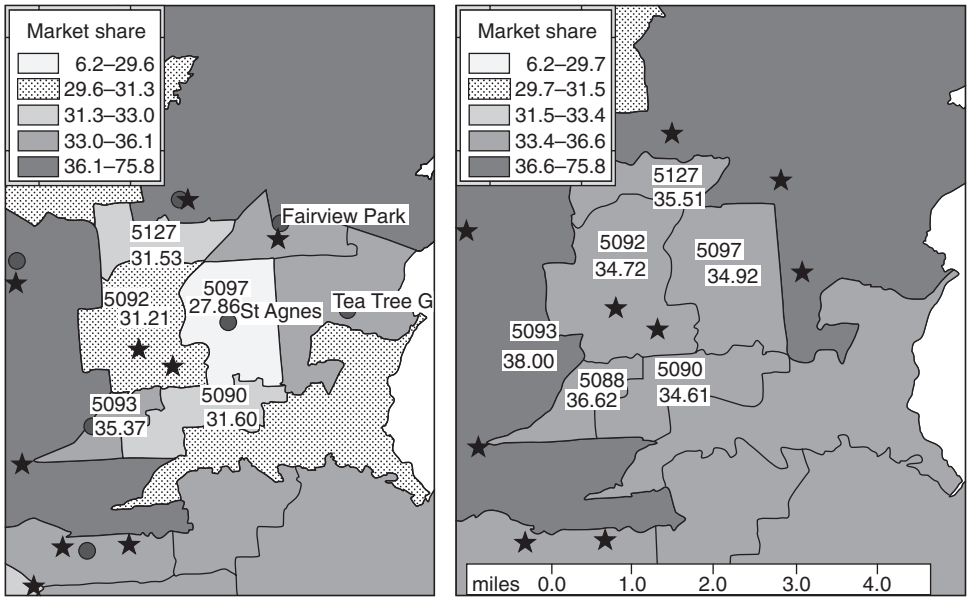


Figure 3.2 *Market shares before and after new branch opening in St Agnes*

Table 3.1 *Impacts on other branches of new branch in St Agnes*

SBSA branches	Net change in sales
Modbury Central	–13
Tea Tree Gully	–10
Clovercrest Plaza	–7
Fairview Park	–6
Golden Grove	–5
Salisbury	–5
Ingle farm	–4
Pava Hills	–3
Nailsworth	–2
Total lost to other SBSA branches	76

the spatial models? Third, there are those applications which seem to challenge the very principles of spatial interaction modelling. We shall look at each in turn.

3.4.2 Modelling Centre Deflections

One of the most interesting retail location issues over the last 20 years or so has been the decline in the number of stand-alone large stores. Although many early

retail warehouses and grocery superstores occupied single-store sites the pattern since the early 1980s has been to locate within centres (warehouse parks and shopping centres often anchored by a major superstore). Similar developments are observable in non-food retail sectors: for example, the trend towards agglomerations of activity in automotive distribution – the so-called auto-mall – and the proliferation of multi-franchised outlets is transforming choosing a new car into something more akin to comparison shopping. For the retailers, whenever outlets are grouped together within city centres or retail parks, then there are potential scale economies to be obtained. In effect, these benefits may be closely related to multi-purpose trip-making, so that customers can also see benefits in visiting more than one destination within a centre or cluster. Alternatively, this may be viewed as a hierarchical choice process in which customers first choose a centre in which to shop, and only then select an outlet within that centre. This structure has been exploited and developed in the creation of the 'competing destinations' model (Fotheringham, 1983, 1986), in which a new variable is included that measures accessibility to other retail outlets. This structure is reproduced within GMAP's generic model, which is explicitly hierarchical in character, so that individual customer flows are modelled to groups of retail outlets (but the attractiveness of the group is related to the attractiveness of its component parts). Individual outlets within each group then compete for market share within the cluster. In a sense, there is also a special case of the model, in which the clusters are constituted from single outlets. This is then the traditional model, which might still be most appropriate for single-purpose trip patterns, say to stand-alone hypermarkets or schools and hospitals in the non-retail sectors.

The increasing use of centres has produced another issue of importance when dealing with classic gravity or spatial interaction models: namely, the way in which the models handle customer deflections when new stores are opened. This discovery was made through working with client B, a northern bank. For this client, financial centres were used instead of individual bank branches. That is, interactions were predicted between residential areas and financial centres (such as town centres containing four or five branches). Revenues to individual branches within the centres was then allocated according to the fascia and its store attractiveness (age of store, size, etc.). The traditional interaction model worked well for predicting revenues and catchment areas for the existing set of stores. However, for new store openings the patterns of deflections seemed intuitively wrong. Let us take the common scenario of client B wishing to open an additional branch in a large town centre where it is already present (along with a number of competitors). All things being equal, we would expect total centre sales to undergo a marginal increase in sales, but we would not expect 400 or 500 new customers suddenly to come in from surrounding towns to go to that new branch. In other words, we would expect sales for the new branch to be generated mainly from the centre's existing branches, deflecting most heavily perhaps from the existing client B branch. What we actually see from the standard model is an increase in centre sales, near incremental sales for the new branch and minimal deflections from the other client B branch in that centre.

In the 'standard' model (Table 3.2, Column 3), the introduction of a new branch has a proportionate effect on the attractiveness of the centre (via an increase in

Table 3.2 *Open a financial service branch in a large centre*

	Baseline	Standard model	GMAP model
Centre sales	18 332	18 782	18 353
Change		450	21
Percentage change		2.4	0.1
Open branch		727	707
Existing branches	5 663	5 578	5 288
Change		85	275
Percentage deflection		1.5	4.8

floorspace – F_j). The extent of the associated sales increase depends on the relative balance between attractiveness and accessibility. In this example the new branch generates 727 new sales, and about two-thirds of these (450) are sales which were previously made in competing centres (for higher values of beta, then we would expect to see less diversion of sales between centres as the respective catchments are more or less fixed; for lower values of beta then sales will follow floorspace, and we would expect more new centre sales). Thus, our new branch takes 450 sales away from branches in neighbouring centres, and the remainder (277) from existing branches within the same centre. The deflections within the existing centre are proportionate to the floorspace of competing financial services organisations. In the baseline, client B has a share of 5663/18 332 = 31%. Therefore client B loses 31% of 277 = 85 sales within the standard model scenario.

In the revised model (Table 3.2, Column 4), the effects are explicitly hierarchical. In the first stage, the flow to centres is modelled as a function of a basket of centre characteristics (quality of access and parking, presence of key retailers and anchor stores, pubs, restaurants and other attractors) in addition to financial services branches. Opening a new financial services branch has only a marginal effect on the character of the centre; hence the vast majority of new sales are at the expense of existing branches within the same centre. A second layer of distinction concerns the loyalty of customers to each financial services brand. Although client B can attract new customers by the addition of new branches, there are diminishing returns to scale – doubling the number of branches will not double sales, for example, because existing customer niches may become saturated. Thus, the deflections from existing branches within the estate of client B are proportionately higher than from its competitors.

In Table 3.3, we can see the same processes for a branch closure within a small centre. In this case, client B does better than expected because many of its customers choose to switch their loyalty to the remaining branches rather than defecting to an entirely new competitor.

Table 3.3 *Close a financial service branch in a small centre*

	Baseline	Standard model	GMAP model
Centre sales	4657	4157	4588
Change		500	69
Percentage change		10.7	1.5
Closed branch	558	0	0
Remaining branches	733	751	837
Change		18	104
Percentage retention		3.2	18.6

The centre effects can be expressed through a revised set of model equations:

$$T_{ij}^k = A_i^k O_i^k \theta_j^k W_j \exp^{-\beta_i^k d_{ij}} \tag{3.7}$$

where:

$$A_i^k = \frac{1}{\sum_j W_j \exp^{-\beta_i^k d_{ij}}} \tag{3.8}$$

to ensure that:

$$\sum_{jk} T_{ij}^k = O_i \tag{3.9}$$

where W_j is now a measure of the attractiveness of centre j:

$$W_j = f\left(X_j^1, X_j^2, \ldots, X_j^n\right) \tag{3.10}$$

X_j^k, $k = 1, \ldots, n$ are variables which contribute to centre attractiveness.

Flows are allocated in proportion to market share within a centre:

$$\theta_j^k = f\left(V_j^{k1}, V_j^{k2}, \ldots, V_j^{km}; W_j\right) \tag{3.11}$$

where V_j^{kl}, $l = 1, \ldots, m$ are attributes for retailer k in centre j.

Ongoing research is also examining other issues to do with impact assessment. Using a set of traditional retail models for the identification of areas poorly served by the major food retailers (so-called 'food deserts') produced a standard set of deflection rates when a new Tesco store was opened in East Leeds (Clarke *et al.*, 2002). The model predicted deflection rates of up to 60% for those stores in close

proximity. However, this case study was unusual in that actual consumer deflections were collected as part of the overall food deserts project. The results of this major household survey (1000 households in the vicinity of the new store) showed more complex deflection patterns than predicted in the model (Wrigley *et al.*, 2003). Two trends were particularly noticeable. First, there was more loyalty to the discount sector than predicted by the models. It seems that some consumers felt uncomfortable in the new Tesco store – they felt more at home remaining in the discount store. Second, there was a greater switching from Tesco's other store in the region. That is, the models under-predicted the amount of switching from the rather distant centre at Crossgates (home of the other Tesco store). These issues are driving the examination of new consumer behaviour variables in the retail grocery model (Clarke *et al.*, 2003). This is another good example of the importance of applied research in terms of driving theoretical extensions of our models.

3.4.3 Incorporating Elastic Demand

A second major issue to emerge from working with client B was the concept of elastic demand. In practice, the assumption of a fixed demand pool (O_i) is not always a sufficient one. This is particularly the case when the levels of demand and supply for a product or service are interrelated. In the case of groceries, for example, it is safe to assume that customers have a reasonably fixed requirement for fruit and vegetables, meat, dairy products and so forth. These requirements are unlikely to be altered by the decision of Safeway to build a new supermarket on the doorstep. That is, although consumers may switch allegiance and frequent the new store, their average weekly budget for food is unlikely to increase significantly. On the other hand, providers of leisure services such as cinemas, sports clubs or restaurants would do well to recognise that these facilities will to some extent generate their own demand. People lacking access to a cinema, for example, will tend simply to substitute other activities, and will have a lower demand for cinema tickets than similar people who have easy access to cinemas.

Modelling frameworks for the treatment of elastic demand have been proposed by Pooler (1994) and Ottensmann (1997). The principle of these applications is that the level of demand in an area is not fixed, but is related to some measure of service provision in that area. The relationship may be regulated through the introduction of a new parameter, which can adopt some very low or zero value for activities such as the sale of retail groceries, to much larger values for a restaurant or cashpoint machine. The latter is the example emanating from client B. Figure 3.3 plots the relationship between ATM location and consumer residence in Leeds.

There are many approaches to the problem of coping with a demand base which is not fixed. One strategy is to develop a hybrid model, since the flows can no longer be strictly considered as origin-constrained. An alternative is to specify the model via a two-stage process: first, estimate the demand; second, allocate the flows. So we might start with:

$$O_i = e_i P_i w_i^{\phi} \Big/ \hat{w}_I^{\phi} \tag{3.12}$$

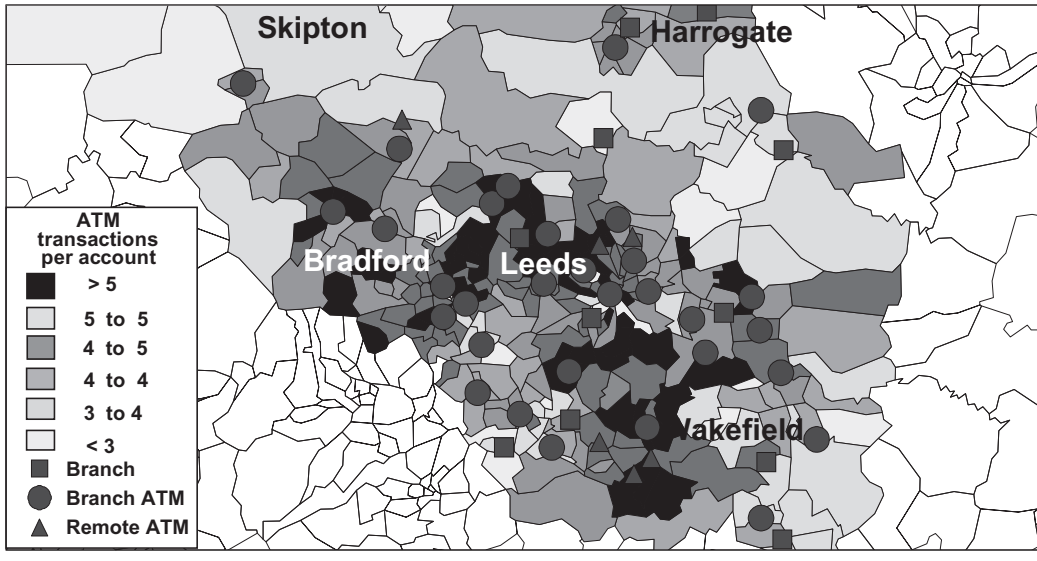


Figure 3.3 *Variations in ATM usage by Leeds residents*

where:

e_i = estimated sales of product l, for organisation k in origin zone i at destination j;
P_i = demand for product l in origin zone i; and
w_i = measure of access to services for residents of zone i, such as the Hansen accessibility:

$$w_i = \sum_{jk} W_j^k \exp^{-\beta_i^k c_{ij}} \tag{3.13}$$

and where:

$\hat{w}_l$ = average accessibility across the whole region; and
ϕ = is a parameter.

Having estimated the demand, we may then continue as previously with the traditional spatial model.

3.4.4 Trips from Work

Even if it is safe to assume that the demand for products and services is relatively fixed, the assumption of a simple flow from the point of consumption (residence) to the point of sale (outlet or retail centre) is an increasingly unsophisticated representation of the consumption process. To an increasing extent, retail consumption is

satisfied through a complex web of multi-purpose trips, including trips to or from places of work, education or leisure.

One implication is that the demand pool needs to be segmented, at least into a portion which may be satisfied from the home, and another portion which is satisfied from the workplace. A journey to work model may be the most appropriate way of assigning customers to workplace locations since the buoyancy of retail sales within a typical city centre owes as much to the proximity of workers as to its attractiveness to suburban neighbourhoods. Our experience suggests that when analysing model outputs in the case of W.H. Smith town or city centre branches, somewhere between a quarter and a third of retail expenditure for these products was being released through the workplace rather than the residence. However, such work-based trips are also highly constrained in the geographical sense, so there will not be many trips of four or five miles undertaken within the lunch hour. There may also be an element of elasticity here between the demand components, so that householders employed within city centres may tend to spend heavily through the working week; others may be restricted to the usual evenings and weekends. This is one of the features which makes petrol forecourts so difficult to model, as we will see in the next section below. Rather than drawing from a 'captive' local market, customers will typically be intercepted at the mid-point on trips which might connect the home, workplace, shop, cinema, nightclub, holiday resort, or any other of a myriad of possible locations.

3.4.5 Conclusions

The implementation of these various model enhancements has been facilitated by the production of a combined software module incorporating a range of application building blocks. This is somewhat more complex than the basic spatial interaction model which was introduced in Section 3.2 above. Thus while a basic spatial interaction model simulation might be effected through less than a page of FORTRAN code, the 'generic model' is composed of several thousand lines of C++ code, because the model is designed to deal with many different data structures and model specifications within different markets. This latter point is illustrated in Table 3.4, which makes explicit some of the issues which we have already discussed above, i.e., that different approaches are more relevant in some circumstances than others. For example, as we saw earlier, elasticity of demand is far more important in the provision of financial services than motor vehicles.

3.5 Approaches to Modelling Difficult Markets

Some retail markets are evidently more difficult to model than others. This idea is summarised in Figure 3.4. Excellent results have been achieved in forecasting outlet revenues for retailers such as supermarkets and DIY retailers. For example, Birkin *et al.* (1996) report on an application to the DIY sector in which average forecasting errors of around 5% were consistently achieved. Of equal importance is the ability

Table 3.4 *Importance of model drivers by retail market*

	Supermarket	Retail	Auto	Finance	Petrol
Spatial aggregation	Fine zones	Medium zones	Coarse zones	Medium zones	Medium zones + network
Brand loyalty	Moderate/ Strong	Moderate	Moderate	Strong (transactions)	Moderate
Attraction components					
Space	5	3	4	1	4
Parking	5	1	3	1	5
Accessibility	5	4	2	3	5
Product range	4	2	4	2	1
Prices	5	5	5	5	5
Adjacencies	1	5	1	5	2
Opening hours	4	3	1	3	5
Complexity of distribution	Simple/ Moderate	Simple/ Moderate	Simple/ Moderate	Complex	Simple
Travel distance	Short	Moderate	Long	Moderate	Long
Trip type	Single purpose	Multi-purpose	Single purpose	Multi-purpose	Distress
Segmentation	Important	Fundamental	Important	Moderate	Marginal

1 = Relatively unimportant
5 = Very important

of the models to produce accurate forecasts of the impact of new store developments on an existing portfolio of stores, and to identify the geographic extent and composition of the store catchment for advertising and promotion. Certain other markets are less tractable. A notorious example is the car market. Here the issue is not so much about the ability of the models to generate accurate estimates of 'market potential', but the fact that the actual performance of dealers varies quite strongly from place to place (not least because of the persuasiveness of individual salespersons!).

In this section we draw upon experience built up from client C – Exxon/Mobil. In the petrol sector, the issues of complexity are more fundamental. Here the normal linkage between an outlet and its local market is less strong, and the range of factors which influence revenue generation is very large. That said, using standard modelling procedures had produced some impressive results in the petrol sector. Table 3.5 shows the results from a model run in Sicily. However, as can be seen, the error margins are much higher than we would expect in other sectors, and there tends to be more outliers. Those petrol stations with particularly poor model predictions turned out to be in key transient locations. To build a model which captures transient demand as well as local demand would need a full transport network model. The daily effects of price changes are also a difficult marketing activity to capture within the models. The petrol

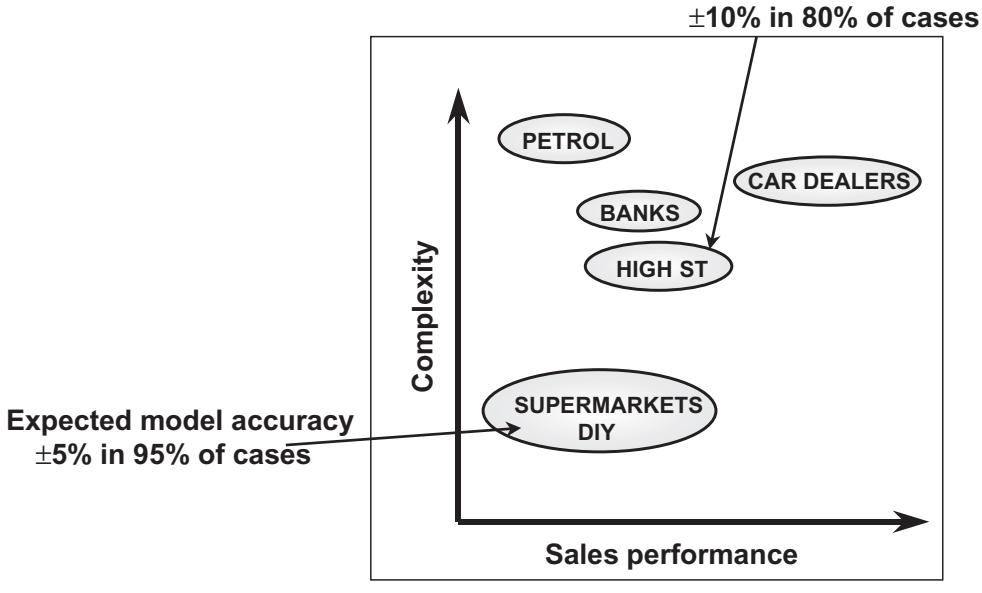


Figure 3.4 *Effectiveness of forecasts by market*

Table 3.5 *Observed versus predicted sales volumes at selected petrol stations in Sicily*

Petrol station location	Actual volume	Predicted volume	Performance index
Corso Tukory	1510	1663	0.91
Piazza Giacchery	1860	1600	1.16
Via Uditore	1002	1050	0.95
Vilae Regina Margherita	1424	1000	1.42
Corso Dei Mille	956	900	1.06
Viale Michelangelo	1089	1150	0.95
Via Aurispa	763	600	1.27
Via Sciuti	1980	2141	0.92
Niscemi	1250	1350	0.93

market is one of the most susceptible to small variations in price. Although it is relatively simple to build in a price variable to standard models (cf. Wilson, 1985), modelling the impacts of daily price changes on both local and transient demand is a more difficult issue.

In these circumstances, we have experimented with the development of multivariate statistical approaches to revenue modelling. The core of our approach is to prepare 'site ratings' for each location. There are two steps to this process: first, to identify the most important factors and to calibrate their relative importance; second, to construct individual scores for different attributes or outlet configurations. Note that while there are some similarities to straightforward regression analysis in this

Table 3.6 *Components of a site rating system*

Factor	Maximum score
Quantity	45
Quality	35
Competition	15
Concentration	5
Total	100

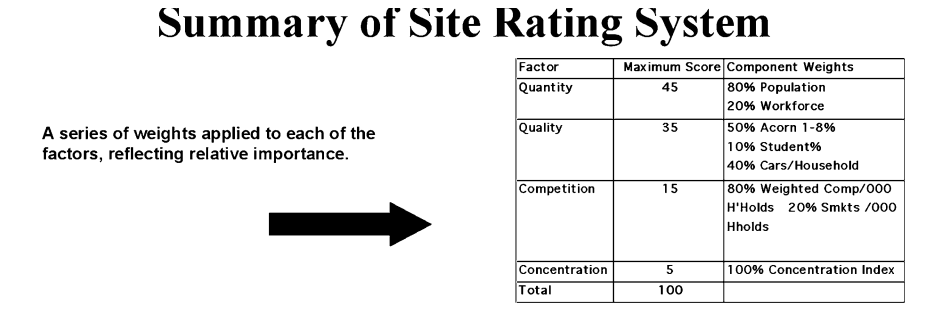


Figure 3.5 *Rating summary*

approach, and that indeed multiple regression may be used in the process of deriving the individual scores, the approach is more closely related to credit scoring (Leyshon and Thrift, 1999), and indeed to business scorecards (Ahn, 2001) in which the performance of an organisation might be evaluated against a whole raft of criteria.

Consider the example shown in Table 3.6. Here four sets of factors which drive convenience store ('C-Store') performance on petrol forecourts have been identified. Note that despite our comments above, the local trade area (or quantity of demand) is still the most important factor (generating both drive-in and walk-in traffic to forecourt shops) but accounts for less than 50% of the overall site score.

Further description of these components is provided in Figure 3.5. For example, the competitiveness of markets is assessed according to the presence of retailers of different types. Each type is in turn weighted according to its relative impact, with supermarkets having the most effect, and other forecourts the least. Summary results are shown in Figure 3.6. A typical application of this type of model would be

Figure 3.6 *Sample results for a site rating system*

expected to explain between 75 and 80% of the variance in C-Store performance. The tools are being extensively used to prioritise investment from a local to a global scale (i.e., not just Otley versus Ilkley, but Mexico versus Thailand!).

It may of course be possible to look for new quantitative methods to solve these difficult problems. The authors have experienced some success with genetic algorithms. These evolutionary algorithms are particularly suitable for network optimisation problems: for example, how to configure a new network of 50 car dealers or supermarkets in an Eastern European country (such problems are being faced with increasing frequency by growing retail groups in the UK and the USA). A conventional solution to problems of this type is described by Birkin, Clarke and George (1995), and a GA version by Birkin and Culf (2001). Unsupervised neural networks are also thought to be useful for problems of pattern recognition. In an abstract sense, the problem of rating sites is of this type, in which one needs to find the pattern of factors and attributes which determines outlet performance. The application of self-organising maps in the style of Kohonen (1984) (cf. Openshaw and Openshaw, 1997) has been used by GMAP to generate more effective solutions to the site rating problem described above. The weakness of this approach is that the ratings tend to be less 'clean' than those prepared through desktop analysis. In practice, a hybrid approach in which the results from some neural network-based approach are then manually tidied up has been found to be most effective (see Birkin, Boden and Williams, 2003).

Consider a simple example. We wish to rate sites according to the number of petrol pumps (two, four or six) and their accessibility (on an integer scale from one to five). The parameters to be optimised via the neural network are the importance of accessibility (relative to number of pumps), and the relative weightings between each pump size. Thus the model would be:

$$R_i = \sum_k \delta(1)_i^k a^k + \gamma \sum_m \delta(2)_i^m b^m \tag{3.14}$$

where

R_i is the 'rating' of petrol station i;
a^k is a size parameter;
b^m is an accessibility parameter;
$\delta(1)_i^k$ is a zero-one variable, one if site i has k pumps, zero otherwise;
$\delta(2)_i^m$ is a zero-one variable, one if site i has an accessibility score of m, zero otherwise; and
γ is a scalar.

An ideal solution might look like this:

$a_1 = 0.77$
$a_2 = 1.02$
$a_3 = 1.48$
$\gamma = 0.43$
$b_m = 0.48, 0.56, 0.84, 1.19, 1.21.$

Whilst this set of scores represent the best available fit to the data, it will often make sense to smooth them into a form which is slightly less accurate but intuitively pleasing and simpler to describe, for example:

Pumps – max 20 points
  2 pumps – 8 points
  4 pumps – 12 points
  6 pumps – 20 points
Accessibility – max 10 points
  1 – 3 points
  2 – 4 points
  3 – 6 points
  4 – 9 points
  5 – 10 points

In this example, if there are an equal number of sites in each category then the simple model retains 99% of the variance in the precise estimates, but involves no decimal points!

3.6 Conclusions

The application of theoretical models within an applied commercial environment has generated many extensions and enhancements. It has also posed challenging problems. To solve these problems may require imaginative extensions of an existing

framework, the introduction of completely new techniques, or a recourse to old-fashioned but robust methods. A distinguishing feature of business is that novelty counts for less than performance. As the geomodelling industry continues to mature we should expect to see a greater plurality of approaches, providing demonstrable benefits to an increasingly wide portfolio of business customers.

References

Ahn, H. (2001) Applying the balanced scorecard concept: an experience report, *Long Range Planning*, **34**(4): 441–61.

Birkin, M., Boden, P. and Williams, J. (2003) Spatial decision support systems for petrol forecourts, in Geertman, S. and Stillwell, J. (eds) *Planning Support Systems in Practice*, Springer, Berlin.

Birkin, M., Clarke, G.P., Clarke, M. and Wilson, A.G. (1996) *Intelligent GIS: Location Decisions and Strategic Planning*, Geoinformation, Cambridge.

Birkin, M., Clarke, G.P. and Clarke, M. (2002a) *Retail Geography and Intelligent Network Planning*, Wiley, Chichester.

Birkin, M., Clarke, G.P. and Douglas L. (2002b) Optimising spatial mergers: commercial and regulatory perspectives, *Progress in Planning*, **58**(4): 229–318.

Birkin, M., Clarke, M. and George, F. (1995) The use of parallel computers to solve non-linear spatial optimisation problems: an application to network planning, *Environment and Planning A*, **27**: 1049–68.

Birkin, M. and Culf, R. (2001) Optimal distribution strategies: one company's experience in the application of regional science techniques to business planning problems, in Clarke, G.P. and Madden, M. (eds) *Regional Science in Business*, Springer, Berlin.

Boyce, D., Day, N. and McDonald, C. (1970) *Metropolitan Plan-making*, Monograph Series Number 4, Regional Science Research Institute, Philadelphia.

Clarke, G.P. and Clarke, M. (2001) Applied spatial interaction modelling, in Clarke, G.P. and Madden, M. (eds) *Regional Science in Business*, Springer, Berlin.

Clarke, G.P., Eyre, H. and Guy, C. (2002) Deriving indicators of access to food retail provision in British cities: studies of Cardiff, Leeds and Bradford, *Urban Studies*, **39**(11): 2041–60.

Clarke, G.P., Douglas, L. and Guy, C. (2003) The impact of new store developments on urban food deserts, *Working Paper*, School of Geography, University of Leeds (forthcoming).

Elliott, C. (1991) Store planning with GIS, in Cadeau-Hudson, J. and Heywood, D.I. (eds) *Geographic Information 199*1, Taylor and Francis, London.

Foot, D. (1981) *Operational Urban Models*, Methuen, New York.

Fotheringham, A.S. (1983) A new set of spatial interaction models: the theory of competing destinations, *Environment and Planning A*, **15**(1): 15–36.

Fotheringham, A.S. (1986) Modelling hierarchical destination choice, *Environment and Planning A*, **18**: 401–18.

Hernandez T. and Bennison D. (2000) The art and science of retail location decisions, *International Journal of Retail, Distribution and Management*, **28**(8/9), 357–67.

Kohonen, T. (1984) *Self-organisation and Associative Memory*, Springer-Verlag, Berlin.

Leyshon, A. and Thrift, N. (1999) Lists come alive: electronic systems of knowledge and the rise of credit-scoring in retail banking, *Economy and Society*, **28**(3): 434–66.

Openshaw S. and Openshaw C. (1997) *Artificial Intelligence in Geography*, Wiley, Chichester.

Ottensmann, J.R. (1997) Partially constrained gravity models for predicting spatial interactions with elastic demand, *Environment and Planning A*, **29**: 975–88.

Pooler, J. (1994) A family of relaxed spatial interaction models, *Professional Geographer*, **46**: 210–7.

Reynolds, J. (1991) GIS for competitive advantage: the UK retail sector, *Mapping Awareness*, **5**(1): 33–6.

Wilson, A.G. (1985) Structural dynamics and spatial analysis; from equilibrium balancing models to extended economic models for both perfect and imperfect markets, *Working Paper 431*, School of Geography, University of Leeds.

Wrigley, N., Warm D. and Margetts, B. (2003) Deprivation, diet and food retail access: findings from the Leeds 'food desert' study, *Environment and Planning A*, **35**(1): 151–88.

4

Location-Based Services for WAP Phone Users in a Shopping Centre

António S. Câmara and António Eduardo Dias

Abstract

Advances in mobile Internet devices together with the recent availability of location awareness for mobile phones represent new opportunities for the implementation of location-based services (LBS). In the commerce arena, LBS can have a crucial role in combining virtual and real infrastructures, thus improving the experience of visiting a commercial area like a large shopping centre. To support these synergies between real spaces and their symbolic representations, a new personal navigation technology has been developed based on mobile phones named DiWay. DiWay facilitates user navigation and bridges the gap between traditional 'bricks and mortar' commerce and electronic or mobile commerce by the use of anchor points. The most relevant of these can be memorised using the phone, a concept called Phonemarks. Furthermore it also provides security and community applications. DiWay is built upon the FluidMapping software framework that facilitates the use of heterogeneous databases and displays the results in any presentation device, including 2.5 or 3G mobile phones. The system is being deployed for one of the largest shopping centres in Europe, the Centro Colombo in Lisbon.

4.1 Introduction

Maps are symbolic representations of real spaces. The gap between reality and its symbolic representations has been shortened in recent years with digital multimedia

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

information systems. Such systems have included maps, aerial and ground photos, video and sound. Reviews of available spatial multimedia systems may be found in Câmara and Raper (1999), Gouveia and Câmara (1999) and Câmara *et al.* (2000). Mobile devices and particularly Internet-enabled mobile phones are helping to bridge that gap further by enabling the confrontation of reality with their digital representations in real time.

Such devices are creating opportunities for new LBS that rely on having a spatial background usually in map form with geo-referenced points of interest and other relevant information. LBS also depend on knowing the position of the user in space. LBS applications for the common mobile phone user include personal and car navigation, point-of-interest search, e-commerce, emergency services and entertainment. LBS are supported by context-aware systems relying on the dissemination of contextual information (Dias *et al.*, 1998; Ryan *et al.*, 1998; Ryan, 1998; Abowd *et al.*, 1997a). Their main focus regarding context is on location, thus falling into the field of location-aware computing systems. These systems have been thoroughly discussed in Dias *et al.* (1996, 1997), Davies *et al.* (2001), Beadle *et al.* (1997, 1998), Abowd *et al.* (1997), Frisk (1998) and Pradhan *et al.* (2001). A good review of location systems for location-aware applications can be found in Hightower and Borriello (2001).

The Colombo Centre (http://www.colombo.pt) is the largest shopping centre located in the Iberian Peninsula. It has over 40 million visitors a year, 420 shops, 65 restaurants, 10 movie theatres and many other facilities. The parking lot has capacity for 6000 cars. In such a large shopping centre, guidance is needed to locate shops and products. A system that facilitates car finding in the parking lot may be also helpful. The need for systems, minimising those problems, is well understood by the companies that manage shopping centres. SONAE, the company that runs Colombo Centre, has developed paper-based diagrammatic maps and signs available in several locations to help locate stores. It has also created a system based on the parking cards and fixed machines that stamp the position of a car in the parking lot.

Both systems have shown limitations as follows:

- paper-based representations cannot be queried and are usually thrown away after use. Moreover, any change in the Colombo Centre's structure requires new representations. These two issues bring up both financial and environmental problems;
- diagrammatic maps also represent a level of aggregation not suitable to detailed information needs from the customers, which are often centred on products;
- information structure on diagrammatic maps is static which limits its use by those with specific search needs; and
- the car location system is expensive to maintain and forces the visitor to find the nearest stamping machine.

Our group has developed a navigation technology, called DiWay, based on mobile phones that overcomes these limitations and, at the same time, presents additional advantages. These include the facilitation of mobile commerce, social and entertainment activities and emergency services. DiWay is based on the use of wireless

application protocol (WAP) using images and text, and the short message service (SMS) relying upon on text messages. It will also accommodate multimedia message service (MMS) sending images, sounds and text upon request or permission of the user. DiWay architecture is supported by a view of geographical information systems (GIS) as being composed by a data tier, a logic tier and a presentation tier. The data tier facilitates the access of geo-referenced data. The logic tier incorporates spatial analysis tools to determine shortest paths, near points of interest and shortest circuits. Finally, the presentation tier enables the visualisation of maps, images and text on the Web, mobile Internet, and SMS and MMS systems. DiWay may be seen thus as a gateway for mobile access to a customised GIS of a shopping centre. Early tests at the Colombo Centre, carried out in cooperation with SONAE, have shown that the system may be adapted to other shopping centres of the SONAE group. The success of that adaptation depends on minimising the burden placed on the user in future developments of DiWay.

This chapter discusses the problem underlying the need for DiWay, and the methodological issues behind DiWay such as micro-detailed geo-referencing, database development, algorithms behind navigation and user interface design. The DiWay application to the Colombo Centre is reported for illustrative purposes. Future developments of DiWay are also listed.

4.2 Problem Definition

The geo-referencing of points of interest in urban areas is becoming increasingly detailed to satisfy user needs. Maps of major cities available on the World Wide Web illustrate this trend. Shopping centres provide the setting to go one step further. Geo-referencing points of interest such as shops is not enough; one may be interested in locating shelves or even products in a shop, and certainly one is interested in having the parking places geo-referenced. The definition of spatial location, at this refined level, facilitates navigation, commerce and social activities.

Exploration of a place is usually associated with the use of symbolic representations such as maps that are then matched against the reality they try to synthesise. This matching is facilitated by the use of ‘anchors’: a discrete number of points of interest displayed in maps that are landmark shops or other easily recognisable locations. In a shopping centre, there are certainly shops more easily noticed than others. From the perspective of different visitors, one may conclude that every existing shop may be an anchor for a certain visitor. By the same token, every product in a shop may be a point of interest and, clearly, the same is true of every parking place in the centre.

The Colombo project has shown that the definition of anchors has been limited by the lack of the availability of back-office data to support the potential uses of the system. That data may be only location (such as in the case of the parking places), but it may also include other data such as the Web address, list of products and prices. In the end, anchors have to be defined by the shopping centre operators. Traditional maps and wayfinding signs of shopping centres are thus limited to aid generic navigation:

- they do not help a visitor to find a particular product on a shelf or remind him or her of the place he or she left the car;
- they do not help him or her in obtaining more information on a product or even where and how to buy it; and
- they are of little use in helping the visitor in meeting other visitors at a certain time in a certain place.

Mobile phones, and particularly Internet-enabled mobile phones, provide an alternative platform for navigation in shopping centres: they can display maps showing the visitor's location, offer directions and present product information. They may be used to buy products, locate friends and in emergency situations. However, the development of navigation systems for mobile phones to be used in shopping centres presents key research problems such as:

- indoor positioning, as traditional mobile phone positioning systems (based on the Global Positioning System or other radio-based triangulation systems) are not successful in indoor spaces;
- the selection of anchor point technology to provide information to the mobile phone user;
- the definition of the most appropriate spatial analysis tools to aid the users in navigation;
- the software framework to support heterogeneous data sources and present information in different platforms (Web, WAP, SMS, MMS, UMTS, VoiceXML); and
- user interface issues in the design of the DiWay system for mobile phones.

4.3 DiWay Methodology

4.3.1 Introduction

The DiWay methodology comprises proposals to solve the research problems mentioned above. It is based on the representation of a shopping centre as a set of graphs, each graph representing a floor. In the parking lot floors, each node represents one parking spot and the arcs are the paths to elevators and stairs. In the shopping floors, each node is a shop, restaurant or other facility. The arcs are the shopping centre streets. Each shop's node may have associated secondary nodes representing shelves or products. Links between floors are represented by special arcs, which have specific properties indicating the type of connection (stair, elevator, escalator, ramp) and the floors they link to. Using these properties the system is able to generate paths according to specific user needs. If, for example, a handicapped person wants to visit the centre using a wheelchair, DiWay will only use floor-linking arcs with an elevator or ramp connection thus providing the user with a personalised path.

Each node is represented by an alphanumerical code defined hierarchically (Figure 4.1). Country, city and shopping centre sub-codes are automatically provided by the operator's location technology, so that the user only has to enter between three (shop) and seven characters (shop + object). In the DiWay methodology, these codes are

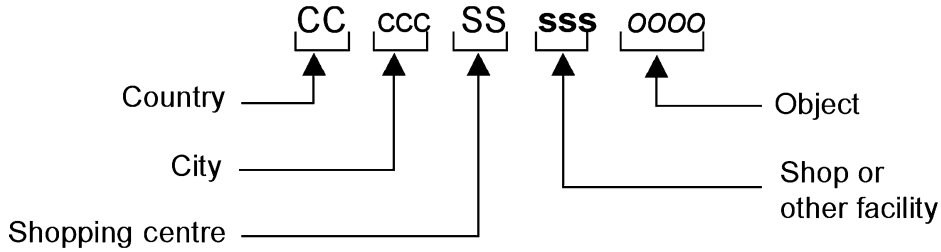


Figure 4.1 *The hierarchical structure of the spatial bar code (SBC)*

named spatial bar codes (SBCs) and identify uniquely an object anywhere in the world. Each node is considered to be an anchor point. It is possible to geo-reference them *a priori*, thus establishing a correspondence between their coordinates and the SBCs, by digitising the plans of the shopping centre.

4.3.2 Indoor Positioning

Anchors may be tags or emitters with associated SBCs. The SBCs may be presented together with the corresponding WAP address or SMS number (Figure 4.2). This concept solves the positioning problems in indoor spaces where both satellite and radio frequency triangulation systems are often inadequate.

4.3.3 Anchor Point Technology

For the time being the best option is to use plastic tags as in Figure 4.2. The user can then use this code to locate him or herself to the system, writing it on a WML form or in the text of an SMS. For the third-generation mobile phones, and even for some 2.5G smart phones, a Bluetooth-based emitter will probably be the best solution (Held, 2000). Using Bluetooth emitters as anchors will allow users to pinpoint their location automatically without having to read or write any SBC. Whenever they ask for some information, DiWay will tell the phone to read the nearest SBC and display the adequate contextual information. Tests with Bluetooth-based anchors are already being conducted using a PDA to simulate a 3G phone.

The anchors may be used to support navigation, as a reference for meetings and emergency services and to facilitate mobile commerce as discussed for the Colombo Centre case study.

4.3.4 Spatial Analysis Tools

The spatial analysis problems involved in shopping centre navigation are essentially the definition of the shortest path between an origin and a destination. Typically, the

Figure 4.2 *Use of a tag to anchor a shop in the Colombo Centre*

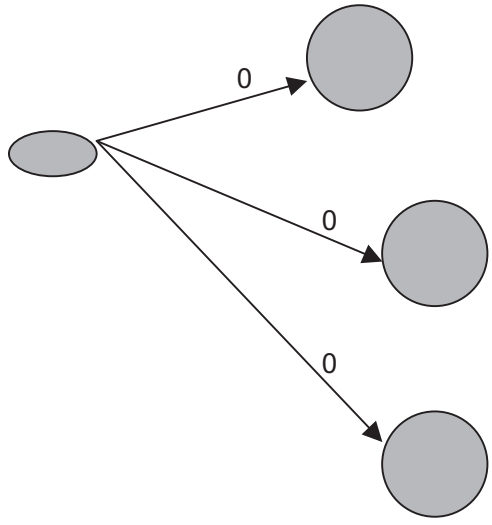


Figure 4.3 *Adaptation of Dijkstra algorithm to solve the problem of having alternative origin points: the distance between the virtual origin point and the potential departing points is 0*

visitor wants to find the route between an entrance of the shopping centre and a shop or a product. Conversely, he or she may be interested in finding the way between the exit and the car in the parking lot. For both problems, DiWay applies the Dijsktra algorithm (see the description in Cormen *et al.*, 1990). The Dijsktra algorithm requires that weights are assigned to each arc (or shopping street). Those weights represent distances. The algorithm then proceeds to determine the shortest distances from the origin to all other points in the network. It also provides the sequences of arcs that are associated with those distances.

DiWay applies a modification of the Dijkstra algorithm to solve two other problems: finding the shortest path between a destination and several possible origin points; and determining the nearest point of interest. The former problem is of interest when one wants to select which entrance to use to be closer to a certain shop. The latter occurs when someone wants to look, for example, for the nearest bookshop to his or her current location.

To solve both problems, artificial nodes and arcs with null weights are considered. When there are several origin points and one destination, an artificial node is inserted as a virtual origin point. The distances between the virtual and the real origin points are zero (Figure 4.3). The Dijsktra algorithm is then applied with the virtual origin point as the starting node. The origin point that is part of the shortest path is the entrance that the user should select.

In the nearest-point problem, several destination points may be considered (i.e., all the bookshops in the centre). In this case, the arcs from the real destination points (the available bookshops) to a virtual destination point have weights equal to zero. The Dijsktra algorithm is applied considering the virtual destination point. The bookshop that is part of the shortest path is the nearest bookshop to the user.

Other analytic problems of interest are node-covering algorithms (see Princeton's Department of Mathematics website on the travelling salesperson problem). One may be interested in defining shortest circuits or paths covering a predefined number of shops. The solution to these problems is not included yet in the DiWay methodology. Examples may include the planning of a visit to a number of predefined shops (i.e., to a bookstore, sports store, furniture shop and clothes store), trying to spend the minimum amount of time departing and arriving at the parking place. Inside a supermarket, one may want to define the minimum path from the entrance to the exit booth while picking up predefined items, provided that these are geo-referenced.

4.3.5 Software Framework

The software platform developed by YDreams to support LBS is called FluidMapping. It was developed on Microsoft.NET, taking advantage of the flexibility, security and performance of this platform. FluidMapping is structured into three tiers: the data tier, the logic tier and the presentation tier (Figure 4.4). These tiers have been designed to:

- handle heterogeneous sources of information therefore allowing the integration with already deployed database services. This is important for a system like DiWay where data are provided by several different entities using different information systems;

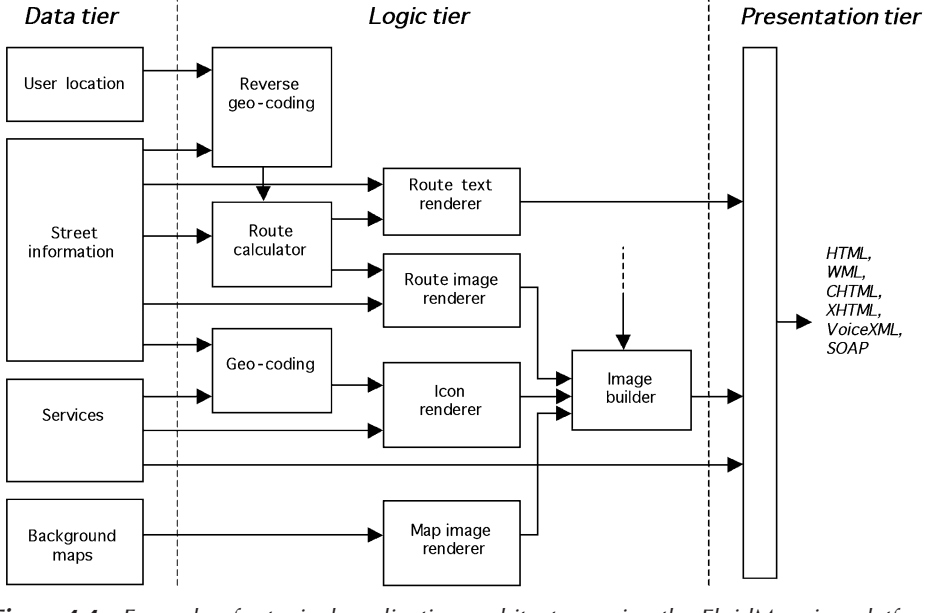


Figure 4.4 *Example of a typical application architecture using the FluidMapping platform*

- be incorporated into locally executed applications (classic architecture), client/server environments (three-tier architecture), or distributed environments (distributed three-tier, or *n*-tier). This is therefore a flexible solution allowing the same service to run locally (e.g., on a kiosk or a PDA) or remotely (e.g., on a mobile phone, Internet-enabled PDA or interactive television) or even as a web-service;
- allow the stratification of products using different combinations of modules. In this way, the same framework can be used to deploy different products or services. Using FluidMapping, DiWay can easily be converted to a museum navigation service; and
- allow the presentation of the services in all wired and wireless devices (SMS, MMS, HTML, WML, CHTML, XHTML, VoiceXML and SOAP).

4.3.6 Data Tier

There are four types of information that are essential for location-based services: user location, street information, background maps and the services (or contents) themselves.

User Location

The user location is what differentiates the LBS from the traditional wired services. The location can be found by automatic technologies such as enhanced observed time difference (E-OTD), time of arrival (TOA), cell ID and assisted-GPS (A-GPS) or simply supplied by the end user. The existing automatic methods can only give the area where the user is located; they still cannot provide the store or object location. If the service needs a more accurate position, the user still has to type it or say it, depending on the device he or she is using. Future outdoors solutions will minimise this problem. DiWay type propositions, based on Bluetooth code emitters, will solve the indoor positioning problem.

Street Information

The street information database includes all the information for the streets. It is used for the calculation of the route between two points, the determination of the nearest point and to find the geographic position of a point given its address (geocoding). This information is created using the FluidMapping Graph Editor.

Background Maps

The background maps are used for the graphical visualisation of the route and position of a point. These maps are related to the street information but are handled separately. For example, WAP uses black and white maps, while the Web may use coloured maps. However, the road information can be exactly the same. The FluidMapping platform has editors for raster and vector maps. When using raster

maps, the adequate maps are automatically tiled to form the background map. Vector maps can either be constructed using the appropriate editor or imported from other applications like MapInfo.

Services

The services databases include the contents that will be geocoded and made available to the end user. The FluidMapping framework only needs the address of the points of interest to handle the contents automatically. The operator usually controls the user location, and the shopping centre's administration, the street information and background maps databases. Third-party companies usually supply the services databases through syndication. This means that the operator may not have control over how the data is supplied. As a result, there can be a great variety of transportation protocols, query methods and data structures involved. The conversion and integration problems are usually significant.

To minimise these problems, FluidMapping adopted the Microsoft BizTalk server solution (http://www.microsoft.com/biztalk/). BizTalk provides powerful and easy-to-use tools to control the flow of information and its conversion into formats that the YDreams platform can interpret. The use of this product considerably reduces the time necessary to integrate a new third-party data provider.

The data tier framework of FluidMapping is developed on top of the latest generation of Microsoft database-access components called ADO.NET. These are very generic components allowing a common data retrieval method for all types of databases, static files or data generated dynamically by other components, simplifying the code of each component.

4.3.7 Logic Tier

The logic tier framework consists of several modules. Key modules were shown in Figure 4.4 and are discussed below. These components use XML to communicate between them thus providing total control over the data flow. This data can also be redirected, at any place, to third-party components.

Geocoding: this module uses the street information database to convert an address into geographical coordinates.

Reverse geocoding: uses the street information database to convert geographical coordinates into addresses.

Map image renderer: generates the background map image given the geographic coordinates, image size and source.

Route calculator: calculates the route between two geographical points. The different methods of calculating routes (shortest, fastest, cheapest) are implemented in different modules so that they can be used separately, depending on the product.

These modules share the same interfaces so their interchange is straightforward.

Route text renderer: processes the route calculation result so that the name of the streets, shops, direction changes and other useful information are added to it.

As for all the other components the output of this module is in XML. For example, the direction changes are coded as tags, instead of text, so that they can later be translated into any language using a simple XSLT rule.

Route image renderer: draws the route so that it can be superimposed on the background map. It provides total control over the line colour, style and width.

Icon renderer: calculates the relative position of the icons to be superimposed on the background map image. It gives total control over the icon design.

Image builder: allows the combination of several layers of images. It gives total control over the layer order, position, size and opacity. This module also enables the addition of other elements to the image such as legend, copyright notice and watermark. The image can be saved in any of the Web file formats (PNG, JPG, WBMP and GIF) and other graphical formats.

4.3.8 Integration

The logic tier framework is currently implemented on the Microsoft.NET framework (http://msdn.microsoft.com/net/) and can take advantage of all the related technologies on the Windows 2000, Windows XT and Windows CE platforms. On client/server architectures, the end user does not need to have any of these platforms to be able to access the services. The framework is located on the server and only the output is sent to the client, meaning that it is independent of the client's platform.

In case the framework is run locally on devices with Windows CE platform, like pocket PCs and auto PCs (http://msdn.microsoft.com/library/default.asp?URL=/library/wcedoc/apcintro/apcintro.htm), it can take advantage of the CE edition of the Microsoft SQL Server database (http://msdn.microsoft.com/library/default.asp?URL=/library/wcedoc/sqlce/portal_6ol0.htm) to store the data on a local database. To keep memory usage down, this database can store just the data related to the services close to the user location. This database can be automatically updated through wired or wireless synchronisation. The framework can also be wrapped in a SOAP compatible Web-service that allows its access to third parties through a custom API.

4.3.9 Presentation Tier

The use of XML on the framework allows the customisation of its output to virtually any device, from regular HTML for web browsers, passing by several wireless types such as WML, CHTML and XHTML, to VoiceXML, that allows the use of voice on a regular cellular phone.

The DiWay interface consists of the display of maps and associated geo-referenced content by request on WAP enabled mobile phones. Typical queries include point-of-interest search (which may comprise nearest point-of-interest search), information on anchor points, directions for personal navigation and emergency services.

The display of maps is limited by:

- the size of most WAP-enabled phone screens; as a result, display clutter usually occurs; and
- the monochromatic nature of those screens; colour cannot be use to differentiate floors, shops or products.

Interaction with WAP phones is further hampered by the multidimensionality of the keyboard (Matilla and Ruuska, 2000): most keys stand for one number and three letters. To minimise these problems, DiWay has incorporated the following principles.

- Minimise display clutter. Two levels are provided: the overall view of the floor and the local area where the visitor is at a certain moment. Both are represented diagrammatically. Text labelling is done for major anchors at the global level. At the local level, text is only used to label the anchor of interest; numerical labelling, complemented by text, is used for adjacent anchors and streets.
- Minimise text input. Menu options are offered for point-of-interest search. Users only need to type three letters to obtain a list of matching names for streets and points of interest.

An additional feature of the interface is the Phonemarks (Figure 4.5). By using the Phonemarks feature the user stores anchor points for later exploration in the fixed or mobile Internet platforms. If those anchor points reflect commercial items this

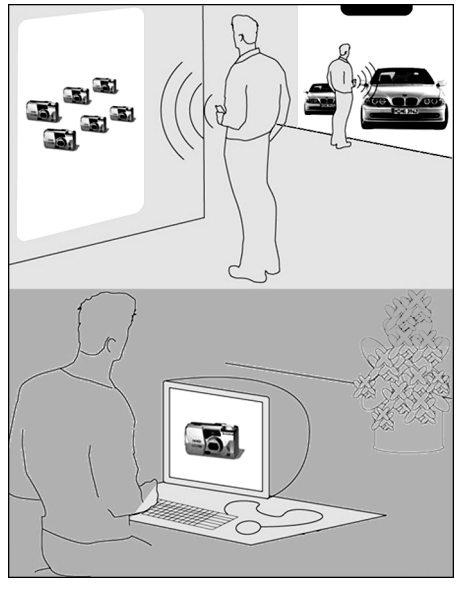


Figure 4.5 *Phonemarks – whilst visiting, users can send codes of preferred objects or shops to a descriptive file. This will allow the user to visit, at a later date, Web sites or WML decks that are associated to these codes*

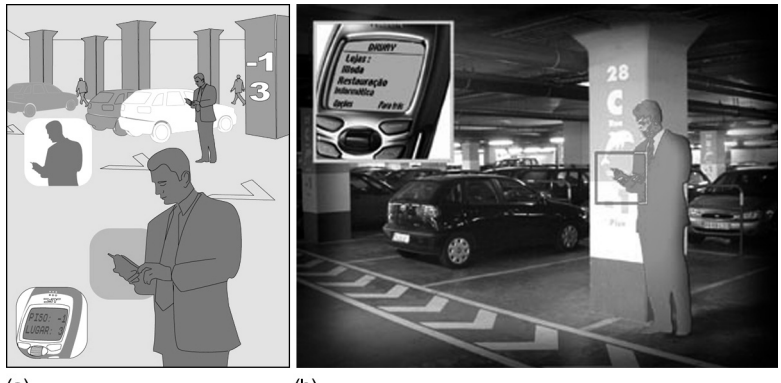
feature may be used for mobile commerce. Phonemarks are also used to store the location of cars in the parking lot.

The DiWay system may also use text messages (SMS) and, in the recently released phones, text, image and sound messages based on MMS. At this point, it is not yet clear if navigation systems based on MMS, that place less of a burden on the user, may overcome mobile Internet alternatives. If flat rates are adopted as in Japan, mobile Internet will become less expensive, and allow for a thorough use of mobile applications.

4.4 Application

DiWay is being implemented for the Colombo Centre in a demonstration project that includes the following features.

- Store and product finders, to find stores that sell a particular product. These are displayed on diagrammatic maps of the shopping centres. Paths from origin to destination are provided on those maps and by textual directions using key anchor points.
- Parking place finder. Using the Phonemarks feature, the visitor stores the spatial code of the place where he or she parked the car. The parking place finder will later help him or her in finding the way back to the car (Figure 4.6).
- Product interaction, to enable the visualisation of detailed product information and facilitate its mobile purchase (Figure 4.7). The user types in the code for the product (or receives its code via a Bluetooth anchor) and accesses all the corresponding stored information. Furthermore, if the store is closed the user can buy the product through DiWay or use the Phonemarks feature to store the product's code for a later review.



(a) (b)

Figure 4.6 *Using Phonemarks to (a) store the position of the car and (b) locate the car in Colombo Centre's parking lot*

Figure 4.7 *Interaction with a product using the mobile phone*

- Casual encounters, to facilitate the meeting of people with identical profiles. This is a community-building feature that uses the centre as a link from the cyberspace to the real world, allowing people to establish virtual communication and eventually meet using the directions provided by DiWay. This works like a treasure-hunting game where the treasures are real people. Profiles are filled on the Web prior to the visit, or through DiWay upon entrance to the shopping centre. In order to stimulate sponsorship of the system, directions will be given using shops as points of interest and the meeting points will be established at main sponsor stores.
- Emergency alerts are possible by pressing e after entering the nearest spatial code (or in the future receiving it by activating its emission from the nearest Bluetooth anchor).
- Information on transportation to and from the centre. This feature includes an automated service to request taxis, without forcing the user to make any phone call. DiWay will both inform the user and the taxi of the most adequate meeting point.

The experience so far has shown that frequent WAP users find the system useful. However, such users total less than 10% of the Colombo Centre visitors that have mobile phones. A much higher percentage regularly use SMS and is expected to adopt MMS. Thus, it is clear that unless third-generation devices enable an easier use of the mobile Internet, DiWay and other navigation systems will be based on MMS systems that will enable the display of maps, sounds, images and even 360 degree views.

4.5 Conclusions and Further Developments

The emergence of mobile computing has opened new opportunities for LBS. Indoor applications of LBS in shopping centres are most relevant to facilitate navigation, mobile commerce and community activities. This chapter has described a methodology developed with those goals in mind for mobile phones that are mobile Internet,

SMS and MMS enabled. The approach, named DiWay, includes detailed geo-referencing of anchor points and other relevant objects, spatial analysis tools and an appropriate interface for mobile phones. DiWay is based on a software platform called FluidMapping. It relies on a view of GIS as composed of modular components. FluidMapping is designed to handle heterogeneous databases and multiple presentation devices such as personal computers, personal digital assistants and mobile phones. It also includes modular spatial analysis components. In this chapter, FluidMappping was applied to a shopping centre represented as a graph where nodes were shops and parking places and the arcs were the shopping streets. The spatial analysis component used in DiWay was based on the Dijkstra algorithm. This algorithm was applied to determine the shortest path and the nearest point of interest. The potential application of node-covering algorithms to determine optimal circuits was also discussed.

The DiWay approach was illustrated with the development of a system for the Colombo Centre in Lisbon. This system has been developed in cooperation with SONAE, the operator of the centre. It was designed to replace the existing methods of assisting visitors in their shopping experience. The experience so far has shown that the system is only beneficial for seasoned WAP users. These make up less than 10% of the total number of visitors to the Colombo Centre. MMS, which enables the visualisation of maps, sounds, images and 360 degree views, may be an alternative for future implementations. DiWay will benefit also from the development of the third-generation mobile phones and the increasing use of Bluetooth chips. Third-generation phones will offer a larger bandwidth and colour. Bluetooth chips will provide an infrastructure to facilitate indoor positioning and avoid dialling codes which is the current positioning solution.

References

Abowd, G., Atkeson, C., Hong, J., Long, S., Kooper, R. and Pinkerton, M. (1997) Cyberguide: a mobile context-aware tour guide, *Wireless Networks*, **3**: 421–33.

Abowd, G., Dey, A., Orr, R. and Brotherton, J. (1997a) Context-awareness in wearable and ubiquitous computing, in *Proceedings IEEE 1st International Symposium on Wearable Computers*, 13–14 October, Cambridge, MA. Available as an html document in http://www.cs.gatech.edu/fce/pubs/iswc97/wear-poster.html.

Beadle, H., Harper, B., Maguire Jr., G. and Judge, J. (1997) Location aware mobile computing, in *Proceedings IEEE/IEE International Conference on Telecommunications (ICT'97)*, Melbourne, April, pp. 1319–24. Available as an html document at http://www.elec.uow.edu.au/people/staff/beadle/badge/ ICT97/ict97.htm.

Beadle, H., Maguire Jr., G. and Smith, M. (1998) Location based personal mobile computing and communication, in *Proceedings 9th IEEE Workshop on Local and Metropolitan Area Networks*, Banf, May. Available as an html document at http://www.elec.uow.edu.au/people/staff/beadle/badge/ieee_lan_98/.

Câmara, A.S. and Raper, J. (1999) *Spatial Multimedia and Virtual Reality*, Taylor & Francis, London.

Câmara, A.S., Scholten, H. and Remedio, J.M. (2000) Spatial multimedia, in Openshaw, S. and Abrahart, A. (eds) *GeoComputation*, Taylor & Francis, London.

Cormen, T., Leiserson, C. and Rivest, R. (1990) *Introduction to Algorithms*, McGraw-Hill, New York.

Davies, N., Cheverst, K., Mitchel, K. and Efrat, A. (2001) Using and determining location in a context-sensitive tour guide, *IEEE Computer*, **34**(8): 35–41.

Department of Mathematics, Princeton U. (2001) Travelling Salesman Web site at http://www.math.princeton.edu/tsp.

Dias, A.E., Romão, T., Pimentão, P. and Câmara, A. (1998) Keeping contextual awareness: the TOI system, in *Proceedings CD-ROM GIS PlaNet'98*, Lisbon, Portugal, 7–11 September.

Dias, A.E., Romão, T. and Pimentão, P. (1997) Traveller's on-line information system – TOI: information system and location modules, in *Proceedings JEC-GI'97 Joint European Conference and Exhibition on Geographical Information*, Vienna, Austria, 16–18 April, IOS Press, pp. 238–47.

Dias, A.E., Santos, E., Pimentão, P., Pedrosa, P. and Romão, T. (1996) Bringing ubiquitous computing into geo-referenced information systems, in Rumor, M., McMillan, R. and Ottens, H. (eds) *Geographical Information – From Research to Application Through Cooperation*, IOS Press, Amsterdam, The Netherlands, pp. 100–3.

Frisk, C. (1998) Location and context awareness in adaptive personal communication services. Available as an html document in http://www.pcs.ellemtel.net/Claes/ pcc/context.html.

Gouveia, C. and Câmara, A.S. (1999) Multimedia and urban planning, in Stillwell, J., Geertman, S. and Openshaw, S. (eds) *Geographical Information and Planning*, Springer Verlag, Heidelberg.

Held, G. (2000) *Data Over Wireless Networks: Bluetooth, WAP and Wireless LANs*, McGraw-Hill, New York.

Hightower, J. and Borriello, G. (2001) Location systems for ubiquitous computing, *IEEE Computer*, **34**(8): 57–66.

Mattila, K. and Ruuska, S. (2000) Designing mobile phones and communicators at NOKIA, in Bergman, E. (ed.) *Information Appliances and Beyond*, Morgan Kauffman, San Francisco.

Pradhan, S., Brignone, C., Cui, J., McReynolds, A. and Smith, M. (2001) Websigns: hyperlinking physical locations on the web, *IEEE Computer*, **34**(8): 42–8.

Ryan, N. (1998) FieldNote Desktop: an experimental spatio-temporal information system, 4th International Colloquium on Computing and Archaeology, Bilbao, May. Available as an html document at http://www.cs.ukc.ac.uk/research/ fieldwork/papers/Bilbao/FieldNote Desktop.html.

Ryan, N., Pascoe, J. and Morse, D. (1998) Enhanced reality fieldwork: the context-aware archaeological assistant, in Gaffney, V., van Leusen, M. and Exxon, S. (eds) *Computer Applications in Archaeology*. Available as an HTML document at http://www.cs.ukc.ac.uk/research/fieldwork/papers/CAA97/ERFldwk.html.

5

Mass Appraisal and Noise: the use of Lifestyle Segmentation Profiles to Define Neighbourhoods for Hedonic Housing Price Mass Appraisal Models

Steven Laposa and Grant Thrall

Abstract

This chapter presents a case study on the application of geographic information systems (GIS) with the mass appraisal of 1400 residential properties located in an urban area. The objective of the case study is to estimate the total value of the residential portfolio using 850 sales from the multiple listing service. The residential properties are in an urban area subject to extensive vandalism, crime and vacant homes. The authors contend that the use of GIS, and special three-dimensional (3D) modelling and visualisation of sale prices, provides a higher explanatory power than typical hedonic residential price models. The study shows how the application of lifestyle segmentation profiles (LSPs) such as ESRI Bis (formerly CACIs Coder/Plus) ACORN geodemographic database can support valuation modelling.

5.1 Introduction

The practice and implementation of mass appraisal models are widespread in the real estate industry. Local tax assessors use a form of mass appraisal modelling to

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

estimate residential property values for property tax assessments (for a discussion and overview, see International Association of Assessment Officers (IAAO), 1999). Implementations of computer-assisted mass appraisal (CAMA) applications are widespread throughout the assessment industry. Mass appraisal models have been used in litigation cases and portfolio investment analysis for similar properties, such as fast food or convenience stores. Mass appraisal is a statistical procedure of defining a representative sample of a larger population with the purpose of estimating the aggregate value of the overall population. According to the Appraisal Institute Foundation, a mass appraisal includes:

- identifying properties to be appraised; defining market area of consistent behaviour that applies to properties;
- identifying characteristics (supply and demand) that affect the creation of value in that market area;
- developing a model structure that reflects the relationship among the characteristics affecting value in the market area;
- calibrating the model structure to determine the contribution of the individual characteristics affecting value;
- applying the conclusion reflected in the model to the characteristics of the property(ies) being appraised; and
- reviewing the mass appraisal results.

Lead author, Steve Laposa, is global Director of Real Estate Research for PricewaterhouseCoopers (PwC). In that capacity, he was responsible for supporting a team assigned to appraise approximately 1400 residential properties in a major midwestern city. Grant Thrall, was invited to contribute his GIS and business geography expertise to the team. The properties were located in an older, downtown neighbourhood. The properties were owned by a client of PwC; the client requested a total value of those properties in order to establish a market value which would be used to contribute those properties for the creation of a new public–private joint venture. Mass appraisal therefore derives a reasonable estimate of the aggregate value of the properties. Individual property appraisal would instead estimate the value of each individual property, which could be aggregated for an estimate of total value. However, the cost and time involved in executing 1400 individual appraisals would be too great and would not therefore meet the clients objective. Mass appraisal generates comparable results at a much lower cost and in a much shorter time horizon. Hedonic mass appraisal models assume there to be a high level of homogeneity in the population. Hedonic mass appraisal models are a special case of the more general hedonic model. Hedonic models rely upon various forms of regression analysis, where a dependent variable is hypothesised to be explained or described by a set of independent variables (for discussion, see Thrall, 2002, Chapter 4).

Residential home values within small local neighbourhoods are generally homogeneous in terms of size, number of bedrooms, lot size, age and other commonly used descriptive characteristics. Large variances between adjoining homes are usually not found. However, large variances between adjoining homes can be found in older urban neighbourhoods. In the United States older urban neighbourhoods often

exhibit great structural and value differences from street to street, and even within a block. So, the assumption of homogeneity in the population breaks down. Older urban, downtown residential neighbourhoods, often contain deteriorating homes due to abandonment, vandalism, fire, crime and lack of standard maintenance. Thus, the average price per square foot over a large sub-market does not necessarily reflect the value of an individual home. This provides a problem for the usual procedures for mass appraisal.

A key ingredient to mass appraisal models is the assumption that variance is small between residential home values within a neighbourhood; the smaller the geographic delineation of the neighbourhood, the smaller would be the expected housing price variance. The assumption of small variance excludes the existence of large price per square foot variance between nearby homes. In other words, the land price surface is expected to be smoothly unfolding across space, and not discontinuous (see Thrall, 1998a, for a discussion on the error of this assumption).

Geographic surface interpolation procedures assume that space is smooth and continuous. These procedures assume that values between locations do not change abruptly, allowing local values to be accurately represented by a larger spatial trend. In many instances, these assumptions accurately represent physical phenomena. However, real estate is more problematic. Ownership and build characteristics change from parcel to parcel. So, spatial trend models may be used to represent general trends, but not individual parcel values. When the variation between parcels is very large, as in our study, large area spatial trends have considerable error.

In this chapter, we deal with two issues of mass appraisal: first, how can one specify an appropriately small geographic neighbourhood in a manner that is both defensible and repeatable between neighbourhoods, and even between cities? Within the resulting small geographic neighbourhood, land prices would be expected to have a small (minimum) variance between dwellings; second, when large per square foot variances between nearby homes do occur, what can explain the variance and how can that explanation be incorporated into a mass appraisal model? This chapter addresses the use of mass appraisal modelling incorporated with GIS and databases. Our contribution to the literature is the use of lifestyle segmentation and 3D GIS applications in addition to standard housing hedonic models. First, we provide a discussion of the explanatory variables we use in our mass appraisal, including lifestyle segmentation profiles which are particularly important for neighbourhood delineation. We then present a discussion of our methodology and our conclusions.

5.2 Explanatory Variables and Lifestyle Segmentation Profiles

It has long been recognised that house prices are highly dependent upon neighbourhood. Homer Hoyt in his renowned study demonstrated that like-kind neighbourhoods were so spatially interdependent that they formed geographic clusters giving rise to Hoyt's classic hypothesis of the sectoral nature of cities (Hoyt, 1933, 1939, 1966). Within a sector, a variety of measurements are correlated with housing prices; the most commonly used in hedonic housing price models are listed in Table 5.1. The first column of Table 5.1, AM, are measurements of access to place specific

Table 5.1 *Independent variables by geographic scale*

Accessibility measures (AM)	House features (HF)	Neighbourhood attributes (NA)
Highway	Living area	Average house size
Central business district	Bathrooms	Size of neighbourhood/subdivision
Major retail	Pool	Floodplain
Minor retail	Fireplace	Census tract land characteristics
Major employment centres	Foundation type	% land area in single-family dwellings
University/colleges	Age	% land area in multi-family dwellings
Hospitals	Heating type	% land area in offices
	Air-conditioning	% land area in industry
	Roof type	% land area in parks
	Wall type	Census tract population characteristics
	Garage type	% non-white
	Lot size	Median income
		School district

phenomena that are often cited as being important in households' decisions to locate which translates into a housing price effect. Features of a house (HF) affect the market value of a dwelling; the larger the square footage, and the more bathrooms, the greater will be the market value of the house. Fireplaces and pools are expensive to build (often, dwellings with greater construction costs will have greater market value) and neighbourhood attributes (NA) generally affect housing prices. Neighbourhoods dominated with large houses are generally expected to have greater housing values than neighbourhoods with smaller houses. Hedonic housing models will have house price or square footage price as the dependent variable, and sets of variables from the three columns of Table 5.1 as independent variables (see Thrall, 2002, Chapter 5 for further discussion, and Thrall, 1998b, 2002, Chapter 3, on how land prices are related to the leading independent variables).

A typical hedonic housing price model (see also Case *et al.*, 1991; Clapp *et al.*, 1995) might be expressed, using terms from Table 5.1, as

$$P = P(AM, HF, NA) \tag{5.1}$$

where AM are a representative set of accessibility measures, HF is a representative set of housing features, and NA is a representative set of variables that characterise a neighbourhood. Regression coefficients are estimated for a hedonic expression based upon equation (5.1). Dummy variables might be added to the hedonic expression of functional (5.1) to capture the difference in how the independent variables affect the dependent variable between the various neighbourhoods. The mechanics of appropriate estimation of functional expression (5.1) are well known and successfully applied in CAMA (computer assisted mass appraisal) for the determination of estimates of market value of individual dwellings. However, two problems can arise that can make for unreliable estimates of a hedonic expression based upon

functional form (5.1): spatial autocorrelation and biased data sets, and appropriate geographic delineation of the neighbourhoods.

Spatial autocorrelation is a measurement of how closely correlated near neighbours are. If near neighbours are highly correlated with one another, then there is said to be a high level of spatial autocorrelation. From the above discussion, housing prices are generally expected to be highly correlated with prices of neighbouring dwellings. After all, it is said that real estate is 'location, location, location', and indeed if that is true, then real estate prices will be dependent upon the spatial association with neighbouring properties. Spatial autocorrelation in hedonic housing models arises because an observation's price highly depends upon the price of a near neighbour, and that dependence declines the further apart are the observations in geographic space. The degree of spatial autocorrelation may be measured. If there are high levels of spatial autocorrelation, then regression coefficients will be biased in an unknown direction and thereby make for unreliable housing forecasts. An accepted and most frequently used method for correcting for spatial autocorrelation is by way of appropriate independent variable choice, where the independent variables capture the spatial trend. The problem has been, however, the identification of independent variables that can capture the uniqueness of the neighbourhood. A neighbourhood can be a geographically contiguous area, or noncontiguous, where similar physical characteristics of the properties may or may not be important. However, usually, similarity among the inhabitants is important when defining a neighbourhood.

Data observations that form the basis of the calibration of a hedonic housing price model are assumed to be unbiased representations of the population that the model is intended to represent. Constraints on experimental design might not allow the data observations to represent the population adequately. Constraints on the experimental design might give rise to small variations among the observations in the traditional independent variables listed in Table 5.1 – namely the experimental design might have constrained the observations to have been selected from a region where the properties were physically similar. In addition, house prices within the region might have substantial sub-market variation.

In this chapter we propose a solution for when housing observations are taken from a housing stock that is highly homogeneous in physical characteristics, taken from the same region of the city, while at the same time whose prices are observed to be highly variable though spatially autocorrelated. The experiment reported on here is to estimate housing prices of 1400 properties that are overall of the same age and physical characteristics, and taken from what colloquially would generally be accepted as the same sector of the city. Yet, the sales of the subject properties revealed that there was a wide variation in actual prices paid for these properties. The traditional hedonic housing price approach, following accepted practices and including those physical variables listed in Table 5.1, did not yield acceptable results. The researchers then were confronted with discovering the underlying cause of housing price differentiation, where those housing prices are not a function of traditionally measured physical characteristics.

Since the observations were taken from a small geographic area, and since the experiments revealed that whatever physical variation there was that existed between

the dwellings were not determinants to revealed house prices, the researchers then hypothesised that some underlying social forces were the cause of variation in housing prices. The problem was how to measure those small geographic-scale neighbourhood-specific social forces in a repeatable unbiased manner? In other words, we preferred to use some automated procedure to define a neighbourhood versus relying upon our own prejudices, biases and limited information to define a neighbourhood. Therefore, we turned to methods and procedures of social area analysis.

While lifestyle segmentation profiles (LSPs) have their roots in academia, they recently have received more attention by practitioners with requirements to use consistent criteria to define neighbourhoods on the basis of the type of people that live there. LSPs provide a multidimensional view of people, in contrast to conventional univariate census variables such as census tract median income. LSPs allow for a differentiation between neighbourhoods on the basis of the social makeup of the inhabitants of those neighbourhoods. If neighbourhoods are defined on the basis of dominant LSPs, then within a geographic region of a city, a neighbourhood need not be spatially contiguous as has been the convention in hedonic modelling with dummy variables representing neighbourhoods. LSPs are used to define a social neighbourhood. It is our contention that LSPs can identify geographically small sub-markets, and thereby capture the small-scale component that gives rise to near neighbour spatial autocorrelation for some properties, and discontinuities for other properties. In our study, we used ESRI Bis (formerly CACI) ACORN LSPs because they have a geographic resolution of a ZIP + 4 code that we considered a necessary and sufficient level of geographic resolution. Other LSPs such as Mosaic (Thrall *et al.*, 2001) have the geographic resolution of the census block or block group, and are appropriate for other forms of analysis. For a discussion of LSPs from Claritas Corporation, see Hartshorn (1992, p. 255) and Weiss (1988).

Social area analysis was developed in the post World War II era, by sociologists interested in the social structure of urban sub-markets (see Shevky and Williams, 1949; Shevky and Bell, 1955; as quoted in Hartshorn, 1992, p. 248). Geographers, adopted the social area analysis work of sociologists (see for example Berry, 1964, 1969, 1970) and used factor analytic methods to classify neighbourhoods. CACI Marketing Systems commercialised the concept through the sale of neighbourhood classification data, which CACI referred to as ACORNs, an acronym for 'A Classification of Residential Neighbourhoods'. CACI Marketing Systems became well known to the business community for its ACORN LSPs. CACI Coder/Plus, the product used in this research, integrates the assignment of geographic coordinates, ACORN LSPs, and ZIP + 4 codes to each record observation based upon its street address. Recently, ESRI (www.esri.com) purchased CACI Marketing Systems and markets the same product under the name ESRI Bis. ACORN categories are accurate and differentiated at the geographic scale of the ZIP + 4 code (for further discussion on the derivation and use of LSPs, see Thrall *et al.*, 2001; and Thrall, 2002).

5.3 Data and Methodology

In order to estimate the values of 1400 homes, the authors collected 850 actual home sales from a local Multiple Listing Service (MLS). They selected actual

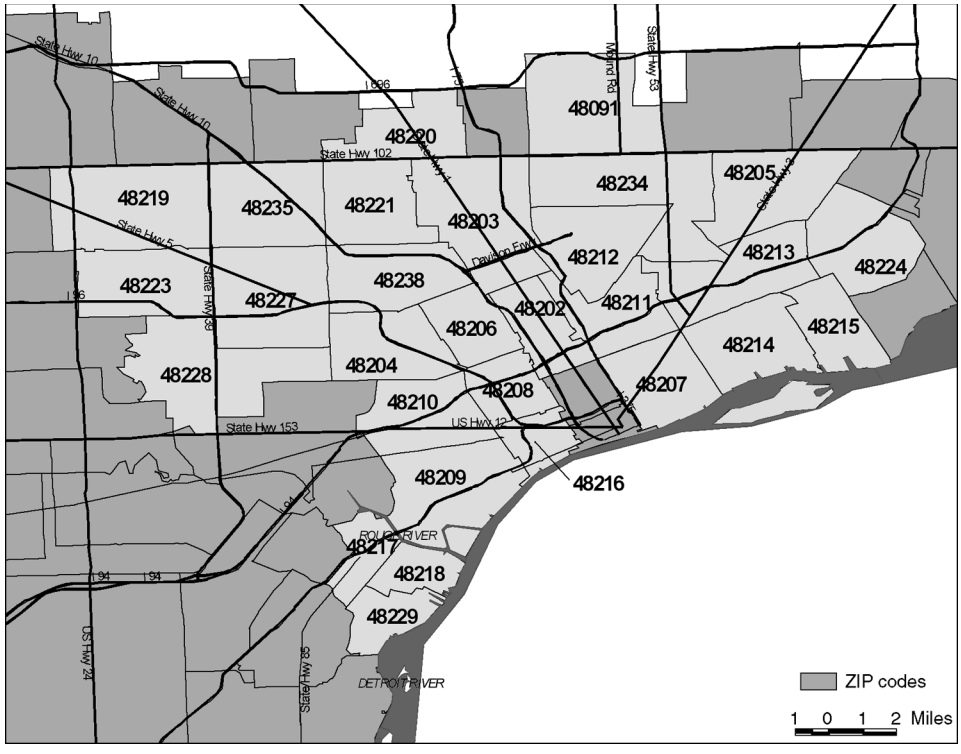


Figure 5.1 *Targeted geographic area by zipcode*

home sales for the prior year and limited their sample to (1) homes sold under $80 000 based on discussions with field inspectors, and (2) geographic proximity to the 1400 homes. Figure 5.1 illustrates the zipcodes selected for the MLS home sale database.

Plate 4 is a geographic visualisation of the MLS properties selected for use as comparables. Addresses of the properties were used to geocode. ESRI Bis (Thrall, 1998b) was used to assign geographic coordinates. ESRI Bis assigns geographic coordinates on the basis of the most accurate coordinate that can be derived between each data record's street address, ZIP + 4, or ZIP code; ESRI Bis is an 'intelligent' geocoder in that addresses will be corrected and standardised using current US postal records. Among the corrections are street address format, name spelling and the addition of ZIP and ZIP + 4 codes. Standardised FIPS codes are also added for documenting the census polygon (block, block group, tract) in which the address is located. Because the neighbourhood is 'older urban inner core', ESRI Bis intelligent use of the US Census TIGER/Line files results in street address assignment of the geographic coordinates (Thrall and Thrall, 1993). ESRI Bis also assigned its ACORN LifeStyle Segmentation Profiles (LSPs) to the data observations; all properties within the same ZIP + 4 are assigned the same ACORN LSP classification. ArcView (www.esri.com) software was used to visualise the geographic data. ArcView was commanded to

display the MLS parcels as a point data file, and to shade those points thematically by three categories of range: \$0 to \$20 K, \$20 K to \$60 K, >\$60 K. Qualitative interpretation of the map includes that there are clusters of observations within each price category; however, observations from all price categories can be observed throughout the map.

There were several steps in our methodology:

- review client-supplied data on the 1400 homes and ensure that critical data fields are populated, i.e., age, size of home, number of bedrooms and condition of the home. Field inspectors coded each property for fire damage, vandalism and whether the property was abandoned (dummy variables);
- geocode the 1400 homes, identify zipcodes, and integrate ACORNs for each property;
- select MLS actual home sales from the zipcodes identified in step 2. Obtain similar data fields as found in step 1 and code with ACORNs;
- run standard pricing models on MLS database, including transformations and stratified models based on price ranges. Test for differences across zipcodes and LSPs; and
- use 3D GIS applications on MLS data.

5.4 Analysis

Figure 5.2 shows the distribution of MLS home prices and correlation between home price and size in square footage. Table 5.2 confirms the lack of correlations between home price to number of baths, number of bedrooms and square footage.

Figure 5.3 is a boxplot of house prices by ACORN categories and the number of MLS home sales by each ACORN (see Table 5.3 for a complete listing of ACORN LSPs). Only eight LSPs were represented in our data set. ANOVA tests confirm significant differences between ACORN categories.

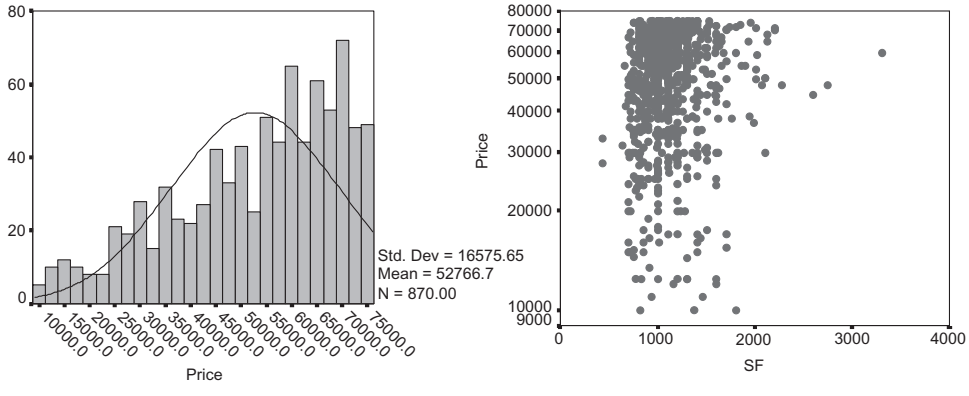


Figure 5.2 *MLS home price distribution and size correlation*

Table 5.2 *Correlation matrix*

		Price	Baths	Bedrooms	SF
Price	Pearson Correlation	1.000	.023	.042	.109**
	Sig. (2-tailed)	.	.495	.214	.001
	N	870	870	870	870
Baths	Pearson Correlation	.023	1.000	.279**	.319**
	Sig. (2-tailed)	.495	.	.000	.000
	N	870	870	870	870
Bedrooms	Pearson Correlation	.042	.279**	1.000	.629**
	Sig. (2-tailed)	.214	.000	.	.000
	N	870	870	870	870
SF	Pearson Correlation	.109**	.319**	.629**	1.000
	Sig. (2-tailed)	.001	.000	.000	.
	N	870	870	870	870

**Correlation is significant at the 0.01 level (2-tailed)

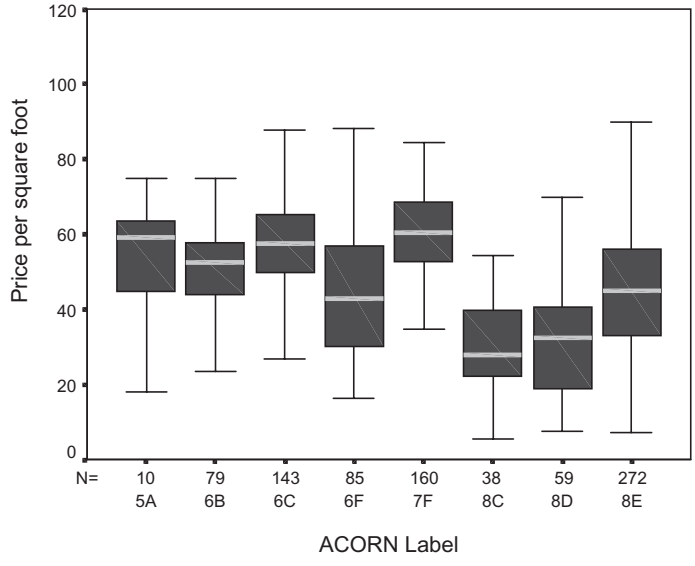


Figure 5.3 *Sale price per square foot by ACORN category*

Table 5.4 presents results from a multiple regression model with price as the dependant variable, and square footage, number of baths and number of bedrooms as independent variables. The regression model's explanatory power is around 1%.[1]

So what is the implication of the model of Table 5.4 having only an R^2 of 1%? First, the traditional hedonic mass appraisal model as described above is inappropriate for neighbourhoods such as that represented in our data set. The traditional hedonic

[1] The authors attempted several other regression models with similar results.

Table 5.3 *ACORN lifestyle segmentation profiles*

ACORN	Description
1A	1A Top One Percent
1B	1B Wealthy Seaboard Suburbs
1C	1C Upper Income Empty Nesters
1D	1D Successful Suburbanites
1E	1E Prosperous Baby Boomers
1F	1F Semirural Lifestyle
2A	2A Urban Professional Couples
2B	2B Baby Boomers With Children
2C	2C Thriving Immigrants
2D	2D Pacific Heights
2E	2E Older Settled Married Couples
3A	3A High Rise Renters
3B	3B Enterprising Young Singles
4A	4A Retirement Communities
4B	4B Active Senior Singles
4C	4C Prosperous Older Couples
4D	4D Wealthiest Seniors
4E	4E Rural Resort Dwellers
4F	4F Senior Sun Seekers
5A	5A Twentysomethings
5B	5B College Campuses
5C	5C Military Proximity
6A	6A East Coast Immigrants
6B	6B Working Class Families
6C	6C Newly Formed Households
6D	6D Southwestern Families
6E	6E West Coast Immigrants
6F	6F Low Income: Young & Old
7A	7A Middle America
7B	7B Young Frequent Movers
7C	7C Rural Industrial Workers
7D	7D Prairie Farmers
7E	7E Small Town Working Families
7F	7F Rustbelt Neighbourhoods
7G	7G Heartland Communities
8A	8A Young Immigrant Families
8B	8B Social Security Dependents
8C	8C Distressed Neighbourhoods
8D	8D Hard Times
8E	8E Urban Working Families
9A	9A Business Districts
9B	9B Institutional Populations
9C	9C Unpopulated Areas

approach does not work because the data do not conform to required assumptions of the modelling procedure, namely that the distribution of both dependent and independent variables are not normal in our data set. The distribution of these variables confirms the lack of homogeneity within the sub-market of the study.

Table 5.4 *Multiple regression model output*

Variables entered/removed[b]

Model	Variables entered	Variables removed	Method
1	SF Baths, Bedrooms[a]	.	Enter

[a] All requested variables entered
[b] Dependent variable: Price

Model summary[b]

Model	R	R square	Adjusted R square	Std. error of the estimate	Durbin-Watson
1	.115[a]	.013	.010	16494.82	.814

[a] Predictors: (Constant), SF, Baths, Bedrooms
[b] Dependent variable: Price

Coefficients[a]

Model		Unstandardised coefficients		Standardised coefficients	t	Sig.	95% confidence interval for B		Collinearity statistics	
		B	Std. Error	Beta			Lower bound	Upper bound	Tolerance	VIF
1	(Constant)	47400.153	2849.657		16.634	.000	41807.110	52993.195		
	Bedrooms	−1090.939	1118.223	−.043	−.976	.330	−3285.683	1103.804	.597	1.674
	Baths	−450.813	1735.245	−.009	−.260	.795	−3856.591	2954.965	.888	1.126
	SF	8.057	2.567	.139	3.139	.002	3.018	13.095	.582	1.719

[a] Dependent variable: Price

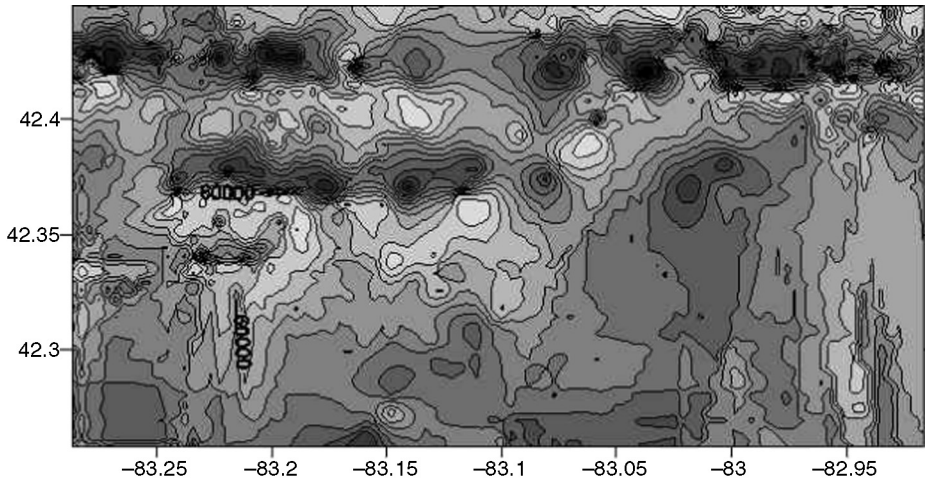


Figure 5.4 *Map of estimated home prices*

Figure 5.4 is a geographic visualisation of the total price surface. The darker shades are isoline peaks, while the lighter shades are isoline valleys. Hence, the higher the price of the MLS comparable property, the darker the area; the lower the price, the lighter the area. Thrall (1998) argued that isoline maps cannot be used to derive estimates of properties within a contour; however, he reasoned that their use was appropriate for qualitative analysis to generate hypotheses about the underlying geographic nature and regularity of the phenomena being studied (in this instance, neighbourhood variation prices of homogeneous properties). Visual analysis of Figure 5.4 reveals a clear east–west street trend, with adjacent streets having abrupt and discontinuous (versus a gradual smooth) change in housing prices. Figure 5.4 was constructed using the same data set as elsewhere in the paper; SURFER (www.goldensoftware.com) was used to derive the isolines. Qualitative analysis of Figure 5.4 gave rise to the hypothesis that prices were highly dependent upon neighbourhood, where neighbourhood was defined as a street face.

The US Postal Service defines unique ZIP + 4 codes along a street face. A second hypothesis arose from the qualitative analysis as well: that since various and nonspatially contiguous ZIP + 4 codes shared the same price isoline, and since the properties were highly homogeneous between neighbourhoods, then the underlying geodemographic characteristics might explain the variation in prices between ZIP + 4 neighbourhoods, and commonality of prices between various ZIP + 4 neighbourhoods.

Figure 5.5 corroborates the qualitative analysis of Figure 5.4. The spatial trend is strongly east–west, with abrupt changes along parallel lines (streets) from north to south,[2] particularly in the northern study area. So, even though the physical

[2] Field inspectors confirmed that houses on main neighbourhood arteries were generally in acceptable quality condition whereas similar sized homes on adjacent streets were of lower quality.

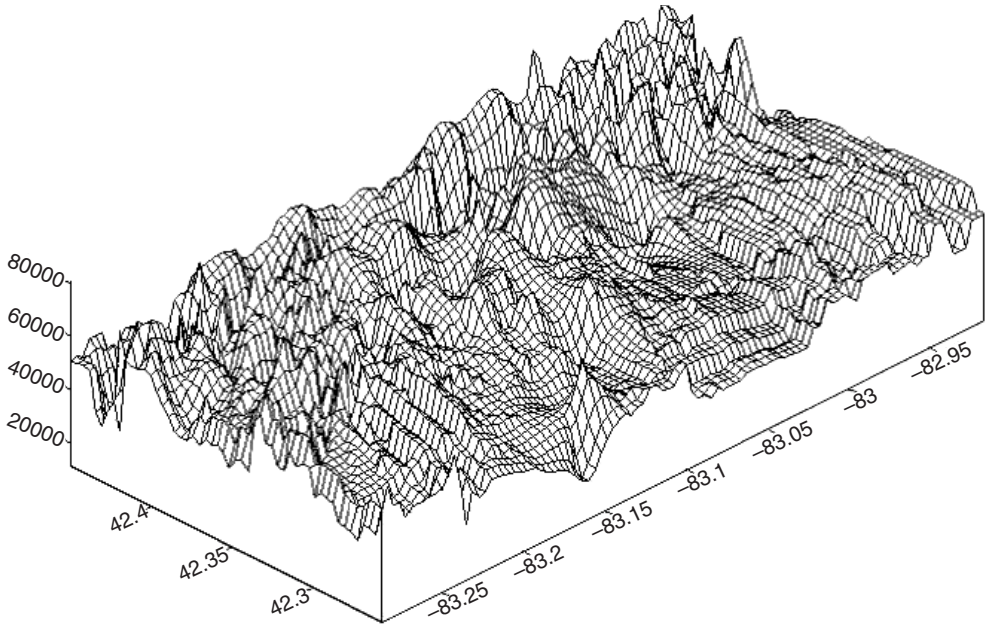


Figure 5.5 *3D map of home prices in study area*

properties were highly homogeneous, there is a strong geographic trend among the properties within what would be considered a sub-market. Therefore, any global averaging of housing values following conventional appraisal techniques would result in an underestimate of properties on some streets in the north, an overestimate of some properties on other streets in the north, and a highly mixed quality of assessment in the southern area of the sub-market.

Plate 5 provides the necessary geographic documentation (Thrall, 1998a) showing the location of the point observations along with the neighbourhoods generated by the isolines. Spatial smoothing can create surfaces even when no observations are present, as evident in the southeastern quadrant of the study area. In other words, the estimate of the surface in the southeast is not reliable. However, because of the apparent geographic regularity and the high density of the observations that give rise to the east–west trend and the abrupt changes from north to south, qualitative analysis here may be considered as reliable. This corroborates our hypothesis that LSPs at the ZIP + 4 scale might capture the underlying cause of spatial variation in prices of the dwellings. In other words, we hypothesise that the variation in prices within LSP categories is less than the variation in prices between LSP categories. Consumers are responding to very small-scale neighbourhood specific geodemographic characteristics, and that response creates a price effect. If so, then neighbourhoods within the study sub-market can be geographically discontinuous, while at the same time sharing LSP characteristic. Plate 5 was created in the same way as the other surface maps, but with optional parameters set differently in SURFER software.

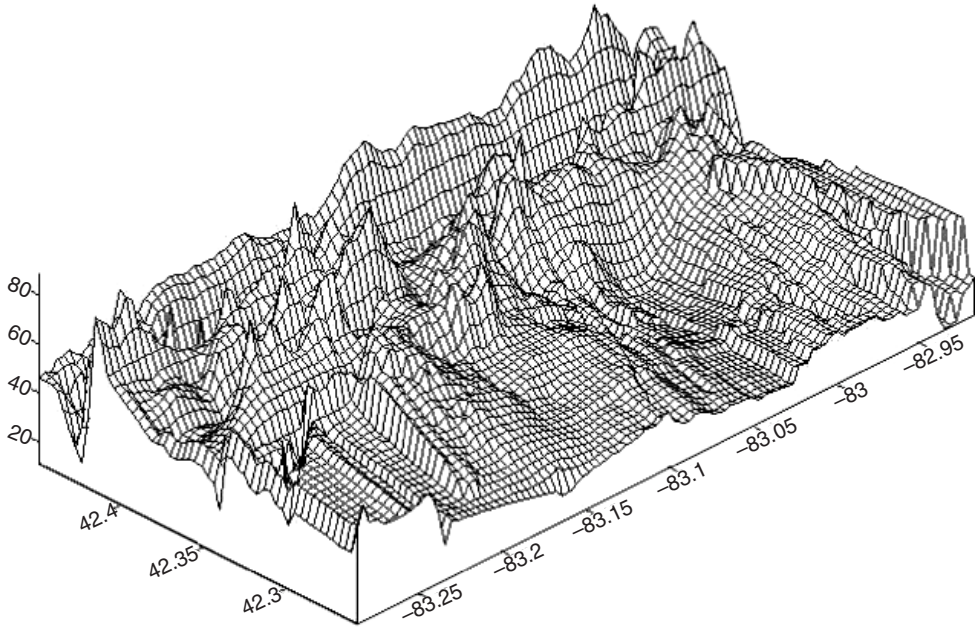


Figure 5.6 *3D model of price per square foot*

Figure 5.6 increases the vertical scale of price thereby allowing the analyst to observe and confirm the qualitative analysis presented in the other figures.

5.5 Conclusion

Practitioner applications often challenge the envelope of tidy academic solutions. Certainly as academics we would not have designed an experiment with an inherently problemsome design as this. However, practitioners are often called upon to provide acceptable solutions to real-world problems that are not as tidy as that derived from an ideal laboratory situation.

The authors support the inclusion of LSPs, such as ESRI's ACORN, in hedonic mass appraisal models. Visualisation with 3D maps assists in improving one's intuition about the geographic problems inherent in hedonic mass appraisal of urban residential neighbourhoods. A long-standing issue in mass appraisal has been how to define appropriately a neighbourhood and sub-market. Here we have demonstrated that LSPs can effectively define neighbourhoods, including neighbourhoods within a larger sub-market, that are not spatially contiguous. That is, within a sub-market, ZIP + 4s with the same or similar LSP can be grouped together to define a neighbourhood. Average square footage prices of houses can be effectively derived from those dwellings within a neighbourhood so defined. Future applications of this procedure might improve suburban residential hedonic mass appraisal that have proved

problemsome because of their heterogeneity. The client accepted the value of our mass appraisal, and used that as a basis to improve their judgement and complete negotiations for the successful transaction.

References

Berry, B.J.L. (1964) Cities as systems, *Papers of the Regional Science Association*, **13**: 147–64.

Berry, B.J.L. (1969) The factorial ecology of Calcutta, *American Journal of Sociology*, **74**: 445–91.

Berry, B.J.L. (1970) *Geographic Perspectives on Urban Systems*, Prentice Hall, New Jersey.

Case, B., Pollakowski, H.O. and Wachter, S.M. (1991) On choosing among house price index methodologies, *AREURA Journal,* **19**: 286–307.

Clapp, J.M., Dolde, W. and Tirtiroglu, D. (1995) Imperfect information and investor inferences from housing price dynamics, *Real Estate Economics,* **23**: 239–69.

Hartshorn, T. (1992) *Interpreting The City: An Urban Geography*, Wiley, New York.

Hoyt, H. (1933) *One Hundred Years of Land Values in Chicago*, University of Chicago Press, Chicago.

Hoyt, H. (1939) *The Structure and Growth of Residential Neighbourhoods in American Cities*, Federal Housing Administration, Washington, DC.

Hoyt, H. (1966) *According to Hoyt – 50 Years of Homer Hoyt (1916–1966),* The Homer Hoyt Institute, Washington DC.

IAAO (1999) *Mass Appraisal of Real Property*, International Association of Assessing Officers, Chicago.

Shevky, E. and Bell, W. (1955) *Social Area Analysis: Theory, Illustrative Application and Computational Procedures*, Stanford University Press, Stanford.

Shevky, E. and Williams, M. (1949) *The Social Areas of Los Angeles: Analysis and Typology*, University of California Press, Los Angeles.

Thrall, G. I. (1998a) Common geographic errors of real estate analysis, *Journal of Real Estate Literature*, **6**(1): 45–54.

Thrall, G.I. (1998b) CACI Coder/plus version 1, Review, *Geo Info Systems*, **8**(6): 43–6.

Thrall, G.I. (2002) *Business Geography and New Real Estate Market Analysis,* Oxford University Press, New York and London.

Thrall, G. I., Casey, J. and Quintana, A. (2001) Geopsychographic LSPs, *GeoSpatial Solutions*, **11**(4): 40–3.

Thrall, G.I. and Thrall, S. (1993) Business GIS data, part two: highend Tiger/Line, *Geo Info Systems*, **3**(9): 66–70.

Weiss, M.S. (1988) *The Clustering of America*, Harper & Row, New York, pp. 4–5; as quoted in Hartshorn, 1992, p. 255.

PART 2
SOCIAL DEPRIVATION

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

6

Targeting Clusters of Deprivation within Cities

Richard Harris and Paul Longley

Abstract

This chapter considers the relative merits of census-based and lifestyles approaches to identify and represent spatially concentrated areas of urban deprivation (and affluence). Using census-based geodemographic indicators and lifestyles data supplied by a UK-based data warehouse company, we appraise three broad approaches: conventional deprivation indicators based on census information; hybrid approaches that link census data to more specific social survey information; and, finally, approaches that make use of private sector lifestyles surveys that can be combined with high resolution imagery of the built environment to more precisely determine social and built conditions. We argue that appraisal of these different methodologies and of the representations of socioeconomic geography they create enhances our understanding of uncertainty in GIS and other geographic modelling. It also casts light on the role and substance of generalisation in urban analysis. We explore these issues in relation to the practice of marketeers working within the geodemographic and lifestyles industries.

6.1 Introduction

Urban geography seeks to identify and understand the geographical patterning of urban landscapes and urban populations in relation to the processes that give rise to those patterns. By forming general understandings of such processes and of the inter-relationships between processes, people and places there is a possibility both to compare the socioeconomic, demographic or environmental structures of different places at particular points in time, and to predict and simulate how settlements might

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

evolve in the future. These understandings are essential to the governance and management of urban environments and systems. However, the detailed richness of any (specific) representation is always depleted by the process of generalisation when hard decisions are made regarding what to deem as irrelevant detail in order to establish baselines of comparability. Rarely is all of what is discarded wholly irrelevant to the process of building representations (Longley *et al*., 2001, Chapter 3).

Donnay *et al.* (2001) make this clear in the context of urban remote sensing and image classification. They see advantages of a generic model in the power of constructing a relatively simple (and understandable) representation of the properties of a large number of objects and in the portability of that general model from one location to another, permitting comparison between places. Yet, the match between a generalised model and the specific patterns observed in any scene normally will not be exact; there is inevitable uncertainty associated with the application of a generic model to a particular setting. Such uncertainty is likely to be greater than that associated with more specific, finely focused case studies. But if the merits of generalisation are accepted as worthwhile, then attention focuses upon how to minimise the uncertainty associated with the practice of abstraction and generalisation; how to ensure general models are 'fit for purpose' and usable.

In the past, outdated, imprecise, coarsely aggregated and indirect data sources have done urban geographers few favours in their search for empirical regularities and general understandings of urban growth, structure and morphology. Given those circumstances it is unsurprising that many geographers have become disenchanted with working with inert spatial mosaics and sometimes absurdly surrogate data, instead channelling their efforts to detailed, qualitative and ideographic approaches. At its best, urban geography continues to generate valuable insights into urban lifestyles and landscapes. However, there is also a gathering sense that the subject has become overly preoccupied with introspective and individualistic case studies and anecdote at the expense of more system-wide understandings of urban environments and systems. It is ironic (or perverse) that the relative abandonment of systematic, quantitative approaches in much of urban geography has come at a time when the data infrastructure for representation building has improved enormously, offering more direct and meaningful sources of information and analysis. In the UK, for example, Ordnance Survey (OS) recently has launched MasterMap, an ambitious project to provide a consistent, national framework for referencing geographic information. The foundation of MasterMap is a database containing information about polygons that together represent 400 million man-made and natural features (for example, individual fields, lakes or pillar boxes) and includes National Grid, GPS and detailed topographic information for all features (see www.ordsvy.gov.uk).

We are not suggesting that only that (large, and increasing) subset of phenomena that can be incorporated in quantitative representation should be deemed worthy of academic study; nor are we implying that better data and better data-handling technologies are a panacea to the limits of GIS-based representation. However, it is increasingly apparent that the worlds of business and commerce, with their pragmatic tactical and strategic concerns, better understand the importance and power of generalisation than much of academia. We view this as an odd, undesirable and some-

what self-defeating state of affairs when policy-makers at many levels of governance (in both the private and public sectors) are increasingly becoming minded to ask geographical questions, and look to geographical data and analysis for the answers. Consider, for example, the UK Government's Policy Action Team (PAT) 18 report on better information which states that 'comprehensive and up-to-date [geographical] information about deprived neighbourhoods is crucial to the National Strategy for Neighbourhood Renewal [. . .] Far too often information is seen as an additional and peripheral issue in regeneration activities, rather than as central to them. In reality, information is one of the key tools in the drive towards a more inclusive society' (PAT 18, 2000, p. 5).

One outcome of the Action Team's report was the setting up of a Neighbourhood Statistics Service (NSS) for the UK (see www.statistics.gov.uk). Yet, despite this important initiative we are actually little closer to answering one of the main questions that acted as a catalyst to creating PAT 18 and the NSS in the first instance. Specifically, at what scale does deprivation and social exclusion exist and persist? That is a question that the information currently provided by the NSS cannot answer because it is constrained to a particular, possibly inappropriate, zonal geography prior to dissemination (namely, to the 1991 UK Census ward geography). Use of the information presumes the census geography is suited to the task of targeting deprivation and measuring variation in socioeconomic conditions – a presumption that we have challenged (Harris and Longley, 2002).

Elsewhere we have followed the lead of business and commerce and looked to the datasets they are using to try and produce more finely attuned and targeted measures of deprivation (Harris and Frost, 2003; Harris and Longley, 2002). We have sought to navigate a path between the polarities of 'scientific but outdated and dull' academic research, and 'unscientific but rich and timely' analysis of private sector databases (Longley and Harris, 1999). In this chapter we continue along that path in the context of studies of deprivation in urban areas. We argue that clearer thinking about the conception, measurement and analysis of deprivation has far-reaching implications for anyone with interests in social exclusion. We focus on establishing the scale at which private sector sources of income information can be utilised to provide robust and comparable estimates of geographical concentrations of deprived (or, conversely affluent) income groups.

6.2 Geographies of Deprivation

The conception of deprivation is very much a contested domain: different conceptions lead to different geographical representations of physical and social conditions. Any representation can come to reflect the priorities of those who do the representing. The nomenclature of deprivation is often linked to notions of poverty, inequality and social exclusion. It can also have an absolute or relative meaning. Few could argue that the estimated average 24 000 people who die each day worldwide from hunger – 75% of whom are children aged under five – are not deprived (The Hunger Site: www.thehungersite.com). However, it is a moot point whether those who (are

said to) wear new clothes, own a microwave and also a VCR can honestly be described as living in poverty (*Daily Telegraph*, 2000 responding critically to a report by the Joseph Rowntree Foundation: Gordon *et al.*, 2000).

In generally affluent societies such as the UK's, few people are deprived of access to food, clean water, health, shelter and education (the UN's definition of absolute poverty). However, that does not mean that levels of wealth, well-being and social opportunity are harmonious across society – they are not. Hall and Pfeiffer (2000, p. 81) write that 'the optimistic high-growth period after World War II saw poverty in retreat; even experts could not imagine that inequality would increase again – yet in many countries, and cities, it has' (they cite London, New York and Los Angeles as examples). The emphasis on inequality is indicative of much deprivation-related research in the UK. For example, Shaw *et al.* (1999) and Dorling *et al.* (2000) focus upon inequalities in health outcomes in Britain. Measures of deprivation assume relative connotations, often defined against mean or median household. But this raises the question of what constitutes an inequality? The answer is necessarily subjective. Ultimately, the way in which deprivation is conceived conditions the way that it is measured (Yapa, 1998).

Even if there was consensus about the physical and social measures and manifestations of deprivation, to identify, measure and target geographical clusters of the deprived 'on the ground' requires consistent and task-relevant geographical information to be available to the analyst or policy maker. The usual areal 'containers' of deprivation in the UK – that is the zones to which social and physical characteristics are ascribed – are (1991) Census wards, containing on average a household population of 2081 but varying from a minimum of 153 households to a maximum of 9591 (Openshaw, 1995b). Are these geographical zones really the best suited for measuring and monitoring deprivation? The decision is not neutral in its effects: the choice of geography can alter the apparent intensity and extent of deprived conditions in various locations (Openshaw, 1984). Scale and aggregation effects have profound impacts upon the representation of measured attributes.

Differences of conception compounded with vagaries of measurement have far-reaching consequences for the identification and representation of deprived areas. Harris and Longley (2002) contrast the outcome of Lee's (1999, p. 178) study of local conditions with those of Gordon and Forrest (1995). The former, calculated using a central government measure (Department of the Environment, 1981), suggested that 40% of the most deprived wards in England and Wales are located in Greater London and South East England. The latter (the 'Breadline Index', described below) suggests that the same two regions actually account for 24% of the most deprived wards.

Harris and Longley (2002) also consider Lee *et al.*'s (1995) review of areal measures of deprivation, including the (then) Department of the Environment's 1981 and 1991 Indices (Department of the Environment, 1981, 1994), Jarman Scores (Jarman, 1983), the Townsend Index (Townsend *et al.*, 1988) and an index developed from the Breadline Britain poverty study (Gordon and Forrest, 1995; Gordon and Pantazis, 1995). Lee *et al.*'s study emphasises how apparent similarities or differences between various indices are a function of the variables that make up a particular index and also the scale at which the comparison is made. Lee *et al.* conclude that the various indices for measuring and mapping deprivation differ in terms of how they are

constructed, what their constituent variables purportedly represent and how they are weighted. Despite this, at the coarse Census district scale of analysis, the different indices produce similar deprivation profiles and identify similar lists of the most deprived areas. However, at the ward level and below, the opposite is true – different indices produce quite different results. Multivariate indices produce better representation of the multifaceted nature of deprivation than univariate measures and can identify different populations and areas as being deprived. Yet, a large proportion of the households identified using univariate measures such as unemployment or overcrowding will also be flagged in multivariate indices.

6.3 Census-based and Hybrid Measures of Deprivation

The indices that Lee *et al.* (1995) review are typical of those that are reliant (either directly or indirectly) upon data collected in the UK Census of Population – the only comprehensive data source for mapping the widely recognised indicators of deprivation or poverty within small areas (Lee, 1999, p. 172). The 'preferred status' of the Census arises from the high proportion of the population that it enumerates and the high public sector quality control standards that govern its collection. Nevertheless, its reputation was tarnished in 1991 because of the 'missing millions' omitted (following the debacle of the Poll Tax: see Openshaw, 1995a, p. 14) and the confusions surrounding whether students should be enumerated at home or at their term-time address (Mitchell *et al.*, 2002). The overall response rate for the 1991 Census was about 97% of Britain's 55.94 million residents (Champion, 1995), but with increased under-enumeration of particular sub-groups of the population, notably city dwelling, young males. The response rate for the 2001 UK Census is estimated at about 98% but with an increased non-response rate for some questions. Questions that in 1991 met with a non-response of less than 1% tended to have, in 2001, a non-response of 1–5%. Questions with a non-response rate of 1–5% in 1991 had a rate of 5–10% in 2001. Almost one in five individuals did not complete the 2001 census question enquiring about their professional qualifications (http://www.statistics.gov.uk/census200/pdfs/dataquality.pdf). The trend to increased under-enumeration is worrying, though in an international context it is worth mentioning that the overall response rate achieved by the 2000 US Census was only 67%. Nevertheless, it is clear that modern lifestyles and attitudes make it increasingly difficult for even compulsory public sector surveys to achieve complete enumeration of the population. Ironically, the rather narrow range of questions which the UK Census asks sheds little light on those modern lifestyles and attitudes (Longley and Harris, 1999).

A particular problem in the context of deprivation analysis is that UK Census data contain no direct indicators of income, which most would agree to be a direct and relevant measure of deprivation. Analysis is also constrained by the 10-yearly interval of Census data collection and the resulting problem of data ageing. In the UK there is still no strategy for either enriching the Census using a wider range of pertinent indicators or improving temporal frequency at the neighbourhood scale. In the US and UK alike there is, however, an open discussion of replacing census

collection with a fusion of other public or private sector databases to obtain the same sorts of information collected by the Census (and more). At their best, the data arising out of such approaches would be more wide-ranging and more frequently updated than conventional censuses. However, the vagaries of data fusion raise important issues of measurement error in the source data and of tracking and quantifying error propagation as the different datasets are merged.

The limitations of the Census for socioeconomic research apply equally to business and service planning applications using geodemographic classifications, such as ACORN, Super Profiles or other census-based typologies (Longley and Clarke, 1995; Birkin *et al.*, 2002). Sleight (1997) describes geodemographics as the 'analysis of people by where they live' – an apt description for area-driven deprivation research also. Indeed, Voas and Williamson (2001) describe geodemographic classifications as no more than 'dressed-up measures of affluence' (conversely, measures of deprivation). Further similarity is found in the efforts of practitioners from both research fields to 'freshen-up' census data by linking them to more up-to-date and relevant but less geographically replete samples of survey data. Batey and Brown (1995), for example, describe how the Census core of the Super Profiles classification is 'cocooned' by an updateable layer of non-census data.

In the context of deprivation measures, the 'Breadline Britain' index is an interesting hybrid measure that has been described by Lee *et al.* (1999) as the best index for identifying the most deprived wards (in Britain) at a national level. Full details of the design and implementation of this index are given by Gordon and Forrest (1995), Lee *et al.* (1995), Gordon and Pantazis (1997a, 1997b), Gordon *et al.* (2000) and also Harris and Longley (2002). In brief, it is based on a quota sample of 1319 individuals aged 16 or above supplemented by an additional 512 quota interviews carried out in low income areas: areas identified as Group G ('Council Estates') of the 1981 ACORN geodemographic system (Gordon and Pantazis, 1997a). Interviewees were asked to identify from a list of 44 items (that covered a range of activities and possessions) those that were felt necessary and important to everyday living (Lee *et al.*, 1995). Items attracting a total of 50% or more support were deemed as socially perceived necessities by 'a democratic majority' (Gordon *et al.*, 2000, p. 13). A poverty line was then set at households lacking three or more necessities. On this basis the sample was classified according to an 'in poverty' (1) versus 'not in poverty' (0) dichotomy. It was assumed that few differences exist between different sections of the population as to what is regarded as a necessary item.

The results of the Breadline Britain sample survey were then generalised to the population at large. The survey data were limited to 1831 individuals. Because no similar information existed for the rest of the population it was not known (at a national level) who else in the population owned fewer than three of the necessities. To resolve this problem a link was created between the sample data and the more restricted range of variables available in the 1991 Census small area statistics by identifying variables common to both. A logit analysis was undertaken at the household scale, with deprived/not deprived as the response variable. The explanatory variables were the socioeconomic indicators of the survey that could be paired with like variables in the census (Gordon and Forrest, 1995). The regression parameters arising

from the analysis were subsequently transferred from the household scale to weight corresponding census variables at both the ward and census district scales (Lee *et al.*, 1995). The resulting Breadline Britain Index is used to estimate the percentage of 'poor households' using the following regression relation:

$$y' = 0.2025x_1 + 0.2174x_2 + 0.1597x_3 + 0.1585x_4 + 0.0943x_5 + 0.1079x_6 \qquad (6.1)$$

where y' is the estimated percentage of poor households (per census ward or district) and x_1 to x_6 refer to the following 1991 census variables respectively: % households not owner-occupied; % households with no car; % lone parent households; % households where household head is of partly-skilled or unskilled occupation; % economically active population unemployed; and % households containing a person with a limiting long-term illness. Each percentage figure pertains to a census ward or a census district, depending on the chosen scale of analysis – the parameter estimates are (wrongly) taken to be independent of the scale of analysis and hence remain constant.

Although the method of the Breadline Index is an innovative means to break out of the straightjacket of conventional, census-based deprivation measurement, the user remains constrained by two main shortcomings. First, the Index is based on a small sample and generalised to the whole UK in a geographically invariant way. A single, national set of parameter estimates is created by the linkage mechanism although it is likely that the weight given to each census variable will be more appropriate in some locations and some scales of analysis than others. Second, the Index is used (and is relatively robust) only at the census ward and district scales. There is increasing evidence that even small areas conceal diversity in population characteristics and social conditions (e.g., Longley and Harris, 1999).

6.4 'Small' Areas and the Problem of Diversity

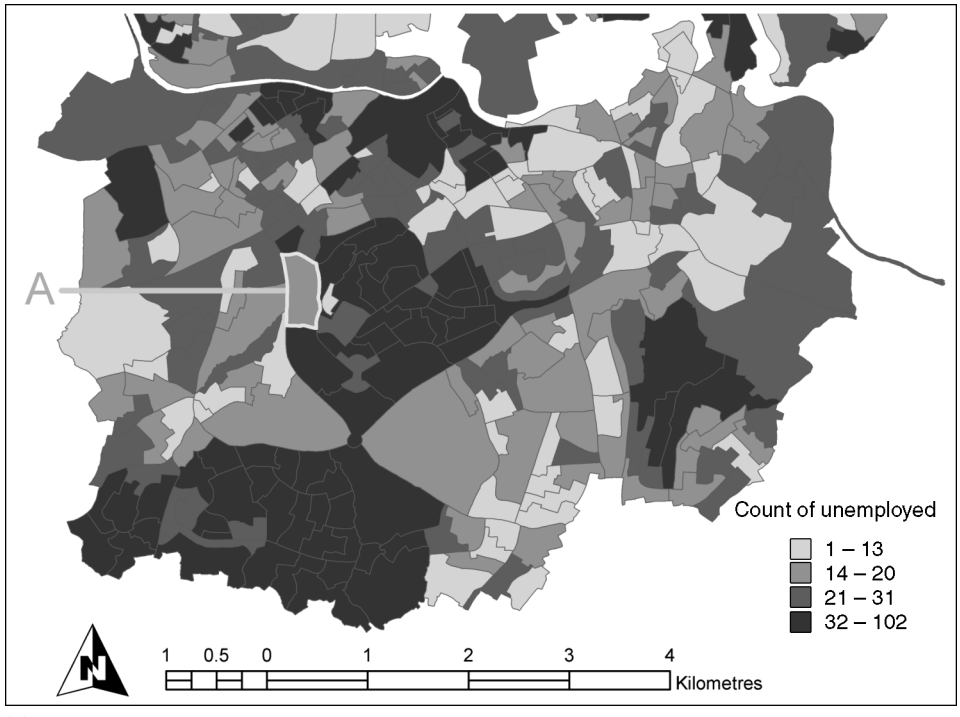
> *'And cities? They are diverse'* (Amin *et al.*, 2000, p. 22).

A significant weakness with the use of census data for deprivation and other area-based analysis is the forced assumption that populations within 'small' areas – meaning census enumeration districts (EDs) or wards – are homogeneous in terms of the socioeconomic conditions in which they live; or, at least, that suppression of within-area diversity is a sacrifice worth making in order to forward the cause of generalisation. Much of our research has contested that assumption (or suppression), focusing on diversity *within* administrative areas. That is not to say that geographical trends and patterns do not exist *between* administrative areas; on the contrary, our argument has been that they do (Harris, 2001; see also Voas and Williamson, 2001). But what of trends and patterns that transcend administrative areas, or are hidden within them? The principal issue is not whether there is a geographical patterning imprinted on the socioeconomic landscape around us – undoubtedly there is – but, instead, the extent to which such trends and patterns are obscured or clarified

by the nature of areal aggregations, including those of the census, or by sources of geographical information that are hardly fit for purpose.

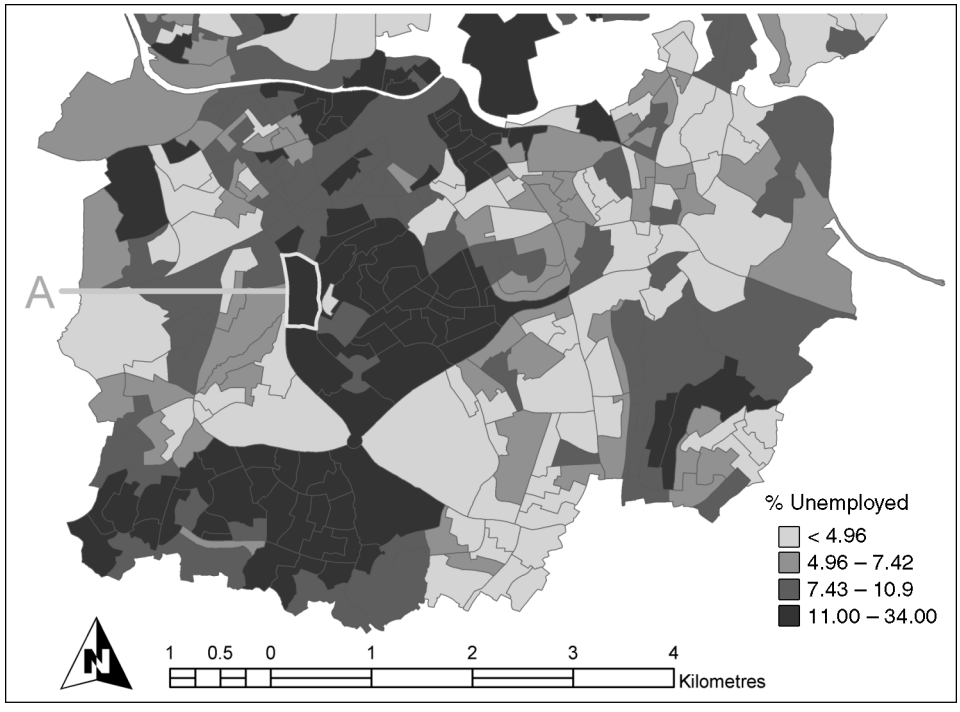
If imperfections and mismatches between spatial distributions are likely to exist, a role for GIS is in quantifying the sources and operation of uncertainties through internal and external validation. Internal validation might, for example, entail experimentation with the scale and aggregation effects arising from different combinations of the smallest available areal units of analysis (e.g., Openshaw, 1984; Alvanides *et al*., 2002). Such experimentation can generate a large range of outcomes but, at the end of the day, there is no incontestable 'optimal' solution. Worse, the results must still take the within-area distributions of the elemental units of analysis as uniform. External validation entails use of ancillary sources that are not constrained by the same (or, ideally, any) zonal geography to reveal more about small area diversity.

A simple example is given by Figures 6.1a and 6.1b, showing the southern half of a map of unemployment in Bristol, for which the unemployment attribute is divided into quartiles. Note that the spatial distribution appears different whether a count of unemployment (Figure 6.1a) or a rate of unemployment (Figure 6.1b) is derived. Consequently it is unclear whether the zone labelled A belongs in a functional sense to the zones to its east or to its west. There is little way of resolving this issue using



(a)

Figure 6.1 *Measures of unemployment in South Bristol, UK*



(b)

Figure 6.1 *Continued*

census data alone. Yet if the area is viewed using high resolution aerial photography, it becomes clear that the area is of mixed land-use: part residential; part industrial; part open space (Figure 6.2). The (small) population of the zone is concentrated to the north-west. Within GIS, a procedure known as dasymetric mapping (Donnay and Unwin, 2001) could be used to mask out unpopulated areas partially in order that census enumeration district (zone) counts might be averaged only over the built-up portion of the zone. In the same spirit, Longley and Mesev (2002) integrate remotely sensed and census data to produce a hybrid representation of the spatial distribution of occupied built structures, and thence create fine-scale density profiles of a city (see also Harris, 2003).

In each of the examples, the ancillary source (aerial photography or satellite imagery) does not identify unique geographical individuals: the element of analysis is determined by the granularity of the image; each individual, pixelated element might enumerate more than one, zero or indeed fractions of human individuals. A different approach is to use ancillary sources that contain individual population records. In this we have taken our cue from the marketing industry, where the era of information commerce and 'one-to-one' marketing has given rise to detailed, disaggregate datasets, commonly about people, their 'lifestyle preferences' and their shopping habits (Longley and Harris, 1999).

Figure 6.2 *High resolution aerial photograph showing land use in zone A. Source: Cities Revealed (1996)*

With the variety of analytical tools now commonplace in GIS software, there exists potential to adopt more flexible and geographically sensitive approaches to classification and modelling of socioeconomic conditions. This provides some defence against the charge that today's planners and policy makers have much better information about the attributes of the built environment from the greatly increased resolutions offered by remotely sensed (RS) imagery (Donnay *et al*., 2001) than they do about the most important part of an urban settlement: the people who live and work there (Harris, 2003).

6.5 Refocusing on Income

To summarise, the lack of conventional survey alternatives has forced reliance on the census as the only viable source of fine-scale socioeconomic data (Green, 1997). Where alternative public sector data do exist they tend to be *ad hoc* and geographically limited, generated in response to the specific needs of individual councils, planning departments and research projects (e.g., Longley *et al*., 1989) and thus not well suited for more general model building or comparative analysis. However, the data blockages to less ambiguous measures of poverty are not insurmountable. Official

(governmental) datasets do not provide the only useful sources of usable information in today's information commerce, although academics and marketeers do appear to have differing perspectives on what is attainable. In the academic domain, for example, Green (1997, p. 181) laments the '. . . dearth of spatially disaggregated [income] data at the local (particularly sub-county) scale'. In a similar vein, Lee (1999, p. 175) notes that the use of unemployment rate as a principal indicator of poverty is largely a response to the lack of household income data. By contrast, private sector companies such as EuroDirect offer commercial products which allow unit postcodes to be selected on the basis of the household income levels found within them (EuroDirect, 1999). A recent study by the market research group CACI also focused upon household income and claimed to give a truer picture of wealth differentials at disaggregate scales than measures provided by the unemployment register (Baldwin, 1999).

There are a number of advantages to the use of private sector 'lifestyles' datasets that can be defined as composite databases created and updated from sources of individual and household data including shopping surveys, loyalty card information, credit ratings and the guarantee returns from consumer products. The advantages include the fact that lifestyles data can be made available to describe households at the scale of the unit postcode. This frees analysis of heterogeneity in social conditions from the out-dated 'mosaic metaphor' which underpins much census-based urban geography (Johnston, 1999; Raper, 2000). The data are also more up to date (shopping surveys of this nature typically are conducted annually) and thus present a more relevant picture of conditions in fast-changing urban systems. Most importantly of all, lifestyles datasets often do have an income question, the nature of which we consider below. This obviates the need to fall back on the indirect surrogate measures that characterise census-based deprivation studies. It is disingenuous to overlook these advantages, even if the quality of the data is at times considered uncertain or unknown (Goodchild and Longley, 1999; Harris and Longley, 2002).

Harris and Frost (2003) argue that despite the uncertain quality of lifestyles data, they can still be used to identify, at the unit postcode level, the location of every survey respondent (in the Borough of Brent, London, in their case study) who indicated their annual family income to be less than £10 000. The respondents to lifestyles surveys are self-selecting from the mailing lists and other sample frames used by lifestyles companies (there is no compunction upon households to complete and return the survey). Nevertheless, if low income respondents are shown to be clustered spatially, then there is cause to believe that there is a clear areal basis to income deprivation, particularly if a reasonable number of the clusters are shown to be within, or in proximity to wards identified as deprived according to more orthodox measures. Lifestyles data and more conventional sources can be used to cross-validate each other, or be combined to produce more data-rich models of the urban fabric (Harris and Longley, 2001).

To map or to create spatial statistics describing any of the data stored in a data warehouse necessarily requires a way of assigning location to each individual or household record. One way of doing this is to use (GB) Ordnance Survey's (OS) Address-Point product to ascribe a UK National Grid reference of up to 0.1 metre precision to the full property address. An alternative is to use OS Code-Point data

which are cheaper for commercial organisations to purchase and are less voluminous to process. Whereas Address-Point assigns (in principle) a georeference to each residential and commercial address in Britain, Code-Point is intended for analysis at the slightly coarser, unit postcode scale (with an average of 12–16 residential delivery points per postcode). The Code-Point file lists, for each postcode in Britain, the grid reference of a point that is located within the footprint of the property closest to the population-weighted centre of the postcode. A Code-Point is defined by first calculating the average Easting and Northing location for all Address-Points within a postcode, and then by identifying the Address-Point closest to that location. The closest Address-Point defines the Code-Point for the postcode (although the grid reference is degraded to a maximum of 1 metre precision). Further comparison between the Address-Point and Code-Point products is made by Harris and Longley (2000).

Code-Point, Address-Point and the MasterMap products each offer ways of ascribing point locations to new souces of attribute data at ever finer levels of granularity. A problem, however, is that points are, by definition, dimensionless and, in a strict sense, occupy zero area on a map. This is clearly a problem if the intention is to use such data to explore and represent socioeconomic distributions! Commonly the problem is 'resolved' by representing a point using a circular symbol, the radius, circumference, area or shading of which can be arbitrary, tailored to enhance the aesthetics of the map or proportional to the value of the phenomenon at the point. Alternative, space-filling methodologies are less cosmetic and may cause permanent changes either to the source data structure or to the attribute table. A variety of kernel, neighbourhood, density and other estimation techniques exist that convert vector points to a near-continuous surface of raster cells, or append a measure of the local intensity of a given phenomenon to the dataset, on a point-by-point basis.

In the context of modelling lifestyles data to produce local indicators of socioeconomic conditions, the advantages of kernel or neighbourhood estimators have been documented by Harris (2003). By analysing each record – each individual or household – as part of a larger and geographically clustered group the sample size is increased and the error component is (in principle) reduced. This approach helps counter the problems of sample and response bias that affect lifestyles datasets to varying degrees (for example, there is evidence that postal shopping surveys tend to miss young, male adults: Longley and Harris, 1999). However, it also raises questions concerning the most appropriate scale at which the geographical clusters should be defined; what is 'the' correct level of aggregation? What neighbourhood, kernel or search window size should be used? This is an important consideration, since the choice has a considerable effect on the amount of smoothing applied to the data. Local estimation techniques are inherently spatial statistics in that the results (the output) are by no means independent of the geographical scale of the analysis (the size of the search window).

In their study of small area income differences within Brent, London, Harris and Frost (2003) choose a circular search window of radius 500 m. But their choice was essentially arbitrary, based only on experience of 'what seems to work'. Even assuming that the chosen scale was indeed appropriate to the analysis, a second question is immediately raised: is the search window size equally appropriate across the entire study region or should it vary geographically in order to adapt to local conditions?

An example of the latter, adaptive approach is Bracken and Martin's population surface algorithm (see Martin, 2002) designed for remodelling census population data into more realistic portrayals of population distributions. The algorithm incorporates a kernel, the size of which is sensitive to local population densities: the kernel is smallest where densities are highest; widest where densities are lowest.

An analogy to fixed versus adaptive search windows is provided by the histogram. A histogram can be plotted using equal class intervals (fixed bin widths) but unless the observations are uniformly distributed between the classes, differing numbers of observations in each class will result. The alternative is to adopt a quantile classification – as in Figure 6.1 – for which the numbers in each class are harmonised by using unequal bin widths (class intervals). The relatively simple procedure outlined by Harris (2003) uses the 'quantile approach', finding the n_i points from a total set of N that are calculated as being at least distance from a single point i (which is also drawn from the set of N). These n_i points, together with point i itself, are then taken to form a 'cluster' or cloud of $n_i + 1$ points and a total population size of m:

$$m = p_{0i} + p_{1i} + p_{2i} + \ldots + p_{ni} \tag{6.2}$$

where p_{0i} is the population at point i, p_{1i} is the population at the (first) nearest neighbourhood point to i, p_{2i} is the population at the second nearest neighbour, and so forth to the nth neighbour. Each of the points is a residential (unit postcode) CodePoint, and the straight-line distance measure does not attempt to accommodate explicitly the effects of built form or spatial structure (e.g., where actual travelled routes differ significantly from the straight-line crow's distance between points or where, in reality, a physical barrier exists that clearly separates two or more points, all be they in seeming close proximity).

The procedure is then repeated for each of the remaining $(N - 1)$ points in turn. The method therefore creates as many clusters as there are points (N). Most households will be allocated to more than one cluster because any individual point can be a nearest neighbour of more than one other. The value of n_i, the number of CodePoints in each group, will vary from cluster to cluster as it is m (the total population size) that is held constant. The value of n_i is governed by both the variable size of postcode populations and the local sample/response rates for the lifestyles survey data that Harris (2003) analysed. Having grouped the households, Harris was able to identify the most frequently occurring (modal) property type within each cluster and also the proportion of households residing in that type. If there are three property classes and the number of households belonging to each class is z_1, z_2 and z_3 respectively, then, assuming z_1 is the modal type:

$$z_1 > z_2 \quad \text{and} \quad z_1 > z_3 \tag{6.3}$$

and

$$p_{z1} = z_1/m \tag{6.4}$$

where

$$m = z_1 + z_2 + z_3 \tag{6.5}$$

or

$$m = p_{0i} + p_{1i} + p_{2i} + \ldots + p_{ni} \tag{6.6}$$

where p_{z1} is the proportion of all households belonging to the modal type.

The proportion – a simple measure of within-cluster homogeneity – was assigned back to the point (i) at the centre of each cluster and this information was used to guide the selection of training samples to classify a remotely sensed image into classes of residential building densities. Training samples were drawn from a Bristol study area where it was most certain that a particular property type existed: areas with the greatest local homogeneity of a dominant property type (Figure 6.3).

If the procedure described by Harris (2003) circumnavigates the issue of what is the correct search window size to aggregate the lifestyles data, it immediately encounters a related problem instead: what is the correct class membership size, m? The Bristol study took a value of $m = 50$ but this was largely chosen to illustrate the analytical methodology rather than for any more considered reason. For the rest of this chapter we therefore consider the effects of changing the value of m on small area measures of modal income and of income homogeneity in London.

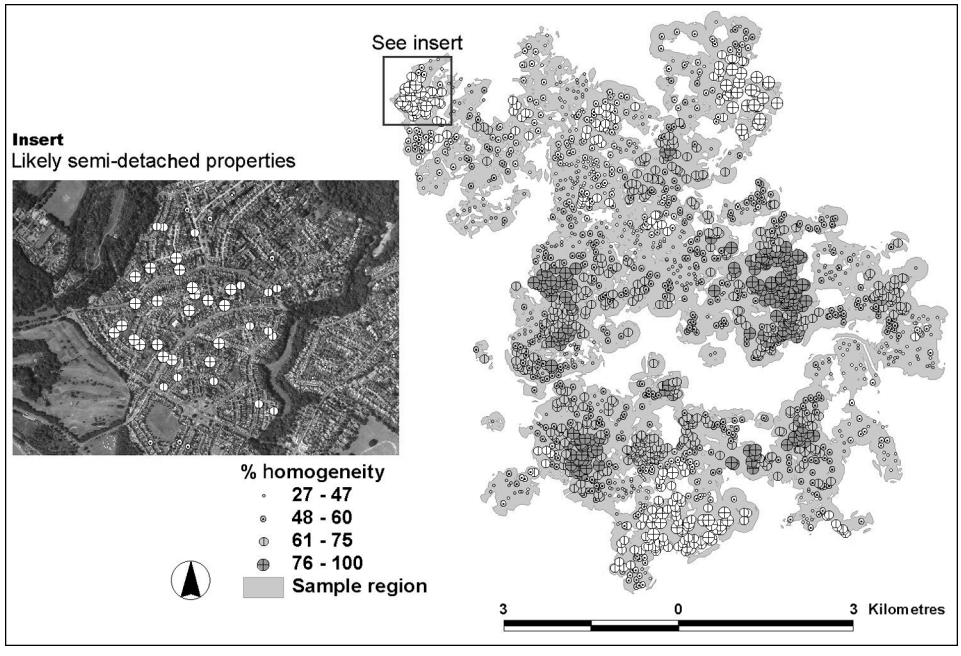


Figure 6.3 *Combining lifestyles and CodePoint population maps to select, from a Cities Revealed image, training samples with high likelihood of being semi-detached properties. Source: Harris (2001)*

6.6 Modelling Income Concentrations in London

Figure 6.4 shows a map of low income clusters in Inner London derived using the methodology described by Harris (2003) but with a value of $m = 75$ and actually undertaken for the whole Greater London region (only Inner London is shown to retain visual clarity of the income distributions). The source data is a commercial lifestyles database compiled from replies to a national consumer survey. That survey was postal, undertaken in England, Wales and Scotland during the spring of 1999, and sent out to a high proportion of households for whom at least one member was listed on the UK electoral register for the three countries. The full lifestyles database contains a wide range of socioeconomic, behavioural and consumer information that describe approximately one million households. We have selected from that dataset those households that gave an indication of their gross household income and have then georeferenced those records, by residential postcode, to an OS CodePoint. This leaves us with a sample of 65 685 households in 42 883 unit postcodes, all of whom

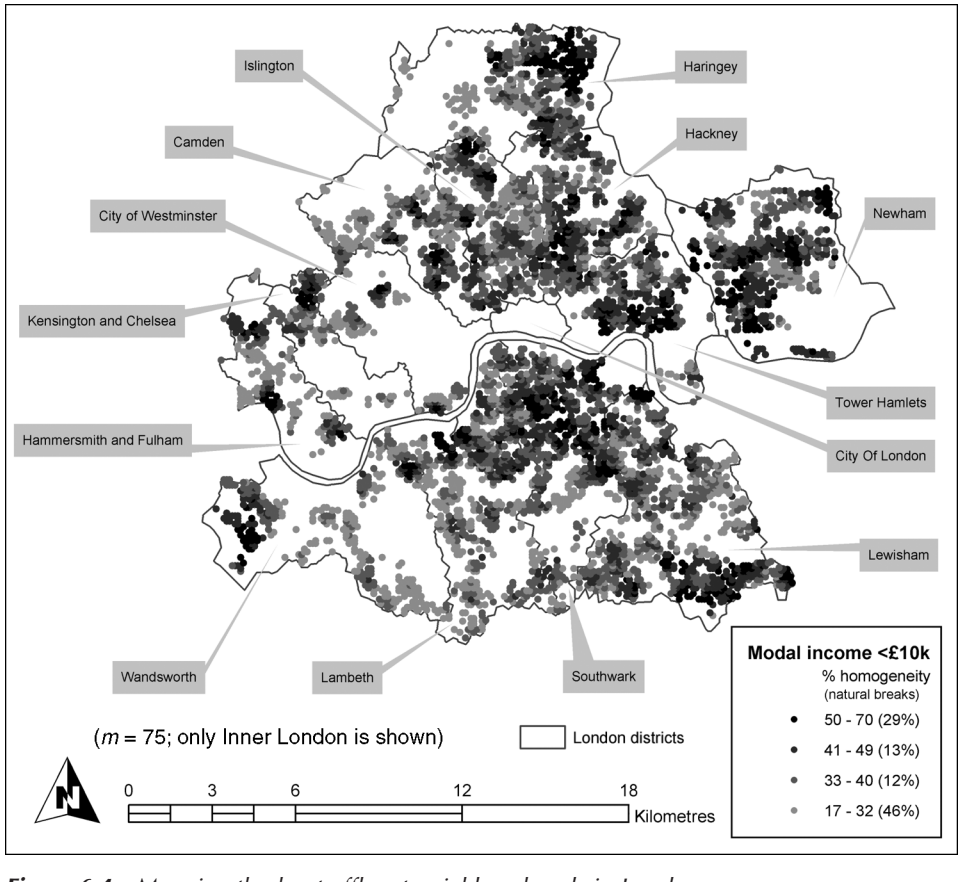


Figure 6.4 *Mapping the least affluent neighbourhoods in London*

have a household income estimated to be in one of six income bands: <£10000; £10000–£19999; £20000–£29999; £30000–£39999; £40000–£49999; £50000 and above. According to Department of Transport, Local Government and the Regions' projections from 1996 household numbers, there are 3.13 million households living in Greater London (http://www.statistics.gov.uk/statbase/Expodata/Spreadsheets/D4524.xls). Hence, the lifestyles income sample is of a size equivalent to 2.1% of households (consistent with Harris and Frost (2003) who estimated the sample size to be just under 2% for the Borough of Brent within London).

Applying the modelling procedure described in Section 6.5 permits the determination of which one of the six income types is dominant in each small area cluster (formed around each of the 42883 CodePoints) and also what proportion of sampled households are of the modal type per cluster. In Figure 6.4 we have grouped the classes to emphasise areas (in Inner London only) where the lowest incomes (<£10000) dominate, and in Figure 6.5 the highest (£50000 or above). The average circular search window required to find the n_i neighbours that give a total population of $m = 75$

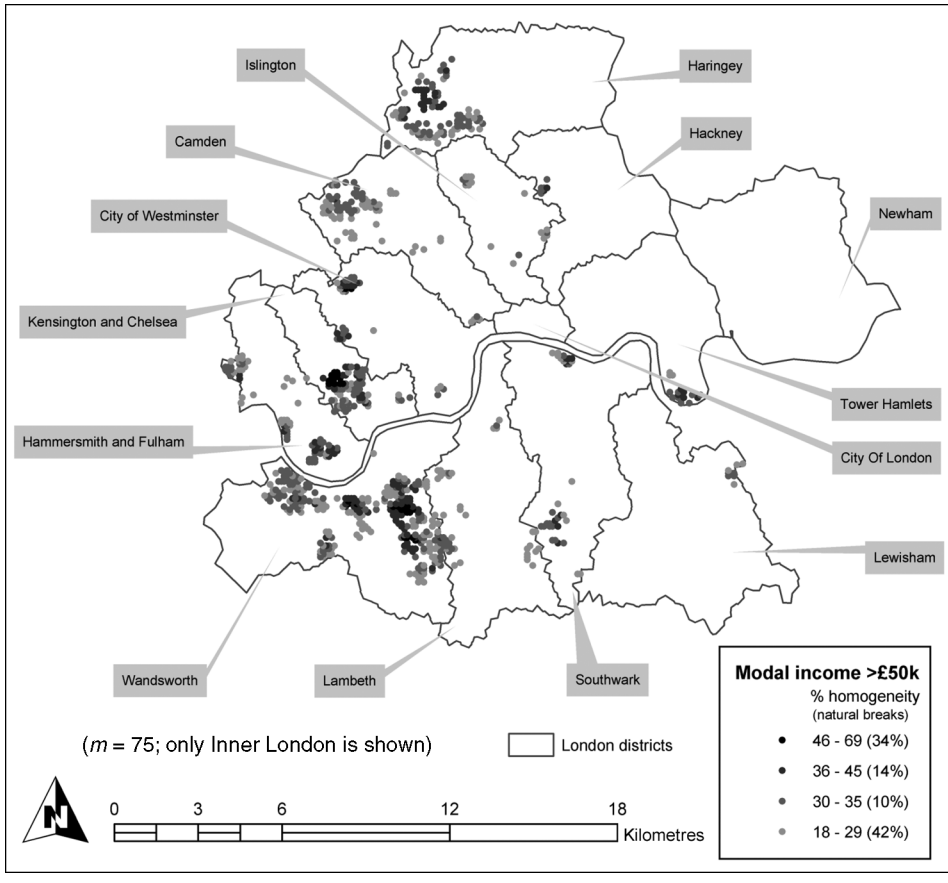


Figure 6.5 *Mapping the most affluent neighbourhoods in London*

per cluster has a radius of 452 metres. We have already noted that there is no reason to suppose that the distribution of lifestyles respondents is evenly distributed across the study region and, accordingly, the radius varies from a minimum of 141 metres to 3999 metres for the set of 42 883 points. The interquartile range is much smaller, however: from 361 metres to 510 metres. Both Figure 6.4 and Figure 6.5 show the proportion of households within the neighbourhood that belong to the modal class: higher proportions suggest spatially concentrated areas of deprivation or affluence.

With respect to our earlier discussion (Section 6.5), why did we choose a value of $m = 75$? In fact, a number of values for m were considered ($m = 25, 50, 75, 100, 200$). The results of the analyses with each of these values are summarised in Table 6.1. The top seven rows of the table (excluding the headers) give summary measures of the search window radii (in metres) required to cluster a total population of m by summation of the n_i records nearest to each of the 42 883 target points. For example, when m is set at 200, the mean average search window size required to gather the n_i points together to meet that population threshold is 768 metres but ranges from 361 metres to 4000 metres. The interquartile range is much smaller: from 608 metres to 854 metres. The general trend, to be expected, is that as the total population size required m decreases, so a smaller search window is required to find the n_i points that will meet the population threshold. Taking the mean, the third quartile (Q3) or the first quartile (Q1) as a measure of the level of aggregation required to meet each

Table 6.1 *Summary of analysis for various values of m*

m	200	100	75	50	25
Mean radius (m)	768	528	452	364	252
Maximum radius (m)	4000	3999	3999	3966	3820
Minimum radius (m)	361	200	141	100	0
Range	3639	3799	3858	3866	3820
Q3	854	600	510	412	300
Q1	608	412	361	283	200
IQ range	246	212	149	129	100
R ($m = 200$)		0.82	0.74	0.63	0.45
R ($m = 100$)	0.82		0.91	0.77	0.55
R ($m = 75$)	0.74	0.91		0.85	0.61
R ($m = 50$)	0.63	0.77	0.85		0.71
R ($m = 25$)	0.45	0.55	0.61	0.71	
Mode: < £10 k (%)	64	57	55	52	47
Mode: £10 k–£19.9 k	20	23	23	24	25
Mode: £20 k–£29.9 k (%)	9	12	13	14	16
Mode: £30 k–£39.9 k (%)	0	1	2	3	5
Mode: £40 k–£49.9 k (%)	0	0	0	1	1
Mode: £50 k+ (%)	6	6	6	6	6
No change from $m = 200$		74	69	62	54
No change from $m = 100$	74		82	70	58
No change from $m = 75$	69	82		77	61
No change from $m = 50$	62	70	77		68
No change from $m = 25$	54	58	61	68	

value of m, then it can be shown that the increase in resolution is in the order of 9:1 as m is decreased from 200 to 25 (the radius is one-third the length, hence the area is one-ninth), and of the order of 4:1 as m is decreased from 100 to 25.

Notwithstanding the apparent increase in precision, the next five rows of Table 6.1 warn that a low value of m is not necessarily to be preferred. These rows record the Pearson correlation coefficients for each set of 42 883 cluster homogeneity values at the five values of m. The homogeneity values give the proportion of the cloud or cluster of $n_i + 1$ points around each of the N points that are of the modal income type for that group of $n_i + 1$ points (i.e., the value p_{z1} in equation 6.4). The set of N proportions is calculated for each of the five values of m. This gives $(m - 1)! = 10$ pairs of cluster homogeneity values, for which the highest correlation, at $R = 0.91$, is obtained between the sets for $m = 75$ and $m = 100$ (Table 6.1). The lowest correlation is (unsurprisingly) between $m = 25$ and $m = 200$. From the table it appears that the homogeneity values for $m = 75$ and $m = 100$ exhibit greatest correspondence with the sets of values calculated for the remaining values of m.

The next six rows of Table 6.1 show the proportion of the N cluster centroids assigned to each of the six income classes at each value of m. Looking across the <£10k row it can be seen that the results are not spatially invariant: as the search windows reduce in size so the proportion of clusters with a modal income in this low income class decreases from 64% ($m = 200$) to 47% ($m = 25$): a 27% decrease. That the results should change with the geography of the problem is not surprising since what we are undertaking is explicitly spatial analysis in the sense defined by Longley *et al.* (2001, p. 278): 'spatial analysis is a set of methods whose results change when the locations of the objects being analysed change'. Here, whilst the location of the points is not actually changing, the boundary of the object class (or cluster) into which they are placed, is. It is this property of geographical contingency that makes it difficult to prescribe the correct value of m. However, bearing in mind the quest for generalisation discussed in our opening section, this is not to say that some values of m are not better suited to particular applications than others. Our criteria remain to construct small area measures of income that do not mask diversity at finer scales but that do aggregate the lifestyles data into representative clusters. Looking again at Table 6.1, and at the distribution of cluster centroids across the income classes, it is seen that a degree of consistency in the results emerge when $m = 75$ or $m = 100$ (we also saw this with the Pearson correlations). Interestingly, the proportion for the highest income class (£50k or above) is constant at 6% for all the values of m tested.

The final five rows of Table 6.1 indicate the movement across the income classes as m is changed. They record, for example, that of the 42 883 points, 74% remain ascribed to the same income class regardless of whether $m = 200$ or $m = 100$. As m is further reduced from 200 to 75, 69% remain in the same income class, and 54% as m is reduced from 200 to 25. The empirical case study confirms the intuitive expectation: that the results of the analytical method depend on the chosen value of m (and hence the search window size). However, the results are by no means unstable. The greatest invariance to scale appears to be when $m = 75$ or $m = 100$. To recall, we chose the former value for the results displayed by Figures 6.4 and 6.5. With $m = 75$, the average search window size has a radius, r, of 452 m (Table 6.1), giving an average

area of 0.64 km^2 over which the lifestyles data are aggregated (= πr^2). This is certainly greater apparent spatial precision than is offered by census ward estimates and also by most census classifications at ED scale. Most importantly, the method offers a direct way of estimating the concentrations of specific income groups using a 'bottom-up' approach that is not constrained to fit within an externally imposed, administrative geography such as that of the census.

6.7 Conclusion

The scale and pace of change in developed countries today is such that we cannot rely upon census data alone. In the UK, the terms of the 1922 Census Act are not interpreted today as providing a mandate for collecting income data. Yet income data, together with richer information on consumption in its broad senses, are central to creation of indicators that can facilitate sensitive spatial policy and rational resource allocation. Such information is, as the report of PAT 18 suggested, essential if the UK Neighbourhood Renewal Unit is to meet its challenge to ensure that 'within 10–20 years no-one should be seriously disadvantaged by where they live' (http://www.dtlr.gov.uk/neighbourhood/).

Developments in remote sensing now allow deprivation arising out of attributes of the built environment to be made consistent with the scale at which housing is in poor condition, or to the scale at which positive or negative externality effects of infrastructure or services are experienced. Yet much remains to be done in creating relevant, timely small area measures of social conditions. Hybrid approaches such as the Breadline Britain Index seek to lever maximum value from the census, yet ultimately only partly compensate for the absence of income measures and sensitivity to local neighbourhood effects. Lifestyles data include an income question, yet the validity of income measures at fine spatial scales can be hampered by the small and geographically variable number of responses received. In our view, it is necessary to undertake some level of aggregation when analysing the lifestyles data, yet to do so in such a way as to be sensitive to local patterns within the data. Aggregating the income information into artificial census units does not meet this challenge so here we have outlined a 'bottom-up' approach to representing income clusters at more precise scales than those offered by more orthodox measures.

The apparent 'Hobson's choice' either of using overly aggregated, surrogate and outdated data on social conditions or abandoning generality and quantitative analysis altogether seems to us a peculiarly academic construct that could be dismantled by an increased, and mutually beneficial, dialogue with business users of geographical information, even withstanding concerns about privacy and ethical issues of data usage (Curry, 1998; Longley *et al.*, 2001, Chapter 19).

Acknowledgements

The research was funded under the (UK) National Environment Research Council, Urban Regeneration and the Environment Programme (http://urgent.nerc.ac.uk/),

grant GST/02/2241. The authors are grateful to Claritas UK for supplying the lifestyles data analysed in Section 6, GeoInformation Group for the Cities Revealed data and Ordnance Survey for the CodePoint data. Any errors arising from the interpretation and analysis of these data are our own.

References

Alvanides, S., Openshaw, S. and Rees, P. (2002) Designing your own geographies, in Rees, P., Martin, D.J. and Williamson, P. (eds) *The Census Data System*, Wiley, Chichester, pp. 47–65.

Amin, A., Massey, D. and Thrift, N. (2000) *Cities for the Many Not the Few*, The Policy Press, Bristol.

Baldwin, P. (1999) Postcodes chart growing income divide, *The Guardian*, October 25.

Batey, P.W.J. and Brown, P.J.B. (1995) From human ecology to customer targeting: the evolution of geodemographics, in Longley, P. and Clarke, G (eds) *GIS for Business and Service Planning*, GeoInformation International, Cambridge, pp. 77–103.

Birkin, M., Clarke, G. and Clarke, M. (2002) *Retail Geography and Intelligent Network Planning*, Wiley, Chichester.

Champion, A.G. (1995) Analysis of change through time, in Openshaw, S. (ed.) *Census Users' Handbook*, GeoInformation International, Cambridge, pp. 307–35.

Cities Revealed (1996) Orthorectified, aerial photography, http://www.crworld.co.uk.

Curry, M. (1998) *Digital Places: Living with Geographic Information Technologies*, Taylor and Francis, London.

Daily Telegraph (2000) The 'absolute poverty' of language, 11 September.

Department of the Environment (1981) *Information Note No. 2*, DoE, London.

Department of the Environment (1994) *Information Note on Index of Local Conditions*, DoE, London.

Donnay, J-P. and Unwin, D.J. (2001) Modelling geographical distributions in urban areas, in Donnay, J-P., Barnsley, M.J. and Longley, P.A. (eds) *Remote Sensing and Urban Analysis*, Taylor and Francis, London, pp. 205–24.

Donnay, J-P., Barnsley, M.J. and Longley, P.A. (eds) (2001) *Remote Sensing and Urban Analysis*, Taylor and Francis, London.

Dorling, D., Mitchell, R., Shaw, M., Orford, S. and Davey Smith, G. (2000) The ghost of Christmas past: health effects of poverty in London in 1896 and 1991, *British Medical Journal*, **321**: 1547–51.

EuroDirect (1999) Promotional material, EuroDirect, Bradford.

Goodchild, M. and Longley, P. (1999) The future of GIS and spatial analysis, in Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (eds) *Geographical Information Systems: Principles, Techniques, Applications, and Management*, Second edition, Wiley, New York, pp. 567–80.

Gordon, D. and Forrest, R. (1995) *People and Places 2: Social and Economic Distinctions in England*, University of Bristol.

Gordon, D. and Pantazis, C. (eds) (1995) *Breadline Britain in the 1990s*, Joseph Rowntree Foundation, York.

Gordon, D. and Pantazis, C. (1997a) Appendix I: technical appendix, in Gordon, D. and Pantazis, C. (eds) *Breadline Britain in the 1990s*, Ashgate, Aldershot, pp. 269–72.

Gordon, D. and Pantazis, C. (1997b) Measuring poverty: Breadline Britain in the 1990s, in Gordon, D. and Pantazis, C. (eds) *Breadline Britain in the 1990s*, Ashgate, Aldershot, pp. 5–47.

Gordon, D., Adelman, L., Ashworth, K., Bradshaw, J., Levitas, R., Middleton, S., Pantazis, C., Patsios, D., Payne, S., Townsend, P. and Williams, J. (2000) *Poverty and Social Exclusion in Britain*, Joseph Rowntree Foundation, York.

Green, A. (1997) Income and wealth, in Pacione, M (ed.) *Britain's Cities: Geographies of Division in Urban Britain*, Routledge, London, pp. 179–202.
Hall, P. and Pfeiffer, U. (2000) *Urban Future 21: A Global Agenda for Twenty-First Century Cities*, E & FN Spon, London.
Harris, R. (2001) The diversity of diversity: is there still a place for small area classifications?, *Area*, **33**: 329–35.
Harris, R. (2003) Population mapping by geodemographics and digital imagery, in Mesev, V. (ed.) *Remotely-Sensed Cities*, Taylor and Francis, London, pp. 223–41.
Harris, R. and Frost, M. (2003) Indicators of urban deprivation for policy analysis GIS: going beyond wards, in Kidner, D., Higgs, G. and White, S. (eds) *Socio-economic Applications of Geographic Information Science Innovations in GIS 9*, Taylor & Francis, London (in press).
Harris, R.J. and Longley, P.A. (2000) New data and approaches for urban analysis: modelling residential densities, *Transactions in GIS*, **4**: 217–34.
Harris, R.J. and Longley, P.A. (2001) Data-rich models of the urban environment: RS, GIS and 'lifestyles', in Halls, P. (ed.) *Innovations in GIS 8: Spatial Information and the Environment*, Taylor and Francis, London, pp. 53–76.
Harris, R.J. and Longley, P.A. (2002) Creating small area measures of urban deprivation, *Environment and Planning A*, **34**: 1073–93.
Jarman, B. (1983) Identification of underprivileged areas, *British Medical Journal*, **286**: 1705–9.
Johnston, R.J. (1999) Geography and GIS, in Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (eds) *Geographical Information Systems: Principles, Techniques, Applications, and Management*, Second Edition, Wiley, New York, pp. 39–47.
Lee, P. (1999) Where are the deprived? Measuring deprivation in cities and regions, in Dorling, D. and Simpson, S. (eds) *Statistics in Society: the Arithmetic of Politics*, Arnold, London, pp. 172–80.
Lee, P., Murie, A. and Gordon, D. (1995) *Area Measures of Deprivation: A Study of Current and Best Practice in the Identification of Poor Areas in Great Britain*, University of Birmingham.
Longley, P.A. and Clarke, G.P. (eds) (1995) *GIS for Business and Service Planning*, GeoInformation International, Cambridge.
Longley, P.A. and Harris, R.J. (1999) Towards a new digital data infrastructure for urban analysis and modelling, *Environment and Planning B*, **26**: 855–878.
Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2001) *Geographic Information Systems and Science*, Wiley, Chichester.
Longley, P.A. and Mesev, T.V. (2002) Measurement of density gradients and space-filling in urban systems, *Papers in Regional Science*, **81**: 1–28.
Longley, P., Smith, R. and Moore, L. (1989) Cardiff House Condition Survey Phase 1: Inner Areas – Final Report, Keltecs Ltd.
Martin, D. J. (2002) Census population surfaces, in Rees, P., Martin, D.J. and Williamson, P. (eds) *The Census Data System*, Wiley, Chichester, pp. 139–48.
Mitchell, R., Dorling, D,. Martin, D. and Simpson, L. (2002) Bringing the missing million home: correcting the 1991 small area statistics for undercount, *Environment and Planning A*, **34**: 1021–35.
Openshaw, S. (1984) The modifiable areal unit problem, *Concepts and Techniques in Modern Geography*, **38**, Geo Books, Norwich.
Openshaw, S. (1995a) A quick introduction to most of what you need to know about the 1991 Census, in Openshaw, S. (ed.) *Census Users' Handbook*, GeoInformation International, Cambridge, pp. 1–26.
Openshaw, S. (ed.) (1995b) *Census Users' Handbook*, GeoInformation International, Cambridge.
Policy Action Team 18 (PAT 18) (2000) *Report of Policy Action Team 18: Better Information*, The Stationery Office, London.
Raper, J. (2000) *Multidimensional Geographic Information Science*, Taylor and Francis, London.

Shaw, M., Dorling, D., Gordon, D. and Davey Smith, G. (1999) *The Widening Gap: Health Inequalities and Policy in Britain*, The Policy Press, Bristol.

Sleight, P. (1997) *Targeting Customers: How to use Geodemographic and Lifestyle Data in Your Business*, Second edition, NTC Publications, Henley-on-Thames.

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and Deprivation: Inequality and the North*, Croom Helm, London.

Voas, D. and Williamson, P. (2001) The diversity of diversity: a critique of geodemographic classification, *Area*, **33**: 63–76.

Yapa, L. (1998) Why GIS needs postmodern social theory, and vice versa, in Fraser Taylor, D.R. (ed.) *Policy Issues in Modern Cartography*, Pergamon, Oxford, pp. 249–69.

7

Assessing Deprivation in English Inner City Areas: Making the Case for EC Funding for Leeds City

Paul Boyle and Seraphim Alvanides

Abstract

One of the key roles of the European Commission (EC) is to encourage a redistribution of wealth throughout the Community to reduce inequalities, both between and within countries. One mechanism for achieving this is the release of Structural Funds to those places that have particularly serious social and economic problems. This chapter reports on the attempt by Leeds City Council to secure Objective 2 status funding from the Commission, as part of a wider bid submitted by the Yorkshire and Humberside region. To be successful, Leeds needed to prove that the socioeconomic situation in its inner city was at least as bad as that in other cities in England. However, to achieve this required some computationally advanced methods beyond the scope of the Council's researchers. In collaboration, a methodology was devised which used some in-house software, embedded within a geographical information system (GIS), to show that Leeds did indeed have a strong case for support from the EC. Here we describe the methods used and the results for Leeds compared with the other cities in England. The chapter ends by commenting briefly on the positionality of researchers who undertake consultancy work of this type and suggests that we rarely spend enough time considering the interesting ethical dilemmas that work of this nature raises.

7.1 Introduction

One of the roles of the European Commission is to stimulate economic growth and encourage a redistribution of wealth that reflects the fact that some parts of the

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

European Union (EU) are relatively better off than other parts. Rather than simply redistributing funds at the national level, there is increasing recognition of the significant imbalances within countries between cities and towns that need to be addressed. This is particularly so as cities and towns are critical to regional development. In identifying an urban agenda, the Commission states that 'the starting point for future urban development must be to recognise the role of the cities as motors for regional, national and European economic progress' (European Commission, 1998a). In many European cities, however, severe social and economic problems hinder economic development and there is a growing realisation that these need to be addressed before economic growth can be achieved: 'The future of the towns and cities in the EU depends on fighting growing poverty, social exclusion and stemming the loss of certain urban functions' (Committee on Spatial Development, 1999). The EC response has been to distribute central funds to cities across the EU based on uniform criteria designed to identify those most in need of support. The Structural Funds Programmes therefore require that cities and towns with significant social and economic problems can be identified and that 'given the crucial role of towns for regional development and EU regional disparities, it is important for the effectiveness of regional policy that this funding be more explicitly related to urban needs and potential in the regions' (European Commission, 1998b).

The research reported in this chapter was undertaken as part of Leeds City Council's (LCC) attempt to secure Structural Funds from the EU by applying for Objective 2 status funding within a broader bid made by the Yorkshire and Humberside region. For LCC, there was a certain irony in making an Objective 2 application, as in recent times the city has been heralded as one of the fastest growing and most economically successful in the United Kingdom (UK). Thus, on the one hand, the LCC was producing documents aimed at persuading companies to move to what was fast becoming one of the most successful manufacturing, financial and legal services centres in the UK while, on the other, it was attempting to justify the case for significant EU funding, based on small areas within the city that were experiencing severe social hardship. The research described here was undertaken on behalf of LCC with the aim of demonstrating that the problems in Leeds were as significant as those of other competing UK cities that already received EU funding, or were themselves making the case for consideration for funding.

The remainder of the chapter is divided into four sections. In Section 7.2, we describe the nature of the problem and the requirements of LCC. We outline the measures used to identify areas with significant socioeconomic problems and explain that while LCC could speak authoritatively about the most deprived areas in Leeds, it was much more difficult to identify similarly deprived parts of other English cities. In particular, identifying small areas that met certain selection criteria, such as level of deprivation, population size and contiguity, all required a relatively complex methodological approach. Section 7.3 therefore explains the potential measures in more detail and the GIS-based methodology that was developed to solve these problems. Section 7.4 presents the results and compares the area chosen to represent the deprived inner city of Leeds with similarly defined areas for the remaining cities in England. Comparisons are made based on a series of different population cut-offs and using different measures of deprivation. Finally, Section 7.5 concludes by arguing

that LCC did indeed have a strong case for Objective 2 status. The Local Authority District of Leeds includes some particularly rich and poor people and, compared with many English cities, ignoring this would be misleading. The geographically refined analysis of socioeconomic conditions in the city that we provided helped to highlight this fact. However, the chapter ends by reminding us that we need to give serious thought to our positionality as researchers when conducting work of this type.

7.2 The Applied Problem and Client Needs

The EC has a series of programmes designed to redistribute resources to those parts of the EU that are most deserving. For example, Objective 1 status areas were defined as those whose per capita gross domestic product (GDP) was less than 75% of the Community average; consequently these were relatively large areas in severe hardship (South Yorkshire was one of the English Objective 1 areas, for example). Objective 3 was available nationally, rather than to geographically targeted areas, and was designed to support education, training and employment policies. Previously, Leeds had been successful in attracting considerable funds under this scheme. Objective 2, which was relevant to this research in Leeds, sought to address the needs of areas experiencing structural difficulties, whether industrial, rural, urban or fisheries-dependent areas, or to promote the restructuring of services (Department of Trade and Industry, 1999). This definition combined Objectives 2 and 5b from previous funding rounds. Leeds did not submit itself for consideration in the previous 1997 round, but was hoping to be considered in the 1999 round.

Before the EC can identify cities, or parts of cities, as deserving of Objective 2 status, a system of measuring the socioeconomic conditions within each is required, such that their relative situations can be compared. Some standard criteria were set out in a lengthy legal document (European Commission, 1999), although these were designed to be indicative rather than a strict list which each application needed to address. Thus, 'urban areas in difficulty' were expected to be densely populated and of a 'reasonable size' (probably over 100 000 population). The application areas were expected to meet at least one of the following criteria:

- a rate of unemployment higher than the Community average;
- a high level of poverty, including precarious housing conditions;
- a particularly degraded environment;
- a high crime rate;
- a low level of education among the population.

Constraints were also imposed by the EC on how the funds would be distributed:

- a ceiling for all types of Objective 2 areas of 18% of the EU population;
- a UK share of that ceiling determined by the population of its industrial and rural areas which qualify as of right and the extent of its structural problems, subject to a maximum of one-third in population coverage at the time;
- a limit of 2% of the EU population to be covered by Urban Areas (the UK hoped to get a larger than average share of this).

It was assumed that applications would be made at NUTS level III (counties or groups of unitary authorities in the UK), but that separate applications should be made by local areas within these broad areas. Thus, Leeds was aiming to be one of a number of areas that was included in a broader submission made by the Government Office for Yorkshire and Humberside.

The criteria identified above that were of particular relevance to this study are those relating to the geographical scale at which areas for support should be identified and the indicators that should be used to identify the level of need. These two criteria are closely related as the choice of geographical units will constrain the data that are available for analysis. Geographically, the two most important criteria were, firstly, that the areas of greatest need should be 'targeted', and secondly, that these areas should be large enough to allow credible and integrated programmes to be mounted. It had been suggested that the most appropriate zones that could be used as building blocks to identify the areas of greatest need across the EU would be NUTS 5 regions or, in UK terms, wards. Whilst the EC did not specify a population size constraint, it was clear that very small areas would not be considered in isolation, and a population of approximately 100 000 was commonly quoted as an appropriate cut-off.

A wide range of indicators, including social, economic, crime and housing measures, were suggested as appropriate for identifying areas of greatest need in the UK, although the manner in which these indicators should be used was not prescribed. Clearly, to be successful, any application for Objective 2 status required that the chosen areas performed badly across a wide range of socioeconomic criteria. Previously within the UK, one of the standard tools that had been used by the Government for identifying need for Objective 2 status was the Index of Local Deprivation (ILD) (Department of Environment, 1995; Department of Environment, Transport and the Regions, 1998). This has been calculated for local authority districts (LADs), of which there were 366 that encompassed the whole of England at the time of the 1991 Census. Through use of the ILD, each district can be ranked according to a standard measure of the material deprivation in that district. The ILD is a useful summary measure that is used in many applied and academic studies to give a feel for the social standing of a particular place. For LCC, the focus of attention was on identifying how Leeds compared with other LADs in England, against whom it was competing for the limited amount of funds.

However, the assumption that LADs are homogeneous areas that can be summarised using a single deprivation measure ignores the diversity of social conditions that exist throughout a city and it is especially biased for cities that are geographically large, encompassing a range of better-off and worse-off areas. The diversity of social deprivation that exists in cities also varies so that an 'average' measure of deprivation will be more biased in some cities than in others. In this round, it was expected that applications would use ward-level deprivation scores as one way of summarising the socioeconomic situation of the local area, partly because this was a useful way of being able to compare submissions from different parts of the country.

Therefore, applicants were encouraged to identify a group of contiguous wards (which aggregate neatly into districts) that was relatively large and that was also an

area of great need. LCC believed that the scale of analysis was particularly crucial in Leeds as LAD-level measures masked the reality of inner area deprivation in the city. Two points are relevant here. Firstly, because Leeds LAD is such a large geographical area, it inevitably encompasses a range of places with different levels of deprivation. Other cities like Manchester or Liverpool are defined more tightly with their LAD boundaries encompassing the inner city, but few of the surrounding, often more wealthy, suburban and extra-urban areas. On the other hand, cities like London include a number of LADs within their boundary, each of which is relatively more homogeneous than the LADs that include entire cities. Secondly, it has been argued for some time that the city of Leeds is unusual because it encompasses a heterogeneous mixture of residential environments compared with other UK cities. Put simply, there is considerable variation in the levels of deprivation within the inner city area such that Leeds has been described as 'two cities', one 'working hard to position itself as the capital of the north', and the other 'where unemployment is high, educational attainment is low and where dependency on benefits is almost universal' (Leeds City Council, 1998).

To address this problem, LCC required an analysis of deprivation conducted at a more detailed spatial scale that would reflect these extremes – an analysis of deprivation *within* the city, rather than a summary deprivation score for the city as a whole. The aim, therefore, was to identify a contiguous set of wards within the Leeds LAD which, when aggregated together, represented a relatively large homogeneous population with a high deprivation score. The EC required that areas with relatively large populations be submitted for consideration, as the aim of the Objective 2 status programme is to target relatively large areas of concentrated deprivation that require aid. However, it was not stipulated that this should necessarily include the entire city. By selecting appropriate wards from within the district boundary, it removes the problem that using the entire LAD includes a mixture of richer and poorer areas that, on average, suggest that the city is not particularly deprived.

LCC was therefore keen that an analysis be conducted that compared such a choice of wards in Leeds with similarly defined groups of highly deprived wards in other English cities. As most city councils purchase census and other data that can be used to construct deprivation indices for the wards within their own cities, LCC was able to carry out such an analysis for Leeds. In fact, they also had deprivation scores for the wards across all of England apparently allowing them to compare Leeds with other cities. However, as we demonstrate below, relatively complex methods were required to make sure that the selection of wards in other cities was as equitable as possible.

There were two problems with this approach which were the reasons behind LCC's request for some external consultancy work. First, LCC did not have the digital boundaries required to identify which wards in other 'competing' cities were *neighbouring*. This was important since a requirement of the EC exercise was that the areas submitted for consideration should be compact. Before LCC could argue that their preferred selection of wards was a better reflection of concentrated deprivation within the city, it was necessary to compare the results of their analysis with a similar analysis in other English cities. LCC were convinced that there was a large area of inner city Leeds that was significantly more deprived than most equivalent areas in other English cities, but an analysis was required actually to prove this. The second

problem was that, even if contiguous wards could be identified, a strategy for optimising the choice of which wards should be aggregated together was potentially onerous. Given that there are 10 000 wards in England and Wales, deciding how to group wards in every city so that the largest and most deprived combination was achieved was not simple. In Leeds, the planners could do this by hand, knowing well the worst areas of their own city, but they did not have such insights into other English cities. The method we adopted for solving this problem is described below.

7.3 Methods of Identification of Deprived Urban Areas in British cities

The first strategy was to investigate how Leeds LAD compared with other English LADs in terms of the 1998 ILD. This index is based on 12 variables, three of which are derived from the 1991 Census, with the remainder coming from other updateable sources, such as unemployment, mortality and insurance company data. Consequently, this index was relatively up to date and, although we may debate the validity of the index itself, it does at least provide a consistent means of comparing all LADs in England. The ILD can be used as the basis of a number of different comparative measures, including:

- the degree of deprivation, or the overall district level score;
- the intensity of deprivation, or the severity of deprivation in the LAD taken as the average score of the worst three wards in the LAD;
- the ward level extent, or the proportion of the LAD population living in wards that are within the 10% most deprived in England; and
- the ED level extent, or the proportion of the EDs in the LA that fall within the most deprived 7% of EDs in England.

Here we focused mainly on the first of these measures. Of course, as argued above, this method of identifying deprived inner city areas is spatially crude. There is no logical reason why the LAD scale is the most appropriate for analysis and, given the considerable variation in the size and homogeneity of LADs, there is a strong argument that a more considered definition of deprived inner city areas should be sought. In particular, the aim here is to identify *concentrated areas of deprivation*, as these are likely to be especially deprived places.

This objective required analyses that build deprived areas from smaller geographical units. LCC therefore required a series of analyses using aggregations of small areas (wards) for Leeds and other competing LADs in England, which would show that Leeds has a significant area within the city with a higher level of material deprivation compared with other places. These groups of wards would include large populations in most LADs, but in most cases their populations would obviously be smaller than the populations of the entire LADs. Fortunately, a ward-level version of the ILD has been constructed which is based on six variables from the 1991 Census. These six variables were: unemployment, children in low-earning households, households with no car, households lacking basic amenities, overcrowded households, and 17 year olds no longer in full-time education. Although this version of the ILD

is not regularly updateable, as it relies on census information which is only collected every 10 years, it does provide an indication of deprivation for small geographical areas.

In consultation with LCC, inner city Leeds was defined based on informed views of the most deprived parts of the city. Of the 12 chosen wards, the least deprived was Armley, with an ILD score of 7.97, which ranked 1062 out of 8620 wards in England. There were 14 wards in Leeds that were at least as deprived as this ward, using the ward-level ILD, although two (Beeston with ILD = 9.60 and Bramley with ILD = 9.50) were excluded from the definition of inner city Leeds used below (Figure 7.1).

Using the deprivation score of 7.97 for Armley as a cut-off, all wards across England with a similarly high level of deprivation could be identified and relevant groups were then aggregated for each city. As can be seen in Figure 7.2, many districts had few wards that reached this level of deprivation. However, it was not a simple matter of aggregating all these wards for each city, as a strategy was required for deciding which to include. Simply using a deprivation score as a cut-off means that isolated wards within LADs would have been included in the list and their

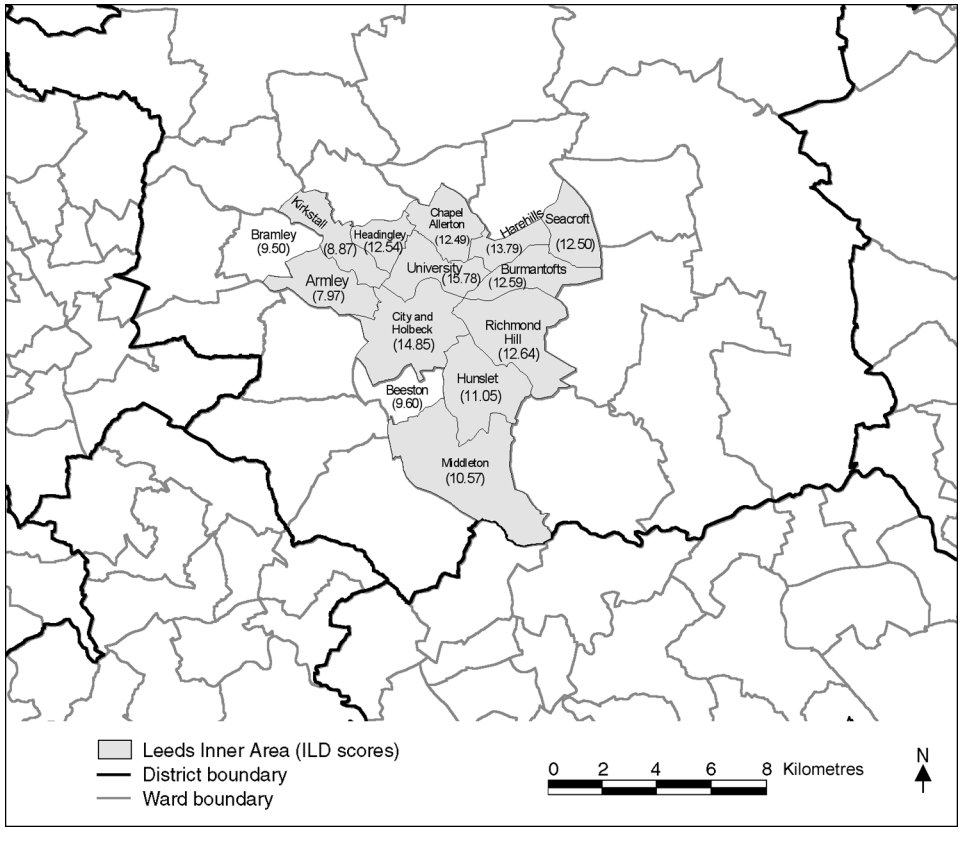


Figure 7.1 *Deprived wards in inner city Leeds © Crown Copyright. Reproduced with permission*

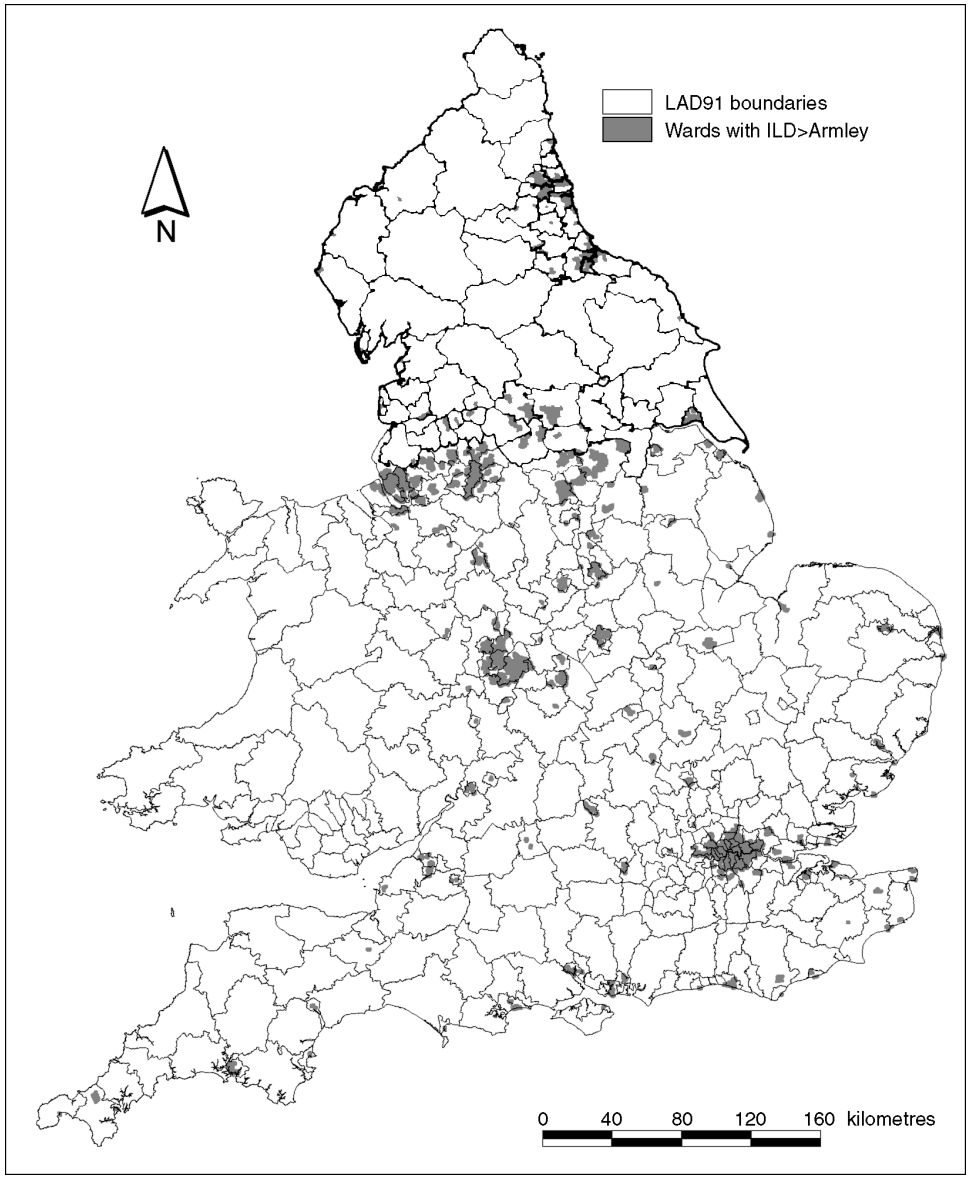


Figure 7.2 *Wards in England that were at least as deprived as Armley ward © Crown Copyright. Reproduced with permission*

inclusion was not acceptable according to the EC guidelines; the group of wards that were chosen needed to be contiguous.

To illustrate this, the districts in West Yorkshire with ILD scores above Armley are mapped in Figure 7.3. Leeds has 14 contiguous wards in total with ILD scores at least as deprived as Armley, including Beeston and Bramley. These 14 wards are clustered to form the contiguous inner area, without any holes or deprived wards outside

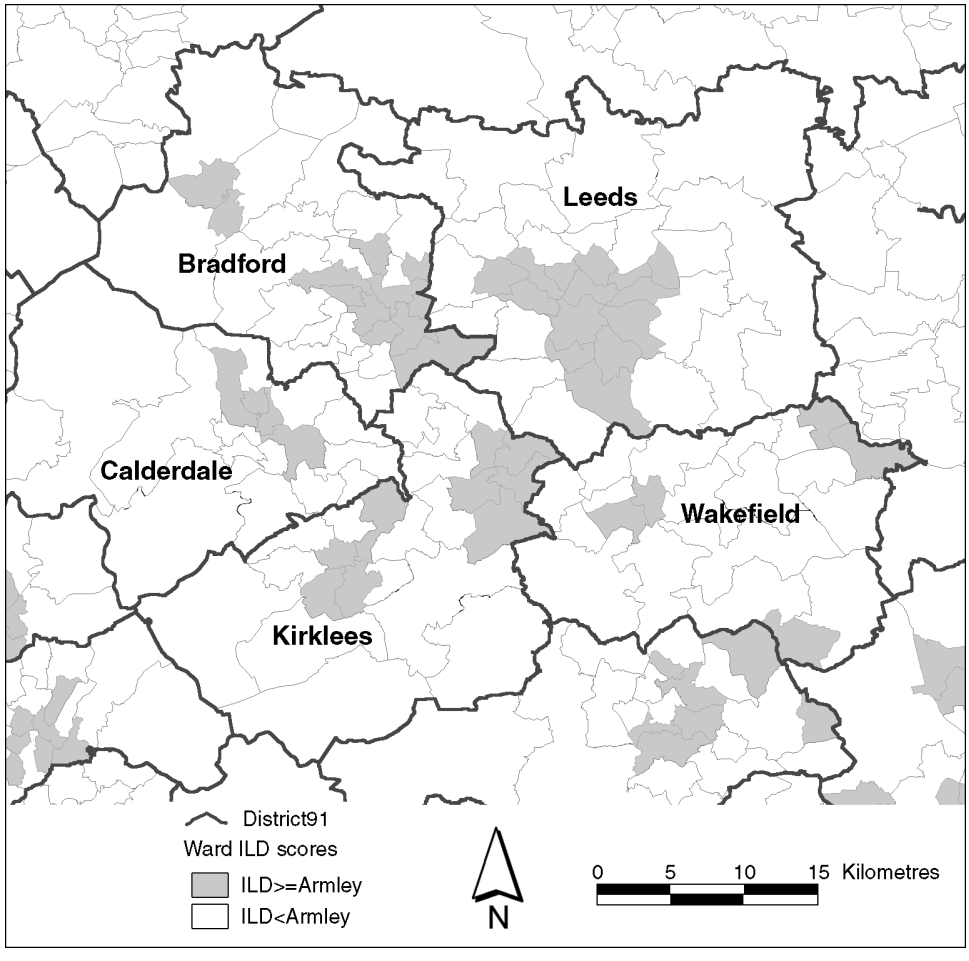


Figure 7.3 *Deprived wards in selected districts in West Yorkshire © Crown Copyright. Reproduced with permission*

this area. This is also the case in Calderdale, where four contiguous wards with ILD scores at least as deprived as Armley cluster to form a deprived area. However, this is not the case in either Bradford or Kirklees, where the deprived wards cluster in two areas, nor in Wakefield where there are three pockets of deprivation. In order to check the contiguity of the selected wards shown in Figure 7.2, our research therefore exploited GIS and some in-house software designed to optimise the selection of contiguous groups of geographical wards based on some pre-determined criteria (maximising deprivation in this case).

The analysis was carried out using semi-automated methods in ArcInfo GIS, utilising the Arc Macro Language (AML) capabilities for topology building, spatial queries and calculations. First, all the wards with ILD scores greater than Armley were selected, as shown in Figure 7.2. Subsequently, the adjacent boundaries of the selected wards were dissolved in order to identify the number of deprived areas within

each LAD. In cases like Leeds and Calderdale, the selected wards would only form a single deprived area, with no internal islands and the selection was acceptable. However, in cases like Bradford, Kirklees and Wakefield, the selected wards would form two or more separate areas, within the same district. These districts were automatically identified within ArcInfo GIS and a number of measures were calculated for each deprived area, such as the total population and the degree of deprivation. If the population within one of the sub-district areas was adequate (as described below), then this area was taken as a representative deprived area for the district, otherwise, further GIS analysis was necessary to secure an acceptable population size. Further analysis was undertaken in the form of aggregations, so that wards with relatively low scores were added to the contiguous set if the ward neighboured another ward with a particularly high deprivation score. These 'lower-than-Armley' wards were known as 'bridging wards'. Thus, the addition of both wards to a cluster increased the average deprivation in that cluster, even though the score for the bridging ward may have been lower than the arbitrary cut-off that had been chosen. This procedure therefore identified 187 large aggregations of contiguous wards, in cities across England.

In the analysis, we also introduced a series of population cut-offs enabling us to identify clusters of wards that were above or below certain thresholds. The deprivation cut-off was agreed with LCC as those wards that were at least as deprived as Armley ward (see above). The population cut-offs were also a subjective choice, but one that was guided by the knowledge that the EC was only interested in funding relatively large areas; deprived pockets were not to be targeted. The software allowed us to identify clusters of wards that were above or below certain populations. While there is a strong case that deprived areas with large populations should be targeted for aid, there are also concentrated areas of deprivation that should not be ignored simply because they have a smaller population than other inner cities. The LCC definition of inner city Leeds had a population of approximately 228 000 but, as suggested above, while 100 000 was an oft-quoted figure for an appropriate size of a deprived place, no guidelines specified that smaller areas would not be considered. Four levels of analysis were therefore conducted to reflect the size of the deprivation problem in the various LADs. Initially, we retained only those sets of contiguous wards that had a combined population of at least 200 000. This is useful as it draws attention to the largest areas of concentrated deprivation in England – places that should certainly be considered as worthy of consideration in attempts to redistribute resources. The second and third definitions used clusters of wards with populations of at least 100 000 and 60 000, respectively. Finally, we also considered those clusters with a maximum population of 60 000 as a cut-off, allowing us to identify the most concentrated areas of deprivation in the various inner cities.

The GIS software facilitated the analysis in many ways. Firstly, it was possible to manage the spatial datasets and carry out calculations in a seamless manner. In addition, the AMLs allowed automation of the process, by completing repetitive spatial queries, building the topology, and producing descriptive measures for population and deprivation scores. Finally, the GIS identified the most appropriate bridging wards, restricting the search of appropriate wards considerably. The diagram in Figure 7.4 illustrates the semi-automated process followed in identifying the optimum district ILDs for selected sets of wards.

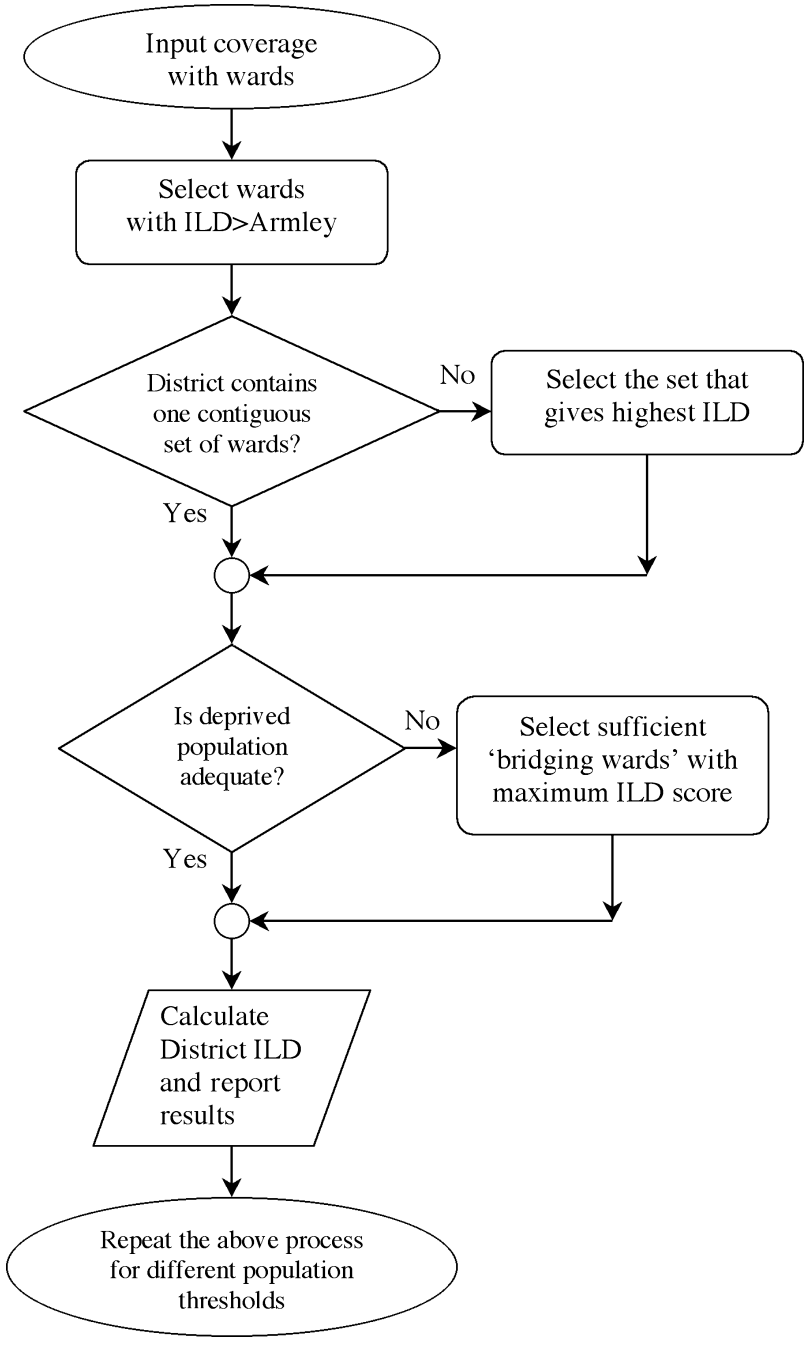


Figure 7.4 *Diagram showing the semi-automated GIS process*

7.4 A Comparison of Concentrations of Deprivation

The results for the *LAD-level* analysis are provided in Table 7.1. The mean deprivation is perhaps the most obvious indicator of overall deprivation. Leeds ranked very low (56th) on this list, demonstrating that it was not among the most deprived LADs in England, using this spatially crude analysis. Note, however, that Leeds ranked seventh in England using the average deprivation in the worst three wards, suggesting that a more sensitive definition of concentrated deprivation may be appropriate for the city.

The next step was to show how Leeds ranked using a simple comparison of *groups of wards* that were included because they had a higher deprivation score than our chosen cut-off value. Table 7.2 provides the cities with the most deprived wards based on this analysis and this is equivalent to the analysis that LCC could have conducted using the information they had to hand. Note, however, that while the wards in Leeds have been chosen because they are contiguous, the wards in the other cities are chosen simply because their deprivation score is higher than a particular cut-off; many of these wards will not have been neighbouring. According to this analysis, Leeds climbed from 56th (Table 7.1) to 11th (Table 7.2). Of particular note is the very large population of the Leeds inner city area compared with other cities. In terms of population size, the group of wards in Leeds was exceeded only by Birmingham, Liverpool, Manchester and Sheffield. Lambeth was the highest-ranking London borough, in sixth position.

The problem with this analysis is that it ignores contiguity for all the cities except Leeds. The aim was to identify a concentrated area of deprivation as the EC had made it known that a smaller number of substantial areas of deprivation were liable to receive funding, rather than a larger number of smaller areas, thus avoiding a 'pepperpotting' of funds. Once a contiguity constraint is introduced it is likely that the deprivation measures will alter for the other cities and, if nothing else, the population of many of the inner city areas will be reduced. For Leeds to make a strong case, it needed to be able to compare its position with that of the other cities in England.

Table 7.3 lists those nine deprived inner city areas with populations of at least 200 000, ranked by the mean deprivation score. In summary, the Leeds inner city area ranked:

- fifth using the ward-level ILD degree score;
- fourth using the ILD ward intensity score;
- third on the maximum index value indicator; and
- fourth on the estimated population size indicator.

Note that the standard deviation of the ILD values is also provided to give an indication of the variation in the deprivation scores for each cluster of wards. Leeds has a fairly high value compared with other clusters (this result is consistent in the following tables).

The second stage of this analysis extended the list of deprived inner city areas to include those that had a population of at least 100 000. These are still large concentrations of deprivation, although many have much smaller populations than the

Table 7.1 *Local authority district rankings on degree, intensity, ward and ED extent measures*

District name	Degree score	Degree rank	Average worst 3 wards	Ward intensity (rank)	% LAD population in worst 10% wards	% LA population in worst 10% wards	% EDs in worst 7%	ED extent (rank)
Liverpool	40.07	1	15.9	2	71.22	11	34.22	6
Newham	38.55	2	14.5	10	100.00	2	55.75	3
Manchester	36.33	3	15.2	4	77.98	8	31.10	8
Hackney	35.21	4	14.7	8	100.01	1	60.79	1
Birmingham	34.67	5	16.6	1	64.61	15	24.88	11
Tower Hamlets	34.30	6	15.1	5	99.99	4	58.26	2
Sandwell	33.78	7	13.3	29	50.02	19	10.95	40
Southwark	33.74	8	14.2	14	88.96	6	39.11	4
Knowsley	33.69	9	14.1	15	72.50	10	38.18	5
Islington	32.21	10	13.8	22	100.00	3	29.93	10
Greenwich	31.58	11	12.2	45	58.27	16	15.24	26
Lambeth	31.57	12	14.7	9	89.42	5	34.09	7
Haringey	31.53	13	14.2	12	79.26	7	30.65	9
Lewisham	29.44	14	13.9	19	53.58	17	18.86	20
Barking and Dagenham	28.69	15	11.0	65	46.40	22	4.51	89
Nottingham	28.44	16	13.8	21	41.30	29	18.94	19
Camden	28.23	17	13.1	32	77.72	9	16.70	23
Hammersmith and Fulham	28.19	18	14.1	17	65.93	14	19.22	16
Newcastle Upon Tyne	27.95	19	12.9	34	39.08	32	19.01	18
Brent	26.95	20	14.1	16	46.86	21	20.76	14
Sunderland	26.90	21	12.2	44	36.65	34	13.06	34
Waltham Forest	26.68	22	14.2	13	66.77	12	16.34	25
Salford	26.64	23	13.3	30	31.43	44	10.72	43
Middlesbrough	26.41	24	12.8	37	53.55	18	22.26	12
Sheffield	26.09	25	14.8	6	36.33	36	13.38	32
Kingston Upon Hull	26.06	26	14.5	11	44.91	24	18.18	22
Wolverhampton	25.94	27	13.5	26	40.87	31	14.66	28
Rochdale	25.13	29	12.3	43	38.63	33	14.29	30
Bradford	25.94	28	15.6	3	42.15	27	21.47	13

Table 7.1 *Continued*

District name	Degree score	Degree rank	Average worst 3 wards	Ward intensity (rank)	% LAD population in worst 10% wards	% LA population in worst 10% wards	% EDs in worst 7%	ED extent (rank)
Wandsworth	25.05	30	11.9	48	36.45	35	8.23	57
Walsall	25.02	31	12.7	39	45.82	23	8.49	55
Leicester	24.95	32	14.0	18	48.46	20	12.95	35
Oldham	24.82	33	13.5	25	34.86	37	14.09	31
Halton	24.69	34	9.7	94	16.52	72	10.26	46
Gateshead	24.58	35	11.7	54	30.30	46	9.47	51
Ealing	24.48	36	11.8	52	29.59	48	7.14	66
Hartlepool	23.72	37	12.0	47	44.85	25	20.57	15
South Tyneside	23.67	38	10.4	76	28.94	50	9.69	50
Doncaster	23.60	39	12.5	40	41.15	30	10.85	41
Coventry	23.48	40	13.4	27	34.77	38	15.00	27
Blackburn with Darwen	23.04	41	12.9	36	32.20	43	19.18	17
Barnsley	22.30	42	10.5	73	23.07	56	6.62	70
Redcar and Cleveland	21.54	43	11.9	49	13.47	81	10.63	44
Wirral	21.25	44	12.8	38	21.40	60	10.81	42
St.Helens	20.98	45	10.9	68	20.91	62	8.70	54
Lincoln	20.70	46	9.3	105	15.57	76	9.88	48
Bolton	20.66	47	13.6	24	29.71	47	11.73	37
Stoke-on-Trent	20.61	48	11.2	60	23.17	55	5.28	81
Stockton-on-Tees	20.41	49	11.9	50	24.90	52	11.51	38
Rotherham	20.23	50	11.3	58	16.71	71	7.05	67
Blackpool	20.14	51	13.7	23	21.99	59	11.99	36
Easington	19.97	52	10.1	82	15.91	75	4.02	99
Tameside	19.78	53	9.8	90	10.11	100	4.08	97
Sefton	19.41	54	11.0	64	32.52	42	9.90	47
Barrow-in-Furness	19.38	55	8.6	117	7.30	122	5.44	80
Leeds	**19.06**	**56**	**14.8**	**7**	**33.30**	**40**	**9.72**	**49**
City of Westminster	19.05	57	13.8	20	66.29	13	18.31	21

Table 7.2 *Inner city area rankings on mean ILD score*

District name	Number of wards	Popn	Popn rank	Index mean	Index mean rank	Average worst 3 wards	Worst 3 wards rank	Index max	Index max rank	Obj 2 status (1 = yes)
Hackney	23	181248	10	13.33	1	14.73	8	14.91	15	1
Newham	24	212170	7	13.32	2	14.50	10	14.58	20	1
Tower Hamlets	19	161064	13	12.91	3	15.06	5	15.66	6	1
Liverpool	26	356476	2	12.71	4	15.94	2	16.65	3	0
Lambeth	20	222576	6	12.64	5	14.70	9	15.00	13	0
Haringey	18	158247	14	12.63	6	14.25	12	14.60	19	1
Birmingham	27	665336	1	12.54	7	16.60	1	17.12	1	1
Bradford	12	182938	9	12.25	8	15.64	3	16.79	2	1
Trafford	2	20260	92	12.23	9	10.60	69	12.83	50	1
Waltham Forest	13	137821	26	12.21	10	14.22	13	14.47	23	1
Leeds	**12**	**227747**	**5**	**12.14**	**11**	**14.81**	**7**	**15.77**	**5**	**0**
Oldham	7	75786	49	12.12	12	13.54	25	14.17	30	1
Sheffield	14	241960	4	12.09	13	14.81	6	15.55	7	1
Islington	20	164686	12	12.05	14	13.80	22	14.29	28	0
Southwark	23	201058	8	11.97	15	14.16	14	14.88	16	0
Hove	2	17073	96	11.94	16	n/a	n/a	12.08	67	0
Wirral	5	75181	50	11.90	17	12.76	38	14.00	34	0
Kingston upon Hull	11	139764	23	11.77	18	14.47	11	15.30	9	1
Blackpool	6	39837	68	11.63	19	13.69	23	14.33	26	0
Brent	18	141111	21	11.60	20	14.13	16	14.75	17	0
Manchester	29	355787	3	11.59	21	15.21	4	15.53	8	1
Salford	8	88185	39	11.55	22	13.27	30	14.30	27	1
Blackburn	8	52043	61	11.52	23	12.88	36	13.80	36	1
Bolton	7	90504	37	11.51	24	13.62	24	16.12	4	1
Knowsley	17	117525	30	11.46	25	14.13	15	15.13	11	0
Wolverhampton	10	121095	29	11.32	26	13.47	26	14.23	29	1

Table 7.2 *Continued*

District name	Number of wards	Popn	Popn rank	Index mean	Index mean rank	Average worst 3 wards	Worst 3 wards rank	Index max	Index max rank	Obj 2 status (1 = yes)
Westminster, City of	15	114009	31	11.32	27	13.84	20	14.54	22	0
Preston	8	53087	59	11.31	28	12.99	33	13.22	44	0
Pendle	2	8959	121	11.26	29	9.71	90	11.51	82	1
Derby	7	76581	47	11.25	30	12.90	35	13.78	37	0
Kensington and Chelsea	10	65902	54	11.24	31	13.41	28	14.04	33	0
Coventry	9	147194	18	11.22	32	13.42	27	15.12	12	1
Sandwell	14	169220	11	11.13	33	13.30	29	14.12	32	1
Tameside	2	22782	89	11.11	34	9.77	85	12.00	69	1
Solihull	4	47026	63	11.10	35	11.45	55	11.99	70	1
Calderdale	4	42574	66	11.09	36	11.56	53	12.63	54	0
Hammersmith and Fulham	16	103306	33	11.09	37	14.12	17	14.39	25	0
Hartlepool	7	37227	72	11.04	38	11.96	46	12.33	61	1
Canterbury	1	4958	150	11.02	39	8.51	112	11.02	90	0
Camden	20	131111	27	11.00	40	13.05	32	13.78	38	0
Rochdale	8	80866	43	10.99	41	12.26	42	13.38	42	1
Portsmouth	5	67191	53	10.97	42	11.80	49	12.56	57	0
Plymouth	7	85181	40	10.97	43	13.19	31	15.29	10	1
Nottingham	16	156161	15	10.97	44	13.81	21	14.69	18	1
Rotherham	4	45752	65	10.96	45	11.32	56	12.96	47	1
Lewisham	17	151027	17	10.92	46	13.85	19	14.57	21	0
Leicester	15	144907	19	10.89	47	13.99	18	14.94	14	0
Milton Keynes	2	16030	100	10.84	48	8.13	127	11.41	83	0
Shepway	2	7319	134	10.83	49	9.78	84	12.33	60	0
Newcastle upon Tyne	14	139753	24	10.75	50	12.92	34	13.20	45	1
Sunderland	12	138739	25	10.74	51	12.20	43	12.27	63	1

Table 7.3 *Inner city areas with at least 200 000 people: rankings on mean ILD score*

District name	Number of wards	Popn	Popn rank	Area	Index mean	Index mean rank	Standard deviation of index	Max index score	Average worst 3 wards
Newham	24	212 170	7	36 360 712	13.32	1	1.13	14.58	14.50
Liverpool	22	296 891	3	62 447 304	12.81	2	1.81	16.65	15.94
Birmingham	24	602 240	1	137 786 944	12.74	3	2.94	17.12	16.45
Lambeth	20	226 824	5	24 457 256	12.64	4	1.55	15.00	14.70
Leeds	**12**	**227 747**	**4**	**75 948 004**	**12.14**	**5**	**2.25**	**15.78**	**14.81**
Sheffield	14	216 687	6	68 008 216	12.09	6	2.18	15.55	14.81
Southwark	23	201 191	8	23 311 460	11.97	7	1.39	14.88	14.17
Manchester	29	353 203	2	102 427 784	11.59	8	2.07	15.53	15.21

Leeds inner city area. Table 7.4 provides details of all 27 of these deprived inner city areas. In summary, the Leeds inner city area ranked:

- tenth using the ward-level ILD degree score;
- sixth using an ILD ward intensity score;
- fourth on the maximum index value indicator; and
- fourth on the estimated population size indicator.

The list was then extended to include contiguous sets of wards with a population of at least 60 000 (Table 7.5). Even on this much longer list, Leeds ranked high up the list:

- eleventh using the ward-level ILD degree score;
- sixth using an ILD ward intensity score;
- fifth on the maximum index value indicator; and
- fourth on the estimated population size indicator.

These analyses provide a clear indication that the case for Leeds is strengthened by using definitions of deprived inner city areas based on contiguous sets of wards, rather than LADs. We argue that this geographically sensitive analysis is an improvement on simple studies based on the relatively arbitrary definitions of LADs. However, there is a strong argument that the most concentrated areas of deprivation are much smaller than the areas identified in these analyses. While it is important to identify places with relatively large populations, the decision about where such a population cut-off should be is crucial.

The aim in the final analysis was therefore to identify deprived areas with a maximum population of 60 000 – clearly a considerable number of people, but far short of the populations in many of the 'inner city areas' identified above. This strategy identified the most *concentrated areas of deprivation* in the inner cities as the mean deprivation values in Table 7.6 demonstrate. These values are inevitably much higher than those provided in the previous table. This list indicates that the Leeds inner city area ranked:

- third using the ward-level ILD degree score;
- fourth using an ILD ward intensity score; and
- fifth on the maximum index value indicator.

Indeed, both Birmingham and Bradford, which ranked higher than Leeds, already had Objective 2 status at the time of the study.

The populations in this list of contiguous inner city areas were large and, we suggested, should be given serious consideration as zones that may have been deserving of Objective 2 status. While they were much smaller than the populations of entire LADs, these areas were considerably more deprived. The results for this concentrated area of deprivation certainly supported the wider Leeds case for eligibility and it is noteworthy that Leeds also had a high ranking based on the average of the worst three wards indicator.

Table 7.4 *Inner city areas with at least 100 000 people: rankings on mean ILD score*

District name	Number of wards	Popn	Popn rank	Area	Index mean	Index rank	Standard deviation of index	Max index score	Average worst 3 wards
Hackney	23	181 248	9	19 501 144	13.33	1	1.04	14.91	14.73
Newham	24	212 170	7	36 360 712	13.32	2	1.13	14.58	14.50
Tower Hamlets	19	161 064	13	19 733 140	12.91	3	1.46	15.66	15.06
Liverpool	22	296 891	3	62 447 304	12.81	4	1.81	16.65	15.94
Birmingham	24	602 240	1	137 786 944	12.74	5	2.94	17.12	16.45
Bradford	10	163 125	12	50 480 368	12.71	6	2.68	16.79	15.65
Lambeth	20	226 824	5	24 457 256	12.64	7	1.55	15	14.70
Haringey	18	160 269	14	21 242 866	12.63	8	1.30	14.6	14.25
Waltham Forest	13	141 537	15	21 951 868	12.21	9	1.41	14.47	14.22
Leeds	**12**	**227 747**	**4**	**75 948 004**	**12.14**	**10**	**2.25**	**15.78**	**14.81**
Sheffield	14	216 687	6	68 008 216	12.09	11	2.18	15.55	14.81
Islington	20	164 686	11	14 881 020	12.05	12	1.38	14.29	13.81
Southwark	23	201 191	8	23 311 460	11.97	13	1.39	14.88	14.17
Westminster	12	102 430	25	9 238 828	11.83	14	1.94	14.54	13.84
Brent	17	122 595	21	18 418 158	11.70	15	1.87	14.75	14.13
Knowsley	15	104 047	24	63 369 060	11.60	16	1.85	15.13	14.13
Manchester	29	353 203	2	102 427 784	11.59	17	2.07	15.53	15.21
Coventry	8	135 572	17	30 821 952	11.50	18	2.10	15.12	13.43
Wolverhampton	10	122 078	22	32 541 922	11.33	19	2.05	14.23	13.47
Sandwell	14	169 112	10	48 745 228	11.13	20	1.55	14.12	13.30
Hammersmith and Fulham	16	109 860	23	12 270 624	11.09	21	2.07	14.39	14.12
Camden	20	136 438	16	16 145 836	11.00	22	1.27	13.78	13.05
Leicester	13	128 510	20	34 664 192	10.76	23	2.00	14.94	13.69
Newcastle upon Tyne	14	130 489	19	47 679 616	10.75	24	1.89	13.2	12.92
Walsall	10	130 839	18	38 011 604	10.69	25	1.61	13.57	12.73
Doncaster	8	100 040	26	66 884 280	10.60	26	1.90	14.14	12.46

Table 7.5 *Inner city areas with at least 60000 people: rankings on mean ILD score*

District name	Number of wards	Popn	Popn rank	Area	Index mean	Index rank	Standard deviation of index	Max index score	Average worst 3 wards
Hackney	23	181248	9	19501144	13.33	1	1.04	14.91	14.73
Newham	24	212170	7	36360712	13.32	2	1.13	14.58	14.50
Tower Hamlets	19	161064	13	19733140	12.91	3	1.46	15.66	15.06
Liverpool	22	296891	3	62447304	12.81	4	1.81	16.65	15.94
Birmingham (a)	24	602240	1	137786944	12.74	5	2.94	17.12	16.45
Bradford	10	163125	12	50480368	12.71	6	2.68	16.79	15.65
Lambeth	20	226824	5	24457256	12.64	7	1.55	15	14.70
Haringey	18	160269	14	21242866	12.63	8	1.30	14.6	14.25
Waltham Forest	13	141537	15	21951868	12.21	9	1.41	14.47	14.22
Nottingham	8	72874	38	17205614	12.19	10	1.78	14.69	13.81
Leeds	**12**	**227747**	**4**	**75948004**	**12.14**	**11**	**2.25**	**15.78**	**14.81**
Lewisham	9	85788	32	11980866	12.13	12	1.82	14.57	13.85
Oldham	7	75493	36	19797938	12.12	13	1.61	14.17	13.54
Sheffield	14	216687	6	68008216	12.09	14	2.18	15.55	14.81
Islington	20	164686	11	14881020	12.05	15	1.38	14.29	13.81
Southwark	23	201191	8	23311460	11.97	16	1.39	14.88	14.17
Wirral	5	70786	41	20803394	11.90	17	1.41	14	12.76
Westminster	12	102430	25	9238828	11.83	18	1.94	14.54	13.84
Brent	17	122595	21	18418158	11.70	19	1.87	14.75	14.13
Knowsley	15	104047	24	63369060	11.60	20	1.85	15.13	14.13
Manchester	29	353203	2	102427784	11.59	21	2.07	15.53	15.21

Bolton	7	87730	31	21653956	11.51	22	2.59	16.12	13.63
Coventry	8	135572	17	30821952	11.50	23	2.10	15.12	13.43
Wolverhampton	10	122078	22	32541922	11.33	24	2.05	14.23	13.47
Plymouth	6	68382	43	14981245	11.31	25	2.53	15.29	13.19
Sunderland	8	84662	33	30128508	11.26	26	1.08	12.27	12.16
Derby	7	72873	39	23252992	11.25	27	1.85	13.78	12.90
Kingston Upon Hull	8	94467	28	29991772	11.24	28	2.63	15.3	14.00
Sandwell	14	169112	10	48745228	11.13	29	1.55	14.12	13.30
Hammersmith and Fulham	16	109860	23	12270624	11.09	30	2.07	14.39	14.12
Camden	20	136438	16	16145836	11.00	31	1.27	13.78	13.05
Birmingham (b)	3	67216	44	18791090	10.99	32	0.34	11.39	10.99
Middlesbrough	14	75419	37	21296718	10.97	33	1.29	13.64	12.76
Southampton	5	61648	47	17079916	10.84	34	2.35	14.43	12.06
Leicester	13	128510	20	34664192	10.76	35	2.00	14.94	13.69
Newcastle upon Tyne	14	130489	19	47679616	10.75	36	1.89	13.2	12.92
Walsall	10	130839	18	38011604	10.69	37	1.61	13.57	12.73
Doncaster	8	100040	26	66884280	10.60	38	1.90	14.14	12.46
Brighton	8	71072	40	15052678	10.43	39	1.51	12.82	11.97
Gateshead	8	68843	42	43460340	10.38	40	1.32	11.95	11.68
Sefton	8	94154	29	22979070	10.36	41	0.72	11.93	10.82
Greenwich	16	99902	27	21164610	10.32	42	1.05	12.68	12.18
Wandsworth	6	64785	46	7615645	10.31	43	1.44	12.15	11.48
Kirklees	5	83973	34	40540520	9.67	44	0.84	10.63	10.26
Lewisham	8	67170	45	11015418	9.56	45	0.99	11.04	10.58
South Tyneside	10	75784	35	20240788	9.56	46	0.90	10.57	10.45
Barking and Dagenham	11	88395	30	20086828	9.52	47	1.15	12.05	10.95

Table 7.6 *Inner city areas with a maximum of 60 000 people: rankings on mean ILD score*

District name	Number of wards	Popn	Area	Index mean	Index rank	Standard deviation of index score	Max index score	Average worst 3 wards
Birmingham	2	51 057	14 300 438	16.55	1	0.81	17.12	–
Bradford	3	53 460	13 670 915	15.56	2	1.78	16.79	15.56
Leeds	**3**	**59 399**	**17 355 309**	**14.81**	**3**	**1.00**	**15.78**	**14.81**
Lambeth	4	50 468	4 133 218	14.62	3	0.27	15.00	14.70
Hackney	7	57 588	5 272 163	14.46	4	0.35	14.91	14.73
Tower Hamlets	6	58 134	5 479 824	14.38	5	0.83	15.66	15.03
Newham	7	59 900	8 348 577	14.35	6	0.15	14.58	14.50
Manchester	5	58 517	13 172 278	14.22	7	0.89	15.53	14.72
Sheffield	3	48 659	11 951 534	14.08	8	1.49	15.55	14.08
Liverpool	4	58 774	11 846 305	13.94	9	2.13	16.04	14.84
Haringey	6	57 575	6 879 391	13.71	10	0.62	14.60	14.25
Southwark	6	55 463	4 655 179	13.54	11	0.82	14.88	14.17
Islington	6	58 207	5 044 179	13.49	12	0.44	14.29	13.81
Waltham Forest	5	57 683	8 383 511	13.46	13	1.25	14.47	14.22
Lewisham	5	50 569	6 529 237	13.45	14	0.73	14.57	13.85
Coventry	3	48 796	11 657 353	13.43	15	1.61	15.12	13.43
Westminster	6	59 857	3 930 741	13.34	16	0.85	14.54	13.84
Bolton	4	49 364	11 665 953	13.15	17	2.01	16.12	13.63
Wolverhampton	4	49 358	14 362 921	13.11	18	0.94	14.23	13.47
Knowsley	8	55 099	45 252 499	13.10	19	1.00	15.13	14.13

Hammersmith and Fulham	7	54818	6661414	12.95	20	1.49	14.39	14.12
Brent	8	58936	8991411	12.93	21	1.14	14.75	14.13
Nottingham	6	53166	11644620	12.88	22	1.16	14.69	13.81
Oldham	5	55414	12821654	12.80	23	1.26	14.17	13.54
Sandwell	5	57821	13426594	12.45	24	1.35	14.12	13.30
Salford	5	45142	12107229	12.41	25	1.39	14.30	13.15
Wirral	4	54909	17531268	12.24	26	1.37	14.00	12.76
Derby	5	51926	16319104	12.12	27	1.29	13.78	12.90
Walsall	4	49586	14565825	12.08	28	1.46	13.57	12.73
Sunderland	5	51216	14480702	12.01	29	0.29	12.27	12.21
Blackpool	5	32315	4822156	12.00	30	2.44	14.33	13.69
Blackburn	7	43996	14785446	11.94	31	1.09	13.80	12.88
Plymouth	5	58340	13461113	11.83	32	2.44	15.29	13.19
Kensington and Chelsea	6	42163	3230210	11.79	33	1.69	14.04	12.93
Doncaster	4	49482	38684399	11.73	34	1.93	14.14	12.46
Newcastle upon Tyne	6	52710	14727652	11.57	35	1.74	13.20	12.78
Camden	9	59032	7156480	11.53	36	0.84	13.08	12.53
Kingston Upon Hull	4	48313	13078984	11.52	37	2.51	13.76	12.67
Leicester	6	59304	11352824	11.51	38	1.56	14.07	12.47
Rochdale	5	52913	16486902	11.37	39	1.41	13.38	12.23
Preston	8	53421	19178555	11.32	40	1.48	13.22	12.99
Bristol	5	58863	12828471	11.30	41	1.37	12.60	12.27
Middlesbrough	11	58932	15249053	11.27	42	1.28	13.64	12.76
Southampton	4	47751	12570230	11.14	43	2.60	14.43	12.06

Finally, we should not ignore the fact that the geographical size of deprived areas will also influence the relative inequalities experienced by people resident there. Living in a small area of deprivation, where the distance to various resources such as cheap shops, good quality food, parks or medical facilities, may not be as detrimental to well-being as living within a large area of deprivation where the distance to such facilities is greater. Leeds ranked third in geographical area among those clusters of wards with a population larger than 200 000 (Table 7.3) and fifth among those clusters with a maximum population of 60 000 (Table 7.6). Even based on this criterion, Leeds ranked high compared with most other LADs.

7.5 Conclusion: A Tale of Two Cities

The analysis that we were asked to conduct formed part of the LCC application for Objective 2 status (Leeds City Council, 2000). The technique showed that Leeds had a strong case and that a substantial area of the inner city could be described as suffering from severe material deprivation. We believed that the LCC was correct in its assertion that the inner city area was relatively more deprived than would be indicated by a crude analysis based on data collected for LADs. Redefining Leeds based on clusters of wards in the inner city area better reflected the relative needs of Leeds compared with other English cities. Indeed, the final results in Table 7.6 demonstrate that Leeds ranks third among the most deprived inner city areas with populations less than 60 000, while Table 7.3 showed that it ranked fifth among the cities with more than 200 000 people. We concluded that Leeds was indeed a 'two-speed economy' and that despite the strong record of economic growth, the benefits were failing to 'trickle down' to the inner city residents.

The analysis was crude in the sense that it relied on a single measure of deprivation to assess the position of Leeds relative to other English cities. Little attention was given to the extent of relative inequalities that may be relevant to the Leeds inner city case. Leeds can be described as having 'two cities' where the relative inequalities between two broad groups – the rich and poor – are quite extreme. There are good reasons to suppose that people take the standards of others around them as the basis for self-appraisal and evaluation (see, for example, Boyle *et al*., 1999, 2001; Elstad, 1998). Those living in deprived conditions may find it harder to cope if they are surrounded by people that are better off than themselves, rather than being surrounded by people in similar circumstances. This appears to be particularly pertinent to the Leeds case; the analyses presented above demonstrate that there is considerable variation among those living in the Leeds LAD – more so than in most other LADs. There is good reason to suppose, therefore, that this may cause the deprivation in this area to have even worse effects than might otherwise be the case.

Finally, as contracted researchers, we should comment on our own positionality in the context of this research. Obviously, we were employed as neutral researchers who could provide an objective insight into the relative position of the Leeds inner city. It was important for LCC to be able to argue that the research was independent and this was the case. We introduced a methodological approach that was implemented as fairly as possible in every English city, and the results did indeed

demonstrate that Leeds' position was worse than a simple district-level analysis would indicate. However, while our choice of population cut-offs was not influenced by LCC, it is unlikely that the results from each would have been reported by them in their application for funding had the results not supported their general argument. This raises an interesting ethical dilemma. Do we as researchers demand that all of our results are used in any publications associated with the work, or is this an unreasonable expectation given that our services have been paid for? In this case, we felt that the use of our analysis was reasonable and that our conclusions were robust enough for LCC to use them in arguments concerning their claim for Objective 2 status. However, consultancy work of this type does raise questions that are perhaps given little attention in many studies.

Epilogue

Leeds City Council were part of the successful Yorkshire and Humberside application for Objective 2 status following the submission of its report which demonstrated, among other things, that Leeds had a significant cluster of deprived people living in its inner city, compared with other cities and towns in England. As a result, the region has been awarded a total of approximately £320 million from the EC, which when matched with UK and private funds that were also secured as part of the bid, brought the total to £870 million. Leeds, and other areas within Yorkshire and Humberside that were part of the successful bid, are now and have been, competing for these funds between 2000 and 2006 under a series of programmes designed to support business, community regeneration and physical developments.

References

Boyle, P.J., Gatrell, A.C. and Duke-Williams, O. (1999) The effect on morbidity of variability in deprivation and population stability in England and Wales: an investigation at small-area level, *Social Science and Medicine,* **49**: 791–9.

Boyle, P.J., Gatrell, A.C. and Duke-Williams, O. (2001) Do area-level population change, deprivation and variations in deprivation affect self-reported limiting long-term illness? An individual analysis, *Social Science and Medicine,* **53**: 795–9.

Committee on Spatial Development (1999) *European Spatial Development Perspective: Towards Balanced and Sustainable Development of the Territory of the European Union,* Office for Official Publications of the European Communities, Luxembourg.

Department of Environment (1995) *1991 Deprivation Index: A Review of Approaches and a Matrix of Results,* HMSO, London.

Department of Environment, Transport and the Regions (1998) *1998 Index of Local Deprivation: A Summary of Results*, DETR, London.

Department of Trade and Industry (1999) *EU Structural Funds: Determining Areas for Objective 2 Eligibility*, DTI, London.

Elstad, J.I. (1998) The psycho-social perspective on social inequalities in health, in Bartley, M., Blane, D. and Davey-Smith, G. (eds) *The Sociology of Health Inequalities,* Blackwell, London.

European Commission (1998a) *Towards an Urban Agenda in the European Union*, Communication from the European Commission.

European Commission (1998b) *Sustainable Development in the EU: A Framework for Action*, Communication from the European Commission.
European Commission (1999) *Council Regulation No 1260*, Communication from the European Commission.
Leeds City Council (1998) A tale of two cities, *New Statesman*, 22–3.
Leeds City Council (2000) *The City of Leeds – A Two Speed Economy,* Leeds City Council, Leeds.

8

GIS for Joined-up Government: the Case Study of the Sheffield Children Service Plan

Massimo Craglia and Paola Signoretta

Abstract

This chapter gives a practical example of the use of GIS in supporting public policy at the local level. The issues it raises in respect to the sharing of data among multiple agencies, and the methodology it puts forward to develop a composite indicator able to identify priority areas for intervention will be of interest to many practitioners, particularly because the project it discusses is set in the context of the preparation of the Children and Young People Service Plan, which is a statutory requirement on all local authorities under Section 17(4) of the Children Act 1989. Moreover, the range of agencies involved in this project reflects many of the partnerships being established at the local level to deliver public policy in a range of fields, from urban regeneration to crime reduction. Therefore, much of the discussion will be applicable beyond the specific application discussed here. The chapter is divided into five sections. The first introduces the project and the local setting, the second discusses the use of GIS to identify children's needs using data from different local sources; the third reports on the case study and the methodology used; the fourth introduces the results obtained, and finally the fifth draws some conclusions.

8.1 Introduction

This chapter examines the result of the *Needs Analysis and Mapping Project* supporting the preparation of the Children and Young People Service Plan (CSP)

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

2000–03 in Sheffield. The formulation of CSPs, and their yearly updates, is a statutory requirement on all local authorities (LAs) established under Section 17(4) of the Children Act 1989. The project was requested in 1998–9 by the Joint Commissioning Group (JCG), a local partnership in Sheffield, to the Sheffield Centre for Geographic Information and Spatial Analysis to support data collection and analysis as input to the preparation of the plan. The Partnership led by the City Council included representatives of South Yorkshire Police, the Department of Social Services (DSS), the Department of Education, the Young Children's Service (YCS), Community Health Sheffield, Sheffield Health Authority, and the City Council's Central Policy Unit and Department of Housing.

The chapter demonstrates the value of GIS in bringing together data coming from different agencies for policy purposes, and puts forward a simple methodology for constructing composite indices that capture different dimensions of need, and enable the identification of key priority areas for intervention. The use of composite indices is now well established for identifying need and targeting resources (see Chapters 6 and 7). These include, for example, the Townsend Index in the health field (Townsend *et al.*, 1988), and the Index of Multiple Deprivation regularly updated by the Department of Transport, Local Government and the Regions (DTLR). However, the methodological approach put forward in this chapter has three advantages. Firstly, that a simple methodology can be more easily understood and adopted by the government agencies themselves. This is important to facilitate the wider use of GIS in practice, and enable regular updates to monitor policy and assess outcomes. Secondly, by using enumeration districts as spatial units instead of the wards, which are used both by the DTLR Index and other recently proposed relevant indices (Prince's Trust, 2000), it is possible to identify some of the pockets of needs that are otherwise masked by the ward frame. Finally, for all its simplicity, we did explore the extent to which the results also stood against other methodologies, which whilst more complex, had the advantage of being grounded on firmer statistical bases. Having carried out the necessary tests, we are satisfied that the proposed methodology yields comparable results, with the bonus of simplicity and adaptability.

This project was carried out in the context of children's welfare but has a number of aspects that make it of wider interest for policy analysis: in the first place, there has been an increasing emphasis in Britain, but also in other countries, on a more integrated approach to policy making, which recognises the increasing complexity of many social issues, and therefore the need for a collaborative multi-agency approach. This often requires the integration of data coming from different agencies on a spatial basis, and the methodology to construct a composite index discussed here has wider applicability. In the second place, whilst inter-agency collaboration is fine in principle, it does not always come easily or naturally to many of those involved. Moreover, there may be just as many difficulties of communication within each agency as there are between agencies. This was well illustrated in the course of this project in relation to the challenges of collecting the data necessary from each agency. These challenges are discussed briefly in Section 8.2 and more fully in Signoretta and Craglia (2002).

The project is set in Sheffield, a city of approximately half a million people located in the north of England. Its economy has been traditionally based on steel

production and cutlery, while its hinterland was also strongly based on coal mining. The restructuring of these industries in the 1980s has had a major impact in the city. Employment in the steel industry alone fell by approximately 90%, from 45 100 in 1971 to 4700 in 1993, while another 10 000 jobs were lost in the mining industry in the surrounding region between 1985 and 1994 (Taylor *et al.*, 1996). These job losses have particularly affected the inner city wards where much of the traditional manual labour was concentrated. According to a study based on the use of a deprivation index similar to Townsend's (Woodhouse, 1999, p. 38), 'the most deprived wards, Manor, Park, Castle and Burngreave, are clearly concentrated in the inner city. They consist of both nineteenth-century industrial housing and post-war local authority estates. The next most deprived category consists of areas that have similar housing stock to the most deprived wards, such as Netherthorpe and Firth Park' (Figure 8.1). Regardless of the period of construction, many of these areas share the common characteristics of having been built when the local authority was under enormous pressure to build fast and re-house population after slum clearance or in the aftermath of World War II. Hence the quality of the construction was often low, and physical decay set in relatively quickly (Crook, 1993). Moreover, parts of these estates have been plagued for more than 30 years by high levels of crime, relative to the otherwise generally lower rates for Sheffield as a whole compared with other conurbations

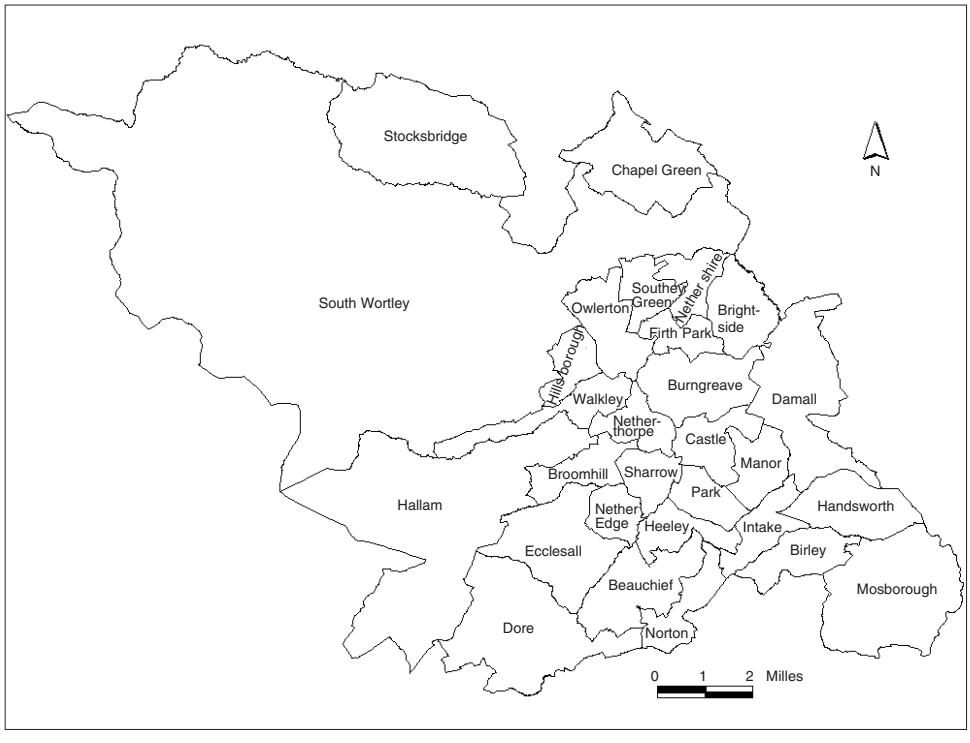


Figure 8.1 *Wards of Sheffield*

in England (Baldwin and Bottoms, 1976). The cumulative impacts of economic restructuring, physical decay and social malaise, including crime, make it all the more important to identify areas where children are most vulnerable, target intervention across many agencies in a coordinated fashion, and monitor outcomes.

With these considerations in mind, the chapter is divided into four further sections. The first sets the project in context and discusses the use of GIS in performing small areas analysis to identify children's needs using data from different local sources to support the Joint Commissioning Group's requirements; the second section reports on the case study and the methodology used; the third introduces the results obtained, finally the fourth draws some conclusions.

8.2 Small Area Analysis for the Identification of Children's Needs

Since New Labour came into power in 1997, there have been numerous policy initiatives aimed at improving children and young people's lives. These policy actions are very important as recent research work points out that experiencing financial disadvantage in childhood has severe implications in adult life (Prince's Trust, 2000). The Children's Services Plans (CSP) which are the focus of this chapter provide the general framework for the strategic planning of children and young people's services. Their importance has therefore increased as a way to help coordinate the many initiatives that are now underway in this field. With this in mind, the Joint Commissioning Group in Sheffield had two aims in setting up the Needs Analysis and Mapping Project with the University. The first was to pool together data about children from the different participating agencies, and set the framework for the yearly updating process; the second was to use the expertise at the University to have a series of analyses made that could inform the preparation of the CSP and in particular help identify areas of priority for intervention. The first aspect of the project highlighted the difficulties that those involved in joined-up initiatives face in trying to build a common base of data to support policy making. These difficulties included organisational issues (such as communication difficulties within agencies, varying organisational priorities among the agencies, and resource constraints), uncertainty on what could be shared in the light of the Data Protection Act, technical difficulties relating to the characteristics of the computer systems (both hardware and software) existing in each agency, and the degree to which they could respond to demands unforeseen at the time of their installation. These problems are discussed more fully in Signoretta and Craglia (2002), whilst the remainder of this chapter focuses on the methodological aspects related to the analysis of the data, and in particular to the development of composite spatial indices that bring together different dimensions of 'need', and help to prioritise areas for investment or additional analysis.

The 1996 Guidance (DfEE, 1996) indicates the content of the CSP for children in need[1] while the latest non-statutory guidance aimed at vulnerable children[2]

[1] Defined as children who will not achieve a reasonable standard of health and development unless they receive services under part III of the Children Act 1989.

[2] Defined as children whose life chances will be jeopardised unless action is taken to meet their needs and reduce the risk of social exclusion.

(http://www.doh.gov.uk/scg/childplan.htm) gives a more strategic role to the plan (defined as the Children and Young People Strategic Plan). The Plan should link vertically to more general plans that set out strategies concerning for instance environment and transport issues, and to more focused centrally prescribed plans such as *Sure Start* aimed at improving the health and well-being of children and their families and area-based programmes such as *Education Action Zones* aimed at raising the achievement of pupils tackling the underlying causes of poor performance and behaviour. Relevant for this work are the latter area-based programmes. These stem from the recognition that areas of deprivation and social exclusion need a multi-agency and cross-sectoral approach for their regeneration and also that the identification of these areas is vital to pursue a programme of local regeneration as proposed in *A New Commitment to Neighbourhood*: *National Strategy Action Plan*, the Government's strategy for the socioeconomic renewal of the most deprived neighbourhoods in England.

The Children and Young People Strategic Plan contains elements such as information about levels and types of needs and gaps between needs and service provision, which allows the preparation and coordination of focused joined-up intervention among the local agencies. A recent report on disadvantaged young people from the Prince's Trust (2000) highlights that mapping is an essential tool for an approach that is based on finding local solutions to local problems. Small area analysis is an important tool to help identify pockets of deprivation. The report looks at a variety of factors to define multiple deprivation for young people and describes several indices to measure disadvantage using broad measures of poverty together with measures more specific to young people. However, the analysis is based on very broad geographical areas (regions, counties, districts and wards) within which severe pockets of deprivations may be overlooked. This aspect severely limits its use for policy intervention at the local level.

The need for small area analysis has also been recognised by the Social Exclusion Unit of the Cabinet Office, and in particular by the Policy Action Team 18 Report on Better Information (PAT 18, 2000). The report recommended that a set of neighbourhood statistics at ward level should be established nationally covering access to services, community well-being, crime, economic and work deprivation, education, health, housing, and the physical environment. Following this recommendation, a new service *Neighbourhood Statistics* was established by the Office for National Statistics in collaboration with central, and local government and other agencies. These small area indicators are used at national level to compare areas targeted by a specific initiative and at local level to evaluate local conditions, to target resources, and to monitor and assess outcomes (http://www.statistics.gov.uk/neighbourhood/patinitiatives.asp). However, the choice of the ward as the basic geographic unit for the dissemination of data about neighbourhoods is problematic. As recognised by a sub-group of PAT18, which focused on Geographic Referencing, wards vary greatly in size and population and their boundaries are continuously modified.

In the UK, the use of deprivation indices is well established and a detailed review was produced by PAT 18 (for more details on indices and associated methodologies, see http://www.cabinet-office.gov.uk/seu/2000/pat18/depindices.htm). For instance, the Townsend Index of Material Deprivation used for health analysis is the sum total

of the standardised Z scores of four census variables. Other indices such as those developed by the former Department of the Environment, now ODPM, for 1991, 1998 and 2000 are based on various approaches. As an example, the latest Index of Multiple Deprivation 2000 (IMD 2000) (DETR, 2000) comprises six ward-level domain indices which were first standardised to indicate the departure of each value from its mean, then transformed to a common distribution and finally weighted. Another example especially relevant for the present work is the already mentioned study by the Prince's Trust (2000) which measures disadvantage for young people using five indices. Each index relates to a different theme (namely population, income and work, care, education and crime) with sub-components each of which is ranked according to a defined classification (for more details, see Prince's Trust, 2000, p. 33). The total score is obtained by adding up the single rankings. The advantage of this methodology is in its simplicity that makes it easier to be employed by government agencies; the downside being that the indices are developed for large units (districts, wards) which may not be very useful for policy purposes at the local level.

With these considerations in mind, this project was set up to use fine-grained socio-economic data available in the city, and use GIS as a way of integrating these data sets, and analyse them over small geographical units to support local area planning. The methodology developed and the data utilised are described in the following section.

8.3 Methodology

8.3.1 The Needs Analysis and Mapping Project

According to the project design, three main questions about children and young people needed exploring; firstly, the significant needs of under 18s in the city which affect their educational achievement, health and welfare; secondly, the distribution of these needs according to age, gender, ethnicity and geography; thirdly, the geographical distribution of children in areas affected by multiple deprivation. The project was organised in two phases: the first for data collection and input in GIS, the second for data analysis and discussion with the partnership.

The project brief contained a list of 33 data sets to collect from the participating agencies reflecting several dimensions of children and young people's lives. The list was based on the *Children's Services Planning: Guidance* (DfEE, 1996) and included information on children and young people relating to the following issues: disabilities, mental health, distribution of under 8 years olds, looked-after children, runaways, adopted children, children leaving care, protected children, young carers and children in conflict with the law. The project steering group also decided that the information was to be collected at the most disaggregated level, focusing on the post-code unit as the most appropriate scale. The time period to which the data referred was limited to one particular period (1997–8) to facilitate the collection due to the restricted time scale of the project. In some instances adjustments had to be made in relation to specific data sets. For instance, the time period chosen for education data was the academic year, rather than the calendar year. In other cases, as detailed below, it was not possible to have a whole year's worth of data but only a snapshot at a

particular time. However, in all cases, questions about particular trends affecting the data were asked of the data owners to ensure that the data sets were not affected by any particular bias.

In spite of the support verbally committed by the partners in the Joint Commissioning Group, who also 'owned' the data, it proved impossible to collect all 33 data sets. After three months of efforts, only 18 data sets were made available, even though all 33 existed, mostly already in digital form. Four data sets were dropped because they were only available in paper format or they were too time consuming to collect; for two variables only sample data were available, while nine data sets were not received. As outlined before, the problems associated with data sharing were largely due to organisational issues and the different level of priority accorded to the project in the participating agencies. Following an assessment of the overall quality and robustness of the 18 data sets received, and discussions with the steering group representing the agencies involved, six data sets were selected for analysis to detect areas where children are affected by multiple needs, and thus identify areas for priority action and funding.

For the present work it was decided that a simple but effective methodology should be employed. Thus the use of statistical tests usually included in specialised software, which may not be available to local government data analysis and research units, was avoided. The six variables utilised represent different dimensions of children's life in terms of health, education, crime and social background, these are:

- children with disabilities and special needs;
- children on the Child Protection Register;
- children in families on income support;
- children with low weight at birth;
- children excluded from school; and
- children and young people in conflict with the law.

Each variable represents a group of young people with specific needs, and their combination points out areas where children's needs are more acute.

Initially each data set was checked for invalid postcodes, and then the remaining postcodes were georeferenced using the ED/Postcode directory provided by Manchester Information & Associated Services (MIMAS, www.mimas.ac.uk) at the University of Manchester. The postcoded data were aggregated at the enumeration district level (using the 1991 boundary provided by EDINA Edinburgh University Data Library, http://edina.ac.uk), since this level would preserve the anonymity of single individuals but at the same time was fine enough to pick up pockets of deprivation within Sheffield. The six variables were combined to produce two indices: one was derived using count data while the other rates, thus reflecting different policy needs, both operational (where are the largest number of children in need?) and strategic (where do we have more or less children in need than we would expect other things being equal?). The indices measure needs of children and young people in Sheffield, not material deprivation as for instance in the Townsend Index, and are *indirect* indicators of deprivation as they represent the victims of particular disadvantaged circumstances (for more details on this issue, see

http://www.cabinet-office.gov.uk/seu/2000/pat18/depindices.htm). A description of the six variables and the denominators used is given below.

8.3.2 Description of the Data Sets Utilised

The data set on *Children with Disabilities and Special Needs* was delivered by Sheffield Social Services and includes four data sets (Children registered disabled with Sheffield Social Services, Children with a user group for disabilities, Children referred during the required period to the Social Work Team for Children with Disabilities, Children and Families User Group with a referral reason of mobility problems, disability problems or mental health). The steering group decided to include all four data sets, as this would give a reliable picture of the problem in Sheffield. Ninety-two per cent of the records could be mapped while the remaining 8% had an invalid postcode. *Children on the Child Protection Register* was also delivered by Sheffield Social Services. Ninety per cent of the records were mapped while the remaining 10% had an invalid postcode. Data on *Children aged 0–17* in families on income support was a snapshot at December 1998. This was because the data set was derived from a system designed for administrative purposes which could not easily respond to the requirements of this project. The data was supplied by the Central Policy Unit at Sheffield City Council. *Children born less than 2.5 kg* was derived from a file supplied by Sheffield Health Authority on the children born in Sheffield in 1998 with details of their weight. The steering group selected the threshold of 2.5 kg because this is the value at which health practitioners define 'under weight'. Ninety-nine per cent of the records delivered were mapped. The data set *Children excluded from school* was based on data delivered by the Department of Education and includes children aged 3–16 permanently excluded from school during the academic year 1997–8. A permanent exclusion applies when pupils are excluded from the school on a permanent basis because the school does not wish them to return. Ninety-nine per cent of the total number of records were mapped. Finally, the Youth Offending Team supplied the data set on *Children aged 10–17 in conflict with the law*. The data set contained the cases brought to court during April 1997 to March 1998. From this initial data set information on young offenders was extracted and 90% of the records were geocoded.

As far as the denominators are concerned four data sets were used: total number of children aged 0–17 (disabled children, protected children, children in families on income support), total number of children aged 3–16 (children excluded from school), total number of children aged 10–17 (children in conflict with the law) and total number of children born in Sheffield during 1998 (babies born less than 2.5 kg). The data were delivered by the Central Policy Unit; 98% of the records were geocoded. The method to combine the variables to obtain the indices is explained in the next section.

8.3.3 Construction of the Two Indices

The methodology deployed included the following steps. Firstly, for each variable we computed rates and counts at the ED level. Rates were calculated as the proportion of observed cases in each ED over the total population at risk in that ED (e.g., number

Table 8.1 *Score attributed to each class*

Class	Score
Worst 20%	5
Second worst	4
Mid-range	3
Second best	2
Best 20%	1
No cases	0

of disabled children aged 0–17 in ED/Total number of children aged 0–17 in ED). Secondly, we classified each variable in quintiles to identify the worst 20% of EDs. This was done for both sets of counts and rates. For all variables, only the EDs with cases were included in the classification and a score from one to five was assigned to the EDs falling in each group (Table 8.1), with the worst 20% having the highest score.

A score of zero was given to the EDs with no cases. In this way, each ED ended up having a score assigned for each variable for both sets of counts and rates. Thirdly, the scores given to each ED were summed and two total scores were derived, one for counts and one for rates. Finally, the two total scores were grouped in quintiles and then mapped. This methodology makes it easy to assign different weights to each variable prior to the construction of the composite index, and thus could be applied in the context of multi-criteria evaluation (see for example Henig, 1996; Hill, 1972). In the case of this project, however, the stakeholders represented in the steering group decided that differential weighting was not warranted, and therefore none was applied.

8.4 The Spatial Distribution of Child Deprivation

As far as the initial three questions are concerned, the project investigated the overall multiple needs of children and young people and their geographical distribution in Sheffield. Gender and ethnicity issues were not incorporated in the analysis due to the incompleteness of the data sets collected but will be part of the Plan update. Out of a theoretical maximum score of 30, which would indicate that an ED was consistently among the worst 20% on all six variables, the maximum score obtained in practice was 27 on counts, and 26 on rates (Figures 8.2 and 8.3).

As far as the total score on rates is concerned the worst EDs (first two categories), appear to be concentrated to the east of the city centre, with a similar pattern when looking at the counts. The figures also highlight the outliers. The outliers common in both indices are located in Southey Green, Firth Park, Manor, Norfolk Park and Park Hill, i.e. many of the areas that have been significantly impacted by economic restructuring as discussed in Section 8.1. It is important though to acknowledge that even within an estate or neighbourhoods there can be significant variations, with problematic areas sitting right next to others that are not. Here lies the value of working with fine-scale data. This issue is all the more apparent if one used wards, as suggested by PAT 18, as the basic unit for neighbourhood data collection and

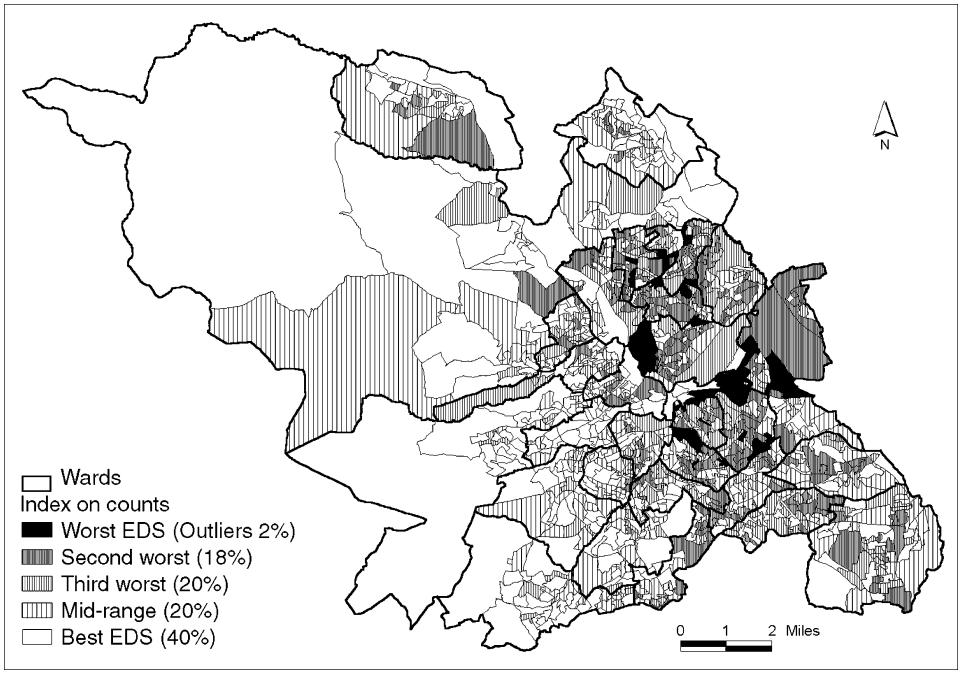


Figure 8.2 *Index based on counts*

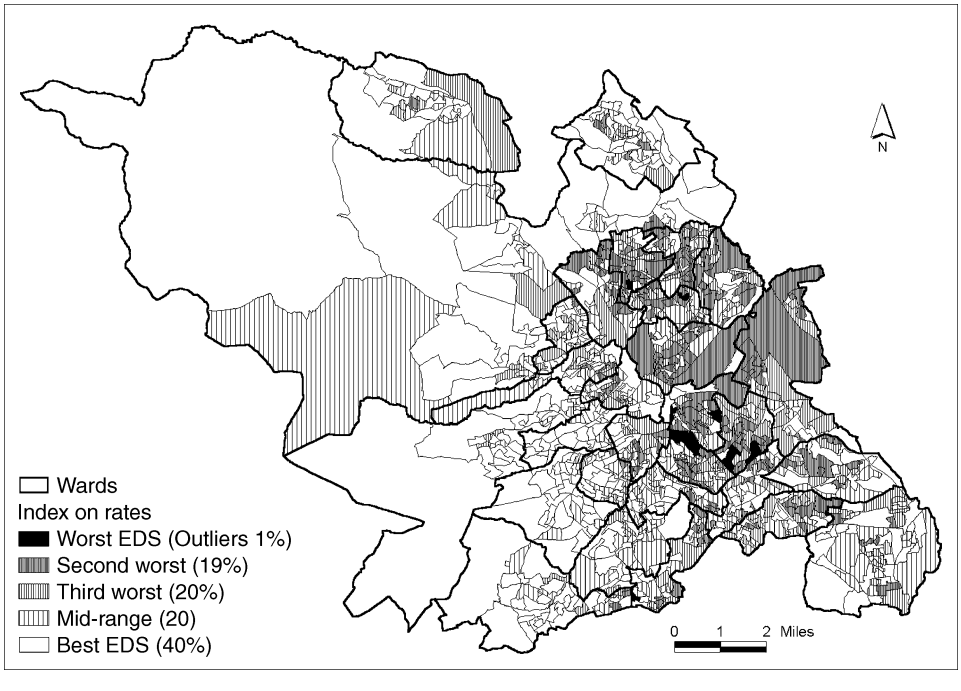


Figure 8.3 *Index based on rates*

dissemination. The figures show this by superimposing the ward boundaries onto the ED framework thus confirming the unsuitability of aggregating data at the ward level for local planning and service delivery purposes. This is yet another example of the Modifiable Areal Unit Problem (MAUP), i.e., of the extent to which the outcome of spatial analysis depends on the nature of the chosen spatial framework. Whilst well known to academics, highlighting this type of problems to the partnership funding the project was particularly useful. Partly as a result of that, the City Council is currently in the process of revising both ward and corporate area boundaries to maximise the homogeneity of key social indicators.

The advantage of the method described above to combine different variables, and identify areas affected by multiple needs lies in its arithmetic transparency, which makes it ideal for use by local authorities and agencies engaged in needs assessment and planning. Moreover, the simplicity of the method makes its outcomes more easily understood by local stakeholders and community groups. The potential drawback is that this methodology disregards any spatial structure in the data and gives no measure of statistical significance. Cognisant of these potential limitations, we undertook further analysis using the Besag–Newell test (Besag and Newell, 1991). This is a test for detecting clusters in rare events over a geographical area that has been divided into small zones (in the present study enumeration districts). For each zone, data on the population at risk must be available and a coordinate assigned to represent the location of the zone – such as the population-weighted centroid. Each event occurring in the ED is then allocated to the centroid. For each ED a population-centroid is available and for each variable and each ED a population at risk (which as in the first method will depend on the particular variable) is available and a count of the number of cases. In this application all EDs that were included in a cluster test attained the 5% significance level (for more details on the procedures followed see Craglia *et al.*, 2003). The results confirmed that the simpler method described in this chapter to create composite indices was more than adequate in identifying the key problem areas in the city. This finding is important as the methodology adopted can be easily employed for similar analysis especially in the context of local government, with some confidence that the results are similar to those obtained with more sophisticated, and complex, statistically based methods.

The results of this project proved very helpful for the planning and programming of priority actions for all the agencies involved, and gives a benchmark to evaluate the effectiveness of strategies over time. The results formed the basis for consultation by the area groups of the Young Children Service with the local communities, and were also utilised for subsequent projects. For example, the data on disabled children supported the Residential Respite Service project and the provision of additional services for disabled children in the northern part of the city. A final result was the setting up of 14 multi-agency groups each with the task to investigate issues relating to children and young people's welfare.

8.5 Conclusions

This chapter has presented a case study based on the *Needs Analysis and Mapping Project* within the context of the preparation of the Children Service Plan 2000–03

in Sheffield. In order to identify priority areas for intervention, the methodology developed has produced two indices based on a total score for counts and rates using six variables. The analysis carried out identified areas where children and young people are most affected by multiple needs. One advantage in the context of local government is that this methodology overcomes the difficulties associated with sophisticated statistical tests that are usually included in specialised software developed and mainly utilised in academic environments. One additional advantage of the present work was the small level (ED) of analysis adopted that allowed the identification of pockets of deprivation within wards which would not have emerged as affected by multiple needs.

This project demonstrated the high value that data collected from several local sources and integrated through GIS can offer to policy making in tackling local problems. The difficulties encountered in sharing the data among agencies have raised awareness among all the participant organisations of the practical difficulties of 'joined-up' working. As a result, more formalised arrangements have been put in place among all the stakeholders with written protocols committing each partner to provide key data sets to the University at regular intervals. In this context, the role of the University has been crucial as a neutral broker, able to collect highly disaggregated data, and develop small area analyses that add value to the data for policy making without infringing the requirements of the Data Protection Act. From a policy perspective, the analysis pointed out that the most deprived areas in terms of the six variables selected are still, not surprisingly, the same areas that started suffering from economic decline over 30 years ago. As this analysis concerned young people, the future for these areas and consequently for the people that live in them looks unpromising unless sustained action is taken. There are now new opportunities to help the local economy through the structural funds awarded to South Yorkshire and Sheffield from the European Union for the period 2001–06. Hence, the added importance of having in place an analytical framework to target intervention and monitor outcomes. GIS has a key role to play in this framework, as shown in this chapter. More practical applications like this will do much to raise further the awareness among policy makers of the opportunities of this technology and method of analysing spatial data.

Acknowledgements

The authors would like to thank the Young Children's Service and the Joint Commissioning Group in Sheffield for providing the data used in this analysis. However, the views expressed are those of the authors and not necessarily of the agencies involved. This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown and the ED-LINE Consortium.

References

Baldwin, J. and Bottoms, A. (1976) *The Urban Criminal,* Tavistock, London.

Besag, J. and Newell, J. (1991) The detection of clusters in rare diseases, *Journal of the Royal Statistical Society*, **154**(1): 143–56.

Craglia, M., Haining, R. and Signoretta, P. (2003) Identifying areas of multiple social need: a case-study in the preparation of Children Service Plans, *Environment and Planning C*, **21**(2) April: 259–76.

Crook, A.D.H. (1993) Needs, standards and affordability: housing policy after 1914, in Martin, D. and Binfield, J.C.G. (eds) *The History of the City of Sheffield 1843–1993*, Society Vol. 2, Sheffield Academic Press.

Department for Education and Employment (1996) *Children's Service Planning: Guidance*, G15/004 3756 1P 10k March 1996 (03).

Department for Education and Employment (2000) *Connexions, The Best Start in Life for Every Young Person*, DfEE Publications, London.

Department for Transport, Local Government, and the Regions (DTLR) (formely DETR, currentlt ODPM) (2000) Index of Multiple Deprivation. Office for National Statistics, http://www.statistics.gov.uk/.

Department of Environment, Transport and the Regions (2000) Indices of Deprivation 2000, *Regeneration Research Summary No 31*, DETR, London.

Henig, M. (1996) Solving MCDM problems: process concepts, *Journal of Multicriteria Decision Analysis*, **5**: 3–21.

Hill, M. (1972) A goals–achievement matrix for evaluating alternative plans, in Robinson I. (ed.) *Decision-Making in Urban Planning*, Sage, Beverly-Hills, CA.

Policy Action Team 18 (2000) *Better Information*, http://www.cabinet-office.gov.uk/seu/2000/pat18/.

Prince's Trust (2000) *Mapping Disadvantage*, Greenhouse, London.

Signoretta, P. and Craglia, M. (2002) Joined-up government in practice: a case study of the children needs in Sheffield, *Local Government Studies*, **28**(1): 59–76.

Taylor, I., Evans, K. and Fraser, P. (1996) *A Tale of Two Cities*, Routledge, London.

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and Deprivation: Inequality and the North*, Croom Helm, London.

Woodhouse, P. (1999) Deprivation and health in Sheffield, *Geography Review*, November: 36–41.

9

The Application of New Spatial Statistical Methods to the Detection of Geographical Patterns of Crime

Peter Rogerson

Abstract

The evolution of geographic information systems has spawned a resurgence in the development of spatial statistical methods. In this chapter, several recent developments in the field of spatial statistical analysis are summarised and applied to burglary data for census tracts in Buffalo, New York. The methods employed here are concerned with the detection of spatial patterns in geographic data. Some of the methods are concerned with finding regions with rates that are significantly higher than those that could be expected under the null hypothesis of no geographic pattern, relative to prior expectations. Others have as their goal the detection of temporal change in spatial data.

9.1 Introduction

As the value of geographic information systems (GIS) has become more evident to researchers and practitioners in an expanding array of disciplines, the demand for spatial analysis, spatial modelling and spatial statistics has witnessed substantial growth. Although the early benefits of GIS derived primarily from their ability to store, retrieve and display spatial information, GIS is now increasingly used in the analysis of spatial data. The addition of modules such as Spatial Analyst and

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

Geostatistical Analyst to ArcView, and Spatial Statistics to S-PLUS represent examples of how vendors are beginning to meet the needs of the new marketplace for geographical analysis. The effects of these changes vary from one subfield to another. In epidemiology, the use of spatial modelling and spatial statistics has soared in recent years. Many new methods for the statistical assessment and modelling of spatial patterns have been developed recently (e.g., Elliot *et al.*, 2000; Lawson *et al.*, 1999; Lawson, 2001). In crime analysis, the development of new statistical methods has perhaps been less apparent, but the use of GIS by police agencies as well as by other local government agencies has been much more widespread than in public health and epidemiology.

This chapter demonstrates how recent advances in spatial statistical analysis may be used in the context of crime analysis. Section 9.2 provides a brief description of the burglary data (from Buffalo, New York) that will be used to illustrate the various methods. Section 9.3 describes several alternative approaches to detecting geographic patterns in the data, and in each case the approach is illustrated by using the burglary data.

9.2 Burglary Data

Data on burglary incidents in the city of Buffalo, New York, for 1998 is used to illustrate the methods. Buffalo has a population of approximately 300 000 residents. Its population has declined steadily over the last several decades as it has coped with the difficulties associated with transforming its economy and becoming less reliant on what was a once-strong manufacturing base. The maintenance of infrastructure and the provision of services are, of course, difficult for declining manufacturing cities. In part reflecting these difficulties, the Buffalo Police Department (BPD) has had to operate with a crime analysis unit that is understaffed. Consequently, the crime analysis unit has a greater focus on database development and reporting, and less of a focus on statistical analysis.

The BPD has simultaneously maintained strong links with the University at Buffalo, and the two have been involved with several joint projects. Within this context, and with the funding of a grant from the National Institute of Justice to (a) develop new methods, and (b) work with local agencies, an effort was made to examine geographic aspects of crime. In particular, we worked with the Crime Analysis Unit to develop new methods for monitoring temporal changes in the geographic patterns of crime.

Information is available on the number of burglaries by census tract, as is information on the coordinates of tract centroids. In addition, the data for 1998 is disaggregated by quarter; a summary of the number of incidents in each quarter is given in Table 9.1. Since there are 90 census tracts in the city, there is an average of approximately 15 burglaries per tract, per quarter.

The data are spatial in the sense that regional frequencies are available for the 90 census tracts that comprise the city of Buffalo. Consequently, the methods discussed in this chapter focus on those that are useful for spatial data. Other methods are relevant for situations where coordinates are available for the point locations of events; these latter methods are not covered in this chapter.

Table 9.1 *Burglary incidents by quarter in 1998, Buffalo*

Period	Number
January–March	1364
April–June	1540
July–September	1645
October–December	1564
Total	6113

Table 9.2 *Chi-square test comparing observed frequency of burglaries with that expected under the null hypothesis of no spatial pattern*

Period	χ^2
January–March	810.3
April–June	925.6
July–September	1027.8
October–December	875.5

Note: The critical value of chi-square, when $df = 89$ and $\alpha = 0.05$, is approximately 112.

9.3 Analysis of Burglary Data Using Alternative Spatial Statistical Methods

9.3.1 Expected Versus Observed Frequencies

The expected numbers of burglaries in each tract were calculated by assuming that there was no spatial pattern in the burglary data, and by conditioning on the total number of burglaries in each quarter. Specifically, expected values were determined by multiplying the total number of burglaries in a quarter by each tract's share of city-wide population. One straightforward way to compare observed and expected frequencies is via the common, aspatial, chi-square test:

$$\chi^2 = \sum_i \frac{(f_{obs,i} - f_{exp,i})^2}{f_{exp,i}} \tag{9.1}$$

where the fs refer to tract i-specific observed and expected frequencies, and the sum is over all tracts. This statistic has a chi-square distribution with degrees of freedom equal to the number of tracts minus one, under the null hypothesis that the number of burglaries observed in tracts is generated by distributing the total number of burglaries randomly across tracts, in accordance with tract population shares.

The results, shown in Table 9.2, indicate a very high degree of statistical significance. In all cases the observed value of the chi-square statistic exceeds greatly the

critical value of 112, and the pattern of crime is thus substantially different from the underlying pattern of population. This is not particularly surprising; it is well known that crime tends to cluster geographically, and the results are consistent with the visual impression one obtains from Figure 9.1. Of interest is the fact that the fit is different in each quarter; the burglary pattern most closely resembles population distribution in the first quarter (even though it is very different from it), and the

Figure 9.1 *Burglary incidences in Buffalo by census tract, 1996*

pattern of burglaries is most different from the spatial distribution of population in the third quarter. We will investigate these differences in more depth in the following sections.

9.3.2 Changes in Spatial Pattern: Redefining Expectations

After finding that the spatial pattern of burglaries differs significantly from that which would be expected if the true probability of a burglary were proportional to population, we now turn our attention to the changes in spatial patterns that occur from one quarter to the next. An alternative way to derive expected frequencies is to use the spatial pattern of crime itself, rather than the spatial pattern of population. One might, for example, choose to use the spatial pattern of burglaries that prevailed during the previous year as a basis for expectations. Then, any significant difference between what is observed in the current year and what is expected (on the basis of the past year) would pertain to *changes* that had occurred in the spatial pattern. Since we do not have access to 1997 burglary data for Buffalo, we will simply use the complete set of data for 1998. Thus the expected frequency of burglary in a census tract, in a given quarter, is now derived as the total number of burglaries in the quarter, multiplied by the fraction of all 1998 burglaries that occurred in that particular census tract. Note that this particular calculation of expectations uses some of the actual data, and hence judgements of statistical significance based on these expectations might be viewed as conservative. An alternative might be to calculate expectations for each quarter based upon the spatial pattern that prevailed during the other three quarters, but that seems to be even more *ad hoc*. Clearly the question of what data to use in the definition of expectations is an important one. Most often one will want to use previous, historical data. Even then, the question of how much historical data (e.g., last quarter, last year or last three years) is a difficult one.

Table 9.3 reveals that only the first quarter, January–March, has a spatial pattern that differs significantly from the pattern that prevailed for the entire year. Of course it is of interest to know *how* the January–March pattern differs from the remainder of the year (e.g., where are burglaries higher or lower than expected?). This question will be addressed in section 9.3.4.

Table 9.3 *Chi-square test comparing observed frequency of burglaries with that expected under the null hypothesis of no spatial deviation from the annual pattern*

	χ^2	Significance (p)
January–March	129.9	0.003
April–June	63.9	0.979
July–September	106.4	0.101
October–December	77.6	0.800

9.3.3 A Spatial Chi-Square Test

When the categories in a chi-square test represent a set of geographic regions, the observed frequencies in each category/region may be spatially dependent, and one approach to account for this dependence is to deflate the usual chi-square statistic in an appropriate manner (Fingleton, 1983, 1986). Cerioli (1997) has indicated how chi-square tests should be modified in the presence of spatially autocorrelated data for 2×2 tables. A specific weakness of the chi-square goodness-of-fit statistic, when applied to spatial data, is that it does not account for *where* the greatest deviations between observed and expected values are, relative to one another. Figure 9.2, taken from Rogerson (1999), shows that both (a) and (b) have the same chi-square statistic, and yet the spatial pattern of deviations is quite different. In (a), there is a distinct cluster of regions where the fit is poor, while in (b), the spatial distribution of deviations is more random. Suppose that the usual, aspatial chi-square statistic is computed for (a) and (b), and it is found that it is almost, but not quite, significant. It would be desirable to have a statistic with which we would conclude that in (a), the combination of the aspatial deviations *and* the spatial pattern of deviations could not have occurred by chance (that is, where the observed *spatial* pattern of frequencies is not likely under the null hypothesis of independent, expected regional frequencies).

Likewise, Moran's *I* statistic might be used to examine the spatial pattern of deviations but, by convention, the diagonal elements of the spatial weight matrix are zero, implying that deviations between the observed value and the expected value within

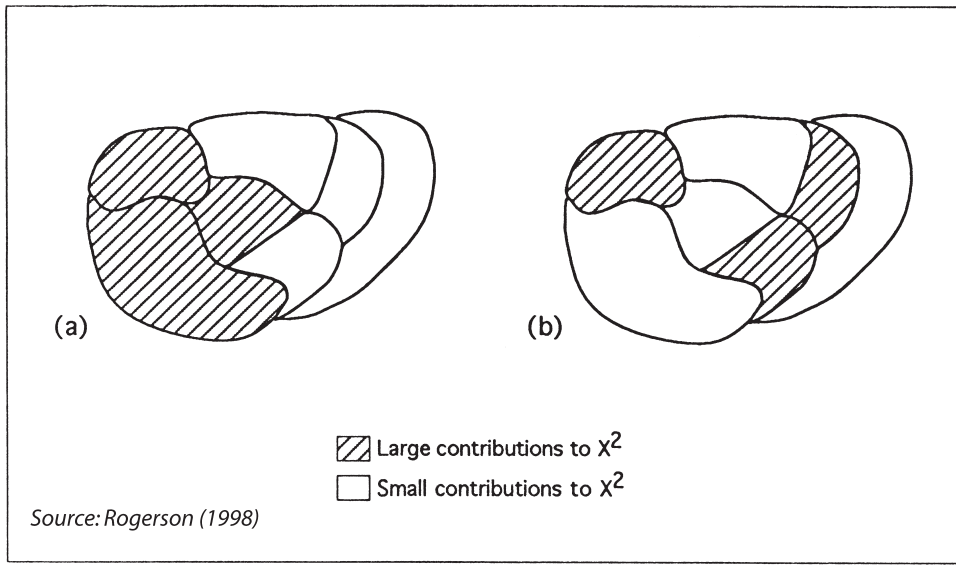


Figure 9.2 *Two patterns of deviations (observed–expected) yielding identical χ^2 values. From Rogerson (1998)*

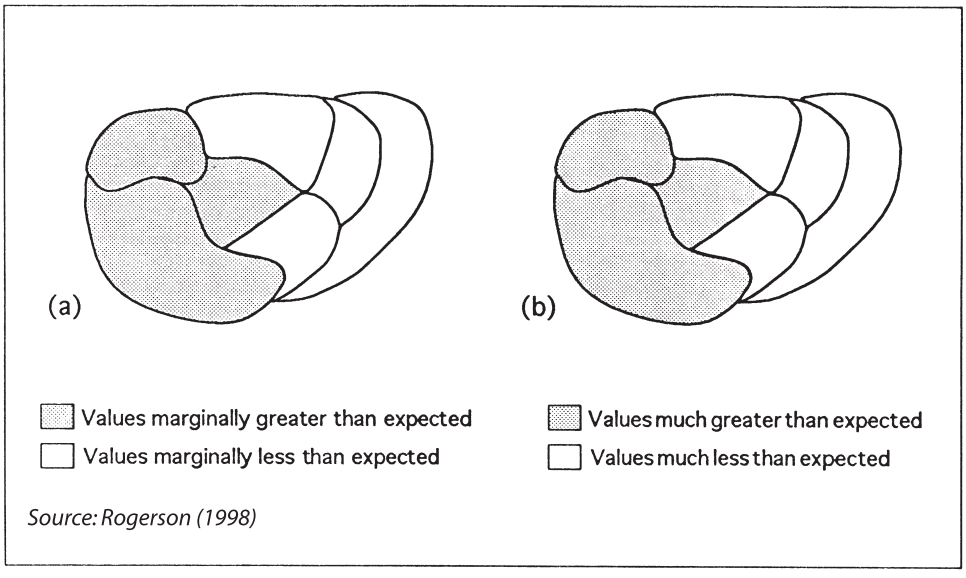


Figure 9.3 *Two patterns of deviations (observed–expected) yielding identical I values. From Rogerson (1998)*

each of the regions are not accounted for. Figure 9.3, also from Rogerson (1999), shows two situations that could have identical I values, and yet in (a) the pattern is consistent with the null hypothesis, while in (b) the deviations are unlikely to have arisen under the null hypothesis. Whereas the chi-square goodness-of-fit test focuses upon individual regions and ignores the geographical pattern of deviations, Moran's I focuses upon the pattern of deviations, but ignores the fit *within* regions.

Rogerson (1998, 1999) introduced the spatial chi-square statistic, R, to construct a test of the null hypothesis that a set of m observed regional frequencies, N_i, $i = 1, 2, \ldots, m$, could have occurred under the following null hypothesis:

$$H_0: E[N_i] = \lambda\xi_i, \quad i = 1, 2, \ldots, m. \tag{9.2}$$

For example, in the case of the spatial pattern of crime, the population of each region is given by ξ_i, the overall crime rate in the population is λ, and the expected number of cases in each region is $E[N_i]$.

Rogerson defines the spatial version of the chi-square goodness-of-fit statistic as

$$R = (\mathbf{r} - \mathbf{p})' \mathbf{W} (\mathbf{r} - \mathbf{p}) \tag{9.3}$$

where there are m regions, and the observed frequencies are used to construct the $m \times 1$ vector of proportions $\mathbf{r}$, where $r_i = N_i/N$ and N is the total number of observed cases. The values $\xi_1, \xi_2, \ldots, \xi_m$ are known quantities, which, for example, might be taken as the populations of the regions. The nonrandom $m \times 1$ vector of expected

proportions is **p**. A natural definition is to have $p_i = \xi_i/\xi$, where $\xi = \xi_1 + \xi_2 + \ldots + \xi_m$ and $\lambda = N/\xi$. Also, **W** is an $m \times m$ matrix containing elements w_{ij} (including elements along the diagonal, where $i = j$):

$$w_{ij} = \frac{a_{ij}}{\sqrt{p_i p_j}} \tag{9.4}$$

The elements of a_{ij} measure the 'connectedness' of region i to region j, and are discussed further below. The spatial chi-square statistic tests the null hypothesis by combining terms that focus on within-region deviations (as does the chi-square goodness-of-fit statistic) and the spatial pattern of deviations found in nearby regions (as does Moran's I).

High values of the statistic indicate that there may be single regions where observed and expected proportions differ significantly, and/or there may be areas where positive (or negative) deviations between observed and expected proportions exhibit significant spatial autocorrelation. Since R may also be written as

$$R = \sum_i \sum_j w_{ij}(r_i - p_i)(r_j - p_j) = \sum_i \frac{(r_i - p_i)^2}{p_i} + \sum \sum_{i \neq j} a_{ij} \frac{(r_i - p_i)(r_j - p_j)}{\sqrt{p_i p_j}} \tag{9.5}$$

it may be conceptualised as the sum of two terms – the aspatial chi-square statistic, and a term that accounts for the spatial pattern of deviations between observed and expected values. High values of R will lead to the rejection of the null hypothesis for at least one of two reasons: either there are large deviations between observed and expected regional values (in keeping with the usual interpretation of the aspatial chi-square test), or the deviations exhibit a significant spatial pattern, or both.

The spatial chi-square statistic is closely related to Tango's (1995) statistic:

$$C_G = (\mathbf{r} - \mathbf{p})' \mathbf{A} (\mathbf{r} - \mathbf{p}) \tag{9.6}$$

where **A** is a $m \times m$ matrix containing the elements a_{ij} that measure the degree to which regions are connected. Tango uses

$$a_{ij} = e^{-d_{ij}/\tau} \tag{9.7}$$

$$a_{ij} = 1 \tag{9.8}$$

where d_{ij} is the distance between regions i and j, and τ is a measure of the importance of distance (where larger values of τ correspond to larger 'spheres of influence' around each region). Tango gives the expected value and variance of C_G as

$$E[C_G] = \frac{1}{N} \mathrm{Tr}(\mathbf{A}\mathbf{V}_p) \tag{9.9}$$

$$V[C_G] = \frac{2}{N} \mathrm{Tr}(\mathbf{A}\mathbf{V}_p)^2, \tag{9.10}$$

where

$$\mathbf{V_p} = \Delta(\mathbf{p}) - \mathbf{pp}' \tag{9.11}$$

with $\Delta(\mathbf{p})$ defined as an $m \times m$ diagonal matrix containing the elements of $\mathbf{p}$ on the diagonal. Furthermore, the null hypothesis may be tested using the chi-square distribution. Specifically, the quantity

$$\nu + \frac{C_G - E[C_G]}{\sqrt{V[C_G]}}\sqrt{2\nu} \tag{9.12}$$

has an approximate chi-square distribution, with ν degrees of freedom, where

$$\nu = \left[\frac{(\mathrm{Tr}(\mathbf{AV_p}))^{1.5}}{\mathrm{Tr}(\mathbf{AV_p})^3}\right]^2 \tag{9.13}$$

The significance of the spatial chi-square statistic, R, may be assessed by replacing $\mathbf{A}$ with $\mathbf{W}$ in equations (9.9)–(9.13).

Results for the burglary data are shown in Table 9.4 for various choices of τ (which can be roughly interpreted as proportional to cluster size). For a Type I error of $\alpha = 0.05$, the statistic is significant only for the January–March data (though note that the statistic would be barely significant at $\alpha = 0.10$ for the July–September data). For January–March, the significance is highest ($p = 0.002$) when $\tau = 0.9$. As τ increases, the results approach those of the aspatial chi-square test. The results here are similar to those for the aspatial chi-square test (Section 9.3.1); when used in an exploratory fashion, by examining different values of τ, the spatial chi-square test has the additional capability of better capturing the scale on which clustering is occurring.

9.3.4 Outliers and Local Tests

When the global chi-square test (spatial or aspatial) is significant, it is of interest to know which categories (in this case, census tracts) contributed to the significant value. Perhaps it was simply one or two tracts that might be then considered as potential

Table 9.4 *Spatial chi-square tests*

	Jan–Mar		Apr–Jun		Jul–Sept		Oct–Dec	
τ	R	Sig.	R	Sig.	R	Sig.	R	Sig.
0.5	.108	.006	.0508	.496	.0590	.195	.0522	.429
1	.097	.002	.0429	.962	.0644	.107	.0535	.623
1.5	.0953	.003	.0416	.978	.0648	.099	.0501	.781
2.	.0952	.003	.0416	.979	.0647	.101	.0497	.799
Optimal	.098	.002	.0408	.147	.0648	.099	.0286	.356
τ	(0.9)		(0.2)		(1.3)		(0.2)	

outliers. Also, when the global test is insignificant, it is still possible that an examination of individual categories will uncover instances where the observed value differs significantly from the expected value (and these might then be considered local hotspots). Fuchs and Kennett (1980) propose the *M* test to find outlying cells in the multinomial distribution. The test may be used in conjunction with the global, aspatial chi-square test (and the related quadrat test). Their statistic is the maximum adjusted residual. For cells $i = 1 \ldots, m$ define

$$z_i = \frac{N_i - Np_i}{\sqrt{Np_i(1-p_i)}} \tag{9.14}$$

A one-sided test is based on max z_i; $i = 1 \ldots, m$, and bounds for the critical value are

$$\Phi^{-1}\left\{1 - \frac{m - [m^2 - 2\alpha m(m-1)]^{1/2}}{m(m-1)}\right\}, \Phi^{-1}\left\{1 - \frac{\alpha}{m}\right\} \tag{9.15}$$

where Φ is the cumulative distribution function for the standard normal distribution. The upper bound can be recognised as the common Bonferroni adjustment. Fuchs and Kennett (1980) show that the test is more powerful than the global chi-square test in the presence of a single outlier. Thus the *M* test can be used in conjunction with the global test – when the global test is significant, the *M* test may be useful in determining whether the global significance is achieved via contributions from many regions, or via contributions from only a small number of outliers. Similarly, when the global test is not significant, the *M* test may still uncover anomalous values that remain hidden in the global statistic, since it is more powerful at detecting outliers.

To determine those census tracts that contributed significantly to the overall chi-square value in the burglary data, the tract-specific local *z*-statistics were calculated using equation (9.13). Using (9.12), with $k = 90$, $\alpha = 0.05$, bounds for the critical values of $M = \max z$ are $\Phi^{-1}(.99943, .999444) = \{3.253, 3.261\}$. The observed maximal *z*-values for each quarter are shown in Table 9.5. In each of the first three quarters, the observed value of $M = \max z$ is significant (though only marginally so in quarters two and three). The most significant value is found in the first quarter – a result that is consistent with the global analysis summarised in Table 9.3. Table 9.5 also depicts the tract number associated with the maximal *z* value; in all four quarters, only one of the 90 tracts had a *z*-value exceeding the critical value, and thus there were no other tracts that contributed significantly to the overall chi-square.

Table 9.5 *Results of M test*

Period	Maximum *z*	Tract
January–March	3.68	69
April–June	3.29	87
July–September	3.30	26
October–December	2.67	75

9.3.5 A Local Version of *R*

A recent development in spatial statistics has been the advent of 'local' spatial statistics that focus more specifically upon the possible existence of spatial association as viewed from a particular local region (e.g., Getis and Ord, 1992; Ord and Getis, 1995; Anselin, 1995). Thus 'hotspots' of significant local positive spatial autocorrelation may be detected, despite the absence of pattern on a global scale. Also, as Anselin has indicated, local statistics are useful in detecting outliers when the global indicator is significant. For example, in their application to international trade, O'Loughlin and Anselin (1996) found that a significant global measure of autocorrelation was due to a small number of outliers, and when these were removed, the spatial pattern of the remaining values was essentially random. The local version of the spatial chi-square statistic, R, for a given location i, is:

$$R_i = \frac{(r_i - p_i)}{\sqrt{p_i}} \sum_j \frac{a_{ij}(r_j - p_j)}{\sqrt{p_j}} \tag{9.16}$$

Under the null hypothesis, the quantity

$$1 + \frac{R_i - \mathrm{E}[R_i]}{\sigma[R_i]}\sqrt{2} = \frac{R_i}{\mathrm{E}[R_i]} \tag{9.17}$$

can be approximated by a chi-square distribution, with one degree of freedom. The expectation is defined as

$$\mathrm{E}[R_i] = \frac{a_{ii}(1 - P_i)}{N}$$

This local test can be thought of as a spatial generalisation of the aspatial M test, so that the assessment of fit in individual sub-regions also includes an assessment of how similar the fit is in nearby sub-regions as well.

Table 9.6 displays the maximum value of R in each of the four time periods. This is done for that value of τ which yields the highest local value of R. The results are similar to those shown in Table 9.5 – the first time period has the most significant local spatial association, around census tract 69. The next two quarters are

Table 9.6 *Local R tests*

Period	Local statistic	Max τ	Observation number
January–March	13.53	τ = high and flat	69
April–June	10.85	τ = 1.7 (flat)	87
July–September	12.358	τ = 0.46	26
October–December	10.175	τ = 0.52	75

Note: (R/E[R]) is approximately chi-square with 1 degree of freedom; under the null hypothesis we use $3.26^2 = 10.63$ as the critical value (where 3.26 is the critical value of a standard normal random variable with $n = 90$ tests).

marginally significant, and the last quarter fails to reject the null hypothesis that any single tract significantly differs from the expected number of burglaries (where expectations are based here upon the annual pattern). For the first two quarters, the value of τ leading to the maximum value of R is quite high; the results of the maximum local R test approach those of the M test as τ gets large.

9.3.6 The Q^2 Statistic for Detecting Change in Multinomial Distributions

When spatial data are available for a number of time periods, it is of interest to test the null hypothesis of no change in spatial pattern, versus the alternative that a shift in pattern occurred at a particular point in time. Srivastava and Worsley (1986) discuss how changes may be detected retrospectively in data that come from a multinomial distribution. Let $\mathbf{x}_t$ be a $1 \times (m-1)$ row vector with elements indicating the number of crimes in each of $m-1$ tracts during period t. There are m tracts in the study region, and so one region is arbitrarily omitted from the vector. For n time periods, the data may be given as an $n \times (m-1)$ matrix, $\mathbf{X}$, consisting of n of these row vectors. Srivastava and Worsley (1986) derive a likelihood ratio statistic that is asymptotically equivalent to the maximum of χ_r^2 over r. Here χ_r^2 is the usual chi-square statistic used for testing for association in the $2 \times (m-1)$ table formed by first segmenting the observations into the two sets $t = 1, 2, \ldots, r$ and $t = r+1, \ldots, n$ and then forming the two aggregate row vectors corresponding to the two sets of observations. The distribution of the test statistic, $Q^2 = \max\chi_r^2$, under the null hypothesis of no change, is given by

$$p(Q^2 > c) = 1 - F_m(c) + q_1 \sum_{r=1}^{n-2} t_r - q_2 \sum_{r=1}^{n-2} t_r^3 \tag{9.18}$$

where $f_m(c)$ and $F_m(c)$ are, respectively, the probability density and cumulative density function of chi-square random variables with m degrees of freedom. Furthermore,

$$q_1 = 2(c/\pi)^{1/2} f_m(c) \tag{9.19}$$

$$q_2 = \frac{q_1}{12}\left[\frac{(m^2-1)}{c} + c - 2m + 1\right] \tag{9.20}$$

and

$$t_r = \sqrt{1-\rho_r} \tag{9.21}$$

$$\rho_r = \left(N_r / N_r^*\right)^{1/2} \Big/ \left(N_{r+1} / N_{r+1}^*\right)^{1/2} \tag{9.22}$$

where N_r is the total number of observations in the first r rows of the table, and $N_r^* = N - N_r$, with the total number of observations equal to N.

Extramultinomial variation may be due either to situations where (a) the probabilities of falling into particular categories are themselves subject to uncertainty, or

(b) where there are correlations between successive rows of the contingency table. To model this, one first calculates

$$\hat{\sigma}_r^2 = \frac{Q - Q_r^2}{(n-2)m}; \quad n \geq 3 \tag{9.23}$$

where Q is the value of χ^2 measure of association between rows and columns for the entire $n \times (m + 1)$ table. Then define $K_r^2 = Q_r^2/\hat{\sigma}_r^2$ as the test statistic. Under the null hypothesis, K_r^2/m has an F distribution with m and η degrees of freedom, where $\eta = (n - 2)m$.

Results are shown in Table 9.7. These results were derived by omitting the last census tract from the analysis; again the results are consistent with those found previously. The most likely 'change point' is between the first quarter and the remainder of the year (p-value = 0.004); no other break-point yields a greater distinction between 'before' and 'after' patterns. The geographic pattern of burglaries during the fourth quarter is very similar to the pattern during the first three quarters ($p = 0.94$).

Once a change point has been found, it is then of interest to see whether other change points exist within either the 'before' or 'after' segments of the data. The bottom part of Table 9.7 reveals that further disaggregation of quarters 2, 3 and 4 does not uncover additional, significant change in the spatial pattern of burglaries.

9.3.7 Spatial Outliers

As we have seen, local spatial statistics may be used to pick out local regions with values that are significantly higher or lower than expected. They are defined for individual regions as a function of the value of the variable in that region, and values of the variable in nearby regions. Local statistics are designed primarily for testing hypotheses of spatial association for particular localities; the issue of multiple testing arises when one wishes to test more than one local statistic for significance. The M test applied to spatial data and the local R test were implemented with a Bonferroni adjustment. However, since there is correlation among tests of local statistics that are near to one another in space, the Bonferroni adjustment is conservative, and a more refined adjustment is desirable. The method summarised in this section addresses this issue.

Bailey and Gatrell (1995) and others have described the use of kernel-smoothing as one way to represent the spatial variability in the mean of a variable of interest. The value at any particular location is taken to be a weighted function of the values

Table 9.7 *Srivastava and Worsley's maximal chi-square*

	χ_r^2	σ_r^2	$K = Q_r^2/\sigma_r^2$	K/m	p-value
Quarter 1 vs Quarters 2, 3, 4	167.2	1.16	143.9	1.62	0.004
Quarters 1, 2 vs Quarters 3, 4	126.6	1.39	91.1	1.02	0.44
Quarters 1, 2, 3 vs Quarter 4	101.3	1.53	66.1	0.74	0.94
Quarters 2, 3 vs Quarter 4	107.3	1.13	94.8	1.07	0.38
Quarter 2 vs Quarters 3, 4	85.8	1.37	78.1	0.70	0.95

in the neighbourhood of the location, with closer locations receiving higher weights. The result is a surface portraying the regional variation in the underlying value, smoothed enough to eliminate the roughness of the image that would result if the original data were used, but not so much that underlying geographic variability is eliminated. Although these images represent a useful visual way to explore data, one often wishes to assess the significance of peaks in the surface. Attempts at such hypothesis testing have been limited to Monte Carlo simulation (e.g., Kelsall and Diggle, 1995, for an application to point data) or to more formal statistical methods that do not control properly for the likelihood of a Type I error (e.g., Bowman and Azzalini, 1997).

In this section, there is a summary of a statistical method for the detection of geographic clustering that is based upon Worsley's (1996) work on the maxima of Gaussian random fields and developed further in Rogerson (2001). The method provides a way to assess the significance of the maximum of a set of local statistics. It also may be viewed as a method that allows for the assessment of the statistical significance of a kernel-based, smoothed surface. Finally, the method is similar in concept to spatial scan statistics (e.g., Kulldorff and Nagarwalla, 1995) since, like the scan statistic, it considers many possible sub-regions and evaluates the statistical significance of the most extreme value. In addition, the method yields a calculable critical value that may be derived without resorting to Monte Carlo simulation methods.

The method assumes that the underlying data are normally distributed, and a Gaussian kernel is then used to smooth individual values and to construct local statistics. The steps involved may be described as follows:

1. Standardise the original observations so that the regional values may be considered as observations from a normal distribution with mean zero and variance one.
2. Construct local statistics z_i for each region using the standardised original observations (denoted y), and weights defined as

$$w_{ij} = \left(\sqrt{\pi}\sigma\right)^{-1} \exp\left(-d_{ij}^{\,2} \big/ 2\sigma^2\right) \tag{9.24}$$

   where d_{ij} is the distance from cell i to cell j and σ is chosen as the standard deviation of a normal distribution that matches the size of the hypothesised cluster. Then define the local statistic as

$$z_i = \sum_j w_{ij} y_j \Big/ \sqrt{\sum_j w_{ij}^2} \tag{9.25}$$

   To account for edge effects, it can be useful to add a guard area around the study region, with values for sub-regions in the guard area assigned from a standard normal distribution.
3. Find the critical value M^* such that $p(\max z_i > M^*) = \alpha$ by using that value of M^* that leaves probability $(1 + .81\sigma^2)\alpha/A$ in the tail of the standard normal distribution, where the study region has been subdivided into a grid of A square cells, each having side of unit length. Alternatively, M^* may be approximated by

$$M^* = \sqrt{-\sqrt{\pi}\ln\left(\frac{4\alpha(1+.81\sigma^2)}{A}\right)} \tag{9.26}$$

Alternative improvements to the Bonferroni adjustment are addressed in Rogerson (2002).

Figure 9.4 displays the areas where the observed number of burglaries is significantly greater than expected (where expected numbers are based upon the background population). Dots on the maps represent census tract centroids. Of interest is the degree to which the shaded regions are changing over time. Although there are no strong visual changes from one figure to the next, the results are consistent with the tests described above, since the clustering shown in Figure 9.4a (January–March) is, visually, somewhat different from Figures 9.4b–d. The clustering shown in Figures 9.4b–d occupies a greater sub-area, and the intensity of clustering (which can be determined from the scale) is stronger in these figures than it is in Figure 9.4a. This is consistent with the findings of the global, aspatial chi-square test, which found that the pattern of crime was most similar to the underlying population in the first quarter.

9.4 Summary

In this chapter, we have reviewed some recent developments in spatial statistical analysis, and their applications to crime analysis. The methods discussed here are for spatial data, and they have as their objective the detection of any deviations from expected geographic patterns. A key element of their use is to pay important attention to the definition of expectations. It is well known that patterns of crime exist, over and above their relation to background population, or to, say, the size of the male population in the young adult age groups. One approach would be to introduce other, known covariates (such as measures of socioeconomic status) in the definition of expectations. If the interest is in changes in patterns over time, the geographic patterns that held during some previous period may be used for the expectation.

The focus in this chapter on the temporal differences in geographic patterns has been one that assumes such detection is retrospective; i.e., a given set of data covering multiple time periods is searched for potential change points. When interest is in the *prospective* detection of change, the goal is to uncover any changes in spatial patterns as quickly as possible. If the time series of data describing the geographic pattern is of sufficient length, and if data on the areal location of successive observations is available, Rogerson (1997) suggests how quick detection may be carried out by combining Tango's statistic with the cumulative sum methods used in industrial process control (see also Montgomery, 1996). When the objective is similar and the data are for point locations, Rogerson and Sun (2001) demonstrate how quick, prospective detection of changes in spatial patterns may be achieved by combining the nearest neighbour statistic with cumulative sum methods.

As the development of spatial statistical analysis and spatial modelling continues to evolve, there will be a continuing need to embed the tools of spatial analysis within GIS. The emergence of packages such as *CrimeStat* (Levine and Canter, 1998) and

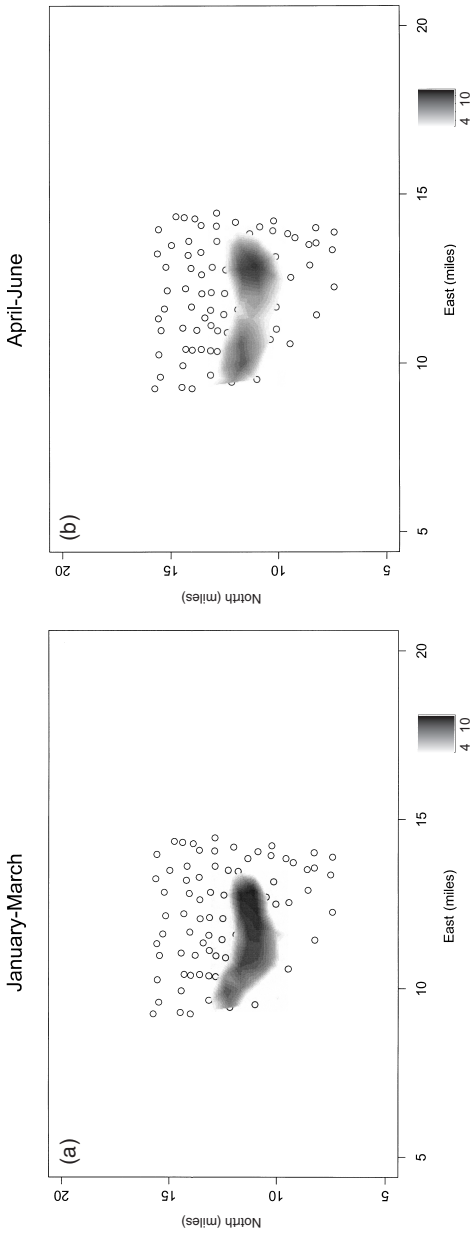


Figure 9.4 *Burglary clusters in Buffalo in (a) January–March, (b) April–June, (c) July–September and (d) October–December*

the Spatial Statistics package associated with SPLUS is an encouraging sign that these needs of spatial analysts will be met.

Acknowledgements

This chapter is based upon work supported by Grant No.98-IJ-CX-K008 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. It was also supported by the National Science Foundation under Award No. BCS-9905900. The points of view in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice or the National Science Foundation. The support of a Guggenheim Fellowship is also gratefully acknowledged. I would also like to thank Daikwon Han for his assistance in producing several of the figures.

References

Anselin, L. (1995) Local indicators of spatial association-LISA, *Geographical Analysis*, **27**: 93–115.

Bailey, T. and Gatrell, A. (1995) *Interactive Spatial Data Analysis,* Longman, Essex (published in the US by Wiley).

Bowman, A. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis,* Oxford University Press, New York.

Cerioli, A. (1997) Modified tests of independence in 2×2 tables with spatial data, *Biometrics*, **53**: 619–28.

Elliot, P., Wakefield, J.C., Best, N.G. and Briggs, D.J. (eds) (2000) *Spatial Epidemiology: Methods and Applications*, Oxford University Press, Oxford.

Fingleton, B. (1983) Independence, stationarity, categorical spatial data and the chi-squared test, *Environment and Planning A*, **15**: 483–500.

Fingleton, B. (1986) Analyzing cross-classified data with inherent spatial dependence, *Environment and Planning A*, **18**: 48–61.

Fuchs, C. and Kennett (1980) A test for detecting outlying cells in the multinomial distribution and two-way contingency tables, *Journal of the American Statistical Association*, **75**: 395–98.

Getis, A. and Ord, J.K. (1992) The analysis of spatial association by use of distance statistics, *Geographical Analysis*, **24**: 189–206.

Lawson, A. (2001) *Statistical Methods in Spatial Epidemiology*, Wiley, New York.

Lawson, A., Biggeri, A., Bohning, D., Lesaffre, E., Viel, J.-F. and Bertollini, R. (eds) (1999) *Disease Mapping and Risk Assessment for Public Health*, Wiley, New York.

Levine, N. and Canter, P. (1998) *CrimeStat*: a spatial statistical program for crime analysis: a status report, paper presented at the National Institute of Justice Cluster Conference of the Development of Spatial Analysis Tools, Washington, DC (http://www.icpsr.umich.edu/NACJD/gis_data.html).

Montgomery, D. (1996) *Introduction to Statistical Quality Control*, Wiley, New York.

Kelsall, J.E. and Diggle, P.J. (1995) Non-parametric estimation of spatial variation variation in relative risk, *Statistics in Medicine*, **14**: 2335–42.

Kulldorf, M. and Nagarwalla, N. (1995) Spatial disease clusters: detection and inference, *Statistics in Medicine*, **14**: 799–810.

O'Loughlin, J. and Anselin, L. (1996) Geo-economic competition and trade bloc formation: United States, German, and Japanese exports, 1968–1992, *Economic Geography*, **72**, 131–60.

Ord, J.K. and Getis, A. (1995) Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis*, **27**: 286–306.

Rogerson, P. (1997) Surveillance systems for monitoring the development of spatial patterns, *Statistics in Medicine*, **16**: 2081–93.

Rogerson, P. (1998) A spatial version of the chi-square goodness-of-fit test and its application to tests for spatial clustering, in Amrhein, C., Griffith, D. and Huriot, J.M. (eds) *Econometric Advances in Spatial Modelling and Methodology: Essays in Honor of Jean Paelinck*, Kluwer, Dordrecht, pp. 71–84.

Rogerson, P. (1999) The detection of clusters using a spatial version of the chi-square goodness-of-fit test, *Geographical Analysis*, **31**: 130–47.

Rogerson, P. (2001) A statistical method for the detection of geographic clustering, *Geographical Analysis*, **33**: 215–27.

Rogerson, P. (2002) The statistical significance of the maximum local statistic, in Getis, A. (ed.) *Advances in Spatial Econometrics*, Palgrave, London.

Rogerson, P. and Sun, Y. (2001) Spatial monitoring of geographic patterns: an application to crime analysis, *Computers, Environment, and Urban Systems*, **6**: 539–56.

Srivastava, M.S. and Worsley, K.J. (1986) Likelihood ratio tests for a change in the multivariate normal mean, *Journal of the American Statistical Association*, **81**: 199–204.

Tango, T. (1995) A class of tests for detecting 'general' and 'focused' clustering of rare diseases, *Statistics in Medicine*, **14**: 2323–34.

Worsley, K. (1996) The geometry of random images, *Chance*, **9**(1): 27–40.

PART 3
TRANSPORT AND LOCATION

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

10

Modelling and Assessment of Demand-Responsive Passenger Transport Services

Mark E.T. Horn

Abstract

An investigation of public transport options was carried out for the central Gold Coast region of Queensland, Australia, using the LITRES-2 public transport modelling system. Several road-based demand-responsive passenger transport services were modelled in combination with conventional buses, trains and taxis, under a range of assumptions about future demand. The demand-responsive options included multiple-hire taxis, 'roving buses' and a 'smart-shuttle' local service between railway stations and their immediate suburban localities. These were shown to be valuable and operationally credible as complements to the conventional timetabled modes. Besides substantive conclusions of this kind, the project raised some methodological issues concerned with project planning and execution, and the capabilities and usage of a modelling system such as LITRES-2.

10.1 Introduction

The responsibility for public transport in Australia lies with state rather than with local governments; nevertheless, many local councils take an active interest in the provision of such services, working in consultation with state transport departments. Such was the case in the project described in this chapter. An Integrated Regional Transport Plan was prepared by the Queensland Transport Department for the south-eastern section of Queensland during the mid-1990s. The Integrated Plan

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

included ambitious targets for increased usage of public transport services, and the achievement of those targets was one of the considerations leading to the initiation of a Public Transport Passenger Study by the Gold Coast City Council (GCCC) in 1997. The Passenger Transport Study included the modelling project described in this chapter, and was in turn one of the main inputs to the Gold Coast City Transport Plan, a joint undertaking of the GCCC and the Queensland Departments of Transport and Main Roads (GCCC, 1998).

The main task of the Gold Coast modelling project was to assess the viability of several proposed road-based demand-responsive transport modes, which were envisaged as supplements to anticipated bus, rail and taxi services in the near and medium-term future: that is (approximately) the period 2000–2010. The work was carried out by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), in collaboration with Booz Allen and Hamilton (BAH). BAH's main role was to supply estimates of key parameters concerned with current and forecast demand, service levels and fares. BAH subsequently also used the results of the project in formulating a wider range of recommendations for the enhancement of passenger services and infrastructure (BAH, 1998).

The main analytical tool used in the project was a specialised transport modelling system called LITRES-2, which embodies research carried out by the author and his CSIRO colleagues over several years (Horn, 2002; see also Smith, 1993; Rawling *et al.*, 1995). No proprietary GIS software was used: all data were stored in ordinary Unix files, and all data preparation was carried out using standard Unix tools, or using editing and display programs developed as adjuncts to LITRES-2. The supplementary software includes a graphical program for browsing spatial inputs (see Figures 10.5 and 10.6); there is also a graphical version of the modelling system itself, which has proved useful in demonstrating and validating the system's operation (see Figures 10.3 and 10.4). Besides the substantive issues under investigation, the project involved extension and refinement of the software, and an effort to assess the software's capabilities in practice.

Section 10.2 of this chapter surveys urban patterns in the Gold Coast, states the objectives of the project, and introduces the transport modes under investigation. Section 10.3 describes the LITRES-2 modelling system and the main categories of data involved. Section 10.4 outlines the scenarios tested in the project and the results obtained in each case. The chapter concludes in Section 10.5 with a summary of substantive conclusions from the project, a discussion of methodological issues, and an assessment of LITRES-2 in comparison with alternative modelling packages.

10.2 Project Objectives

The City of the Gold Coast comprises a stretch of Pacific seashore and estuarine hinterland commencing some 32 km south of the Brisbane central business district, and extending another 60 km south to Coolangatta on the New South Wales border. Figure 10.1 covers the more densely settled parts of the region, running 40 km from north to south between Oxenford and Coolangatta. During the past 50 years the region has evolved from a collection of small farming, fishing and holiday towns,

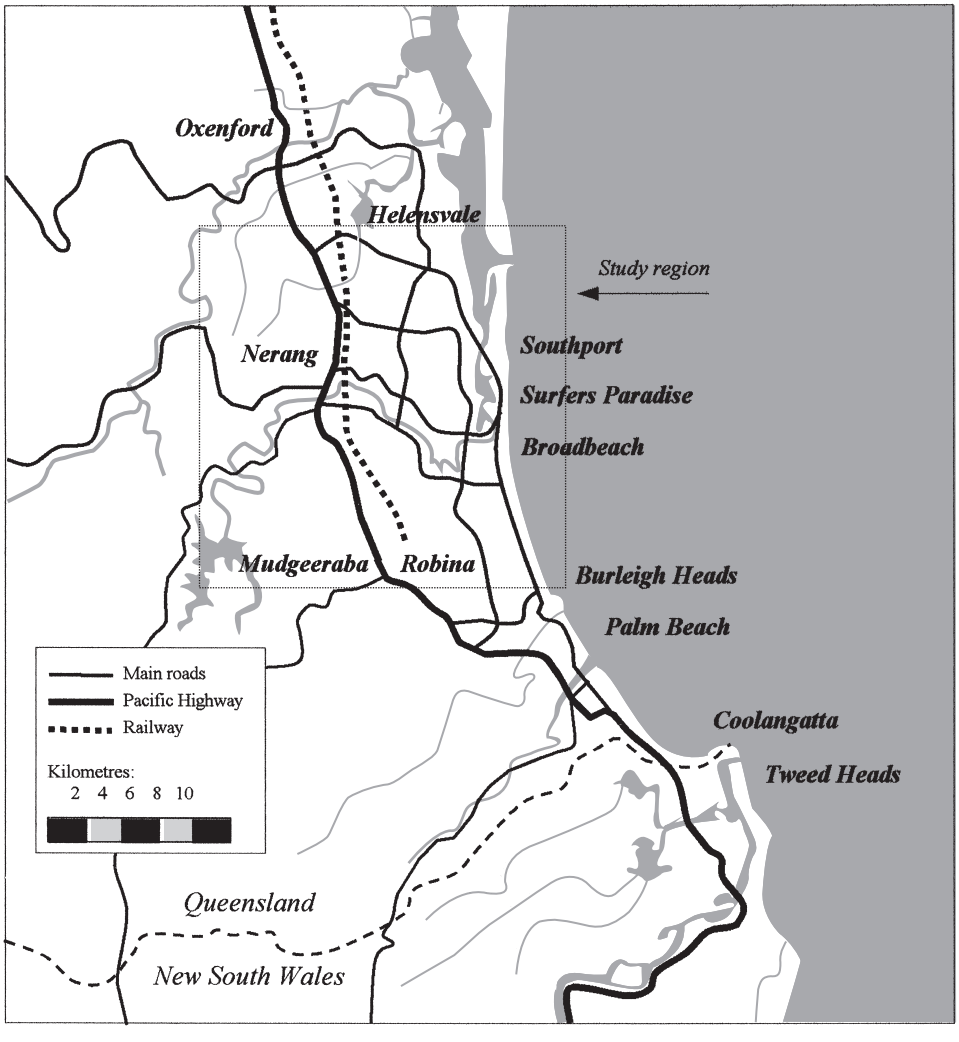


Figure 10.1 *Sketch map of the Gold Coast region*

into a leading tourist destination, favoured also as a residential location by many elderly people on account of its mild climate. Substantial numbers of residents commute to work in Brisbane, and the region now houses a diversified range of industries in its own right. It is one of the fastest-growing urban areas in Australia, with a permanent resident population already exceeding 425 000.

A chain of high-density urban development (e.g., apartment buildings, hotels and other tourism facilities) follows the main north–south coastal road, concentrating especially in the north around the regional centres of Southport and Surfers Paradise. West of the coast there are extensive low-density residential areas built on the banks of estuaries and canals. This waterfront zone is bounded on its western

edge by the Pacific Highway, which traverses regional centres at Helensvale, Nerang and Mudgeeraba, from which suburban developments extend farther westward into the foothills of the coastal ranges.

The Gold Coast City Council initiated the project with the general objectives of reducing growth in car traffic and congestion, improving accessibility within the region for residents and visitors, reducing dependence on the private car and the need for new roads and parking facilities, and improving environmental sustainability. These objectives were to be pursued by increasing the range of choice, convenience and integration of the public transport system, improving the accessibility and affordability of public transport and, in general, improving the balance between private and public modes of transport (BAH, 1998).

In preliminary discussions with the GCCC and BAH, it was decided that the analysis should concentrate on a study area measuring approximately 21 by 15 km. This covers the central part of the Gold Coast region, and includes all the main regional centres mentioned earlier. In some respects the region is well served by public transport: there are high-frequency bus services along the coastal strip, and a railway line has been built from Brisbane to Helensvale, with extensions to the south under way at the time of the project. Even so, current levels of passenger transport usage are very low (GCCC, 1998, Section 4.3), and the Gold Coast's generally dispersed settlement pattern constitutes a difficult environment in which to sustain geographically comprehensive bus or rail services.

The public transport modes covered by the project (see Table 10.1) included the existing bus network, the railway line with its projected extensions, and a variety of demand-responsive services. The attraction of the latter in the context of a region such as the Gold Coast is that they can be operated more flexibly than fixed-route services, and thus should be more effective in serving spatially dispersed urban areas. They may be envisaged as variants of conventional taxi services, designed to achieve economies of scale through the use of larger, shared vehicles. The demand-responsive modes are characterised by the use of recently developed technologies to handle communications between vehicles and a central dispatching facility, and to allow prospective travellers to make enquiries and bookings. Beyond that, their viability will often depend on the use of automated, centrally located scheduling software to obtain an effective and efficient deployment of vehicles.

Table 10.1 *Characteristics of transport modes*

Transport mode	Pickup and setdown	Advertised service	Routing
Walking	Stop to stop	–	Euclidean or Manhattan path
Bus, Rail	Stop to stop	Timetabled	Fixed
Taxi	Door to door	Free-range	Demand-responsive
TaxiMulti	Door to door	Free-range	Demand-responsive
RovingBus	Stop to stop	Free-range	Demand-responsive
SmartShuttle	Stop to stop	Timetabled, stops or zones	Demand-responsive

The main demand-responsive modes modelled in the Gold Coast project are outlined below.

- *Taxis*: single-hire, door-to-door taxi service. As with the other demand-responsive service modes, provision is made for advance bookings. The traditional 'step-in' style of taxi service was not modelled in the Gold Coast project.
- *TaxiMulti*: multiple-hire taxis guaranteeing service within prescribed travel-time standards.
- *Roving Bus*: multiple-hire services like TaxiMulti, but picking up and setting down passengers at designated stops.
- *SmartShuttle*: a bus service based on an advertised timetable but run on demand. That is, the actual route taken by a SmartShuttle bus needs to traverse only the locations for which demand has been registered, and the service can be omitted altogether on occasions when no demand is registered. The timetabled locations may be points (e.g., bus stops), or sets of points (in spatially defined zones), or a combination of the two.

10.3 Methodology

The LITRES-2 modelling system provides a fine-grained representation of demand and transport services, with particular focus on the operational characteristics of demand-responsive modes (Horn, 2002). The system's modelling architecture is outlined below, before proceeding to the data framework and modelling strategy used in the Gold Coast project.

10.3.1 Modelling Framework

LITRES-2 applies a microsimulation strategy to the representation of events and processes evolving over a given period of time (some alternative approaches are discussed at the end of this chapter). By contrast with more statistically based approaches, the parameters are all empirically based, and a formal calibration of the system is thus not required. A primary distinction between simulation and control modules (Figure 10.2) provides a clear conceptual basis for the system, and is consistent with the intended future usage of the control modules in a real-time context. In particular, the request-brokering and journey-planning modules embody algorithms that are applicable in a public transport information system, while the fleet-scheduling module comprises procedures designed to manage bookings and deployment for a fleet of demand-responsive vehicles in real time. The simulation modules embody representations of real-world processes, and are the source of the control systems' knowledge of critical real-world phenomena such as passenger requirements and vehicle movements. By contrast, the role of the control modules is to influence the simulated processes, this influence being exerted by means of messages (e.g., dispatch commands) sent to the simulation modules.

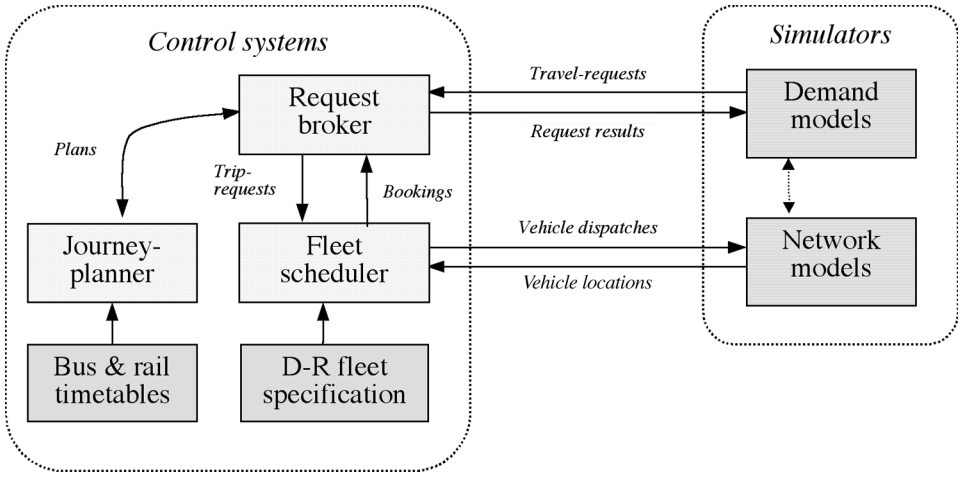


Figure 10.2 *LITRES-2 modelling framework. Reprinted from Horn, M., Multi-modal and demand-responsive passenger transport systems,* Transportation Research A, ***36****(2): 167–88. Copyright 2002, with permission from Elsevier Science*

The main simulation modules are demand simulators and a network simulator. The demand simulators convert predefined models of aggregate demand into a stream of travel-requests, which is sent to the control modules. The network simulator tracks the movements of demand-responsive vehicles through the road network, and interacts mainly with the fleet-scheduling module. Several additional simulation models (not shown in Figure 10.2) provide for the modelling of randomly occurring contingencies, such as vehicle breakdowns and delays, and passenger bookings and no-shows. The simulation models are executed in a temporal context provided by the C++Sim toolkit (Little and McCue, 1993), which ensures that timed processes evolve in a chronologically consistent time-sequence. Outside the simulation modules, the time-sequence is manifest as time-stamps on messages sent to the control modules.

Aggregate demand for passenger transport is represented as a collection of predefined demand models, each associated with a particular market segment, parameters for party size (i.e., the number of people travelling together in a given journey), and a given time period (e.g., 8:00 to 9:00 a.m.). Two types of demand model were used in the Gold Coast project. An origin–destination (OD) model is a matrix of which a given element specifies the total number of journeys from one zone to another during the period covered by the model. A source–sink (SS) model consists of an array of which a given element specifies the aggregate demand for travel from a given point to a given zone, or vice versa: such a model provides a straightforward way to represent demand associated with a major regional facility, such as a shopping complex or tourist venue, as an overlay on the more uniformly distributed patterns associated with an OD model.

The task of a demand simulator is to disaggregate a demand model in a randomised fashion, so as to generate a stream of travel-requests: each such request refers to a person (or party of people) belonging to a particular market segment, requesting travel between a given origin and a given destination point, within a given journey-envelope specifying an earliest departure and a latest arrival time. As a

typical example of the simulation techniques employed in LITRES-2, we shall consider the demand-simulation process for an element (i,j) of a given OD model. This element specifies the total number of journeys k_{ij} to be made from zone i to zone j during the period (*tlo*,*thi*) covered by the model.

In summary, there are k_{ij} travel-requests to be generated for element (i,j), consistent with the aggregate parameters mentioned above but differing from each other in a random fashion. To generate the starting-point for one such request, we refer to the bounding rectangle of zone i (i.e. the smallest rectangle enclosing all parts of the zone), with width and height w_i and h_i respectively. We use a random-number generator to obtain a pair of numbers (p,q), both in the range (0,1), and hence generate a point (x,y), with x located at $p \cdot w_i$ units from the left-hand edge and y at $q \cdot h_i$ units from the bottom of the bounding rectangle. We check that (x,y) lies within the boundaries of the zone and, if not, we obtain a new pair of random numbers and repeat the process. The same technique is used to obtain a destination point in zone j for the travel-request. The commencement-time *tstart* for the request is interpolated at random within the model period (*tlo*,*thi*); that is, given a random number r $(0 \leq r < 1)$, $tstart = tlo + r\ (thi - tlo)$. The end-time *tterm* is found by reference to walking-time and travel-delay parameters of the market segment s associated with the OD model: $tterm = tlo + wmax_s + ttmin_{ij} \cdot kmax_s$, where $wmax_s$ and $kmax_s$ are obtained by uniform random interpolation in ranges specified for market segment s, and $ttmin_{ij}$ is the shortest possible travel time by road from i to j.

10.3.2 Journey Planning

The travel-requests generated by the demand modules are passed, one by one, to a request-broker module, whose task is to plan an optimal journey that satisfies the requirements of each incoming request. A journey consists of a sequence of one or more legs; a journey-leg carried by a vehicular mode is called a trip. A single-leg journey may be walked all the way or carried by a door-to-door vehicular service (i.e., taxi or TaxiMulti). A multi-leg journey involves a sequence such as 'Walk → Bus → Train → SmartShuttle → Walk'; or more precisely, 'Leave origin-point at 10:10 a.m., walk to bus stop, catch 290A bus at 10:15, alight at Nerang at 10:25, take the train from Nerang to Robina at 10:30, alight at Robina at 10:40, take SmartShuttle at 10:45, alight at local stop at 10:50, walk to destination, arriving at 10:56 a.m.'. An example is shown in Figure 10.3.

The planning of multiple-leg journeys is carried out by a journey-planning module, an optimisation procedure embodying a branch and bound algorithm. The journey-planning process is subject to behavioural constraints specified for each market segment, such as the maximum distance that the traveller is prepared to walk at a single stretch. The availability of service is determined by timetable data, or in the case of demand-responsive modes, by reference to the fleet-scheduling module. The objective in planning a journey is to minimise the generalised cost of travel, defined as the sum of any fares, together with the imputed cost of time for the person or persons making the journey. This imputed cost is estimated using valuation factors for the various time components of the journey, namely waiting at the origin, waiting

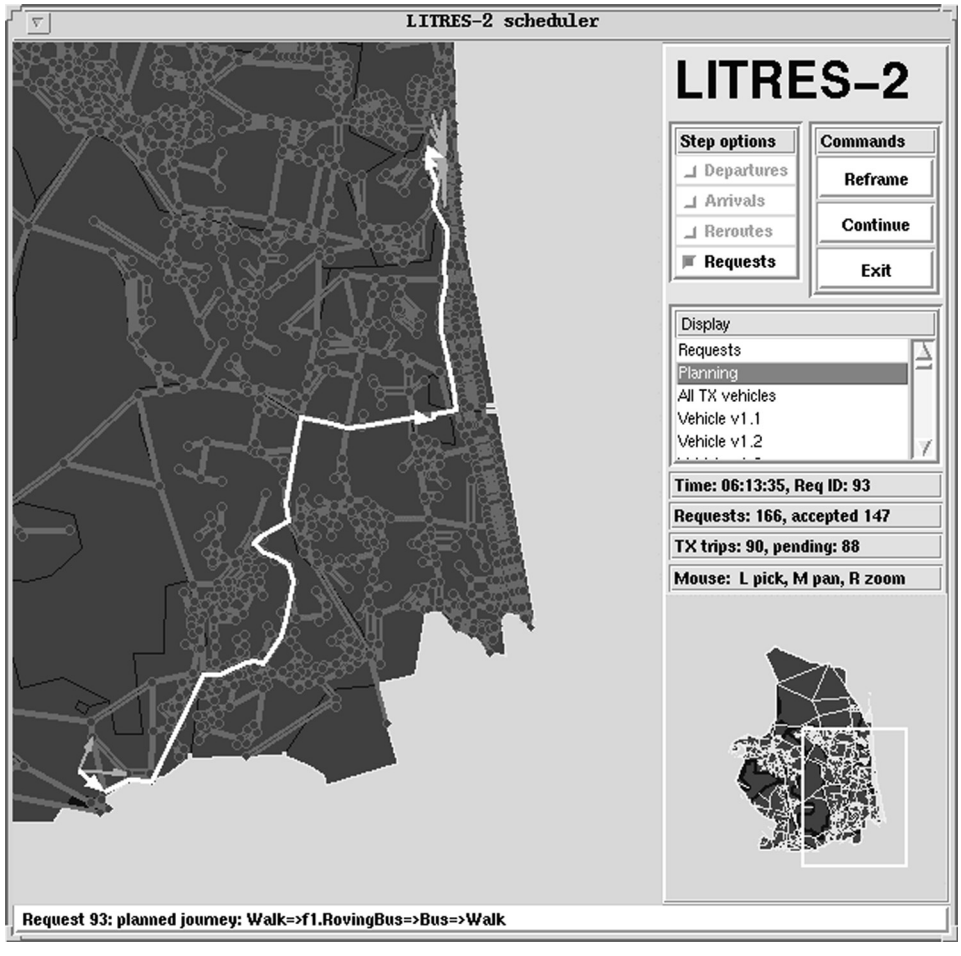


Figure 10.3 *Planning a journey*

at interchanges, walking and travelling. Travel times are estimated as follows. For a walked leg, the time is simply the length of the leg divided by the walking speed associated with the market segment associated with the traveller; while for a leg carried by a timetabled mode, the travel time is obtained from the timetable. For a leg carried by a non-timetabled mode (i.e., RovingBus, TaxiMulti or taxi), travel time is calculated by reference to the mode's service-quality standards (see Section 10.3.4 below).

10.3.3 Fleet Scheduling

The LITRES-2 fleet scheduler manages the deployment of a fleet of demand-responsive vehicles, by planning each vehicle's itinerary in response to trip-requests received over time from the request-broker. Figure 10.4 shows one such itinerary,

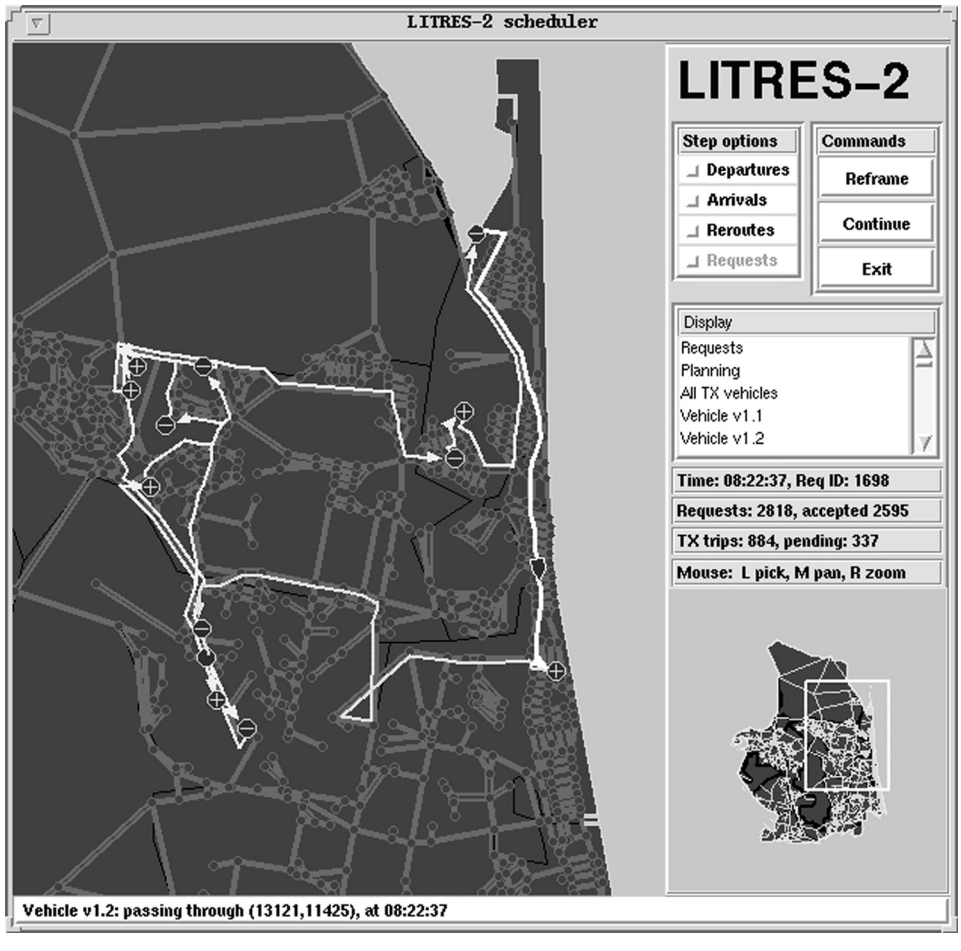


Figure 10.4 *Itinerary of a demand-responsive vehicle*

including the vehicle's current location and its planned pickup and setdown points ('+' and '–' respectively). A trip-request refers to a particular transport mode, and has explicit time-windows specifying early and late limits for departure from the start and arrival at the end of the trip: these limits are defined by the journey-planning module by reference to the service-quality standards pertaining to the mode. The overall objective when choosing an implementation for a trip is to produce favourable operational outcomes for the fleet operator (e.g. by minimising fuel costs). The finding of shortest-time paths between each successive pair of stops in a vehicle's itinerary is a scheduling sub-task, and is discussed below in connection with the road network model. A distinction is made between the routes and timings planned by the scheduling module, and on the other hand the actual trajectory of a vehicle as it moves through the road network. In particular, the fleet scheduler sends dispatch

commands (e.g., 'Cab 99, proceed at 10:10 to location (x,y)') to the network simulation module, which simulates the movement of each vehicle through the road network, and reports back to the scheduler when a vehicle arrives at a pickup or setdown point or is affected by an incident such as a breakdown.

10.3.4 Inputs and Outputs

The Gold Coast project was planned as a series of scenarios, each specified as a set of inputs to a run of LITRES-2. In particular, a Base scenario provided a check of LITRES-2's predictions against current conditions, while other scenarios were constructed as variants of the Base scenario, so as to investigate the performance of various postulated sets of passenger service options over different levels of demand. A LITRES-2 scenario is defined by input data of the following kinds.

- A model of the road network is used as a basis for estimating travel times, and for routing vehicles in the fleet-scheduling and network simulation modules. The network model comprises a set of nodes and a set of directed links connecting the nodes, together with link speeds and distances. This is a format used in many commercially available GIS, but the intensive use of network calculations has required special attention to the algorithms and storage strategies used for network calculations. For the Gold Coast project, a single speed was defined for each link and path durations were precalculated: subsequent extensions allow for fluctuation in link speeds over the course of a day. The techniques used to find shortest-time paths and their timings are variants of Dijkstra's label-setting algorithm, with allowance for off-network travel between the (Euclidean) starting point of a path and a nearby network node, and vice versa at the end of the path.
- A division of the study region into polygonally bounded small-area units called zones provides spatial reference for demand models. Zonal decompositions are used also as bases for defining fare structures and zone-based SmartShuttle services.
- Behavioural characteristics of travellers are defined in market segments. A market segment defines travel parameters for a particular set of travellers, such as elderly people. These parameters include a set of acceptable modes, valuations of time pertaining to generalised costs, a walking speed and a maximum walking distance. In addition, there are parameters for acceptable travel duration, and for the 'notice period' between the time when travel is requested and when it is to take place.
- As indicated earlier, aggregate demand is represented in the form of origin–destination (OD) and source–sink (SS) models, each of which is associated with a market segment, party-size parameters, and a time-range.
- Specifications of timetabled services include locations of the points served (e.g., bus stops and railways stations), and service routes (sequences of points). The specifications for each route include a transport mode (e.g., bus), a fare structure, and a timetable.
- Demand-responsive services are defined by specifying the vehicles in a demand-responsive fleet or fleets. Each vehicle has a passenger-carrying capacity; a list of

the modes offered; and an operating shift, specified as the period when the vehicle is in service, with a corresponding pair of terminal points. Associated with each demand-responsive mode are parameters specifying a fare structure, pickup and setdown times, and service quality. The service-quality standards comprise a waiting-time limit and a travel factor, the latter specifying an upper limit on the ratio of actual to minimum possible travel time: this ratio is one for single-hire taxi, and more than one in the case of the multiple-hire modes.

The main outputs from LITRES-2 are outlined below. The volume of these outputs makes them somewhat unwieldy in their raw form, and their interpretation in practice requires considerable care in the selection and examination of a smaller number of key indicators (for this purpose the outputs are now provided in a coded format, to facilitate transfer to spreadsheets).

- *Travel statistics* include the total numbers of passengers carried and journey-legs completed, with average travel times, distances and costs, for the whole simulation period and for nominated time-intervals. These statistics are broken down by leg-sequence, by market segment, and by transport mode.
- *Scheduling statistics* include the total numbers of passengers carried, with mean trip time and distance per passenger, for the demand responsive fleet(s), for each mode, and for each vehicle. As with the travel statistics, statistics are aggregated for the whole simulation period and also over nominated time-intervals. They include measures of efficiency (e.g., the ratio of travel time to the least possible travel time, relevant in the case of multiple hire modes) and of vehicle occupancy (i.e., the average number of passengers on board at any given time).
- *Route-usage statistics* include the total numbers of passengers carried on each timetabled mode, and the totals for each route and each service run along each route, with average trip times and occupancy ratios. Again, statistics are aggregated for the whole simulation period and for nominated time-intervals.

10.4 The Case Study

The Base and other scenarios were planned in advance, and then adjusted and extended in the course of the project. The most demanding and time-consuming task was to fine-tune the considerable range of data inputs in the Base scenario, so as to obtain a credible match between model outputs and known conditions (Horn *et al.*, 1998). The simulations themselves incurred little real delay: the run time for each scenario was generally between 20 to 40 minutes, on a Sun UltraSparc workstation.

10.4.1 The Base Scenario

An eclectic approach was required in the preparation of the 'Base scenario', involving the use of sources whose applicability to the project was often somewhat

indirect. Outlined below are the main sources for the data categories described earlier, and the adjustments applied to them in preparing the Base scenario.

The Road Network

The GCCC provided a road network data set covering nearly all roads in the region, and comprising 4166 nodes and 9472 links. This was somewhat excessive in terms of the computational resources available; furthermore, it included a great deal of geometric detail which would be irrelevant to the analysis carried out by LITRES-2. The network was therefore reduced in size through the application of a filtering program, followed by manual correction of a few remaining anomalies. The filtering program was developed especially for this purpose. It removed 'superfluous' dead-end links (i.e., each link to a node with no other links inward or outward), and 'superfluous' intermediate nodes (i.e., each node with exactly one inward and one outward link). Aggregate distances were preserved in this process (e.g., where a chain of links representing a curved section of road was reduced to a single link). The reduced network comprises 1686 nodes and 4515 links, and is illustrated in Figure 10.5.

Link speeds were defined initially by applying speed limits corresponding with the road classifications provided with the network. These speeds were then systematically scaled down in the light of a preliminary simulation of taxi system performance, and ground speed checks carried out by BAH. For off-network vehicular travel, a uniform speed of 25 km per hour was assumed.

Zones

A zonal division of the study region was needed to provide spatial reference for bus fares and demand models. The zones were defined as those used by the Surfside Bus Company in determining fares, with unpopulated areas removed. These excisions – mainly in the estuarine parts of the region – were made in order to minimise distortions in the representation of demand, which is assumed within LITRES-2 to be distributed uniformly within each zone.

Market Segments

BAH identified six market segments as a basis for modelling demand. The market segments may be characterised informally as follows:

- *Mainstream*: regular users of public transport when service is readily available.
- *Captive old and captive young*: people who have no real alternative to public transport.
- *Tourists*: tourists who choose to use public transport.
- *Car-oriented*: people who currently travel by car but may decide to switch to public transport if service of sufficiently high quality is available.
- *Rail commuters*: people who live within the corridor served by the railway and its planned extensions, and are likely to commute by train to workplaces in Brisbane.

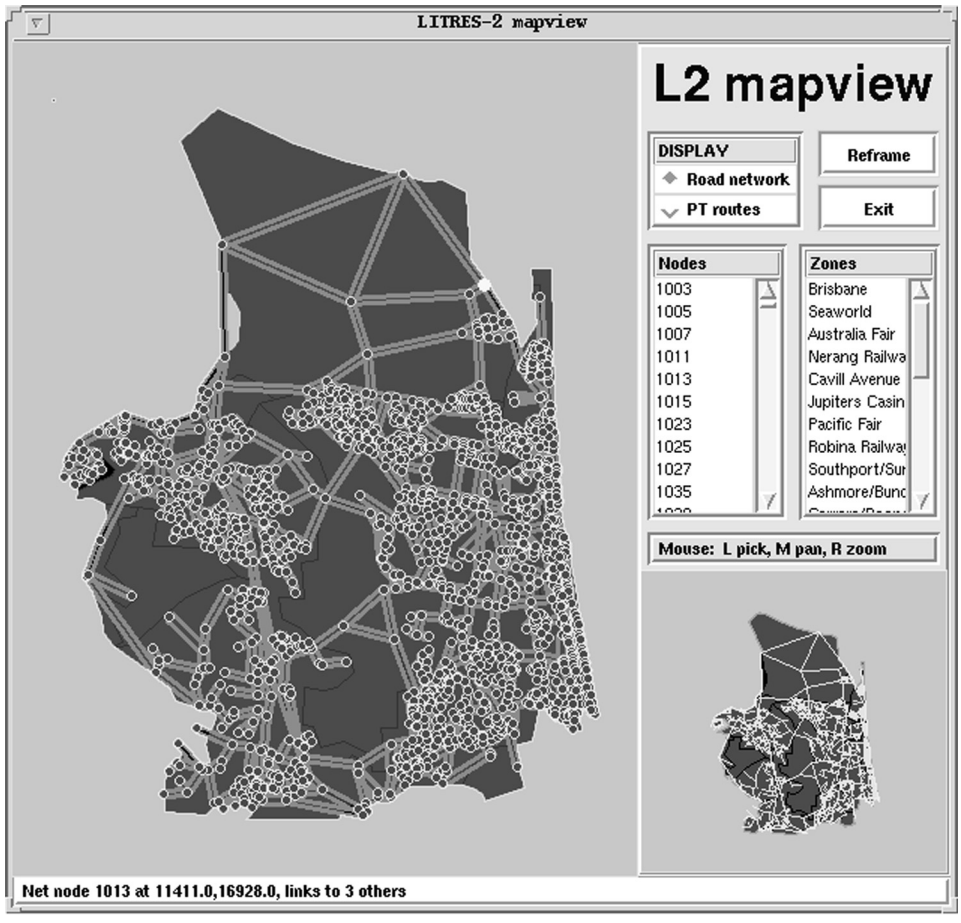


Figure 10.5 *The road network*

Demand Models

Estimates of current demand were obtained from a variety of sources (BAH, 1998). Demand specifically associated with regional attractors (e.g., shopping malls, casinos and funfairs) was modelled in the form of source–sink models, using estimates from economic studies of these facilities. Other demand was modelled in several series of origin–destination matrices, specified (as for the source–sink models) at hourly intervals over a 24-hour period (notionally a weekday). These OD series were compiled by GCCC staff from the Surfside Bus Company's fare-collection data, Queensland Rail passenger data, and estimates of private car journeys from an earlier study in that region. The Base demand amounted to 21 148 travel-requests (31 722 travellers) during a 24-hour period. In temporal terms the patterns of demand were generally similar to those seen in other metropolitan areas, but with smaller peaks around the start and end of the conventional working day, and with a continuation

of activity into the evening (or the early morning, in some of the entertainment districts).

Timetabled Transport Services

The geographic locations of bus stops and railway stations were digitised, and a small proportion of them were designated as interchange points for transfer on multi-leg journeys (this restriction proved to be unnecessary, since subsequent testing yielded very similar results with transfer allowed at every bus stop). The published timetables were transcribed and then elaborated, as follows. First, each route was defined as a bi-directional pair (e.g. an advertised north–south route 290 yielded two routes, 290N and 290S); second, variant routes implied by the timetables (e.g., with skip-stop service during peak periods) were generated explicitly; third, arrival times at stops without explicitly advertised times were interpolated in proportion to distances along the routes; and fourth, services on the projected extension of the railway line were extrapolated from existing timetables. Figure 10.6 shows a snapshot of the graphical browser, with a list of the routes connecting a designated pair of bus stops shown in the right-hand panel, and with one of those routes highlighted on the main map display window.

Demand-Responsive Services

About 220 taxis currently operate in the wider Gold Coast region, under the aegis of Regent Taxis Limited. Some of the drivers specialise in traditional 'step-in' service from cab-ranks, and some – with larger vehicles – specialise in multiple-hire work from specially designated cab-ranks; most drivers, however, rely mainly on Regent's central dispatching centre for most of their work (Horn *et al.*, 1999). The simulation involved a taxi fleet about half the size of the actual Regent fleet, reflecting the approximate proportion of taxi activity encompassed by the study region. As with the timetabled modes, service details and fares for taxis were matched to existing conditions as closely as possible. The new demand-responsive services (TaxiMulti, RovingBus and SmartShuttle) were conceived as intermediate in terms of service quality and fares between taxis and buses. The fare structure in each case was defined as a base fare and a distance-based component, with group discounts for parties of two or more people.

10.4.2 Future Transport Scenarios

Low, medium and high demand scenarios were constructed by scaling up demand from the Base scenario to represent (notionally) the range of conditions possible in the study region within the next decade. Different scaling factors were applied to the various demand components, yielding estimates of 42 291, 53 997 and 65 799 passengers respectively for the three scenarios, compared with 31 722 in the Base scenario. To avoid excessive optimism regarding the prospects for new transport services, the

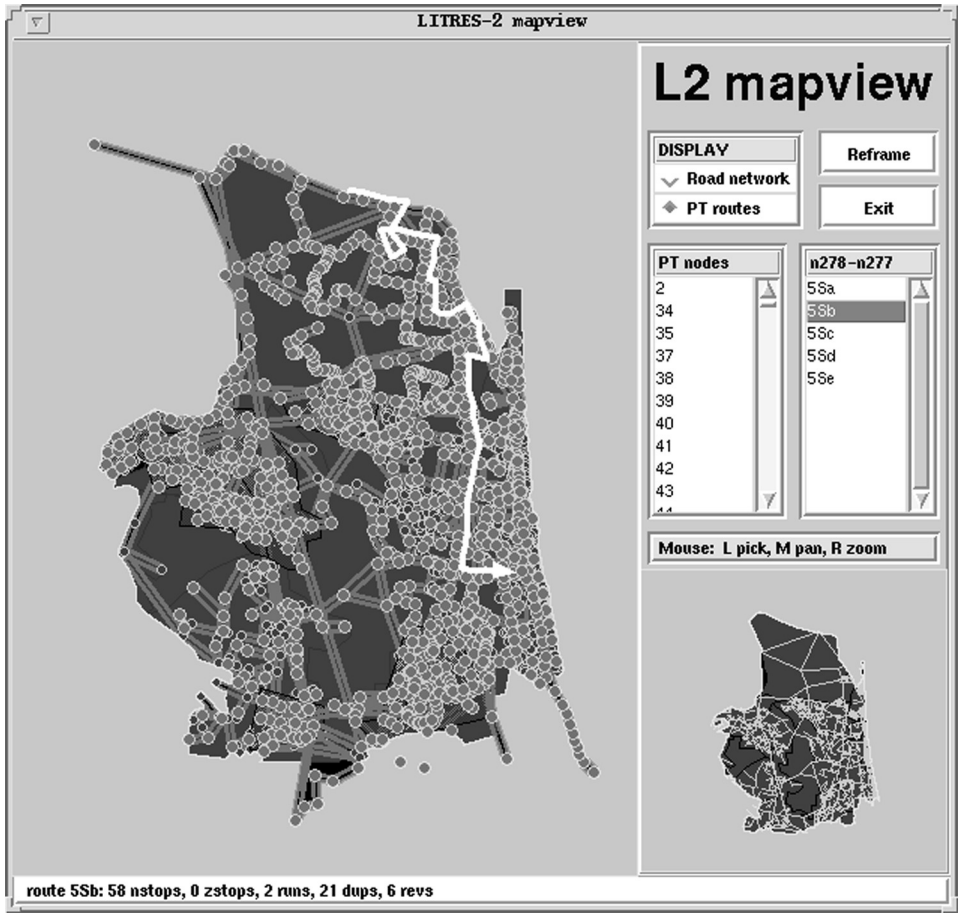


Figure 10.6 *Bus and rail routes*

analysis focused mainly on the *low* demand scenario. A more extensive study might well encompass a wider range of possibilities in this respect, for example with geographical differentiation in the rates of growth in demand.

Fixed-Route Services

The assumptions in the future scenarios regarding rail, bus and taxi services were the same as those in the Base scenario. As might be expected, the simulations indicated that bus usage would increase substantially under increasing demand. In fact, there is probably scope for augmentation of bus services and routes in some areas: an investigation of such possibilities was beyond the scope of the project, but – with the Base scenario already available – would require little additional effort.

Multiple-Hire Services

Demand-responsive services were modelled in each scenario as a single fleet of vehicles, with each vehicle dedicated to a single mode. The low, medium and high demand scenarios included 40, 60 and 80 mini-buses respectively, divided equally between TaxiMulti and RovingBus service. The simulations showed these vehicles carrying several passengers for much of the time, although they were rarely close to their maximum carrying capacity (Figure 10.7). High occupancy ratios are obviously desirable from a fleet operator's point of view, but saturation of fleet capacity prevents additional passengers from using the fleet; thus if the fleet were larger, it might be chosen by passengers who currently use other modes. Consequently, the simulated patronage of multiple-hire services in the main scenarios does not fully reflect the

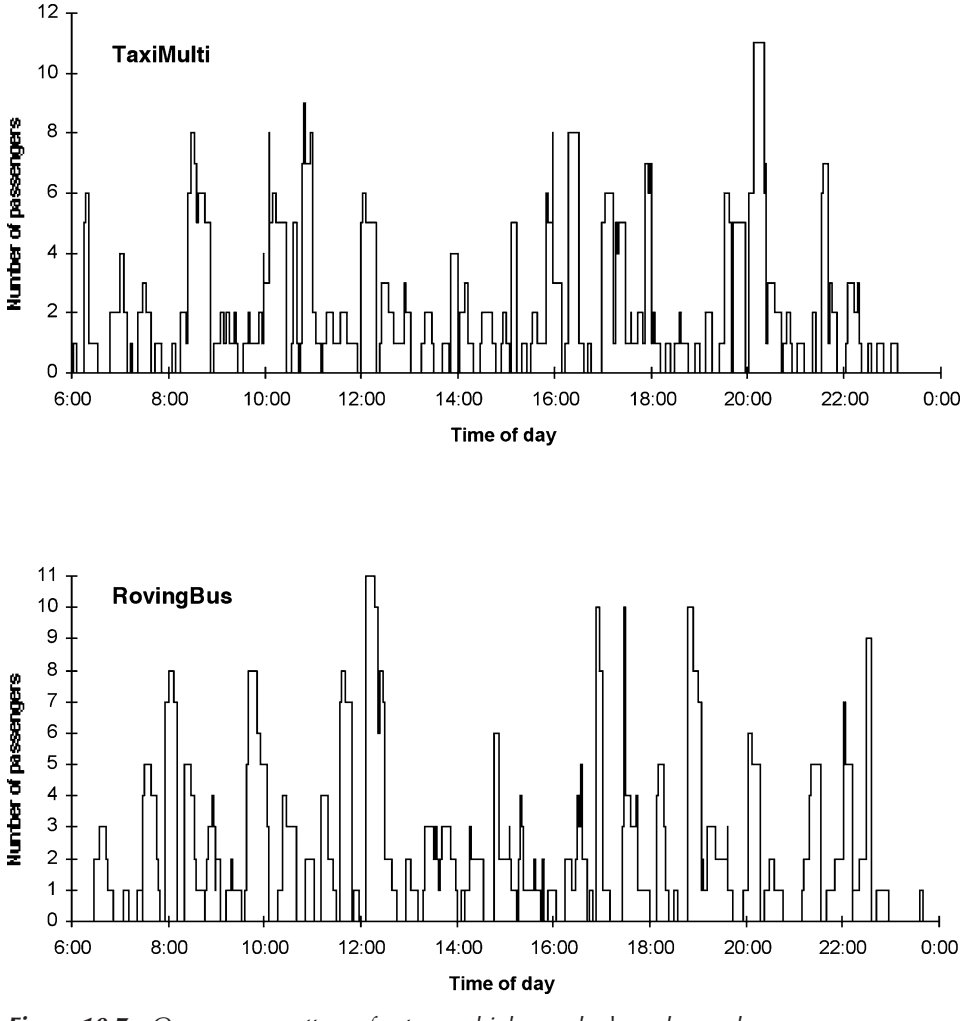


Figure 10.7 *Occupancy patterns for two vehicles under* low *demand*

attractiveness of these services to the travelling public, a deduction that was borne out by supplementary simulations with larger fleet sizes. More generally, the 'carrying capacity' of a fleet is a somewhat ambiguous concept, since 'heavily loaded' or 'saturated' conditions are characterised by a gradual decline – rather than a complete disappearance – in the additional trips that can be accepted under marginal increases in demand: the limit, where no further passengers can be accepted, is arbitrary, fluctuating over time and sensitive to arbitrary variations in current fleet deployment. A reasonably precise estimate of fleet capacity may be obtained, however, by allowing a progressive stretching of service standards under saturated conditions (Horn *et al.*, 1999; Horn, 2002).

Results from the *low* demand scenario (Figure 10.8) indicate that the TaxiMulti and RovingBus services should generate gross earnings per vehicle greater than those obtainable from single-hire taxis, and carry roughly twice as many passengers, provided that fleet sizes are matched to demand as discussed above. These and other results also suggest that trips on multiple-hire modes are on average comparatively long. The trip-makers highlighted here presumably are those who wish to travel between extremities of the region, and are deterred from using alternative services by their unavailability or excessive slowness; thus the multiple-hire modes may be particularly useful in spatially dispersed conditions.

Analyses of usage for RovingBus and TaxiMulti suggested how these modes might be managed in practice. For example, given the temporal patterns illustrated in Figure

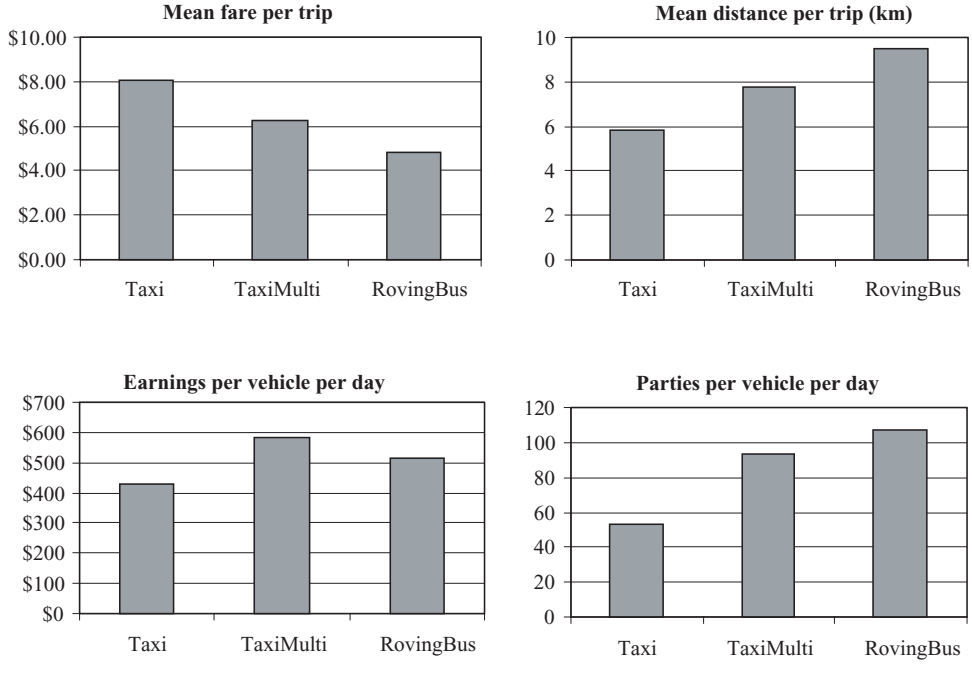


Figure 10.8 *Total trips on demand-responsive modes under* low *demand*

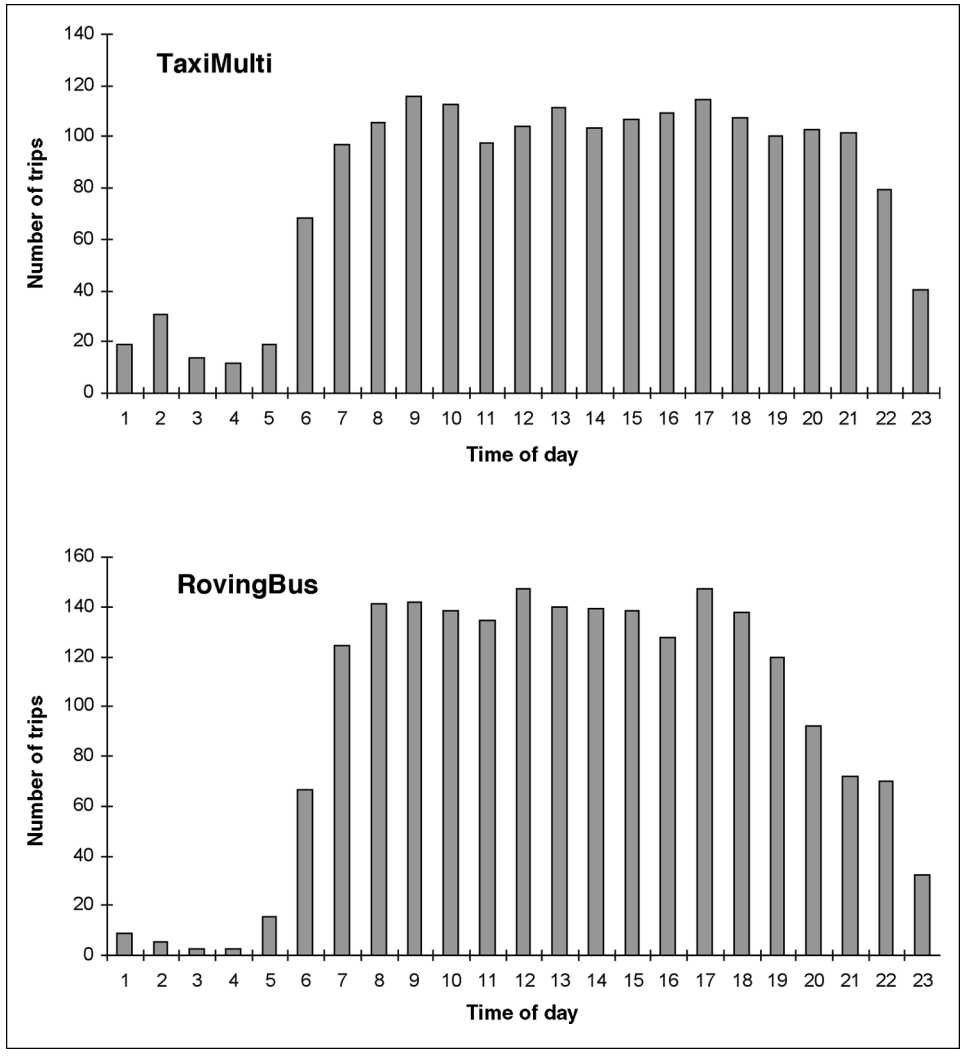


Figure 10.9 *Trips by TaxiMulti and RovingBus services under* low *demand*

10.9, an adequate service might be maintained by replacing RovingBus with TaxiMulti late at night, the latter perhaps run with a fleet reduced to one-third its daytime size. It may be observed also that although each vehicle was assumed to be able to accommodate 11 passengers in the low, medium and high demand scenarios, the actual occupancies rarely reached this level (Figure 10.7). A separate series of tests indicated that with capacity reduced to seven passengers per vehicle, the reductions in fleet performance and fleet-wide carrying capacity would be fairly small (of course this ignores less tangible considerations such as the sense of crowding that passengers would experience under pervasively nearly full conditions).

SmartShuttle

As indicated earlier, a SmartShuttle service is advertised in the form of a timetable, but the actual routes taken by vehicles are defined on demand. A 'feeder' role was envisaged for SmartShuttle in the Gold Coast, with vehicles plying between local railway stations and their residential catchment areas. The following account concerns a series of tests focused on the new Nerang railway station and its catchment area, which was divided into five SmartShuttle zones. One SmartShuttle vehicle was provided per zone, and timetables were designed in coordination with the railway timetable so that (notionally) each vehicle could tour its zone between visits of the mainline train, generally at 30-minute intervals. For comparison, an approximately equivalent set of fixed-route bus services was designed for the same area.

Two series of simulations were run to test these arrangements, one with SmartShuttle feeders and one with buses, the level of service in each case being held constant while the 'rail commuter' component of demand was scaled up incrementally from the Base scenario. These simulations showed that the SmartShuttle vehicles could serve nearly all prospective passengers under Base scenario assumptions while satisfying the journey-planning conditions discussed in Section 10.3 above; by contrast, fixed-route buses would leave about 8% of travellers beyond walking distance. With multipliers of 100–500% applied to base commuter demand, the proportion of requests that could be feasibly satisfied by SmartShuttle would decline steadily from 100% to 85%; the simulations suggested that the corresponding proportion for buses would be approximately constant across the same range of demand. On closer examination, however, it became apparent that the latter figure significantly overstated the performance of buses, which LITRES-2 currently models in less detail than demand-responsive vehicles; in particular, no allowance is made for passenger-capacity constraints or for delays incurred when picking up and setting down passengers. Thus on local routes and under moderate demand, small buses running in SmartShuttle mode would almost certainly be more convenient for travellers than the same vehicles run in conventional fixed-route mode.

Mode Usage by Market Segment

Figure 10.10 shows aggregate mode choice results for each market segment in the *low* demand scenario. In summary, *car-converts* are relatively heavy taxi users because they place high value on their own time. *Tourists* are heavy bus users, their activities being concentrated in the coastal strip where the bus service is frequent and convenient. The *captive old* are willing to walk only short distances, and so are the heaviest users of TaxiMulti.

10.5 Conclusions

10.5.1 Effectiveness of Transport Services

The effectiveness of the new demand-responsive service modes as demonstrated in the project has led to a recommendation that they be implemented on a pilot basis

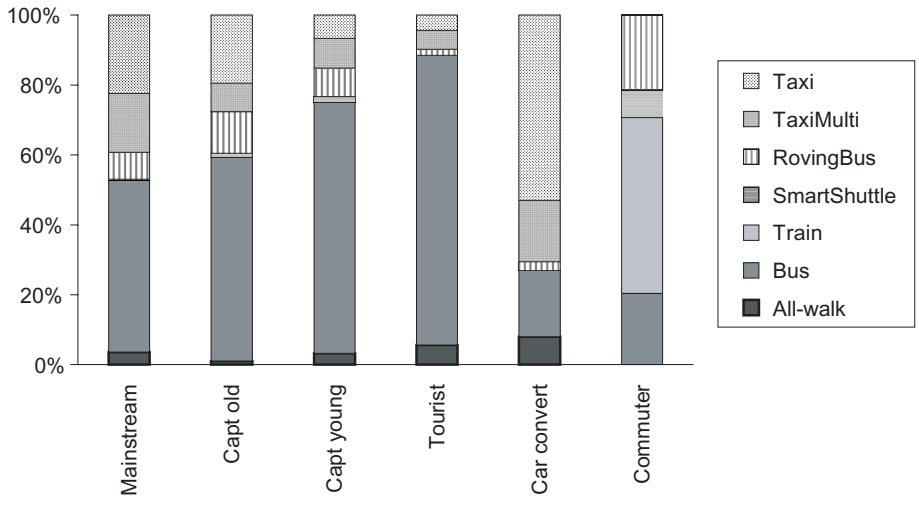


Figure 10.10 *Mode usage by market segment under* low *demand*

(GCCC, 1998, Section 5). The recommendation has not so far been implemented, presumably due to reluctance on the part of the Queensland Government to commit the substantial resources that would be required. In general, however, the project confirmed the benefits of a coordinated, multi-modal array of transport services. Those benefits were apparent in the way that the various modes complemented each other with respect to service quality, geographic coverage and market segments: even RovingBus and taxi, which are closely related in logistical terms, were mutually competitive only to a marginal extent.

Multiple hiring of vehicles in demand-responsive fleets was shown to be effective in terms of service quality, fleet economics and market share. It appears that TaxiMulti and RovingBus may best cater to longer journey-legs for which bus services are either unavailable or too slow to meet travellers' requirements, while taxis are more suited to shorter legs. SmartShuttle was investigated only to a limited extent, but as a feeder to railway stations and other regional interchanges, this mode appears to have advantages over a conventional scheduled bus service.

A critical requirement for TaxiMulti and RovingBus is the achievement of reasonably high occupancies in order to obtain substantial economies of scale over conventional taxi services. Occupancy rates in turn are directly sensitive to the travel factor, which defines the maximum allowable deviation from direct routing (see Section 10.3.4 above, and Rawling *et al.*, 1995). Increasing the travel factor above the value of 130% assumed in the Gold Coast project would provide broader time-windows and hence greater flexibility for scheduling, making it easier to achieve high vehicle occupancies. This would be a desirable outcome, but a reduction in fares might then be needed to compensate travellers for longer travelling times.

Taxi services were not a major concern in the Gold Coast project reported here, but were investigated more thoroughly in a related study that concentrated on the Regent taxi fleet over its full geographical range (Horn *et al.*, 1999). That study was

concerned mainly with operational issues such as the balance between phone bookings and traditional 'step-in' services at cab-ranks, the impact of advance notice on carrying capacity and waiting times, and temporal staggering of shift changeovers. The availability of historical booking records made data preparation much easier than in the wider-ranging project reported here; the methodology was also more straightforward, since the focus was on individual initiatives and changes, rather than broader combinations of assumptions and proposals like those discussed elsewhere in this chapter.

10.5.2 Methodological Considerations

Data Preparation and Validation

For CSIRO and its collaborators the most difficult challenges posed by the Gold Coast project were those arising in the preparation of a credible Base scenario, especially with respect to data items – market-segment parameters, demand models, and road speeds – defining behavioural characteristics of transport system users. Here there was no alternative but to make the most of whatever data sources were available, often in an indirect way or by reference to prior experience in other locations. The empirical foundation of the project cannot therefore be regarded as definitive, and the conclusions are best seen as indicative rather than strictly predictive. Even so, detailed examination of Base scenario outcomes, and sensitivity analyses of key variables (Horn *et al*., 1998), showed a satisfying conformity with known aspects of transport system operation.

Apart from these substantive concerns, the effort involved in data preparation constituted a large proportion of that invested in the project as a whole, and it seems clear that an effort of this kind might be difficult to justify in a once-off consultancy exercise. Conversely, however, the project suggests a framework within which public authorities might accumulate information on transport demand and system usage, the long-term benefits in this respect lying in the ease with which new proposals can be investigated once a suitable database is established.

Investigative Strategy

Some other methodological issues were concerned with the relations between modelling – the central task of the project – and other phases of the planning process. In a naive view, design or planning may be regarded as proceeding in a linear sequence from the enunciation of a problem (e.g., as a set of objectives and conditions), to the formulation of a proposal (e.g., by proposing new transport services), to assessing the implications of that proposal (e.g., by means of a LITRES-2 simulation), and then to evaluating those implications (e.g., by interpreting simulation outcomes). However, in any substantial planning exercise it can be expected that there will be backtrackings or feedback cycles to complicate a simple sequence like this, and it must be admitted that the extent of these complications was underestimated when the Gold Coast project was planned. The lesson here is that modelling should not be

treated as an isolated task: although the broad outlines of transport options could be sketched in advance, the development of those options as credible proposals was intertwined with the modelling process and with the interpretation of its outputs.

For example, the main simulations of multiple-hire demand-responsive services involved the testing of a small number of proposals, each comprising specifications of fleet size, the carrying capacities of vehicles, the fares to be charged, service-quality standards, and so on. In fact these proposals themselves emerged from a preliminary series of simulation runs, in which a range of plausible values was explored for each of the main parameters. There was a similarly close relationship between design and simulation in the tests of SmartShuttle feeder services, which involved the additional design task of formulating 'benchmark' bus services for comparative purposes.

A further point concerns the criteria applied in assessing simulation outcomes. Besides convenience, accessibility, and so on, it emerged during the course of the project that the clients believed that new passenger services should be able to operate with little or no public subsidy. The desire for operationally attractive performance entailed considerable effort in tuning the various proposals, and led to the exploration of demand-responsive services on a smaller scale than might have been undertaken otherwise. In retrospect, the rationale for this approach seems questionable, given the emphasis elsewhere on the benefits of consistency and integration in the provision of transport services. Operational viability is certainly a significant planning consideration, but to require that demand-responsive services should be fully self-supporting could prejudice their availability at the lower end of the socioeconomic spectrum. It would be unfortunate indeed if an effort to modernise transport services were to involve the relegation of traditional fixed-route modes to a second-class or 'safety-net' role.

Modelling Tools

Experience in the Gold Coast project shows that effective usage of LITRES-2 requires a considerable level of interpretive intelligence, as indeed with any modelling system of comparable scope. This is illustrated by the counter-intuitive results encountered in the SmartShuttle tests, which highlighted an anomaly in the level of detail applied to fixed-schedule as opposed to demand-responsive vehicles (see Section 10.4 above). The anomaly can be removed by applying the same operational constraints and detailed simulation to buses as to demand-responsive vehicles, and further attention might be given also to the system's representation of passenger behaviour (Horn, 2000). In addition, an interface between LITRES-2 and a GIS might facilitate the preparation of input data and the analysis of modelling outputs.

There appears to be no immediate prospect of overcoming LITRES-2's reliance on external estimates of passenger transport demand (Horn, 2002). In this respect it is interesting to contrast the microsimulation approach adopted in LITRES-2 (see also Tong and Wong, 1999), with network-equilibrium models such as EMME/2, in which probabilistic methods are used to estimate the distribution of *all* journeys, including those made by private car (Spiess and Florian, 1989). From a planning perspective, the main differences between the two approaches have to do with granularity and scope: where the equilibrium models are coarse-grained and comprehensive,

microsimulation is typically fine-grained and specialised (consider for example the practical difficulties attending a microsimulation of all road traffic).

It is interesting also to consider the use of GIS buffering capabilities in estimating 'catchment areas' and latent demand in the vicinity of transport routes, bus stops, railways stations and so on. This is clearly pertinent to analyses concerned with accessibility, especially where bus or rail routes are to be planned in conformity with official standards (e.g. maximum walking distances to bus stops). But accessibility by itself must be reckoned a weak indicator of transport system quality, since it is concerned merely with identifying the transport services that are physically available to people in a given area, rather than the services that those people are likely to use in practice. This criticism applies even to some interesting recent research involving the generalisation of temporal and operational aspects of accessibility (O'Sullivan *et al.*, 2000).

The foregoing discussion has highlighted some strengths and weaknesses of LITRES-2 in the context of projects such as the one described in this chapter. In summary, LITRES-2 emerges as a very suitable means for assessing passenger transport operations and performance, with unique capabilities with respect to demand-responsive services. Considerable effort must be devoted to data collection and validation; nevertheless, once that investment is made, it is possible to explore a wide range of passenger transport planning options, to a level of detail that is not available elsewhere.

Acknowledgements

Acknowledgement is due to the author's collaborators in the work described in this chapter: Ken Deutscher, Melissa Doueihi, Garry Glazebrook, Mac Hulbert, Bella Robinson and John Smith.

This chapter includes some material adapted from the report cited as Horn, Robinson and Smith (1998), and included here by permission of the Gold Coast City Council. Enquiries with respect to the contents of the report or their dissemination should be directed to the Council's Transportation Planning Branch.

This chapter includes some other material that has been adapted from an earlier article written by the author. The material is reprinted from Horn, M.E.T. (2002) Multi-modal and demand-responsive passenger transport systems: a modelling framework with embedded control systems, *Transportation Research A*, **36**(2): 167–88. Copyright with permission from Elsevier Science.

References

BAH (Booz Allen and Hamilton) (1998) *Gold Coast Public Transport Study: Proposed Strategy and Implementation Plan*, Report submitted to Gold Coast City Council.

GCCC (Gold Coast City Council) (1998) *Gold Coast City Transport Plan.*

Horn, M.E.T. (2000) A framework for modelling passenger movement in urban public transport systems, in *Proceedings, 22nd Conference of Australian Institutes of Transport Research*, Canberra.

Horn, M.E.T. (2002) Multi-modal and demand-responsive passenger transport systems: a modelling framework with embedded control systems, *Transportation Research A*, **36**(2): 167–88.

Horn, M.E.T., Robinson, B. and Smith, J.L. (1998) *LITRES2 – Public Transport Simulation Research Project*, Report submitted to Gold Coast City Council, CSIRO Australia.

Horn, M.E.T., Smith, J.L. and Robinson, B. (1999) Taxi fleet performance under flexible operating conditions and with improved scheduling procedures, in *Proceedings, Fourth International Conference of ITS Australia*, Adelaide.

Little, M.C. and McCue, D. L. (1993) Construction and use of a simulation package in C++, *Computer Science Technical Report #437*, University of Newcastle upon Tyne, July 1993 (also in *C Users' Journal*, **12**(3), March 1994).

O'Sullivan, D., Morrison, A. and Shearer, J. (2000) Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach, *International Journal of Geographical Information Science*, **14**(1): 85–104.

Rawling, M., Smith, J.L. and Davidson, I. (1995) Modelling of personal public transport in a multi-mode service environment, in *Proceedings International Conference on Application of New Technology to Transport Systems*, ITS Australia, Melbourne, v1, pp. 167–78.

Smith, J.L. (1993) An activity oriented urban passenger transport model, Proceedings, *Australian Transport Research Forum*, **18**(2): 993–1006.

Spiess, H. and Florian, M. (1989) Optimal strategies: a new network assignment model for transit networks, *Transportation Research B*, **23**(2): 83–102.

Tong, C.O. and Wong, S.C. (1999) A stochastic transit assignment model using a dynamic schedule-based network, *Transportation Research B*, **33**(2): 107–21.

11

The South and West Yorkshire Strategic Land-use/ Transportation Model

David Simmonds and Andy Skinner

Abstract

A major multi-modal transport study is currently being carried out in South and West Yorkshire. This chapter outlines the land-use/transport/economic modelling system that has been developed for the study. This system is being used at the strategic level to forecast future land uses and travel demands, and to examine the impact of alternative transport and development plans.

11.1 Introduction and Client Needs

The South and West Yorkshire Multi-Modal Study (SWYMMS) is one of a series of studies which were proposed in *A New Deal for Trunk Roads in England*, published by the Department for Transport in July 1998 (Department for Transport, 1998). The overall aims of the study are to make recommendations for:

- an integrated and sustainable plan for the strategic road, rail and water networks in the study area; and
- a plan of specific interventions to address the most urgent key strategic problems in the study area through to 2021.

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

The study is required to address the Government's aims for transport, which refer to economy, accessibility, safety, integration and environment. It is also required to address a set of study-specific objectives, which are

- to reduce congestion on the motorways and A1;
- to re-establish the primary role of the trunk network for strategic traffic flows;
- to facilitate sustainable economic regeneration of depressed areas, especially the Objective 1 area of South Yorkshire and the Objective 2 area of West Yorkshire; and
- to sustain economic growth in other parts of the study area.

At the time of writing,[1] the study, which is being carried out for the Government Office for Yorkshire and the Humber, is still ongoing. Further details of the team, the work and the findings of the study can be found at www.swymms.org.

At an early stage in the development of the study, it was decided that it would be necessary to use both a strategic model and a detailed model in the analysis of transport options. The need to use both reflects the fact that with present technology (in terms of modelling methods and computing power) transport planners are forced to choose between:

- models which allow detailed representation of the origins and destinations of travel (zone system) and of the transport network, but represent only a few of the choices which households and businesses make in response to changing travel conditions; and
- models which represent more of the choices and linkages between choices, and hence are more appropriate to forecasting reactions to change, but which incorporate much less details about the spatial system and the transport network.

The study-specific requirements imply that it is necessary to study not only the impacts of transport change on transport use, but also the interactions between transport change and local economic performance. It therefore needs to examine how demand for transport is affected by land-use changes and how changes in transport may affect land use – with particular reference to impacts on economic performance. The consideration of transport must take account of both planned interventions and other changes such as worsening congestion.

In order to examine these changes and their linkages systematically, particularly to compare the consequences of different strategies, it was recognised that a strategic land-use/transport model was required, i.e. a model considering a wide range of responses to change with relatively little spatial detail, and that this should be used as the main tool in developing the proposed strategy. At the same time, it was also necessary to use more conventional, detailed models to examine alternative changes to the transport infrastructure in the process of developing the recommended *plan* of interventions within the preferred strategy. To ensure maximum consistency

[1] July 2002.

between the two levels, the outputs from the strategic model are used as a way of generating the levels of travel demand which are assumed in the detailed model.

In the remainder of this chapter, we explain the components and structure of the strategic transport and land-use/economic models in Sections 11.3 and 11.4 respectively and briefly recognise the interaction between them in Section 11.5. Section 11.6 demonstrates how the model is used for testing particular policies. We begin, however, with an explanation of the overall design of the model.

11.2 Model Design and Application

11.2.1 Scope of the Model

The strategic model draws heavily on the study team's previous experience with a number of similar models developed since 1995 (see, for example, Copley *et al.*, 2000; Simmonds and Skinner, 2001). The approach adopted is that of a linked model system which brings together distinct transport and land-use/economic models. The basis of the approach is that the land-use/economic model generates the activities (and in some cases the interactions between activities) which generate the demand for transport; and that the transport model represents the interaction between demand and supply in transport, and generates information about the ease or difficulty of travel which influences subsequent land-use and economic change.

A significant advantage of this approach is that because the land-use/economic and transport models remain distinct entities, it is very easy to run each one on its own. This helps in the development stage, where to a considerable extent the two models can be developed in parallel. It is also useful in the application stage: the transport model can be used on its own to test the transport impacts of different alternatives under common land-use forecasts, and the land-use/economic model can be run to test (for example) the land-use impacts of different economic scenarios assuming no change in transport conditions.

The transport model was implemented using MVA's START package (Roberts and Simmonds, 1997). The land-use/economic model was implemented using DSC's DELTA package (see Simmonds and Still, 1999; Simmonds, 2001). The MapInfo GIS package is used extensively in mapping the inputs and outputs of the model. In the following sections we describe the design of the South and West Yorkshire Strategic Model (SWYSM) model and illustrate its application.

11.2.2 Scope of the Model – Space and Time

The focus of the study is on the 'motorway box' formed by the M62, A1(M), the M18 and the M1. It was recognised that the forecasting process needs to consider not only what happens within this box, but also to consider in equal detail what happens in a significant area around it, and in less detail what happens further away. Accordingly the model covers the whole of South and West Yorkshire plus a ring of adjoining areas such as Skipton and Harrogate; this is the fully modelled area (Figure

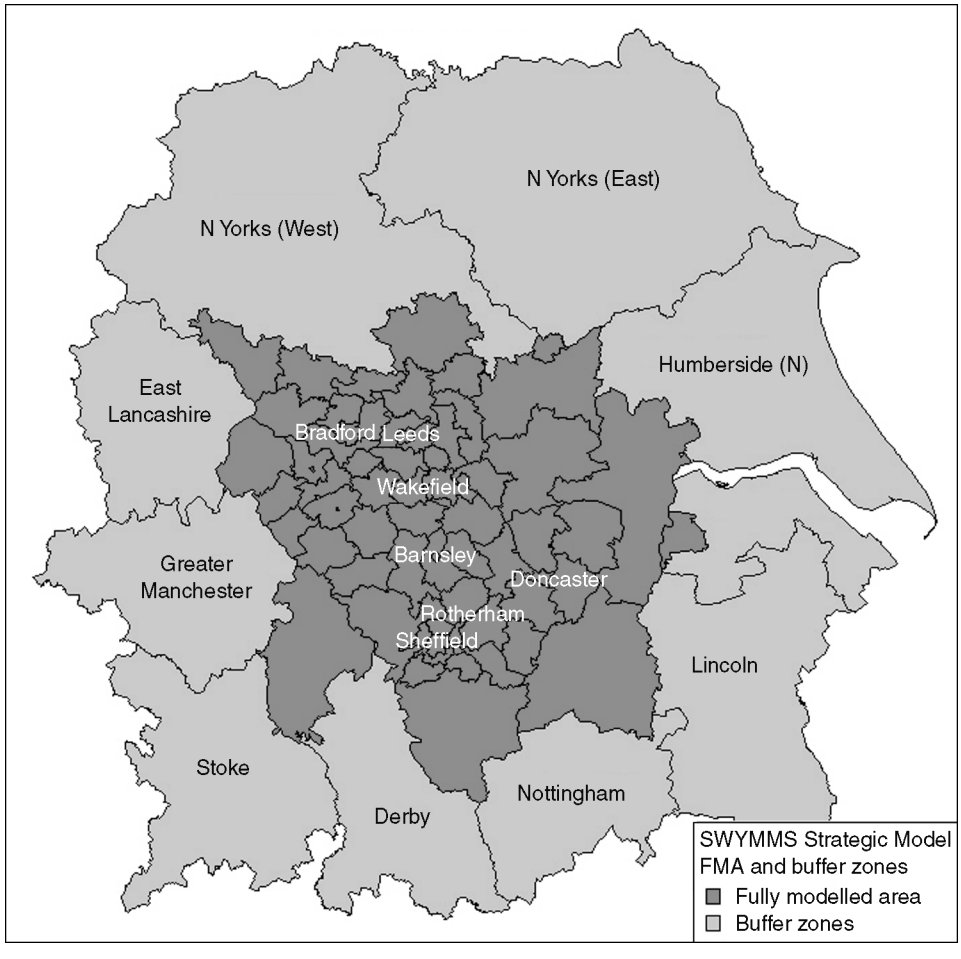


Figure 11.1 *SYMMS strategic model: fully modelled area and buffer area*

11.1). Around this, a wider ring of adjoining areas (the 'buffer area', extending westwards to Manchester and eastwards to Hull) is covered in less detail, with larger zones and fewer model processes working. Beyond these is a set of 'external zones' used to represent movement to and from the rest of Britain.

Within the fully modelled area, the zones were defined initially as 1991 wards, or groups of wards, within district boundaries; these conditions facilitate the use of data both from the 1991 Census of Population and from other sources. The grouping of wards into zones within districts was intended to distinguish the main urban centre of each district (in the event, some subdivision of wards was required), and then as far as possible to create zones of reasonable size and shape in relation to the transport networks. The process was constrained by the number of zones which is compatible with practical computing times and memory requirements, especially in the transport model where both time and memory increase as the square of the number

of zones. In total, there are 92 zones in SWYSM – in contrast with the 570 zones in the more detailed highway and public transport models. The model was set up so that the base year is 2000 and the forecasting process works forward to 2020; this involved an element of forecasting to create the 2000 base.

11.3 Transport Modelling: The START Package

Unlike many conventional transport models which have to be calibrated to reproduce the base year situation, START is incremental in approach: matrices of trips are estimated outside the model to describe the base year pattern of travel, making use of all available data. The model then modifies the pattern of travel in response to factors 'external' to transport – mainly changes in land-using activities such as the number, mixture and car-ownership of residents, the number and type of jobs, etc. and to changes within the transport system itself. The following sections summarise the preparation of the data for the base situation, the external forecasting model and the transport model proper.

11.3.1 Transport Model Base Data

Data on base year travel by motorised modes were obtained from the SWYMMS detailed highway and public transport models, which were in turn developed mainly from the extensive surveys carried out during 2000. Walking and cycling, which are not represented in the more detailed models, were estimated separately. Most of the START supply representation was obtained by aggregating data already assembled in the detailed models. The primary sources of data for the detailed models were surveys of road traffic and public transport trip making undertaken in spring 2000. Approximately 140 road traffic interview sites were employed, covering all entry points to the motorways and the A1 within the study area. Public transport passenger surveys were undertaken in the centres of the 10 largest study area towns, with details of non-central movements provided via operator ticketing systems. Cartographic network data for the highway system were created using GIS techniques, and public transport networks were coded from timetables.

11.3.2 External Forecasting Model

The external forecasting model (EFM) predicts changed travel patterns due to zonal changes in land use, including population, employment and car ownership. The model requires considerable detail about present and future land-use by zone (age groups, household types, etc.), since such factors are known to influence the amount and the type of travel. In SWYSM, all of these are forecast by the land-use model (see below).

The EFM assumes that future residents of a given age, employment status, car-ownership level, etc. will, if other factors remain equal, wish to make the same trips

as comparable present residents. It works by factoring the different trip matrices in proportion to the numbers of residents by type and zone, i.e., in proportion to changes in the land uses that produce and attract trips. The output of the EFM is therefore an artificial travel pattern – what would happen if the land-use changes occurred but the time and money costs of travel were unchanged? The impact of time and money changes is considered in the transport model proper.

11.3.3 The Transport Model

The transport model (Figure 11.2) contains the demand model, which adjusts travel in response to changes in transport conditions; and the supply model, which adjusts transport conditions in response to changes in the pattern of travel. These are

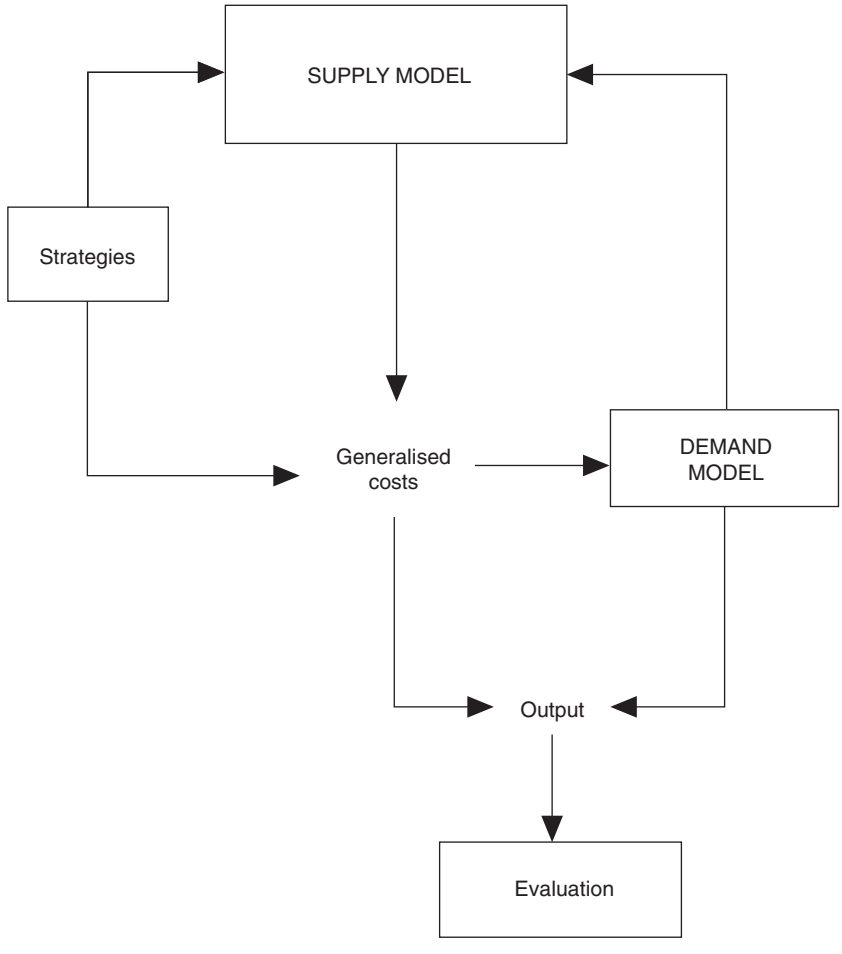


Figure 11.2 *The structure of the START transport model*

automatically run in turn until the travel pattern and transport conditions are consistent with each other. The key variable linking them is generalised cost, which is used to describe every journey within the modelled area by a single value bringing together money, time, inconvenience or discomfort, etc.

The *Demand* component of the transport model deals with all of the key choices available to travellers. These are:

- frequency – how often to travel;
- destination – where to travel to;
- mode – car, bus, rail, tram, walk, cycle, taxi;
- time of day – when to travel;
- route – highway or public transport route alternatives; and
- parking – type, location and duration.

These choices are organised into a hierarchy. The most sensitive choices are dealt with first, at the bottom of the structure, and the least sensitive choices at the top of the structure. The relative sensitivities, and hence the hierarchy, differ between trip purposes.

Changes in the pattern of freight movement are forecast by the land-use/economic model. Only road freight is currently represented in SWYSM; this is equivalent to assuming that the road share of freight movement for each commodity and each origin-destination (OD) pair will remain constant. Growth in road goods vehicle movements is forecast by DELTA as a result of changing land use and economic conditions. The prospects for modal shift in freight are being considered within SWYMMS, but outside SWYSM.

The *Supply* component of the transport model ensures that the level of service of the transport system, as defined by monetary costs and the various components of travel time, is compatible with the flows on the various transport networks. The highway supply definition is based upon the following key features:

- links that represent the speed/flow relationships for flow groups in each zone (typically defined in terms of inbound and outbound radial, clockwise and anticlockwise orbital);
- motorway and other 'limited access' links defined as specific roads; and
- a set of fixed car routes for each OD pair, including intra-zonal movements, defined in terms of distance travelled upon each link (with allocation of trip proportions to routes carried out via the demand model).

A similar arrangement is applied to public transport, with the addition of specific rail network links. The public transport definition allows for mixed mode trips. Crowding relationships are included so that generalised costs increase as demand approaches capacity. Fares and frequencies can be made responsive to changes in patronage. A number of indicators of the environmental impact of transport are calculated within the transport model, and some of these influence subsequent land-use change. In addition, for SWYMMS, more extensive analysis of environmental impacts was carried out by Environmental Resources Management as an input to the strategy appraisal process.

11.4 Land-Use/Economic Modelling: The DELTA Package

The DELTA land-use/economic modelling package has been developed by David Simmonds Consultancy since the mid-1990s. It was designed as a practical tool to represent urban and regional change and to interact with an appropriate transport model. The approach draws upon previous modelling experience and upon the wide range of relevant research carried out in geography, urban economics, etc. (for a fuller description of the model, see Simmonds, 2001).

Since land uses take a long time to respond to transport changes, the land-use model needs to represent change over time, in contrast with the transport modelling which describes transport supply and demand at particular points in time. Another requirement is to recognise that different processes operate at the urban and the regional levels: for example, different factors affect the total economic activity *in* an area and the location of employment *within* it. The model accordingly contains urban processes which change in or between the 92 zones, and regional processes in which the units are (approximate) travel-to-work areas – typically districts, in SWYSM.

The urban processes represent both changes in buildings and changes in activities. The processes of physical change are:

- development: the amount of floorspace by zone and type (residential, retail, office, industrial); this is driven by the economic scenario and the modelled property market, and controlled by inputs measuring what is allowed by the planning system; and
- the housing quality model, which models the way in which an area may decline to slum status, or be revived from slum to high-quality.

The urban activity sub-models are:

- the transition model, which represents household (and hence population) changes in terms of movements through a simplified lifecycle;
- the car-ownership model, which predicts the proportions of households by type and zone owning 0, 1 or 2+ cars, mainly in response to increasing incomes;
- the location model, which locates or relocates a proportion of households and of employment in each year, and also models the property market within which location occurs; and
- the employment status model, which updates the work status of residents and the commuting pattern in response to the spatial changes in households and employment.

At the regional level, there are three models, all of activities. These are:

- the migration model moves households between areas;
- the investment model allocates investment to areas (taking account of changes in accessibility and property costs); and
- the production/trade model (a spatial input–output model) estimates production by sector and area and the patterns of trade between areas.

The processes are considered in a fixed sequence within each one-year step, as shown in Figure 11.3. However, there are also numerous time lags between the different

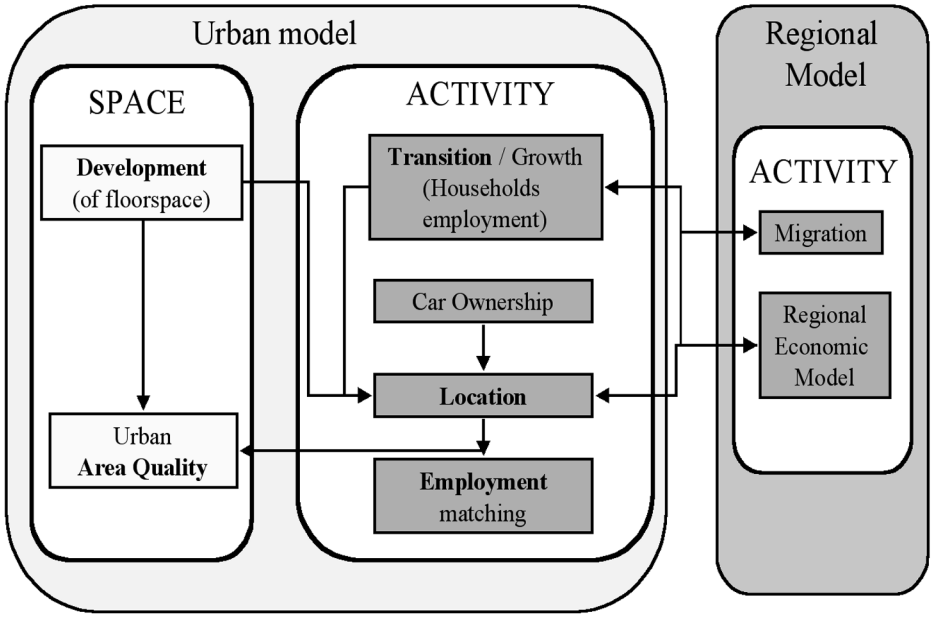


Figure 11.3 *The DELTA sub-models: land-use/economic model structure*

processes which are equally or more important to the overall performance of the model.

11.5 Land-Use/Transport Interaction

The model operation over time is a sequence of five one-year steps through DELTA, followed by a run of START to represent the resulting state of the transport system (Figure 11.4). Land-use changes have immediate impacts on transport, through the EFM described earlier. At the area level, changes in the costs of trade (goods and passengers) affect the location of production and of new investments. At the urban level, generalised costs are used to calculate various accessibility measures which affect the location of households and businesses within each area. In addition, the environmental impacts of transport influence the locational preferences of households.

11.6 Use of the Strategic Model

11.6.1 The Strategy Testing Process

As with most models, the strategy testing process using SWYSM is comparative. A strategy option is examined by appropriate changes to the model inputs; running the

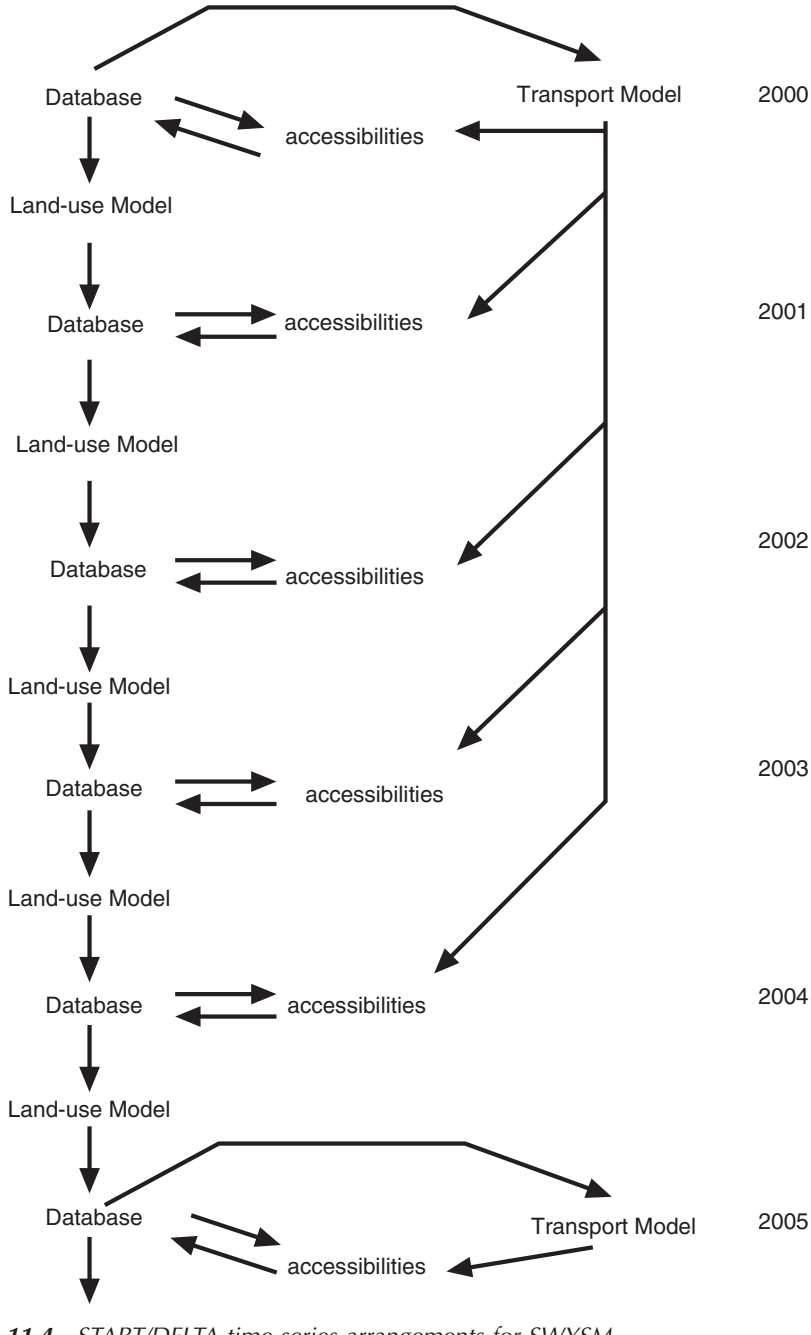


Figure 11.4 *START/DELTA time series arrangements for SWYSM*

model over time; and comparing the resulting outputs with those from a previously run reference case. This process applies whether transport options, land-use options or combinations of the two are to be tested.

Transport policy changes are introduced into START either as direct adjustments to costs or as changes in the characteristics of the transport supply. They influence land-use and economic change through the accessibility and environmental values that are passed from START to DELTA. Land-use policy changes are introduced into DELTA via a special policy file, which is used to specify changes in the amount of land available for each type of development in each zone in each year. Policies may not have immediate effects: for example, an additional allocation of land may not be used immediately or at all.

An important characteristic of using a land-use model is that land-use policies are tested in the same way as transport policies. This is not always the case. Whilst it is conventional for transport policies to be tested by forecasting their effect taking account of user preferences and interactions, land-use planning often assumes simple linkages between policy changes and outcomes. Formal land-use modelling allows a more rigorous approach to be taken.

11.6.2 Strategy Testing in SWYMM

The strategic model was used extensively in the Strategy Development phase of SWYMMS. This was divided into three main themes: travel reduction options; economic development options; and combined transport/economic development options. For the first theme the transport model was frequently used on a stand-alone basis, as direct transport effects were dominant. The full model was used for investigation of options for the second and third themes. The following paragraphs illustrate the results from just one part of this exercise.

11.6.3 Example Strategy Results

One of the tests carried out under the economic development options theme (identified as test ED3 within SWYMMS) was to look at the combined impact of two packages of road building and widening schemes. The packages were a group of schemes aimed at economic regeneration, and a group of schemes aimed at the relief of traffic congestion in areas where economic regeneration is a priority concern.

These schemes were concentrated in the South Yorkshire area between Sheffield and Doncaster, plus improvement to the A1 north of Doncaster; they are illustrated in Figure 11.5. They were introduced into the transport supply in 2005, and the model system was run to 2020. The following discussion of their impacts is based on comparison with the reference case; except for the road schemes, there were no other differences in transport or land-use policy between the ED3 test and the reference case.

The changes in the road networks are introduced into the model through the supply side of the transport model in 2005 (and in all the subsequent transport model years). These changes have the effect of allowing faster travel by road for some parts of

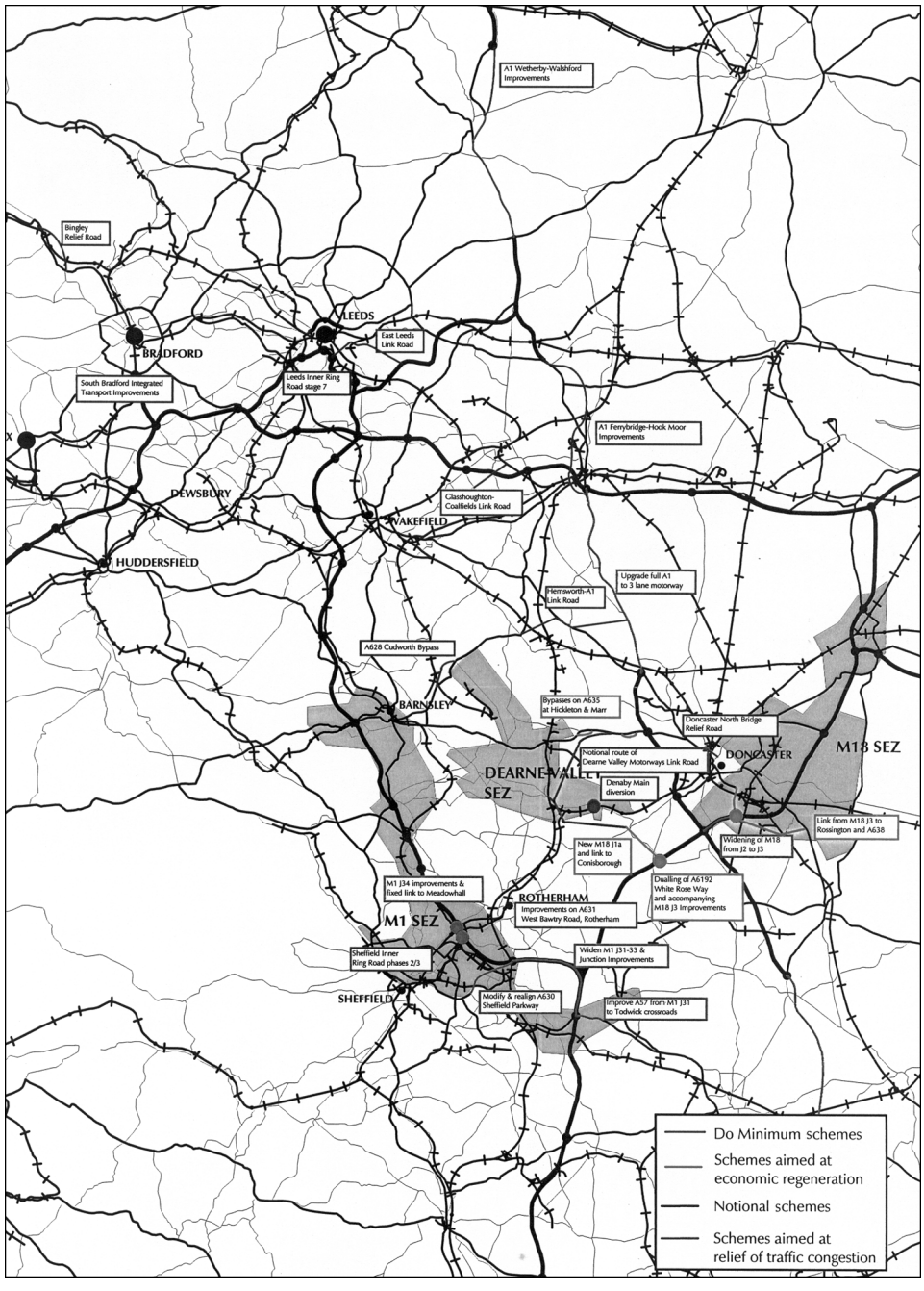


Figure 11.5 *The location of ED3 road building/improvement schemes*

journeys in South Yorkshire (including some journeys between origins and destinations outside South Yorkshire). The journey time reductions come about partly from higher road design speeds, and partly from increases in capacity which reduce the effects of congestion.

As a result of the journey time reductions, the generalised cost matrices output from the transport model show very slight reductions for affected zone pairs relative to the equivalent matrices for the reference case. As a result, there are some very slight changes in trip patterns, including a slight increase in commuting between Doncaster and Sheffield/Rotherham.

All of these effects are very slight – mostly measured in fractions of 1% change from the reference case. The average cost of delivering goods from the Doncaster area to the Sheffield/Rotherham area, for example, falls by slightly less than 1%. These effects are apparent in the data passed to the land-use model at the end of the 2005 run of the transport model (that is, they are slightly different from the equivalent data in the reference case). The differences from the reference case will affect the land-use/economic model in a number of ways. At the area level:

- the changes in cost or generalised cost of travel will have some impact on the pattern of trade, and this will have an immediate impact on the level of production of each sector in the affected areas; and
- the same changes will also affect the accessibility of each area and hence its attractiveness to investment: the change in the level of investment (or reinvestment) will affect the output of each local economy in the longer term.

The effect on investment is, as one would expect, slight, and markedly different across sectors. Figure 11.6 shows the percentage changes in capacity by sector in the Doncaster area in 2020, i.e., the accumulated effect on investment (relative to the

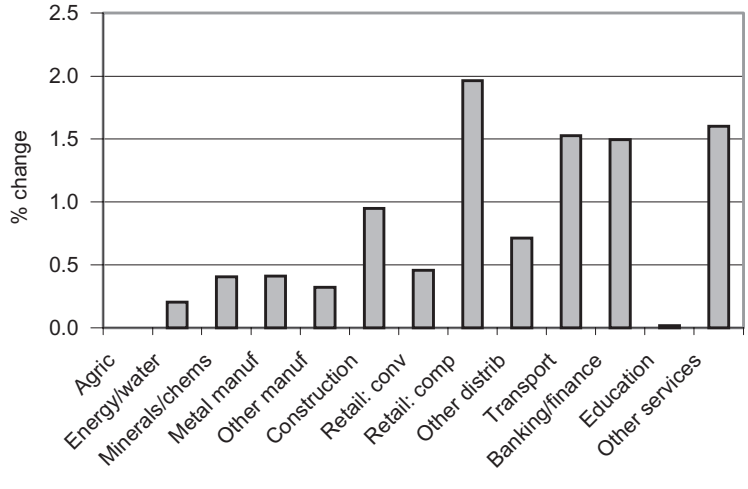


Figure 11.6 *Changes in productive capacity, Doncaster area, 2020, test E3 compared with reference case*

reference case) of slightly better accessibility over 15 years. It can be seen that the greatest impact on capacity is a 2% increase for comparison retailing, followed by increases of around 1.5% for transport and communications, banking/finance and other services.

Most of this increase occurs over the period 2006–2015, and the increased investment is largely responsible for the resulting slight increase in economic activity in the Doncaster area. As is shown in Figure 11.7, total (all-sectors) value added in the area gradually increases very slightly compared with the reference case, reaching a gain of slightly more than 1% after 10 years.

Figure 11.7 also shows the much smaller positive impact on the Barnsley and Sheffield/Rotherham areas. It should be noted at this point that the model system reflects what has become known as the 'two-way road effect', that is, that transport improvements intended to benefit a specific area may assist the economic regeneration of that area, or they may increase the problems of the area by making it easier to serve that particular market from other locations. In the present case, the concentration of schemes in South Yorkshire does have a (very slight) positive effect in that sub-region, though it will be seen below that not all of the indicators are positive for all the districts affected. This comes about mostly through the improvement in accessibilities making South Yorkshire in general, and Doncaster in particular, a marginally more attractive location for investment in various sectors.

The gain in value-added in Doncaster in 2006, the first year after the opening of the road schemes, is rather larger than that in any subsequent year. This is due to the initial impact on the pattern of trade, as Doncaster businesses find it marginally easier to win custom elsewhere as a result of the decreased transport costs. Note that within the results for later years there are multiplier effects:

- the additional production due either to the initial impact on the pattern of trade or to the additional investment creates additional intermediate demand for goods and services; and

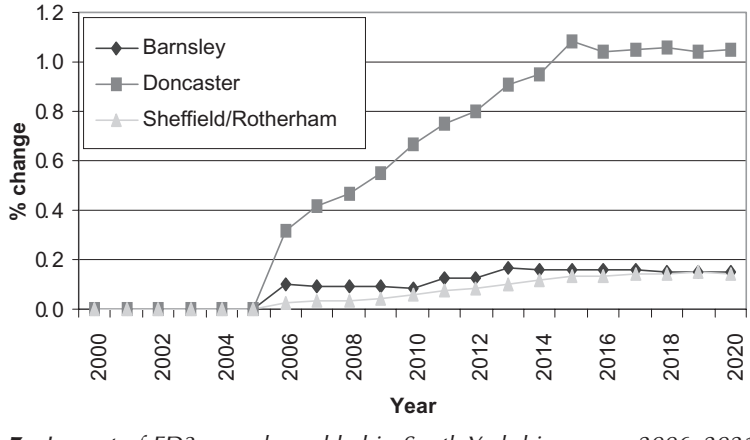


Figure 11.7 *Impact of ED3 on value added in South Yorkshire areas, 2006–2021*

- the increases both in capacity and in production create additional employment which leads to increases in consumer final demand.

The additional employment opportunities affect the pattern of migration, leading to a slight increase in population of the Doncaster area relative to the reference case, which in turn helps to retain expenditure within the area and to reduce the 'leakage' of multiplier effects which would otherwise occur.

Returning to the aspects of the model which are directly affected by the outputs from the transport model, at the urban (zonal) level:

- the local pattern of location of households and businesses will be affected by the changes in accessibility resulting from the changes in generalised cost (with further consequent impacts through competition in the housing and commercial floor-space markets); the pattern of household location will also be affected to some extent by the environmental changes resulting from the changed traffic patterns; and
- the likelihood of working-age individuals being in or out of work will be slightly modified, with consequent impacts on residential location.

These effects and their further consequences (for example, in the development model) work through over a number of years.

The increases in economic activity generate increases in employment. These are shown by district for the South and West Yorkshire authorities, in Figure 11.8. Note

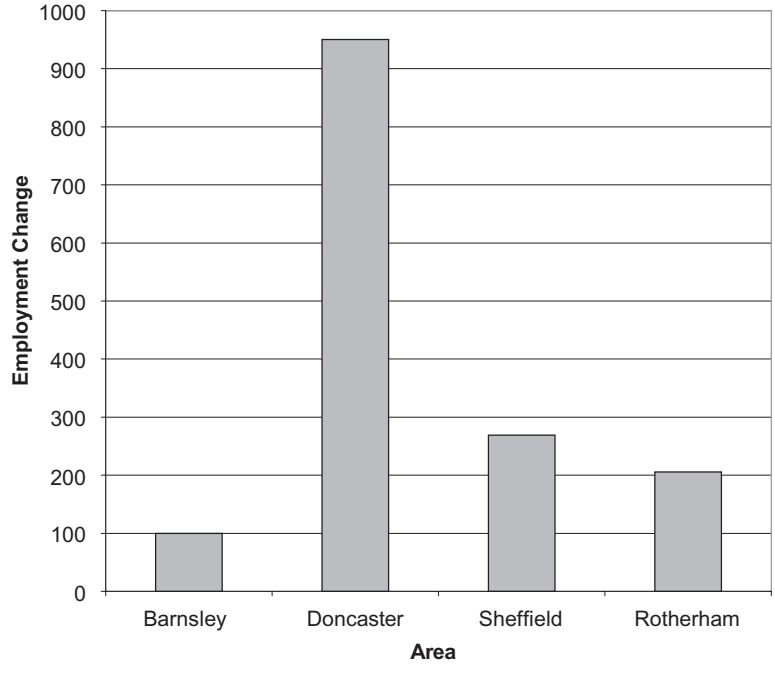


Figure 11.8 *Impact of ED3 on employment in South Yorkshire authorities, 2020*

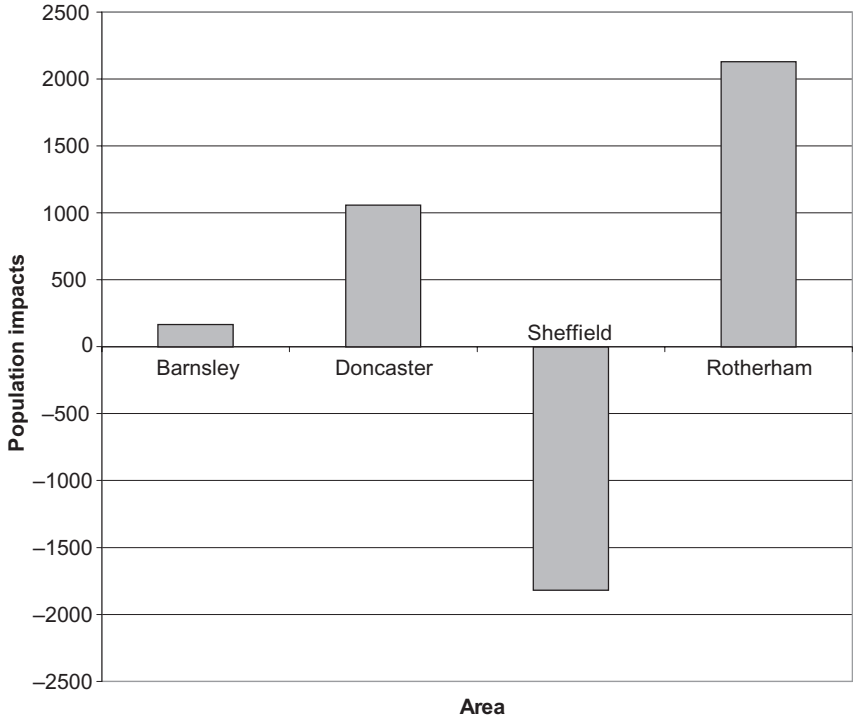


Figure 11.9 *Impact of ED3 on population in South Yorkshire authorities, 2020*

that whilst there are some very slight negative impacts on parts of West Yorkshire, the gain to South Yorkshire is greater, i.e., the gains to South Yorkshire are largely at the expense of other areas rather than of West Yorkshire.

The growth in employment in Doncaster makes a slight improvement to the relative employment prospects there, and this, as already noted, leads to some changes in migration patterns. The overall population impacts by 2021, inclusive of migration effects, are shown in Figure 11.9. Whereas the largest change in employment was in Doncaster, followed by Sheffield and Rotherham, the largest change in population is in Rotherham, with half as much growth in Doncaster and a negative effect on Sheffield.

The localised differences between the employment and population impacts illustrate the scope for transport improvement to modify commuting patterns as well as economic activity. One can see that there is a risk that measures which help to capture additional employment may also promote an exodus of population, and in a small way that is happening to Sheffield in this particular case.

It has to be kept in mind that the impacts considered here are very slight when considered as percentage changes though, as one would expect, there is greater variation between individual zones than between local authorities. This is illustrated in Figures 11.10 and 11.11, which show the percentage impacts of the ED3 test in 2021 on employment and population. The impacts of this particular test are small enough

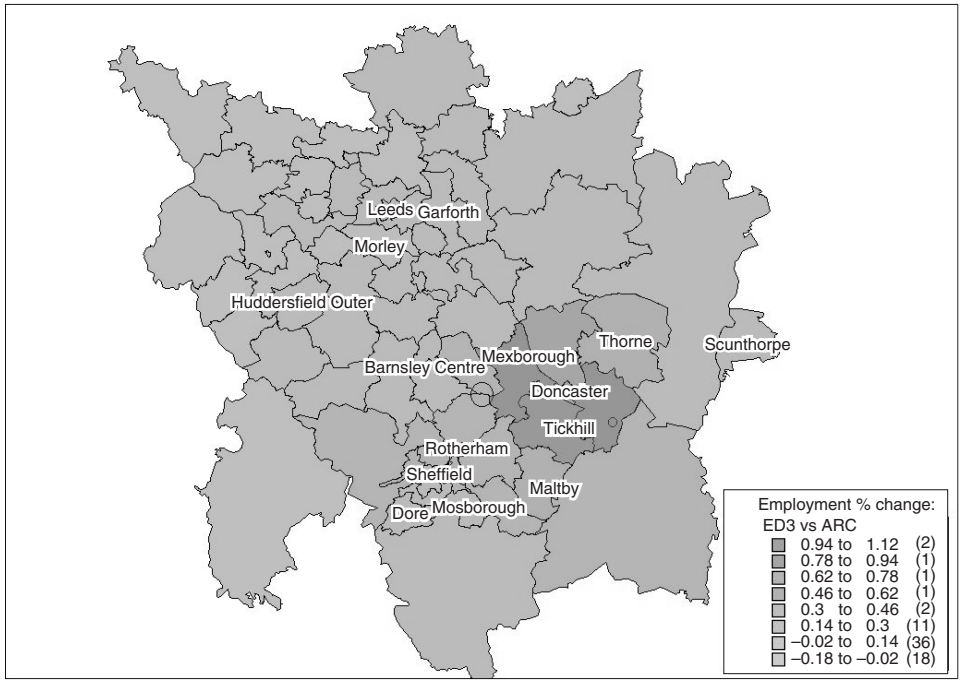


Figure 11.10 *Impact of ED3 on employment by zone, 2020*

that there are no significant impacts either in the land-use impact (little change in residential or commercial rents) or in transport (little additional congestion due to the growth in economic activity and population). In other tests, however, we found that these induced effects could be significant. Indeed we found that land-use policies aimed at encouraging growth in certain locations without complementary transport measures to provide the necessary capacity could actually have negative effects, that is, could actually have the effect of discouraging growth in their target locations.

11.7 Discussion

11.7.1 The Role of the Model in the Study

One of the objectives of using the strategic land-use/transport interaction model in SWYMMS was to provide a formal, quantified analysis of the likely impact of transport and other schemes intended to promote economic regeneration. As we have seen in the ED3 results discussed above, the strategies designed to improve access to regeneration areas and to relieve congestion in their vicinity would have very slight but generally positive impacts in South Yorkshire (particularly the Doncaster area), whilst being neutral for West Yorkshire. Whilst the impacts identified are very small percentage changes from the underlying reference case, it is worth remembering both

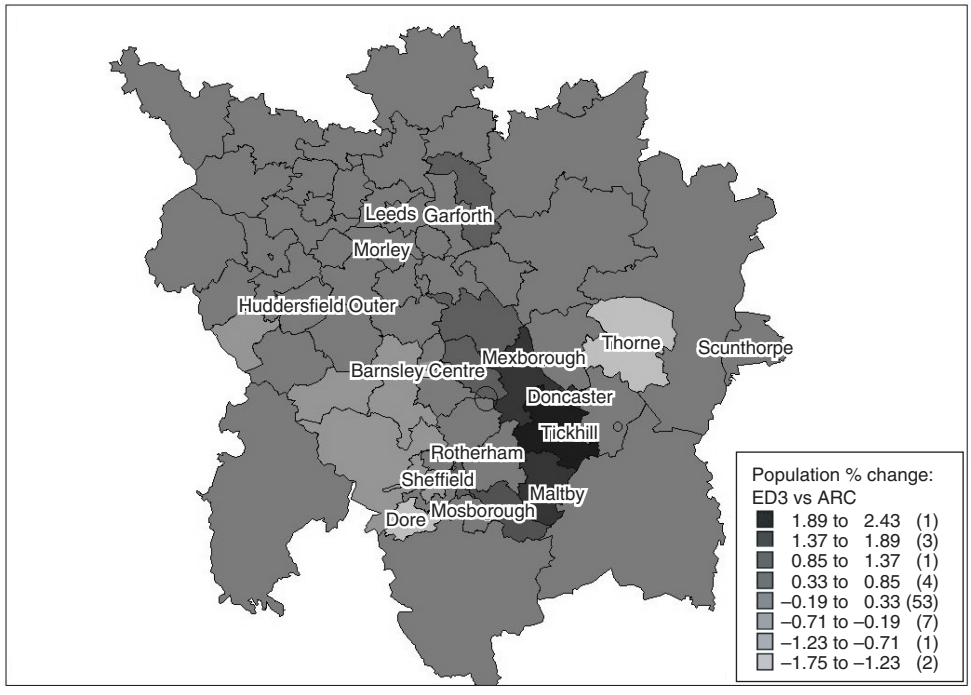


Figure 11.11 *Impact of ED3 on population by zone, 2020*

- that the formal analysis of such impacts may serve as much to guard against exaggerated claims about the economic consequences of transport strategies as to support reasonable claims (whether the consequences are expected to be positive, as in the ED3 case, or negative as might be expected for some demand management measures); and
- that in many cases the small percentages nevertheless represent politically significant numbers of people and jobs.

It is also appropriate to note that the impacts predicted by the land-use/transport interaction model do not always go in the expected direction. An example of this arose in testing another strategy within SYWMMS, which involved substantial (and entirely hypothetical) incentives to the development of retail, office and industrial floorspace in town/city centres and selected Strategic Economic Zones of South Yorkshire. (Details can be found in the SWYMMS Scenarios and Strategies Report, as test ED6.) No other changes were introduced. The development incentives did lead to significant extra construction in the zones affected, but, to the initial surprise of the modelling team, the impact on total employment in South Yorkshire was negative. Detailed examination of the model outputs showed that this was due to

- much of the additional development in South Yorkshire being in congested locations;

- businesses relocating into the additional floorspace thus moving into congested locations and contributing to making congestion even worse (with no provision in the strategy for transport improvements); and
- the resulting increases in congestion, and the increases in the proportion of activity in the most congested locations, making South Yorkshire slightly less competitive as a location for production and slightly less attractive as a location for investment.

Other planners with whom we have discussed these results have described them as an entirely unsurprising and intuitively obvious example of what is likely to happen if land-use and transport plans are not properly integrated!

11.7.2 Model Development

Without embarking on a general review of land-use/transport interaction models, or of the SWYMMS project, it is appropriate to note some of the technical characteristics of the modelling process adopted, and to consider how these are being developed in subsequent projects and may develop further in future.

As noted earlier, the strategic transport model was supported by more detailed highway and public transport models, which were used both to prepare the aggregated descriptions of transport supply used in the strategic model and as the modelling tools for more detailed analysis of highway and public transport impacts in the subsequent development of the plans through which the preferred strategy may be implemented. The possible integration of the detailed and strategic levels of modelling into a single model system, combining the land-use and time-of-day responses currently unique to the strategic model with the more detailed zone system and network representation of the detailed models, is one of the possibilities for future application. This would of course be more demanding of computer resources, but might be justified in terms of avoiding the need for complex interfaces between the two levels, and in terms of the benefits of having a finer zoning system in the land-use/transport interaction modelling. To some extent this approach is already being adopted in work by the authors' firms for the Central Scotland Transport Corridor Studies, where a land-use/transport model with over 700 zones is being used.

Another important aspect is the treatment of land-use planning in the modelling system. The treatment of planning policy and of the planning process in land-use/transport interaction models is typically one of extreme simplification, requiring inputs which tend to imply a very simple 'zoning' system of planning rather than the more subtle approach which operates in the UK. Current developments of the DELTA model are seeking to improve on this, particularly to represent the ways in which the planning process can influence the type of development within broad categories such as 'housing', and in particular to represent the possibility that planning activities in regeneration areas can introduce developments radically different from the existing stock (such as new luxury housing in decaying inner-city industrial areas). More broadly, attention is also being given to the possibility of modelling the planning process as a responsive system working within a quantified policy context,

rather than simply quantifying the expected outputs of the planning system – the amounts of development which will be permitted – as is done at present.

11.8 Concluding Comment

We believe that the SWYMMS strategic model has made a valuable contribution to the study for which it was commissioned, which may well continue to be used on behalf of the various regional and sub-regional planning bodies in relation to other strategic issues. It is representative of the state-of-the-art in forecasting land-use/transport interaction for major UK studies, but further developments in this field can be expected and indeed are already emerging.

Acknowledgements

The authors are grateful to the numerous organisations which contributed to the study and to the work described here; to the editors for the invitation to publish this paper, and to the Government Office for Yorkshire and the Humber for permission to do so; and to their respective colleagues under the overall direction of Dr Denvil Coombe for their strenuous efforts in data collection and in the development and operation of the model. The authors remain responsible for this description of the model and of the use made of it.

The views expressed are not necessarily those of the Government Office or of the Secretary of State for Transport, Local Government and the Regions.

References

Copley, G., Skinner, A., Simmonds, D. and Laidler, J. (2000) *Development and Application of the Greater Manchester Strategy Planning Model*, European Transport Conference, 2000, PTRC, London.

Department for Transport (1998) *A New Deal for Trunk Roads in England*, http://www.dft.gov.uk/itwp/trunkroads/.

Roberts, M. and Simmonds, D. (1997) A strategic modelling approach for urban transport policy development, *Traffic Engineering & Control*, **38**: 377–84.

Simmonds, D. (2001) The objectives and design of a new land-use modelling package: DELTA, in Clarke, G. and Madden, M. (eds) *Regional Science in Business,* Springer, Berlin.

Simmonds, D. and Skinner, A. (2001) Modelling land-use/transport interaction on a regional scale: the Trans-Pennine corridor model, *Proceedings of the European Transport Conference 2001*, PTRC, London.

Simmonds, D. and Still, B. (1999) DELTA/START: adding land use analysis to integrated transport models, in Meersman, H., van de Voorde, E. and Winkelmans, W. (eds) *Proceedings of the 8th World Conference on Transport Research, Volume 4: Transport Policy*, Pergamon, Amsterdam.

12

The Relocation of Ambulance Facilities in Central Rotterdam

Stan Geertman, Tom de Jong, Coen Wessels and Jan Bleeker

Abstract

This chapter considers an important strategic planning question – the relocation of emergency service facilities – as the focus for applied research and, in so doing, exemplifies the reciprocal relationship between academic (or fundamental) and applied research in the field of geo-information and communication technology (geo-ICT). The underlying indicators of accessibility used in the application are introduced initially and a set of optimal locations of ambulance facilities in the study area of central Rotterdam are determined for the client using a planning support tool known as *Flowmap* in conjunction with a proprietary GIS. Some further experiments to improve the solution are also presented.

12.1 Introduction

Since the early 1990s, interest in information systems for storing, managing, processing, analysing and displaying geographical data has boomed. The last decade has witnessed some new areas of application for this 'geo-information and communication technology' (geo-ICT). The expansion started in sectors such as defence, forestry and mining and has extended into geo-marketing, insurance, car navigation and telecommunications. Moreover, geo-ICT expanded in size, functionality and diversity (Tomlinson, 1987; Burrough, 1986; Ottens, 1993; Steinitz, 1993). Increasing numbers of commercialised information systems were put on the market and new systems, or updated versions of old systems, have been developed, offering novel

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

capabilities (e.g., Internet-GIS) as well as traditional functions in a way not previously available (e.g., three-dimensional mapping/animation) (Geertman, 1999).

Research activities within the field of geo-ICT can be broadly divided into those that are fundamental and those that are applied. Fundamental research is frequently referred to as 'blue skies' research and is often undertaken in an academic context. Examples of fundamental research include studies of ontological and/or semantic issues (e.g., Winter, 2001), advances in the use of open technologies (e.g., OCG, 2002), the creation of new error propagation methods (e.g., Burrough and McDonnell, 1998), and the development of geographical modelling methodologies such as those based on artificial intelligence (Openshaw, 1997; Fischer, 1999). Applied research involves the application of these new technologies in real-world contexts, frequently in public sector planning at local or regional level. Whilst the distinction between fundamental and applied research is drawn for purposes of clarification, most institutions involved with geo-ICT related research undertake a mixture of both types of activity. This has led to the critical view amongst commentators in the higher education sector that there is a growing amount of research, which is money-driven, and that fast, clear-structured and short-term applied research is taking priority over more fundamental longer-term research, which is suffering as a consequence. However, there are some advantages of applied work, not least of which is the necessity to develop policy-relevant tools for understanding problems and providing solutions. Consequently, an appropriate combination of more theoretical fundamental research activity and more applied developmental work seems desirable. At Utrecht University in the Netherlands, for almost two decades now, various research groups have been working with geo-ICT. Although some of these are more oriented towards fundamental research and others towards applied work, in general all groups show elements of both and, in certain cases, their research activities have resulted in the commercialisation of geo-ICT products (such as *Flowmap* and PCRaster) and in the creation of (quasi-) private organisations (such as Nexpri and Carthago).

Flowmap, for example, is a stand-alone software package dedicated to the analysis of spatial interaction data, data that describes movement of people, goods, energy or information through space on a regular or irregular basis usually via some form of transport network. An example of regularly occurring interaction is commuting whereas irregular interaction might be residential migration. *Flowmap* has capabilities for the selection and display of this kind of data, together with various methods and models to analyse historical flows and predict potential interaction. These latter capabilities make *Flowmap* a very suitable tool to incorporate in planning support systems, especially in the field of facility/service localisation (Geertman *et al.*, 2002). Although *Flowmap* has many GIS functions, it is not designed as a full GIS package, but rather as an extension for specific spatial analyses that are difficult or impossible to realise within mainstream GIS packages.

Nexpri, the Netherlands Expertise Centre for Geographical Information, can be regarded as an example of a spin-off organisation. It began its existence in 1989 as a subsidy-based institution related to five Dutch universities, but in 1996 it became an entirely market-oriented organisation related to Utrecht University and is soon to change its status to that of a private company. Its activities primarily concentrate on two fields of applied geo-ICT research: one concerns the organisational issues related

to the implementation and functioning of geo-ICT in organisations; the other concerns different kinds of location-oriented research questions such as the identification of the best location for a new dealer in the branch network. There is a strong interrelationship and active interaction between the research groups at the University and the applied research undertaken by Nexpri.

In this chapter we discuss one particular example of the benefits of this reciprocal interrelationship between fundamental and applied Geo-ICT research and endeavour to demonstrate how the two-way interaction between the two research communities can be of considerable advantage to both. The example involves an important strategic planning question – the relocation of ambulance facilities. Nexpri was contracted by the Rotterdam Municipal Health Authorities to undertake a study of the provision of ambulance services in the Rotterdam Harbour area and to identify how the location of ambulance stations might be altered to improve the efficiency and effectiveness of the service.

Initially, attention is focused on some of the underlying concepts related to the research question in general terms, i.e., how can facility siting be performed and how to measure accessibility. Thereafter, the real-world planning question will be elaborated in more detail, followed by an explanation of the way in which the *Flowmap* software has been applied to solve this planning question. Subsequently, reflections on the application will be presented and recommendations made about future applications of this instrument to these kinds of research questions. Finally, some remarks are provided on the experience of working in this reciprocal interrelationship between fundamental and applied geo-ICT research.

12.2 Facility Siting with Accessibility Analysis

12.2.1 Introduction

Facility siting is a form of applied research aiming at finding the absolute or relative optimal location or set of locations for a predefined facility or group of facilities. Facility location is closely related to the concept of geographical accessibility that can be defined as the effort taken to get to some place and the level of acceptability to the target group. So the concept of accessibility incorporates the transport link between origin and destination and the ability to travel by the target group, as well as the characteristics of the destination and the objective of the trip. For example, what is accessible by private car may be inaccessible by public transport; what is within an accessible walking distance for younger people may be effectively inaccessible for senior citizens; what is an acceptable effort to travel to buy antique furniture is likely to be unacceptable when shopping for daily groceries.

Moseley (1979) has put the concept of accessibility into a schematic framework in which activities require transport and transport involves people; from this it follows that actual accessibility depends on the properties of each of the components in this scheme. However, accessibility can be measured in many different ways. Robinson (1990) distinguishes between composite measures, comparative measures and time–space approaches. The time–space approaches are most useful for studies based

on the perspective of the individual: 'Is service C accessible for a person living in place P with mode of transport T?' The other approaches are more useful for measuring aggregate accessibility. Composite measures are a combination of several different individual measures like 'How many buses arrive before 8.00 a.m.?' or 'Is there a Sunday bus service?'. Comparative measures also take differences into account between social groups like car users or public transport users. In addition the comparative measures can be further subdivided into potential and actual accessibility measures (Hilberts and Verroen, 1992). Moreover, accessibility can be measured for many individual locations or these measures can be combined into an aggregate measure for a whole region. The latter is relevant, for instance, when one wants to measure the effect of infrastructure changes on the accessibility of (or inside) a larger region. Finally, accessibility measures can be based upon a single-purpose trip as presented in the scheme above or on multipurpose trips (also known as combination trips) where for instance the accessibility of a shopping centre is based on its location relative to the home–work trip of the customer (Dijst and Vidacovic, 1995).

Despite its geographical dimension, facility siting is not a field of study restricted to human geography. Other disciplines, such as operations research, have made contributions too and still do. Therefore, before discussing the methodology of accessibility analysis in depth, we discuss some alternative approaches for research on facility siting: the optimisation approach from operations research; the buffer/overlay approach from geographical information science; and the values of accessibility potential that have their roots in quantitative geography.

12.2.2 The Optimisation Approach

Whenever a facility-siting project is considered, there is always the question of whether or not it really involves an optimisation problem and requires an optimisation approach. This is because there are several reasons why such an approach is not appropriate (Floor *et al*., 2001):

- not all factors of relevance in a public planning or policy environment can be included in an optimisation approach;
- the typical 'hard' and 'singular' result of an optimisation approach often is very difficult to incorporate in a public planning or policy environment;
- optimal results can be highly sensitive to the choice of spatial aggregation and also to temporal changes; a good example of this is the sensitivity to the modifiable area unit problem (MAUP) as described by Fotheringham *et al.* (1995);
- there is not necessarily one optimal solution, but many different optimal solutions depending on the chosen target functions (like minisum or minimax);
- although exact solution methods are known for these target functions, for most practical applications they require a prohibitive amount of computer memory or time, in which case usually some approximation method has to be used;
- it is very difficult to optimise using transport network-based impedance functions instead of airline distances. The traditional optimisation approach is exemplified by the ALA and NORLOC algorithms described by Cooper (1964) and Törnqvist

et al. (1971) for locating services in a plane and Maranzana (1964) for locating services on a transport network. There is no known published example of locating services in a plane using distances via a transport network.

In a recent publication (de Jong *et al.*, 2001) we argue that the only useful part optimisation techniques can play in a public planning or policy environment is to provide a yardstick against which the outcome of a project can be measured in terms of accessibility.

12.2.3 The Buffer/Overlay Approach

In an earlier study (de Jong and Ritsema van Eck, 1997), the traditional GIS approach was outlined. It mainly involves the use of buffer techniques to determine a straightforward 'accessible' area followed by overlay techniques to clip out this buffer area and to present the remainder as the 'inaccessible' area. This approach can be criticised for a number of reasons. Firstly, transport networks are not taken into account. Not only does travel speed differ depending on mode of travel and network quality but also natural (waterways, steep slopes) or man-made (railway tracks, airfields) barricades will have a diminishing effect on accessibility. Secondly, all sites within the buffer zones have similar weight, so a location with all opportunities at the maximum allowed distance scores just the same as a location with all opportunities at minimum distance. Depending on the type of facility siting that is being undertaken, other shortcomings of the buffer/overlay approach may also be of relevance: competition with already established service outlets is not taken into account; and depending on the subject of analysis, sometimes the limited capacity of service outlets cannot be ignored.

12.2.4 Potential Values

To overcome some of the shortcomings of the GIS approach, Geertman and Bosveld (1990) used the potential value technique based on a transport network instead of using buffers. This can be seen as a first attempt to incorporate some accessibility analysis into GIS. Potential values are the best-known measure for the nearness of clients/customers/potential users and were first introduced by Hansen (1959) to describe accessibility. Potential values can be calculated as follows:

$$Pot(i) = \sum_j D_j \cdot C_{ij}^{-\beta} \tag{12.1}$$

in which:

$Pot(i)$ is the potential value for client site i;
D_j is the weight (size or number) of services provided for service location j;
C_{ij} is the distance between origin site i and destination site j; and
β is the distance decay parameter.

This equation can be used to find good sites for new residential areas, for example, based on the number of jobs offered at the destination locations. Sites relatively close to a large D_j score a higher potential value than sites further away. However, being relative measures, potential values are difficult to interpret. This problem can be overcome by calculating two sets of potential values where the distance decay parameter differs by exactly one. After dividing the outcome of the one set of values by the other the result will be an accessibility measure in the form of a weighted average distance for each origin site to all destination sites (Geertman and Ritsema van Eck, 1995).

Also in practice we see that, because of the ease with which potential values can be calculated within a raster-based GIS, the measure is often calculated by using just the geometric distances between cell centroids. As these distances are comparable to the simplest form of airline distances, the first criticism mentioned above regarding the buffer/overlay approach is in this case valid too. There is no satisfactory way in raster-based GIS to handle the effect of the discontinuous aspects of travel via a network, like multilevel overpasses and limited access to highways. Therefore, in our opinion, raster-based GIS are unsuitable for any realistic analysis involving calculations of effects of travel via transport networks.

There are some other criticisms of the potential value approach that are harder to overcome. In gravity modelling the distance decay function can be calibrated using the (log) mean trip length to represent distance behaviour correctly. A similar mechanism is not available when it comes to the calculation of potential values (often simply a power function is selected with distance decay parameter 2 in an unfounded analogy with Newton's Law of Gravity). Moreover, the results are highly sensitive to opportunities at extremely short distances (Ritsema van Eck and de Jong, 1996). As these particular distances are often arbitrary (intrazonal distance estimation or depending on the location of the centre point that represents a whole zone), this renders any outcome potentially unreliable.

12.2.5 The New Methodology of Accessibility Analysis

A comparative potential accessibility measure that overcomes most of the criticisms mentioned above can be developed. Moseley's schematic framework involving population linked to activities by a transport network can be redefined in GIS terms as origin locations linked to destination locations by transport network distances. Based on this adapted scheme, we can use GIS to calculate, for each destination location, which of the origin locations are within reach (and sum up the selected attribute values of those locations) and vice versa. This creates an accessibility measure called the proximity count or cumulative opportunity index. The proximity count is a simple count of locations that are within a predefined distance range. Instead of counting the number of locations, a weight variable for each location (such as the number of jobs) can be summed. For each site, all locations are counted that meet the distance requirement:

$$PCount(i) = \sum_{j^*} D_j$$
$$(j^* = \text{over all } j \text{ such that } C_{ij} < \text{Range}) \tag{12.2}$$

where $PCount(i)$ is the proximity count for origin site i.

A location or location related value is not uniquely allocated to any site. In fact, locations can be counted for more than one site. The proximity count therefore gives an indication of the potential (job) market not taking any competition effects into account. It can be used to determine which sites reach above certain (minimum) threshold values. Of course the choice of distance is critical: if it is set too large then all origins will be counted for all destinations and all proximity counts will end up equal. As an alternative, the threshold value can be preset and the threshold distance can be calculated as the minimum distance at which enough locations are to be reached (de Jong and Ritsema van Eck, 1997). Proximity counts and threshold distances are especially useful for determining whether or not the potential market size of a proposed site (service or retail) is large enough (de Jong *et al.*, 1991). In the next section, we exemplify this with the help of a real-world strategic planning question, the relocation of ambulance facilities.

12.3 The Relocation of Ambulance Facilities

In the Netherlands, as in many other countries in the western world, increasing the efficiency of public services is one of the main solutions for handling decreasing funds for public facilities. One of the areas in which the cutback in public funds is translated into a quest for improved efficiency is in the number and the location of ambulance service stations. The Dutch national government is striving to unify smaller service areas into larger service regions. This forces different ambulance services to collaborate, which in turn leads to a more efficient and effective use of ambulance provision, and a reduction in costs. One of the newly formed service regions is the ambulance service region of the Rotterdam Harbour area (Figure 12.1).

Ambulance provision is divided into three categories. The first category consists of the urgent journeys (so-called A1 journeys) where, according to the law, an ambulance has to arrive at its destination within 15 minutes of the Central Post for Ambulance (CPA) journeys receiving a call for help. The second category involves less urgent journeys (A2 journeys) where there is no strict time limit other than to arrive 'as quickly as possible'. The last category consists of so-called 'ordered' journeys (B journeys) consisting mainly of pre-arranged journeys between hospitals and pre-arranged journeys of patients from home to hospital and vice versa (e.g., for radiation therapy). In 1998, 74 478 ambulance journeys occurred in the study area of which 45% belonged to the A1 category and 41% to the B category.

The existing ambulance services and the CPA asked the Municipal Health Services of the Rotterdam Harbour area to investigate which ambulance stations in the Rotterdam Harbour area should be established within the new configuration and to determine how many ambulances would be needed in the area to fulfil the requested ambulance provision. To answer this question the Municipal Health Services contracted Nexpri to undertake an analysis of the effectiveness of current service provision and to recommend how ambulance services within the Rotterdam Harbour area might be improved.

The locations of the 16 existing ambulance stations and the classified total demand for A1 journeys in 1998 for each four-digit postal area in the Rotterdam Harbour

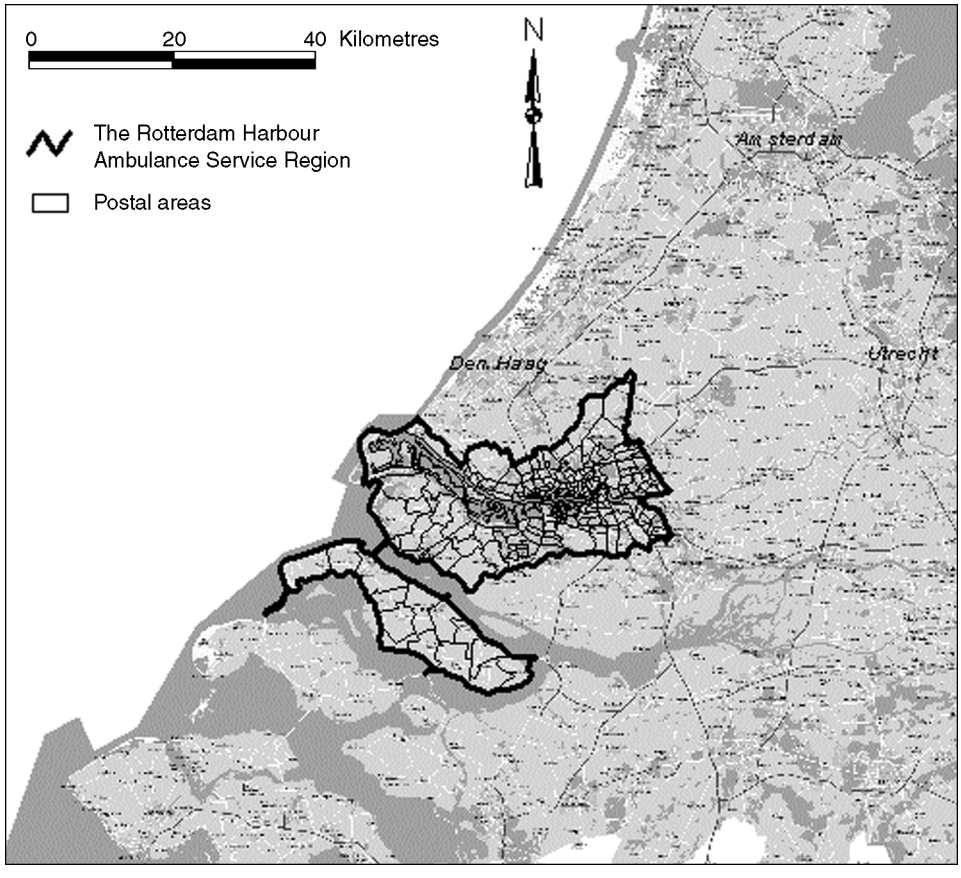


Figure 12.1 *Rotterdam Harbour ambulance service region*

area are shown in Figure 12.2, indicating the high demand intensity (counted number of calls for A1 journeys) in the centre of Rotterdam (the eastern part of the Rotterdam Harbour area) and decreasing towards the perimeter of the area.

Initially, a measure of the effectiveness of the existing ambulance services provided from the 16 stations was calculated with the help of ArcView and ArcView Network Analyst. The service area associated with each ambulance station was constructed by calculating an 8.5 minute travel time buffer based on the road network. To determine the required locations of stations to be able to reach patients within 15 minutes in response to an A1 call for assistance, a 95% confidence interval was calculated based on the Poisson distribution of A1 arrival times. The upper 95% confidence limit was assigned at the legal norm of 15 minutes. This resulted in a mean arrival time of 8.5 minutes within the confidence interval. The time period of 8.5 minutes takes into account the problem of congestion at peak hours as well as the average delay time (2 minutes) between the incoming call for assistance at the ambulance station and the moment the ambulance begins its journey.

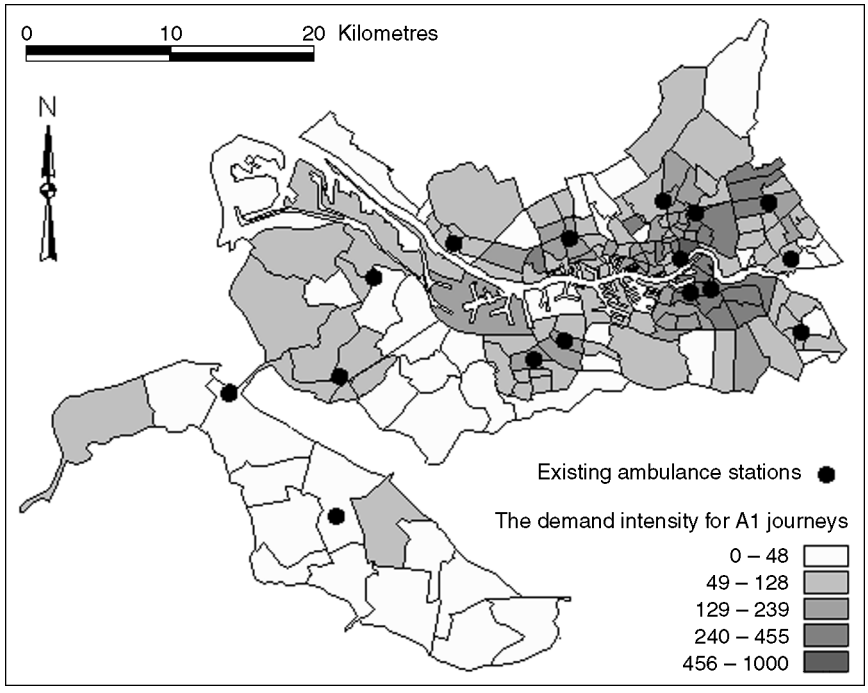


Figure 12.2 *Location of existing ambulance stations and the counted demand intensity for A1 journeys in 1998 per four-digit postal area*

Within the city centre of Rotterdam, the ambulance stations have overlapping service areas. Moreover, the travel time zones indicate that some areas in the more peripheral areas are not covered at all (Figure 12.3). The conclusion of this is that, when taking the threshold value of 15 minutes as a maximum for A1 journeys (8.5 minutes as an average), some parts of the Rotterdam Harbour area are in fact over-served by existing ambulance services and some are under-served. This implies that a more efficient distribution of ambulance services over the Rotterdam Harbour area might be identified.

As a starting point for this redistribution of ambulance station locations, four pre-requisites must be taken into account:

- each part of the Rotterdam Harbour area is potentially suitable as the location of a new ambulance station, so no areas are excluded initially;
- the new locations for the ambulance stations should be situated as close to demand as possible;
- the overlap of service areas of the ambulance stations should be minimised; and
- the total range of the new ambulance stations should cover at least all the urban areas within the Rotterdam Harbour area and preferably also the remaining rural areas within this region.

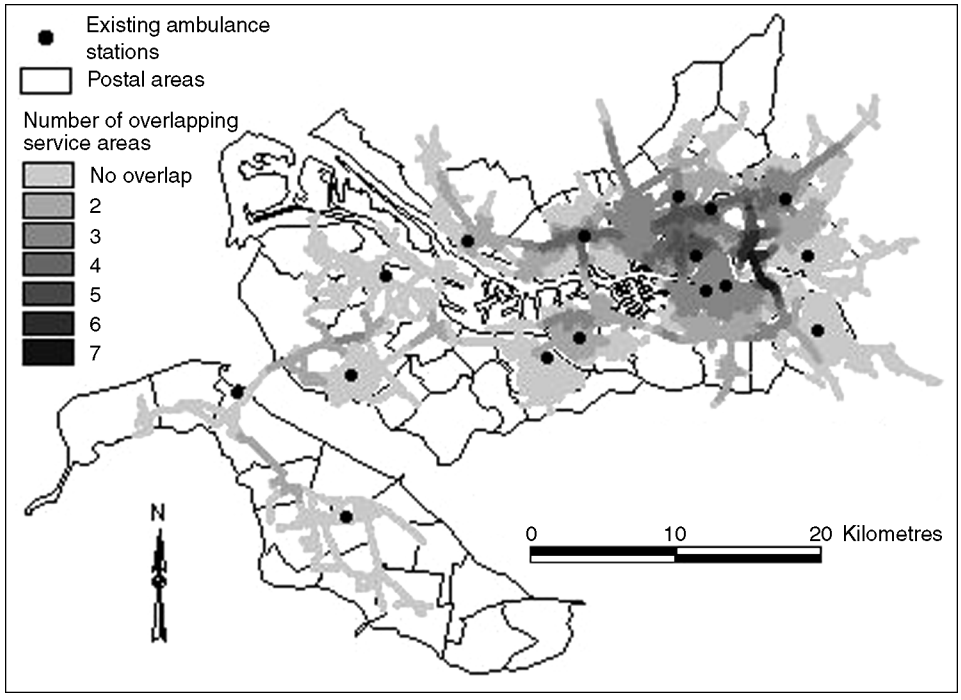


Figure 12.3 *Service areas for current ambulance stations*

Taking these prerequisites into account, network-based analyses have been performed with the help of ArcView and *Flowmap*. ArcView has been used primarily for database management, map-making, and the requisite analytical GIS functions of buffering and network analysis. *Flowmap*, on the other hand, has performed all the additional network calculations and measures that cannot be executed within the proprietary GIS (for additional information concerning *Flowmap*, see http://flowmap.geog.uu.nl). The demand for ambulance facilities is represented by the intensity of requests for ambulance service for each individual postal area. Moreover, use has been made of a hexagonal grid, corresponding in size with the average four-digit postal area. The grid cells represent potential new ambulance locations in the Rotterdam Harbour area. The choice for a hexagonal grid was made because the distance from the centroid of a square grid to the centroids of its neighbouring square grids is different (the diagonal is longer than the vertical or the horizontal), while with a hexagonal grid this distance is equal. In addition, a detailed road network has been used that links all the four-digit postal areas and hexagonal grid cells. Each of the component parts of the road network possesses a travel time, which also takes into consideration the delay time as a consequence of congestion.

Two accessibility measures, the proximity count and the proximity coefficient, have been computed for each of the hexagonal grid cells using *Flowmap*. The proximity count gives an indication of the demand for ambulance services within a user-defined

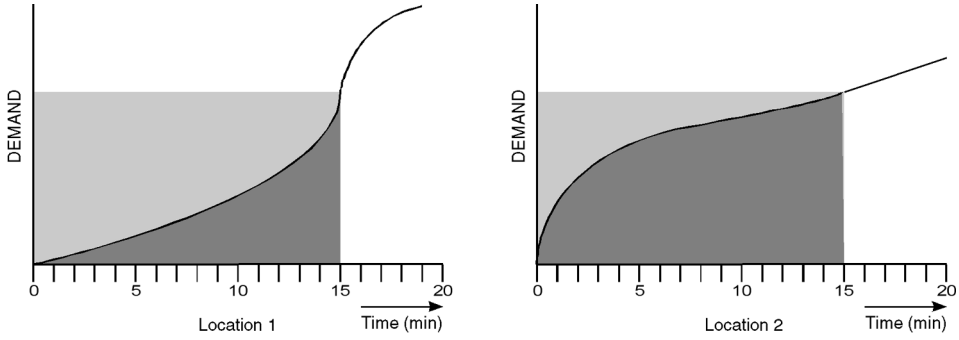


Figure 12.4 *The cumulative location profiles of two alternative sites*

travel time from each hexagonal cell. In this case, the proximity count indicates the intensity of requests for ambulance services within an average travel time of 8.5 minutes. This proximity count can be presented in a map, but also in a so-called location profile (Figure 12.4). A location profile for a potential ambulance station indicates the distance (in this case expressed in travel time) between that location (hexagonal cell) and the cumulative requests for the ambulance service. In fact, the shape of the location profile provides some information on the accessibility of each specific location. A convex profile (as shown in location 2 in Figure 12.4) indicates that the site is relatively close to the demand for ambulance services; a concave location profile (as shown in location 1) indicates that the site is relatively far away from the location of demand for the service.

This relationship can also be expressed in one measure, a so-called proximity coefficient, calculated by expressing the area under the graph (dark shaded area) as a percentage of the total area (dark and light shaded areas). Hexagonal grid cells with high proximity coefficients possess high accessibility to demand and are therefore very suitable sites to locate ambulance stations. The inverse is true for grid cells with low proximity coefficients. Figure 12.5 shows the proximity coefficients – classified on natural breaks – for each of the hexagonal grid cells within the Rotterdam Harbour area (in fact it appears that other classification methods in ArcView showed similar results).

Based on the proximity count maps and location profiles, the optimal future locations of ambulance stations were assigned by trial and error. The proximity coefficients were used to locate the ambulance stations so as to maximise accessibility and minimise the overlap between service areas. It appears that in and near the city centre of Rotterdam these coefficients possess a high value, so represent sites with good accessibility, as we would expect. The siting of ambulance stations in the (south) western part of the Rotterdam Harbour area is mainly driven by the achievement of the fourth prerequisite: that the range of the new ambulance stations should cover at least all the urban areas within the study area and preferably also the remaining rural areas within this region. Figure 12.6 illustrates the proposed sites for the ambulance stations.

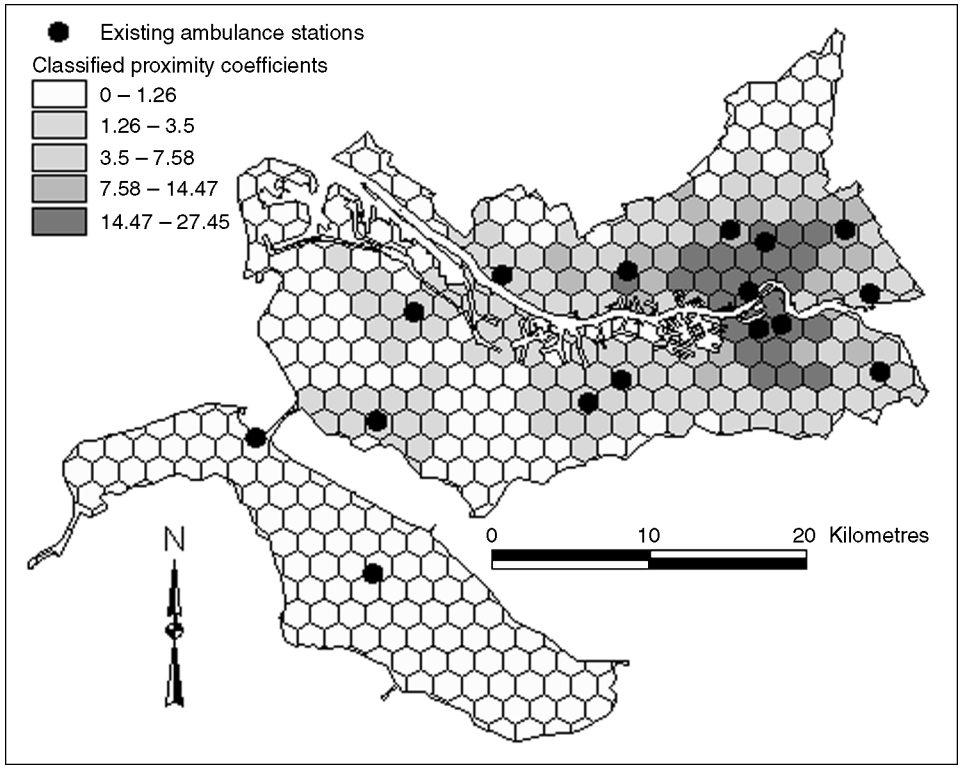


Figure 12.5 *Classified proximity coefficients for hexagonal grid cells in the study area*

This allocation process resulted in a total of 13 newly located stations (three stations less than the initial situation), with a reduction in the level of over-provision (less overlapping service areas) and an improvement in under-provision (fewer areas that do not fulfil the threshold of 15 minute A1 journeys) (Figure 12.7).

The results of this analysis were delivered to and accepted by the Municipal Health Services at the end of 2000. The existing ambulance services, the CPA and the Municipal Health Services of the Rotterdam Harbour accepted the results of the study. The political debate about the reconstruction of the ambulance service network is still in progress.

12.4 Further Developments

The analysis conducted for the client shows the solution of a real-world planning problem by applied geo-ICT research with the help of more fundamental research than had been required to create *Flowmap* in the first instance. Although acceptable to the Rotterdam Municipal Health Services, the research undertaken was not particularly innovative and begs the question of whether it is possible to improve on the proposed facility locations.

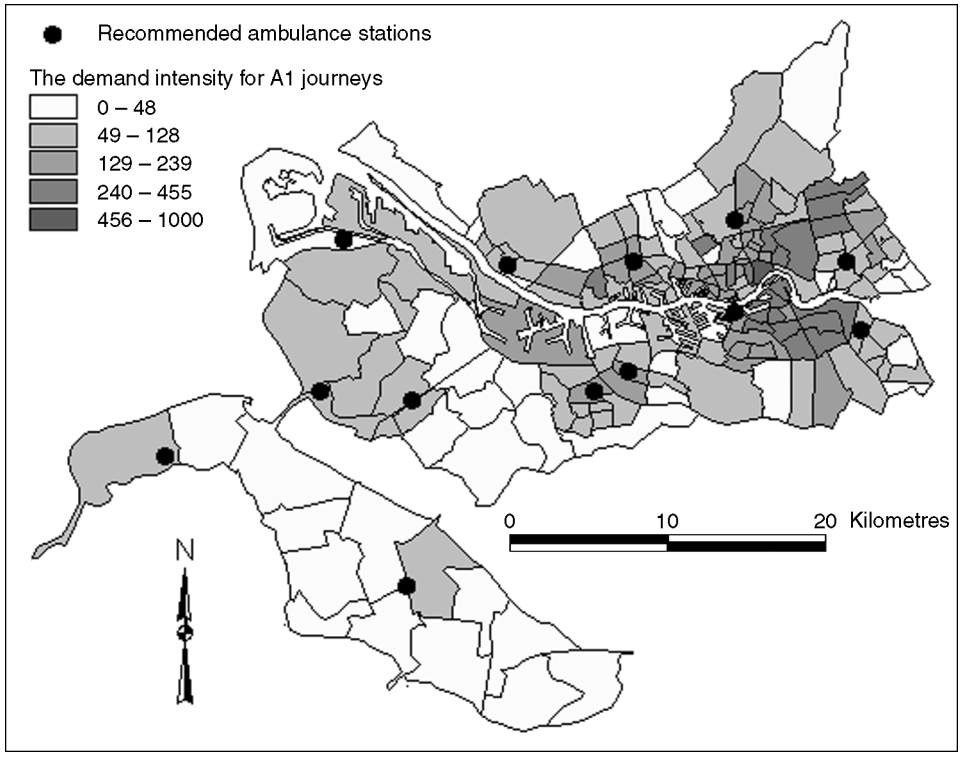


Figure 12.6 *The proposed sites for ambulance stations*

One development is to make use of the so-called 'alternate' algorithm to examine the distribution of ambulance stations. This alternate algorithm is an approach to optimisation that can be adapted to a GIS environment when dealing with the impedance associated with transport networks. As such it can be considered a combination of different approaches to the problem of facility siting. The basic principle of this adaptation is to 'tessellate' space into a large number of cells and to calculate the optimisation criterion for all the cells. The most extreme value found in a catchment area is the best location for a new site. Where the best site differs from the current site it replaces the current site. Thus, new catchment areas and best sites should be calculated until no more substitution takes place. This procedure does not guarantee a true optimal solution; there is some sensitivity to the MAUP problem and the procedure may end in a suboptimal solution. However, in general, it gives a good indication of what can be achieved through optimisation.

When performing an optimisation, it is necessary to be aware that there is not just one optimal solution, but there may be several solutions, depending on the 'target' function that is to be minimised. In the case of ambulance stations, for example, one can handle at least two different target functions. On the one hand, there is a good case for minimising the mean unweighted travel cost, ensuring the best accessibility for the whole area. On the other hand, one can minimise the mean travel cost

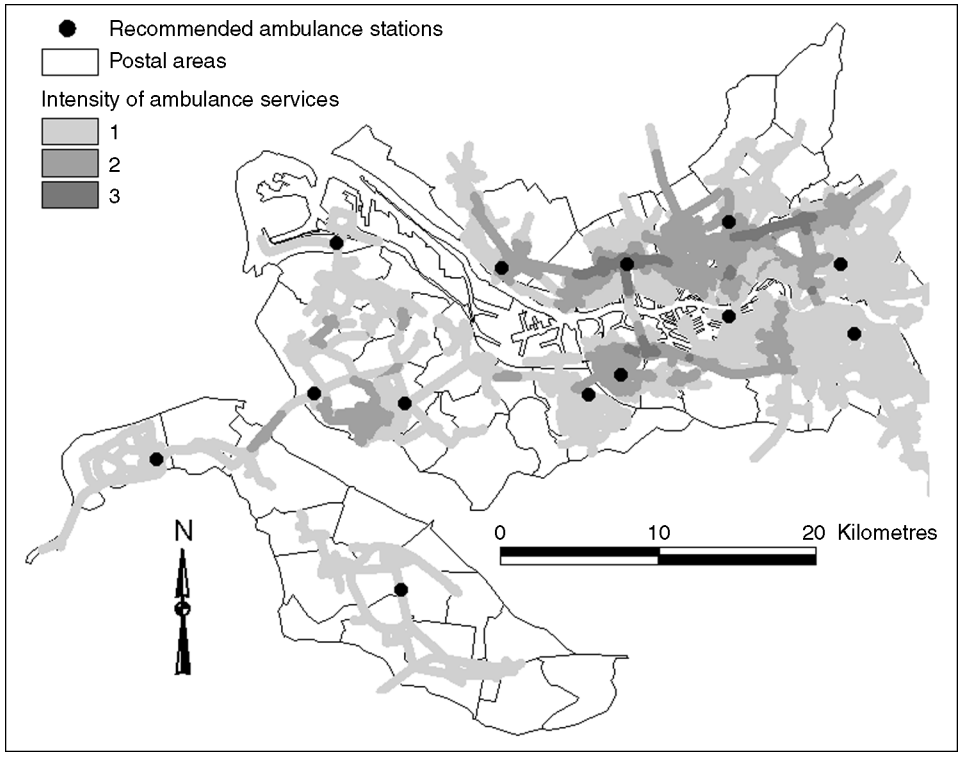


Figure 12.7 *Service areas of proposed ambulance stations*

weighted by the distribution of demand, ensuring the best accessibility where demand is highest. In the first case, sites tend to be found towards the centre of their catchment areas. In the second case, sites are more oriented towards the centre of gravity of demand within the catchment area. These two points may very well not overlap and as a consequence the choice of target functions can result in very different solutions.

The consequences of handling different target functions can be exemplified using the problem of siting ambulance stations. We have divided the study area into almost 1400 hexagonal cells each with a radius of 50 metres. Each hexagon has been allocated to the nearest of the proposed 13 ambulance stations. The study area is thus divided into 13 catchment areas as shown in Figure 12.8a.

Based on these initially proposed 13 catchment areas, some access time statistics have been calculated (Table 12.1). These statistics represent the base position shown in the second column. If the target criterion is to minimise the mean area access time without weighting, the catchment areas in Figure 12.8b are derived and the access times change accordingly. The mean area access time is reduced from 9.76 to 8.26 minutes by relocating 12 of the 13 proposed ambulance stations. This optimisation procedure can be repeated again but on this occasion under the criterion of minimising the mean travel cost weighted with the distribution of demand (Figure 12.8c).

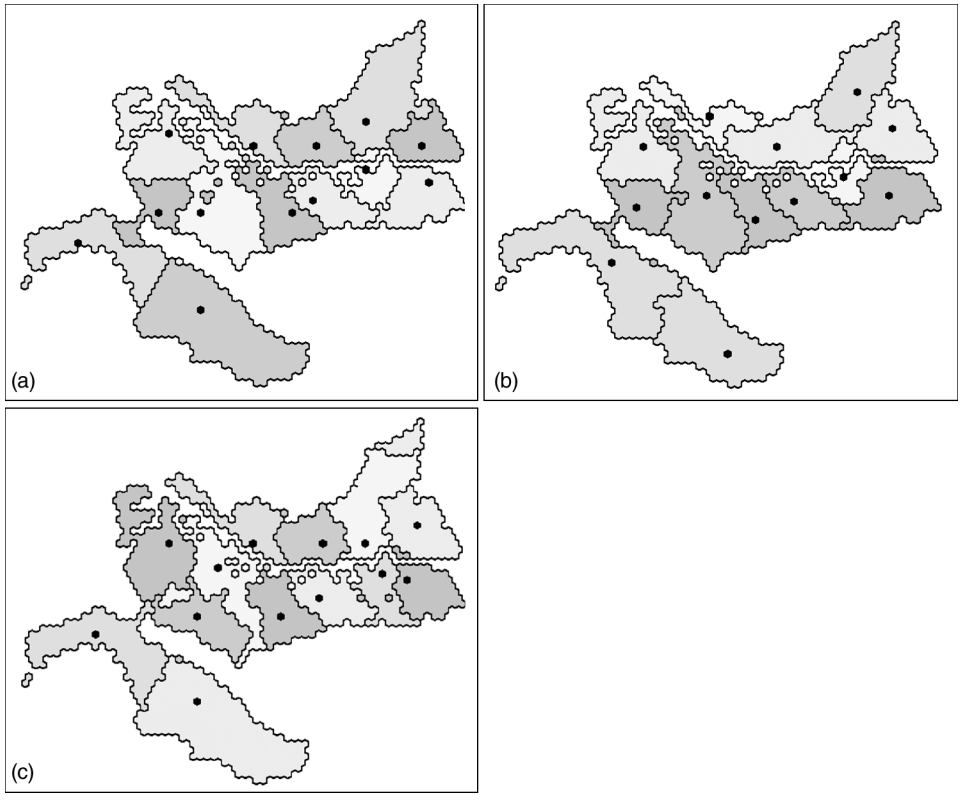


Figure 12.8 *Catchment areas: (a) proposed initially, (b) derived from minimising unweighted mean travel costs and (c) derived from minimising weighted mean travel costs*

Table 12.1 *Access times under different target criteria*

	Initial proposals (minutes)	Minimise mean area access time (minutes)	Minimise mean demand access time (minutes)
Mean area access time	9.76	8.26	9.29
Mean demand access time	7.09	6.70	5.62
Worst-case access time	43.22	38.24	34.59

This ensures the best accessibility where demand is highest. Based on the resulting catchment areas, the mean demand access time is reduced from 7.09 to 5.62 minutes by moving 12 of the proposed 13 ambulance stations.

This approach, using the alternate algorithm, can be applied in all cases where the relocation of service facilities, possibly in combination with an addition or reduction, is at stake. The added value of this kind of analysis is not so much the maps shown

in Figure 12.8, but primarily the statistics that go with these maps. Similar statistics can be calculated for all sorts of scenarios based on the relocation of fewer ambulance stations, in which case the optimal statistics provide a yardstick to measure the effect of each of the scenarios. Although of academic interest to continue such analysis until the optimal location has been found, it should be stressed that in practical situations accessibility rarely is the only factor playing a part in the decision-making process. As a consequence, the trend surface that can be constructed around the optimal location is of far greater importance as it can be combined in a multi-criteria fashion with other criteria to arrive at a decision space that provides the decision-maker with some direct insights.

12.5 Conclusions

This chapter has provided a particular example of the benefits of a reciprocal relationship between fundamental and applied research in the context of public service planning. An analytical tool (*Flowmap*) and a proprietary GIS have been used to generate a set of proposals for the relocation of ambulance stations in central Rotterdam. Some additional developments have been reported to show more state-of-the-art capabilities and possibilities.

In relation to the specific policy problem, efficiency has been improved by reducing the number of ambulance facilities from 16 to 13. Moreover, the effectiveness has been improved by a reduction of the areas experiencing either over-provision or under-provision of service. In this way the goals of the applied research have been achieved and have become input into the policy process. The outcome of the political process is still unclear, although the upcoming pressing financial situation undoubtedly will push in the direction of an efficiency operation in which the outcomes of this research will be of help.

As far as the interrelationship between fundamental and applied geo-ICT research is concerned, two final conclusions can be drawn. First, without the input of the former and its outcome (*Flowmap*), the research question would have had to be answered in a far more rudimentary way. In fact, *Flowmap* offered different tools that substantially improved our insight into the planning problem and the generation of a solution. Second, the research performed can only be attained by individuals working closely together on fundamental and applied research. It is the direct contact that is of additional value to both parties and which will result in novel solutions. This collaboration, the solution that has been produced and the insights generated through the research process, give impetus to undertake further applied work to support public service planning.

Acknowledgement

The authors would like to thank the editors of this book for their fruitful comments on an earlier version of this chapter.

References

Burrough, P.A. (1986) *Principles of Geographical Information Systems for Land Resources Assessment*, Monographs on soil and resources survey 12, Clarendon Press, Oxford.

Burrough, P.A. and McDonnell, R.A. (1998) *Principles of Geographical Information Systems*, Oxford University Press, Oxford.

Cooper, L. (1964) Heuristic methods for location-allocation problems, *SIAM Review*, **6**: 37–52.

Dijst, M. and Vidacovic, V. (1995) Individual action space in the city: an integrated approach to accessibility, Paper presented at the conference 'Activity-based Approaches: Activity Scheduling and the Analysis of Activity Patterns', May 25–28 1995, Eindhoven.

Fischer, M.M. (1999) Intelligent GI analysis, in Stillwell, J., Geertman, S. and Openshaw, S. (eds) *Geographical Information and Planning*, Springer, Heidelberg, pp. 349–368.

Floor, H., de Jong, T. and Ritsema van Eck, J.R. (2001) The adaptation of optimization techniques for usage in combination with Planning Support System; including a case study of siting an Islamic secondary school in the city of Utrecht, The Netherlands, Paper presented at the Cupum 2001 Conference, Honolulu.

Fotheringham, A.S., Densham, P. and Curtis, A. (1995) The zone definition problem in location-allocation modelling, *Geographical Analysis*, **27**: 60–77.

Geertman, S. (1999) Geographical information technology and physical planning, in Stillwell, J., Geertman, S. and Openshaw, S. (eds) *Geographical Information and Planning*, Springer Verlag, Berlin, pp. 69–86.

Geertman, S.C.M. and Bosveld, W. (1990) A GIS-based model for goals achievement analysis and design, *Proceedings of the First European Conference on Geographical Information Systems*, 356–65, EGIS, Utrecht.

Geertman, S., Jong, T. de and Wessels, C. (2002) Flowmap: a support tool for strategic network analysis, in Geertman, S. and Stillwell, J. (eds) *Planning Support Systems in Practice*, Springer, Heidelberg, pp. 155–76.

Geertman, S.C.M. and Ritsema van Eck, J.R. (1995) GIS and models of accessibility potential: an application in planning, *International Journal of Geographical Information Systems*, **9**: 67–80.

Hansen, W.G. (1959) How accessibility shapes land use, *Journal of American Institute of Planners*, **25**(1): 73–6.

Hilberts, H.D. and Verroen, E.J. (1992) Het beoordelen van de bereikbaarheid van lokaties. *INRO-TNO Rapport INRO-VVG 1992-21*, Delft.

Jong, T. de and Ritsema van Eck, J. (1997) Threshold surfaces as an alternative for locating service centers with GIS, in *JEC-GI '97 Proceedings. Vol 1*, IOS Press, Amsterdam, pp. 799–808.

Jong, T. de, Ritsema van Eck J. R. and Toppen F. (1991) GIS as a tool for locating service centers, *Proceedings of the Second European Conference on Geographical Information Systems*, EGIS, Utrecht, pp. 509–19.

Jong, T. de, Maritz J. and Ritsema van Eck, J. (2001) Using optimisation techniques for comparison of the accessibility criteria of facility siting scenarios; a case study of siting police stations in South Africa's Bushbuckridge area, in *Proceedings of the 4th Agile Conference on Geographic Information Science*, Masaryk University Brno, Czech Republic.

Maranzana, F. (1964) On the location of supply points to minimize transport costs, *Operational Research Quarterly*, **15**: 261–70.

Moseley, M.J. (1979) *Accessibility: The Rural Challenge,* Methuen, London.

OCG (2002) Implementations of OCG specifications. http://www.opengis.org/cgi-bin/implement.pl.

Openshaw, S. (1997) Building fuzzy spatial interaction models, in Fischer, M.M. and Getis, A. (eds) *Recent Developments in Spatial Analysis. Spatial Statistics, Behavioural Modelling and Computational Intelligence*, Springer, Heidelberg, pp. 360–83.

Ottens, H. (1993) The application of GIS technology in Europe, in Harts, J., Ottens, H. and Scholten, H. (eds) *EGIS '93 Proceedings*, EGIS Foundation, Utrecht, pp. 1–8.

Ritsema van Eck, J.R. and de Jong, T. (1996) Data points in GIS-based network analysis, *Stepro Working Paper*, Faculty of Geographical Sciences, University of Utrecht, Utrecht.

Robinson, G.M. (1990) *Conflict and Change in the Countryside*, Belhaven Press, London.

Steinitz, C. (1993) Geographical information systems: a personal historical perspective, the framework for a recent project, and some questions for the future, in Harts, J., Ottens, H. and Scholten, H. (eds) *EGIS '93 Proceedings*, EGIS Foundation, Utrecht.

Tomlinson, R.F. (1987) Current and potential uses of geographical information systems: the North American experience, *International Journal of Geographical Information Systems*, **1**: 203–18.

Törnqvist, G., Nordbeck, S., Rystedt, B., and Gould, P. (1971) *Multiple Location Analysis*, Lund Studies in Geography Series C, Number 12, Lund University, Lund.

Winter, S. (2001) Ontology – buzzword or paradigm shift in GI science? *International Journal for Geographic Information Science*, **7**: 587–90.

13

A Probability-based GIS Model for Identifying Focal Species Linkage Zones across Highways in the Canadian Rocky Mountains

Shelley M. Alexander, Nigel M. Waters and Paul C. Paquet

Abstract

Road fragmentation is a concern for the survival of sensitive species within and adjacent to protected areas in the Rocky Mountains of North America. The Idrisi Decision Support Module and Dempster–Shafer theory are used to establish potential linkage zones for lynx and wolves across the unmitigated sections of major highways in Alberta. Wolves and lynx acted as focal species for large and medium carnivore guilds, respectively. Linkage zones, derived from probability surfaces, showed minimal overlap between these two species. This suggests that the spatial requisites of crossing structures for wildlife are guild-specific at a minimum and mitigation should be designed to capture multiple guilds. Spatial probability modelling is a powerful method for representing habitat associations, determining ecologically important linkage zones and generating testable hypotheses. The uncertainty inherent in resource management and planning is reduced and managed by using such innovative spatial modelling tools.

13.1 Introduction

Road, rail, power and telecommunication networks create a grid upon landscapes commensurate with the extent of human endeavour. These networks are vital to our economy and provide a means to expand the human experience and improve travel

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

convenience. Associated with these 'benefits' is the ecological cost of habitat fragmentation, which occurs when contiguous patches of habitat are divided into smaller, isolated units (Andren, 1994). Most wildlife species have evolved resilience, to the extent that they can endure ecological change caused by natural fragmentation (e.g., floods, rockslides, rivers) (Weaver *et al.*, 1996). However, human fragmentation tends to occur with an intensity and extent unprecedented in natural systems, and some species fail to persist when confronted with this human disturbance.

Of the previously listed networks, roads carrying traffic may have the greatest cumulative disturbance effect. In addition to habitat loss and isolation, vehicular traffic can kill animals that attempt to cross between isolated habitats, and may cause others to avoid crossing roads for fear of detection. In ecology this phenomenon is termed 'the barrier effect' (Noss and Csuti, 1997). The barrier effect can alter animal community composition, create meta-populations, reduce biological diversity and increase the threat of extinction (Trombulak and Frissell, 2000; Spellerberg, 1998; Forman and Alexander, 1998; Bascompte and Solé, 1996).

In recent years, 'road ecology' has become more prominent in ecological investigations, as shown by conference series such as the International Conference on Wildlife Ecology and Transportation (Evink *et al.*, 1998a,b, 1999) and the 'Roads and Rails' workshop series (Columbia Mountain Institute of Applied Ecology, Revelstoke, British Columbia 1997–2001). Reducing wildlife mortality on roads has been one focus of previous studies, as roadkill can pose a serious threat to species with small populations or low resilience. Low resilience results from life history traits, which include, among others, late age of first reproduction, low reproductive rate, and low fecundity (Weaver *et al.*, 1996). For example, a species that cannot reproduce rapidly enough to offset road mortality will suffer population decline. Over time and across space, such a decline could force a population into extinction. Road fragmentation has been identified as an extinction concern for carnivores in and adjacent to protected areas in the Rocky Mountains (Noss *et al.*, 1996). Specifically, roadkill has been noted as one of the largest known sources of mortality for wolves within and outside of protected areas of the Canadian Rockies (Noss *et al.*, 1996; Banff-Bow Valley Study, 1996).

The creation of linkages between isolated patches of habitat is one method for improving biotic exchange across barriers (Forman and Alexander, 1998; Swart and Laws, 1996; Fahrig and Merriam, 1995). In the present study region (described below) few sources of empirical data document the effects of traffic on wildlife movement or the spatial design requisites for placing linkages (highway mitigation) across highways. The present research was designed to address these deficiencies in knowledge and understanding of landscape connectivity for multiple species in the Canadian Rocky Mountains.

In this study, wildlife movement was documented in four different road fragmented environments in the Rocky Mountains of Alberta. Wildlife movement was measured, using snow tracking surveys, across and adjacent to the Trans-Canada Highway (TCH) (14 000 vehicles: Annual Average Daily Traffic, AADT) and the 1A Highway (4000 AADT) in Banff National Park, and the Highway 40 (5000 vehicles: Summer Average Daily Traffic, SADT) and Smith Dorrien/Spray Trail (2500 SADT) in Kananaksis Country. In the study area (described subsequently), little empirical

Table 13.1 *Summary of objectives*

Objectives	General hypotheses/methods
Identify landscape attributes predicting wildlife movement for individual species	Species preferences are defined by a probability of relationship to habitat attributes
Test for species interactions	Species interactions will be analysed using GIS
Create a linkage zone model using evidential reasoning	Application of IDRISI GIS and D–S weight-of-evidence modelling

research exists on multiple-species crossing behaviour before road expansion, on the effects of traffic on wildlife movement or on the spatial requisites for crossing structures.

The modelling effort presented in this chapter is unique. This work provides an original application of Bayesian (Dempster–Shafer, D–S) inference to determine optimal road crossing sites (i.e., linkage zones across highways), generates additional working hypotheses about the efficacy of wolves and lynx as focal species, and creates a platform for the integration of spatial wildlife data in highway planning. The matrix in Table 13.1 specifies the objectives of the research relative to developing the linkage zone model, and presents the general hypotheses or methods by which objectives were met.

The following section explores the literature pertaining to road fragmentation effects, wildlife habitat preference, modelling habitat selection using GIS and the integration of Bayesian analysis in defining wildlife linkage zones across highways.

13.2 Background Literature

13.2.1 When Roads and Wildlife Corridors Collide

This research focused on the interface between wildlife movement corridors and highways, and the identification of optimal sites for placing wildlife crossing structures (i.e., mitigation). A number of key literature reviews have summarised research on road-related, ecological disturbance effects, including Trombulak and Frissell (2000), Jackson (1999), Spellerberg (1998) and Forman and Alexander (1998). *The Proceedings of the International Conferences on Wildlife Ecology and Transportation*, ICOWET (e.g., Evink *et al.*, 1998a,b, 1999) offer examples of recent advances. Research trends have focused on mortality, pollution/contamination effects, and mitigation.

Highway mitigation, which refers to the implementation of measures that reduce wildlife mortality and/or facilitate movement across highways, and spatial trends in mortality remain at the forefront of road fragmentation studies. Mitigation effectiveness, the barrier effect and placement requisites remain little studied (Spellerberg, 1998; Forman and Alexander, 1998). Community studies that address movement

requisites of multi-species are rare or non-existent, with the exception of the present study. The use of GIS has featured prominently in the recent *Proceedings of ICOWET* (Evink, 1999) and non-spatial decision support systems have become focal (Alexander and Waters, 1999; Smith, 1999; Maurer, 1999).

13.2.2 GIS Applications in Mitigating Road Effects

Habitat modelling has been advanced by the advent of geographic information systems (GIS), which greatly increase the scale, complexity and precision of spatial analysis. GIS is an excellent medium for predicting patterns of biological diversity, identifying areas of conservation significance and assessing the habitat potential of unstudied sites (Lenton *et al.*, 2000). In the region, studies have applied GIS in defining habitat selection by wildlife (e.g. Alexander and Waters, 2000a; Paquet, 1993). Relatively few have examined the spatial attributes of road crossing sites (Alexander and Waters, 2000b; Singleton and Lehmkuhl, 1999). Beyond the present research, no studies have examined multi-species interactions or predicted spatial linkages at a community level. GIS analysis has been used to identify mortality hotspots in determining placement of highway mitigation, although mortality hotspots have never been demonstrated to reflect crossing preferences. They simply measure mortality, not crossing attempts. Intuitively, these sites may reflect road alignment or curvature effects, rather than animal preferences. These road features may increase the likelihood of vehicle–wildlife collision and give the illusion that more animals cross at these sites.

Modelling the spatial requisites of movement corridors can be used to ameliorate the barrier effect, by confirming mitigation sites that have a high probability of selection for use by wildlife. Jackson (1999) summarises the types of mitigation popularly employed including modified drainage culverts, wildlife/drainage culverts, upland culverts, oversize stream culverts, expanded bridges, viaducts, wildlife underpasses, wildlife overpasses and fencing. Efficacy studies have demonstrated that species tolerances to structural design are highly variable (Jackson, 1999). Expanded bridges (open spans), viaducts and wildlife overpasses appear to be the most effective across communities in the Rocky Mountains of Banff National Park (Alexander and Waters, 1999). Placement of mitigation is argued as the most important component of mitigation success, followed by mitigation design. Jackson (1999 and all included references) lists the following efficacy criteria: size and openness, placement, human disturbance, substrate, vegetation cover, moisture, hydrology, temperature and light.

Barrier effects and mitigation efficacy vary by species (Trombulak and Frissell, 2000). For terrestrial mammals, GIS analysis helps identify the spatial attributes (habitat selection) likely to be suitable for movement corridors for individual species, guilds or community assemblages of species. Habitat selection is detailed in the next section and the use of GIS to infer habitat selection is presented in Section 13.3.3.

13.2.3 Habitat Selection: Spatial Attributes that Define Movement Corridors

Habitat selection modelling is predicated on the assumption that wildlife 'select', 'prefer' or 'avoid' specific habitat attributes (e.g., they may select closed vegetation for security or avoid roads). The selection of habitat may be inferred by examining use versus the availability of habitat. Peek (1986) identified 'ultimate factors' and 'proximate factors' as two facets of habitat selection. Ultimate factors include food, cover or shelter from enemies and adverse weather, nesting or denning sites and inter-specific competition. Proximate factors may include stimuli of landscape, terrain and competitors.

To exemplify, habitat selection by wolves was described as 'a complex interaction of physiography, security from harassment, positive reinforcement (e.g., easily obtained food), population density, available choice, and distribution theory' (Carroll *et al.*, 2000). In a multi-species environment, such selection must be explored for individual species. Alternatively, the interactions between species may be investigated to determine the degree to which a given focal species may act as a surrogate for others. In this study, relationships to habitat metrics and the spatial interaction of species were investigated for each study species. These results are the topic of a different paper (Alexander *et al.*, 2003, in prep). The latter analysis allowed inference on which species is the 'best surrogate species' and identified three focal species, representing three functional ecological scales: marten – micro-scale; lynx – meso-scale; and wolf – macro-scale.

13.2.4 Bayesian Approaches in Ecological Modelling

Ecologists are revisiting the use of classical statistics in the evaluation of ecological phenomena (Ellison, 1996). Ellison (1996) argues that this introspection is due to the fact that the central assumptions of classical statistics are violated in ecology: true randomisation is difficult, replication is often small, misidentified or nonexistent and ecological experiments rarely are repeated independently. Alternatives to classical statistical tests have been presented, and include variants on probability theory (e.g., Bayesian probability, Dempster–Shafer theory) and information theory (e.g., Bayesian information criterion, Akaike's information criterion) (Wade, 2000; Anderson *et al.*, 2000).

The application of information and Bayesian theory is neither new, nor confined to conservation biology, but is popular in current ecological research (Anderson *et al.*, 2000). For examples of relevant ecological applications, see Bartel (2000), Anderson *et al.* (2000), Dragicevic and Marceau (2000), Urbanski (1999), Ellison (1996) and Pearl (1987). The integration of Bayesian theory into GIS for use in wildlife management and ecosystem planning is far less developed. One notable example is the work of Lenton *et al.* (2000), which predicted species biodiversity in Africa.

A central assumption of these alternative approaches is that testing null hypotheses adds little to scientific understanding (Anderson *et al.*, 2000; Ellison, 1996). Ellison (1996) argued that 'we normally expect to accept a single ecological alternative hypothesis by rejecting a statistical Ho: otherwise we would not have done the

experiment in the first place'. Anderson *et al*. (2000) add that nearly all null-hypotheses are false on *a priori* grounds and that *p*-values are dependent on sample size (i.e., with a large enough sample one can always reject a null-hypothesis), even though such statements might be controversial. They argue that the former points to the difference between statistical significance and biological importance. Testing prior hypotheses (Ellison, 1996) or multiple working hypotheses (Anderson *et al*., 2000) has been suggested as an alternative to classical statistical approaches. Prior hypotheses are based on the current state of knowledge about underlying distributions of the data. These priors are compared with new information to see if the new information adds to the current knowledge. Multiple working hypotheses (Anderson *et al*., 2000) exhaust all possible combinations of explanatory variables for a given phenomena, such as species distribution. In this case, the range of hypotheses (i.e., combination of predictive variables) is compared using an information criterion, to provide a rank that indicates the optimal combination of variables.

13.2.5 Integrating Bayesian Theory and GIS in Habitat Assessment

Research on wolves in Banff National Park (BNP) indicated that landscape features, such as slope, aspect, elevation and vegetation type can influence movement (Paquet, 1993). For example, steeper slopes require greater energetic expenditure for an animal to cross and may result in the displacement of movement around a steep feature. Alternatively, dense vegetation may provide escape or hiding cover and may change a movement path when an animal fears detection or predation. A disproportionate use of physiographic attribute classes is assumed to reflect a 'preference' for those landscape elements. A 'preference' suggests that the animal is selecting a specific attribute class over another. Attribute classes are aggregated values within an attribute. For example, for the variable 'aspect', which ranges from 0 to 360 degrees, the 'aspect classes' are groupings of aspect values, such as those labelled north, which includes all aspect values between 337.5 and 22.5 degrees.

The statistical evaluation of preference is often handled with techniques, such as logistic regression, which treat species occurrence as a binary value representing presence as (1) or absence as (0) (Pearce and Ferrier, 2000; Bartel, 2000). Although there are variations in analytical procedures, often a site's presence and absence have been defined by stating a threshold count of 'wildlife sign' that indicates presence. Sites that have a number of observations of a species that exceed a threshold are considered 'presence' sites (1), those below the threshold are 'absence' sites (0). For instance, if the greatest number of species observed for all sites surveyed is 10, the analyst may specify that any sites with greater than five observations indicate presence (value 1) and any sites where there were fewer than five are discarded from the sample (value 0).

Binary data (1,0) are then subjected to techniques such as logistic regression. In this approach, the imposed threshold that defines the difference between presence and absence is arbitrarily defined: a value of 1 might be five tracks, 20 tracks or any number in between; a value of 0 might be assigned to any sites with fewer than five tracks. This assumption that sites with fewer than five tracks are not 'used' may be

an ecologically incorrect assumption. Moreover, Conroy and Noon (1996) argued that although presence provides affirmation of occurrence, absence does not provide any information – it could be an artefact of the sampling effort.

Binary approaches do not represent adequately the spectrum of preferences observed within landscape attributes, may needlessly discard information and are inherently indefensible because, in most cases, habitat attractiveness and preference for movement corridors exists as a continuum. For example, species may use sub-optimal habitat during regular foraging or movement (Tischendorf and Fahrig, 2000) because habitat is most often heterogeneous. In this study, highly suitable lynx habitat was spatially discontinuous and lynx traversed areas of low habitat suitability when moving between highly suitable patches. As an alternative to binary approaches, detailed above, we used a Bayesian classification approach that treats habitat use as a continuous probability. This approach considered relative track density by attribute classes and the probability of use was indicated by the weight-of-evidence of a trend in habitat use, based on the difference between observed and expected track counts for each attribute class. Thus, instead of specifying a threshold value, all track density values are considered important in predicting the presence of a species.

Fuzzy set theory is well suited to problems that deal with habitat analysis, where there is 'no clear separation between areas that are suitable and those that are not' (Bartel, 2000; Eastman, 1999; Zadeh, 1965). Although typically applied to spatial boundaries, such as problems related to soil classification (Dragicevic and Marceau, 2000), fuzzy set theory is extended easily to describe the fuzziness in species associations with habitat types. Fuzziness, as a measure of inconclusiveness (i.e., lack of exact separation in selection), is handled well with Bayesian probability or its variant D–S theory (Eastman, 1999). D–S theory was developed by Bayes' researchers to handle absence of information (Gordon and Shortliffe, 1985). The distinction between Bayesian probability theory and D–S theory is elaborated in the methods section below.

13.3 Study Area and Methodology

13.3.1 Introduction

Research was conducted in Banff National Park (BNP) and Kananaskis Country (KC), Alberta. Figure 13.1 shows the study area. BNP and KC are approximately 110 kilometres west of Calgary. BNP is 6640 square kilometres in area and is the most heavily visited national park in Canada with over 5 million visitors per year (Banff Bow Valley Study, 1996). KC is a 4250 square kilometres Forest Landuse Zone that is reserved for hunting, ranching, resource development, recreation and tourism activities. Peter Lougheed Provincial Park (PLPP) is a 514 square kilometres park in the northwestern corner of KC, where hunting and resource development are not permitted, with the exception of ongoing maintenance of existing hydro-electric projects. The study region is characterised by rugged mountainous terrain, steep valleys and narrow (2–5 kilometres), flat valley bottoms.

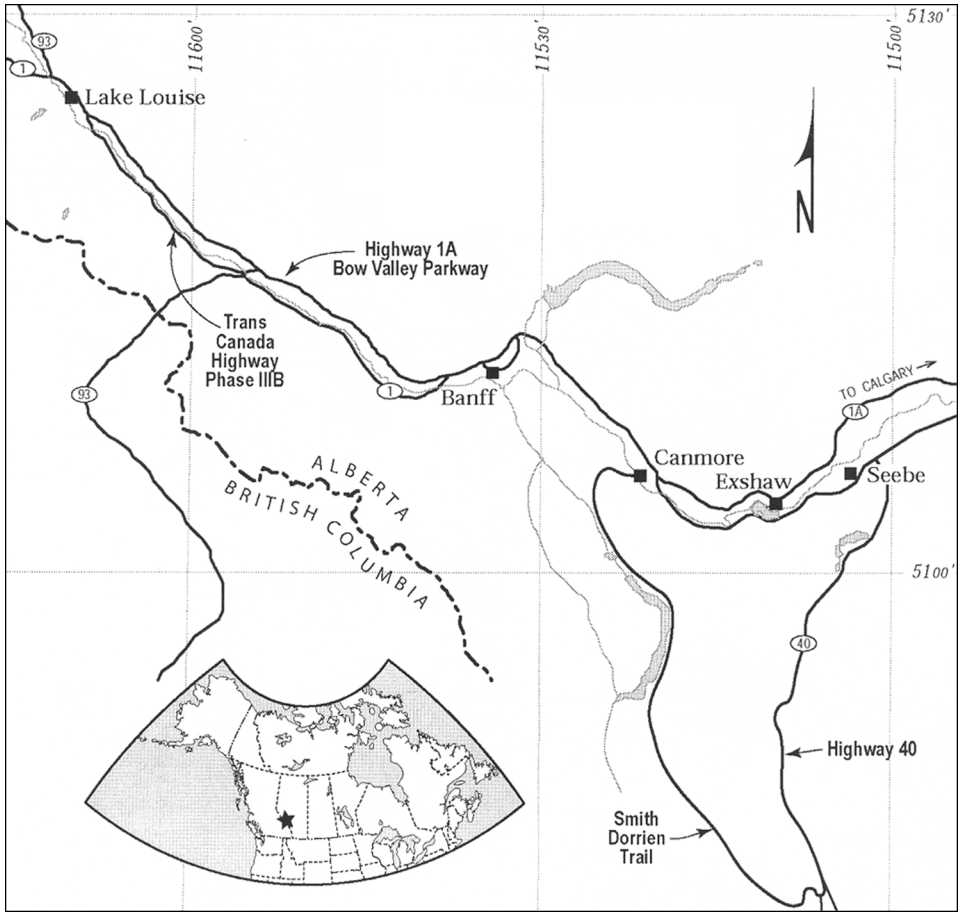


Figure 13.1 *Study area map for wildlife linkage zone model*

In BNP, track data were collected on the non-twinned (two lane), unfenced section of the Trans-Canada Highway (TCH) from Castle Junction to the British Columbia border, and along the entire length of the Bow Valley Parkway (1A). This section is termed Phase IIIB of the TCH 'twinning project' in BNP. Apart from the TCH and 1A the study area transportation corridor contains the Canadian Pacific Railway (CPR) main line, Trans-Alta power line, and a few seasonal campsites and tourism facilities. The BNP study roads do not have structures that facilitate the movement of wildlife over or under the highway or to reduce wildlife–vehicular collisions (i.e., it is not fenced). The previous structures are termed 'highway mitigation'.

In KC, research was conducted on Highway 40 (Hwy 40) and the Smith Dorrien/Spray Trail (Figure 13.1). Highway 40 is a paved two-lane highway with moderate traffic volumes, and the Spray Trail is a gravel two-lane road primarily offering access to primitive backcountry activities. The service areas of both

transportation routes are under pressure for recreation and tourism development, which will result in increased traffic volumes. With the exception of wildlife crossing signs, neither route is mitigated for wildlife.

13.3.2 Methodological Framework

Figure 13.2 shows the methodological framework for production of the final linkage zone model. This image is useful for understanding each of the following steps of model creation described below.

Research was conducted during the winter months (from November to April) over three years (1997–2000). Roads were surveyed between 18 and 48 hours after the end of a snowfall. Tracks in snow were observed from a field vehicle, while driving slowly (15–20 km/h) and verified on foot (Van Dyke *et al*., 1986). Tracks entering or exiting the road right of way were recorded for 13 terrestrial species, including the two focal species presented in this paper, wolf (*Canis lupus*) and lynx (*Felis lynx*). Tracks of other species were observed but not reported herein, for coyote (*Canis latrans*), fox (*Vulpes vulpes*), cougar (*Puma concolor*), bobcat (*Felis rufus*), marten (*Martes americana*), fisher (*Martes pennanti*), wolverine (*Gulo gulo*), elk (*Cervus elaphus*), moose (*Alces alces*), sheep (*Ovis canadensis*) and deer (*Odocoileus virginianus* and *Odocoileus hemionus*). Tracks were only observed in the right of way, as vegetation obscured tracks beyond this area. It was assumed that tracks entering this area were of animals attempting to cross the road.

Data collected at crossing sites included an Easting/Northing (UTM), which was collected with a handheld GPS (Garmin II, non-differentially corrected). Other data collected included species type, number of individuals and a range of behavioural parameters that are not analysed here. Repeat road surveys were conducted, as conditions permitted, approximately 3–4 days after initial surveys until the next new snowfall. Transects of one-kilometre length were fixed perpendicular to each road treatment and surveyed for tracks of the above study animals and in addition for hare, weasel and squirrel. Again, only the results of wolf and lynx are used for the probability mapping section in this chapter. Transects are straight survey lines across a study area, in this case marked every few metres with flagging tape on trees. Flagging tape was placed every 50 metres, which allowed researchers to repeat survey exactly the same area. Forty transects were surveyed randomly in Banff National Park and twenty in Kananaskis Country, between 18 and 108 hours after snow and immediately following each road survey (Thomson *et al*., 1988). Transects were surveyed on foot and required an extended survey period relative to the road survey. A one-metre resolution differential GPS (Trimble Pathfinder) was used to collect UTM coordinates for every 50 metre interval start point, for every transect. These data were used to geo-reference track counts for import into the Idrisi GIS (Eastman, 1999) for spatial modelling. Data were standardised by sampling effort, to minimise the effects of sampling bias. For example, if one site was surveyed five times and another three times, the total track count for each site was divided by the number of samples per site (i.e., five and three) to standardise between the transect sites.

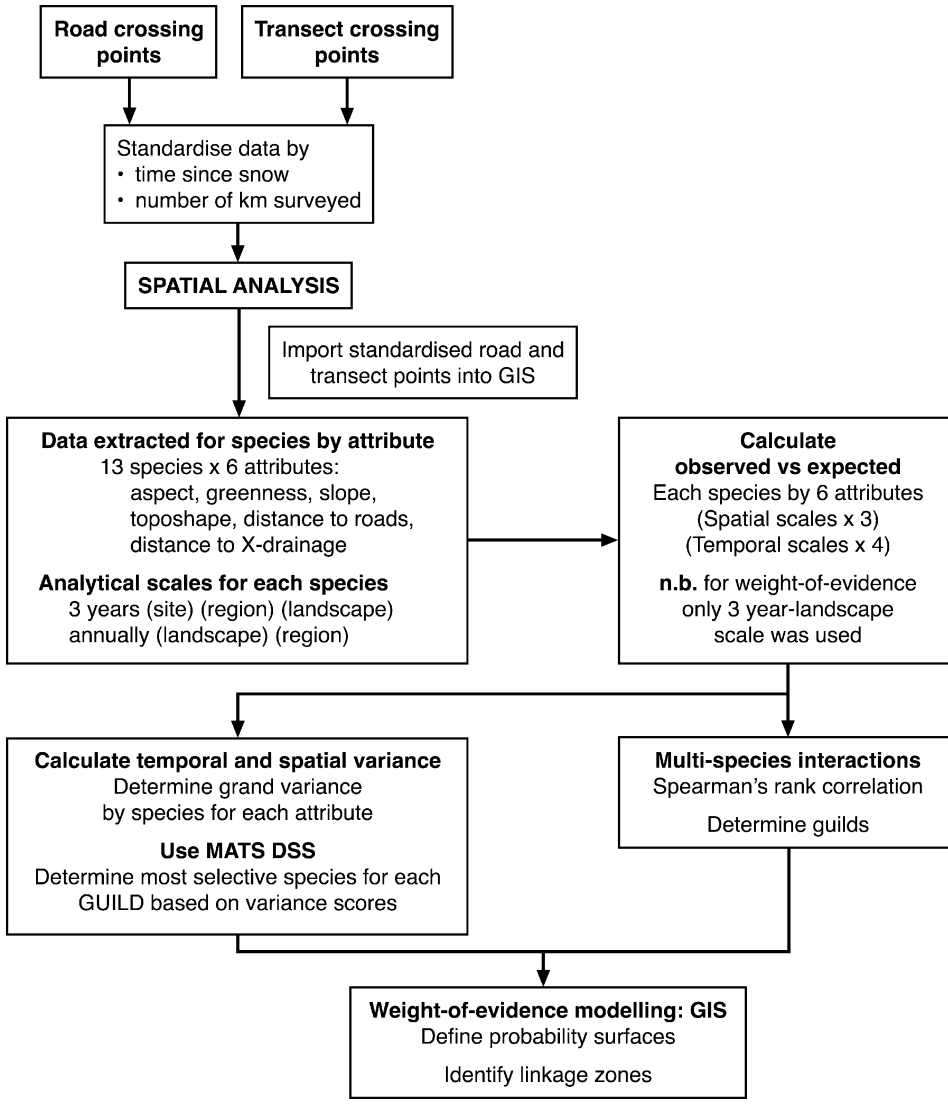


Figure 13.2 *Flowchart of model development*

13.3.3 Inferring Selection from Use Density

Density of tracks by attribute classes (i.e., track density in each aspect class) was determined by using the EXTRACT function in Idrisi (Eastman, 1999). With the EXTRACT function, the total track density by each attribute class is summarised in table format. These data tables were saved as text files and imported into Excel (MS Office 97).

A study frame (i.e., the spatial extent of area actually surveyed) was used to extract the distribution of class types within each attribute (e.g., north class or aspect). The study frame was a Boolean image, in which any site with a value of one (1) was an area that was sampled, any site with a zero (0) value was not sampled. The use of the frame is essential to avoid issues of 'framing bias' and 'ecological fallacy', which can yield incorrect results about habitat selection (Verbyla and Chang, 1994). The output from this operation provided a summary of total cell counts by attribute class. The extracted study frame values (i.e., the distribution of attributes samples) were used to determine the proportional representation of each attribute class in the sampling frame, which was then used to determine the expected number of tracks for each attribute class, as a proportion of total tracks observed. A value was then derived to indicate over- and under-use of attribute classes. This value was determined by subtracting the expected from the observed track density and dividing by the expected density (*N*) for each sub-class. Table 13.2 shows the distribution of marten points by aspect class for the three years. Column numbers are indicated at the top.

The first column of Table 13.2 indicates the attribute classes (here using the example of aspect classes). In this case, the aspect classes represent groupings of aspect values that are similar. For example see the row North (column one) in Table 13.2, where north aspect is defined as all slopes that face the directions 337.5 to 22.5 degrees around north (0 degrees); northeast is 22.5–47.5 degrees, and so on. Column two shows the total number of tracks observed by aspect class, as tabulated from the EXTRACT function in Idrisi. Column three is the expected track count, the derivation of which is discussed below. The fourth column shows the total sampled raster cells for each aspect class, followed by the proportional representation of the total number of cells (column five). The cells represent locations on the landscape that were sampled. The attribute value on any cell is an average over the cell (or pixel) resolution, which in this case is 30 metres. The proportion in column five was used to derive the expected track count (column three) by multiplying the proportion by the total number of marten tracks observed. Column five sums to 1.00 when the error

Table 13.2 *Species, track counts by aspect class*

1 Aspect Class	2 Observed	3 Expected	4 Total Aspect	5 Prop	6 Diff(o-e)	7 Over/Under use	8 RANK
Background	67	0	0		67	0	
North (337.5–22.5)	219	263.5402	8031	0.076234	–44.5402	–0.16901	7
Northeast	556	335.1761	10214	0.096956	220.8239	0.658829	1
East	285	205.7523	6270	0.059518	79.24768	0.385161	3
Southeast	184	145.3721	4430	0.042052	38.62794	0.265718	4
South	466	281.1948	8569	0.081341	184.8052	0.657214	2
Southwest	730	642.4592	19578	0.185840	87.54083	0.136259	5
West	305	662.7062	20195	0.191699	–357.706	–0.53977	8
Northwest	92	286.9047	8743	0.082992	–194.905	–0.67934	9
Flat – No aspect	553	633.8944	19317	0.183365	–80.8944	–0.12761	6
Total	3457	3457	105347				

point (67 points in the Background class 0) are considered. The difference between observed and expected values appears in column six, and finally, the observed minus expected value is divided by the total expected (column three) in each class, to produce the values in the seventh column. This provides a relative index of observed versus expected use for comparison between scales of analysis and between species.

In Table 13.2, over-use is indicated in column seven by a positive value and under-use by a negative value. It was assumed that over or under-representation of tracks by attribute class has biological significance and represents a continuum of use by species of attribute classes. Column seven values were ranked by their relative level of over- and under-use. Over-use defines a greater amount of use of a habitat class than expected, based on its availability. Alternatively, under-use implies less use than expected, based on availability. To exemplify, in column eight, the highest rate of under-use (rank 9) is found in the class northwest, and the highest rate of over-use (rank 1) is found in class northeast. The rank scores in column eight provide a means to treat the nominal data as a linear variable, when determining the probability of detection for each attribute class and creating a probability surface in the GIS, for the aspect variable. This procedure was carried out for all 13 species sampled (refer to Section 13.3.2).

13.3.4 Dempster–Shafer Weight-of-Evidence Modelling

A univariate analysis of track occurrence by landscape attribute indicated that the probability of detecting carnivores is a function of slope, aspect, greenness and toposhape. Two other variables, distance to cross drainages and distance to roads were evaluated, but yielded no trend in selection or avoidance. Slope and aspect are surfaces derived from a DEM (30 metre resolution). Greenness was derived from a Landsat TM image using a Tasselled Cap (TC) analysis. TC is a simple linear index produced by taking the weighted average of the input bands from a TM Image (Richards and Jia, 1999). TC Greenness is proportional to vegetation productivity, where higher greenness values are assumed to represent better suitability. Toposhape is a measure of surface curvature that defines the landscape as a series of critical features, including peaks, ridges, saddles, flat areas, ravines, pits and five types of hillsides – convex, saddle, slope, concave and inflection (Eastman, 1999; Warntz and Waters, 1975).

Probability functions describing optimal conditions for habitat linkage zones were derived from the index of selection and avoidance. These are represented by the over-use and under-use observed for certain habitat classes, such as those shown in Table 13.2, previously. The decision rules and probability functions are described below.

Probability surfaces were constructed for wolves and lynx using the FUZZY module in Idrisi (Eastman, 1999). Although 13 species were studied in this research, only the probability surfaces for wolves and lynx are presented in this chapter. For habitat relationships and probability maps relating to other species, refer to Alexander (2001). An example of how decision rules are translated to probability curves is shown in Figure 13.3, which shows the aspect classes and ranks, which were identified previously in Table 13.2 (see also the discussion in Srinivasan and Richards,

Aspect	Rank	Sign
North	7	-
Northeast	1	+
East	3	+
Southeast	4	+
South	2	+
Southwest	5	+
West	8	-
Northwest	9	-
Flat	6	-

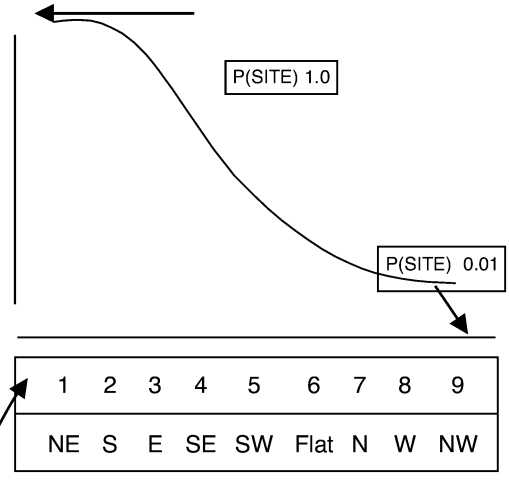


Figure 13.3 *Species use by aspect*

1990). The sign indicates the selection value (see description of Table 13.2), and the rank at which the sign changes may be used as an inflection point on the probability curve.

In the Idrisi GIS software, fuzzy sets can be expressed by sigmoidal, j-shaped, linear and user-defined probability curves. The curve indicated in Figure 13.3 is a monotonically decreasing sigmoidal function. For each probability curve, an associated hypothesis must be stated. From the curve, a hypothesis may be stated that the probability of detecting a marten site *p*(site) decreases with increase in rank.

The BELIEF module in Idrisi was used to aggregate all the 'lines of evidence' (slope, aspect, greenness and toposhape – as represented by their various hypotheses, e.g., species are more likely to be found on slopes with certain aspects as described in Figure 13.3). BELIEF aggregates these various 'lines of evidence by applying rules of combination based on Dempster–Shafer (D–S) Weight-of-Evidence modeling' (Eastman, 1999). D–S theory is a variant of Bayesian probability theory (Srinivasan and Richards, 1990), which allows us to decompose a large body of evidence into its component parts. The importance or influence of these component parts can then be assessed as probability statements that subsequently can be recombined to determine the probability of finding a particular species at a given location. If this species is then found in association with others, such as a guild of wildlife species, these locations, when adjacent to roads, may be especially attractive as sites for the mitigation of highway induced wildlife mortality. In this case, the evidence was aggregated using BELIEF, to create a linkage zone model, which indicates regions with the highest probability of detecting a given species. These high probability sites should indicate optimal locations for placing crossing structures.

D–S theory defines hypotheses in a hierarchical structure. For example, if the frame of discernment (Srinivasan and Richards, 1990), or decision frame using Eastman's

(1999) terminology, contains two hypotheses [A,B], the structure of the hypotheses for which D–S incorporates evidence includes all possible combinations of these hypotheses i.e. [A], [B], [A,B]. All three hypotheses are recognised because often evidence supports one or other of this combination of hypotheses. D–S theory employs the following rule of combination:

$$\begin{aligned} m(Z) &= \sum m_1(X) \times m_2(Y) \text{ when } X \cap Y = Z \\ &= 1 - \sum m_1(X) \times m_2(Y) \text{ when } X \cap Y = \Phi \end{aligned} \tag{13.1}$$

If

$$\sum m_1(X) \times m_2(Y) = 0 \text{ for } X \cap Y = \Phi \tag{13.2}$$

Then

$$m(Z) = \sum m_1(X) \times m_2(Y) \text{ when } X \cap Y = Z \tag{13.3}$$

where $m(Z)$ = basic probability assignment (the mass of support) for the hypothesis [Z] and X and Y are independent variables (Eastman, 1999; Gordon and Shortliffe, 1985).

D–S theory varies from Bayesian probability theory, in that it recognises the existence of ignorance due to incomplete information – as indicated by the previous non-site hypotheses. In Idrisi, D–S theory is 'able to determine the degree to which we, with our current state of knowledge (evidence), cannot distinguish among a subset of hypotheses (a condition known as ambiguity) or among all hypotheses (a condition known as ignorance)' (Eastman, 1999). For example, the mass of support for the probability of detecting wolves by slope is based on the empirical evidence that wolves prefer lower slopes. The weight of this evidence tends to support the non-use of steep slopes (>30 degrees), given the paucity of use in sites with higher slope steepness. Low slope alone, however, may not be enough to explain preference for shallow slopes, thus the hypothesis is expressed as a non-site (i.e., high slopes lack wolves) rather than as a site hypothesis that would state directly that low slopes have a high probability of being wolf sites. By contrast, all other lines of evidence, toposhape, greenness and aspect, were used to support the site hypothesis.

The BELIEF module is used to construct a knowledge base from the input data (i.e., all four predictive attributes for each species) and the hypothesis that each data layer supports (i.e., site or non-site, as discussed above). All four lines of evidence are integrated into a knowledge base by combining the various lines of evidence and producing basic probability assignments for each of the hypotheses. The BELIEF module will produce maps of the probability that an animal will be found at a given location based on the three lines of evidence that support the habitat hypothesis. The line of evidence, in this case slope, which supports the non-habitat hypothesis, produces a 'plausibility' map. The difference between the belief and plausibility maps produces the belief interval that may be useful for determining where to look for new sources of support for the habitat hypothesis. In the present chapter we are

primarily concerned with our existing evidence for species existence with respect to their habitat properties.

A first iteration of probability surfaces was constructed using four attributes (aspect, greenness, slope and toposhape). A lack of clear linkages was evident, and maps were re-analysed. A second iteration of probability maps was then developed without greenness (i.e. with three attributes: aspect, slope and toposhape). Greenness was removed because it is highly correlated for all species, and the initial probability image showed little variation in greenness within one kilometre of roads. Finally, data assimilation techniques were used to improve the identification of corridors (linkage zones) in a third iteration of maps. Data assimilation techniques involve using regional data to 'fine-tune' global models. Climatologists frequently use this method to enhance global weather models (Khattatov *et al*., 2000). In the third iteration of the linkage zone model, known species sites were 'assimilated' to calibrate the final model geographically. The assimilation was done by adding a layer depicting distance from known sites of the focal species, represented by a decay function where suitability of habitat decreases with distance from observed sites. This calibration was used because the first probability surface was generated using habitat associations from the three-year landscape scale analysis, a global model.

13.4 Results and Discussion

13.4.1 Introduction

In this section the results of three iterations of the linkage zone model are presented. The initial belief surface combined the four variables that showed non-uniform use by the focal species (i.e., slope, aspect, greenness and toposhape). A second surface was developed in response to an examination of the first surface. Specifically, conclusions regarding the lack of usefulness of 'Greenness' resulted in the development of iteration two, which included only slope, aspect and toposhape. In the final iteration (3), the model was enhanced by including known species occurrences. This is an assimilation approach that was investigated because there is a general understanding that, in movement, wildlife may be more inclined to use sites they have previously used more than those they have not. This was an exploratory iteration to determine the effect on the final model.

13.4.2 Iteration 1: Slope, Aspect, Greenness, Toposhape

In the first iteration of BELIEF maps, the valley bottoms showed uniformly high probability of detection for lynx and wolves. This probability was high for any areas below rock or cliff faces. Although a few sites of high probability could be identified crossing the valley (i.e., linkage zones), there was little visible separation between high and low classes in the valley bottom. This lack of separation was more pronounced in the wolf probability surface. Consequently, we reviewed the probability surfaces

and found that greenness was the only surface with extremely low geographic variability within 1 km of roads. Thus, the effect of adding greenness to the probability model appeared to be that of enhancing suitability of the entire image and reducing separation on the valley floor. This effect was not expected, but is reasonable given that all valley areas are highly productive, relative to the surrounding peaks. However, as the objective of the research was to identify suitable linkages across roads (i.e., the valley floor) the lack of separation in greenness at lower elevations is problematic. Greenness was removed from the calculation in the second set of map iterations and linkage zone separation was improved.

13.4.3 Iteration 2: Slope, Aspect, Toposhape

In this second iteration of the BELIEF module, non-suitable lynx habitat linkages were evident in the valley bottoms, illustrating the constrained movement opportunities for lynx at the interface with roads. Areas that exhibit low probability of detecting lynx were areas of flat slope and aspect. The region adjacent to the TCH Phase IIIA (southern side) showed enhanced suitability relative to other regions in the Bow Valley. Ground validation showed a combination of old-growth spruce and young, dense pine stands. In this unique situation, there is a spatial convergence of two habitat niche requirements for lynx, namely, denning and hunting. This type of spatial arrangement in forest types is not common in the Bow Valley study frame.

The second iteration of the wolf probability image reconfirms the topographic effect of the landscape on wolf habitat suitability (Weaver *et al.*, 1996). The highest probability of detection is limited to extreme valley bottoms. Lakes appear in the images as high probability sites for wolves, which is appropriate in winter only, when lakes are frozen and used as travel routes (Paquet, 1993). In some cases, glaciers were identified as high probability wolf sites, because they have low slope angle. This is a consideration for future research that may look at linkages outside the valley floor. The identification of glaciers, lakes and rivers as highly probable linkages between valleys is misleading, as they may be inhospitable for travel by wolves. A glacial and river hydrology overlay could be used to remove these areas, but is not necessary for the present research.

The limitation of wolves to the valley floor, as depicted in the probability image provides insight into the reasons why wolves have a difficult time persisting in road fragmented mountain terrain (Weaver *et al.*, 1996). In addition to finding the valley floors easier for movement, wolves are obligate predators of ungulates, which as a rule are human tolerant, valley bottom dwellers. These constraints placed on wolves increase their interaction with road surfaces in the valley and, consequently, increase their probability of being killed.

13.4.4 Iteration 3: Slope, Aspect, Toposhape, Observed Species-sites

In the third iteration of probability surfaces, shown in Figures 13.4 (lynx) and 13.5 (wolves), the data assimilation method was used. The probability of detection was

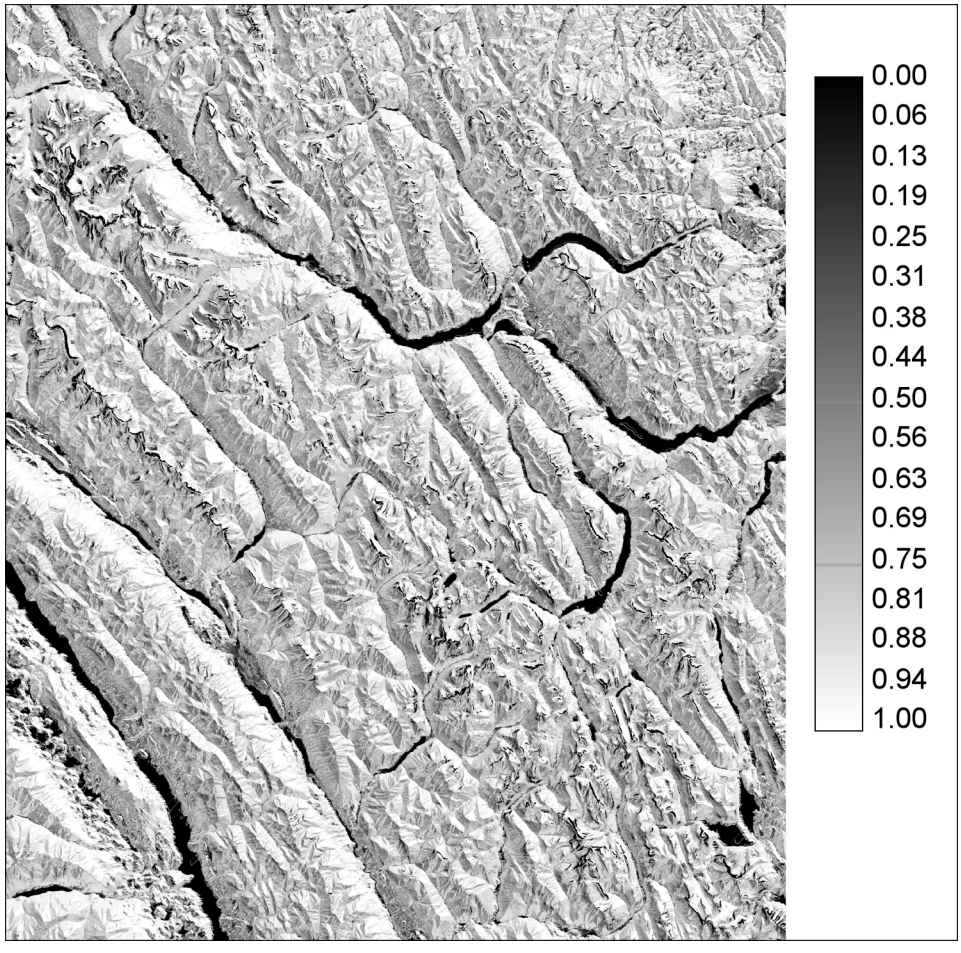


Figure 13.4 *Probability of lynx*

assumed to decline rapidly with distance from known (observed) points. Hence, the probability decayed over a distance from zero to 300 metres. The data assimilation created a more regionally specific (species-precise) predictive surface. However, the inclusion of this data enhanced sites within the existing linkages, but did not change the overall linkage zones.

Cross-valley linkages with higher than 95% probability (from Figures 13.4 and 13.5) were more frequent for lynx than for wolves, and crossing sites for wolves have wider spatial extent than lynx. By comparison Figure 13.4 (lynx probability) shows less contiguous highly suitable habitat for lynx than that for wolves, shown in Figure 13.5. The highest potential habitat for lynx is dispersed and is not confined to the valley floors. This offers lynx the benefit of not interacting with human disturbance on a continuous basis. However, lynx may be required to traverse very unsuitable patches of habitat (i.e., valley floors), in order to access other high quality habitat

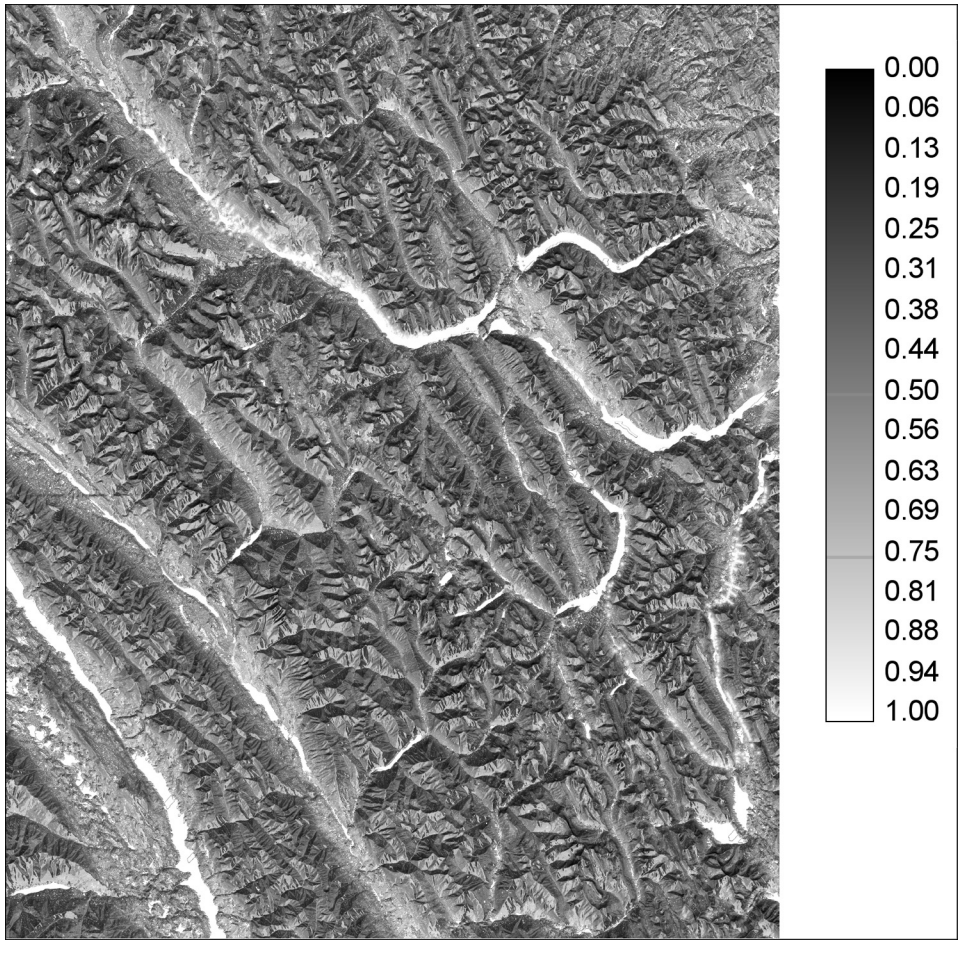


Figure 13.5 *Probability of wolves*

during dispersal or times when local prey abundance is low. Figure 13.5 also shows that wolves are highly constrained to the valley floors.

A comparison of Figures 13.4 and 13.5 also yields evidence that fewer connections are needed across highways for wolves, relative to lynx. The limited number of crossing opportunities suggests that the placement of mitigation will be critical to wolves. Lynx movement may be accommodated by placement of mitigation at a greater frequency. The suggestion to place lynx structures at higher frequency than wolves is grounded in the arguments of body size–home range and functional scale requirements (Holling, 1992). Lynx have smaller home ranges, thus, more may be accommodated in the same space as that required by wolves (assuming equal suitability for both species).

Plate 6 shows regions where there is >90% probability of detecting three focal species. Marten are included in this figure for demonstration purposes (i.e., how the

methodology may be extended to multiple species). Plate 6 was created by combining three Boolean images – one for each species. Marten were coded in Boolean as [1,0], wolves as [3,0] and lynx as [5,0]. This allows the images to be combined, such that the output may be used to detect sites for individual species (e.g., marten at zones with value 1), or to identify sites of various combinations of species. For example, if the user would like to identify a site where marten and wolves intersect in the landscape, then he/she would look for sites with the value 4 (i.e., 1 – marten + 3 – wolf). Likewise, sites where all three species may be detected have a value of 9 (1 + 3 + 5). This figure shows only the study region in Banff, west of Castle Junction. Evident in Plate 6 is the limited frequency and high dispersion of sites where all focal species have greater than 90% probability of crossing (i.e., the yellow sites). This suggests linkage zones must be species- or guild-specific, supporting previous arguments about connectivity requisites (Tischendorf and Fahrig, 2000; Conroy and Noon, 1996).

In major linkage zones where there exist fragmented regions of overlap, mitigation design might enhance the chances of both species crossing. For example, in the region directly west of Castle Junction there is some overlap between species. Mitigation design in this region should address the species-specific habitat needs that improve movement probability, such as maintaining a wide swath of land. Corridor width is one of the over-riding indicators of corridor success (Harrison and Bruna, 1999). Such design would increase the functional connectivity of the landscape (Tischendorf and Fahrig, 2000).

13.5 Policy Implications

As a result of the research findings, a number of policy implications have been presented to management agencies in the region, which are summarised in Table 13.3. These policy recommendations should serve as general guidelines for the spatial requisites of highway mitigation in the Rocky Mountain ecosystem.

Table 13.3 *Policy implications of the research*

1	Placement of mitigation should occur in high probability crossing areas, as defined on the GIS predictive model, herein.
2	Mitigation should be between 250 and 1000 metres in diameter (Alexander and Waters, 2000).
3	Using the above, mitigation should use elevated and buried highway sections, as often as possible. Exclusion fencing will be important in non-corridor areas.
4	Mitigation should be species- or guild-specific to maintain functional connectivity and ecosystem integrity.
5	Lynx and wolves may be used as focal species for mitigation (Alexander, 2001).
6	Higher resolution GIS images (e.g. <30 metres) are needed to manage appropriately for smaller functional levels (i.e. marten).
7	The adoption of GIS probability modelling and decision support is a cost-effective means of identifying crossing sites and provides a traceable process for decision makers.

13.6 Conclusion

This chapter examined the use of probability modelling in determining sites with the highest probability of detecting wildlife along four road systems in the Rocky Mountains of Alberta, Canada. The principal goal of the research was to locate optimal sites for placing mitigation structures (overpasses, underpasses or elevated highways) in order to restore structural and functional habitat connectivity for wildlife. Structural connectivity was addressed in this research by using spatial modelling and probability theory to find the optimal linkages in the landscape. Functional connectivity was addressed by identifying which species were good representatives of two levels of ecological organisation (i.e., meso- and macro-scales). Maintaining structural connectivity for two levels of ecological organisation alone will not achieve functional connectivity, if the design attributes of crossing structures are not appealing to the species involved (e.g., too narrow or too sparsely vegetated). This requires further examination of individual species' tolerances to habitat structure and geometry.

Biophysical attributes that predict movement were tested for non-uniform use by each of 13 study species. Attributes included aspect, greenness, slope, distance to drainages, distance to roads and toposhape (curvature). Strong evidence was found that aspect, greenness, slope and toposhape were useful predictors of habitat selection. It was observed that these preferences were well defined using probability functions. Moreover, in the analysis of track density estimates, this technique was argued to be superior to classical statistical approaches such as logistic regression or compositional analysis.

It was expected that greenness would be a useful predictor, based on work with grizzly bears (Mace *et al.*, 1999). However, in this case greenness was not suitable as a predictor because of the frame of analysis. There was little separation in greenness classes within the region of 1 kilometre from roads. Consequently, the inclusion of greenness obscured most linkage zones by applying an overwhelming 'greenness' suitability to habitat near roads. Therefore, greenness was removed from the analysis.

Structural connections were evident along the valley in BNP and these varied by species. Zones of overlap for both species that had high probability of use, were not abundant in the valley. This confirms the hypothesis that mitigation must be species-specific, or at least, guild-specific. Further tests should examine the periodicity of placement that may best suit each species/guild. In addition, design constraints should be examined. Other metrics should be tested for use in predicting species occurrence. For example, some species are known to associate with forests that have high structural complexity at the ground. Expert knowledge on relationships between metrics and finer scale predictors could be gleaned and incorporated into the weight-of-evidence model.

The strength of this modelling technique lies in its flexibility, its explicitness and its lack of assumptions associated with the more traditional statistical methods. Future data could be used to assess the predictive power of the model. Future analysis or evidence may suggest that other metrics are important, and these could be incorporated into the model. Moreover, if relationships within attributes (by class) are found to change, the associated probability functions can be modified. Human preferences could also be included in this model, at a later date. For example, highway

structural engineers may want to evaluate the costs of mitigating each identified linkage zone, based on the site criteria. The flexibility of this tool enhances its suitability for use in resource management, because it makes the tool practical. The model is explicit in its assumptions and rules. Explicitness is a desirable quality for resource management agencies, such as Parks Canada, who are subject to public scrutiny and who must be accountable for decisions, and for replication and adaptation of the scientific research.

Acknowledgements

This project was made possible by the generosity of the following financial contributors: the Natural Sciences and Engineering Research Council of Canada, the University of Alberta and A.C.A. Biodiversity Fund, the Province of Alberta Graduate Fellowship and the University of Calgary, Alberta Environmental Protection, Canadian Pacific Corporation, Banff National Park Wildlife and Highways Divisions, Dr Margaret P. Hess, Mr Edward L. Alexander, the Alberta Sport-Recreation-Parks and Wildlife Foundation, Employment Canada, the Western Forest Carnivore Society and the Paquet Fund for Wildlife Biology.

References

Alexander, S.M. (2001) A Spatial Analysis of Road Fragmentation and Linkage Zones for Multi-Species in the Canadian Rocky Mountains: A Winter Ecology Study, PhD, Department of Geography, University of Calgary, Alberta, 350pp.

Alexander, S.M. and Waters, N.M. (1999) Decision support methods for assessing placement and efficacy of road crossing structures for wildlife, in Evink, G.L. *et al*. (eds) *Proceedings of the Third International Conference on Wildlife Ecology and Transportation*, Missoula, MT, ICOWET, September, pp. 237–46.

Alexander, S.M. and Waters, N.M. (2000a) Modelling wildlife movement and planning mitigation in highway transportation corridors, CD-ROM publication, Fourth International Ecological Modelling Conference, Banff, Alberta, September.

Alexander, S.M. and Waters, N.M. (2000b) The effects of highway transportation corridors on wildlife: a case study of Banff National Park, *Transportation Research Part C*.

Alexander, S.M, Paquet, P.C. and Waters, N.M. (2002) *Quantification of the Road Barrier Effect on a Wildlife Community in the Rocky Mountains of Canada*, in preparation.

Anderson, D.R., Burnham, K.P. and Thompson, W.L. (2000) Null hypothesis testing: problems, prevalence, and an alternative, *Journal of Wilderness Management*, **64**(4): 912–23.

Andren, H. (1994) Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review, *Oikos*, **71**: 355–66.

Banff-Bow Valley Study (1996) Banff-Bow Valley: At the Crossroads, Technical Report of the Banff-Bow Valley Task Force, prepared for the Honourable Shiela Copps, Minister of Canadian Heritage, Ottawa, ON.

Bartel, A. (2000) Analysis of landscape pattern: towards a 'top down' indicator for evaluation of landuse, *Ecological Modelling*, **130**: 87–94.

Bascompte, J. and Solé, R. (1996) Habitat fragmentation and extinction thresholds in spatially explicit models, *Journal of Animal Ecology*, **65**: 465–73.

Carroll, C., Paquet, P., Noss, R. and Strittholt, J. (2000) *Modeling Carnivore Habitat in the Rocky Mountain Region: A Literature Review and Suggested Strategy*, prepared for WWF Canada.

Conroy, M.J. and Noon, B.R. (1996) Mapping of species richness for conservation of biological diversity: conceptual and methodological issues, *Ecological Applications*, **6**: 763–73.

Dragicevic, S. and Marceau, D.J. (2000) A fuzzy set approach for modelling time in GIS, *International Journal of Geographical Information Science*, **14**(3): 224–45.

Eastman, J.R. (1999) *Idrisi 32. Guide to GIS and Image Processing, Volume 1,* Clark Labs, Clark University, MA, p. 170.

Ellison, A.M. (1996) An introduction to Bayesian inference for ecological research and environmental decision-making, *Ecological Applications*, **6**(4): 1036–46.

Evink, G.L., Garrett, P. and Zeigler, D. (eds) (1998a) *Proceedings of the Second International Conference on Wildlife Ecology and Transportation (ICOWET)*, FL-ER-73-99, Florida Dept. of Transportation, Tallahassee, Florida.

Evink, G.L., Garrett, P. and Zeigler, D. (eds) (1999) *Proceedings of the Third International Conference on Wildlife Ecology and Transportation (ICOWET)*. Florida Dept. of Transportation, Tallahassee, FL, Missoula, MT.

Evink, G.L., Garrett, P. Zeigler, D. and Berry, J. (eds) (1998b) *Proceedings of the International Conference on Wildlife Ecology and Transportation, FL-ER-69-98*, Florida Dept. of Transportation, Tallahassee, Florida, p. 263.

Fahrig, L. and Merriam, G. (1985) Habitat patch connectivity and population survival, *Ecology*, **66**(6): 1762–8.

Forman, R.T.T. and Alexander, L.E. (1998) Roads and their major ecological effects, *Annual Review of Ecology and Systematics*, **29**: 207–31.

Gordon, J. and Shortliffe, E.H. (1985) A method for handling evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence*, **26**: 323–57.

Harrison, S. and Bruna, E. (1999) Habitat fragmentation and large-scale conservation: what do we know for sure? *Ecography*, **22**: 225–32.

Holling, C.S. (1992) Cross-scale morphology, geometry and dynamics of ecosystems, *Ecological Monographs*, **62**: 447–502.

Jackson, S. (1999) Overview of transportation related wildlife problems, in Evink, G.L. *et al.* (eds) *Proceedings of the Third International Conference on Wildlife Ecology and Transportation (ICOWET)*, FL-ER-73-99, Florida Dept. of Transportation, Tallahassee, Florida, pp. 1–4.

Khattatov, B.V., Lamarque, J.-F., Lyjak, L.V., Menard, R., Levelt, P.F., Tie, X.X., Gille, J.C. and Brasseur, G.P. (2000) Assimilation of satellite observations of long-lived chemical species in global chemistry-transport models, *Journal of Geophysical Resources*, **105**: 29135.

Lenton, S.M., Fa, J.E. and Perez del Val, J. (2000) A simple non-parametric GIS model for predicting species distribution: endemic birds in Bioko Island, West Africa, *Biodiversity and Conservation*, **9**: 869–85.

Mace, R.D., Waller, J.S., Manley, T.L., Ake, K. and Wittinger, W.T. (1999) Landscape evaluation of grizzly bear habitat in western Montana, *Conservation Biology*, **13**: 367–77.

Maurer, M. (1999) Development of a community-based, landscape-level terrestrial mitigation decision support system for transportation planners, in Evink, G.L. *et al.* (eds) *Proceedings of the Third International Conference on Wildlife Ecology and Transportation (ICOWET)*, FL-ER-73-99, Florida Dept. of Transportation, Tallahassee, Florida, pp. 99–110.

Noss, R.F. and Csuti, B. (1997) Habitat fragmentation, in Meffe, G.K. and Carroll, R.C. (eds) *Principles of Conservation Biology.* Second edition, Sinauer Associates, Sunderland, MA, pp. 269–304.

Noss, R., Quigley, H., Hornhocker, M., Merrill, T. and Paquet, P. (1996) Conservation biology and carnivores: conservation in the Rocky Mountains, *Conservation Biology*, **10**(4): 949–63.

Paquet, P.C. (1993) *Summary Reference Document – Ecological Studies of Recolonizing Wolves in the Central Canadian Rocky Mountains*, prepared by John/Paul and Associates for Parks Canada, BNP Warden Service, Banff, AB, 215pp.

Pearce, J. and Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression, *Ecological Modelling*, **133**: 225–45.

Pearl, J. (1987) Evidential reasoning under uncertainty, in Shrobe, H.E. (ed.) *Exploring Artificial Intelligence: Survey Talks from the National Conferences on Artificial Intelligence,* Morgan Kaufmann, Palo Alto, CA, pp. 381–490.

Peek, J.M. (1986) *A Review of Wildlife Management*, Prentice-Hall, Englewood Cliffs, NJ.
Richards, J.A. and Jia, X. (1999*) Remote Sensing Digital Image Analysis: An Introduction*, Springer, New York.
Singleton, P. and Lehmkuhl, J. (1999) Assessing wildlife habitat connectivity in the Interstate 90 Snoqualmie Pass Corridor, Washington, in Evink, G.L. *et al*. (eds) *Proceedings of the Third International Conference on Wildlife Ecology and Transportation (ICOWET)*, FL-ER-73-99, Florida Dept. of Transportation, Tallahassee, Florida, pp. 75–84.
Smith, D. (1999) Identification and prioritization of ecological interface zones on state highways in Florida, in Evink, G.L. *et al*. (eds) *Proceedings of the Third International Conference on Wildlife Ecology and Transportation (ICOWET),* FL-ER-73-99, Florida Dept of. Transportation, Tallahassee, Florida, pp. 209–30.
Spellerberg, I.F. (1998) Ecological effects of roads and traffic: a literature review, *Global Ecology and Biogeography Letters*, **7**: 317–33.
Srinivasan, A. and Richards, J.A. (1990) Knowledge-based techniques for multi-source classification, *International Journal of Remote Sensing*, **11**(3): 505–25.
Swart, J. and Laws, M.J. (1996) The effect of habitat patch connectivity on samango monkey (*Cercopithecus mitis*) metapopulation persistence, *Ecological Modelling*, **93**: 57–74.
Thomson, I.D., Davidson, I.J., O'Donnell, S. and Brazeau, F. (1988) Use of track transects to measure the relative occurrence of some boreal mammals in uncut forest and regeneration stands, *Canadian Journal of Zoology*, **67**: 1816–23.
Tischendorf, L. and Fahrig, L. (2000) On the usage and measurement of landscape connectivity, *Oikos*, **90**: 7–19.
Trombulak, S.C. and Frissell, C.A. (2000) Review of ecological effects of roads on terrestrial and aquatic communities, *Conservation Biology*, **14**(1): 18–30.
Urbanski, J.A. (1999) The use of fuzzy sets in the evaluation of the environment of coastal waters, *International Journal of Geographical Information Science*, **13**(7): 723–30.
Van Dyke, F.G., Brocke, R.H. and Shaw, H.G. (1986) Use of road track counts as indices of mountain lion presence, *Journal of Wilderness Management*, **50**(1): 102–9.
Verbyla, D. and Chang, K. (1994) Potential problems in using GIS for wildlife habitat research, *GIS World 1994 Conference Proceedings*, Vancouver, BC, February.
Wade, P. (2000) Bayesian methods in conservation biology, *Conservation Biology*, **14**(5): 1308–16.
Warntz, W. and Waters, N.M. (1975) Network representations of critical elements of pressure surfaces, *Geographical Review*, **65**: 476–9.
Weaver, J.L., Paquet, P.C. and Ruggiero, L.F. (1996) Resilience and conservation of large carnivores in the Rocky Mountains, *Conservation Biology*, **10**(4): 964–76.
Zadeh, L.A. (1965) Fuzzy sets, *Information and Control*, **8**: 338–53.

PART 4
NATIONAL SPATIAL PLANNING

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

14

Modelling Migration for Policy Analysis

Phil Rees, A. Stewart Fotheringham and Tony Champion

Abstract

This chapter reports on the development of an applied model of internal migration in the UK to support central government policy making. A two-stage modelling procedure was developed and parameters calibrated on a time series of historical data. Details of the data sets and the modelling are presented, together with illustration of the software package named MIGMOD that was constructed as a planning support system to assist government officials in predicting the impact of alternative scenarios.

14.1 Introduction

This chapter describes the results of a project carried out by a team (see Acknowledgements) based in the Universities of Newcastle upon Tyne and Leeds for the UK government during 2000 (Phase 1) and 2001 (Phase 2). Phase 1 was led by Tony Champion and Phase 2 by Stewart Fotheringham. The UK government department sponsoring the project (see Acknowledgements) changed its name and composition twice during the project and was known successively as the Department for Environment, Transport and the Regions (DETR), the Department of Transport, Local Government and the Regions (DTLR) and the Office of the Deputy Prime Minister (ODPM). For simplicity we will refer to the various organisations as 'the Department'. The chapter is published with the Department's permission but represents the views of the authors and not the Department.

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

The chapter describes the aims of the project, the type of spatial model constructed and calibrated and explains the choices in model design that were made during the course of the project. Exposing the choices explored and the different model designs used in Phase 1 and Phase 2 will, we hope, provide valuable guidance to other researchers seeking to model the flows of migrants over time between a large number of origin and destination areas. The main output of the project is a user-friendly software system known as MIGMOD (Migration Modeller) which we also describe. The original aim of the project was that it might help inform policy decisions, and we have undertaken simple policy scenarios. These are illustrative and indicative of the model's performance, rather than a true reflection of the effect of policy changes. This account of the work complements the overview provided by Champion *et al.* (2002).

The chapter is divided into a further seven main sections. Section 14.2 places migration within the UK in the broader context of population change and explains why the Department needed a model of migration. Section 14.3 explains the aims that the model was designed to achieve. Section 14.4 outlines the 'state space' of the model: what spatial units were used, what migrant groups were studied and what time periods were used in which to measure migration. Most of the choices were 'data-driven'. Section 14.5 describes the nature of the dependent variable being modelled and the set of determinant variables used in the two stages of the model (out-migration from origins, distribution from origins to destinations). Section 14.6 describes the model of out-migration used, known as the Stage 1 model, which is used to identify the factors that contribute to varying rates of out-migration from origins across England and Wales. Section 14.7 describes the Stage 2 model, a model of destination choice by migrants, which is used to determine the factors that lead to varying levels of destination attractiveness. In Section 14.8 we describe how the two models are integrated within a user-friendly software system for migration modelling termed MIGMOD. We also discuss the revisions to model design made between Phase 1 and Phase 2 of the project.

14.2 Context: The Role of Internal Migration in Population Change

Population change is often a good indicator of regional economic prosperity. Regional policy is often concerned with the direction of population shift: decline is a trend to be avoided because it leads to destruction of social and physical capital. Population change is also one of the main ingredients in household change, although the two are not necessarily highly correlated as evidenced by the decrease in average household size in recent decades in the UK leading to a greater household growth than population increase. Population change is driven by components which can have very different geographical patterns.

Table 14.1 presents estimated regional components of population change for Government Office regions of the United Kingdom in 1999–2000. Only three regions – Scotland, the North East and West Midlands – experienced such population decline, while London and the southern regions showed the highest rates of increase. The

Table 14.1 *Components of population change for Government Office regions (England) and countries, UK, 1999–2000*

Country and Government Office Region (England)	Rates (per 1000 population)				
	Natural increase	Net external migration	Net internal migration	Other changes	Total change
Wales	–1.0	2.0	2.1	0.0	3.1
Scotland	–1.2	0.8	–0.8	0.2	–1.0
Northern Ireland	4.1	0.5	–0.5	–0.6	3.5
North East	–0.8	0.9	–1.7	0.1	–1.6
NW & Merseyside	0.1	2.8	–1.1	0.1	1.9
Yorkshire & the Humber	0.6	1.2	–0.1	0.5	2.2
East Midlands	0.5	–0.6	3.8	0.3	4.0
West Midlands	1.1	0.2	–1.4	0.1	0.0
East	1.4	1.7	4.0	0.3	7.5
London	5.9	15.6	–9.1	–0.1	12.3
South East	1.1	1.5	2.1	0.0	4.6
South West	–1.2	3.2	6.2	–0.3	8.0
United Kingdom	1.0	3.2	0.0	0.1	4.3

Source: Computed from the Office for National Statistics, *Population Trends 108*, Tables 1.7 and 8.1 and *Key Statistics and Vital Statistics, Local Authorities, England and Wales, 2000*, Table 9a Estimated mid-2000 resident population, persons all ages: components of change.

Notes:

1. External migration includes asylum seekers and visitor switchers.
2. The components are based on mid-year population estimates current at the time of writing. The population estimates for mid-years from 1981 through 2000 will be revised in the light of the publication of the 2001 Census results. Prior to the 2001 Census the best estimate for the mid-2001 population was that in the 2000-based national projection (Shaw, 2002) at 59.987 millions. This compares with a 2001 mid-year estimate published in October 2002 of 58.837 millions based on the 2001 Census 'One Number' of 57.789 millions.

contributions of the three components of population change, natural increase (births minus deaths), net external migration (inflows from the rest of the world minus outflows to the rest of the world) and net internal migration (inflows from the rest of the UK minus outflows to the rest of UK) vary across the regions. Natural decrease characterises Wales, Scotland, the North East and the South West. Natural increase is low elsewhere except in Northern Ireland and London. Net external migration makes a positive contribution to population change in all but one region but a massive contribution to the growth of London's population. From Table 14.1 we can also see that internal migration is a vital component that shifts populations between regions because there are both gaining and losing regions. For example, Scotland, the North East, North West and West Midlands lost between 4000 and 8000 people through internal migration, while Northern Ireland and Yorkshire and the Humber had small losses. London lost substantially through internal migration though this was more than compensated for by its gain from overseas. The other regions, particularly the East, South West and East Midlands gained through internal migration.

If more balanced population growth is a policy goal, then this means improving survival chances in high mortality regions, encouraging couples to have more children in certain regions or spreading the influx of immigration across regions more evenly or reducing migration outflows from certain regions and decreasing migration inflows into others. Of these options, the latter would seem less controversial and more acceptable as a solution to population imbalances. In order to stem population losses from depressed regions, and to stem the increasing population pressures in economically successful regions, it is necessary first to understand how such migration imbalances arise. One of the best ways to achieve that understanding is to build a model of migration behaviour and to investigate the determinants of both out-migration and destination choice. Such a model is reviewed in this chapter.

14.3 Aims of the Project

14.3.1 Introduction

The primary aim of the project was to provide the Department with an initial migration model aimed at informing policy. The model was designed to enable the investigation of the first-round quantitative impacts of alternative economic and policy scenarios on gross flows of population between regions. For the longer term, the intention was that the model would be capable of further enhancement, notably in order to gain a better understanding of the role of migration in influencing household change. It was clear from the outset that there were insufficient reliable data for regions, counties or local authorities on the migration of whole or part households or individuals migrating between households to construct a model that would directly connect to the Department's Household Projection model. The Department was well aware that migration flows produced feedback effects on areas (spirals of decline resulting in out-migration sometimes create opportunities for renewal or heavy in-migration resulting in denser settlement which lowers the attractiveness of destinations) but the team and the Department agreed early on that such modelling was over-ambitious until simpler cross-sectional or time series models had been calibrated and verified.

The 'problems' to be solved were of two kinds. The first was to find ways of reducing heavy migration losses from northern and midland cities in the UK and from other areas with poor economic track records (former coalfields, the Thames Gateway region, rural Wales). Heavy out-migration from these regions means shrinking demand for housing and other consumption goods, leading to a downward spiral of economic and social development, and loss of valuable social capital (Walker, 2002). However, we should also recognise that inter-area migration is also a means of shifting people to areas of economic opportunity. The second problem was to avoid the consequences of over-stimulation of successful regional economies. High in-migration might put pressure on house prices and make housing unaffordable to workers outside the dynamic sector, as is very much in evidence in and around London currently.

14.3.2 Detailed Objectives

The principal objectives were to develop an initial policy model which:

- was capable of handling gross migration flows between regions;
- could be used to examine the impact of alternative scenarios;
- could provide advice on urban–rural shifts;
- was compatible with household projections methodology;[1]
- was flexible, easy to use and transparent; and
- was capable of further enhancement, especially at a finer spatial scale.

It was agreed at the outset that, in developing this initial model, there was a need for pragmatism, notably in relation to data availability and feedback effects. In particular, this initial model would not attempt to incorporate the impacts of migration or other changes on the determinant variables. *Phase 1* of the modelling work established the framework, set up the database, developed a prototype working model and produced a standalone version for use by the Department. *Phase 2* reviewed model robustness, conducted a full range of diagnostic tests of determinants and investigated some counterintuitive results from the prototype model in detail. The overall model structure remained the same in both phases but different model forms were adopted in Phase 2 after extensive winnowing of the very large determinant dataset and after experimentation with multiplicative or additive equations linking dependent migration and independent determinant variables and with linear and quadratic and exponential functions of the determinant variables.

14.3.3 The Modelling Framework

The central features of the approach were:

- the modelling of out-migration from each area by reference to a set of determinant variables (Stage 1 of the model);
- the use of a model to distribute migrants between destinations (Stage 2 of the model);
- the development of an operational, user-friendly combination of Stages 1 and 2 that enables the model user quickly to set up and run a range of 'what if?' scenarios and to view the large volume of inputs and outputs (Stage 3 of the model); and
- the development of a selection of scenarios of determinant variables reflecting desired policy options (Stage 4 of the model).

[1] The ODPM household and ONS population subnational projections use local authority (LA) zones, whereas the migration model described here uses FHSAs. FHSAs are aggregations of LAs for most of the 1983–98 period, though from 1996 Health Authority reorganisation meant small inconsistencies in the aggregation. Changes in Health Service organisation since 1998 mean that the data series will not extend and future work will need to employ the New Patient Migration data series (ONS, 2002). These new data are available directly for the LAs used in subnational projections.

The modelling team discussed and agreed this framework with the Department very early on in the project. In fact, the rationale for this two-stage approach was reviewed in Champion *et al.* (1998). The main argument for modelling the out-migration flow separately from the destination choice is that different data sets were available for the two types of decisions. In order to model out-migration determinants, the team had available to it 15 years of annual out-migration data which allowed some time effects to enter the modelling framework (see below). In order to model the determinants of destination choices, an origin–destination matrix of migration flows was needed and this was only available for seven time periods which was not sufficient to allow time-related variables to enter the model. There is also evidence that migrants often conceive of migration as a two-stage process with worsening conditions at an origin eventually reaching a threshold level at which people decide to leave and then conditions at various locations being examined in order to decide on a destination. The model structure adopted conforms to the consistency principles for spatial interaction models first defined by Wilson (1967, 1972). The equation structures used in Phase 1 and Phase 2 versions of the migration model are discussed below in Section 14.4.

The calibration and testing of the model has embraced several dimensions of space, time, age and gender. Decisions needed to be taken on the zones to employ in the model, the time horizon over which the model would be calibrated and the disaggregation of migration into sub-categories for which different models were fitted. These decisions are discussed in the next section of the chapter.

14.3.4 Additional Investigations

In the development of the initial migration model, the project included three discrete modules of optional work. These were (i) improving the treatment of international migration, (ii) providing a population projection framework and (iii) examining the household dimension of migration. Details of this additional work can be found in the Phase 1 Final Report (Champion *et al.*, 2002).

14.4 The State Space

14.4.1 Spatial Units

For which areas should a migration model be developed? This is really two questions: what scale of spatial unit should be used and what collection of units should the model be calibrated for? Because the Department wished to know how migration affected household change, the ideal zone to have chosen would have been the local authority (LA). However, the UK did not have an annual time series of migration data at the LA scale until 1998 when migration data based on the patient registers came on stream. Data are available in the decennial censuses (1971, 1981, 1991 and soon 2001) at LA scale though the number and definition of LAs differs between

1971 and 1981 and between 1991 and 2001. The time period between the censuses is too long to allow any great understanding of the temporal influences on migration patterns.

An alternative is to use migration data available on an annual basis from the National Health Service Central Register (NHSCR). The zone system used in this series are the Family Health Service Authority (FHSA) areas as defined between 1983 and 1998 (see Figure 15.2 in Champion *et al.*, 2002) in England and equivalent authorities in Wales, Scotland and North Ireland. FHSAs were coincident with shire counties (England and Wales), metropolitan districts (England), groups of London boroughs, Area Health Boards (Scotland) and Health Boards (Northern Ireland). Ideally, the migration model would have been constructed using all of these areas which cover the whole of the United Kingdom. Unfortunately, there are gaps in the flow statistics between areas in Scotland for a number of years and between areas in Northern Ireland and in Great Britain. In Phase 1 of the model construction, therefore, it was decided to treat Scotland and Northern Ireland as two zones only. FHSAs in Wales were included as separate zones in the analysis. Results for Wales were not required directly by the Department because Wales was not its responsibility but the team argued strongly that migration links between places in England and places in Wales were strong and their exclusion would bias results for a good part of England. The flows between England, Scotland and Northern Ireland were much weaker and use of single zones for these countries would not seriously bias model results. The system to be modelled at the outset consisted of migration flows from 100 origins to 100 destinations disaggregated by age and sex.

In the event, one or two changes to this set of origins and destinations were necessary because of difficulties in estimating determinant variables for Scotland and Northern Ireland comparable to those for England and Wales. The statistics available for Scotland and Northern Ireland did not match sufficiently well to merit inclusion of these zones in Stage 1 of the model, so that only 98 zones were used. For Stage 2 there were fewer difficulties because the variables were cross-sectional in nature and only needed for 1996–97. However, major difficulties arose in the prediction of migration flows to Northern Ireland. With the set of determinant variables used in Stage 2 of the model, we consistently and massively overpredicted flows to Northern Ireland and therefore underpredicted flows to other zones. Missing from the model were variables that measured the relative attractiveness of zones based on perceptions of their political history and threat of external violence. In the absence of such measures, we confined the Stage 2 distribution model to 99 zones. In Phase 2 of the work these decisions were reviewed and the number of zones reduced to 98 in both stages of the model.

The effect of excluding some sub-national areas from the model calibration could not be tested with the data sets we were using but is an issue that could be addressed using 2001 Census data when they are available, as the 2001 Census Origin-Destination Statistics will provide comparable and comprehensive flow statistics between all LAs in the UK. There is also the question about what effect does immigration from, and emigration to, external zones have on migration between internal zones (whether those external zones fall within the UK or without). We did include

estimated measures of external immigration and emigration (Stillwell *et al.*, 2001) in the internal migration model but their close correlation with variables measuring 'urbanness' meant their exclusion from the final determinant set to avoid problems of collinearity. The results of the model calibrations of immigration rates to destination zones suggest that previous immigrant stocks, the numbers of term-time students and population size of zone were the main attractors and that immigration and internal migration operated in different 'worlds'.

14.4.2 Time Series

For which years should the migration model be calibrated? Should we use as many years as the available information allow or strive to calibrate for particular periods? NHSCR flow data by origin, destination, age and sex are available for the 98-zone system from 1975–76 (mid-year to mid-year) until 1998–99 (at the time that the work was carried out). However, there are two major discontinuities in the series that make it necessary to confine analysis to the years 1983–84 to 1997–98. Prior to 1983–84, the data were published only as aggregations of the full origin–destination–age–sex (ODAS) array of flows, namely as three sub-arrays of migration flows by origin–destination (OD), by origin–age–sex (OAS) and by destination–age–sex (DAS). Between 1983–84 and 1997–98 the data were supplied as frequency counts classified by origin–destination–age–sex (ODAS). The age classification was by single years of age, whereas prior to then it had been by five-year age group. In the years 1996–98 local governments were reorganised successively in Scotland, Wales and parts of England. After 1 April 1998 the migration data are recorded for these new LAs (see Wilson and Rees, 1998, for details of the links between pre- and post-1998 LA geographies). So there were available 15 mid-year to mid-year counts of ODAS-classified migration flows. How could these data be used and with what kind of migration models?

The first idea was to model the out-migration flows or rates as a time series. All the advice of our econometric colleagues (Barmby, Tremayne – see acknowledgements) was against that approach: at least 40 years of data were necessary. A decision was therefore made to pool the data for the different years into one set of observations for model calibration purposes. However, because we decided after experimentation that it was best to have a one-year lag between the migration outcome and the time reference of the determinant variable, the time series of observations used dropped to 14 (1984–85 to 1997–98) for Stage 1 of the model (the origin prediction model). In Phase 1 of the model fitting we did not use a time trend as one of the explanatory variables but subsequent examination of the time series of out-migration rates for ages 16–19 and 20–24 revealed that a steady rise in out-migration took place from 1990–91 onwards, due primarily to the expansion of higher education in the 1990s. A linear time trend was therefore introduced to Stage 1 of the model in the Phase 2 work.

One argument that was put to us after several presentations of the model was that we should have explored the possibility that the determinants of migration

would behave differently in times of economic boom and bust, and that different outcomes would result. Some analysts have interpreted the influence of the economic cycle on migration as follows: regions enter booms and busts at different times and the cyclic effects spread outwards from inception points. So, for example, the recession of 1981–3 saw downsizing of manufacturing industry and job losses first in the northern regions and then spread south increasing north to south transfers. The boom of 1988–1990 stimulated house price rises and outward migration in the South East which spread to the rest of the country tipping the north into positive migration gains for a year or so before the traditional pattern of north to south losses reasserted itself. These effects are ones we did not explore, given the need to develop a robust and operational model in a short time frame for the Department.

For Stage 2 of the model, time is less critical an issue. We calibrated the distribution model for a single year only but checks on alternative years indicated that the structure of migration distribution from origins to destinations is very stable, being dominated by the 'competing destinations' variables of distance (with a contribution from contiguity), destination accessibility and population size (Fotheringham, 1986, 1991). The year chosen for the final calibration was 1996–97 rather than 1997–98 because boundary changes were beginning to affect some migration flows in 1997–98.

14.4.3 Age Groups

The review of knowledge about the patterns of migration and the reasons for moving suggested that these would vary significantly with a person's position in the life course. It is not usually possible with aggregate data to place migrants into exact life course stages (e.g., retired from work or still working) but usually information is available on the age of migrants. Ages can then be grouped into approximate life-course stages. How many should there be and where should the boundaries be drawn between stages?

Since the NHSCR data were available for single years of age up to age 85+, it would theoretically have been possible to fit 86 different models, one per age group. However, previous empirical work (Stillwell *et al.*, 1992) had established that most of the variation between ages could be described using a small number of grouped ages. Initially, some eight age groups were proposed: 0–16, 17–19, 20–24, 25–29, 30–44, 45–59, 60–69 and 70+. These correspond to childhood/schooling ages, the ages at which adolescents leave home for higher education, the ages at which students leave higher education for their working and partnership careers, the ages when they look for career advancement, the family formation ages, the later working ages (quiescent in terms of migration), the ages of retirement when locations are reassessed and the older ages when events such as loss of spouse/partner or increasing infirmity precipitate migration. We examined the correlations between the flow matrices at ages around the bounds of these life-course stages and moved 16 year olds into the late adolescents group. We also amalgamated the last two ages as the flow matrix for the

70+ age group was too sparse for reliable calibration. The final age groups chosen for the model were therefore seven in number: 0–15, 16–19, 20–24, 25–29, 30–44, 45–59 and 60+.

14.4.4 Gender

There is a growing interest in the differences between men and women in their migration behaviour, though the differences are only expressed when they are not in marriage, partnership or in families. The Government Statistical Service now requires all analyses to include a gender dimension and this we did in the migration model, though the differences in results are generally not great.

14.4.5 Summary

Choosing the state space that we did meant that we would be calibrating the Stage 1 and Stage 2 parts of the migration model for seven age groups and two genders or 14 migrant groups in total. However, there were further choices to be made. Should the out-migration model be calibrated for each origin separately or for all origins together or for groups of origins?

The option of calibrating the Stage 1 model for each origin was ruled out first: there were too few observations (14) per origin to produce reliable parameter estimates. We calibrated models in Phase 1 with grouped origins using Government Office Regions as the grouping criterion and compared the results with a model calibrated on all origins collectively. As would be expected, the former produced more accurate predictions of out-migration rates but a lot of the improved predictive performance came from the ability to estimate separate intercept terms for each zone. How should the variation of these intercepts be interpreted? It appeared that much of the variation in the intercepts was masking variation in some of the relationships between out-migration rates and origin attributes which was thus misleading. It appeared that some origin attributes had little effect on out-migration rates but this was simply because the effect of differential levels of those origin attributes was subsumed within the varying intercept values. Hence we opted for an 'all zones together' calibration for the 14 migrant groups and somewhat reluctantly dropped the idea of separate models for different clusters of origins.

The situation for the Stage 2 destination choice model was different. Here we had, for each origin, 97 destination flows upon which to calibrate the distribution model. By selecting origin-specific distribution models we could allow the determinants to have different influences on the outcomes for each origin. For example, we found that house prices (in the Phase 2 model) had coefficients that were both negative and positive depending on origin (other things being equal), while distance has negative coefficients in all cases (no origin had migrants who preferred to move to distant places). For Stage 2, therefore, it was possible to calibrate 98×14 or 1372 separate models (for just one year).

14.5 Data

14.5.1 Database Assembly

The major element of the work in Phase I of the project was the assembly of the data needed for calibrating the Stage 1 and Stage 2 models. Three broad types of data were required: (i) data on migration flows, (ii) data on the distance between each pair of the 100 'zones' and (iii) data on the characteristics of the 100 'zones' and their populations. We were guided in selection of determinant variables by the review undertaken in Champion *et al.* (1998), paying particular attention to the inclusion of variables that were directly or indirectly connected to the Department's policies for urban, rural and regional development.

In practice, several obstacles were encountered. Few datasets are published for all four countries of the United Kingdom, so data either had to be obtained as a series of separate tasks or had to be estimated. No datasets present statistics for the 100 zones used, so either data had to be aggregated from smaller units or values imputed from larger areas containing the zones. Data sources did not always provide information for all years. Some variables were found to need denominator data from a separate source, thus doubling the task (e.g. for crime rate defined as total number of offences per household).

Despite these problems, data were obtained for 139 potential determinants of out-migration and 69 potential determinants of migration destination choice. Much of this achievement was due to the valuable assistance given by a considerable number of persons and agencies (see Acknowledgements). Even so, it proved necessary to resort to estimation in order to complete the data series for all 98 FHSAs of England and Wales and for all relevant years. The data series for Scotland and, even more so, those for Northern Ireland were less comprehensive. In Phase 2 of the project, the initial 139 determinant variables were reduced through careful qualitative review and examination of multicollinearity to 49 potential determinants for Stage 1 of the model and the 69 potential determinants of destination choice were reduced to 23. The variables in the two datasets are listed in Table 14.2, which also gives the source of the data collected on each variable. Both data sets represent a huge increase in size from those normally used in migration models.

14.5.2 Within-UK Migration

As outlined in Section 14.4, the migration data are derived from the NHSCR and made available by Office of National Statistics (ONS). For the purpose of this study, data files have been constructed for the 100 zones for the 14 age–sex groups for mid-year to mid-year periods as follows:

- the numbers of out-migrations and in-migrations for 15 years 1983–84 through to 1997–98 are converted to rates based on the respective population groups of each zone; and

Table 14.2 *List of variables used in the (Phase 2) Stage 1 model*

Variable name	S1	S2	Meaning	Source
AIR	✓		Principal component of no2nox, ozone (high values associated with areas of poor quality)	AEA Technology. Values taken dataset point nearest to FHSA population-weighted centroid
AIRPORTS		✓	Access to scheduled flights to non-domestic destinations in 1997	Input data from CAA.
ASDIVRT	✓		% divorced – age-specific	1991 Census
ASUNEM		✓	Age-specific unemployment rate	NOMIS database; ONS mid-year estimates
ASUNEM_L	✓		Age-specific unemployment rate lagged	As ASUNEM
ASUNEM_L_Y	✓		Y variable for age-specific unemployment rate lagged	ASUNEM & DIJNETW
CTAX		✓	Average council tax for households	DETR Local Government website
CLIMATE	✓	✓	Principal component of frosty days, sunny days, rainfall, July temperature (high values associated with warmer drier areas)	University of East Anglia. Values taken dataset point nearest to FHSA population-weighted centroid
COMMUT	✓		% long distance commuters	1991 Census
CONTIG		✓	Contiguity dummy, equals 1 for pairs of zones sharing a boundary, otherwise 0	Derived from map of zones
CRIME	✓	✓	Principal component of household insurance premiums, index of crime a serious problem, total rates of offences recorded by police (high values associated with high crime areas)	Insurance companies, police records
DESTACC		✓	Sum of alternative destination populations divided by network distance from origin	ONS population estimates and DIJNETW
DIJNETW		✓	Network distance between each pair of zones	Computed from district level map by Duke-Williams and Alvanides
EMPGRO		✓	Employment growth	1991 Census and Survey of Employment
EMPGRO_L	✓		Employment growth lagged	As EMPGRO
EMPR		✓	Employment rate	As EMPGRO
EMPR_L	✓		Employment rate lagged	As EMPGRO
EMPR_L_Y	✓		Y variable for employment growth lagged	EMPR & DIJNETW
GCSE_L	✓		% 16 year olds obtaining 5+ GCSEs at Grade C or above lagged	DfEE, Scottish Office, DENI
HHINC		✓	Household income	Estimated by Bramley & Smart 1996, Bramley 1996
HHINC_L	✓		Household income lagged	As HHINC
HHINC_L_Y	✓		Y variable for household income lagged	HHINC & DIJNETW

HPRICE		✓	House price	Nationwide Building Society & CURDS, University of Newcastle
HPRICE_L	✓		House price lagged	As HPRICE
HPRICE_L_Y	✓		Y variable for house price lagged	HPRICE & DIJNETW
LISTED	✓	✓	Listed buildings (divided by dwellings 1998)	National built heritage organisations
LISTED_Y	✓		Y variable for listed buildings	LISTED
LOWI_L	✓		Index of relative share of low earners lagged	Estimated from New Earnings Survey
LOWI_L_Y	✓		Y variable for index of relative share of low earners lagged	LOWI_L & DIJNETW
MGINTR_Z_L	✓		Building Society mortgage interest rate lagged	Building Societies average
NOCENT	✓		% no central heating	1991 Census
NOCENT_Y	✓		Y variable for % no central heating	NOCENT & DIJNETW
NONWH	✓		% non-white	1991 Census
NONWH_Y	✓		Y variable for % non-white	NONWH & DIJNETW
OCCMIG	✓		Occupational migration index (predicted number of migrants based on occupation divided by residents – age-specific	Based on analysis of 2% SAR, then applied to 1991 Census LBS.
PARDOM	✓	✓	Students at parental domicile divided by number of residents, average of start and end mid-year estimates	HESA, UCAS and 1991 Census (Bailey and Rees 2001)
PNBU	✓	✓	New housing development (units) on land in former urban use divided by total new housing development	DETR & Ordnance Survey Land Use Change Statistics (LUCS)
PNRL		✓	% net relets social sector	DETR HIP1 series
PNRL_L	✓		% net relets social sector lagged	DETR HIP1 series
POPN		✓	Number of residents, average of mid-year estimates	ONS Population estimates
PQPR		✓	% new build completions private sector	DETR HIP1 series
PQPR_L	✓		% new build completions private sector lagged	DETR HIP1 series
PQPR_Y_L	✓		Y variable for % new build completions private sector	PQPR & DIJNETW
PQSR		✓	% new build completions social sector	DETR HIP1 series
PQSR_L	✓	✓	% new build completions social sector lagged	DETR HIP1 series
PQSR_Y_L	✓		Y variable for % new build completions social sector lagged	PQSR & DIJNETW
PVAC		✓	% vacant all sectors	DETR HIP1 series
PVAC_L	✓		% vacant all sectors lagged	DETR HIP1 series
PVAC_Y_L	✓		Y variable for % vacant all sectors lagged	PVAC & DIJNETW
RGDPCH_Z_L	✓		Real GDP annual % change lagged	ONS Economic Trends, Wilcox Housing Finance Review
RGURB	✓	✓	Land area subject to urban development of all types divided by population	DETR & Ordnance Survey Land Use Change Statistics (LUCS), and ONS mid-year estimates
RGURB_Y	✓		Y variable for land area subject to urban development	RGURB & DIJNETW

Table 14.2 *Continued*

Variable name	S1	S2	Meaning	Source
RLAPSC	✓	✓	Index of private sector stock in poor condition	Index used in HNI system, based on EHCS and proxies. England only.
RLASC	✓	✓	Index of local authority stock in poor condition	Index used in GNI system by DETR, based on stock profile and EHCS. England only
RLASC_Y	✓		Y variable for index of Local Authority stock in poor condition	RLASC & DIJNETW
SCENIC	✓	✓	Access to National Parks, Areas of Outstanding Natural Beauty and Heritage Coasts	Provided by DETR. Weighted averages of ward values.
SINGELD	✓		Proportion of single elderly households	1991 Census
SINGLE	✓		% single	1991 Census
TER	✓		% housing terraced	1991 Census
TERMT		✓	Students at term-time address divided by number of residents, average of start and end mid-year estimates	HESA, UCAS and 1991 Census (Bailey and Rees 2001)
TERMT_L	✓		Students at term-time address divided by number of residents, average of start and end mid-year estimates, lagged	HESA, UCAS and 1991 Census (Bailey and Rees 2001)
TIME_Z	✓		Time trend (year counter)	1 = 1984, 2 = 1985 etc.
TPOPN_Y_L	✓		Y variable for total population lagged	ONS Population estimate, TPOPN & DIJNETW
UNEMP_Z_L	✓		National claimant unemployment rate % lagged	DfEE, ONS population estimates
UNIVPL_Z_L	✓		Total number of students in England and Wales lagged	HESA Statistics
URBAN_L	✓		Principal component of bright lights indicator, % flats, % shared dwellings, population weighted average ward density of population, immigration rate, persons per room, HWCH, urbanisation factor (high values indicate more urbanised areas)	See text and Champion *et al.* (2001)
VACDRL	✓	✓	Vacant and derelict land divided by Total dwellings 1998 (HIP1)	Based on new National Land Use Database. Covers England only.
VACDRL_Y	✓		Y variable for all vacant and derelict land	VACDRL & DIJNETW
VISITS		✓	Visitor numbers to tourist attractions in 1996 divided by 1991 Census population	British Tourism Authority
WIDOW	✓		% widowed	1991 Census

Notes:
1. The suffix _L indicates a lagged variable. The suffix _Y indicates a regional variable. The suffix _Z indicates a national variable.
2. S1 = Stage 1 of the model: out-migration from origins. S2 = Stage 2 of the model: destination choice.

- the numbers of out-migrations from each of the 100 zones moving to each of the other 99 zones for each of the seven years 1990–91 through to 1996–97.

This NHSCR migration dataset is derived from the re-registration of individuals with family doctors (GPs) in another FHSA area. It therefore has the advantage of providing a continuous record of migration within the UK. As such, it has for many years been the main source of within-UK migration data used by ONS for its official population estimates for England and Wales. This is not to say that the NHSCR data provide a full and perfect record of migration. There is a well-recognised undercount of young adult males in the NHSCR datasets caused by this group of migrants being relatively less likely to re-register with their local GP. This causes the measured total out-migration values for these cohorts from many FHSAs to be smaller than they should be. If not corrected, this will lead to biased model calibrations if the undercount is not uniform across all FHSAs (which it is not).

To appreciate the nature of the problem, consider the comparison in Table 14.3 of the ratios of males/females in the SMS (Special Migration Statistics compiled as part of the 1991 Census) and the 1990/91 NHSCR for four sample FHSAs. These figures support the suspicion that young males have a tendency to postpone re-registering with a GP until later years: the ratios of males/females are much lower in the NHSCR dataset than in the SMS for the age groups 16–19, 20–24 and 25–29. The undercount, however, is not spatially stable and varies considerably across the FHSAs. Consequently, a model of out-migration totals for each migrant group will have problems predicting values for males in the age groups 16–19, 20–24 and 25–29 given the problems in the recorded values. There might be an associated problem of males re-registering at a destination after several years and hence inflating the figures in later age groups. There is some evidence for this in the Table 14.3 for Newcastle and North Tyneside, for example. Can we use the information on sex ratios in the SMS (which are presumed to be unbiased) to produce better migration matrices from the NHSCR data?

To achieve better estimates of migration from the NHSCR data, we assume that there is little or no bias in the recording of female migration in the NHSCR and that the ratio of total out-migrations by males/females for each origin in the NHSCR

Table 14.3 *Male/female ratios in migration outflows by age in Census and NHSCR data for 1990–91 for four FHSAs*

FHSA	Dataset 1990–91	Ages 0–15	16–19	20–24	25–29	30–44	45–59	60+
Newcastle	SMS	.92	.98	.99	1.12	1.18	1.16	.58
	NHSCR	.97	.65	.75	1.03	1.3	1.31	.67
North Tyneside	SMS	.93	.90	.87	1.12	1.23	1.14	.64
	NHSCR	1.02	.70	.67	.91	1.36	1.08	.68
Barnet	SMS	1.08	.75	.78	1.02	1.16	1.11	.65
	NHSCR	1.09	.61	.65	.76	1.07	1.00	.71
Kensington & Chelsea	SMS	1.06	.87	.73	.86	1.17	1.38	.76
	NHSCR	1.10	.69	.50	.60	.93	1.21	.81

should be the same as that observed in the SMS. This means that for each age group and for each FHSA the following equality holds:

$$M(n)/F(n) = M(s)/F(s) \tag{14.1}$$

where $M(n)$ and $F(n)$ are the numbers of male and female out-migrants, respectively, from the NHSCR and $M(s)$ and $F(s)$ are the respective values from the SMS. Consequently, we can construct a better estimate of $M(n)$, $M(n)^*$, by deriving separately for each origin the revised number of male out-migrants:

$$M(n)^* = [M(s)/F(s)]F(n) \tag{14.2}$$

For example, in the case of Newcastle, the 1990/91 NHSCR records a total of 214 male out-migrants. Using the above adjustment, the new total would be 0.98×330 (the number of 16–19 female movers from Newcastle recorded in the NHSCR) giving a value of 323. In a similar way, the total number of males 20–24 leaving Newcastle would rise from 1416 to 1869. This process was repeated for each origin for all seven age groups of migrants and across all 15 periods of the NHSCR. The assumption is made that the ratios of males/females is relatively constant through time and is recorded relatively accurately via the SMS.

These adjusted matrices were then used in the Stage 1 modelling procedure described below. To examine the impact of the adjustment, we calibrated a version of the Stage 1 out-migration model with the two sets of out-migrations (unadjusted and adjusted). Although the results are broadly similar, there are some differences in the estimated parameters and in seven cases different conclusions would be reached about the effect of individual variables on out-migration. In five of these seven cases the parameters change in terms of being significant at the 95% level and in the other two the sign of the parameter changes.

14.5.3 Migration Determinants

The data on the determinants of migration, both out-migration and destination choice, fall into three broad groups: (i) a relatively small set of *national* economic indicators; (ii) a much larger set of purely cross-sectional *zonal* variables; and (iii) a set of variables that are measured both over space and over time. The selection of many of the variables was informed by the findings of the review *The Determinants of Migration Flows in England* (Champion *et al.*, 1998), which identified seven factors influencing migration rates and places' attractiveness to migrants. These were demographic, cultural and social, labour market, housing, environmental, public policy and impedance. Below we provide a brief justification for seeking variables to represent each of these categories in the models.

14.5.4 Types of Determinant Variables

The two most important distinguishing features of the data assembled to represent the determinants of migration are whether the variable is national or local, and

whether the observation is in the form of a time series or a single point in time. For any cross-sectional variable we can also compute a *regional* version of the variable (see below) which measures the incidence of that variable in regions other than the one from which out-migration rates are measured. In this way, in order to model out-migration rates, we can capture the pull effects of conditions in other regions as well as the conditions from which out-migration takes place.

All the national-level variables are time-series, with observations for each of the 14 years being modelled in Stage 1. In Table 14.2 these are denoted with the suffix 'Z'. A number of the local-level variables are also time-series in nature. Where this is the case, we have used a one-year time lag for the equivalent variable in the Stage 1 modelling to avoid problems of circularity. Lagged variables are denoted with the suffix 'L'. For the remainder of the local variables, there is only one observation time point. The most common reason for this is that the data are taken from the Population Census and therefore relate to 1991. Fortunately, most of the variables affected are ones that change only rather slowly and the year 1991 is around the mid-point of the Stage 1 modelling period.

Regional variables (labelled with Y suffix) are calculated for the Stage 1 out-migration model. These variables are meant to capture the possible pull effects on out-migration caused by conditions elsewhere in the country. After experimentation with alternatives, the following formula was used for the regional variables:

$$Y_i = \left[\sum_{j\,j\neq i}(X_j/X_i)d_{ij}^{\beta} \Big/ \sum_{j\,j\neq i} d_{ij}^{\beta}\right] \tag{14.3}$$

where X_i is the value of X at location i and X_j represents the value of X at one of the other FHSAs. The formula produces a distance-weighted average ratio of X_j to X_i where nearby locations are weighted more heavily in the calculation than more distant ones. The value of β was taken as –2 which, from experience, gives a reasonably differentiated surface of Y values. Values of β less negative than this give a surface which is smoother; values of β more negative will give a spikier surface. Values of $Y_i > 1$ indicate that X_i is generally *smaller* than its neighbours. Values of $Y_i = 1$ indicate that X_i is generally very similar to its neighbours. Values of $Y_i < 1$ indicate that X_i is generally *larger* than its neighbours. One point to note with the formula for the regionalised variables is that the value of Y will tend to infinity as X_i tends to zero. In the few instances where this occurred, small values were substituted for zero. Also, the formula should not be used if the X values take both negative and positive signs.

14.5.5 National Indicators

A number of factors can be considered to affect the overall volume of migration taking place in the UK. These are time-series variables that are used only in the Stage 1 (migration generation) modelling. They are primarily related to the state of the economy: although we considered a wider set in Phase 1 of the work. The national variables that survived for input to the Stage 1 model in Phase 2 were GDP change (RGDPCH_Z), unemployment rate (UNEMP_Z_L), mortgage interest rate

(MGINTR_Z) and the number of university places (UNIVPL_Z_L), together with a time trend indicator (TIME_Z). These variables influence people's willingness and ability to move house or to set up home independently.

14.5.6 Demographic Variables

These variables are area-specific, with each zone being potentially able to take a different value. To varying extents, most of them are considered relevant to both model stages, because they are considered to influence both the extent to which people are likely to leave a zone and, for these migrants, the choice of one destination zone over another. The most important demographic factor influencing the propensity to move house – the age of a zone's population, with its linkage to stage in the life course – is already accounted for by out-migration being modelled for the 14 separate age–sex groups.

It was, however, considered important to have a direct measure of students in higher education, because of their high level of longer-distance mobility. Considerable effort was therefore put into generating two variables for students (Bailey and Rees, 2001). One variable, PARDOM, measures their presence at their parental/vacation address (from which most would be leaving for university and to which many might be expected to return). The other, TERMT, is for their term-time address (to serve initially as destination and then, at the end of their studies, as origin). Note that such special treatment does not need to be given to Armed Forces personnel, the other distinctive high-mobility subgroup, because their movements are not covered by the NHSCR.

The other specifically demographic variables included in the model relate to total population size (POPN), rate of immigration from outside the UK (incorporated in the URBAN composite variable), incidence of divorce (ASDIVRT), distribution of the population by marital status (SINGLE, WIDOW) and composition by household type (SINGELD).

14.5.7 Socioeconomic and Cultural Variables

The likelihood of people moving house, and therefore possibly leaving an area, is strongly related to their socioeconomic and cultural attributes. In particular, past research has highlighted the migration differentials between groups based on ethnicity, occupation/social class and income. One key variable presented under this heading, therefore, is the proportion of the population that is non-white in ethnic origin (NONWH, NONWH_Y), used in the Stage 1 model. Several measures were assembled to represent occupation/social class including the proportions in groupings by Social Class and by Socioeconomic Group, but after a data reduction exercise just one was used for the Stage 1 (migration generation) modelling. An age–gender specific index (OCCMIG) has been generated to represent the relative chances of out-migration from each zone based on the occupational structure of its population. Several measures of income were assembled and, after investigation

for multicollinearity, measures of average household income (HHINC_L and HHINC_Y_L) and the concentration of those on low incomes (LOW_L, LOWI_Y_L) were input to the Stage 1 model calibration and HINC was used in the Stage 2 model. We also included an educational measure, GCSE_L, in the Stage 1 model inputs.

14.5.8 Labour Market Variables

Labour market factors are seen as potentially important both in prompting out-migration from an area and in influencing people's destination choices. They would normally be interpreted as measures of the overall economic environment of the area, but some can also stand as factors affecting the propensity of individuals to move away from an area. A few variables were assembled to represent directly the labour market situation, including levels of employment (EMPR_L and EMPR_Y_L at Stage 1 and EMPR at Stage 2) and changes in jobs (EMPGRO_L at Stage 1, EMPGRO at Stage 2) and unemployment (ASUNEM_L and ASUNEM_Y_L at Stage 1 and ASUNEM at Stage 2). Beyond this, a number of the variables included above (under the 'social and cultural' heading), particularly those related to occupation and income, can also serve as indicators of the economic performance of areas, as can measures of the tightness of the housing market.

14.5.9 Housing Variables

Housing factors form a critical element underlying migration patterns in the UK, but they have complex interactions with migration and need especially careful treatment. On the one hand, some housing measures such as high house prices and low vacancy rates can reflect the strong economic performance of an area and indeed of neighbouring areas within commuting distance. On the other hand, these factors can directly influence the opportunities for in-migration, with high house prices acting as a deterrent and high vacancy rates as an attraction. Even here, however, the effect may not be as simple, as high house prices may lead to more rapid in-migration in anticipation of future price rises and very high vacancy rates may serve to undermine confidence in an area and deter people from moving there. The size, composition and quality of the housing stock can also influence both the level and the type of migration. Housing tenure, besides being a reflection of the social composition of an area, is also known to affect migration patterns, most notably the well-documented problems that people moving between local authority areas have in accessing council housing. A large number of housing variables were therefore assembled for the model. House prices are measured by HPRICE_L and HPRICE_Y_L at Stage 1 and by HPRICE at Stage 2. Vacancy rates are captured by PNRL_L, PVAC_L, PVAC_Y_L at Stage 1 and by PNRL, PVAC at Stage 2. House building is represented by PQPR_L, PQPR_Y_L, PQSR_L, PQSR_Y_L at Stage 1 and PQPR and PQSR at Stage 2. Housing condition is measured by NOCENT, NOCENT_Y,

RLAPSC, RLASC and RLASC_Y at Stage 1 and RLAPSC, RLASC at Stage 2. Housing tenure is input through TER at Stage 1.

14.5.10 Environmental Variables

In the current era within advanced economies, environmental factors play a major role in people's residential moves, both in prompting exits from areas and in acting as 'pull' factors. The term is used here in its broadest sense, covering all the physical, economic, social and political aspects that affect both the everyday quality of life and the longer term trends in life chances. This heading can therefore be considered to include most of the factors mentioned under the other headings above, insofar as they bear upon the overall quality of an area and of the neighbourhoods that it comprises.

Among those not specifically considered above are variables relating to derelict and vacant land (VACDRL and VACDRL_Y at Stage 1 and VACDRL at Stage 2), variables relating to the pattern of development, such as the proportion of new housing on brownfield land (PNBU at Stages 1 and 2), variables relating to population density, settlement size and level of urbanisation (combined into a composite variable URBAN), variables relating to crime (CRIME at Stages 1 and 2), variables relating to climate (CLIMATE at Stages 1 and 2) and air quality (AIR at Stage 1). Also included under the environment theme are variables relating to physical attractiveness, such as accessibility to scenic areas (SCENIC at Stages 1 and 2), and listed buildings (LISTED at Stages 1 and 2 and LISTED_Y at Stage 1), number of visitors (VISITS at Stage 2), the extent to which an area acts as a dormitory for commuters to relatively distant jobs (COMMUT at Stage 1) and accessibility to international air passenger connections (AIRPORTS at Stage 2). A 'bright lights' indicator measuring access to theatres and concert halls was incorporated into the URBAN composite.

14.5.11 Public Policy Variables

Public policy variables relevant to migration behaviour include not only direct interventions such as migration incentives and immigration policy but also indirect influences through the uneven effects of government grants, local taxes, defence spending, higher education expansion and the amount and location of land approved for house building. Two variables relating to expenditure on local government services, namely overall expenditure per capita and expenditure relative to the Standard Spending Assessment, were explored in Phase 1 but dropped because of multicollinearity. Level of council tax (CTAX) was included in the Stage 2 model.

In general, it was believed more satisfactory to estimate the role of public policy, past or anticipated, by reference to variables representing the aspects that public policy seeks to alter. For instance, the migration impact of a regional development initiative can be assessed by reference to, for instance, the number of extra jobs, while the impact of a policy that alters the availability of land for

housebuilding can be studied via changes to the number of housing completions. This is the approach adopted for assessing the consequences for migration of alternative policy scenarios.

14.5.12 Impedance Variables

The term 'impedance' is used here to refer broadly to the friction of distance, which is the single most important factor affecting movers' choice of destination. For a whole set of reasons, people changing address tend to move over short distances; indeed, three-fifths of movers recorded by the 1991 Population Census moved less than 5 kilometres. All other things being equal, places that are further apart therefore tend to have less to do with each other than places that are close together. Secondly, areas that share a boundary tend to have more migration between them, because this migration will include a proportion of short-distance moves from one side of the boundary to the other. Thirdly, areas situated in high population-density regions are likely to be less attractive to migrants, everything else being equal, because of increased spatial competition between destinations.

These three spatial processes are allowed for in the Stage 2 (migration distribution) component of the model, using respectively:

- a network-weighted distance variable, DIJNETW, based on distances between the population 'centroids' of pairs of zones, calculated on the basis of shortest surface route rather than straight-line distance, so as to take account of the effect of estuaries;
- a contiguity variable, CONTIG, taking the value of 1 for zones that share boundaries with each other, 0 for other pairs of zones;
- a destination accessibility variable, DESTACC, which measures the degree of spatial competition faced by a destination from nearby destinations.

The distance variable was used indirectly in the Stage 1 model as it is used in the computation of all regional variables.

14.6 Stage 1: A Model to Predict Total Out-migration from Each Origin

The model consists of two stages which recognise that migration results from two factors: the desire to leave an origin and the choice of destination. The model therefore consists, firstly, of a model to predict migration outflows from each area and, secondly, of a model to allocate the flows of migrants from each origin area to each of the alternative destinations. This section outlines the Stage 1 model. The volume of results from this exercise is substantial and beyond the limitations of this chapter. They are left for detailed discussion elsewhere. However, the end product of the project was the combination of both the Stage 1 and the Stage 2 model results into a user-friendly software package. We describe some sample results from this software to show how the overall package operates. We also describe the changes in both the

Stage 1 and Stage 2 modelling procedures as we moved from the prototype in Phase 1 to the final product in Phase 2.

14.6.1 The Phase 1 Out-migration Model

The volume of out-migration from an origin zone i is predicted as

$$O_{it}^{m} = omr_{it}^{m} P_{it}^{m} \tag{14.4}$$

where O_{it}^{m} is the total out-migration of migrant group m from zone i in time interval t, omr_{it}^{m} is the out-migration rate from origin i in time unit t for migrant group m (one of the 14 age–sex groups) and P_{it}^{m} is the population of migrant group m at risk of migrating from origin i during time interval t (usually estimated as the mid-interval population or the average of start of interval and end of interval populations). The out-migration rate is in turn modelled, in general form as

$$omr_{it}^{m} = f\left(X_{it/t-1}^{m}, Y_{it/t-1}^{m}, Z_{t-1}\right) \tag{14.5}$$

where omr_{it}^{m} is the out-migration rate for migrant group m from zone i in time interval t, $X_{it/t-1}^{m}$ is a vector of origin attributes in either year t or $t-1$ (lagged by one year); $Y_{it/t-1}^{m}$ is a vector of distance-weighted attributes describing the situation in other areas in either year t or $t-1$ and Z_{t-1} is a vector of attributes describing the national economic situation as it affects the overall volume of migration in year $t-1$ (lagged by one year).

In total, 14 models were calibrated, one for each migrant group defined on the basis of seven age groups and the two sexes. The models were calibrated using NHSCR data on total outflows from each of the 98 FHSA areas in England and Wales. Once calibrated, the Stage 1 model allows the total volume of out-migration from each area to be predicted, given conditions of that area, of surrounding areas and of the national economy.

What specific form should equation (14.5) take? Should the form be multiplicative or additive? Should we use logged functions of the variables (the multiplicative option) or leave them unlogged? Should non-linear forms of the variables such as quadratic forms be considered? In Phase 1, we adopted a multiplicative model with the specific form:

$$omr_{it}^{m} = \exp(K^{m}) A^{m} V_{1}^{alm} \ldots V_{53}^{a53m} \exp(a_{54}^{m} V_{54}) \ldots \exp(a_{58}^{m} V_{58}) \tag{14.6}$$

where the term $\exp(K^{m})$ is the intercept in the log–log regression, specific to each age and sex group. The age and sex specific adjustment factor A^{m} is necessary because of a statistical bias in the intercept estimate in a log–log regression. The adjustment factor is obtained by constraining the total predicted out-migration of a specific migrant group for all 98 origins to equal its observed value for the 1996/97 time period. The model contained 58 explanatory variables. The variables V_{1i} to V_{53i} were

explanatory variables which became log functions in the log–log regression, while the variables V_{54i} to V_{58i} could not be logged due to negative values and were represented in the exponential form. The regression parameters, represented by power coefficients, a_v^m, are specific to each variable v, and to each migrant group m.

The Stage 1 model developed in Phase 1 has a number of drawbacks, which were addressed in Phase 2 of the project. The first was the strong possibility of multicollinearity between determinant variables. This was addressed by both qualitative and quantitative means resulting in a much-reduced set of potential explanatory variables and a small set of composite indices constructed using principal components analysis. The second difficulty was the rather awkward combination of power and exponential relationships. This was addressed by switching to a straightforward linear model. The third problem was the assumption that a monotonic relationship existed between a dependent variable and determinant variables. This was addressed by introducing quadratic forms of each variable entered into the models. The fourth drawback of the Phase 1 model was a tendency for out-migration from London FHSAs to be underpredicted. This was handled through the introduction of a London dummy variable. The London effect we think results from the different nature of internal migration in London, with migration flows being within a labour/housing market area rather than between such market areas. The fifth was that the model form made no allowance for possible time trends in migration (not accounted for by the set of time varying determinants used). A time trend was introduced to address this problem. The sixth difficulty was the arbitrary way variables were selected for inclusion in the model. In Phase 1, we had aimed at a common model across all migrant groups and selected the variables based on a majority of migrant group models reporting significance for particular variables. In Phase 2, stepwise inclusion based on significance of parameters was used for each age–sex group to determine a set of 14 age–sex-specific final models.

14.6.2 Investigations of the Phase 1 Out-migration Model Problems

Modelling Time Trends

To gain an understanding of the level of variation in out-migration volumes both across space and over time, we graphed out-migration volumes over time for each of the seven age groups (males and females combined) for all 98 FHSAs. Figure 14.1 illustrates trends in out-migration by the age groups used in the model for one FHSA, Newcastle. The main findings from the 98 graphs are as follows.

- There is a great deal of stability over time in the out-migration rates for the age groups 0–15; 30–44; 45–59; and 60+. This is true across the country and for many FHSAs the relationships are essentially flat. The main variations in the rates for these age groups are cross-sectional.
- Migration rates tend to be highest for the age groups 20–24, 25–29 and 16–19 and lowest for the age groups 60+ and 45–59.

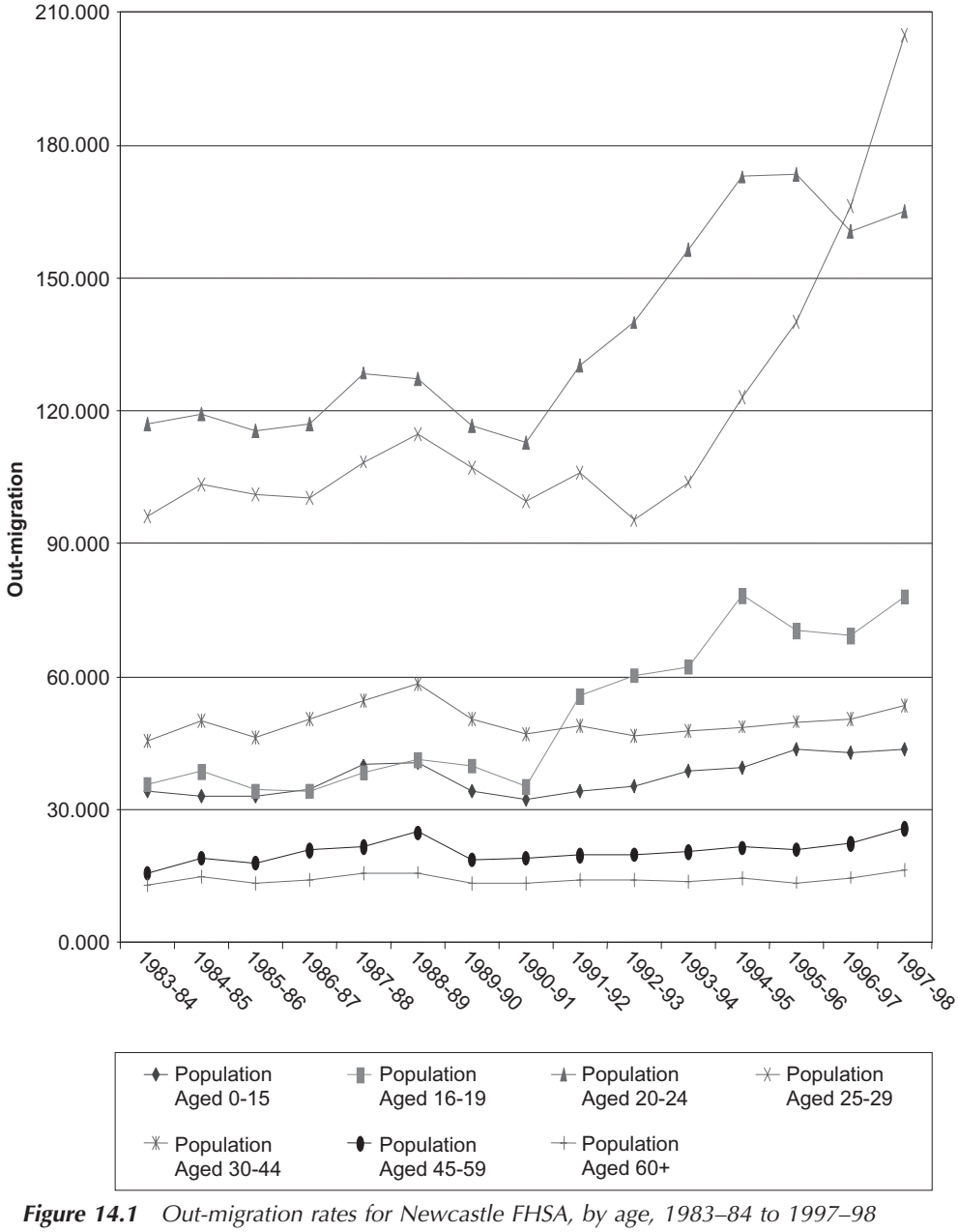


Figure 14.1 *Out-migration rates for Newcastle FHSA, by age, 1983–84 to 1997–98*

- There has been a rapid and strongly linear rise for almost all FHSAs in out-migrations by the 16–19 age group since 1990. Before this date, the level of out-migration for 16–19 year olds was relatively constant. The rise is due to the rapid expansion of higher education in the 1990s. We therefore added a variable to the Stage 1 model, UNIVPL_Z, the total number of university places available in a year.
- For FHSAs with relatively large student populations, such as Newcastle, concomitant with the rise in out-migration by the 16–19 age group through the 1990s, has been a rise in out-migration by the 20–24 and the 25–29 age groups. This is highly likely to be due to graduated students leaving university.
- There is little evidence to support any particular time trend otherwise in out-migration rates over time.

As well as plotting out-migration rates over time, all the other time-varying variables were also plotted over time to get a better 'feel' for the data. Many trends appeared to be highly erratic over time which, to anticipate the results of the Stage 1 model calibrations, probably explains the lack of explanatory power in some of these variables. The Stage 1 model contains some time series variables that might need special treatment in the model because of trends in some of the observations and some of the determinants. In Phase 1, no special treatment was given to these variables.

There are different strategies, of varying degrees of complexity, for handling time series variables in regression. To a large extent, the choice of a particular strategy depends on asymptotic properties of the time series variables, both dependent and independent, that are difficult to assess, especially in the case here with only 15 different time periods. Basically, the issue is whether any trends that are observed are stochastic or deterministic. Different strategies need to be followed depending on the answer to this question: stochastic trends are better handled using a differencing approach; deterministic trends are better handled using a de-trending approach. Although there is a vast literature on this subject, there is no real consensus on how to proceed in our situation where we have only 15 time periods and no real way of assessing the nature of any time-dependency. This severely limits what we can do. It should also be noted that although we are using 15 years of data to estimate certain relationships in the out-migration model, the majority of determinants are cross-sectional. However, despite these caveats, we conducted the following experiment.

- We added a linear trend to the out-migration model. This is the simplest and perhaps most effective way to account for time trends in both the dependent and independent variables. A non-linear trend, such as a quadratic, could be added if the data suggest such a trend exists although there is no real evidence for such a function.
- We transformed all time-varying variables into first differences by subtracting the value in time $t - 1$ from the value in time t. One obvious result of this procedure is that we were left with only 14 time periods. Variables that are constant over time were not first differenced.

A 'plausible' Stage 1 model was constructed and the experiment run with males 30–44 migrants (the largest volume of migrants). The following calibrations of this model were then undertaken:

- a 'baseline' model with no time trends taken into account;
- a model with a linear time trend added; and
- a model using first differences.

The results demonstrated the superiority of using a linear time trend. The use of first differences gave a very poor model fit and most of the cross-sectional variables were statistically insignificant. There was a modest improvement in goodness-of-fit in using a linear time trend over using no time trend. As anticipated, the two models yield broadly similar results for the parameter estimates of the cross-sectional variables but there are some variations in parameter estimates for the time-varying variables, which are more accurately estimated in the model including the time trend. Several other forms of the Stage 1 model were calibrated with and without the time trend and all the results supported the conclusion that adding a linear time trend provided slightly better results. Consequently, the Stage 1 models results used in the MIGMOD software were obtained from models containing a linear time trend.

OLS or WLS?

In Phase I, the Stage 1 models were calibrated with WLS with extra weight given to data pertaining to health areas with large populations. The justification for this was that the dependent variable in the regressions is migration rate which might be less reliable for origins with small populations. However, the justification may not be particularly strong in this case as the FHSAs all have fairly large populations. Consequently, we examined the efficacy of WLS compared with OLS on a sample of models. The results of applying the two calibration techniques were virtually identical so that there is little rationale for using WLS over OLS. Consequently, we employed OLS for the Stage 1 model calibrations, which avoids the need for a somewhat speculative definition of the weighting variable.

Unlogged or Log–Log Format?

In Phase I, the Stage 1 models were calibrated in log–log format. This had the disadvantage of producing a biased estimate of the intercept which had to be corrected after the calibration had taken place by the addition of an adjustment factor to the model format. This was done originally because the log–log format appeared to give slightly better predictions of out-migration. To examine whether this advantage still held, given the change in the Stage 1 models, we performed the following experiment.

Two plausible Stage 1 models were identified that contained the variables listed in Table 14.2. These models differed only in that one contained a linear time trend and the other did not. Both models were then calibrated by unlogged regression and then

Table 14.4 *Goodness of fit (R^2) for model tests*

Model	Unlogged	Log–Log
Without time trend	0.924	0.907
With time trend	0.928	0.918

again in a log–log format. The adjusted R^2 values from the four calibrations are shown in Table 14.4. These results suggest that an unlogged regression is superior to a log–log format. This also has the added advantage of producing an unbiased estimate of the intercept so that an adjustment factor is not needed in Stage 3 (the front-end) to balance the predicted values. Results from the calibrated models in the form of scatterplots of the standardised residual against the standardised predicted value confirmed the validity of the unlogged form of the model. The plots exhibit no evidence of heteroscedasticity and no evidence of any non-linearity.

Lagged or Unlagged Time-Varying Variables?

Some of the variables in the Stage 1 dataset vary over time. It is an issue as to whether these variables are entered in the model with their current year values (unlagged) or the previous year's values (lagged). The answer relates to whether migrants respond more to current conditions or to conditions in the previous year (the justification for the latter being a time lag between the stimulus to move and the move itself). It is difficult, if not impossible, to argue convincingly for either approach and so the question then becomes an empirical one – which form of the variables generates the more accurate model?

To answer this question a plausible Stage 1 model was calibrated for Males 30–44 using the variables described in Table 14.2. In one version of the model the time-varying variables were lagged; in another they were unlagged. The lagged model produced an adjusted R^2 value of 0.952, while the unlagged model R^2 value was 0.935. It was therefore decided to use lagged versions of the time-varying variables.

14.6.3 The Phase 2 Out-migration Model

From the resolution of the set of issues discussed above, the final form of Stage 1 of the model was defined. The Stage 1 model relates the adjusted out-migration rate for migrant group m in FHSA i at time t to a series of independent variables which describe the economic, social and geographic conditions of i. The independent variables are of three types: cross-sectional variables (some of which also vary over time); regional variables; and national time series variables. The linear model of out-migration rates has the basic form:

$$omr_{it}^{m} = \kappa^{m} + \sum_{p} \alpha_{p}^{m} X_{pit}^{m} + \sum_{q} \beta_{q}^{m} Y_{qit}^{m} + \sum_{r} \gamma_{r}^{m} Z_{r}^{m} + \varepsilon_{it}^{m} \qquad (14.7)$$

To this were added quadratic terms, a linear time trend and a London dummy:

$$\begin{aligned} omr_{it}^{m} = {} & \kappa^{m} + \sum\nolimits_{p} \alpha_{p}^{m} X_{pit/t-1}^{m} + \sum\nolimits_{q} \beta_{q}^{m} Y_{qit/t-1}^{m} + \sum\nolimits_{r} \gamma_{r}^{m} Z_{rt-1}^{m} \\ & + \sum\nolimits_{p} \delta_{p}^{m} \left(X_{pit/t-1}^{m} \right)^{2} + \sum\nolimits_{q} \eta_{q}^{m} \left(Y_{qit/t-1}^{m} \right)^{2} + \sum\nolimits_{r} \theta_{r}^{m} \left(Z_{rt-1}^{m} \right)^{2} \\ & + \psi^{m} T_{t} + \zeta^{m} LD_{i} + \varepsilon_{it}^{m} \end{aligned} \tag{14.8}$$

where omr_{it}^{m} is the predicted out-migration rate for migrant group m in FHSA i in time interval t, X represents a cross-sectional variable (some of which can also be time-varying); Y represents a regional variable; and Z represents a national time-varying variable, T is a time counter and LD stands for London dummy (= 1 if the FHSA is in London, = 0 otherwise). The regressions coefficients are denoted by Greek letters and ε_{it}^{m} is the error term for each zone, time and migrant group combination. The model is calibrated separately for each of 14 migrant groups (males and females in each of seven age groups).

14.7 Stage 2: A Model of Destination Choice

14.7.1 The Model

The Stage 2 model is a migration destination model that distributes the total number of out-migrants from FHSA i to each of the 97 destination FHSAs based on the characteristics of each destination FHSA and the separation between the origin and each destination. It is a spatial choice model of the destinations chosen by migrants from an origin. The model to be calibrated has the general form:

$$M_{ij}^{m} = O_{i}^{m} \prod\nolimits_{p} X_{pj}^{\alpha pim} d_{ij}^{\beta im} \Big/ \sum\nolimits_{j} \prod\nolimits_{p} X_{pj}^{\alpha pim} d_{ij}^{\beta im} \tag{14.9}$$

where O_{i}^{m} is the volume of out-migration of type m from origin FHSA i (known when the model is calibrated, predicted by the Stage 1 model when predictions are being made at Stage 3); X_{pj} represents an attribute of FHSA j that affects the choice of j by migrants from i; and d_{ij} is the distance between i and j. The X variables are raised to powers, α_{pim}, specific to each variable p, origin i and migrant group m, while the distance variable is raised to the power β_{im} specific to each origin i and migrant group m. The parameters of this model indicate the sensitivity of migration flows to particular destination characteristics: in essence they indicate what features of a destination make it attractive to migrants and which features make it unattractive. For example, a relatively large score on an attribute with a positive parameter estimate would make a destination attractive to migrants, *ceteris paribus*, while a relatively large score on an attribute with a negative parameter estimate would make a destination unattractive to migrants, other things being equal. This model has a long history in the analysis of spatial interaction patterns (see *inter alia*, Fotheringham and O'Kelly, 1989, and Fotheringham *et al.*, 2000, Chapter 9).

The model is calibrated separately for each of the 98 origins and each of the 14 migrant groups. For any origin, $O_i^m/\Sigma_j \Pi_p X_{pj}^{\alpha pim} d_{ij}^{\beta im}$ will be a constant (k_i^m) so that the origin-specific model is then simply,

$$M_{ij}^m = k_i^m \prod_p X_{pj}^{\alpha pim} d_{ij}^{\beta im} \tag{14.10}$$

Note that there will be $98 \times 14 = 1372$ sets of model parameters to be calibrated in this highly specific spatial interaction model.

14.7.2 Poisson Regression or OLS?

In Phase 1, we calibrated the Stage 2 model by Ordinary Least Squares regression by taking logs of both sides of the equation to make it linear-in-parameters. Although convenient, this has two problems:

- The dependent variable M_{ij}^m is a count of migrants and therefore is likely to be Poisson distributed rather than normally distributed. This means that one of the key assumption of OLS is probably not met. This is particularly a problem where the migration matrices contain lots of small flows as is the case here.
- A practical problem arises in taking logs of both sides of the model when M_{ij}^m is zero (as are many of the recorded flows). Taking the log of 0 yields negative infinity. The solution to this used in Phase 1 was to add a small increment (0.5) to all flows before taking logs. Although tests showed this does not appear to bias the results greatly, nevertheless it is a somewhat arbitrary procedure.

An alternative, and perhaps superior method of calibration, is to use Poisson regression which assumes the conditional mean of the variable we are trying to model has a Poisson distribution and avoids the need for making some approximation to zero flows. The use of Poisson regression for calibrating migration distribution models has been demonstrated in a number of papers (*inter alia*, Flowerdew and Aitkin, 1982; Flowerdew, 1991).

In order to compare the performances of Poisson and OLS regression for migration destination choice models, we conducted a series of experiments comparing Poisson and OLS calibration results on the same datasets. A relatively simple, but quite plausible, Stage 2 model was postulated containing the following explanatory variables (the model was constructed from the original set of variables from Phase 1): Population, Destination accessibility, Distance, Contiguity, Urban, Beach, Rain, Council Tax and Household Income. This model was calibrated separately for each of the 98 origins using migration data for males 30–44. The performance of the calibrated models in replicating migration flows was assessed using the modified psi statistic, ψ (Knudsen and Fotheringham, 1986), which demonstrated the superiority of Poisson regression in terms of goodness-of-fit to the observed data. Consequently, all the Stage 2 models were calibrated with Poisson regression.

14.7.3 Variables for Stage 2

In Phase 1, the Stage 2 dataset contained 69 variables. This set was greatly reduced following a qualitative assessment of each of the variables in the dataset and an examination of multicollinearity amongst the independent variables. This left a final set of 27 variables in the Stage 2 model which are listed in Table 14.2.

14.7.4 Migration Data

As with Phase 1, the migration data used in all calibrations of the Stage 2 model were for the 1996/97 time period. Previous results in Phase 1 suggested that the calibration results were either stable over time or that the 96/97 period provided a good average of the years since 1990 (see Champion *et al.*, 2001).

14.7.5 Modelling Procedure

The procedure for variable selection was different from that employed in Phase 1. In Phase 1 initially, a separate model was calibrated for all origins for each migrant group using all 69 explanatory variables. Then, to determine the final model form for each migrant group, variables were selected from the 69 if they were significant for >9 origins and the model containing only those variables was re-calibrated. Instead of this largely *ad hoc* procedure, in Phase 2 we calibrated the model containing the 27 variables in the Stage 2 dataset for each of the 98 origins and for each migrant group. This avoids the need for any *ad hoc* selection procedure and allows us to make comparisons of the parameter estimates across the 14 migrant groups.

One final note should be mentioned on how the Stage 2 modelling procedure differed from that in Phase 1. It was agreed with the Department that we should drop the modelling of migration flows from Scotland so that we had only 98 origins for which migration flows are estimated endogenously in Stage 3 of the procedure.

14.8 Stage 3: The MIGMOD Software

The out-migration and destination choice models described above (Stages 1 and 2) are combined within a user-friendly software package termed MIGMOD built around the GeoTools mapping software developed at the University of Leeds. MIGMOD contains all the data described above as well as the estimated parameters from the calibrations of the Stage 1 and 2 models. As shown below, the user has access to these in both tabular and, where feasible, mapped form at the click of a button. We now describe some of the features of the MIGMOD software.

The parameters from the calibrated Stage 1 model in equation (14.8) and the calibrated Stage 2 model in equation (14.10) were entered into MIGMOD along with

data on observed migrations and the determinants of both out-migration and destination choice. Within MIGMOD the level of out-migration from each origin is then computed by first estimating the out-migration rate from each origin and for each migrant group within that origin and then multiplying these rates by the relevant population size. This produces estimates of out-migration volumes for each migrant group within each origin. These out-migration numbers are then allocated across the destinations on the basis of the attributes of these destinations according to the Stage 2 model. The user can then alter conditions in any origin or set of origins and/or in any destination or set of destinations and examine the impact such changes have on the resulting migration patterns.

The starting point of MIGMOD is the initial menu as shown in Figure 14.2. This gives the user four options:

- view any of the coefficients from either the Stage 1 or the Stage 2 models;
- examine various aspects of the observed and predicted migration matrices;
- develop policy-related scenarios in which some combination of the independent variables in the Stage 1 and 2 models is changed and a new set of predicted migration flows is obtained which can then be compared with the original set of flows; and
- view any of the time-varying variables in a series of graphs over time.

For instance, selecting 'View Coefficients' produces a set of options on which coefficients you want to view. The user has a choice of a table of Stage 1 parameter estimates, as shown in Figure 14.3, or a matrix of origin-specific Stage 2 parameter estimates for any one of the 14 migrant groups as shown in Figure 14.4 for the

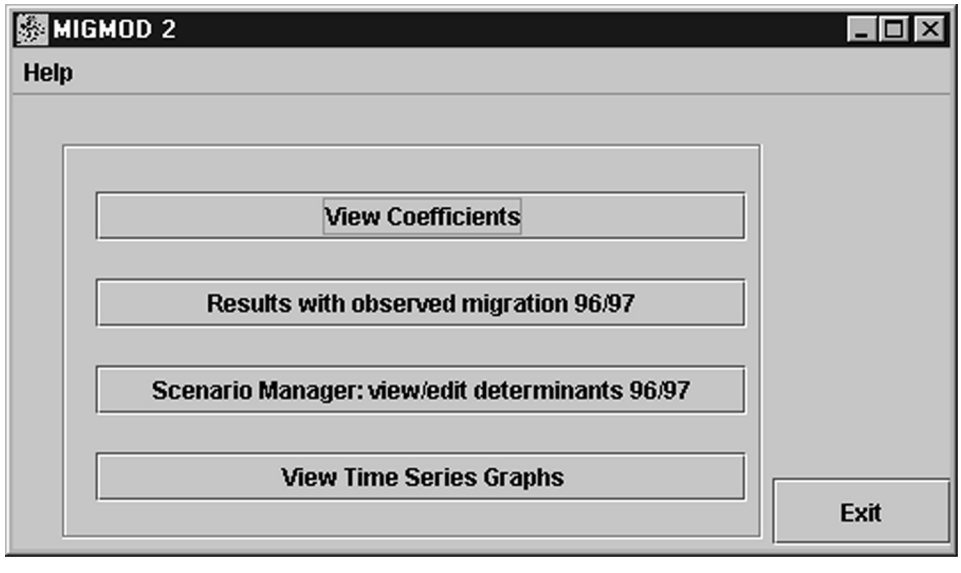
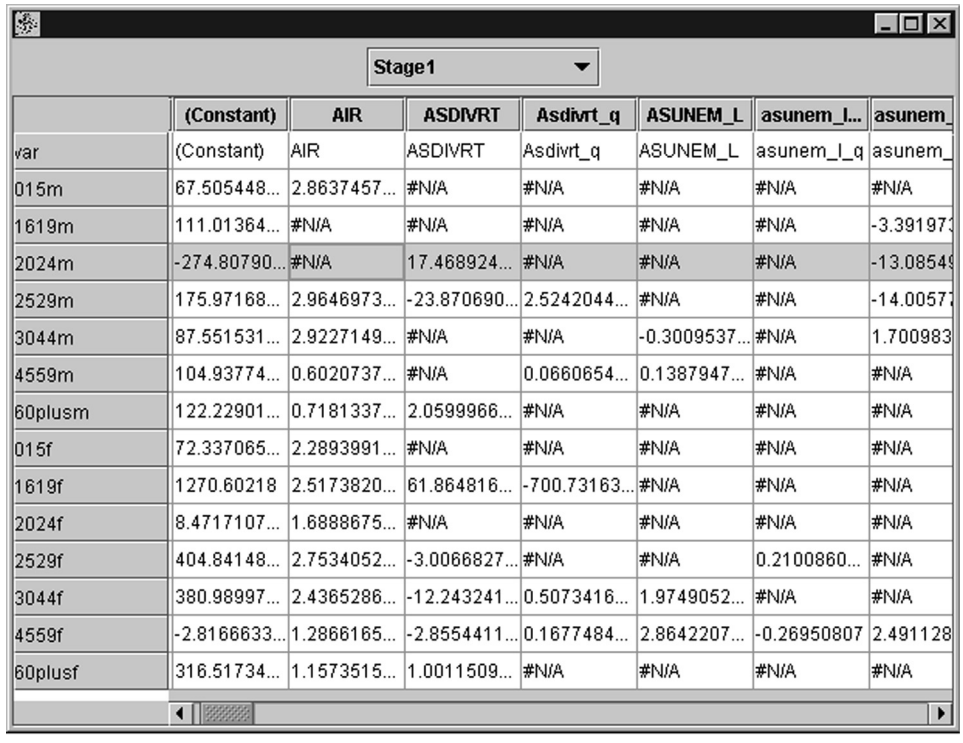


Figure 14.2 *MIGMOD main menu*



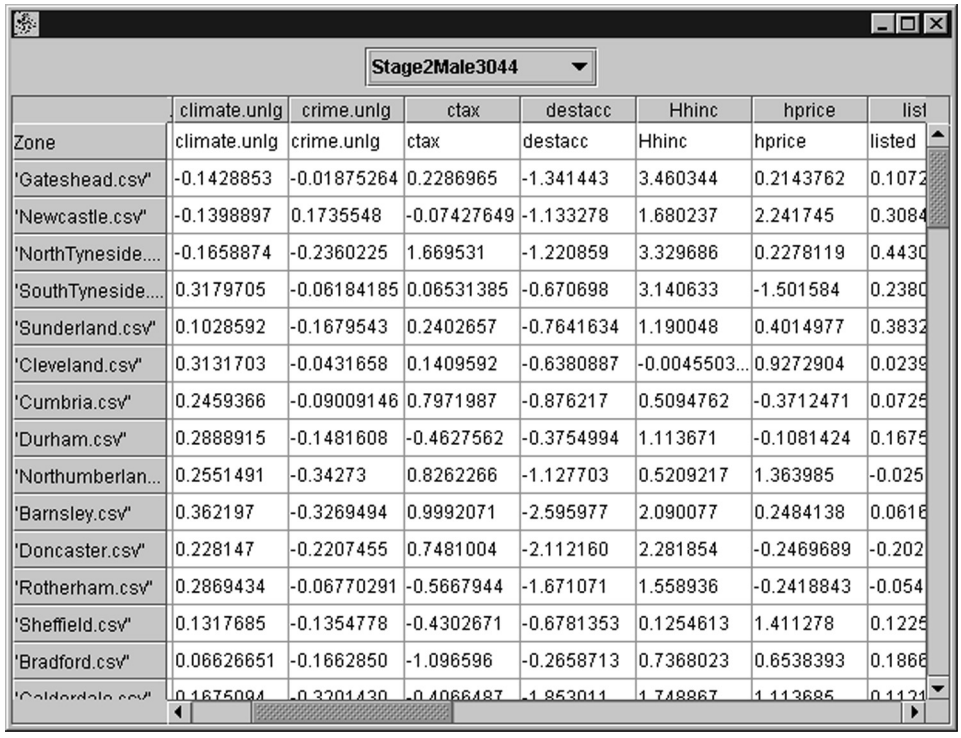
Stage1

	(Constant)	AIR	ASDIVRT	Asdivrt_q	ASUNEM_L	asunem_l...	asunem_
var	(Constant)	AIR	ASDIVRT	Asdivrt_q	ASUNEM_L	asunem_l_q	asunem_
015m	67.505448...	2.8637457...	#N/A	#N/A	#N/A	#N/A	#N/A
1619m	111.01364...	#N/A	#N/A	#N/A	#N/A	#N/A	-3.39197
2024m	-274.80790...	#N/A	17.468924...	#N/A	#N/A	#N/A	-13.0854
2529m	175.97168...	2.9646973...	-23.870690...	2.5242044...	#N/A	#N/A	-14.0057
3044m	87.551531...	2.9227149...	#N/A	#N/A	-0.3009537...	#N/A	1.700983
4559m	104.93774...	0.6020737...	#N/A	0.0660654...	0.1387947...	#N/A	#N/A
60plusm	122.22901...	0.7181337...	2.0599966...	#N/A	#N/A	#N/A	#N/A
015f	72.337065...	2.2893991...	#N/A	#N/A	#N/A	#N/A	#N/A
1619f	1270.60218	2.5173820...	61.864816...	-700.73163...	#N/A	#N/A	#N/A
2024f	8.4717107...	1.6888675...	#N/A	#N/A	#N/A	#N/A	#N/A
2529f	404.84148...	2.7534052...	-3.0066827...	#N/A	#N/A	0.2100860...	#N/A
3044f	380.98997...	2.4365286...	-12.243241...	0.5073416...	1.9749052...	#N/A	#N/A
4559f	-2.8166633...	1.2866165...	-2.8554411...	0.1677484...	2.8642207...	-0.26950807	2.491128
60plusf	316.51734...	1.1573515...	1.0011509...	#N/A	#N/A	#N/A	#N/A

Figure 14.3 *Stage 1 parameter estimates viewed through MIGMOD*

migrant group 'Males 30–44'. A very useful feature of MIGMOD is that if any column of data or parameter estimates relates to a set of regions, as the Stage 2 parameter estimates do, then simply double clicking on the column heading automatically generates a map of the data or parameter estimates. For example, a map of the spatial distribution of the house price parameter in the Stage 2 model for Males 30–44 is shown in Figure 14.5. This shows the interesting pattern that migrants from relatively wealthier origins primarily in the south of England view high house prices at a destination as an attractive feature whereas migrants from relatively poorer origins view high house prices as a deterrent. More information on such parameter variations is given in a future publication and can be found in the government report on this project (Fotheringham *et al.*, 2002).

Rather than viewing the calibration results, the user can select the option of viewing the original migration data, in which case the window shown in Figure 14.6 appears. This allows the user several options to view different facets of the migration data. Basically the user establishes two migration data sets termed 'Observed' and 'Baseline'. The 'Baseline' button is used to define one migration set which will henceforth be referred to as 'The Baseline'. The user is given a set of options ranging from the observed data, the predicted migration matrix from the baseline model and the predicted migration matrix from other runs of the model using different scenarios of



Stage2Male3044

	climate.unlg	crime.unlg	ctax	destacc	Hhinc	hprice	list
Zone	climate.unlg	crime.unlg	ctax	destacc	Hhinc	hprice	listed
'Gateshead.csv'	-0.1428853	-0.01875264	0.2286965	-1.341443	3.460344	0.2143762	0.1072
'Newcastle.csv'	-0.1398897	0.1735548	-0.07427649	-1.133278	1.680237	2.241745	0.3084
'NorthTyneside....	-0.1658874	-0.2360225	1.669531	-1.220859	3.329686	0.2278119	0.4430
'SouthTyneside....	0.3179705	-0.06184185	0.06531385	-0.670698	3.140633	-1.501584	0.2380
'Sunderland.csv'	0.1028592	-0.1679543	0.2402657	-0.7641634	1.190048	0.4014977	0.3832
'Cleveland.csv'	0.3131703	-0.0431658	0.1409592	-0.6380887	-0.0045503...	0.9272904	0.0239
'Cumbria.csv'	0.2459366	-0.09009146	0.7971987	-0.876217	0.5094762	-0.3712471	0.0725
'Durham.csv'	0.2888915	-0.1481608	-0.4627562	-0.3754994	1.113671	-0.1081424	0.1675
'Northumberlan...	0.2551491	-0.34273	0.8262266	-1.127703	0.5209217	1.363985	-0.025
'Barnsley.csv'	0.362197	-0.3269494	0.9992071	-2.595977	2.090077	0.2484138	0.0616
'Doncaster.csv'	0.228147	-0.2207455	0.7481004	-2.112160	2.281854	-0.2469689	-0.202
'Rotherham.csv'	0.2869434	-0.06770291	-0.5667944	-1.671071	1.558936	-0.2418843	-0.054
'Sheffield.csv'	0.1317685	-0.1354778	-0.4302671	-0.6781353	0.1254613	1.411278	0.1225
'Bradford.csv'	0.06626651	-0.1662850	-1.096596	-0.2658713	0.7368023	0.6538393	0.1866
'Calderdale.csv'	0.1675094	-0.3201430	-0.4066487	-1.853011	1.748867	1.113685	0.1121

Figure 14.4 *Stage 2 parameter estimates viewed through MIGMOD*

explanatory variables. The 'Observed' button is used in a similar manner to define a different set of migration data to that defined as 'The Baseline' and this is termed 'The Observed'. The 'Operation' bar then allows the following matrices to be displayed:

- View the 'Observed' Matrix;
- View the 'Baseline' matrix;
- View the 'Observed' – 'Baseline' matrix;
- View the absolute differences between 'Observed' and 'Baseline'; and
- View the percentage differences between the 'Observed' and 'Baseline'.

To this point all the operations are based on the cells of the full matrix. The button labelled 'Matrix' gives the user the option of viewing outflow totals, inflow totals and net flows for any of the migration matrices. The button labelled 'All Ages' allows the user to define which of the seven age groups should be displayed and one option here is to use 'All Persons'. The final option entitled 'Persons' allows the user to define whether the migration data displayed are for males or females or both. The use of MIGMOD for developing scenarios for policy evaluation is described elsewhere (Champion *et al.*, 2002; Fotheringham *et al.*, 2002).

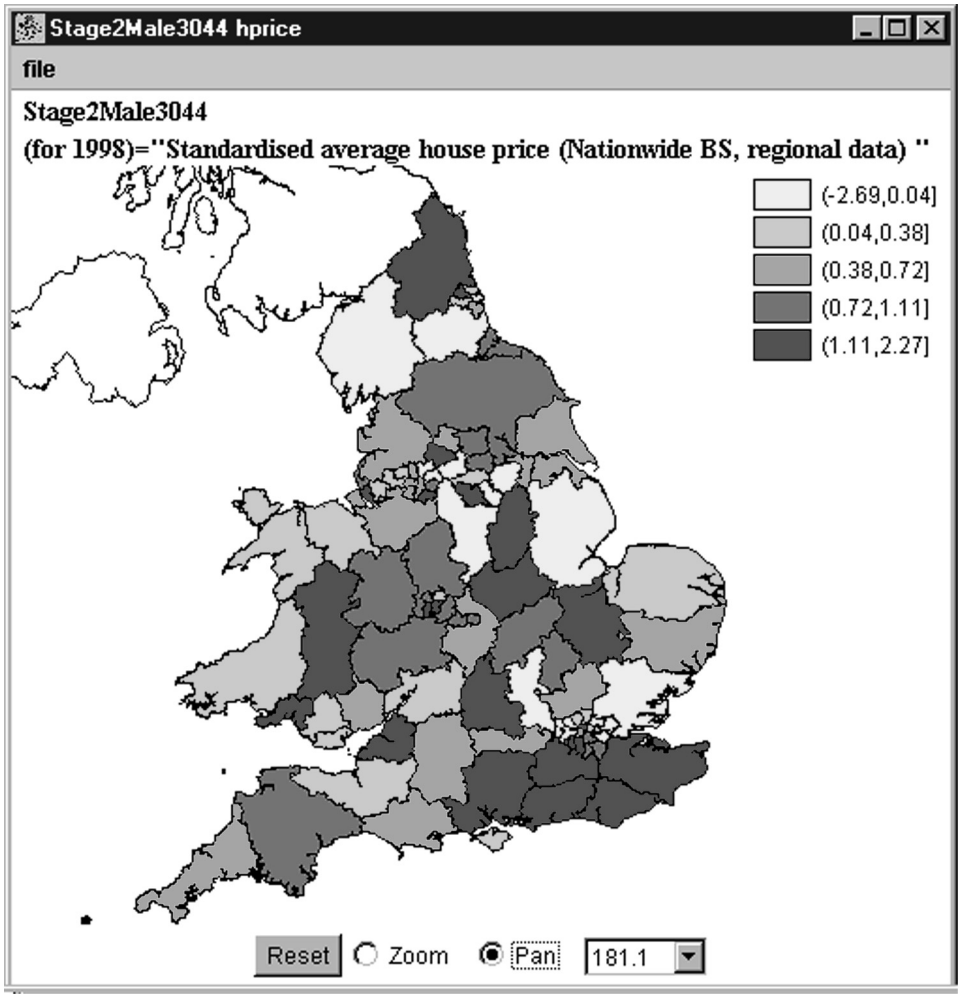


Figure 14.5 *The spatial distribution of the house price parameter in the Stage 2 model*

14.9 Summary

This chapter has described the structure of a very large spatial interaction model, constructed to explain the flows of migrants between 98 zones in England and Wales. The model was constructed in two phases and many changes were made in the second phase to overcome difficulties present in the first phase model. The two model stages, prediction of out-migration flows from origins and their distribution to destinations, were combined in Stage 3 into an operational model for the sponsoring Department so that scenarios related to policy could be implemented. Full details are provided in the Final Reports to the Department (Champion *et al.*, 2001; Fotheringham *et al.*, 2002) and in Champion *et al.* (2002). The authors believe that the model of

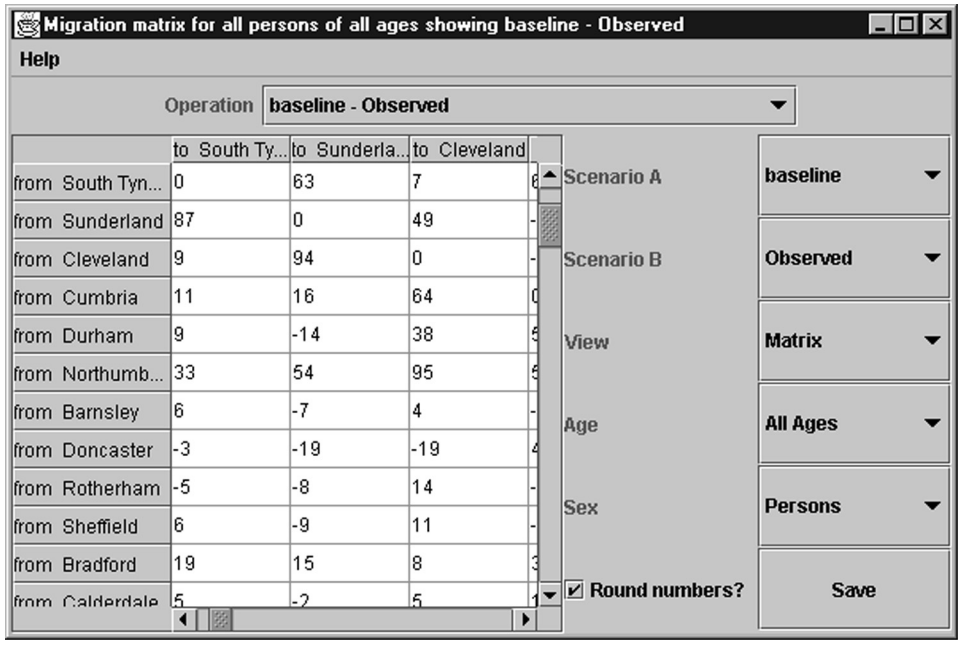


Figure 14.6 *The migration data viewer in MIGMOD*

migration constructed here to be one of the most comprehensive and meticulously researched such endeavours, which we hope will find utility in the formulation of English regional development and housing development policy in the future.

Acknowledgements

The Team

The migration model was constructed by a research team from four organisations: the University of Newcastle upon Tyne, University of Leeds, Heriot-Watt University/Edinburgh College of Art and Greater London Authority (formerly London Research Centre). The authors were supported in their work by other members of the team in data assembly, modelling and scenario development: *Newcastle* – Tony Champion (who led Phase 1), Seraphim Alvanides, Tim Barmby, Chris Brunsdon, Martin Charlton, Mike Coombes, Simon Raybould, Andy Tremayne; *Leeds* – Adrian Bailey, Oliver Duke-Williams, Heather Eyre (now with Education Leeds), James Macgill, John Stillwell; *Edinburgh* – Glen Bramley (who was responsible for scenario development), Hilary Anderson, Tania Ford; *Greater London Authority* – John Hollis and *University of Southampton* – Tom Wilson (now at the University of Queensland, Brisbane). The authors would also like to thank DETR's steering group led by Andrew Parfitt in 2000 and DLTR's steering group led by Stephen Rosevear in 2001,

who managed the contract so efficiently, and members Dorothy Anderson and Jenny Wood (ONS) who were helpful in identifying and supplying data.

The Sponsor

This chapter describes work carried out for the Department of the Environment, Transport and the Regions (DETR), restructured and renamed Department of Transport, Local Government and the Regions (DTLR) in June 2001 under contract RADS 5/9/22 (2000) and a variation (2001). The full report on Phase 1 and Phase 2 has been published by the Office of the Deputy Prime Minister (ODPM, 2002).

The Operational Model (MIGMOD, version 2.0)

An interface to the model was developed at the University of Leeds, primarily by James Macgill of the Centre for Computational Geography with help from Heather Eyre and Phil Rees and delivered to the Department on CD as a self-install and standalone software suite with accompanying observed migration datasets, the determinant variables, the model parameters, a set of predicted migrations and a system for developing alternative scenarios. MIGMOD is the intellectual property of the Department (currently ODPM). Readers who would like a copy of the operational model should seek permission from the ODPM.

Data Contributors

The research team is indebted to the many individuals and organisations that provided advice and help in the assembly of data. We are especially grateful to those at the Department who played a facilitating role in putting us in touch with data holders: Andrew Parfitt, Dominic Bryant, Russell Harris, Stephen Rosevear. In addition to those listed below against individual variables, Kerrick Macafee (the Department) and Glenn Everett (DTI) were also very helpful. We would also like to acknowledge the help and advice provided by members of the national statistical offices, especially in relation to data on international migration: *ONS*: Roma Chappell, Lucy Vickers, Alastair Davies, Denise Anderson, Brett Leeming; *GRO-Scotland*: Francis Hay; *NISRA*: Robert Beatty, Maire Rodgers, Maureen Ferguson. The following persons helped with particular variables. They are listed by variable with affiliation and name: agric, ONS (AES/ACE), Notice Ref 195R, via NOMIS (DETR), Carole Sutton, Simon Connell; air, AEA Technology, Emma Linehan; asunem, Employment Service/LFS via NOMIS, Andrew Craven (NOMIS); bank, ONS (AES/ACE), Notice Ref 195R, via NOMIS (DETR), Carole Sutton, Simon Connell; cars, DETR, Andrew Ledger; climat, Hadley Centre (UEA), David Viner; constr, ONS (AES/ACE), Notice Ref 195R, via NOMIS (DETR), Carole Sutton, Simon Connell; divrate, ONS, John Haskey; dlgexp, DETR, Welsh Office, Scottish Executive, Somerset CC, CIPFA, Stephen Greenhill, Ade Akinfolajimi, Justin Vetta, Stuart

Neil, Ingrid o'Malley, John Rae; earn, ONS (NES), Derek Bird, Claire Nichol; ecact, Employment Service/LFS via NOMIS, Andrew Craven (NOMIS); empgro, ONS (AES/ACE), Notice Ref 195R, via NOMIS (DETR), Carole Sutton, Simon Connell; empr, ONS (AES/ACE), Notice Ref 195R, via NOMIS (DETR), Carole Sutton, Simon Connell; expssa, DETR, Welsh Office, Scottish Executive, Somerset CC, CIPFA, Stephen Greenhill, Ade Akinfolajimi, Justin Vetta, Stuart Neil, Ingrid O'Malley, John Rae; frosty, Hadley Centre (UEA), David Viner; gcse, DfEE, Mark Franks; Catherine Blackham; hhinc, ONS (Regional Accounts), Dev Virdee, David Lacey; ijsa, DSS (ASD), Katie Dodd, Kath Williams, Jacky Javan; ilchip, DETR (1991 ILC), Gillian Smith, Sam Mason; immigr, ONS, Roma Chappell; jlytemp, Hadley Centre (UEA), David Viner; jvacs, Employment Service/LFS via NOMIS, Andrew Craven (NOMIS); listed, CADW, DETR, DoE(NI), RCME, Welsh Assembly Office, Christopher Brown, Lisa Moralee, Gerry Cash, Donnie Mackay, Adam Al-Nuaimi, lowi, ONS (NES), Derek Bird, Claire Nichol; manuf, ONS (AES/ACE), Notice Ref 195R, via NOMIS (DETR), Carole Sutton, Simon Connell; no2nox, AEA Technology, Emma Linehan ; ozone, AEA Technology, Emma Linehan; pnbu, DETR (LUCS), Bob Garland, Barinder Hundle; pnrl, DETR (LA HIP1 returns), Welsh Ofice, Trevor Steeples, Shadan Yusuf, Henry Small, Fiona Leadbitter;pqpr, DETR, Oscar Yau; pqsr, DETR, Oscar Yau; psr, DETR (LA HIP1 returns), Welsh Ofice, Trevor Steeples, Shadan Yusuf, Henry Small, Fiona Leadbitter; punfps, DETR (LA HIP1 returns), Welsh Office, Trevor Steeples, Shadan Yusuf, Henry Small, Fiona Leadbitter; pvac, DETR (LA HIP1 returns), Welsh Ofice, Trevor Steeples, Shadan Yusuf, Henry Small, Fiona Leadbitter; rain, Hadley Centre (UEA), David Viner; regcmp, SBS, Ian Dale; rgdppc, ONS (Regional Accounts), Dev Virdee, David Lacey; rgurb, DETR (LUCS), Bob Garland, Barinder Hundle; rlapsc, DETR (GNI/HNI Indicators), Tina Golton; rlasc, DETR (GNI/HNI Indicators), Tina Golton; rpdi, ONS (Regional Accounts), Dev Virdee, David Lacey; rtoff, Home Office (Crime Statistics), David Covey; scenic, DETR, Stephen Hall; sehcr, DETR (SEH), Jeremy Groves; stmort, ONS, Tim Devis, Justine Fitzpatrick; sunny, Hadley Centre (UEA), David Viner; unempr, Employment Service/LFS via NOMIS, Andrew Craven (NOMIS); vacdrl, DETR (NLUD), Bob Garland; vaclb, DETR (NLUD), Bob Garland; vacul, DETR (NLUD), Bob Garland; voting, Plymouth University, Brian Cheal.

References

Bailey, A. and Rees, P. (2001) Students in the migration model – the estimation process. Technical Appendix 6 in Champion *et al*. (2001) *Development of a Migration Model: Final Report to the Department of the Environment, Transport and the Regions under Contract RADS 5/9/22*, Department of Geography, University of Newcastle, Newcastle upon Tyne.

Champion, T., Bramley, G., Fotheringham, A.S., Macgill, J. and Rees, P. (2002) A migration modelling system to support government decision-making, in Stillwell, J. and Geertman, S. (eds) *Planning Support Systems in Practice*, Springer Verlag, Berlin, pp. 257–78.

Champion T., Coombes M., Fotheringham A.S., Eyre H., Macgill J., Rees P., Stillwell J., Wilson T., Bramley G. and Hollis J. with S. Alvanides, M. Charlton, T. Ford, S. Raybould, A. Bailey, O. Duke-Williams and H. Anderson (2001) *Development of a Migration Model: Final Report to the Department of the Environment, Transport and the Regions under*

Contract RADS 5/9/22. March 2001. Department of Geography, University of Newcastle, Newcastle upon Tyne. (The Phase 1 Final Report. Published in Office of the Deputy Prime Minister (2002)).

Champion, T., Fotheringham, A.S., Rees, P., Boyle, P. and Stillwell, J. (1998) *The Determinants of Migration Flows in England: A Review of Existing Data and Evidence*, Department of Geography, University of Newcastle, for the Department of Environment, Transport and the Regions. Accessible via www.geog.leeds.ac.uk.

Flowerdew, R. (1991) Poisson regression modelling of migration, in Stillwell, J. and Congdon, P. (eds) *Migration Models: Macro and Micro Approaches,* Belhaven Press, London, pp. 92–112.

Flowerdew, R. and Aitkin, M. (1982) A method of fitting the gravity model based on the Poisson distribution, *Journal of Regional Science,* **22**: 191–202.

Fotheringham, A.S. (1986) Modelling hierarchical destination choice, *Environment and Planning A,* **18**: 401–18.

Fotheringham, A.S. (1991) Migration and spatial structure: the development of the competing destinations model, in Stillwell, J. and Congdon, P. (eds) *Migration Models: Macro and Micro Approaches,* Belhaven, London, pp. 57–72.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2000) *Quantitative Geography: Perspectives on Spatial Data Analysis,* Sage, London.

Fotheringham, A.S., Barmby, T., Brunsdon, C., Champion, T., Charlton, M., Kalogirou, S., Tremayne, A., Rees, P., Eyre, H., Macgill, J., Stillwell, J., Bramley, G. and Hollis, J. (2002) *Development of a Migration Model for England and Wales: Analytical and Practical Enhancements. A Report on the Modelling Procedures and Results Connected with the Research Contract Variation to RADS 5/9/22.* March 2002. Department of Geography, University of Newcastle, Newcastle upon Tyne (The Phase 2 Final Report).

Fotheringham, A.S. and O'Kelly, M.E. (1989) *Spatial Interaction Models: Formulations and Applications*, Kluwer Academic, Dordrecht.

Knudsen, D.C. and Fotheringham, A.S. (1986) Matrix comparison, goodness-of-fit, and spatial interaction modeling, *International Regional Science Review,* **10**: 127–47.

Office of the Deputy Prime Minister (2002) *Development of a Migration Model*. The University of Newcastle upon Tyne, The University of Leeds and The Greater London Authority/London Research Centre. Office of the Deputy Prime Minister, London. ISBN 1 85112 583 3.

ONS (2002) Report: internal migration estimates for local and health authorities in England and Wales, 2001, *Population Trends,* **109**: 87–99.

Shaw, C. (2002) 2000-based national population projections for the United Kingdom and its constituent countries, *Population Trends*, **107**: 5–13.

Stillwell, J., Rees, P., Eyre, H. and Macgill, J. (2001) Improving the treatment of international migration. Technical Appendix 7 in Champion *et al.* (2001) *Development of a Migration Model: Final Report to the Department of the Environment, Transport and the Regions under Contract RADS 5/9/22*, Department of Geography, University of Newcastle, Newcastle upon Tyne.

Stillwell, J, Rees, P. and Boden, P. (eds) (1992) *Migration Processes and Patterns. Volume 2, Population Redistribution in the United Kingdom,* Belhaven, London, p. 307.

Walker, D. (2002) Get off yer bike, *The Guardian*, 30 July, p. 13. This draws material from Geographic Mobility, a presentation by John Muellbauer to the Cabinet Office (accessible via www.piu.gov.uk).

Wilson, A.G. (1967) A statistical theory of spatial distribution models, *Transportation Research,* **1**: 253–69.

Wilson, A.G. (1972) *Papers in Urban and Regional Analysis*, Pion, London.

Wilson, T. and Rees, P.H. (1999) Linking 1991 population statistics to the 1998 local government geography of Great Britain, *Population Trends*, **97**: 37–45.

15

Modelling Regional Economic Growth by Means of Carrying Capacity

Leo van Wissen

Abstract

In this chapter a new model of spatial economic growth is proposed. It is based on the ecological concept of *carrying capacity*, defined for a geographical system of regions. The use of ecological concepts refers to the dynamic evolution of populations over time. Here, we view the development of economic markets as an ecological process of growth towards the carrying capacity level. The carrying capacity is given an economic interpretation, and is equal to the market demand for a product. In addition, a spatial dimension is introduced that gives, for each location in a spatial system, the disequilibrium between carrying capacity (demand) and supply. The chapter shows a number of characteristics of the model and demonstrates how the model may be used empirically. Some empirical results are given from an application of the model in the field of industrial demography.

15.1 Introduction

Regional economic growth can be modelled in many ways. The question why certain regions grow faster than others has attracted many geographers and economists in the past. Amstrong and Taylor (2000) give a concise overview of a number of these models and methods, such as the regional input–output model, the neoclassical regional growth model, the regional multiplier model, or the economic base method. Most existing models share one common characteristic: they are essentially

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

macro-economic models, applied at the meso-scale of the region. As a result, they share many of the characteristics of macro-economic models: in particular they are either implicitly or explicitly based on the notion of equilibrium. This equilibrium view of the world has been challenged in recent years from different viewpoints. After all, it seems more logical that regional models of economic growth should focus on change instead of equibrium and structure. Change in these models is a necessary but short episode between two states of equilibrium, or can only be deduced from the difference between two equilibrium states.

One of the most important criticisms against this essentially neo-classical view of the world was raised by Nelson and Winter (1983), who advocated a new evolutionary paradigm in economics. Their models were based on an evolutionary concept of economic change. It emphasises change instead of equilibrium, process, instead of structure. Important economic processes are selection, adaptation, inertia and competition. Within sociology, a somewhat similar field has developed, which focuses on processes of change in populations of organisations. This field is called organisational ecology, or simply demography of firms or industrial demography (Hannan and Freeman, 1989; Carroll and Hannan, 2000), in which populations of firms evolve through selective processes of entry and exit.

Demography of firms is also becoming more widespread in economic geography (van Dijk and Pellenbarg, 2002; van Wissen, 2002). In principle it may offer an alternative view on regional economic growth, based on regional variations in birth, death, migration and firm growth processes. In the past it has been used largely as a descriptive tool for regional employment change (see e.g., Gudgin, 1978), but more recently analytical models of regional economic growth have been developed as well (van Wissen, 2000).

In analytical models of firm population evolution in a regional economic setting, there are a number of problems that have to be solved. These problems derive from (1) the translation of demographic processes to the level of firm populations; (2) the introduction of economic mechanisms of demand and supply in demographic processes; and (3) introducing the spatial dimension in firm demographic models (van Wissen, 2002). The first problem involves issues such as definitions, model structure, the modelling of firm birth, death and other demographic components, and the impact of age, period and cohort, or other primary variables on demographic events. Many of these issues have been treated in van Wissen (2000). This chapter deals with a possible solution to the second and third problems.

The second problem deals with the notion of the market. Even if equilibrium is not taken as the starting point, there is still the impact of the 'invisible hand of the market' on firm founding and failure, and growth. One of the most important differences between human and firm demography is the existence of demand and supply interactions or, in ecological terms, a feedback mechanism between the number of firms in a population and population growth. In short, the larger the population of firms and the larger the supply of products in the market, the fiercer the competition and the lower the population growth. Population ecologists call this feedback between population size and growth the density dependence model (Carroll and Hannan, 2000). The ecological notion of *carrying capacity* is a fruitful starting point for introducing an invisible market hand into these models. *Carrying capacity* is an

ecological concept. In an ecological framework a population (generally of animals) will grow if its current size is smaller than the carrying capacity for that population. If the actual size is larger than the carrying capacity, the population will decrease in size. We will build upon this ecological concept and extend it to apply to a multi-sector economy.

The third problem involves introducing the spatial dimension into the model. In most regional economic models, as well as in most population ecology models, the notion of space is not very well developed. In the simplest case, regions are small-scale variants of nations, with only imports and exports as signs of life outside of the region. In other cases, such as in the multi-regional input–output table, a limited number of regions interact, but the influence of distance remains implicit in the model. In recent years, Krugman and colleagues developed a series of economic models that take the spatial dimension seriously (Krugman, 1995; for an introduction see Neary, 2001). Although these models have many attractive theoretical properties, it is difficult to apply them other than in a highly stylised world, where only two regions exist, or which has a linear or circular form. The ecological model to be developed also contains an explicit spatial dimension. In population ecology, populations interact with each other. These interactions occur between different populations as well as between different regions. The spatial dimension essentially implies that nearby populations have higher propensities to interact than others.

In short, this chapter proposes a model of regional economic change, based on the ecological concept of *carrying capacity*. It also contains elements of regional input–output models, thus establishing a link with more traditional regional economic growth models. Thirdly, the model involves inter-sectoral as well as spatial interactions. Ironically this resulting model has a much larger applicability than the field of firm demography. Although its structure is based on an ecological concept, the resulting model can be used in a population ecology or evolutionary framework, but it may also be used as a regional economic model of change without reference to these fields.

This chapter is organised as follows. First, in Section 15.2, we introduce the basic concept of *carrying capacity*. In the following sections, we extend this concept to a multi-sector economy (Section 15.3) and include a spatial dimension (Section 15.4). In Section 15.5 a general matrix notation is given. A number of model characteristics are investigated in Section 15.6 using a simulation framework. The concept of spatial carrying capacity is not only relevant from a theoretical point of view. On the contrary, it can be applied in an empirical setting as well. Section 15.7 contains empirical evidence of the usefulness of this approach. Section 15.8 concludes.

15.2 The Concept of Carrying Capacity

Carrying capacity denotes the maximum number of users that can be sustained by a given set of resources in a specified territorial unit. In organisational ecology it is defined as the maximum number of organisations of a specified type that the market, or society, can sustain (Hannan and Freeman, 1989). The term can be given a clear

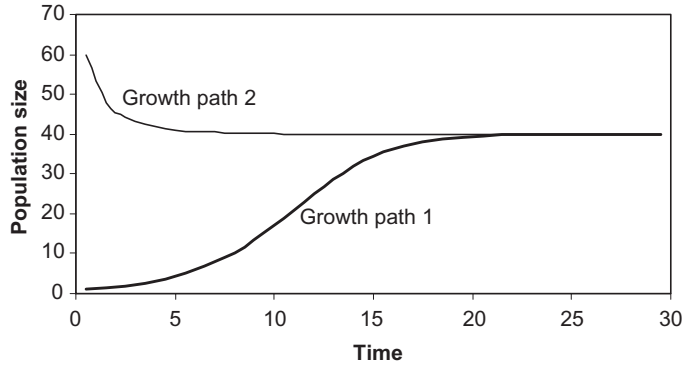


Figure 15.1 *Population growth following a logistic growth curve towards the level of the carrying capacity, starting from two different initial values*

mathematical interpretation, most notably in the example of the logistic growth curve describing the temporal growth path of the number of members of a population in a closed environment; for instance, the number of rabbits on an island. If natural growth of a population in discrete time between time t and $(t + 1)$: $X(t + 1) - X(t)$, denoted as $\dot{X}(t)$ is proportional to the size of that population then we have geometric growth:

$$\dot{X} = g(X(t)) \tag{15.1}$$

where g is the growth rate of the population.[1] If growth is positive this leads to an ever-increasing size of the population. It is clear that this growth curve is not realistic for most, if not all, species. A population cannot grow with a fixed growth rate g indefinitely. Given its dependence upon external resources, such as food or living space, a population is limited to a certain maximum, the level of which is conditional on the level of the resources and the technological level of the population to exploit these resources. Let the maximum size of the population be K, then the geometric growth model may be modified to the logistic growth model by making the growth rate not only dependent on the intrinsic growth parameter g but also dependent on the difference between the actual size of the population, X, and the maximum size, or carrying capacity, K:

$$\dot{X}(t) = gv(K - X(t))X(t) \tag{15.2}$$

The additional parameter v is the adjustment speed of the process. Figure 15.1 depicts two possible growth logistic paths resulting from this equation for fixed g. If $X(t)$ is very small compared with K then $K - X(t)$ is large and growth is high. With increasing size of the population over time, the gap between the maximum and actual size

[1] Formally the growth rate should be multiplied by the average size of the population between t and $t + 1$: $\frac{1}{2}(X(t) + X(t + 1))$, but for simplicity we use the initial population.

of the population narrows and consequently the growth rate $g(K - X(t))$ becomes smaller. If $X(t)$ is equal to K then growth is equal to zero. If, for some reason, the actual size of the population is larger than the carrying capacity, $K - X(t)$ is negative and the growth rate is negative as well.

Hannan and Freeman (1989) applied this logistic growth model (15.2) to longitudinal data on growth in the number of units of organisations. In their approach they included a number of explanatory variables in their model but, at the same time, they circumvented the problem of modelling the carrying capacity K explicitly. Instead, they estimated a second-degree polynomial to longitudinal growth data in order to find an empirical estimate of the maximum level of the population. This can easily be shown to be possible if we rewrite equation (15.2) as follows:

$$\dot{X}(t) = gvKX(t) - gvX(t)^2 = aX(t) - bX(t)^2 \tag{15.3}$$

The value of K may be found as $K = a/b$. Its precision depends on the precision of the coefficients a and b and the estimate is therefore in general not very reliable. But even if the value of the carrying capacity could be estimated precisely, there are a number of objections to model equation (15.3). First, the level of the carrying capacity is not static, but conditional on the size of resources and the level of technology. These change over time and therefore the size of the carrying capacity cannot be treated as static either. In other words: K should be $K(t)$. Secondly, there is no explicit content in this interpretation of the carrying capacity. It is simply a coefficient to be estimated from observed growth curves. Instead, an explicit model of carrying capacity is necessary that captures the essential underlying factors. More specifically, we require a model of carrying capacity based on economic notions.

15.3 An Economic Interpretetion of Carrying Capacity

The starting point is the population growth function equation (15.2). Firms operate in a multi-sector economy where they interact with each other. Interactions occur between industries in the form of intermediate supply and demand relationships, other interactions involving information, transfer of knowledge, and so on. Interactions also occur between industries and consumers. The latter type of relationship is self-evident but the inter-industry linkages are important as well. Each firm, as a member of a certain industry, has a potential network of linkages with other industries. Some linkages are beneficial for the firm: for instance, when these linkages are deliveries from one firm to the other. Some linkages are of the competitive type: the same firms compete in one market. The existence of other industries in the neighbourhood with whom it may interact is therefore a significant growth stimulating or inhibiting factor for this firm. At the same time, the growth of this particular firm is in itself again a potential growth-enhancing factor for other industries with inter-industry linkages with this firm, or a growth-limiting factor for similar firms who are in competition with this particular firm in the same market. In short, the size and composition of the population of firms in the neighbourhood is a determining factor for the growth potential for an individual firm. The effect of the population of firms

in the neighbourhood upon the individual firm is realised through interaction linkages between firms. The concept of carrying capacity, as defined here, captures this growth potential of a local population of firms due to the size and composition of firm populations in the neighbourhood.

The concept of carrying capacity is here defined formally as the level of demand exerted by intermediate and final demand categories in the market. This notion is close to the economic interpretation of the concept as given by Carree and Thurik (1999). They propose a model of entry and exit of firms in markets as adjustments to a disequilibrium situation. The equilibrium situation is called the carrying capacity (Carree and Thurik, 1999, pp. 986–7). It is the benchmark number of firms which drives the direction of the change in the actual number of firms. Carree and Thurik relate the carrying capacity in a local market to the level of consumer demand. Here we use a specification, which is close to an input–output model, involving both intermediate and final demand. First, we present the situation involving only one region.

The actual level of supply will adjust to market demand. If supply is less than demand, there is an opportunity for increased production: birth rates go up, death rates go down, and incumbent firms will grow. If supply exceeds market demand, then the level of production will decrease: firm birth levels will drop, death rates will increase and incumbent firms will diminish in size. The relation between change and the level of production, $\dot{X}(t)$ and $K(t)$ is then given by equation (15.2) above. However, in a multi-sector economy the carrying capacity is specific for each sector s ($s = 1, \ldots, S$). Thus, the level of $K_s(t)$ is determined by both intermediate and final demand in the region:

$$K_s(t) = \sum_{r=1}^{S} a_{sr} X_r(t) + c_s B(t) \tag{15.4}$$

where a_{sr} is the technical coefficient in the input–output table of sector r on sector s, c_s is the per capita final consumption of goods produced by sector s, and $B(t)$ is the size of the population at time t. Equation (15.4) is an elementary input–output equation without investments and exports, that specifies the interrelations among producers and between producers and consumers in the national economy. Note that here final consumption is explicitly related to the size of the population B using the consumption coefficients c. Equation (15.4) does not include exports and investments. They might be added as additional terms. Note that imports are implicit in the input–output specification. Sectors that rely partially on imports will have technical coefficients that do not add up to 1: the demand for this sector will only partially be met by internal supply.

The parameters of the demand equations of the input–output model may be derived from the national input–output table and the national accounts. So, equation (15.4) gives the level of the carrying capacity as a function of intermediate and final demand. Since the input–output model is an equilibrium model, $K(t)$ and $X(t)$ are equal to each other, and small changes in $B(t)$ will lead to a new equilibrium value for the $X_s(t)$. Here, we allow both $B(t)$ and the $X_r(t)$ to change, and specify a

logistic adjustment function of $X_s(t)$ to its equilibrium value (carrying capacity) $K_s(t)$:

$$\dot{X}_s(t) = g_s v_s (K_s(t) - X_s(t)) X_s(t) \tag{15.5}$$

Note that the growth rate g and the speed of adjustment v are sector-specific.

15.4 Introducing Spatial Interactions into Carrying Capacity

There is another source of disequilibrium in the market. In a multi-regional economy, local variations may exist in supply and demand. The input–output model is generally used to determine levels of production given an increase in one of the final demand categories. For economic demography with a spatial dimension, we need an instrument to determine demand–supply ratios at the local level. Of course we cannot assume that each region or local area is an island without interactions with other zones. Producers will deliver their goods to retailers and other producers in other regions; consumers will make shopping trips to retailers in other zones as well; in other words, there exist complicated spatial patterns of transport flows of goods between producers and other producers (inter-industry linkages), and between producers and consumers. In order to deal with these spatial demand–supply interactions, we introduce the notion of the *spatial demand field*. A spatial demand field defined at location j ($j = 1, \ldots, I$) is a probability density function that gives, for every location i ($i = 1, \ldots, I$) in the neighbourhood of j, the probability that the good m, that is required at location j as either an intermediate input by a producer or as final input by a consumer, is produced at location i. This conditional probability $p_m(i|j)$ is not uniform over all production locations i, but decreases with distance between i and j. In other words, the form of the demand field is in general not flat but a dome with its top in the centre located at the demand origin j and decreasing with increasing distance from the centre. In addition, since the good is produced somewhere, it holds that $\Sigma_i p_m(i|j) = 1$. A function that is often used for modelling probability functions is the logit function. Used as a spatial interaction function with distance as its only argument, it has the following form:

$$P_m(i|j) = \frac{\exp -\beta_m d_{ij}}{\sum_k \exp -\beta_m d_{kj}} \tag{15.6}$$

where d_{ij} is the distance between locations i and j, and β_m is a distance deterrence parameter which is specific to the good m being demanded. If instead of the linear distance the log of the distance is used, the exponent in both nominator and denominator is replaced by $d_{ij}^{-\beta_m}$. A high value of β_m indicates a strong distance decay effect and a steep slope of the spatial demand field, implying that with increasing distance it soon becomes highly unlikely that the required good m is produced at i. A small value of β indicates a small distance decay effect and thus a relatively flat spatial demand field: even at larger distances between i and j it is still likely that goods

demanded by producers and consumers at j are produced in i. A value of $\beta = 0$ implies a perfectly flat spatial demand field and the absence of any distance effect: at every location i it is equally likely that the good m demanded in j is produced in i. It should be clear that β is specific for every economic good m, since the form of the demand field varies with the character of the good. The spatial demand field of grocery products is very small (high β), whereas the demand for specialised services is often exerted at the national or even international market (β is very small or zero). Note that this system, just like the input–output model, is demand-driven. Firms and households require inputs or consumer goods m, and look around for suppliers to meet their demand. In which locations they look around for producers in alternative locations i is specified by the demand field $p_m(i|j)$. For practical reasons we do not work with individual goods m, but with goods produced by the aggregate sector s.[2]

How does this relate to the model of carrying capacity introduced here? First we note that the concept of carrying capacity refers to production capacity. In the spatial framework just introduced we make a distinction between locations i, where production takes place, and locations j, where goods are demanded or consumed. Of course in the recursive structure of the input–output framework producers are at the same time demanding intermediate goods as well, so each location may belong to both categories. In order to distinguish between production and consumption characteristics of locations, we introduce the total level of demand exerted by consumers and firms in location j:

$$D_s(j,t) = \sum_{r}^{S} a_{sr} X_r(j,t) + c_s B(j,t) \tag{15.7}$$

Note that this is quite similar in form to equation (15.4). There is one important difference, however. In the spatial framework of equation (15.7) we make a distinction between the level of demand and the level of production in each region, because, unlike the closed system of equation (15.4), they are not equal. The link between demand $D_s(j,t)$ and production capacity $K_s(i,t)$ is a spatial interaction relation, as given by the spatial demand field (15.6). The carrying capacity $K_s(t)$ introduced in equation (15.4) is the aggregate market demand, summed over all intermediate and final consumers. Thus, in a spatial setting, if we sum over all locations j, where these intermediate and final demanders are located, we arrive at the total market demand for goods produced by sector s in location i. The required production capacity of sector s at location i for intermediate and final demand in j at time t is given by:

$$K_s(i,j,t) = p_s(i|j) D_s(j,t) \tag{15.8}$$

and the total required production capacity, or carrying capacity at location i is the sum over all locations j:

$$K_s(i,t) = \sum_{j}^{I} p_s(i|j) D_s(j,t) \tag{15.9}$$

[2] Formally: s is the index of the class of goods Ω_s to which the single good m belongs: $m \in \Omega_s$.

By comparing the carrying capacity $K_s(i,t)$ with the actual level of production $X_s(i,t)$ in region i at time t the likely direction of local economic development of sector s located in i may be derived. The direction and size of growth can be modelled in terms of a logistic growth path as follows:

$$\dot{X}_s(i,t) = g_s v_s (K_s(i,t) - X_s(i,t)) X_s(i,t) \tag{15.10}$$

This equation states that growth in production in sector s in location i in the unit time interval t, $\dot{X}_s(i,t)$, is determined by the intrinsic growth rate g_s of sector s, multiplied by the overall speed of adjustment of the process, v_s, and modified by the effect of the gap between carrying capacity $K_s(i,t)$ and actual level of production $X_s(i,t)$.

The inter-sectoral and spatial interaction linkages developed here reflect a specific form of spatial agglomeration effect. A certain mix of production in a region may increase the level of the carrying capacity of the region and therefore the growth rate of the regional economy. Other mixes will only increase the actual level of production, while keeping the carrying capacity level constant, which results in lower growth rates in the region. The model therefore produces local variations in growth rates.

In summary, the advantages of this definition of spatial carrying capacity for economic demography are the following:

- the spatial input–output representation is consistent with the national input–output table;
- if the national input–output table represents an equilibrium situation, then the spatial input–output model gives for each location i the local disequilibrium in demand and supply for each sector. Since the system is consistent with the national input–output table, excess demand in some locations will be mediated by excess supply in other locations;
- inter-sectoral linkages as well as spatial interaction linkages between sectors are included in the model;
- the model presents a link between spatial economic disequilibrium on the one hand and macro-economic indicators of the national economy on the other;
- the information concerning $K(i,t)$ and $X(i,t)$ is of interest, since it describes the spatial distribution of demand and supply, as well as local market disequilibrium for each economic sector;
- in addition K and X are instrumental in driving the local spatial economic development for each sector, as hypothesised in equation (15.10); and
- the combination of inter-sectoral and spatial interaction linkages reflects the mechanisms of spatial agglomeration effects. As a result of this, local variations in economic growth rates will occur as an endogenous process.

We will test some of these ideas in Sections 15.6 and 15.7 below. First, however, the dynamic concept introduced in equation (15.10) will be discussed a little further in the next section, since it can be reformulated in other terms, giving it a more familiar structure.

15.5 The Model as a Dynamic Difference Equation for its Subpopulations

Equation (15.10) is a specific example of a system of logistic growth equations for interacting populations (Wilson, 1981, pp. 54–5). The system may be formulated in matrix terms as a set of difference equations. For convenience we make no explicit distinction between intermediate and final demand. In matrix terms, the system has the following form:

$$\dot{\mathbf{X}}(t) = \mathbf{G}[\mathbf{K}(t) - \mathbf{X}(t)]\mathbf{X}(t) \tag{15.11}$$

where $\dot{\mathbf{X}}(t)$ is an $S \times I$ growth matrix with elements $\dot{\mathbf{X}}_s(i,t)$ and $\mathbf{G}$ is an $S \times S$ diagonal matrix of sector-specific relative growth rates and adjustment speeds $g'(s,s) = g_s v_s$. The matrix of carrying capacity $\mathbf{K}(t)$ with dimensions $S \times I$ is calculated as:

$$\mathbf{K}(t) = [\mathbf{AX}(t) + \mathbf{cb}(t)]\mathbf{M} \tag{15.12}$$

Here, $\mathbf{A}$ is an $S \times S$ matrix of technical coefficients a_{sr} from the input–output table, $\mathbf{c}$ is an $S \times 1$ vector of consumption coefficients, and $\mathbf{b}(t)$ is a $1 \times I$ population distribution vector. The elements of the $I \times I$ matrix of spatial discounting factors $\mathbf{M}$ are calculated as:

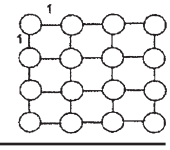
$$m(i, j) = p(j|i) \tag{15.13}$$

where $p(j|i)$ is the probability that a good demanded in j is produced in i; the functional form is given in equation (15.6). The interpretation of equations (15.11)–(15.13) is similar to (15.6)–(15.10). Note, however, that in this matrix formulation the distance deterrence function is generic for all economic sectors. In more realistic formulations this is not tenable. Equation (15.11) is the logistic growth equation for S $\times$ I submarkets. Equation (15.12) is the model for carrying capacity. $\mathbf{AX}(t) + \mathbf{cb}(t)$ is the demand exerted by firms and consumers, located in j, for goods produced by sector s. The matrix $\mathbf{M}$ defines the spatial demand field around j. The row elements of $\mathbf{M}$ sum to 1.

Difference equation systems of the form of (15.11) have been studied extensively in the sphere of organisational ecology (see, for example, Tuma and Hannan, 1984). Although each equation is linear, the resulting behaviour of the system is intrinsically non-linear and not always stable. A complete treatment of all properties of this system will not be given here. General accounts of coupled difference equations involving spatial interactions are provided in Wilson (1981). A number of features of the model should be discussed here, since they are related to the nature of sectoral and spatial interactions and the consequences for overall system stability. From equation (15.11) it is clear that equilibrium occurs whenever the carrying capacity is equal to supply: $[\mathbf{AX}(t) + \mathbf{cb}(t)]\mathbf{M} = \mathbf{X}(t)$. Unfortunately analytical solutions cannot be found for this system of equations, except for trivial cases. In the next section some results of a series of computer simulations with varying characteristics of intersectoral and spatial interactions will be given and discussed. The simulation results are illustrative for the structural characteristics of the model.

Table 15.1 *Input parameters of simulation*

Parameter		Value
ν	=	0.02
G	=	diag [0.1, 0.2, 0.4]
A	=	$\begin{bmatrix} 0.30 & 0.07 & 0.12 \\ 0.45 & 0.40 & 0.15 \\ 0.15 & 0.13 & 0.31 \end{bmatrix}$
c	=	[0.20, 0.40, 0.60]
β	=	–1.00
M	=	See equation (15.6) for specification. Distances between zones based on 4×4 square pattern, spaced 1 unit apart horizontally or vertically, and $\sqrt{2}$ units diagonally



15.6 Computer Simulations of the Model

For the computer simulations we used a three-sector economy plus population for 16 regions, configured in a four by four square regional pattern. Interregional distances were calculated using Euclidean distance. The input parameters for the system are given in Table 15.1. We will focus on the following key parameters: ν is the speed of the system, the sectoral interactions a_{sr}, and the spatial interactions, determined from the logit function (15.6), where the β_s coefficient determines the demand field.

Figure 15.2 gives the initial distribution of the population and each of the three sectors over the 16 zones. Figure 15.3 presents the equilibrium configuration after 300 iterations. If the evolution of the system is smooth, the equilibrium distribution does not reflect the initial distribution of the sectors. Using the speed of adjustment as given in Table 15.1, the evolution of the system is indeed smooth, and the initial highly skewed spatial distribution of the sectors is replaced by a more balanced distribution that reflects the combined effects of different parameters. A number of conclusions may be drawn from the equilibrium distribution. First, the distance effect favours economic development in more centrally located zones over peripheral zones. Second, the equilibrium configuration of each sector reflects, to a certain extent, the population distribution or final demand. The population impact is stronger for sector 3 than for the other two, since this sector is more dependent on final demand. The correlation between the population and sector 3 distribution is 0.78, against values around 0.69 for the other two sectors. The final distributions of each of the sectors are very similar, although the level is different, with correlations around 0.98. Formally, the spatial distribution of each of the sectors is a weighed average of the distribution of the other sectors, with weight matrix **A**.

As can be observed from Figures 15.2 and 15.3, the chosen input parameters have resulted in a model that has evolved to a steady state. This steady state is equal to the level of the carrying capacity for each (i,s) combination. However, if the speed of adjustment is increased (or, equivalently, the intrinsic growth rate of sectors), the

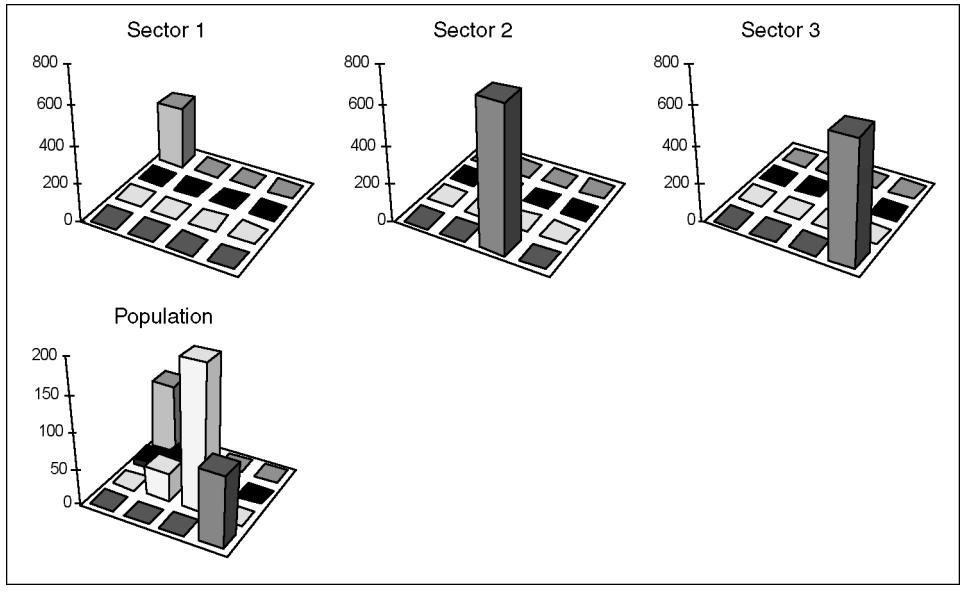


Figure 15.2 *Initial distribution of sectors and population*

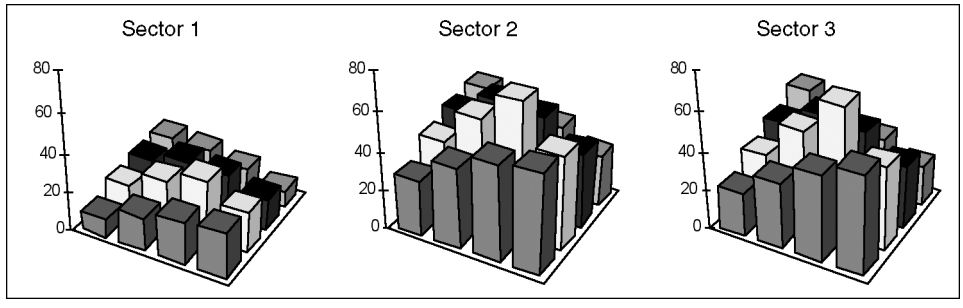


Figure 15.3 *Equilibrium configurations of sectors*

resulting growth of each sector is no longer guaranteed. Figure 15.4 shows an interesting example for one region. In the slow mode with a system speed of 0.02 the region develops from an initial value of 0.5 to the equilibrium value of 60. Increasing the speed to 0.10 results initially in smooth but fast growth up to size 50. At this point, growth decreases substantially and small oscillations set in that become larger after reaching the slow mode equilibrium of 60. After this point, growth oscillates around the value of 60. With speeds higher than 0.10, oscillations set in from the start and are much larger. Oscillations may also occur in a regular pattern; for instance with a certain periodicity. These oscillations are the result of an overreaction of the subsector to a large carrying capacity. This overreaction leads to a growth

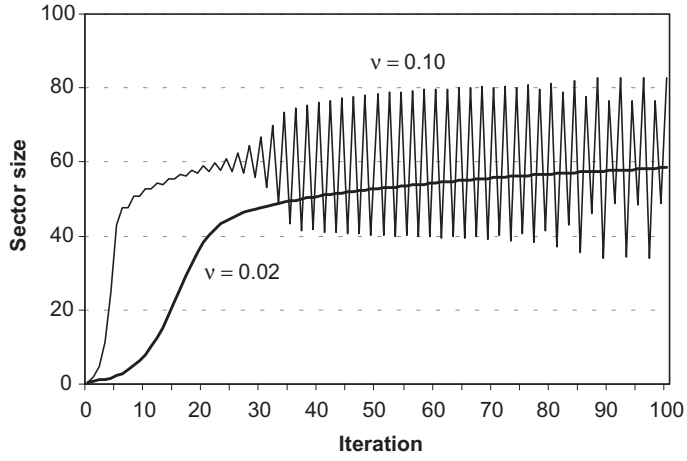


Figure 15.4 *Development towards a steady state and oscillatory behaviour using different speeds of adjustment*

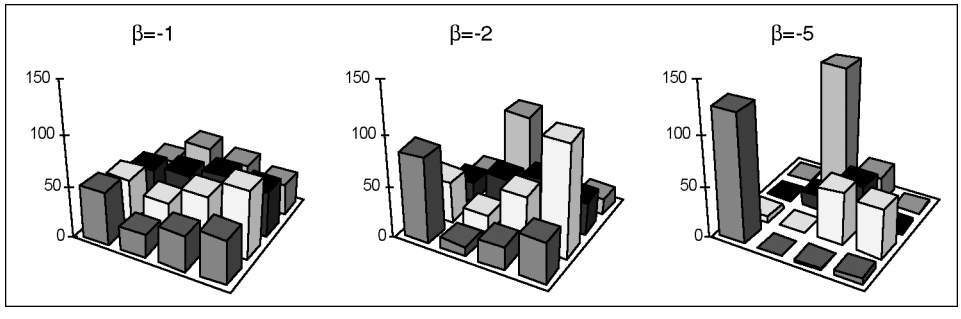


Figure 15.5 *The effect of distance friction on the equilibrium configuration of sector 3*

of the subsector above the carrying capacity. In the next period this has to be compensated by a large negative growth, and so on. Thus, the growth speed of a sector may have serious impacts upon the stability of the regional economic system as a whole.

Increasing the effect of distance will result in the emergence of a few locations where most of the activities are concentrated (Figure 15.5). Interestingly, the regions where concentration takes place are not the regions where initially activities were concentrated. Moreover, with increasing distance friction, the system no longer converges towards the steady state, but starts to oscillate. The configuration shown for the case of $\beta = -5$ in Figure 15.5 is not an equilibrium state. Conversely, interactions without distance friction will result in a completely flat surface and smooth behaviour towards the steady state.

From Table 15.1, which shows the technical coefficients used in the simulations and per capita final consumption, it is clear that sector 3 is most directly related to

final demand, whereas sector 1 is the most independent sector of the three. Sector 2 is dependent on sector 1 for deliveries but also dependent on population. The effect of a change in technology will change the spatial distribution. For instance, a change of the technical coefficient $a(2,1)$ from 0.45 to 0.10, implying that for the production of one unit of good 1 much less input of good 2 is needed than before, has the result that the equilibrium value of sector 1 increases on average by 30%, whereas the equilibrium value of sector 2 decreases on average by 20 per cent. The effect on sector 3 is modest: a slight increase in equilibrium value of 7%.

These examples show that a spatial equilibrium of a multi-sectoral economy depends crucially upon the key interaction parameters of the economic system: speed of adjustment and intrinsic growth speed, sectoral interactions, and spatial interactions. The speed of adjustment of economic activities and the spatial deterrence effect have potentially highly disturbing effects, including non-equilibrium situations. In the next section we will relate some of these findings to empirical observations of individual firm dynamics.

15.7 Empirical Implementation of the Model

The concept of carrying capacity as developed in the previous sections may be implemented in empirical modelling. The required amount of information is only limited. The parameters necessary for calculation of the spatial carrying capacity are:

- the technical coefficients of the input–output table a_{sr} collected in the matrix **A**;
- per capita consumption coefficients c_s, collected in the row vector **c**; and
- sector-specific distance parameters β_s.

In order to calculate the level of the carrying capacity $\mathbf{K}(t)$ we also need:

- the matrix $\mathbf{X}(t)$ of production of activity s at location i at time t;
- the population distribution over regions at time t, $\mathbf{b}(t)$; and
- inter-zonal distances $d(i,j)$.

Technical coefficients and per capita consumption coefficients are readily available in most countries, even at the regional scale. The distance function parameter β_s may be estimated using some form of optimisation. In an empirical application in the Netherlands, which will briefly be discussed below, a simple line search algorithm was used for the distance parameter for each sector, using the Pearson correlation coefficient between sectoral growth $\dot{X}_s$ and $V_s = K_s - X_s$ as the criterion variable. The results are presented in Table 15.2.

All values of β are positive, which implies that interaction between sectors decreases with increasing distance between them (the function uses $-\beta$ as the coefficient value). The smallest coefficients are found for the recreational sector, transport and banking. The value of 0 would imply a completely flat demand field. This is plausible for transport, but less so for banking and certainly not correct for the recreational sector. The largest value, and a very small demand field, is found for

Table 15.2 *Estimated distance deterrence parameters* β_s *and resulting correlations between sectoral growth* $\dot{X}_s$ *and* $V_s = K_s - X_s$

Economic sector		Coefficient	Correlation *r*
1	Agriculture	0.25	0.26
2	Energy, mining	0.30	0.09
3	Industry	0.70	0.45
4	Construction	2.00	0.14
5	Wholesale	2.00	0.75
6	Retail	5.00	0.18
7	Recreation, hotels, restaurants	0.00	0.02
8	Transport communications	0.00	0.81
9	Banking, insurance, business services	0.00	0.32
10	Other services	5.00	–0.07
11	Public sector	0.15	0.07

retailing and other (domestic and health) services, which is again quite plausible. Industry has an intermediate value of 0.75. The resulting correlations between sectoral growth and the difference between carrying capacity and the actual level of production are in general positive and significant, but not very high. These results are in line with the results obtained by Feser and Sweeney (2000). Although using a different approach, they found empirical evidence that economic clustering, as indicated by the technical coefficients of the input–output table, is associated with spatial clustering as well.

The model of carrying capacity was applied at the regional level in the Netherlands (see van Wissen, 2000 for details). Equation (15.11) is a deterministic difference equation defined at the meso level of a regional multi-sector economy. The spatial level in the empirical application was the municipal level ($I = 685$), and 11 economic sectors were distinguished. However, the model was given a micro interpretation. The growth rate of sector s at location i at time t may be viewed as an underlying latent variable that determines the growth potential of individual firms of sector s in location i. Firms in this application are equal to business establishments. The data were taken from a register of business establishments in the Netherlands, called LISA. LISA is a longitudinal administrative register that includes all addresses in the Netherlands where paid work is being carried out. Although the agricultural sector and self-employment are not completely represented, and the province of North Holland is not covered, this register is used as the most important source for spatial information about business establishments in the Netherlands (see Kemper *et al.*, 1996). Here we used the data from the years 1990 and 1991. Individual firms are the agents of economic growth through new firm formation, firm closures, relocations and firm growth. In this application economic growth was operationalised in terms of birth and death probabilities of firms, as well as firm growth for incumbent firms. Within this demographic framework, birth and death probability functions, as well as firm growth functions, were estimated in which the variable $V_s(i,t) = K_s(i,t) - X_s(i,t)$ was used as an explanatory variable. The coefficient should be positive for firm birth and firm growth and negative for firm death. The results of this application, as

presented by van Wissen (2000), were moderately favourable for the model. For firm birth and growth the expectations were indeed confirmed, with positive and significant coefficients. However, for firm death the coefficient was negative, as it should be, but insignificant.

15.8 Conclusions

In this chapter we introduced the notion of spatial carrying capacity, based on a model of sectoral and spatial interactions between submarkets of the economy. The resulting model can be viewed as a dynamic spatial model of growth in a multi-sector economy. The characteristics of this model were illustrated in a simulation context using a deterministic difference equation. The speed of adjustment of the process, the spatial deterrence function and the inter-sectoral interactions turn out to be crucial variables in shaping the spatial equilibrium of the economic system. Moreover, the existence of equilibrium is not guaranteed. The simulations showed that a very dynamic economy, with high speed of adjustment, may result in oscillatory behaviour of the level of economic production. A very strong distance effect has the same oscillating effect. Moreover, with high distance friction, economic activities tend to concentrate in just a few regions, leaving other regions largely empty. The simulations also showed that a change in technology, resulting in a reduction of the technical coefficients and thus in reduced inter-sectoral interactions, changed the equilibrium distribution without entering instability.

The concept of carrying capacity can be applied empirically. The required amount of information is limited, and usually readily available. Sector-specific distance parameters have to be estimated. In an empirical application in the Netherlands the estimated distance parameters were to a large extent, although not completely, according to theoretical expectations. Moreover, there are first signs that the concept of carrying capacity as developed here may be useful in explaining and predicting regional sectoral growth.

In conclusion, the concept of carrying capacity as defined in this study deserves more attention, both theoretically and empirically. From a theoretical point of view the model has many interesting properties and through the relation with the input–output table the link with macro-economic variables is established easily. The model is one way of making the concept of agglomeration economies operational, by explicitly taking into account inter-industry linkages as well as their spatial representation. The model may be used at the meso level of sectoral and spatial submarkets. The usefulness of the approach at this level is not studied in depth yet, but from the theoretical, simulation and empirical results reported in this chapter, further explorations in that direction look promising.

REFERENCES

Amstrong, H. and Taylor, J. (2000) *Regional Economics and Policy*, Third Edition. Blackwell Publishers, Oxford.

Carree, M.A. and Thurik, A.R. (1999) The carrying capacity and entry and exit flows in retailing, *International Journal of Industrial Organization*, **17**: 985–1007.
Carroll, G.R. and Hannan, M.T. (2000) *The Demography of Corporations and Enterprises*, Princeton University Press, Princeton, NJ.
Feser, E.J. and Sweeney, S.H. (2000) A test for the coincident economic and spatial clustering of business enterprises, *Journal of Geographical Systems*, **2**(4): 349–74.
Gudgin, G. (1978) *Industrial Location Processes and Regional Employment Growth*, Saxon House, Westmead.
Hannan, M.T. and Freeman, J. (1989) *Organizational Ecology*, Harvard University Press, Cambridge, MA.
Kemper, J., Kloek, W. and van Wissen, L . (1996) Databronnen voor economische demografie, *Planning Methodiek en Toepassing*, **48**: 3–11.
Krugman, P. (1995) *Development, Geography and Economic Theory*, MIT Press, Cambridge, MA.
Neary, J.P. (2001) Of hype and hyperbolas. Introducing the new economic geography, *Journal of Economic Literature*, **39**(3): 36–61.
Nelson, R. and Winter, S. (1983) *An Evolutionary Theory of Economic Change*, Cambridge University Press, Cambridge, MA.
Tuma, N.B. and Hannan, M.T. (1984) *Social Dynamics. Models and Methods*, Academic Press, London.
van Dijk, J. and Pellenbarg, P. (eds) (2002) *Demography of Firms Spatial Dynamics of Firm Behaviour*, Nederlandse Geografische Studies no. 262, KNAG Utrecht/Groningen.
van Wissen, L.J.G. (2000) A micro-simulation model of firms: applications of concepts of the demography of the firm, *Papers in Regional Science*, **79**: 111–34.
van Wissen, L.J.G. (2002) Demography of the firm: a useful metaphor?, *European Journal of Population*, **3**: 263–79.
Wilson, A.G. (1981) *Catastrophe Theory and Bifurcation, Applications to Urban and Regional Systems*, Croom Helm, London.

16

Planning a Network of Sites for the Delivery of a New Public Service in England and Wales

Mike Coombes and Simon Raybould

Abstract

The Family Law Act (1996) aimed to introduce across England and Wales a new service of Information Meetings (IMs) for people considering divorce. Attendance would be compulsory, which meant that the service would have to be provided in places that left few people objecting to attending one. The research reported here identified a suitable pattern of sites at which IMs could be held. Several methods of analysis were needed: a regression model estimates the spatial distribution of potential IM attenders; survey data analysis measures people's willingness to travel to attend IMs; multi-criteria analysis establishes the list of sites which meet the necessary criteria; and location-allocation modelling identifies which of these sites would minimise the travelling which would be required of attenders. The chapter summarises this mapping and modelling research programme, and highlights the key elements of sensitivity in the results. One important part of the process is to identify which are the issues to be resolved by the researchers and which are the decisions that remain with the client government department.

16.1 Introduction

When the Family Law Act was passed in 1996 some of its most innovative aspects were put 'on hold' until a programme of research studies had provided the basis for deciding how to implement these innovations (Walker, 2001). In particular, Part 2 of the Act made provisions for Information Meetings (IMs) that would be compulsory

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

for married people intending to initiate divorce proceedings. The compulsory nature of the IMs implies that they would have to be delivered in such a way that few people would be likely to object to attending one. The research programme involved evaluating a range of alternative *pilot* IMs – for example, varying the length, style and content of the meeting – in different regions. This chapter summarises one distinct research strand, identifying a set of potential IM sites whose locations would make it unlikely that many people would object to attending IMs specifically because they could not readily reach any of the IM sites.

The research described below aimed to provide the information needed by the Lord Chancellor's Department (termed 'the Department' here) in order to decide the number – and location – of sites at which IMs would need to be provided. It should be noted at the outset that the term 'site' here denotes an approximate location (e.g., the Victoria area of Westminster): the research did *not* aim to choose specific premises as venues for IMs. In practice, the Department's policy decision would centre on judgements as to the likely impact on married people if, for example, the government felt it was no more obliged to provide all areas with a local IM site than it is to provide a university in every locality. By providing IMs at relatively few sites the implementation costs should be minimised, but at the risk of increasing the number of objections by people who found the sites hard to reach. The research challenge is thus to identify how many sites would be needed – and where these should be – so that few people would have to travel far to reach one.

There are many existing public services which are provided through networks of local sites: examples range from the various kinds of court provided by the Department to other local public services provided by central government (e.g., Job Centres). These networks have typically evolved over many years, gradually adjusting to alterations in the services provided and also to changes in the pattern of demand such as the development of new towns. Each of these networks evolved differently to meet its own service's needs, so none of them provides an ideal blueprint for the selection of IM sites. Although the networks of sites used by other public services evolved over long periods, sophisticated methods of analysis have increasingly become available to help maximise the sites' accessibility to their users. For example, Thomas *et al.* (1991) illustrated the value of a location-allocation model in helping to inform decisions on which of the North West region's courts to close while minimising any increase in the distances people have to travel to reach their nearest court. The basic elements of a location-allocation modelling strategy are:

- mapping the spatial distribution of demand for the service;
- measuring people's reluctance to travel to access the service;
- identifying candidate sites for providing the service; and
- selecting the optimal sites from these candidates.

It is the final step which uses the location-allocation model itself and here it becomes most clear how the results' optimality depends on certain assumptions, including those which are embedded in the three previous steps.

This chapter reports location-allocation modelling to tackle the highly unusual proposal to 'roll out' a full national network of sites rapidly for a new public service:

this is clearly more a case of a 'big bang' than the process of evolution which is characteristic of the public sector. The rapid development of a new network is rather more familiar in the private sector than in public services, although even the vigorously growing commercial services such as multiplex cinemas have in fact been 'rolled out' over several years. It is notable that location-allocation modelling has been extensively adopted for site selection in the private sector (Ghosh and Harche, 1993). In the private sector, the key requirement is usually that the sites are close enough to large concentrations of people so as to maximise the likely profitability of the network. By contrast, the public service emphasises more the minimising of differences in service provision across the country. The general framework of location-allocation modelling provides for choices to be made between such different objectives, so that each analysis can be 'customised' to meet the specific objectives of that particular application.

The analyses in this chapter build up the information that is needed for a location-allocation model relevant to identifying sites for IMs. By way of clarification, it should be stated at the outset that all the analyses – the three preparatory stages and then the location-allocation modelling itself – rely upon a geographic information system (GIS) which can provide a high level of resolution through recognising up to 10 000 separate areas across England and Wales (these areas are the local government wards existing at the time of the 1991 Census). Fotheringham *et al.* (1995) showed that the size and shape of the zones used in a location-allocation model can influence the results of the analysis, but here the level of detail provided by nearly 10 000 zones means that there is very little loss of accuracy in a national-level model when the analysis treats each ward's population as if it were all located at the ward's population-weighted centroid. Straight-line distances between the respective centroids are similarly assumed to provide robust estimates of the distances between wards. All models simplify some of the complexity and peculiarities of the real world, the important task is to portray accurately enough the major factors that will shape the evaluation of the alternative scenarios to be explored (Densham and Goodchild, 1990).

The first preparatory analysis is outlined in Section 16.2 and is termed *demand distribution* because it attempts to estimate, for each local area of the country, the number of people who may attend IMs. It is necessary to make broad estimates of the 'population at risk' of attending IMs because the Department has not suggested restricting attendance to an identifiable category of people, such as those people who have petitioned for divorce. A key problem to be faced is that there is insufficient information available on the current whereabouts of relevant population groups (e.g., couples who have been married between five and 10 years). The information that *is* available is modelled to provide plausible estimates at the highly localised scale; this level of detail is necessary for the location-allocation modelling to identify a set of sites that will be as near to as many people as possible. The second preparatory step – described in Section 16.3 – analyses data from the pilot IMs on the extent to which people are deterred from attending IMs if they have to travel further. The results of this *distance deterrence* analysis provide an essential input to the location-allocation model. If people are very strongly deterred, then the model is designed to focus on reducing the minority of very long journeys in more remote areas; if

distance deterrence is not so severe, then the emphasis will be more on making smaller reductions to the journeys of very many more people through the provision of additional sites in heavily populated areas. Section 16.4 of the chapter describes the third preparatory set of analyses that provide a set of suitable *candidate sites* for the location-allocation model. For example, the mountainous centre of the Lake District is unlikely to be a suitable site to serve the population of Cumbria even though it is equidistant from all the county's main population centres (which lie in an arc around the periphery of the Lake District from Kendal through Barrow and Workington to Carlisle). The type of site favoured by people who attended IMs during the pilot phase of implementation provides the basis for identifying the criteria that sites need to satisfy. These candidate sites will then be input to the location-allocation model so that only suitable sites could be included in an eventual network. Section 16.5 describes the location-allocation modelling itself, then Section 16.6 provides a brief evaluation of the results. The chapter ends with some conclusions about this particular research, as an example of modelling in a dynamic policy context.

16.2 Estimating the Demand Distribution

The first step in the modelling research is to produce a set of local area predictions of the numbers of people likely to attend IMs in any year. An annual national total was estimated by the Department separately, and the local area predictions are used to 'share out' this total national demand between the local areas. For the modelling here to choose the best IM sites, this national total value is not actually relevant: the other inputs to the model ensure that central London would be the 'first choice' site regardless of whether the national annual demand was a hundred or a million people.

The best starting point for a model to estimate the likely local distribution of IM attenders is the number of divorces registered at each of the county courts across the country, so the modelling used the 180 county courts for which the number of divorces in the mid-1990s was known (between 165 000 and 170 000 a year in England). This dataset provided only a count of divorces at each court: there is no available dataset with postcoded addresses of divorce petitioners, so assumptions are needed about where people who divorce at any specific court actually live. The first step of this process was to define court catchment areas in terms of 1991 wards, for which areas it would then be possible to compile Population Census data as the basis for the model's independent variables. The catchment area definitions sought to reflect the guidance that the Department publishes, in which around 25 000 'places' are listed and allocated to individual courts. Substantial efforts were needed to allocate each ward to a court area, because there is no linkage between the 25 000 'places' listed in the directory and either wards or any other set of standard statistical or administrative areas. A further necessary step was to group together some county courts, because divorce petitions are not heard in every court. Information on other patterns of movement, notably commuting, was used for this grouping: for example, the Caerphilly county court area was grouped with Cardiff because this city is the predominant destination of out-commuters from Caerphilly and the adjacent

Table 16.1 *Sample of divorce petitioners at North East courts*

	Female (%)	Male (%)	All (%)
Petitioned in 'local' court	77.6	72.5	76.2
Petitioned in 'adjacent' court	17.3	12.3	15.9
Petitioned elsewhere in region	2.3	11.6	4.9
Petitioner from outside region	2.8	3.6	3.1
N	*353*	*138*	*491*

villages. The output from this step is an estimated catchment area of each divorce court.

Although there is no national dataset about where the people who divorce at each court in fact live, anecdotal evidence suggested that people do *not* always use their local court. The addresses of a sample of people who petitioned at courts in some of the North East pilot areas was postcoded as part of the research. Table 16.1 presents the evidence from this dataset as a test of the appropriateness of the catchment areas that have been estimated here. Over three-quarters of petitions were from people who did indeed live within the estimated catchment areas. The proportion was slightly less for male petitioners, but they are notably in the minority in practice (McCarthy and Simpson, 1991, established that men were more likely to move away from the marital home during the divorce process). Of those who petitioned at North East courts but did *not* live within that court's estimated catchment area, most lived just across that estimated catchment area boundary in one of the adjacent areas. A substantial degree of 'boundary hopping' is in fact inevitable with courts so closely spaced as are, for example, Newcastle and Gateshead (bearing in mind that both North and South Shields also have divorce courts). In many parts of the Tyneside conurbation neighbouring people have petitioned at different courts, so adjustments to the estimated catchment area boundaries could barely, if at all, reduce 'boundary hopping' in such areas.

This interpretation is borne out by the results in the other areas where a court sample was collected. Those areas that cover a similar urban/rural mix to the North had a very similar proportion of court sample members petitioning at their 'local' court. More heavily urbanised areas such as Greater Manchester and, most especially, London had noticeably higher proportions of petitioners 'boundary hopping' due to the very close spacing of courts there. All the same, the only practical assumption here is that the numbers of people moving around the country to petition for divorce roughly cancel each other out, so that the number registered at each court should be a fairly close approximation to the number of petitioners who live in that court area and – for the purposes of this research – also should provide a close approximation to the number of potential IM attenders.

Using these 180 court catchment areas, a regression model was developed which used the courts' number of divorces as the dependent variable, and their catchment areas' population size and characteristics as the independent variables. This model could then be used to 'shift scales' down to the ward by predicting the number of divorces likely to arise from each ward's population size and characteristics. These ward-level predicted values will then be turned into 'shares' of the national total of predicted values, which in turn will provide the basis for locally 'sharing out' the total number of people expected to attend IMs nationally.

The dependent variable is *the number of divorces per thousand married men*: the variables were generally defined in terms of male characteristics because some of the potentially important factors (such as unemployment) are not so well measured for women. There is not a huge literature on the factors that may be important in shaping the rate of divorce in this country, although Clarke and Berrington (1999) recently provided a valuable summary. Research on the subject has also been hindered by the limited official statistics available, because divorce registration procedures date back many years and do not, for example, record either party's date of birth and this hampers some longitudinal analyses of divorce (cf. Kiernan and Mueller, 1999).

The age of married people – and the duration of their marriage – are identified in the literature as major influences on the likelihood of divorce. The modelling experimented with a number of detailed age groups variables, with the final model eventually selecting a variable relating to the 16–44 age group. Ethnic minority populations were represented by four variables which measured the proportion of men aged 20–54 who were (a) Afro-Caribbean or (b) Pakistani/Bangladeshi or (c) Indian/Chinese or (d) Irish (see Berrington, 1996 and Beishon *et al*., 1996, for the background research). As well as a variable on the proportion of men in low-skilled occupations, two alternative measures of male unemployment were made available to the model as options to represent the issue of poverty and/or insecure employment, which is known to be linked to divorce (Böheim and Ermisch, 2001). Three other employment-related variables were explored, all representing factors that the literature suggests to be positively associated with increased rates of divorce (e.g., Gibson, 1996):

- the proportion of employed men who have security-related occupations;
- the proportion of married women who either are self-employed or have full-time jobs; and
- the proportion of all households that included two earners but no children.

A variable on the proportion of younger men in the area who are married was among the variables to represent demographic factors that could represent contextual influences on the local level of divorce. The two final variables assessed were both more familiar features of area-based modelling. A housing tenure variable measured the proportion of households that are owner-occupying (this has been argued to be negatively associated with the probability of divorce). The other variable tested was a measure of the level of urbanisation in each area (Coombes and Raybould, 2001).

One important feature of area-based modelling such as this is that two variables which have similar distributions across the areas analysed can 'cancel each other out' in their effect on the dependent variable. As a result, modelling the difference between areas' populations can yield rather different results to modelling the behaviour of the individuals who make up those populations. A particularly clear example of this 'cancelling out' involves the ethnicity variables. It is known that Afro-Caribbeans have higher divorce rates than white people, whilst the Pakistani and Bangladeshi groups have low rates. At the scale of analysis undertaken here, the same areas – in practice, the more rural areas – tend to have populations that include very few people from either of these broad minority ethnic groups. As a result of this similarity of the geographical patterns of the Afro-Caribbean and the Pakistani/Bangladeshi variables, they tend to 'cancel each other out' in the court area analyses and so neither emerges as significant in the model.

After considerable experimentation on technical aspects of the modelling, and also on the precise detail of the independent variables, the robust model to emerge included significant variables, all taking the expected signs, on the age and class/occupation of local populations *plus* a factor lowering the rate according to the proportion of the younger men in the area who are currently married. The latter variable suggests a contextual effect that produces greater resistance to divorce in areas where the proportion of young men who are married is higher than the national average. To check the sensitivity of the modelling, the analyses were run on each of two separate years' data, as well as the two in combination. The results were encouragingly consistent in the selection of variables as significant, and in the calibration of those variables. There is also no clear pattern to the residuals from the model which would call for additional variables to be included. Yet in practice the model's predictions do not widely diverge from an alternative set of results based purely on the number of married people living in each area. Although it is rather disappointing that the efforts devoted to developing this model have not provided a dramatic improvement in the accuracy of predictions vis-à-vis those which could have been obtained by assuming a simple per capita divorce rate, it might also be seen as reassuring that the estimates here do not substantially depart from 'bedrock' demographic data. As a result, this model has been used here to predict the ward-level distribution of the numbers expected to attend IMs nationally: this then provides the demand measure for the location-allocation model.

16.3 Estimating the Level of Distance Deterrence

The next step is to model the degree to which people are reluctant to travel to reach an IM site. This reluctance can be estimated here by considering the behaviour of people in the pilot areas. For each person registering their interest in attending an IM during the pilot phase, their *nearest* IM site was identified by calculating straight-line distances between their home postcode and the locations of all IM sites which provided at least one IM after the date on which the person registered.

Table 16.2 examines both these distances to nearest sites, and the actual distances travelled to pilot IM sites. The first row presents the distribution of distances, for the

Table 16.2 *Distance from meeting sites*

	Distance band (km)					
	0–1	1–2	2–5	5–10	10–20	over 20
% people who attended IMs:						
distance to IM site attended	2.7	7.8	29.0	26.8	20.9	12.8
distance to nearest IM site	4.5	10.0	35.4	30.7	16.0	3.4
% non-attenders:						
distance to nearest IM site	4.7	10.9	36.1	29.7	15.9	2.8

people who *did* attend IMs, between the IM site which they attended and their home address. The next row shows the distances between the same people's home locations and their nearest meeting site. The difference between the values in the two rows is rather modest, suggesting that the vast majority of people attended their nearest meeting site (although nearly 10% did in fact travel over 20 km when there was a nearer meeting site). The third row shows the distance to the nearest site for those *non-attenders* whose home location is known. Comparing the second and third rows suggests that the distance to the nearest meeting is not necessarily the major influence on people's likelihood of attending: if non-attendance had been very substantially fuelled by remoteness of meeting sites, then a noticeably higher proportion of non-attenders would have been found to live in areas which are a longer distance away from the nearest IM site.

Table 16.2 can also be seen to provide an overview of the kinds of distances which people are prepared to travel to attend IMs. More than one in eight of the people attending pilot IMs travelled more than 20 km from their homes, and over a third attended sites at least 10 km away. People who had attended pilot IMs were interviewed, and given other opportunities to feed-back their views, but very few references were made to the remoteness of IMs posing a substantial difficulty. As a result, it might be hesitantly concluded here that for *most* people a journey of up to 20 km is unlikely to pose a major problem.

There was, perhaps surprisingly, little difference between the median distance travelled by women and men (6.3 and 6.6 km respectively). The median distance travelled by car users (7.3 km) was around 50% longer than that for the users of each of the other methods of travel except train, hinting at the higher levels of mobility which are commonplace among more affluent people. This issue is explored here by using the Townsend Index of deprivation (Townsend *et al.*, 1988), which measures average levels of deprivation (positive scores), or affluence (negative scores), for each ward in the country. People who registered an interest in attending one of the pilot IMs provided their home postcode for the research so the Index value of their home area can be identified. Table 16.3 draws upon these calculations for all people attending pilot IMs, grouping them according to their home area Townsend Index values. The first row shows a notable pattern, in which people living in the most deprived neighbourhoods are found much more often to live near to IM sites. This result will be

Table 16.3 *Home neighbourhood and likelihood of attending IMs*

	Townsend Index values					
	Up to –4	–4 to –2	–2 to 0	0 to +2	+2 to +4	Over +4
Distance from nearest pilot IM (km)	14.1	10.4	11.1	12.0	10.4	4.2
Probability that attended	0.85	0.83	0.82	0.81	0.77	0.69
(% of all who registered)	(12.4)	(28.7)	(23.5)	(15.3)	(10.0)	(10.0)

mainly due to many deprived areas being in inner cities, which are thereby near to the pilot IM sites whose locations were mostly in town or city centres.

The relationship between the level of deprivation of people's home neighbourhood and the likelihood of them attending IMs can be examined directly with the evidence about people who had made themselves known to pilot projects. Table 16.3 shows that the majority of people known to the research are from non-deprived areas: the three left-hand columns, which cover people from non-deprived areas, include around 65% of all people in the pilot research dataset. The second row then reveals, for each type of area, the proportion of the people who expressed an interest in attending who did indeed attend a meeting. There is a consistent, if unspectacular, tendency for the more affluent groups to be more likely to attend a meeting whilst those from the most deprived areas are the least likely to have attended.

The extent to which people's reluctance to travel to attend IMs increased with their distance from the nearest site provided the basis for the distance deterrence model here. The first step was to categorise people who expressed an interest in attending a pilot IM according to the distance they lived from an IM (viz. the first category includes all who lived under 1 km from their nearest site, the second category includes those at least 1 km from a site but less than 2 km away, and so forth). For each category, an appropriate 'typical' distance was assumed (e.g., for the first category this was 0.3 km, and for the second 1.3 km). Not everyone who registered an interest did eventually attend an IM and, in practice, people who lived further from an IM site proved to be less likely to attend. The distance deterrence model uses the proportion of people, in each distance category, who did *not* attend an IM as the dependent variable in a non-linear regression against 'typical' distances (using SPSSx). The beta value from the regression is then taken as the distance deterrence coefficient in the location-allocation modelling.

Ever since the early gravity models, the default distance deterrence value has been 2 and, in practice, empirical evidence has quite often found that the model which best fits the data on people's resistance to travelling has the distance squared. Rather higher distance deterrence is found for very 'everyday' activities, whereas the 'one off' nature of IM attendance suggests people may be more willing to travel further. On this basis, a distance deterrence exponent value nearer to 1 might be expected here. In fact, the modelling found the exponent's value to be slightly below 1 when the whole research sample population was analysed.

The distance deterrence analyses were also run on sub-groups of the data: the research sample had been partitioned according to Townsend Index scores, so as to identify separately those who live in deprived/average/affluent areas. These sub-groups' exponent values show the variation which would have been expected, with the deprived group's value found to be 1.2 whilst the affluent group's was 0.9 (and the average group's is 1.0). Given that the policy concern is that the IM sites should be accessible to *all* groups, it seems appropriate to use the more deprived group's deterrence value in the location-allocation model. In other words, the IM sites were then selected on criteria which, in effect, assumed that everyone attending IMs has the higher resistance to travelling which was found among those living in more deprived areas. This approach also protects the analysis somewhat from the probability that many of the people included in the research data do not appear to be so deterred from travelling longer distance simply because they are 'volunteers' who come from more affluent backgrounds. Even so, the relatively low level of the deterrence values which have been found here were broadly to have been expected, given that IMs are concerned with such major issues in people's lives.

A footnote to these analyses is that the distances calculated here have all, in effect, assumed that people have travelled from home to IMs. Table 16.2 had shown that a notable minority of people did *not* attend the IM site nearest to their home, which suggests that the travelling to IMs may often be undertaken as part of a multi-purpose journey. In particular, people may quite often travel from work rather from home (e.g., in their lunch-break). As a result, people's decisions about 'how far is too far' may quite often be based on calculations which are rather different from the distance-from-home measures which have been estimated here.

16.4 Identifying Candidate Sites

One of the key concerns in identifying suitable IM sites is the more restricted level of mobility of some members of the population. A site in the countryside equidistant between three nearby towns might seem to be a geographically 'fair' solution, but would this be appropriate? In particular, would such a site be suitable for people without access to a car? In the absence of comprehensive public transport service information, the search for suitable IM sites *could* be restricted to places within a certain distance of a major road, but the assumption that all such roads have a reasonable level of public transport services would not be a particularly safe one. By contrast, it *is* likely that the 'high street' of a town and city centre, or a major sub-urban centre, will feature quite high levels of public transport services from its catchment area. This approach to site selection was reinforced by the findings from the pilots, in which people stressed their preference for IM sites located in busy areas to which people go for a variety of purposes.

Considerable data analysis was required to implement this strategy because there is no 'off the shelf' list of appropriately defined town centres. Following substantial experimentation, the criteria for a candidate site were set as the requirement for the location to have at least two of the following characteristics:

- it has a Benefits Agency office;
- it has a market (open or covered); and
- it is a substantial focus for commuting flows.

These characteristics were chosen partly because they were likely to be associated with places frequented by less affluent people and so be more likely to identify centres which have better public transport provision.

The list of Benefit Offices used here does not include 'neighbourhood' Benefits Agency offices, which would have provided a much longer list but would not have been suitable for indicating 'high street' locations. Market towns were identified as locations holding a retail market – on days other than Sundays or Bank Holidays – more often that once a month. No source provides an up-to-date comprehensive list, so the method adopted here was to include any town recorded as holding a market in one or more of four different sources (ranging from a general-purpose gazetteer of towns to the developing Internet site of the National Market Traders' Federation). This inclusive approach minimised the risk of 'missing' a town – which thereby might not be identified as a candidate site – purely due to incomplete data.

The identification of commuting foci was automated. The first step was to analyse the data on flows between wards in the 1991 Census commuting dataset. Of over two million such flows, the program distinguished those of at least 10 (note: these are in fact flows which involve over 100 people, due to the sample frame of this dataset). The next step was to identify wards whose centroids are so close together they can be considered to be parts of the same location. Wards whose centroids are less than 1.5 km apart are in effect, linked together. (This device side-steps the tendency in some towns for a major employer to be located a short distance away – and thus possibly in a different ward – from the High Street where the Benefit Office and/or market is likely to be located.) Wards whose centroids were no more than this distance apart were then grouped together for the purpose of counting the number of large flows to them. The criterion adopted was that a (grouping of) ward(s) had to have at least three large commuting flows to be considered a substantial commuting focus here. Figure 16.1 illustrates the results of the analysis in the North East region. It can be seen that this multi-criteria approach tends to find the 'high street' centres of most small to medium size towns. The locations qualifying with just one of the three criteria are mostly suburban parts of the larger towns (e.g., Newcastle). Over 2000 potential sites nationally possess at least one of these three characteristics, of which 647 had at least two and so make up the list of candidate sites. Together these 647 provide a very comprehensive coverage of the country, with only a few rather remote upland areas more than 25 km from all of them.

16.5 Location-Allocation Modelling

All the research detailed so far in the chapter has been necessary preparation for the location-allocation modelling. As the core of the challenge which was faced, it needs to be described here in a little detail. The location-allocation modelling used the

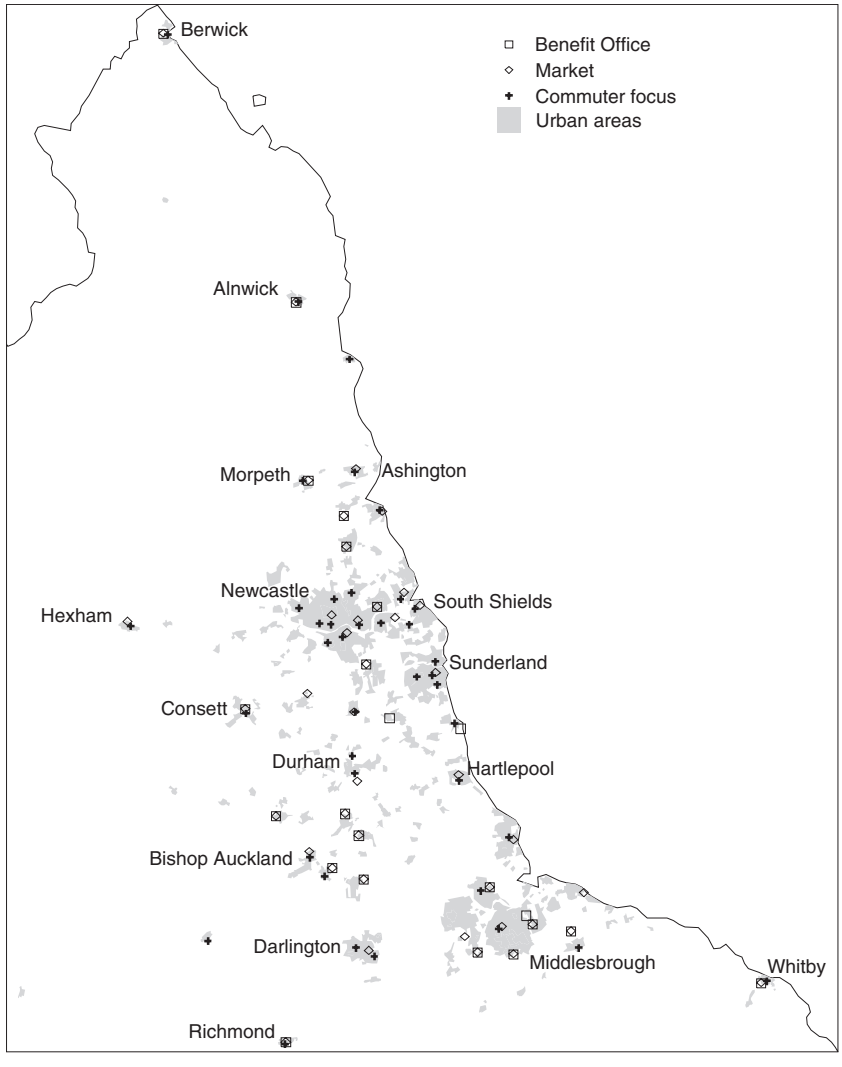


Figure 16.1 *Candidate sites in North East England*

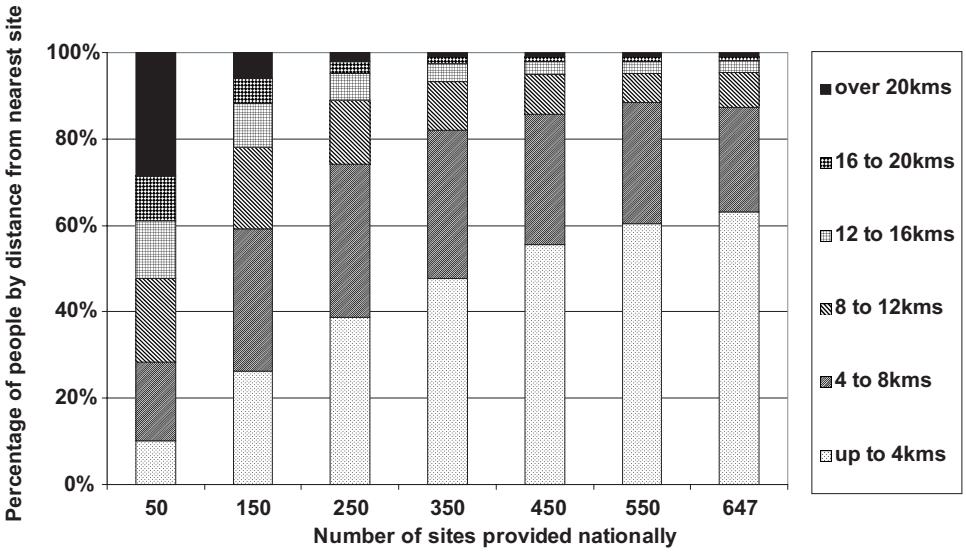


Figure 16.2 *Distance to nearest site, depending on the number of sites*

ARCPLOT module within the ARC/info package. The 647 candidate sites were input, along with the demand distribution model's estimate of demand from each ward, plus the distance deterrence coefficient. The analyses used *mindistpower* as the *locatecriteria*: this means that the analyses sought to choose sites which minimise the total distance travelled, but with these distances weighted by the distance deterrence coefficient. The optimisation procedure needed to be as fast as possible, given the very large number of wards and sites involved, so the default heuristic GRIA was applied. The allocation results were exported to Excel for post-processing and presentation. An additional point of information is that the Department did not set a minimum workload – a number of people to use each site over a year – although this could be expected to reduce costs (Densham and Rushton, 1996). For the record, it appears that this analysis (with nearly 10 000 separate demand locations and nearly 650 candidate sites) is one of the largest ever applications of location-allocation modelling.

The strategy for the modelling was to carry out many different analyses, each reflecting a different scenario from a wide range of options. As has been stressed from the outset, the key policy decision here is 'how many sites should IMs be provided at?' Figure 16.2 summarises the results from the analyses, in terms of the distances people would have to travel, depending on the Department's choice over the number of IM sites. It can be seen that increasing the number of sites much above 250 makes surprisingly little difference to the proportion of people attending who have to travel rather long distances. As an extreme extension to these analyses, it was found that over 2500 sites would have been needed for less than 5% of people to have to travel over 8 km (5 miles). The decision was made by the Department to explore in detail the 250 site option, which provides 95% of potential attenders with a site no more

than 16 km (10 miles) from their home. One notable advantage of this decision for the modelling was that it then became unnecessary to build in site capacity constraints (cf. Murray and Gerrard, 1997), because this number of sites meant that even the busiest one was estimated to have fewer people attending it over the course of a year than had been separately estimated to be a maximum practical workload for one site.

There were no data available, on local patterns of journeys made for comparable reasons, on which to base the predictions of where people would go to attend IMs (cf. Hodgson, 1990). The basic modelling was rooted in the assumption that people would attend IMs after making a journey from home. As noted earlier, a substantial minority of the research sample of people appeared to have travelled to the IMs from work, while others combined the trip with a visit to shops or other facilities. This confirmed the wisdom of only allowing 'high street' locations as plausible sites; it also meant that the distribution of jobs, rather than resident population, provided a better 'starting point' for calculating the distances travelled by some people.

In the absence of a reliable estimate of the proportion of people travelling to IMs 'as if from work' (rather than from home), five scenarios were run with different break-downs of the total attending (viz. 100% 'as if from work' rather than from home; 75% 'as if from work' and 25% from home and so forth). It was notable that it was in the conurbations where the results were most sensitive to the choice of scenarios, with additional city centre locations more likely to be selected when the proportion assumed to be travelling from home was lower. Even so, the overall level of sensitivity to this parameter setting was quite low: town and city centres are usually the optimal locations for IMs *whether or not* people are travelling mostly from the suburbs, so the locations of the selected sites tend not to change greatly between the different scenarios. It was also found that slightly fewer people have to travel longer distances for options in which a proportion of the people are travelling 'as if from work' due to the fact that, on this basis, more of them are assumed to start their journeys in or near town and city centres.

Figure 16.3 shows the location of sites, chosen by any of the five scenarios – with differing proportions assumed to travel 'as if from work' – in the North East region. Sites shown by a diamond in a square were selected by all the five scenarios; those shown as just a square were selected by three or four, while sites represented by diamonds were selected by only one or two scenarios. The remaining candidate sites, shown simply by crosses, were not selected by any of these five analyses. As already indicated, the areas with the most instability tend to be in and around the larger cities. Even there, the effect was often simply to shift the chosen site a mile or two, helped by the fact that quite a large number of the 647 candidate sites are located just a short distance away from each other in the major cities. Looking at the national picture, it was unsurprising that the London area showed the highest level of instability, although again very many of the changes there simply shift the selected site a relatively short distance. Given these location-allocation modelling runs – each finding the optimal 250 sites based on one of five different scenarios in terms of the proportion of people travelling 'as if from work' – the final step was to identify the sites selected by at least three out of the five runs. Figure 16.4 shows the 249 sites selected in this way, and draws attention to the rather few parts of the country, most

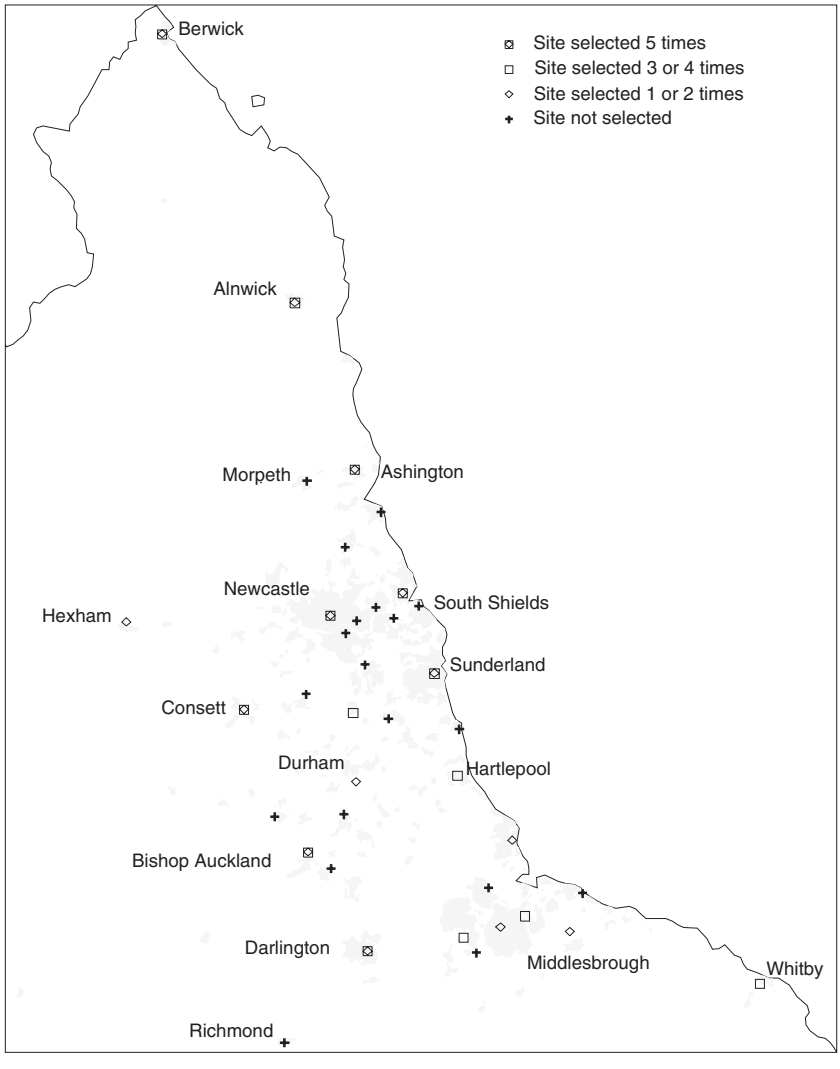


Figure 16.3 *Site selection in North East England*

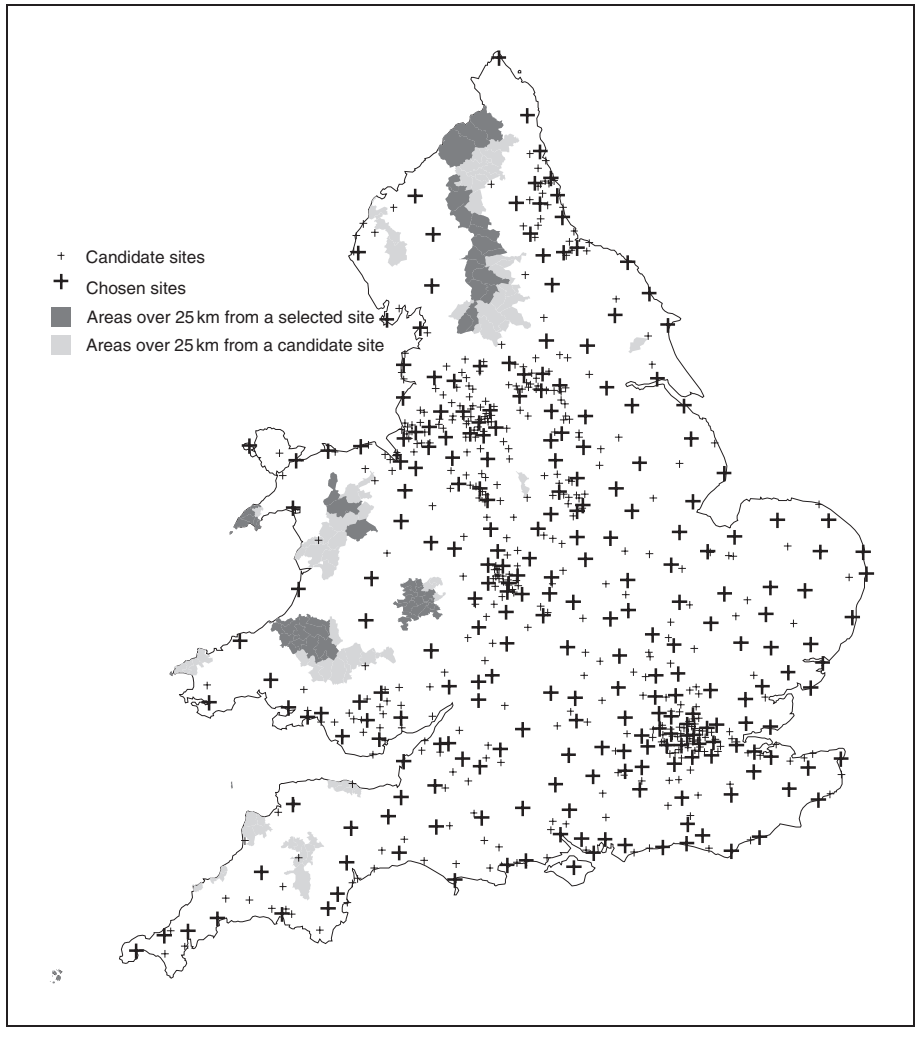


Figure 16.4 *Candidate and selected sites*

of which are in fact only thinly populated, where people would face journeys of over 25 km to their nearest IM site.

16.6 Evaluation of Selected Sites

One way of assessing the results from the modelling is to compare the 249 sites' accessibility with that of the network of Divorce Courts and, in addition, the Benefit

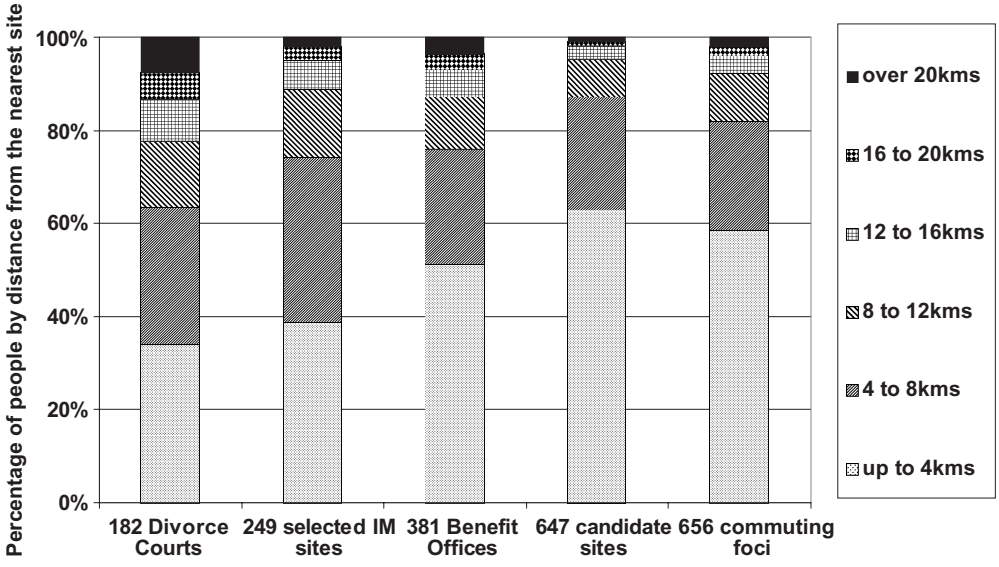


Figure 16.5 *Distance to nearest of the 249 sites, and to the sites of comparator networks*

Offices and commuting foci which were part of the analysis to identify candidate IM sites. Figure 16.5 shows that only the two sets of well over 600 sites substantially surpass the 249 selected IM sites in the proportion of people who would have less than 4 km to travel. In fact the location-allocation model has identified a network of sites which does seem to have minimised the number of people having to travel very long distances. These comparisons give a substantial degree of reassurance that the 249 sites identified here make up *at least* as 'fit for purpose' a network as do these long-established sets of sites whose levels of accessibility seems to attract little public criticism.

16.7 Conclusions

Evaluation suggests that the site identification method which has been developed proved able to identify an appropriate network of sites. This is not necessarily to argue that the set of 249 sites is definitely the best possible network of sites for IM provision: the aim was the more modest one of answering questions which are popularly phrased 'what if . . . ?'. More formally put, the results from the analyses are based on a specific set of hypotheses, assumptions and policy preferences. The modelling has been explicitly designed to allow different scenarios to be explored and, as a result, to assess the sensitivity of the results to changing from one option to another. Figure 16.2 provided the clearest illustration of this sensitivity testing, but many other permutations have been explored to build up the central case results which have been described here. The analytical challenge, which has been met, was to test scenarios with numbers of sites numbered in the hundreds – and nearly 10 000

'demand points' – when location-allocation models are more often applied to a single city or urban region (e.g., Yeh and Man, 1996).

For reasons unrelated to the modelling, the government has subsequently decided not to proceed with a national implementation of IMs. The new proposal is for a set of Family Advice and Information Networks: new research has begun to evaluate this alternative way of providing information to divorcing couples. The decision to change policy – and so not to implement the provision of IMs as outlined here – was based on concerns over just what couples need at times of possible family breakdown. The decision did not turn on any of the geographical issues discussed in this chapter. In its own terms, this research successfully met its objectives. Perhaps the single most important lesson to be drawn was that the results are viewed with more confidence, especially by policy-makers who are unfamiliar with these geographical issues and techniques, if a range of scenarios are presented. Handling uncertainty 'upfront' in this way was found helpful especially, of course, if the results did not vary too radically between the different scenarios. The research also proved, in practice, that uncertainty over policy shifts cannot be handled in a similar way.

Acknowledgements

We are grateful to the Lord Chancellor's Department who funded the research underlying this chapter, as part of a major research programme directed by Professor Jan Walker at Newcastle University's Centre for Family Studies (NCFS) (*see *Information Meetings and associated provisions within the Family Law Act 1996: Final Evaluation Report*, http://www.lcd.gov.uk/family/fla/fullrep.htm). We also fully acknowledge the inputs of numerous colleagues in the research programme, both within NCFS and elsewhere, whose contributions helped to ensure that the methodological developments briefly reported here were as appropriate as possible in meeting the overall objectives of the research programme. We also thank Dr Paul Densham (UCL) for his comments on the modelling work carried out.

All the GIS-based analyses conducted here depended on the use of ARC/Info software from ESRI(UK) who very kindly agreed to its use not only for the authors' academic research but also for this contract-funded research.

References

Beishon, S., Modood, T. and Virdee, S. (1998) *Ethnic Minority Families,* Policy Studies Institute, London.

Berrington, A. (1996) Marriage patterns and inter-ethnic unions, in Coleman, D. and Salt, J. (eds) *Ethnicity in the 1991 Census: Demographic Characteristics of the Ethnic Minority Populations,* ONS, London.

Böheim, R. and Ermisch, J. (2001) Partnership dissolution in the UK – the role of economic circumstances, *Oxford Bulletin of Economics and Statistics,* **63**: 197-208.

Clarke, L. and Berrington, A. (1999) Socio-demographic predictors of divorce, in Simons, J. (ed.) *High Divorce Rates: The State of the Evidence on Reasons and Remedies,* Research Series 2/99 (Vol. 1), Lord Chancellor's Department, London.

Coombes, M. and Raybould, S. (2001) Public policy and population distribution: developing appropriate indicators of settlement patterns, *Environment and Planning C: Government and Policy,* **19**(2): 223–48.

Densham, P. and Goodchild, M. (1990) Spatial decision support systems, *National Center for Geographic Information and Analysis Technical Paper 90:5* Santa Barbara, California.

Densham, P. and Rushton, G. (1996) Providing spatial decision support for rural public services that require a minimum workload, *Environment and Planning B: Planning & Design,* **23**: 553–74.

Fotheringham, A.S., Densham, P. and Curtis, A. (1995) The zone definition problem in location-allocation modelling, *Geographical Analysis,* **27**: 60–77.

Ghosh, A. and Harche, F. (1993) Location-allocation models in the private sector: progress, problems and prospects, *Location Science,* **1**: 71–106.

Gibson, C. (1996) Contemporary divorce and changing family patterns, in Freeman, M. (ed.) *Divorce: Where Next?,* Dartmouth, Aldershot.

Hodgson, M.J. (1990) A flow-capturing location-allocation model, *Geographical Analysis,* **22**: 270–9.

Kiernan, K. and Mueller, G. (1999) Who divorces?, in McRae, S. (ed.) *Changing Britain: Families and Households in the 1990s,* Oxford University Press, Oxford.

McCarthy, P. and Simpson, B. (1991) *Issues in Post-divorce Housing,* Avebury, Aldershot.

Murray, A.T. and Gerrard, R.A. (1997) Capacitated service and regional constraints in location-allocation modelling, *Location Science,* **5**: 103–18.

Thomas, R., Robson, B. and Nutter, R. (1991) Planning the work of county courts: a location-allocation analysis of the northern circuit, *Transactions of the Institute of British Geographers,* **16**: 38–54.

Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and Deprivation: Inequality and the North,* Croom Helm, London.

Walker, J. (2001) Divorce reform and the Family Law Act 1996, in Walker, J. (ed.) *Information Meetings and Associated Provisions within the Family Law Act 1996: Final Evaluation Report (Volume 1),* pp. 17–34. (http://www.lcd.gov.uk/family/fla/fullrep.htm)

Yeh, A.G.-O. and Man, H.C. (1996) An integrated GIS and location-allocation approach to public facilities planning – an example of open space planning, *Computers, Environment and Urban Systems,* **20**: 339–50.

17

New Methods for Assessing Service Provision in Rural England

Martin Frost and John Shepherd

Abstract

This chapter describes work undertaken on behalf of the Countryside Agency to use detailed information on the location of services and households to provide evidence that underpins policy development in two interrelated areas within England. The first is based on household access to services and facilities in rural areas while the second examines the role of small and medium-sized towns in acting as local centres of service provision. Commonly used geographical information system (GIS) procedures including the identification of nearest services and the use of point in polygon allocations are employed to measure levels of access to services and to generate service profiles of the towns. The results show that measuring access to services produces a generally more favourable view of rural service provision than the 'presence and absence' ratios traditionally used by the Agency. The profiles of small and medium-sized towns show high levels of variability suggesting a rather chaotic form for the concept of 'market towns' that in recent years has been a strong component of public policy.

17.1 Introduction

Although rural service provision has been a longstanding policy issue in England (for an early statement see, for example, Moseley, 1979), there has been an acceleration of interest in the 'state of the countryside' in recent years as the declining fortunes of farming, combined with the loss of many local shops and services in rural areas, has reinforced a recognition by the Government that problems of poverty and exclusion are found in both rural and urban areas of the country. The Government's view

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

is captured in the introduction to the Rural White Paper, published in the Autumn of 2000 by the Department of Environment, Transport and the Regions, as follows: 'The challenge for rural communities is clear. Basic services in rural areas are overstretched. Farming has been hit hard by change. Development pressures are considerable. The environment has suffered' (DETR, 2000, p. 9).

The Rural White Paper (subsequently referred to as the RWP) goes on to set out a range of initiatives and targets relating to service provision, housing, transport, farming and the quality of the environment. In an era of 'evidence-led' policy development it is, perhaps, a little surprising that the level of empirical underpinning of many of these initiatives was, and still is, rather limited. In this chapter, we describe how the use of geo-referenced information analysed through GIS procedures has been used in research undertaken for the Countryside Agency to develop a more effective range of evidence in two key areas of Government initiatives: levels of access to rural services, and the role of market towns as centres of local service provision and economic development in rural areas.

In both areas of development, the work has been based on using information on the presence of individual service outlets in which the location of the outlets is either defined directly by a six-figure national grid reference (in the case of some commercial retail data such as that supplied by CACI) or has an address containing a unit postcode to which a grid reference can be attached. A range of data has been accumulated from both commercial sources and public agencies that locate many of the services identified by the Countryside Agency as of key importance in influencing the quality of life in rural areas of England.

The prime concern of the Agency, however, has been to progress from the identification of 'presence and absence' of services in two directions. The first is to assess, at the level of individual households, the geographical proximity that households have to their local services. The second is to use the range of services present within England's market towns to assess their 'vibrancy' in providing services and facilities to households living in surrounding areas.

In both developments, a combination of standard GIS procedures has been used to achieve these objectives. The analysis of household access to local services rests on the ability to identify an accurate geographical position for small groups of households using the postcode of their addresses and being able to locate (for each targeted service) the distance separating them from their nearest outlet. In the case of market towns, the emphasis is on being able to locate services within the built-up areas of towns that generally do not conform to recognised administrative boundaries (such as wards or local authority boundaries). This requires a point in polygon approach that has been supplemented in some of the analyses with the addition of buffer areas around the physical limits of each town that allow for recent expansion.

The chapter is divided into two parts. In the first part, we review the traditional ways of capturing information about the location of rural services as a basis for describing how this has been improved by adopting GIS approaches linked to digitally encoded national sources of information. In the second part, a discussion of the concept of market towns is followed by a demonstration of the profiling of these towns that is based on the capacity to locate key services within their boundaries.

17.2 Access to Services in Rural Areas: Identifying the Problems

The RWP generated no less than 15 headline indicators by which the state of rural areas within the UK was to be assessed. One of these was headed 'Geographical availability of services' and committed the Countryside Agency to maintain regular monitoring of the proportions of rural households living at specified distances from key services. The current list of key services monitored by the Agency includes banks and building societies, cash points, supermarkets, petrol stations, primary schools, secondary schools, doctor's surgeries, job centres, libraries and post offices. Although the notion of households' access to services may appear to be a simple one, producing measurements on a national scale that might be suitable to support the monitoring specified by the government has, traditionally, been far from easy. The problem has three components. The first is knowing where services are located, the second is knowing where households are, and the third is estimating the distances separating the two.

17.2.1 The Location of Services – The Rural Survey Approach

The traditional source of information on the location of services in rural areas has been the periodic surveys of rural services conducted by the Rural Development Commission (RDC). These were conducted in 1991, 1994 and 1997, and relied on parish clerks to complete questionnaires detailing the nature and extent of service provision within their parishes. The questionnaires were ambitious. They requested details on shops, pubs, post offices, banks, educational facilities, health facilities, medical facilities and recreational facilities, together with aspects of housing, local transport and welfare provision.

The basic data collection unit for all surveys was the civil parish which, in addition to its role in the reporting of other official data (notably from the Census of Population), constitutes an important focus of local democratic politics and policy action. The surveys covered only 'rural' parishes that were defined as those with a resident population of less than 10 000. The surveys recorded counts of the number of service outlets and facilities that lay within each parish. The results for the 1991, 1994 and 1997 surveys were published solely as paper documents (Rural Development Commission, 1992, 1995, 1998) with tabulations of the proportions of parishes containing a particular service or facility or a particular number of service outlets. Although this approach utilised the only comprehensive network of potential data collectors in rural areas, the parish clerks, its use in determining the location of rural services for the purpose of assessing levels of access was weakened by two key problems:

- *geographical coverage*: not all parish clerks filled in the forms or filled in the forms completely. Although considerable effort was put in by Rural Consultative Councils to achieve high response rates, in some counties response rates could be as low as 65%, even when an overall response rate close to 80% was achieved by the 1997 survey; and

- *urban/rural interaction*: limiting the surveys to parishes containing fewer than 10 000 residents ignored the fact that many rural areas are close to towns or cities (which do not, in any case, contain defined parishes) and have access to the services and facilities contained in the urban areas.

Either one of these shortcomings means that it is not possible to use these records as a way of identifying consistently how far rural households have to travel to access their nearest service or facility. For many households the nearest facilities might lie in a neighbouring town or in a neighbouring parish which did not respond to the surveys.

17.2.2 The Problem of Locating Households

Using parishes – or indeed any administrative areas – as the sole basis for analysing information on services in rural areas also accentuates the problem of identifying the location of households accurately enough to permit sensible analysis of their distance from local services and facilities. Counts of residents and households within parishes are only generated by the decennial Censuses of Population. Even then, the value of these counts is diminished by the size and irregular shapes of many parishes. There is great variation in parish sizes (about 30 current parishes are of fewer than 100 hectares while 15 are larger than 10 000 hectares) and, for many parishes, their elongated form makes it problematic to distinguish a meaningful centroid.

17.2.3 Assessment of Access from Survey Evidence

The combination of parish survey-based evidence on the location of services in rural areas, combined with the use of parish populations to assess the distribution of households, has severely limited the degree to which previous rural surveys have been used to measure levels of access to services. This is clear in the terminology used in the report on the 1997 Rural Services Survey (RDC, 1998) in which the term 'provision' of services is used consistently instead of any measure of 'access'. All tables within the report show the numbers of service outlets cross-tabulated against the population size of the parishes in which they fall, with many of the key headlines within the report summarising the proportions of parishes that do *not* contain a particular service or facility. A good example to illustrate some of the difficulties that may arise from this is to consider post offices as a key rural service. In 1997, 43% of responding parishes in the rural services survey did not contain a post office (RDC, 1998). By 2000, this figure had risen to 46%, reflecting the progressive closure of small rural post offices (Countryside Agency, 2001). Yet, using the GIS procedures outlined in the next section of this chapter, it was estimated that nearly 94% of households living in rural areas had a post office within 2 km of their homes.

The key variable in explaining the difference between these two sets of figures – one based on provision, the other based on access – is the uneven distribution of population across parishes. In the 1991 Census of Population, about 13% of all

parishes had 100 or fewer residents, rising to 27% of parishes if the threshold is increased to 200 residents. In 2000, 88% of parishes with fewer than 100 residents did not contain a post office with an equivalent figure of 78% for those with between 100 and 200 residents. A concentration on provision alone is clearly weighted by the significant number of parishes that were defined many centuries ago but now contain few residents and correspondingly few services.

17.3 Using GIS Procedures to Measure Access to Rural Services

The development in recent years of a greater availability of digitally encoded lists of services and households that permit accurate geo-referencing, together with the software to link the two together, has transformed our ability to measure access to services at detailed spatial scales throughout the country. This can be seen in the contents of the Rural Services Survey of 2000, the first to integrate a parish-based survey with GIS-based calculations of household access to services. The key services used in these calculations are shown in Table 17.1.

All of these lists are supplied with the unit postcode of the service outlet. Where necessary, the Postcode Address File (PAF) was used to attach national grid references to each record (all CACI-sourced records are already geo-referenced). Additional data cleaning, using the Royal Mail's Address Manager to locate individual addresses where postcode details were inaccurate, was necessary to maximise the number of successfully matched postcodes. Address Manager software allows users to enter portions of an address and identifies an accurate matching unit postcode where an initial postcode has been wrongly recorded within the original digital lists.

Table 17.1 *Sources of digital lists of service locations within the Rural Services Survey, 2000*

Service	Data source	% successfully grid referenced (after cleaning)
Primary schools	DfES (Register of Educational Establishments)	100
Secondary schools	DfES (Register of Educational Establishments)	100
Post offices	Postcode Address File	98.8
Benefit offices	Benefits Agency	99.7
Job centres	Employment Service	99.5
Doctors' surgeries	Binley's Register of General Practitioners	99.5
ATMs (cash machines)	LINK Network plc	90.0
Banks and building societies	CACI Ltd	Already referenced
Hospitals	Yellow Pages and Dept. of Health	100
Petrol stations	CACI Ltd	Already referenced
Supermarkets	CACI Ltd	Already referenced

Source: RSS2000 Technical Report.

From Table 17.1, it can be seen that the rate of success in attaching grid references was generally high. Only ATMs posed any significant problems at this stage of the analysis.

Once grid-referenced, software was constructed to identify, for each unit postcode in England, the location of its nearest outlet for each service. This was assessed on the basis of the straight-line distance connecting the centroid of the postcode (using Royal Mail's Address Manager with 1 m centroid accuracy) to the grid reference of the service outlet. Whilst there are some limitations with this methodology over road lengths, straight-line calculations were used as the most practicable way of computing distances for about 1.3 million residential postcodes within England. After each unit postcode had been assigned a distance to its nearest service outlets, two avenues of development were used in the Rural Services Survey of 2000. In the first, the pattern of access was presented visually through plotting individual postcodes as points to generate maps at national and regional scales showing areas of relatively high or low access. An example showing the shortest distance to a hospital for all non-commercial postcodes in England is contained in Plate 7 (taken from Countryside Agency, 2001). Each dot on the map (there are about 1.2 million dots) represents a unit postcode with its colour indicating distance to the nearest hospital.

The real purpose of the map is to permit visualisation of those areas that are relatively remote from hospital services, picked out in dark blue in this example. Key areas in this category include North Cornwall and West Devon, the Welsh borders together with Lincolnshire and East Yorkshire. In addition, the sparsely settled highland areas stand out as blank, focusing attention towards those areas that are both relatively remote *and* relatively intensively settled.

In the second avenue of development, the fact that the number of households within each unit postcode can be estimated from its counts of residential delivery points was used to estimate, for standard administrative units (wards through to counties, regions and England as a whole), the proportions of households within each area having specified levels of access to key services.

An example of one of the products of using this approach is shown in Table 17.2. Developing the example of access to post offices used earlier in the chapter, this shows the breakdown by Government Office Regions of the proportions of rural households living within specified distances from their nearest post office.

The households used in this calculation are all those living within wards defined as 'rural' by the Countryside Agency. The results further confirm the arguments made earlier in the chapter that considering provision of facilities at a parish scale does not provide a meaningful view of the degree to which the great majority of households within rural areas live quite close to their nearest post office.

In these examples, using GIS linked to the increasing availability of machine-readable lists of services that can be, or already are, geo-referenced, combined with an ability to locate unit postcodes and their constituent households, has greatly improved our ability to address the three key issues of estimating access to services outlined earlier. It has provided a direct and generally cost-effective route to determining where services are located, where households are located and of estimating the distances separating the two. It has achieved this with comprehensive national

Table 17.2 *Proportions of rural households living within a specified distance from their nearest post office*

Government Office Region	0–2 km (%)	2–4 km (%)	>4 km (%)
East	92.3	7.0	0.7
East Midlands	93.8	5.7	0.5
North East	93.2	5.6	1.1
North West	95.0	4.4	0.6
South East and London	93.4	6.3	0.3
South West	93.9	5.8	0.4
West Midlands	91.9	7.7	0.4
Yorkshire & Humber	94.4	5.0	0.6
England	93.5	6.0	0.5

Source: Rural Services in 2000 (Countryside Agency, 2001).

coverage eliminating the problems of non-response to surveys and the problems of limited geographical coverage omitting services contained in the larger urban areas. The approach outlined here parallels that used in the construction of the access to services domain within the Indices of Multiple Deprivation (IDM) 2000, defined over a smaller range of services, further confirming the degree to which GIS methods are becoming central to the construction of monitoring of an increasing number of key Government policies and performance targets.

17.4 The Role of Market Towns

17.4.1 What are Market Towns?

The debate over access to services in rural areas has developed a further dimension over the last two years with the recognition that the bases for efficient and sustainable provision of a wider range of services in the countryside are likely to be found among the small and medium-sized towns. Compared with the need for provision of certain basic services within, or accessible to, villages and hamlets, market towns are perceived as having a wider economic and social role in rural areas, that is 'as a focus for growth in areas which need regeneration, and more generally as service centres and hubs for surrounding hinterland, exploiting their potential as attractive places to live, work and spend leisure time' (DETR, 2000, p. 73).

However, whilst the general perception of a market town may be deeply ingrained in the English psyche – viz. a free-standing, small to medium-sized place with a market square, a reasonable provision of professional, business and retail services serving a rural catchment area and, probably, a tourism function based upon the presence of historic buildings – in reality, the so-called 'market towns' are now a far more varied set of elements in the settlement system. In recent years, the profound changes that have taken place in the rural economy and its demographic and social

structure have been partly driven by and partly mirrored in the functional evolution of 'rural' towns (Newby, 1985; Champion, 1989; Hodge and Monk, 1987; Marsden *et al.*, 1993). Moreover, the direction and pace of that evolution has been different in different parts of the country, depending on such contextual circumstances as levels of regional prosperity, location relative to larger urban places, access to motorways and other major roads and national planning policies (Shepherd and Congdon, 1990; Errington, 1997).

The RWP itself acknowledges the inherent variety suggesting that market towns are 'small rural and coastal towns, many of which serve a rural hinterland whether or not they have ever had traditional rural markets. Some may have grown up around a canal or railway junction or as a coastal resort while continuing to be an important commercial and leisure focus for a rural hinterland' (DETR, 2000, para 7.2.1). However, the first statement of a formal definition is given, with no further rationale, in the comment that 'there are over 1,000 towns in England with populations between 20,000 and 2,000' (ibid.). In particular, there is no hint here of a relationship between population size and the functional 'endowment' serving to help define market towns for policy purposes.

17.4.2 Problems of Estimating Indicators of Market Town Prosperity

Capturing the diversity of market towns in empirical terms is neither easy nor always totally convincing but, in attempting to do so for policy development and monitoring purposes, GIS have an indispensable role to play. In general terms, GIS have two main contributions to make in relation to the study of smaller urban places:

- they facilitate the integration of point and boundary data that make it possible to identify service levels available within and around market towns; and
- they permit the estimation of market town characteristics based upon the linkage of data captured at different geographic scales and for different types of data collection unit.

It is these contributions that we focus on here. A third area of analytical contribution, that of distance analysis, which is important in characterising the location (distance) of market towns relative to larger towns and cities to which they may look for jobs and major services/leisure opportunities and smaller places which they themselves supply with 'lower order' goods and services, is not covered.

In practical terms, and in the context of data specifically for England and Wales, the key problems of conducting empirical analyses of small and medium-sized urban places qua 'market towns' are as follows:

- the key dataset on urban places in England defines such places in terms of contiguous urban land uses (producing 'Urban Areas') and not, as in the case of the USA, for example, in terms of urban functions or the provision of jobs or services to the surrounding area (ONS/GRO, 1997; Office of Management and Budget, 1998);

- the boundaries of Urban Areas, which were last defined in 1991 are, at the time of writing, significantly out of date whilst, in some parts of the country in particular, it was known from other sources that there had been extensive residential development on the outskirts and within the catchment area of market towns (Bibby and Shepherd, 1997);
- those datasets which were available to characterise market towns (principally, though not exclusively, the Census of Population), are reported in spatial units (wards, districts, etc.) that, in terms of geographic scale, are better suited to analytical work on larger urban places rather than small ones and this difficulty is considerably magnified for those places in the lowest range of the market town 'hierarchy', i.e., in the region of 2000 residents;
- problems of the way market towns are defined as urban places and of the 'fit' of data reporting units to this definition make it very difficult, on the basis of quantitative data alone, to distinguish between 'true' local service centres and places with a significant or overwhelming dormitory function; and
- finally, no readily available information exists upon which to estimate the catchment areas linked to each town, a task which is important not only for 'free-standing' places, but to distinguishing or estimate overlapping market town catchments in areas with complex urban settlement patterns (ERM, 2000).

What follows relates entirely to the study of English (and possibly Welsh) market towns. In Scotland, different datasets are available which render certain aspects of studies of rural towns more straightforward.

17.4.3 Market Town Size and Regional Distribution

Although the 20 000 to 2000 population range for the definition of a market town is implicitly accepted as an indicator rather than a strict definition, its existence has nevertheless entered into the vocabulary of policy on market towns and has resulted in the list of places for consideration in strategic analyses of market towns within this range. The RWP gives a template for service availability in the East Midlands which defines 'Larger Market Towns' as having populations between 10 000 and 25 000 and 'Smaller Market Towns' as those with populations between 2000 and 10 000 (DETR, 2000, Table 7.1). However, the Countryside Agency's market town headline indicator of economic health focuses on Urban Areas in the 20 000 to 2000 range. This endemic flexibility in definition and the lack of certainty as to the character of market towns due to data problems, both imply the need for equally flexible (and powerful) approaches to policy analysis of which GIS is one such tool. Analytical flexibility is inherent, for example, in the need for awareness of:

- towns larger than 20 000 population which clearly form 'market town' functions in relation to their catchment area but which, because of their larger size, may be inappropriate for market town policy initiatives;

- places that are around the 2000 population level which, because of lack of services or the nature (lack of) settlement in their periphery, may not have an important catchment area function; and
- places that were (perhaps some way) below the 2000 population level at the time of definition but which nevertheless do serve a local catchment area.

There are two main ways in which small market towns may be omitted from current analyses: places which have grown significantly since the base population data (from the 1991 Census) were derived and places which were not originally defined for technical reasons related to the geographical 'capture' of population data within urban area boundaries. Bishop's Castle in Shropshire is a good example of the latter, a small, even classic market town, which fails to meet the population criterion because a 100 m buffer around the urban area fails to capture a sufficient number of census enumeration districts (ONS/GRO, 1997, p. 3).

As noted above, the identification of a place as a market town depends upon a boundary dataset for urban areas that relies upon a strict land-use definition of urban places. Using these data, there are 1210 market towns defined strictly on the '2000 to 20 000' principle within England. They range from Stow-on-the-Wold (with a 1991 population of 1999) to Thetford, in Norfolk, with a population of 20 058. This range excludes some towns which clearly serve a rural hinterland but which are only moderately larger, e.g., Newton Abbott (23 801), Chippenham (25 961), Sevenoaks (24 489) and Bishop's Stortford (28 403). On the other hand, it includes places which have either an extremely restricted 'catchment area' function (e.g., Westhoughton, population of 23 307, essentially a part of Greater Manchester) or are an appendage of a larger place (e.g., Bessacar, population of 19 907, now predominantly a suburb of Doncaster).[1] For the purpose of most analyses reported here, we have included essentially free-standing urban places in the population range 2000 to 30 000. Note, however, that this now incorporates some evident non-market towns such as Addington, an 'out-county' estate just outside Greater London, but just excludes Bury St Edmunds in Suffolk, a very clear candidate for true market town status.

The distribution by population size of market towns in this range is shown in Table 17.3. Over half the candidate market towns have fewer than 5000 residents and four-fifths (80%) have less than 10 000 population. In view of the dominance of market towns in the lower end of this distribution, there is a strong case, at the policy level, for closer scrutiny of the impact of size on market town type, service provision and catchment area range. At the level of empirical analysis, it further emphasises the need for awareness of the technicalities of definition (especially those, as noted above, in the linkage of census and other areas to urban land use boundaries) and especially any peculiar local instances of the fit of such areas to boundaries. Here, painstaking perusal of GIS mappings becomes *de rigueur*.

The absolute and proportional distributions of the size of market towns by region are shown in Tables 17.4 and 17.5. There are significant differences between regions in the proportional table, which indicates different regional settlement patterns based

[1] Bessacar (and no doubt others like it) is included on a technical point relating to the way discontinuous parts of larger urban areas are defined.

Table 17.3 *The distribution of market towns in England by population size*

Population size	Number	%
Over 20 000	64	5.0
15 000–20 000	80	6.3
10 000–15 000	111	8.7
5 000–10 000	309	24.3
2000–5000	710	55.7
Total	1274	100.0

Table 17.4 *English market towns by region and population size, 1991*

Region	2–5000	5–10 000	10–15 000	15–20 000	>20 000	Total
East of England	138	45	17	12	10	222
East Midlands	95	42	11	9	4	161
North East	41	23	7	5	6	82
North West	42	35	11	4	11	103
South East	147	49	18	23	15	252
South West	105	53	22	12	10	202
West Midlands	57	22	7	5	6	97
Yorks & Humbs	85	40	18	10	2	155
Total	710	309	111	80	64	1274

Table 17.5 *The proportion of market towns by population size and region, 1991*

Region	2–5000 (%)	5–10 000 (%)	10–15 000 (%)	15–20 000 (%)	>20 000 (%)	Total (%)
East of England	62.2	20.3	7.7	5.4	4.5	100.0
East Midlands	59.0	26.1	6.8	5.6	2.5	100.0
North East	50.0	28.0	8.5	6.1	7.3	100.0
North West	40.8	34.0	10.7	3.9	10.7	100.0
South East	58.3	19.4	7.1	9.1	6.0	100.0
South West	52.0	26.2	10.9	5.9	5.0	100.0
West Midlands	58.8	22.7	7.2	5.2	6.2	100.0
Yorks & Humbs	54.8	25.8	11.6	6.5	1.3	100.0
Total	55.7	24.3	8.7	6.3	5.0	100.0

upon different economic circumstances. Such differences have significant implications for market town policy (especially in relation to regeneration and equity issues), the interpretation of indicators and the evaluation of individual market town initiatives. Tables 17.4 and 17.5 indicate a generally even distribution of market towns by population size across regions, with the exception of the North West which is under-represented in the category of towns of 2–5000 and over-represented by towns with

more than 20 000 population. However, this may have as much to do with the underlying definition of a town using land use criteria as with functional attributes and town–catchment area interactions.

The analyses that follow are the first stages in an attempt to identify some order in this group of settlements as a step towards identifying effective indicators of their condition and of identifying towns which are of particular importance to a rural hinterland. It is essentially experimental work that is trying to close the gap between policy and understanding.

17.5 Using GIS to Estimate Service Provision in Market Towns

The fact that the Urban Area boundaries and associated populations defining market towns were 10 years out of date at the time of the analyses was a cause for some concern. As previous studies had shown, small and medium-sized towns were the fastest growing places between 1971 and 1981 and the forces that had generated that growth were unlikely to have abated by much, if at all, in the subsequent decade (Champion, 1989; Shepherd and Congdon, 1990). The problem in gaining some idea of the rate of change among this particular set of towns lay in the scale difference between the areas for reporting inter-censal population estimates (i.e., wards with populations assigned from pro-rating local authority district estimates), and the smaller market towns which formed the majority of the overall size distribution of these places. One simple way to obtain population change information would have been to assign wards to market town urban areas on the basis of a boundary intersection/coincidence rule, for example, incorporating those wards which had any part of their boundary shared with or crossing that of an urban area. In practice, however, intersecting ward boundaries are not only a very poor fit to Urban Areas, but can also lead to complex overlapping of ward boundaries in locations where there are geographic clusters of the smallest market towns, thus making it difficult to allocate population growth to particular places.

An illustration of some of these difficulties can be seen in Figure 17.1. This shows both urban areas and ward boundaries centred on the Braintree urban area in Essex that in 1991 had a resident population of just over 30 000 people. Even at this scale, right at the top of the market town range, Census wards divide the town unevenly and often contain significant areas outside the built up area. In addition, the detailed rules by which the 1991 Urban Area boundaries were fixed often lead to rather ragged boundaries, particularly where urban sprawl has taken place. This can be clearly seen in the ward to the north of the centre of Braintree where a straggle of urban development protrudes into what would otherwise be an undeveloped ward. On the west side of the map another case can be seen where a smaller urban centre is divided into two more or less equal parts by the boundaries of two large rural wards.

Our initial approach, therefore, was to focus on the immediate environs of market towns by defining a consistent 500 metre buffer for all towns in the dataset and to use this to capture residential delivery points (RDPs) from Royal Mail's Address Manager database. This procedure had been used in the development of the Rural

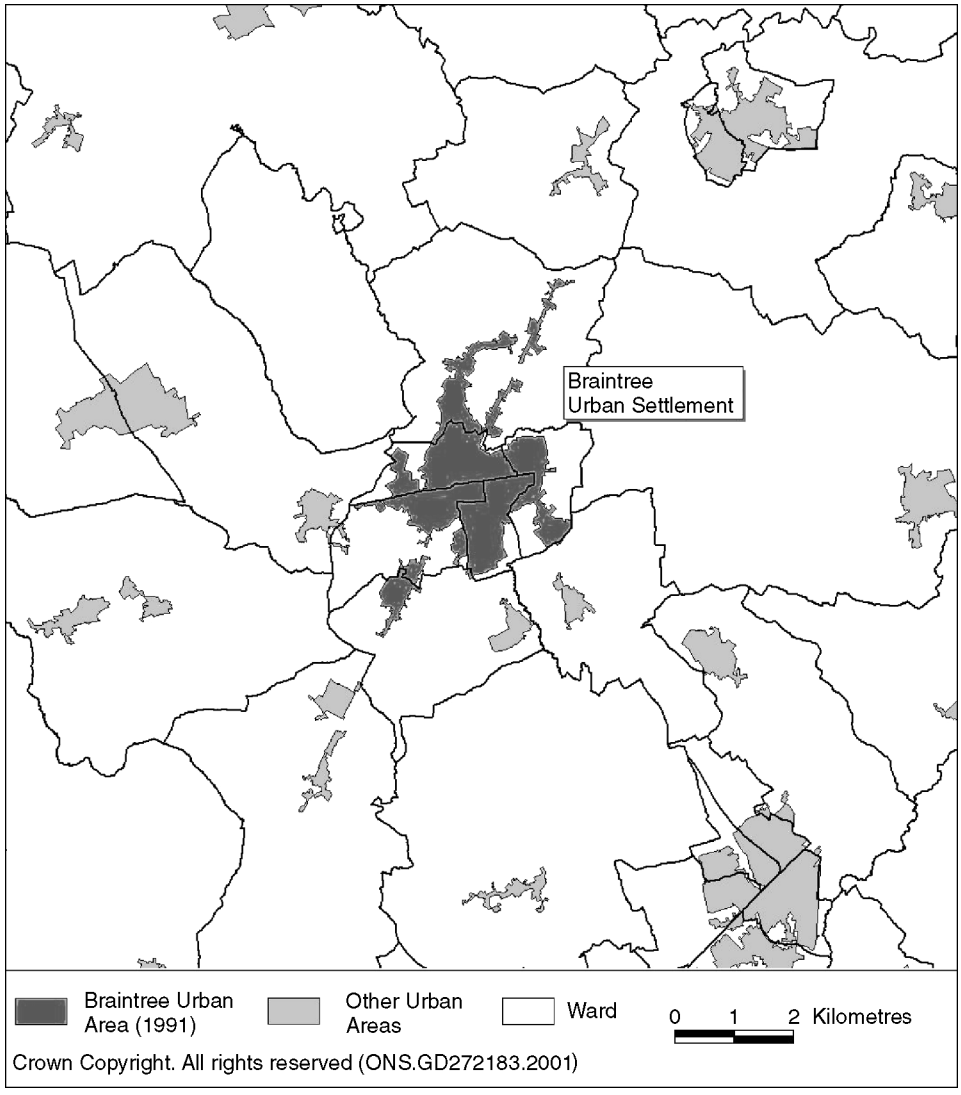


Figure 17.1 *A large market town and 1998 ward boundaries. From Countryside Agency. © Crown Copyright*

Settlement Gazetteer (RDC/HC, 1998) and in earlier studies using postcoded information to track population change (Raper *et al.*, 1992, pp. 220–33). In this case, however, there was no attempt to transform RDPs into population on the basis of regional or local differences in household size since, again, information on the latter, being derived from the 1991 Census was also likely to be inaccurate. In addition, there was no attempt, though it could be done, of differentiating between total household

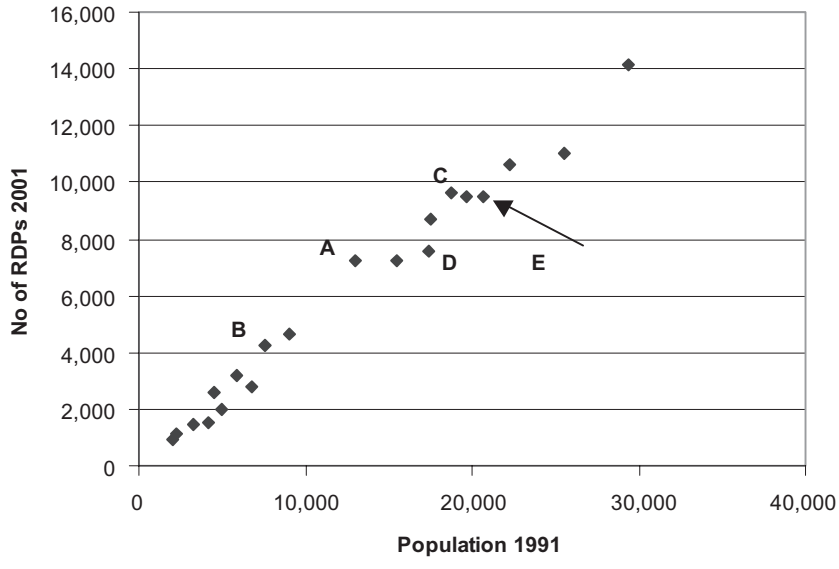


Figure 17.2 *Selected market towns; population 1991 (Urban Area) and number of Residential Delivery Points 2001 (Urban Area plus buffer)*

change (i.e., RDPs within Urban Area + buffer) and purely urban edge change (i.e., RDPs within the buffer).

Figure 17.2 shows a close relationship between population in 1991 and the number of RDPs in 2001. Of potentially more interest here, however, are the apparent departures from a 'straight-line' relationship reflecting, in some part at least, differences between market towns in levels of household growth in the period 1991–2001. These departures emphasise the importance of the economic structures of market towns and their regional contexts in accounting for population change. Stronger than expected growth can be seen in some of the small and medium towns in the south (e.g., East Dereham (A), Norfolk; Sherborne (B), Dorset; Spalding (C), Lincs) while those falling short of expectations are predominantly northern towns with a strong industrial or mining base (e.g., Spennymoor (D), Durham; Workington (E), Cumbria).

The first test rests on the well-established concept of urban hierarchy. We would expect certain medium to high level key services to be located in some of the larger towns in this group. The services selected from those where digital lists were available were banks (and building societies), secondary schools, supermarkets, large DIY outlets, Benefit Offices and Job Centres. To allow for fringe growth around the DETR urban boundaries an additional 500 metre buffer was added to each settlement.

These results demonstrate the considerable size ranges over which these services are found (Table 17.6). Only Benefit Offices and principal DIY outlets seem to have any power as threshold indicators of key centres and even for these it is limited. Most services are found throughout the size hierarchy, although with a tendency to be more completely and intensively present in the larger settlements.

Table 17.6 *Presence and absence of selected services*

	Largest town without service	Smallest town with service
Bank	15514	2000
Secondary school	18336	2012
Supermarket	15514	2052
DIY chain	19907	4591
Benefit Office	19907	6466
Job Centre	19907	2245

Sources: DfEE, CACI, Countryside Agency.

Given this range of occurrence, the initial approach to identifying the more important local service centres was based on overall relationships between population size and the number of service outlets found in each town rather than pursuing the concept of population thresholds. For a selection of important local services (secondary schools, banks and building societies, and supermarkets), the number of services present in each town was predicted using simple regression equations based on population size. The results were as follows:

$$Y_{bi} = -0.32^{**} + 0.00042^{**}P_i + e_{bi} \ (R^2 = 0.58) \text{ for banks} \tag{17.1}$$

$$Y_{si} = 0.02 + 0.00018^{**}P_i + e_{si} \ (R^2 = 0.59) \text{ for supermarkets} \tag{17.2}$$

$$Y_{di} = 0.45^{**} + 0.00013^{**}P_i + e_{di} \ (R^2 = 0.41) \text{ for doctors' surgeries} \tag{17.3}$$

where Y = count of services present, P = resident population in 1991, i = 1:1272 market towns and ** represents $p < 0.01$.

The residuals from these equations were then summed across all three services to derive an indicator of consistent over- or under-endowment of services. This is then divided by the population of each town to assess the intensity of these deviations.

$$I_i = \sum e_{ti} / P_I \tag{17.4}$$

where I = the intensity indicator and t = 1,3 service categories.

The purpose of this approach is to identify towns where there are significantly more services than would be expected from their local populations, suggesting a broader service centre specialisation. Table 17.7 shows the 12 towns with the strongest signs of positive specialisation in service provision.

Most of these towns appear as reasonably free-standing local service centres. The exception is Hemingford Grey which is part of a cluster of Urban Areas around Huntingdon/St Ives. Overall these suggest that there might be some path here to identifying the elusive market towns. This view is reinforced by considering the 12 towns with the lowest scores (Table 17.8). Many of these towns are either close to, or part of, larger urban or suburban settlements while some like Hatfield and New

Table 17.7 *Towns with strong service centre functions*

Town	Over-representation score
Stamford	2.30
Uttoxeter	2.41
Evesham	2.54
Tiverton	2.73
March	2.75
Tewkesbury	2.76
St Austell	3.16
Beverley	3.41
Diss	3.99
Wetherby	3.99
Hemingford Grey	4.03
Windermere	4.25

Table 17.8 *Towns with weak service centre functions*

Town	Under-representation score	Comments
New Addington	–3.84	near Croydon
Maghull	–3.74	outskirts of Liverpool
Hatfield (Yorks)	–3.55	close to Doncaster
Westhoughton	–3.08	close to Wigan
Golborne	–2.88	close to St Helens
Borehamwood	–2.75	fringes of London
Hoylake	–2.56	close to Birkenhead
Winsford	–2.49	more difficult to explain
Nailsea	–2.35	close to Bristol
New Rossington	–2.28	close to Doncaster
Trowbridge	–2.17	fairly close to Bath
Aylesford	–1.94	close to Maidstone

Rossington were developed partly as colliery settlements in addition to being close to a regional service centre.

Finally, we use this approach to examine the 'service level endowment' scores for a selection of classic free-standing small and medium-sized market towns. These are shown in Table 17.9.

It will be noticed that few of these scores are particularly strong but all are positive. Of the 1275 market towns defined in simple population terms, 735 show negative scores on this indicator, so to achieve 10 positive scores from the selection of towns is an interesting result. It confirms a little further that this approach has potential value in sifting the mixed group of towns that are now the focus of considerable policy effort.

Table 17.9 *Scores for free-standing towns*

Town	Town scores
Ely	0.07
Framlingham	0.08
Harleston	0.48
Bicester	0.24
Malton	1.51
Market Drayton	0.65
Bourne	0.21
Horncastle	0.56
Market Rasen	0.40
Thirsk	1.24

17.6 Conclusions

This chapter has demonstrated that relatively simple analytical approaches involving standard GIS procedures such as nearest neighbour measures and point in polygon allocations can, in combination with digital lists of services and households, be used to underpin the development and implementation of rural policies in England. At the moment, this work is in its early stages of development and will continue to be developed in coming years to provide a firmer base from which to specify and operate plans for greater rural support. At this stage, possibly the most interesting conclusions are not contained in the work itself but relate more to its timing.

In spite of public agencies widely proclaiming the virtues of evidence-led policy development, the political momentum of measuring access to services and, particularly, the use of the market town label preceded any systematic and thoughtful investigation of the concepts underlying the terms. All of the evidence points to the fact that little initial thought was given to how these terms might be effectively operationalised to produce measurements that would provide policy development with meaningful support.

In this setting, the application of GIS methods has provided a flexible way of responding to an empirical need generated by the speed of policy announcements and development of political agendas outstripping the capacity of public agencies to explore and comprehend relevant conceptual frameworks. We would argue that there is a risk in this flexibility. The power of modern software, linked to the increasing availability of digitally encoded information, can easily come to be seen as offering a quick fix to any empirical need thrown up in the policy development process. On the evidence of current practice, not all of these needs will be amenable to sensible analysis. The chaotic conception of market towns comes very close to this in bundling together towns with dissimilar functions, characters and histories into a common grouping linked only by their population size. If only some geographical analyses for exploring the nature of the current settlement and service-based urban hierarchy had

been undertaken before, rather than after the term was coined, things might have been different – and better!

Acknowledgement

The research contained in this chapter was conducted by the South East Regional Research Laboratory, Birkbeck College London under contract to the Countryside Agency. The views expressed are those of the authors alone.

References

Bibby, P. and Shepherd, J. (1997) Projecting rates of urbanisation in England, 1991–2016: method, policy application and results, Town Planning Review, **68**(1): 93–124.

General Register Office (Scotland) (1995) *Census Key Statistics for Localities in Scotland*, HMSO, Edinburgh.

Champion, A.G. (1989) *Counterurbanisation in Europe Vol 1: Counterurbanisation in Britain*, Royal Geographical Society, London.

Countryside Agency (2001) *Rural Services in 2000*, Countryside Agency, Cheltenham.

Department of the Environment, Transport and the Regions (2000) *Our Countryside: The Future*, Cm 4909, HMSO, London.

Environmental Resources Management (ERM) (2000) *Indicators for Assessing the Character and Vulnerability of Market Towns in North West England*, for the Countryside Agency.

Errington, A.J. (1997) Rural employment issues in the peri-urban fringe, in Bolman, R. and Bryden, J. (eds) *Rural Employment an International Perspective*, CAB International.

Hodge, I.D. and Monk, S. (1987) Manufacturing employment change within rural areas, *Journal of Rural Studies*, **3**: 65–69.

Marsden, T. J., Murdoch, P., Lowe, M., Munton R. and Flynn, A. (1993) *Constructing the Countryside*, UCL Press, London.

Moseley, M.J. (1979) *Accessibility: the Rural Challenge*, Methuen, London.

Newby, H. (1985) *Green and Pleasant Land? Social Change in Rural England*, Wildwood House, London.

Office of Management and Budget (1998) Alternative approaches to defining metropolitan and non-metropolitan areas, Federal Register, **63**(244), December 21.

Office for National Statistics website (www.statistics.gov.uk).

Office for National Statistics/General Register Office Scotland (1997) *1991 Census, Key Statistics for Urban and Rural Areas*, HMSO, London.

Raper, J., Rhind, D. and Shepherd, J. (1992) *Postcodes the New Geography*, Longman, London.

Rural Development Commission/Housing Corporation (1997) The Rural Settlements Gazetteer, *Research Paper 26*, RDC, London.

Rural Development Commission (1992) *1991 Survey of Rural Services*, RDC, Salisbury.

Rural Development Commission (1995) *1994 Survey of Rural Services*, RDC, Salisbury.

Rural Development Commission (1998) *1997 Survey of Rural Services*, RDC London, Salisbury.

Scottish Executive, Ministerial Committee on Rural Development (2000) *Rural Definitions*.

Shepherd, J. and Congdon, P. (1990) *Small Town England: Population Change Among Small and Medium Sized Urban Areas 1971–1981*, Progress in Planning, **33**(1).

18

Forecasting River Stage with Artificial Neural Networks

Pauline Kneale and Linda See

Abstract

Forecasting models are required to determine continuous river stage for drought and water resource management, and for real-time forecasting of extreme flood events. The models use telemetred data from raingauges and river stage gauges located at key points in the catchment. Stage forecasts are traditionally made using physical models, but these large-scale systems have not proved to be accurate enough to provide timely flood warnings. Artificial neural networks (ANNs) provide an alternative modelling technology for building flood forecasting models. Artificial neural network software has been developed and installed for operational use by the Environment Agency in Northumbria, UK. This chapter describes its use in practice and compares the results using two feedforward neural network algorithms. Results from three experiments on three different sizes of catchment are compared. Overall the results suggest that ANNs are useful for operational purposes and represent an improvement over ARMA and regression modelling schemes.

18.1 Introduction

There is a continuing risk of flooding in the UK from rivers and from the sea. Some five million people live in two million properties that are situated in flood risk areas in England and Wales. Although there are over 36 000 kilometres of flood defences, the risk of flooding is too low in some areas for further physical defences to be considered cost effective. Therefore, timely flood warnings are required to give businesses and residents the opportunity to protect themselves and their properties. This

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

requires real-time forecasting models to be in place which are capable of providing accurate river stage forecasts. Stage is the depth of flow above the river bed which is then used to calculate flow rates and volumes. The UK Environment Agency is responsible for providing flood warning services on main river channels, and they, in conjunction with DEFRA and local and regional authorities, are responsible for giving advice to planners.

Forecasting river stage has proved problematic. There are many examples of the continuing development of statistical and deterministic models but accuracy and reliability remain issues in a forecasting area where the events of primary interest, i.e., extreme floods, are infrequent and non-linear (Cameron *et al.*, 2000; Horritt and Bates, 2001; Mellor *et al.*, 2000; Stewart *et al.*, 1999). Artificial neural networks (ANNs) provide a potentially useful alternative to traditional flood forecasting approaches with rapid development times relative to large-scale physical models. ANNs are biologically inspired computational models that map a set of inputs, in this case rainfall and river stage data, to a set of outputs, such as river stage at downstream stations. The network achieves this task by learning the patterns in an historical dataset through a training or calibration process. Once trained the neural network can be utilised in a predictive capacity with unseen, independent input data and tested for accuracy using independent validation data. The results should be reliable provided the training dataset is representative of all the relationships being modelled, which is always an issue with flood forecasting. The largest flood event may be the next one and occur tomorrow.

There are many opportunities for neural network use in hydrology, and its application is becoming increasingly widespread (Dawson and Wilby, 2001). River stage forecasting is just one hydrological situation where ANNs are being adopted, as exemplified by Abrahart *et al.* (2001), Campolo *et al.* (1999), Chang and Hwang (1999), Dawson and Wilby (1998), Golob *et al.* (1998), Openshaw *et al.* (1998) and See and Openshaw (1999, 2000). In an early study, Hsu *et al.* (1995) showed that the non-linear ANN model provided a better representation of the rainfall–runoff relationship of the medium size Leaf River basin near Collins, Mississippi than the linear ARMAX (AutoRegression Moving Average with eXogenous inputs) time-series approach or the conceptual SAC-SMA (Sacramento Soil Moisture Accounting) model. They acknowledged one key difference in modelling philosophy; ANNs do not provide models that have physically realistic components and parameters and so are no substitute for conceptual watershed modelling. But they point out that ANNs provide a viable and effective alternative to the ARMAX time-series approach for developing input–output simulation and forecasting models in situations that do not require modelling of the internal structure of the watershed. In forecasting river stage there is no absolute requirement for physically realistic variables or mirroring of the catchment structure. Thus, ANNs are particularly useful as forecasting tools in situations where the relationships between the key variables are non-linear or cannot be adequately described by dynamic equations, or where the data are limited and more complex process models cannot be adequately calibrated.

Elsewhere in hydrology ANNs have been used to forecast water quality: phosphorus levels in 927 catchments across the USA (Omernik, 1997), river pH (Moatar *et al.*, 1999), Cryptosporidium and Giardia concentrations (Neelakantan *et al.*, 2001) and PAHs in water samples (Cancilla and Fang, 1996). Other examples include the

prediction of fish numbers and habitat preferences in rivers (Lek and Baran, 1997; Mastrorillo *et al.*, 1997; Gozlan *et al.*, 1999) and algal numbers by Maier *et al.* (1998), Recknagel *et al*. (1997) and Whitehead *et al.* (1997).

While the range of hydrological applications using ANNs is encouraging, there is a pressing need to trial models cited in the literature, many of which are implementations in experimental or historical mode, to operational real-time settings. This work represents one of the first attempts at making this technology transfer. This chapter describes the development and application of operational neural network flood forecasting software, which was undertaken in collaboration with the Northumbria office of the UK Environment Agency (EA). Neural network theory and a description of the software tools are provided. Three case studies illustrating the application of the software highlight the ability of the networks to forecast events from limited records and beyond the range of the training data. Models were built using data from a seven-month winter period and tested using data from three further periods of continuous river stage data. The results were also compared with more conventional statistical models, multiple linear regression and autoregressive moving average models (ARMA).

18.2 Client Requirements

River stage forecasting, whether for engineering design or flood warning purposes, is one element in a wider hydrological modelling framework (Figure 18.1). A forecasting model sits between the data sources and a decision-making process. Although hydrological modelling has been developing since the 1920s, and numerical modelling has brought major advances since the 1960s, the accuracy of results is not yet satisfactory. Flood forecasting systems in current use are generally based on either distributed modelling ideas or on mass transport equations, which essentially move blocks of water down the river (HEC-RAS, 2001; ISIS, 2001; MIKE 11, 2001; NOAH 1D, 2001; RFFS, 2001; SSIIM, 1997; van Kalken *et al.*, 2001). The complexity of natural systems and the number of processes that continuously interact to influence river levels make traditional modelling based on mirroring natural processes with process-based equations very difficult. Where processes can be modelled numerically, there are constraints imposed by the spatial availability of data and the cost of collecting this data in real time.

Any modelling system is a considerable financial issue for an agency or company. Investment is required in the numerical model, the acquisition of historic and real-time input data and trained analysts to interpret the results. For real-time forecasting this involves a team of analysts available 24 hours a day, to interpret results and implement flood-warning procedures. Once a forecasting system is established within an organisation, moving to alternative systems will only happen if there are seen to be clear advantages, preferably in both financial and forecast accuracy. The cost of moving to a new system and the different modelling approach of ANNs are barriers to the acceptance of the technology.

At present the EA uses a number of conventional hydrological models. In a report, Marshall (1996) stated that 'the development/calibration of the models has, with few exceptions, not yet reached the stage where warnings can be reliably based on the

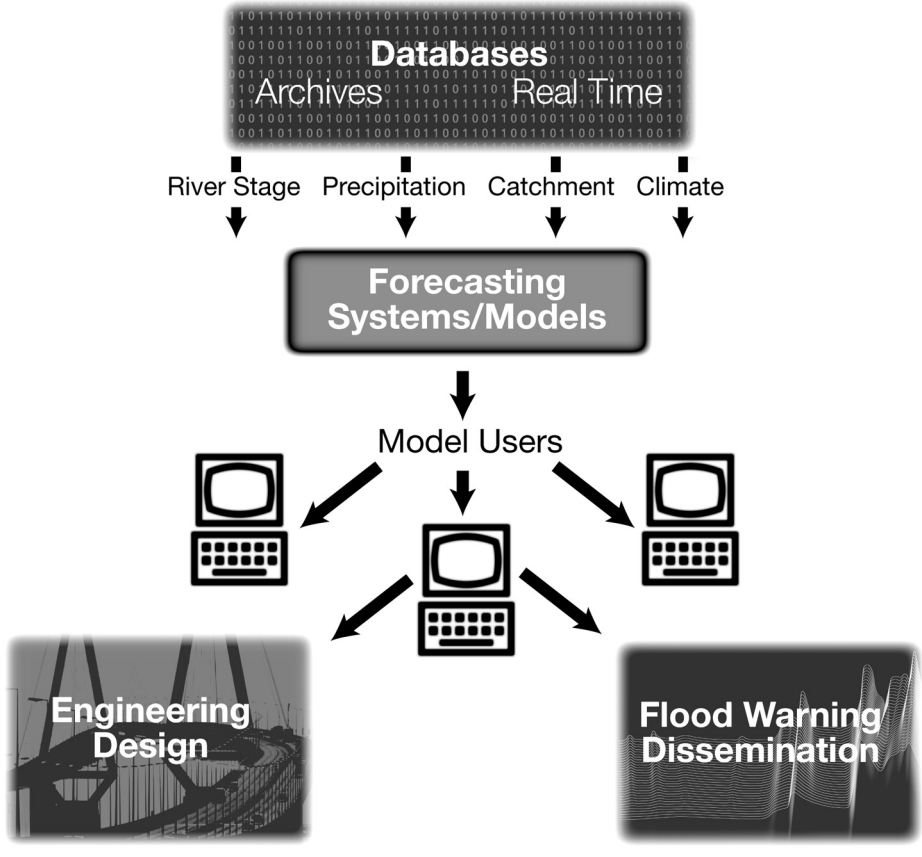


Figure 18.1 *Generalised framework for hydrological forecasting*

forecasts produced. The systems provide considerable help to duty officers by collecting, processing and displaying telemetry and radar data on a region-wide basis, but their operational performance as forecasting systems cannot yet be measured' (p. 7). Although none of the models in the regional offices were considered fit for operational purposes they are still in use at present. This leaves a gap in the market for alternative technologies that have the potential to provide affordable, accurate forecasting that is operationally straightforward. The neural network software developed for the EA represents an initial attempt to fill this market gap.

18.3 Methodology and Software Tools

18.3.1 Artificial Neural Networks

The basics of ANNs are discussed in detail by many authors (see e.g., Bishop, 1995; Gurney, 1997; Haykin, 1998); only a brief description is therefore provided here.

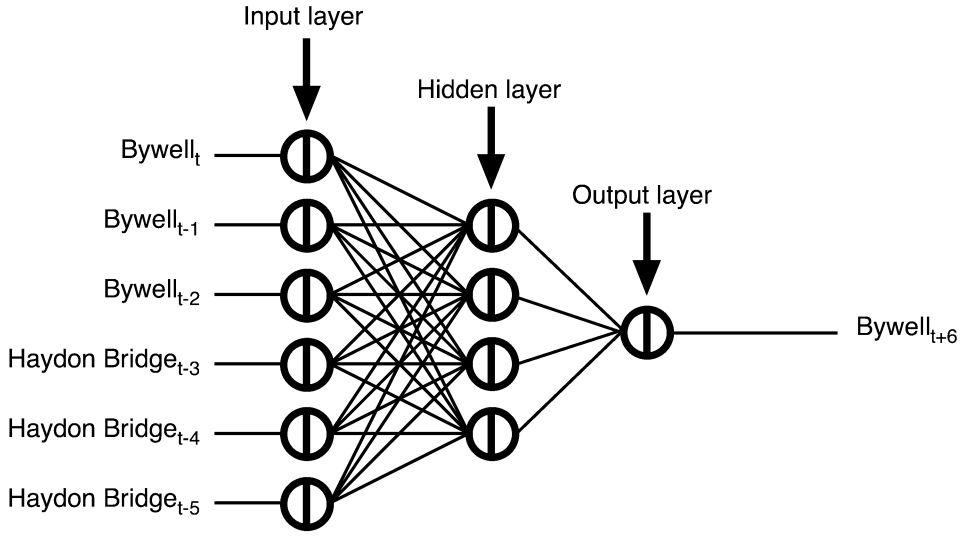


Figure 18.2 *Basic structure of a neural network*

ANNs perform a mapping of inputs to outputs via a series of simple processing nodes or neurons. The ability to map a function is derived via the configuration of these neurons into a set of weighted, interconnected layers as shown diagrammatically in Figure 18.2, where the neurons in the first and last layers have a one-to-one correspondence to the input and output values, respectively. Bywell and Haydon Bridge are two of the stage gauging stations on the River Tyne. Bywell is the lowest gauge on the non-tidal river where a forecast is required, and Haydon Bridge is upstream on the tributary River South Tyne (Figure 18.3). In this neural network mapping, the inputs are river stage at the two stations at the current time t, and t_{-1}, an hour before, and t_{-2}, two hours before. The inputs can be any combination of variables that are thought to be important for predicting the output for a specific case, in this case river levels at the forecast and upstream stations and rainfall would be appropriate. Thus, ANNs are capable of incorporating a large number of variables in a flexible manner. The input layer is connected to the hidden layer, which is further connected to the output layer where the forecast is produced. It is the programme-defined weights that are updated in the network learning phase (Beale and Jackson, 1990).

The task of any neuron, k, is to integrate the n inputs, x_i, from an external source or from other neurons within the network as follows:

$$S_k = \sum_{i=1}^{n} w_{ik} x_i \tag{18.1}$$

where w_{ik} are the weights between neurons i and k, and S_k is the sum of the weighted inputs. The neuron then further processes this sum via an activation function, such as the sigmoid:

$$f(x) = \frac{1}{1 + e^{-S_k}} \tag{18.2}$$

To learn the relationships in the data the network is trained by iteratively adjusting the network weights in such a way that the network converges to an acceptable solution. Developed by Rumelhart and McClelland (1986), backpropagation is the first reliable algorithm for achieving this task and its development represented a major historical breakthrough in neural network research.

Backpropagation is a variation of the gradient descent optimisation algorithm. It is also known as the generalised delta rule for weight updating and is given by the following equation:

$$\Delta w = -\eta g \tag{18.3}$$

where η is a learning rate parameter greater than zero. The gradient, g, is calculated in two passes through the network. In the forward pass, the network output for pattern p is calculated. The error between the calculated output, x_k, and the desired value, z_k, is defined as:

$$E_p = \sum_k (x_k - z_k)^2 \tag{18.4}$$

In the backward pass, the partial error associated with the pattern is backpropagated through the network to update the weights. Weight updating can take place after each pattern or after each training cycle in which all the patterns are presented to the network. The network is trained until a stopping criterion has been satisfied such as an increase in error in the validation data set. When this situation occurs, further training will result in an overtrained network and an inability to generalise to independent data. The backpropagation training algorithm can be slow to converge so a momentum term can be added to the weight update equation to speed up the training process (Bishop, 1995).

A second, faster training algorithm incorporated into the forecasting software is Quickprop, developed by Fahlman (1988). In contrast to backpropagation, Quickprop is a second-order method because it uses information provided by the second derivative of the error surface to take larger steps down the gradient of the error function. The method assumes that the shape of the error surface is parabolic and it tries to move from the current position to the minimum of the parabola at each step. Quickprop computes the derivatives in the direction of each weight; firstly, the gradient is computed with regular backpropagation and then a direct step to the error minimum is attempted. The weights are updated as follows:

$$\Delta w(t) = \frac{g(t)}{g(t-1) - g(t)} \Delta w(t-1) \tag{18.5}$$

where $g(t)$ is the partial derivative of the error function at the current time step and $g(t - 1)$ is the last partial derivative.

Despite significant computational and methodological advances, there are fundamental problems involved in the selection of an optimal neural network architecture. As Abrahart *et al.* (1999) show, the choice of algorithm is a subject of lively debate. To simplify the choice of algorithms, only backpropagation and Quickprop are currently incorporated into the flood forecasting software developed for the EA, as outlined in the next section. The networks are also provided with default settings so that the users can apply the software with minimal knowledge of ANNs. In the future, additional algorithms will be built into the software.

18.3.2 Description of the Neural Network Software

The PC-based neural network model was developed for operational flood forecasting purposes by the University of Leeds and implemented by the Northumbria Environment Agency's flood defence and water resources department in December 1999 (Kneale and See, 1999; Kneale *et al.*, 2000a and b).

The software was designed so that the EA operators can develop and maintain their neural network models independently. It is written in Visual BASIC v.6 as a standalone program. The program can either use historical or real-time rainfall and stage data to forecast stage, and was programmed so that the EA staff could use it with data for any of the main rivers in the Northumbria region. The interface is designed to be user friendly. It contains a series of interactive lists that guide the operator through each stage of the forecasting procedure. There are three main components: an interface for specifying the data, a neural network component and a graphing feature.

The Data Component

The data interface provides a diagrammatic river network which permits the user to specify the forecasting station, the forecasting time horizon, station inputs to the network and the location of data files. A list of steps is provided on the right-hand side of the screen to guide the user through the process. Once the input data and files are specified, the user can build a training data set for the network. The following functions are available within the data component:

- *Data pre-processing*: There are two processes: (a) creation of hourly averages from the 15-minute stage data available from the historic records as used here, but some users might choose (b) first-order differencing of the data series prior to building the training data set in order to render it stationary.
- *Station selection and forecasting horizon*: The forecasting station is selected along with the forecasting horizon, ranging from 1 to 24 hours.
- *Build the training data set*: This function builds the training dataset from each of the station inputs specified by the user.
- *Travel time specification*: This allows the user to change the average travel times between stations.

The Neural Network Component

The neural network component allows the user to set up a neural network, choose an algorithm and specify training parameters. The trained network can be saved to a file or loaded and used to make forecasts. The functions available within the neural network component include the ability to:

- *Configure the network*: Specify the number of inputs and hidden nodes. Set training parameters if defaults are not desired. Choose one of two available network training algorithms: backpropagation or Quickprop.
- *Specify data set*: Indicate the name of the training or testing data set.
- *Select mode*: Real-time operational forecasting or training on historic data.
- *Train the network*: Train the network for the number of iterations specified by the user.
- *View the network performance*: Examine the training or testing statistics.
- *Save/load a network*: Save the trained net or load an existing one.
- *Forecast stage*: Use a trained network to forecast the stage at the user specified station.

As mentioned previously, the software has two neural network algorithms: standard backpropagation (Beale and Jackson, 1990) and Quickprop (Fahlman, 1988). In practice a researcher exploring the use of neural nets will look at alternative algorithms and choose the one which is most efficient for each investigation. Including two algorithms gives users the opportunity to experiment with different solutions and explore relative performance accuracy.

The Graphing Component

The graphing component provides the user with an interface to graph any input data set or to view network forecasts on a historical or real-time basis. The results required for hydrologists are a graph of the actual and forecast stage in metres on a common time frame. In some hydrological modelling the forecasting procedure checks the actual stage against the forecast value and then calculates the next step ahead based on the updated, real stage. In ANN modelling there is no step ahead updating, the forecast is continuous.

18.3.3 Use of the Software

There are two main ways to use the software. In the first mode, a single model is created with a set forecasting horizon. For example, networks can be built to predict river stage from 1 hour ahead up to 24 hours ahead. Predictions are always made from t_0 to produce a single forecast x hours ahead, where x is the forecasting horizon chosen by the user. To produce a continuous set of forecasts at three-hour intervals, for example, the user would need to build eight networks with increasing forecasting horizons. The second use is the more conventional forecasting model in which

a single network is created with a short forecasting horizon, such as 1 hour, and stage and rainfall information are used to produce continuous forecasts over the next 24 hours.

It is important to recognise that a neural network must be developed for each specified forecasting station, so errors are not propagated downstream as occurs in a flow routing model. The programmes do not use any of the forecasted information as inputs to create additional forecasts. All the graphs in this chapter are validation forecasts created using independent data, and there is no step updating and correction as one would expect in a conventional hydrological forecasting model. In real-time operation the model can be run to make a new forecast every 15 minutes or whenever further real-time stage or rainfall data become available. The neural network forecasting model runs in seconds and can be retrained with new data in minutes. The retraining time is governed entirely by the time taken to access the real-time input data. It has the potential to be a very useful operational forecasting tool.

18.4 Case Studies

18.4.1 Background

The Tyne catchment in Northumberland, north east England, UK, has two main channels: the North and South Tyne. The River Nent and the East and West River Allens are important, ungauged lateral inflows (Figure 18.3). The underlying geology of the catchment largely consists of Carboniferous limestone and Millstone Grit. The Pennine upper reaches respond very quickly to rainfall inputs, the time between rainfall and rising river stage can be minutes. There is an approximate lag time of 4 hours between the raingauge at Alston and the river gauging station at Haydon Bridge. Flood alerts are often issued to the flood-risk areas along the river. Bywell is the lowest non-tidal gauging station on the River Tyne.

In these examples forecasts of stage are made for flood events at Alston, Haydon Bridge and at Bywell. Two types of forecasting are undertaken: event-based and continuous. Event-based forecasting involves extracting flood events from the historical record and using these small datasets to develop neural network models. Continuous forecasting uses a dataset over a continuous time period such as the winter period. Alston and Bywell are used to assess the value of ANNs for event-based forecasting using different forecasting conditions. Alston is a good example of a flashy headwater catchment where runoff starts at almost the same time as it rains. Bywell is a lowland station downstream from the North and South Tyne confluence. It is critical in providing flood warnings for Newcastle upon Tyne. In the experiment using Bywell, stage is forecast using only data from stations on the River South Tyne. Haydon Bridge is used for long-term forecasting based on continuous input data rather than individual events from the training record.

For the event forecasts the digital river stage archives (1985–2000) were searched to find the highest stage events at the input stations and Bywell. For the Alston experiments the input data were stage at Alston and rainfall at the Alston raingauge. For the Bywell forecasts, data from Haydon Bridge, Featherstone, Alston and the

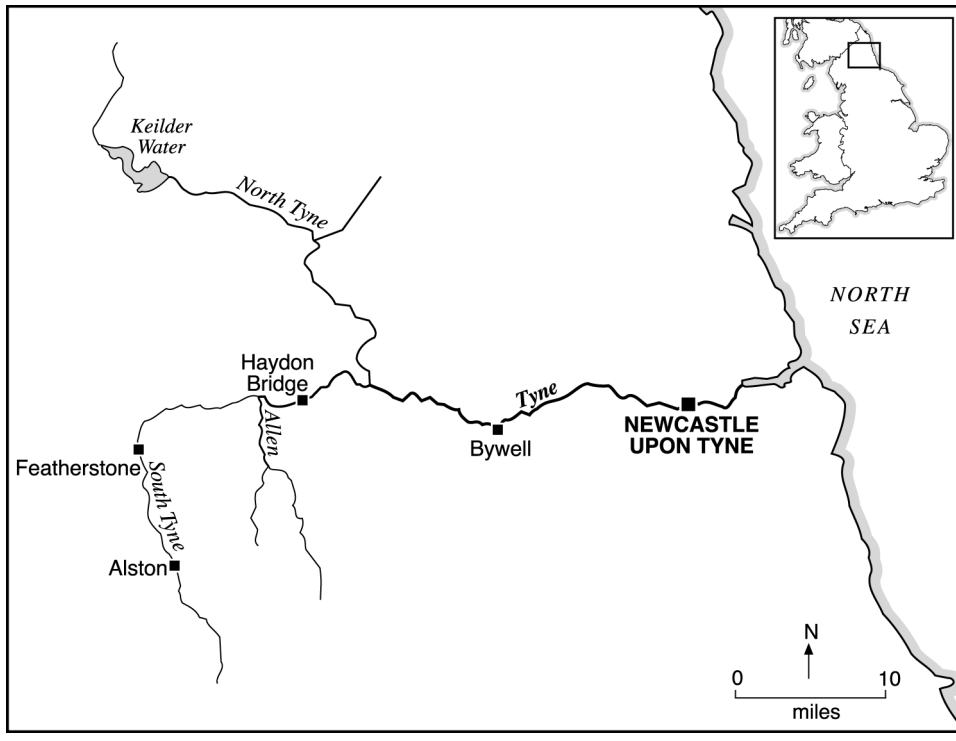


Figure 18.3 *Map of the catchment*

raingauge at Alston were used. In each case the 11 largest events were ranked. The odd-numbered events were used as the training series and the even-numbered events used for independent model validation. Networks were generated using both the backpropagation and Quickprop methods and for various lead times and training epochs (number of training iterations). The limited size of the event training data sets was deliberate to see how well the forecasts were made with small datasets.

For the long-term continuous forecasts at Haydon Bridge, the input data were stage at Haydon Bridge, Featherstone, Alston and the raingauge at Alston. Again backpropagation and Quickprop routines were compared.

Evaluation of the accuracy of forecasts by hydrologists is problematic. There is no single statistic that is uniquely helpful, primarily because the data of interest is frequently the rare high flow events in the record. For real-time on-line forecasting eyeballing the forecast stage is crucial, but where longer records are available statistical methods, root mean square error (RMSE) and the Nash and Sutcliffe (1970) index are standard hydrological procedures (Cameron *et al.*, 2002; Legates and McCabe, 1999). In addition to these measures we look at the accuracy of forecasts with respect to the real warning times which operators use on this river to notify the public of a flood event.

18.4.2 Event Forecasting for Alston

Alston is the headwater gauging station on the River South Tyne. It is a Pennine, upland moorland and rural catchment, with very flashy runoff patterns. The hydrographs (graphical representation of changing stage on a river) rise and fall within hours as water runs rapidly from the hillside to the river. In this type of catchment flood warning times are very short so any increase in warning capability is an advantage. For the EA flood warning team, the accuracy in forecasting the peak is not as significant as having a good forecast of the rising limb which triggers flood warning actions downstream.

A series of experiments have been undertaken using the approach outlined above, i.e., the flood events were extracted and ranked, the networks were trained on the odd-numbered events and validated on the even-numbered events, using both backpropagation and Quickprop algorithms to produce forecasts for 1 and 2 hours ahead. Two further variations were introduced. In the first case, models were developed to forecast stage at Alston using the stage data alone. In the second case, both the stage data and the rainfall records from a continuously recording, 'tipping bucket' gauge nearby, were input to the model.

The results showed that adding information from the raingauge significantly improved the forecast timing in this flashy catchment. Where a forecast is made using the gauging station data alone, the network cannot forecast a rise in stage until it observes a rise in stage, so the forecast of rise is delayed by at least the time step used in the calculations. The earlier inputs from the raingauge kick-start the forecast rise in river level. Examples of validation forecasts on the largest event in the record (Figure 18.4a) and a more typical event (Figure 18.4b) illustrate the forecasting ability of the networks. In these examples, the data resolution is 1 hour but these forecasts could be improved by using 15-minute stage and rainfall input data.

Similar results were also obtained in a parallel study for Kilgram, a comparable flashy Pennine station on the River Ure in the headwaters of the Yorkshire Ouse system. At both Alston and Kilgram, the Quickprop algorithm performs slightly better than backpropagation. While the peak is not always hit precisely, the warning levels on the rising limb were either hit on time or hit slightly ahead of time. This is an advantage because the sooner a warning can be given the longer local people have to respond. In future work, including rainfall data from real-time radar images could improve the quality of flood forecasts.

These results suggest that ANNs can be useful in flashy headwater areas with short response times and if further raingauge data are available, possibly from adjacent catchments, that could also improve the forecast. They also indicate that where data are limited, as in this example where only seven flood events were used in the training series, a neural network can provide a useful forecast.

18.4.3 Event Forecasting for Bywell

Two contrasting approaches were taken in this case study. Firstly, as in the Alston experiment, the highest 11 stage events for which there are data at the four input

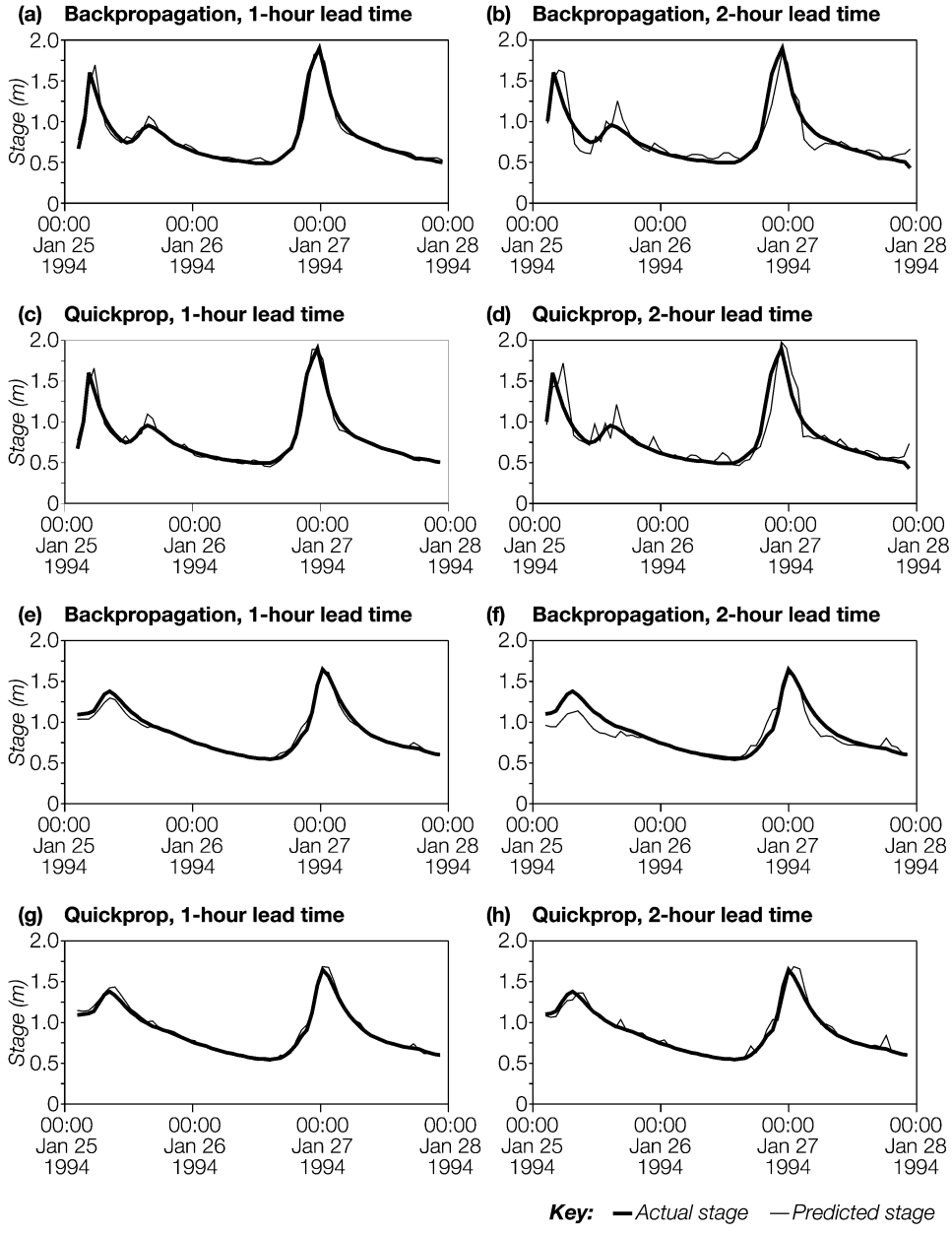


Figure 18.4 *Validation forecasts for events 1 (a–d) and 2 (e–h) at the Alston stage gauge, using stage at Alston and the raingauge data from Alston as inputs, and trained for 3000 epochs using the odd events*

stations and Bywell were ranked. The odd-numbered events were used as the training series and the even-numbered events were used for independent model validation using both the backpropagation and Quickprop methods for 1 and 2 hour lead times. The first flood warning alarm stage at Bywell is 3.5 m.

The peak forecast for the largest event in the record is acceptably modelled in both shape and timing (Figure 18.5a and b). The 3.5 m warning stage is forecast more accurately by the backpropagation algorithm for both 1 and 2 hours ahead. The lowest event on record (Figure 18.5b) is, by contrast, overpredicted at the peak and the warning times are forecast up to a couple of hours early. This may be in part due to the slower travel times of lower flood peaks down to Bywell and that this was the lowest event in the validation record. However, this peak height would not cause a flooding problem in the catchment.

This experiment was then extended to develop models for a longer lead time, i.e., 6 hours. However, the results proved less than satisfactory. The peaks and total runoff were underestimated and the forecast time of rise was delayed by 2 hours. While the hydrograph shape was correct, this model is operationally less useful, although under extreme circumstances it may be better than nothing. The fast response time of the catchment, where travel from Alston to Bywell is of the order of 6 hours, means that looking from 6 hours out there has not been enough significant rainfall and rise in stage to indicate that a significant storm runoff could be expected. The Quickprop algorithm performed better than backpropagation here. Under normal forecasting circumstances the network would be created using stations on both main rivers and from the reservoir at Kielder. If forecasts of this type are to be valuable then experiments that look at the radar rainfall in adjacent catchments should be integrated with the neural network data to provide longer lead time warnings.

In the second approach the event data were used to see how the neural network models would handle flood event data outside the training envelope. The critical question here is can neural nets cope when asked to model stage patterns that are higher or lower than the data used to develop the model? In this experiment the challenge is greater in that the model is given no guidance from stage levels at Bywell itself, or from the whole of the North Tyne system. The three highest flood events were excluded from the training series. The neural network was generated on events ranked 4–11 and used to forecast stage at Bywell for the three unseen events. Results for the third largest and largest unseen events are shown in Figure 18.6a and b. The training data has peaks ranging from 3.05 to 4.57 metres. The validation forecasts are for events peaking at 4.75–5.88 metres.

Encouragingly, Figure 18.6a shows a good forecast of both peak and timing for an event with stage 18 cm higher than seen in the training set. For the highest event on record with a peak of 5.88 metres, and without knowledge of the North Tyne inputs which contribute 40–50% of the stage, the forecast is very respectable for the rising limb but not surprisingly in three of the four forecasts underestimates the peak (Figure 18.6b). The results suggest that the network can cope with forecasting rising limb responses for extreme stage events. The accuracy of the peak and the modelling of alarm levels would seem to be appropriate for routine forecasters' purposes.

In modelling terms, it was to be expected that given a limited number of flood events and a situation where the validation data are outside the range of the

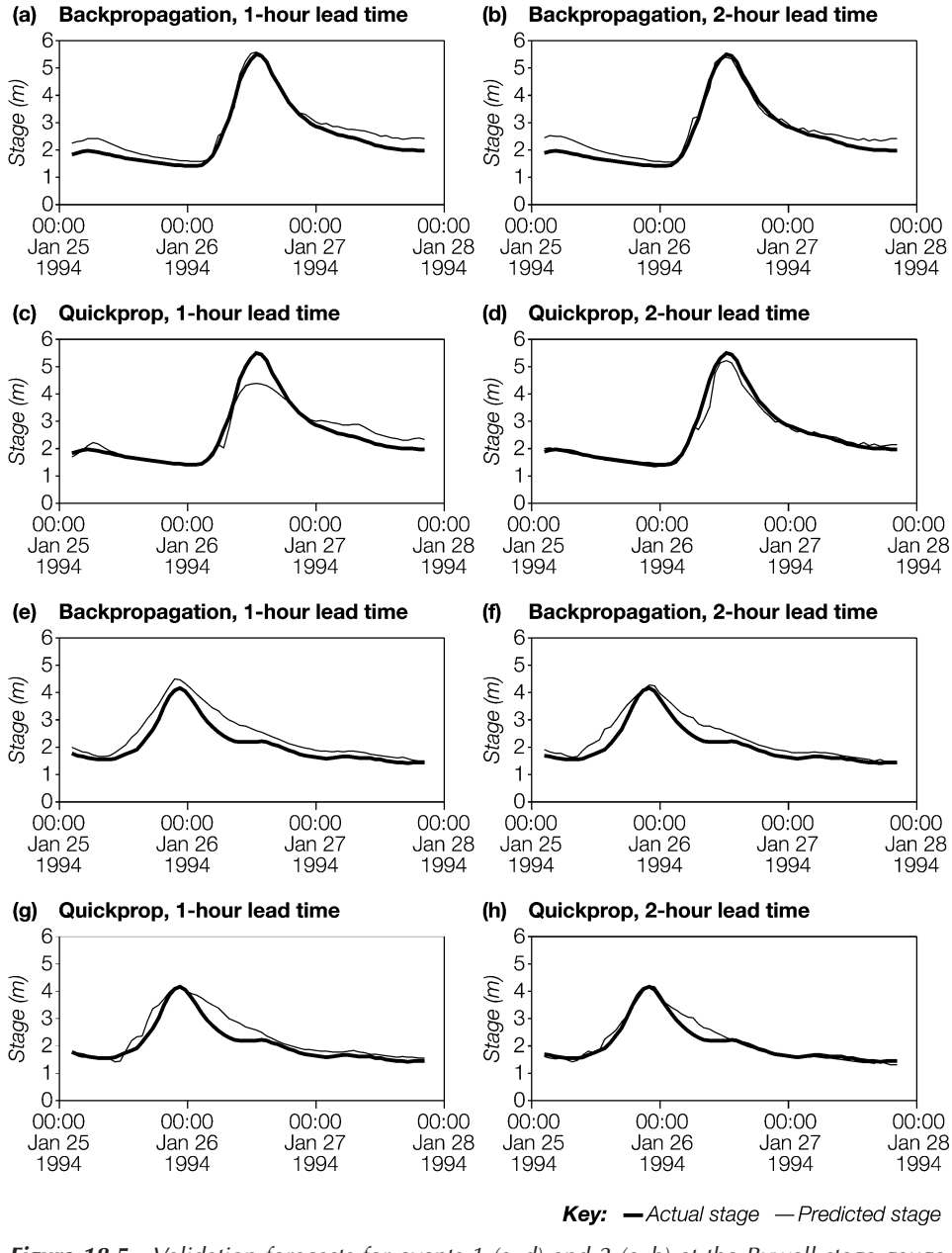


Figure 18.5 *Validation forecasts for events 1 (a–d) and 2 (e–h) at the Bywell stage gauge, using stage at Alston, Featherstone and Haydon Bridge and the raingauge data from Alston as inputs, and trained for 3000 epochs using the odd events*

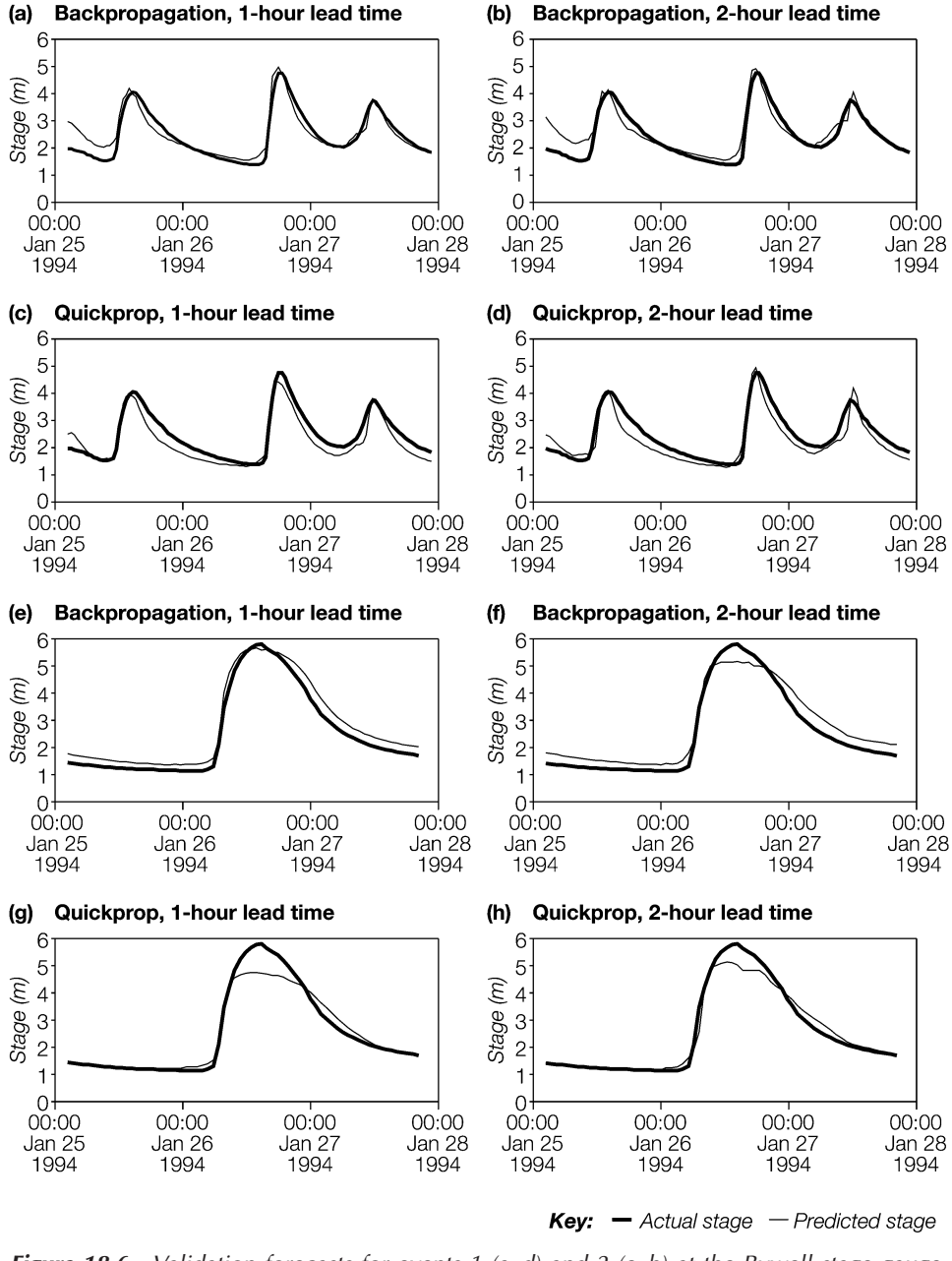


Figure 18.6 *Validation forecasts for events 1 (a–d) and 2 (e–h) at the Bywell stage gauge, using stage at Alston, Featherstone and Haydon Bridge and the raingauge data from Alston as inputs, and trained for 3000 epochs with the events ranked 4–11*

training data set that the model would underpredict very high events (Figure 18.6b) and overpredict a low event (Figure 18.5b). Ideally models would be developed with longer training sets, but where new gauging stations are installed and records are short, these results indicate that neural network models are useful even with limited data.

18.4.4 Continuous Forecasting for Haydon Bridge

In this set of experiments, models were developed to forecast river levels at Haydon Bridge. The inputs used were stage at Haydon Bridge and Featherstone, and stage and rainfall at Alston. The ANNs were trained for 3000 epochs using both the back-propagation and Quickprop algorithms, various experiments having shown that there was minimal improvement in the networks' ability to forecast at larger numbers of epochs. The networks were trained on data from October to April of 1996–97 and validated using data for October to April of 1994–95, 1995–96 and 1997–98. Table 18.1 shows the summary evaluation statistics for all four years for the two ANNs. The networks are compared with results from models developed using multiple linear regression and ARMA (AutoRegressive Moving Average) techniques.

The Nash–Sutcliffe and R^2 statistics are encouragingly high for the neural network models. The multiple linear regression and ARMA models do slightly less well. RMSE values, indicating the absolute error in stage between the actual and forecast levels, are of the order of centimetres for all models. Quickprop produces the best model followed by backpropagation, multiple linear regression and then the ARMA model.

Table 18.1 *Evaluation measures for long-term calibration of neural networks, multiple linear regression and ARMA models*

Algorithm	Historical period	Nash–Sutcliffe coefficient	*R*-squared	RMSE (m)
Backpropagation	Winter 1994/95	0.914	0.966	0.097
	Winter 1995/96	0.891	0.912	0.071
	Winter 1996/97	0.965	0.971	0.056
	Winter 1997/98	0.939	0.953	0.065
Quick propagation	Winter 1994/95	0.981	0.984	0.052
	Winter 1995/96	0.962	0.967	0.043
	Winter 1996/97	0.971	0.971	0.054
	Winter 1997/98	0.979	0.979	0.041
Multiple regression	Winter 1994/95	0.918	0.941	0.108
	Winter 1995/96	0.843	0.876	0.086
	Winter 1996/97	0.916	0.916	0.092
	Winter 1997/98	0.936	0.937	0.071
ARMA	Winter 1994/95	0.784	0.805	0.176
	Winter 1995/96	0.813	0.830	0.093
	Winter 1996/97	0.717	0.751	0.168
	Winter 1997/98	0.776	0.798	0.132

Table 18.2 *Accuracy and timing in predicting the operational level of 2.0 metres*

Event	Observed stage	Backpropagation	Quickpropagation	Regression	ARMA
1994/11/13	2.07	2.20	2.18	2.12	2.09
		(+1.0)	(+1.0)	(+2.0)	(+3.0)
1994/12/08	2.21	2.48	2.17	2.04	2.21
		(0.0)	(0.0)	(+1.0)	(+2.0)
1994/12/11	2.03	2.16	2.03	2.02	2.32
		(+3.0)	(+1.0)	(+4.0)	(+4.0)
1994/12/31	2.26	2.28	2.06	*	2.22
		(–1.0)	(0.0)	(N/A)	(+2.0)
1995/01/31	2.04	2.14	2.10	2.23	2.27
		(+1.0)	(+1.0)	(+2.0)	(+3.0)
1995/02/12	2.29	2.15	2.27	*	2.34
		(0.0)	(0.0)	(N/A)	(+2.0)
1995/02/19	2.10	2.19	*	*	2.06
		(+1.0)	(N/A)	(N/A)	(+2.0)
1995/02/22	2.61	2.08	2.52	2.08	2.74
		(0.0)	(+1.0)	(+1.0)	(+3.0)
1995/03/01	2.13	2.06	2.18	*	2.30
		(0.0)	(+1.0)	(N/A)	(+3.0)
1995/03/26	2.01	2.09	2.02	*	2.09
		(–1.0)	(+1.0)	(N/A)	(+2.0)
1995/10/03	2.02	2.34	2.00	2.12	2.25
		(+0.0)	(+1.0)	(+0.0)	(+2.0)
1995/11/15	2.37	2.25	2.40	2.14	2.02
		(+1.0)	(+1.0)	(+2.0)	(+2.0)
1996/02/18	2.18	2.31	2.30	*	2.03
		(0.0)	(+1.0)	(N/A)	(+2.0)
1998/01/01	2.09	2.24	2.08	*	2.05
		(0.0)	(+1.0)	(N/A)	(+2.0)
1998/01/08	2.17	2.17	2.39	2.19	2.36
		(0.0)	(+1.0)	(+2.0)	(+3.0)
1998/03/07	2.28	2.58	2.21	2.10	2.37
		(0.0)	(0.0)	(+2.0)	(+3.0)
1998/04/03	2.12	*	2.12	*	2.06
		(N/A)	(0.0)	(N/A)	(+2.0)

Note: *Forecast did not reach the operational stage at all.

Table 18.1 summarises the results for the entire seven-month winter series, but in event forecasting the accuracy with which the networks predict the operational alarm stage of 2.0 m on the rising limb of the hydrograph is also very important; Table 18.2 details the results from 17 independent, validation events. The Quickprop neural network proved to be the most accurate of the four models in terms of predicting the timing of the alarm levels. The ARMA model consistently predicted the operational stage 2 to 3 hours late.

These results show that the multiple linear regression model did not reach the alarm levels on eight of the 17 events and was generally 2 hours late at other times.

If there were no significant inputs between the gauging stations then one might expect a linear relationship between stage at Featherstone and Haydon Bridge, and a linear routing model or regression equation should work well. However, the River Allens join the Tyne between Featherstone and Haydon Bridge, as does the River Nent just downstream of Alston adding significantly to the discharge. The ability of ANNs to handle non-linearities, for example from ungauged tributaries, is its strength over alternative linear approaches.

In addition to improved accuracy, a clear advantage of the neural network software is that model development times are rapid, even over a long continuous time series. In these experiments the training runs took approximately 12 minutes for 3000 epochs, and validation of the models with the three years of winter data was virtually instantaneous. Multiple linear regression equations were also very fast to develop but their use appears to be limited. The ARMA models, on the other hand, required a great deal of experimentation and the results were not very satisfactory.

18.5 Conclusions

The incidence of river flooding is increasing and forecasts of global warming indicate that this trend is likely to continue. Building new flood defences to contain higher flows, defending vulnerable sites and domestic and commercial insurance claims all have serious financial implications. Accurate forecasting systems are required for real-time day-to-day emergency forecasting and for the planning and design of new defences. ANNs are flexible mathematical structures capable of identifying complex non-linear relationships between input and output data sets. Their use in hydrological forecasting is still very new but it has been shown here that they can provide a viable alternative approach to statistical models.

This software has been in use since December 1999. Operators report that it is easy to use and straightforward to manage in both training and real-time operations. The experiments described here are designed to test the system response given difficult hydrological conditions with the emphasis on testing to see whether the alarm levels on the rising limb of the hydrograph can be satisfactorily forecast. The results would give the operators time to generate warnings at the first alarm stages.

Efficient use of a neural network model requires a model trainer who builds appropriate networks and selects the best algorithm. To find the lowest error solution may take many runs, but training runs are quick to do, typically less than 10 minutes per run. Given this high operational speed, a series of experiments can be defined to ensure that alternative algorithms and stage regime models are fully investigated. Forecasters can therefore make a mature selection of networks for operational use on the basis of wide experience. Real-time forecasts with established networks are effectively instantaneous once the data are downloaded. Forecasters must therefore take the time to investigate the effects of model architecture to ensure the validity of the results. However, having established a suite of models for operational use, these may be used in the flood room by additional operators after a short period of training (one day at most). Moreover, the experience with the software to date indicates

that it is a viable technology that should be considered as an additional tool for hydrological forecasting.

The software was developed in response to a growing demand for user-friendly access to ANNs but integrated within a single, user-customised software environment. These approaches are already used commercially in areas such as database mining and financial market prediction, but this software represents the first commercially available, operational hydrological forecasting tool that allows for the development of river stage forecasting models. In addition to incorporating other neural network algorithms, it is envisaged that the software will eventually be linked to a GIS in order that forecasters can visualise the flood wave as it travels down the catchment, both from a plan and a 3-D perspective. The spatial element will also allow forecasters to calculate flood inundation. The potential for constructing spatial decision support systems that incorporate the expert knowledge that is currently embedded in human forecasters remains an exciting area for future development.

References

Abrahart, R.J., See L. and Kneale, P.E. (1999) Using pruning algorithms and genetic algorithms to optimise network architectures and forecasting inputs in a neural network rainfall-runoff model, *Journal of Hydroinformatics,* **1**: 103–14.

Abrahart, R.J., See, L. and Kneale, P.E. (2001) Investigating the role of saliency analysis with a neural network rainfall-runoff model, *Computers and Geosciences*, **27**(8): 921–8.

Beale, R. and Jackson, T. (1990) *Neural Computing: An Introduction*, Adam Hilger, Bristol.

Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.

Cameron, D., Beven, K., Tawn, J. and Naden, P. (2000) Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation), *Hydrology and Earth System Sciences*, **4**(1): 23–34.

Cameron, D., Kneale, P.E. and See, L. (2002) An evaluation of a traditional and a neural net modelling approach to flood forecasting for an upland catchment, *Hydrological Processes,* **16**: 1033–46.

Campolo, M., Soldati, A. and Andreussi, P. (1999) Forecasting river flow rate during low-flow periods using neural networks, *Water Resources Research,* **35**(11): 3547–52.

Cancilla, D.A. and Fang, X. (1996) Evaluation and quality control of environmental analytical data from the Niagara River using multiple chemometric methods, *International Association Great Lakes Research*, **22**: 241–53.

Chang, F.J. and Hwang, Y.Y. (1999) A self-organization algorithm for real-time flood forecast, *Hydrological Processes,* **13**(2): 123–38.

Dawson, C.W. and Wilby, R. (1998) An artificial neural network approach to rainfall-runoff modelling, *Hydrological Sciences Journal,* **43**(1): 47–66.

Dawson, C.W. and Wilby, R. (2001) Hydrological modelling using artificial neural networks, *Progress in Physical Geography,* **25**(1): 80–108.

Fahlman, S.E. (1988) An empirical study of learning speed in backpropagation networks, *Technical Report, CMU-CS-88-162*, Carnegie Mellon University.

Golob, R., Stokelj, T. and Grgic, D. (1998) Neural-network-based water inflow forecasting, *Control Engineering Practice*, **6**: 593–600.

Gozlan, R.E., Mastrorillo, S., Copp, G.H. and Lek, S. (1999) Predicting the structure and diversity of young-of-the-year fish assemblages in large rivers, *Freshwater Biology*, **41**: 809–20.

Gurney, K. (1997) *An Introduction to Neural Networks*, UCL Press, London.

Haykin, S. (1998) *Neural Networks*, Prentice Hall, London.

HEC-RAS (2001) HEC-RAS water surface profile model, Boss International, online http://www.bossintl.com.hk/html/hec-ras_details.html.

Horritt, M.S. and Bates, P.D. (2001) Predicting floodplain inundation: raster-based modelling versus the finite-element approach, *Hydrological Processes,* **15**(5): 825–42.

Hsu K.L., Gupta, H.V. and Sorooshian, S. (1995) Artificial neural-network modeling of the rainfall-runoff process, *Water Resources Research,* **31**(10): 2517–30.

ISIS (2001) ISIS Flow / Hydrology, Wallingford Software Ltd and Halcrow Group Ltd, online http://www.wallingfordsoftware.com/products/isis.asp.

Kneale, P.E. and See, L. (1999) *Developing a Neural Network for Flood Forecasting in the Northumbria Area of the North East Region*, *Environment Agency*: *Final Report*, University of Leeds.

Kneale, P.E., See, L. and Kerr, P. (2000a) Developing a prototype neural net flood forecasting model for use by the Environment Agency, *Proceedings 35th MAFF Conference of River and Coastal Engineers 2000*, 11.03.1–11.03.4.

Kneale, P.E., See, L., Cameron, D., Kerr, P. and Merrix, R. (2000b) Using a prototype neural net forecasting model for flood predictions on the Rivers Tyne and Wear. British Hydrological Society, 7th National Hydrology Symposium, 3.45–3.49.

Legates, D.R. and McCabe, G.J. (1999) Evaluating the use the 'goodness-of-fit' measure in hydrologic and hydroclimatic model validation, *Water Resources Research*, **35**: 233–41.

Lek, S. and Baran, P. (1997) Estimations of trout density and biomass: a neural networks approach, *Nonlinear Analysis: Theory, Methods and Applications,* **30**(8): 4985–90.

Maier, H.R., Dandy, G.C. and Burch, M.D. (1998) Use of artificial neural networks for modelling cyanobacteria *Anabena* spp. in the River Murray, South Australia, *Ecological Modelling*, **105**: 257–72.

Marshall, C.T. (1996) Evaluation of River Flood Forecasting Systems, R&D Technical Report W17, Environment Agency, Bristol.

Mastrorillo, S., Lek, S. and Dauba, F. (1997) Predicting the abundance of minnow *Phoxinus phoxinus* (Cyprinidae) in the River Ariège (France) using artificial neural networks, *Aquatic Living Resources,* **10**: 169–76.

Mellor, D., Sheffield, J., O'Connell, P.E. and Metcalfe, A.V. (2000) A stochastic space–time rainfall forecasting system for real time flow forecasting II: Application of SHETRAN and ARNO rainfall runoff models to the Brue catchment, *Hydrology and Earth System Sciences*, **4**(4): 617–26.

Mike 11 (2001) MIKE Flood watch, A management system for real-time flood forecasting and warning, Danish Hydraulics Institute, online http://www.dhisoftware.com/mike11/index.htm.

Moatar, F., Fessant, F. and Poirel, A. (1999) pH modelling by neural networks. Application of control and validation data series in the Middle Loire river. *Ecological Modelling,* **120**(1): 141–56.

Nash. J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models, Part 1: a discussion of principles, *Journal of Hydrology,* **10**(3): 282–90.

Neelakantan, T.R., Brion, G.M. and Lingireddy, S. (2001) Neural network modelling of Cryptosporidium and Giardia concentrations in the Delaware River, USA*, Water Science and Technology,* **43**(12): 125–32.

NOAH 1D (2001) NOAH 1D Modelling System, Newcastle Object-oriented Advanced Hydroinformatics, online http://www.ncl.ac.uk/noah/.

Omernik, J.M. (1997) *Non-point Sources Stream Nutrient Level Relationships: A Nationwide Study*, EPA-600/3-77-105, Corvallis Environmental Research Laboratory, Office of Research and Development, US EPA, Corvallis, Oregon.

Openshaw, S., Kneale, P., Corne, S. and See, L. (1998) *Artificial Neural Networks for Flood Forecasting: Final Report*, MAFF Project OCS967P, University of Leeds.

Recknagel, F., French, M., Harkonen, P. and Yabunaka, K. (1997) Artificial neural network approach for modelling and prediction of algal blooms, *Ecological Modelling,* **96**(1): 11–28.

RFFS (2001) Real-Time Flood Forecasting System, Institute of Hydrology, Wallingford, online http://www.nwl.ac.uk/ih/www/products/iproducts.html.

Rumelhart, D.E. and McClelland, J.L. (eds) (1986) *Parallel Distributed Processing*, MIT Press, Cambridge, MA.

See, L. and Openshaw, S. (1999) Applying soft computing approaches to river level forecasting. *Hydrological Sciences Journal*, **44**: 763–78.

See, L. and Openshaw, S. (2000) A hybrid multi-model approach to river level forecasting, *Hydrological Sciences Journal*, **45**(4): 523–36.

SSIIM (1997) SSIIM 1.4, SINTEF Group, online http://www.sintef.no/nhl/vass/ssiim.html.

Stewart, M.D., Bates, P.D., Anderson, M.G., Price, D.A. and Burt, T.P. (1999) Modelling floods in hydrologically complex lowland river reaches, *Journal of Hydrology*, **223**(1–2): 85–106.

Van Kalken, T., Huband, M., Cadman, D. and Butts, M. (2001) Development of a Flood Forecasting Model for the Welland and Glen Rivers in East Anglia, UK, online http://www.dhisoftware.com/uc2001/Abstracts_Proceedings/Papers01/122/122.htm.

Whitehead, P.G., Howard, A. and Arulmani, C. (1997) Modelling algal growth and transport in rivers: a comparison of time series analysis, dynamic mass balance and neural network techniques, *Hydrobiologia*, **349**: 39–46.

CONCLUSION

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

19

Undertaking Applied GIS and Spatial Analysis Research in an Academic Context

Robin Flowerdew and John Stillwell

Abstract

This is the final chapter in a book that has aimed to demonstrate the utility of applying quantitative methods to solve real-world problems in human geography. The collection clearly exemplifies the wide range of applications and flags up the potential for future endeavour. Yet doing applied research within the confines of an academic environment is neither easy nor straightforward. This chapter concludes the book by describing the setting and some of the applied quantitative work done at the North West Regional Research Laboratory and at the Yorkshire and Humberside Regional Research Observatory during the 1990s. Some of the constraints under which the organisations operate are discussed, along with more detailed examples of four applied projects. Finally, some comments are made on the future scope for commercial applications of geography and their implications for the discipline.

19.1 Introduction

The chapters contained in this volume have reported on a spectrum of applications of GIS, spatial analysis and geographical modelling for a variety of clients in a range of different contexts. Whilst certain contributions refer to projects undertaken by individuals working as consultants for companies in the private sector (such as PricewaterhouseCoopers, David Simmonds Consultancy) the majority of the applications have been undertaken by researchers employed by organisations created

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

within their universities specifically to provide an environment in which applied research can flourish (such as CURDS, Nexpri, SERRL), or by academics working from within their respective university departments. These university-based studies reflect the extent to which 'applied geography' has emerged within the discipline of geography and how quantitative and computer-based approaches have been at the forefront of this development over the last 20 years in particular.

In the sections of this chapter that follow, we concentrate on three aspects of applied research. Firstly, we review the rationale and structure of two different institutional models for applied research established in neighbouring regions in the north of England, the North West Regional Research Laboratory (NWRRL) and the Yorkshire and Humberside Regional Research Observatory (ReRO). Secondly, the types of projects undertaken by each of these organisations are reviewed and selected GIS or spatial analysis-based applications are detailed for exemplification. Finally, based on our personal experiences of working for the NWRRL and ReRO respectively, we identify some of the major benefits arising from applied research but set these positive factors against a number of constraints or difficulties that have been associated with the conduct of this type of research.

19.2 New Regional Research Institutions for Applied GIS and Spatial Analysis

19.2.1 Introduction

Government funding for the social sciences, including human geography, comes via the Economic and Social Research Council (ESRC), renamed from the Social Sciences Research Council because of Government scepticism in the 1980s about their findings (which were often at variance with the underpinnings of Government policy). Then (and now) the ESRC received a tiny proportion of the total science budget, and had to argue its case to maintain this share. It saw its best chances of obtaining funding through emphasising that some parts of the social sciences were relevant to business and industry, and that social sciences too could adopt the technological 'laboratory' style of the physical sciences. GIS came along at just the right time and met the criteria. Arguably at least GIS was part of human geography and hence social science; requiring 'equipment' and 'laboratory facilities' (just like 'proper science'); seemingly objective and free from questions of values and interpretations; and with the potential to be relevant to a wide range of clients and 'customers'.

At the same time, the ESRC also attempted to exert more influence on the future of the social sciences, increasing the extent to which it decided what areas to put funding into, as opposed to merely being responsive to the ideas of practising social scientists. This was done through creating research initiatives or programmes, whose general topics were announced in advance with a request for proposals, several of which were to be funded. In addition to those proposals that were funded, this strategy had the effect of encouraging other research groups to think about the issues and

in some cases to appoint relevant staff. Unsuccessful ESRC research programme proposals have often been funded through other ESRC research money or by other funding agencies, so the research programme approach can be regarded as levering more research in the specified area than is actually funded.

19.2.2 The Regional Research Laboratories

It was in 1987, prior to the publication of the Chorley Report on the development of a national GIS policy, that the ESRC began the trial phase of its Regional Research Laboratory (RRL) initiative, moving in 1988 to the main phase. Eight RRLs were created in the main phase:

- the South East RRL (based at Birkbeck College London and the London School of Economics);
- the Midlands RRL (based at the Universities of Leicester and Loughborough);
- the North East RRL (based at the University of Newcastle upon Tyne);
- the North West RRL (based at Lancaster University);
- the Urban and Regional Policy Evaluation RRL (based at the Universities of Liverpool and Manchester);
- the Wales and South West RRL (based at the University of Wales Institute of Science and Technology and the South West Regional Computing Centre at the University of Bath);
- the Scottish RRL (based at the University of Edinburgh); and
- the Northern Ireland RRL (based at the University of Ulster).

In addition, the RRL initiative provided support for a Co-ordinator (Ian Masser, at the University of Sheffield) and a Technical Advisor (Mike Blakemore, at the University of Durham). The individual RRLs were constituted by a team of academics with permanent contracts with their universities, supplemented by appointed research staff. In five cases, the RRL was run by geographers, who also played a major role in a sixth, and people whose disciplinary affiliation was to geography, although sited in a planning department, dominated another. Over time, the composition of these RRLs has changed, as a result of moves of key people from one site to another and changes in degrees of involvement. Another trend was for the multi-site RRLs to split, so that the roles of the LSE, Loughborough and Bath diminished, and the Urban and Regional Policy Evaluation RRL split into two separate operations at Liverpool and Manchester. Other groups doing GIS research have also been granted 'RRL status', so groups from Glamorgan, Glasgow and York, among others, are part of RRL.net, the organisation that exists to coordinate the former RRLs in their post-ESRC situation.

The intention of the RRL Initiative was to provide resources for the groups receiving funding to allow them to become 'well-founded' laboratories for GIS work

(Masser, 1988). Money was provided for equipment and for staff, but the RRLs were always expected to become self-supporting through working on a commercial basis with clients in government and industry. At the end of the main RRL funding in 1991, the RRLs remained in existence but have followed different paths, some effectively disappearing while others have grown in size through a mixture of strategies.

There were always some ambiguities and contradictions in the roles that they were expected to play, reflected in part in the name. Part of the original idea was that each RRL would be particularly responsible for collecting and providing access to datasets about its own region, and for liaison with potential clients in that region. This function worked reasonably well in the RRLs with responsibility for Scotland, Wales and Northern Ireland, where there were government and commercial offices with devolved responsibilities with which they could work. The English RRLs, however, tended to work with datasets and organisations at the national level, and although they may have built up closer relationships with local government or other organisations in their own regions, their roles quickly became demarcated in terms of subject specialisation rather than region.

The term 'Research' also relates to an ambiguity in the role of the RRLs. The requirement to become self-supporting appeared to demand a concentration on commercial and consultancy work. However, the potential 'users' of the research, especially those from the private sector, were very concerned that the RRLs might become subsidised competitors for their own companies. Hence there was also a requirement on the RRLs to conduct basic research in GIS, which might lead to innovations that could be developed commercially by private sector organisations. Many or most of the RRLs applied successfully for GIS-related funding for what were clearly research projects, and indeed ESRC funded several projects in this general area, including a collaborative research programme with the Natural Environment Research Council (NERC).

It is clear that the RRL designation 'laboratory' was part of the ESRC's attempt to provide the social sciences with a more scientific image. Although RRL staff seldom wore white coats and handled test tubes, the RRLs did usually possess large rooms where several people worked together at a range of computers. For many in human geography this may have been a change from the model of the single scholar working in isolation in his or her own office with occasional trips to the library or archive. In the next section we outline the ReRO model which provided an alternative approach to bringing together academics from differing institutions to work together on applied projects.

19.2.3 The Yorkshire and Humberside Regional Research Observatory

No RRL status was awarded to a university in Yorkshire and Humberside. This was surprising given the large group of applied geographers at the University of Leeds and the international status of the modelling work carried out there under the leadership of Alan Wilson. In fact a consultancy business named GMAP had already been set up within the University and had begun to flourish, using spatial

interaction modelling to advise companies on locational decisions and building intelligent GIS (Clarke *et al*., 1995; Birkin *et al*., 1996). However, in 1990, separate from its commercial consulting activities under the GMAP banner, the Yorkshire and Humberside Regional Research Observatory (ReRO) was created to provide clients in the region with access to a region-wide data and analysis service.

Initially, ReRO was a joint venture between the School of Geography at the University of Leeds and the Centre for Urban Development and Environmental Management (CUDEM) at Leeds Polytechnic (later Leeds Metropolitan University). The aim was to create a network of research centres, or 'Associate Research Units' (ARUs) as they became known, each offering a range of skills and expertise in data collection and processing, problem-solving, strategy formulation and policy analysis. Such a network was regarded as being advantageous in that research teams might be assembled at relatively short notice from different ARUs with the skills and local knowledge required to respond to any particular request for research from the public or private sector. Two further units, the Policy Research Unit at Leeds Polytechnic, and GMAP Ltd, became part of the network at the outset and research units from other universities in the region joined the network subsequently, including the School of Geography and Earth Sciences at the University of Hull, the School of Social and Professional Studies at the University of Lincolnshire and Humberside, the Regional Office, University of Sheffield, the Policy Research Centre (PRC) and the Centre for Regional Economic and Social Research (CRESR) at Sheffield Hallam University, the Institute for Transport Studies at the University of Leeds as well as groups of researchers at the Universities of Bradford, Huddersfield and York. Thus, through its network of ARUs, the Observatory was able to bring together not only subject specialists with skills in data processing, spatial analysis and policy evaluation, for example, but also academics and researchers with a wealth of knowledge and understanding of their own local geographical areas, thus creating a rich base of experience and expertise and a unique resource with which to address the range of strategic, developmental and research issues related to Yorkshire and Humberside.

The general structure of ReRO (Figure 19.1) indicates that the ARUs represented the formal members and, through a Committee of Members, constituted the strategic development body for the Observatory through its annual meeting. ARUs paid an annual subscription for the benefits of being part of the research network. The executive function of the Observatory was managed through the Board of Directors consisting of five senior academics from the ARUs. The day-to-day management and operation was handled through the ReRO Office housed in the School of Geography. In the early 1990s, the Director was supported by a secretary, a research associate, a graphics design technician and a data manager, all of whom were appointed on a part-time basis.

Research Council funding was never secured by ReRO, whose staff, running and development costs were financed entirely from subscriptions from the member ARUs, overhead earnings (20%) from research contracts and consultancy work, sales of ReRO publications and a limited amount of sponsorship from companies such as Coopers & Lybrand Deloitte, Barclays Bank, Halifax Building Society and Northern Foods.

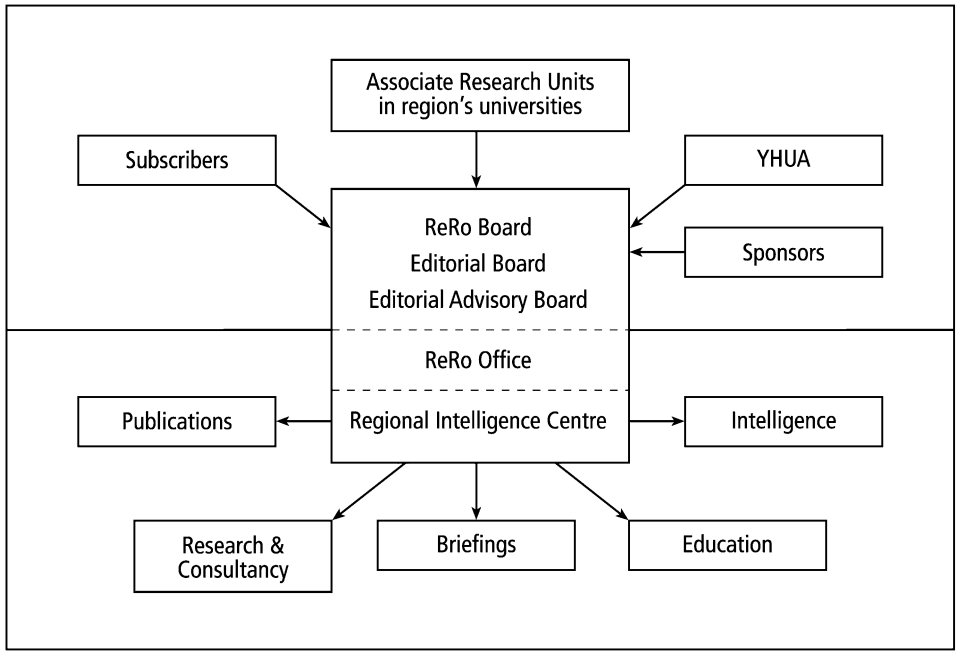


Figure 19.1 *The structure of the Yorkshire and Humberside ReRO*

19.3 Applied Research Activities

In the following two sub-sections, we discuss the range of research activities by the North West RRL and the Yorkshire and Humberside ReRO and present two more detailed examples of projects undertaken at each of the two centres.

19.3.1 Projects at the North West RRL

One very important category of research activity at the North West RRL has been the larger and longer term projects funded by research councils, some of which have been theoretical in nature. They are particularly valuable because they have tended to run for quite a long time, promoting financial stability and job security for the staff employed as researchers. They have usually brought in a substantial amount of money and should have led to opportunities for publication in the academic literature (though the need to chase the next project has sometimes meant that fewer papers reach publication than should be the case). Work in this area has been initiated by RRL staff and allows the RRL to build alliances with others around the university (or outside). The NWRRL has worked on projects involving spatial statistical models, areal interpolation and the modifiable areal unit problem. There have also been research council funded projects that have been more specific, such as evacuation modelling in the context of emergency planning.

In the more applied context, work has been done with Eurostat on GIS software for EU applications, with the European Environmental Agency on environmental databases, with the Home Office on developing GIS software for the Safer Cities Programme, and with the Department of Environment, Transport and the Regions (DETR) on pollution modelling using the critical loads approach. Much of the RRL's work, however, is made up of much smaller and shorter term projects, often based on the employment of relatively straightforward GIS capabilities for practical purposes. Such projects may only bring in a few hundred or thousand pounds, but are relatively simple to accomplish. A common type of project involves the mapping of data of various types, usually involving the conversion of unit postcodes to grid references. Sometimes the work involves aggregation of the data to areal units and production of choropleth maps. Another type of small project is the construction of a spatial database, for example the Lancaster Urban Archaeology Database which involved creating and linking georeferenced records for archaeological finds in the city of Lancaster. Sometimes the work is basically data analysis, not necessarily involving GIS skills. The RRL has also been asked to assist in the design of spatial sampling, involving the collection of data on household waste and the analysis of rural poverty among older people. Another frequent task has been to solve routing problems, for example through examining the extra travel distances or times if a particular facility is closed down or another one is opened. This type of analysis has been important in the study of health care and leisure facilities as well as commercial service delivery systems. A particularly important set of tasks has been the study of visibility, especially in relation to the location of wind farms, where the RRL has developed close links with a number of companies, providing technical expertise for planning enquiries as well as relevant information for siting issues.

Two examples of applied research provide an indication of some of the advantages and disadvantages of being an RRL in particular contexts. The first is an ESRC-funded project on using GIS to measure remoteness in the context of issues related to funding for public services, and the second is a large public financed project on the problems for North West England of changing demographic trends.

Remoteness of Services

This project was primarily academic rather than commercial, though its genesis was very much applied. The original beginning of interest in this area was a set of consultancy projects with local authorities relating to the Standard Spending Assessment (SSA) system. The latter was designed to provide a fair basis for central government to disburse funds to local government with the objective of providing standard levels of service across England. The formulae used were open to question on several grounds (Flowerdew *et al.*, 1994). In particular, several rural authorities felt that the SSA calculations did not adequately reflect the additional costs they incurred because of the time taken to deliver services to clients. The project was based on the idea that GIS provided a methodology for calculation of the actual distances that service deliverers would have to travel to reach the entire population. If assumptions were made about how service delivery centres were located, estimates of the costs in service delivery could be derived. Calculations of this type could be done for

all local authorities, on the basis of several sets of assumptions about the locations of service delivery points. These estimated costs could then be compared with existing allocations. The ESRC agreed to support the project, which employed one research assistant half time for a year.

In the event, the project suffered somewhat from being carried out alongside several other commercially funded projects. It was difficult for the researcher concerned to spend 50% time on this project when there were other shorter term projects requiring significant input to a tighter deadline. More pre-processing of data was required than anticipated before the calculations could be started, and the latter proved to be more time-consuming than expected. Although the availability of experienced staff may have been an advantage of being conducted within an RRL, the necessity of carrying out the work at the same time as other commercial projects proved to be a problem. A further problem arose in writing up the results of the project, which conflicted with the necessity for all staff members to proceed with tendering for new projects to maintain the RRL's income flow, and to carry out those projects for which tenders had been successful. This was exacerbated by the Research Council's policy of paying overheads at rates much lower than commercial ones. Such projects are still worth doing for an RRL, however, because they are worthwhile for the group's academic profile, even if they do not bring in much additional money.

Ageing Population

Another project undertaken by the RRL involved studying the implications of changing population structures for North West England. This project was sponsored by a consortium of regional-level bodies, including the North West Regional Assembly, the North West Development Agency, the Government Office for the North West, the National Health Service Executive North West and Age Concern. The main aims of the project were to produce population projections by age at the district level for the North West and to investigate the consequences of these projections for a range of economic, social, health and planning issues.

Population projections are available from the Office of National Statistics for the population as a whole until 2025. However, for the purposes of this project it was important to project conditions at the local level for a longer period, and also to project the populations at the local level for population subgroups, including ethnic groups and the disabled population. These projections raised a number of interesting academic issues as well as being highly relevant to policy questions. A crucial issue in the reliability of projections turned out to be the effect of students, whose migration behaviour, before and after moving away from home for higher education, is particularly difficult to estimate given available data (Simpson, 1995). It seems likely that the use of census data may underestimate student population in university cities and the use of National Health Service registration data may overestimate it. As a result, the methodology used in this context produced wildly disparate projections for places with large student populations.

Producing local-area projections by ethnic group and disability status also raised methodological issues. Arbitrary decisions had to be made to produce these

projections and the unavailability of all the desired data meant that simplifications and unrealistic assumptions were necessary. For example, it was difficult to make projections for places that had a very small population in one ethnic group but had experienced migration of ethnic group members in or out. If migration for the ethnic group was assumed to be proportional to the population of that ethnic group, this could result in a very rapid and totally unrealistic rise (or a rapid disappearance) in the numbers of that group. Accordingly, we assumed that the numbers of migrants rather than their proportion would remain constant. Whereas government has, quite understandably, refused so far to make projections by ethnic group, we felt constrained to produce the best projection we could, even though we knew it was unrealistic in several respects.

In any exercise of this nature, it is vital to draw a distinction between projection and forecast. The nature of a projection is that it is based on the continuation, in some sense, of current trends. This is quite unlikely in many situations and, in some cases, it is more than likely that an undesirable projection will generate some sort of policy response to prevent it actually coming true. Of course there are many other factors, predictable or otherwise, which will mean that the assumption of conditions remaining constant will not be accurate.

19.3.2 Applied Research through Partnership

In summary, ReRO developed a portfolio of activities that can be grouped broadly under five headings:

- research and consultancy: through the assembly of multi-institutional and multi-disciplinary teams;
- intelligence: through the interpretation and analysis of data rather than mere cataloguing of current data sources;
- education and training: through a European initiative to provide a pan-European course in training of regional managers;
- publication: through a regular journal, *The Regional Review*, and a periodic bulletin; and
- networking: both amongst its member ARUs but also with relevant bodies in the region and beyond.

The majority of these services are provided through collaboration between one or more of the ARUs and examples of research contracts are listed in Table 19.1. In most cases, the applied work was focused at the regional scale and was carried out at a time when the regional tier of agencies and administrative institutions was embryonic. Work has been commissioned by central government departments, national agencies, regional agencies, local government bodies, charitable organisations and private sector companies.

The view at ReRO was that urban and regional planning and policy-making had substantial needs for information and intelligence. Good regional data capture, maintenance and analysis were considered imperative as the basis upon which to

Table 19.1 *Examples of ReRO research contracts*

Title of research project	Funding body
Regional Textile Database Analysis: Annual Report	Barclays Bank and six textile companies
Yorkshire and Humberside Universities: A Potential for the Region	Yorkshire and Humberside Universities Association
Profiles of English Regional Arts Boards	The Arts Council
The Impact of Inward Investment into Yorkshire and Humberside	Department of Trade and Industry and the European Commission
The Role of Leeds in the Regional Financial System	Leeds Development Agency and Leeds TEC Ltd
Research into the Application of Bio-technologies in North Yorkshire	Yorkshire and Humberside Development Association
Regional Technology Plan	The Regional Technology Network and Yorkshire and Humberside Partnership
Research into School Performance	Government Office for Yorkshire and the Humber
Developing a Strategic Framework for Yorkshire and Humberside: Review of Existing Policies and Strategies	The Regional Assembly for Yorkshire and Humberside

understand problems and to evaluate feasible solutions. Yet it was evident that few guidelines were available as to which particular demographic, economic, environmental or social variables should be used to measure regional performance. This vacuum has been highlighted by the debates on measuring 'sustainability' that preoccupied various organisations in the 1990s. ReRO always placed a high premium both on the collection of data and on 'adding value' to that data. On the one hand, this involved maximising the availability of computer technology to manipulate, analyse and display data. On the other hand, it also meant using expert local knowledge to provide reliable interpretation of the cross-sectional and time series datasets that have been assembled. As with the North West RRL, several of the smaller projects undertaken by ReRO involved collecting or assembling data, geocoding postcode-based records and mapping spatial distributions of phenomena such as inward investment establishments, poorly performing schools or organisations receiving funding from the Prince's Trust. Further details of two projects are provided to exemplify the nature of the applied GIS-based and spatial analytical work.

Adjustment of Population Projections

In the early 1990s, the data available from the 1991 Census provided local authorities with the opportunity to evaluate the accuracy of the 1989-based sub-national population and household projections for 1991 produced by the Office of Population Censuses and Surveys (OPCS, 1991) and the Department of the Environment (DoE, 1991) respectively. ReRO was commissioned to carry out this assessment for the region by the Yorkshire and Humberside Regional Planning Conference (representing the interests of all the local authorities in the region) and the Regional Office of the DoE. In

addition, the ReRO research team, comprising staff from the Universities of Leeds and Sheffield Hallam, were asked to utilise the findings in the formulation of a range of alternative population and household projections which could subsequently be used to inform the assignment of land for the future development of housing, part of the process of formulating Regional Planning Guidance (RPG).

It was found that, overall, the 1989-based projections for the region as a whole underestimated the mid-1991 estimates by 0.5% of its 4.9 million inhabitants but that differences between projections and estimates varied appreciably by local authority and by age group. In North Yorkshire, for example, overprojection of the total population was due to significant overprojection of those living in the county of young working age, offsetting an underprojection of those aged over 40. In Humberside, on the other hand, the population projections were lower than the population estimates for mid-1991 in almost all age groups, providing a substantial underprediction of the population of the sub-region as a whole. One of the key factors in explaining the differences was net migration. A methodology was developed to update the existing 'official' population and household projections for local authorities rather than constructing a new projection model (ReRO, 1993). The procedure for adjusting the population projections involved applying revised population base factors and adjusting the net migration assumptions on the basis of three scenarios about the range of possible futures. A modelling methodology was then used to convert the population projections into a revised set of household forecasts and involved subtracting the institutional populations from the projections, applying marital status factors (the proportions of a given population group that are either single, married or widowed/divorced) to the male and female projected populations by age group, and finally applying headship rates (the proportion of members of a population group defined by age, sex and marital status that are heads of households) to the relevant age–sex–marital status group projections. The details of the methodology and results of the project have been reported in Gore and Stillwell (1994).

The experience of undertaking the project was both rewarding and frustrating for the researchers involved. A huge database was created, a large amount of new demographic information about the region was uncovered, and an innovative solution was derived to a difficult problem. However, it became very clear to the research team at an early stage in the project that what was required to do a proper job was a full multi-regional population projection model linked to a regional model for household projections which would generate outputs that could be converted into projections for a variety of different uses. Frustration arose from the inability, due to time and cost constraints of the project, to deliver a fully fledged regional model that would become an immensely useful tool to support this type of analysis in the future. Further experience of undertaking applied quantitative research in population geography for local and regional authorities has been reported in Stillwell and Rees (2001).

Educational Performance

In the mid-1990s, ReRO was contracted by the Government Office for Yorkshire and the Humber to undertake a study of the educational performance of primary and secondary schools throughout the region. Recent studies by Gordon (1996) and

Clarke and Langley (1996) had shown the extent of spatial variation in school performance and preliminary work at the Local Education Authority (LEA) level had indicated huge variations in performance in Yorkshire and Humberside at Key Stages 2 (age 11) and 4 (age 16). At one end of the spectrum, the average numbers of pupils achieving level 4 in English, Mathematics and Science at primary schools in North Yorkshire were 66.4%, 66.5% and 73.5% respectively whereas equivalent figures for Kingston upon Hull, at the other end of the performance league table, were 55.8%, 54.5% and 63.1%. Similarly, the average percentage of pupils attaining five or more A* to C GCSE grades was 52.7% in North Yorkshire and only 24.1% in Kingston upon Hull (Stillwell and Langley, 1999). This geographical disparity is explained by a number of factors collectively summarised as levels of social disadvantage. These factors include children's personal characteristics, family structures, socioeconomic features, educational levels and ethnicity.

The focus of the project was to examine the geographical variations in achievement in more detail by looking at the performance of individual schools. This involved filtering out special schools, non-state schools and schools for which sets of examination results were incomplete and then geocoding the schools remaining in the database according to their unit postcodes. The locations of the schools were mapped against a set of ward boundaries and the best and worst performing schools were overlaid on a ward-based choropleth map of deprivation as measured by the Department of Environment's Index of Local Conditions, a measure based on seven deprivation indicators from the 1991 Census (unemployment, children in low earning households, overcrowded housing, housing lacking basic amenities, households with no car, children in unsuitable accommodation and educational participation at age 17). The spatial coincidence between schools performing badly and higher levels of deprivation was particularly significant. However, one of the key difficulties in linking examination performance and social deprivation is the use of ward-based socioeconomic characteristics as measures of school catchment area, when it is well known that social conditions vary within wards and that the size of school catchment areas differs widely from location to location. Consequently, one of the most important recommendations of the project was to use GIS buffering techniques to identify enumeration districts within a defined radius of each school and to compute new socioeconomic indicators based on the areas located within the buffer (see Stillwell and Langley 1999, for details). Gender-specific data was also made available to the researchers to look at the extent to which girls across the region were performing better than boys. Remarkable variation exists in the gender-gap, even at LEA level, where the East Riding was ranked 18th out of all 129 LEAs with 12.5% difference between the proportion of girls and boys getting five or more grades at A* to C, and North East Lincolnshire was ranked 124th, where the difference was only 3.9% (Bright, 1998). The project found no spatial relationship between the size of the gap and level of examination performance.

19.4 Cultural Differences, Constraints and Opportunities

The overviews of selected projects presented in the previous section exemplify some of the work undertaken within two different research institutions specifically

established to conduct applied research. In this section, some generalisations regarding the relative nature and priorities of applied work in contrast to academic work are offered and some of the benefits and problems associated with applied research are considered. Here we focus primarily on the role of the RRL although many of the points made relate to applied research in an academic context more generally.

19.4.1 Differences Between the Applied and Academic Research Cultures

There can be little doubt that operating within the framework of an RRL had a major impact on the nature of the work done and on the balance between theoretical and applied work. This is partly because an RRL has a particular remit to work in the GIS field, but largely because of the financial arrangements under which it has had to operate since the conclusion of RRL funding. The key issue of course is that the RRL employs a number of staff funded to work on GIS projects in general rather than on any one specific project. The first priority in considering possible work is therefore to provide enough money to pay the staff, and if possible, to give them a reasonably long-term degree of security. One major difference between RRLs and commercial consultancy companies is that the director of the RRL will usually have job security through a continuing contract with a university. The negative aspect of this is that he or she will also have academic responsibilities, and must be prepared to fit consultancy work around teaching commitments, marking examinations, and other non-applied research.

The differences in time scales between academic and commercial research are well known. Academic projects have a long lead time and last for a considerable period, sometimes for several years. Commercial projects are likely to come up with little prior notice, insufficient time to prepare the tender (especially when it may involve staff in a number of different institutions), no lead time between winning the project and starting work, and usually a very tight time scale to final report. Academic time scales allow posts to be advertised so that a new employee can be brought in with appropriate skills. For commercial projects, it is essential to have staff available already with the right sorts of skills and with time available to work on the project intensively and immediately. Generally universities find it difficult to operate in this way, but in an RRL it is easier, provided that staff have the willingness and the ability to try their hand at a wide range of projects. Certainly, the more commercially orientated environment makes it very hard for any on-the-job training to be accomplished or for RRL funding to be used for supporting research students.

The academic environment still tends to be fairly casual about punctuality. In the commercial world, however, clients are more concerned that work be produced on time. In many circumstances, their business needs mean that a sub-optimal answer provided on time may be more useful than a better answer which does not arrive until two or three weeks after it was needed. Sometimes it may in fact be more important to say something than to say the right thing. Decisions have to be made to fixed deadlines, and sometimes a client may be worse off if no decision is made than a decision that is flawed. Whereas in the academic world, few people would want to argue for their point of view unless they had carefully considered the arguments and the available data; in the commercial world there is seldom time for reflection.

The *modus operandi* of applied research creates difficulties because work is often concentrated at certain times of the year and absent at other times. Also, because most commercial work is on a tight timetable, it is relatively hard to schedule it to iron out the peaks and troughs. This causes difficulties in that RRL staff may be overworked at one time and at a loose end at another. The latter should not be a problem, in that slack periods can be used for writing up work that has been done already (though this seems surprisingly difficult to do for commercial projects), or for seeking out new business. It may also be possible to bring in casual labour to help out at busy periods. Often a university is well equipped to do this because it contains well-trained students who may be prepared to work part time and short term on an interesting project. On the other hand, if RRL staff have different skills and interests, there may be times when one staff member is needed to work on several different projects at the same time while another has nothing to do, only for the situation to be reversed a few months later.

The establishment of the RRLs coincided with the development of Masters level courses in GIS, and these courses have been fruitful sources of researchers for the RRLs. This is particularly so where, as at Edinburgh and Leicester, successful MSc courses are taught in a department where an RRL is located. It has been relatively easy to persuade many of the best MSc students to stay on to work on RRL projects, or to become research students in the department. Even in other RRLs, such as the North West at Lancaster, finishing MSc students have been a good source of able and well-trained researchers, who have often been keen to acquire experience of the mixture of theoretical and applied research projects characteristic of RRL research. At Leeds, whilst GMAP recruited new graduates, masters' students and PhD students (particularly from geography) in considerable volume from year to year, the philosophy of ReRO was to assemble teams of established researchers from institutions within the region.

The best solution to recruitment and staffing problems is through establishing economies of scale. The more people there are, the easier it will be to accommodate changes in the workload, and to find somebody suitable to work on a short-term short-notice contract. Unfortunately, however, this requires a bigger overall turnover and one which is not sufficiently guaranteed to allow permanent appointments to be made. Over time, and with a consistent record of high-quality work, it becomes possible to build up the volume of work, but there is still going to be a workflow problem. Furthermore, writing tenders and submitting proposals is a difficult and time-consuming activity, which will normally fall to the leader of the RRL, whose task becomes increasingly difficult as the size of the operation increases.

The working environment for applied research has often in the past been at odds with a university's normal expectations of security for research staff. The researchers themselves need the promise of job security to commit themselves to working in an RRL, but university finance offices are reluctant to promise such security unless money is in hand or contracts have been won sufficient to pay the staff for whatever length of contract is specified. In a situation where income is irregular and projects may appear at very short notice, RRL directors may have to convince their institutions to take the risk that sufficient money will be brought in to pay for new appointments.

Another implication of the ambiguous position of an applied research community that sits between the academic and commercial worlds is that there may be a conflict between staff's research interests and the portfolio of funded projects. Sometimes the most interesting ideas may arise from unfunded or underfunded work; developing these ideas may only be possible by neglecting income-generating projects. Often, the most socially worthwhile projects are sponsored by organisations with very limited resources. Applied research staff may also feel an obligation to colleagues and students; a lot of time at the North West RRL, for example, has been spent in helping students with GIS problems, and in exploring possible GIS aspects of the research of colleagues with other interests. Although it is important to show colleagues in the same university how GIS and spatial analysis may help them with their research, there is sometimes a risk of setting up a free and unresourced GIS consulting service.

Academics are trained to make sure they provide proper acknowledgement of other people's ideas, and that any claims not substantiated within their papers must be referenced, so that other academics may verify their truth or trace their intellectual provenance. Nobody in commerce or government will have time to do this, and referencing sources is usually unimportant. It may even produce a negative impression, as it may be taken to emphasise the academic nature of the project, using the word 'academic' in its sense of 'not leading to a decision; unpractical' (Onions, 1972). The style of reporting on a project is also an important difference between academic and commercial requirements. Academics are used to hedging their conclusions with all manner of reservations and *caveat*s, always aware that what they are saying may depend on context and may be matters of probability, nuance or weight of evidence. Such mealy-mouthed conclusions are not popular in the outside world, where a plain and definite statement is required and a short executive summary of any major report is paramount since this is all that the key decision-maker is likely to read. The conclusions, and indeed the description of the project, must also be short, succinct and definite. When working on issues related to government policy, the 'taxi ride principle' applies; apparently there is little point in saying anything that is too long or complex for a civil servant to explain to a minister during the taxi journey from the Ministry to the House of Commons!

A further simplification is the tendency for decision-makers and the media to ascribe work not to the individuals who conducted it, nor to the RRL or ReRO, but to the university with which they are affiliated. Academics may be taken aback or overwhelmed with responsibility if their analyses and policy recommendations are ascribed not just to them, but to their universities. This is especially so if they are aware that many of their colleagues would not share their views about the issues involved.

From the academic's point of view, publication, especially publication in refereed academic journals, is an important part of career development. In fact this is much more important than dissemination of research that may be of practical relevance. Academics are notoriously bad at writing about the practical application of their work for the layperson to understand. This issue is particularly familiar to one of the authors after 12 years of editing a publication (*The Yorkshire and Humber Regional Review*) trying to inform policy-makers of regionally relevant research. Equally, applied projects may be difficult to write up for academic journals for a variety of

reasons. In terms of the points just discussed, the academic may need to reinsert the nuances and the qualifications that were taken out to meet the client's needs for clarity and definitiveness. There may also be problems relating to confidentiality concerns. A client may not want his or her company's problems or operating policies published to the world in an academic journal, and anonymising the client may be insufficient if describing the situation makes the identity obvious. If the academic analysis indicates problems with the client's activities, the client will not usually want these known. Equally, if the research has identified new and better ways of doing something, considerations of commercial confidentiality may mean that the new methods cannot be described explicitly lest competitors adopt them, destroying any commercial advantage that the clients may have gained.

There may also be problems from the academic side. Academic research is usually best regarded if its findings are general. Results specific to one place or one company are less likely to be regarded as important by journal editors and referees. If important details are withheld or anonymised for reasons of commercial confidentiality, the academic article is further weakened. The perspective may also need to be different. In most fields, academic work would be expected to present a neutral review of situations and problems, and an even-handed discussion of what can be done about them. In contrast, the commercial report will necessarily be based on the position of one specific actor, the client, and will be primarily concerned with what the clients can do for their own benefit rather than for the good of society as a whole.

19.4.2 Benefits and Constraints

Despite the differences and problems discussed above, most participants in the research efforts of the RRLs and ReRO would probably agree in seeing many benefits to involvement in commercial projects. First, the wide range of applications is highly stimulating and the requirement to produce solutions is very challenging. Certainly some commercial problems are repetitive and formulaic, but very frequently there are interesting differences which require variation of a standard approach, and in some cases require revisions of an existing method which may shed new light on the problem from an academic perspective. A very common situation occurs when a fairly standard situation is encountered but there are limitations in the data available to develop a solution, and interesting intellectual problems arise in trying to make the optimal use of whatever data can be found. Sometimes, new and very valuable datasets are unearthed which can be used in other research projects or contexts.

Second, some intellectual satisfaction can be gained from the problem-solving nature of much applied research. Admittedly the problems must be solved with limited time and sometimes limited data, and some academic researchers may find this frustrating because they recognise what might be possible with greater resources. However, there are compensations in the finite nature of the problems. A researcher can feel he or she has reached the best possible solution within the constraints available. In contrast, much academic research is inevitably open-ended; one always feels

slightly unsatisfied, in the knowledge that further research or further thought could produce a fuller or more sophisticated understanding of the point at issue.

For both the authors, quantitative geographers trained in the 1970s, applied research in GIS and spatial modelling has often involved a journey back in time, to some of the operational research and statistical models that were then at the cutting edge of geography. Though fashions have changed within the discipline and not many geographers get excited by location-allocation, spatial interaction models or multivariate classifications, they do have a commercial value. So do some of the skills that quantitative geographers have, less valued perhaps within academic geography than they were, but able to command a considerable price in the commercial marketplace, particularly if combined with some more advanced computer programming skills.

The use of the very term 'applied geography', or of 'applied GIS and spatial analysis' more specifically, begs some important questions. If geography is to be applied, what is it to be applied to? Who is to be doing the applying? Who will benefit from the application? If geography is to be applied for money, for the support of geographers working in higher education, then the answer must be that it will be applied for the benefit of those who have the resources to pay for it. This will usually be business or government, and it will usually be for their own benefit. It is rare for applied research to be performed in the interests of poor or excluded groups within society, simply because they are poor and excluded and hence unable to commission such work. The Joseph Rowntree Foundation is an exception to the general rule. Even within commercial organisations, it is the larger and more aggressive who are most likely to be clients. If geography is applied to help such companies with store location or market penetration, it will inevitably be at the expense of small businesses and corner shops with which we might otherwise tend to sympathise. In the RRL and ReRO context, however, ethical considerations of this type must be set against the ethical imperative to provide work and earn funds for the institution concerned.

It could also be argued that elements of applied geographical methods may themselves have questionable ethical effects. In constructing market areas or accessibility measures, for example, it is usual to adopt drive times as measures of accessibility. If facilities are located on this basis, their location pattern will no doubt make sense as far as car users are concerned, but there is a risk that those reliant on bus or rail travel, or other transport modes, will not be served adequately by the facilities. Further, for environmental reasons, we may not be happy to recommend a system that relies on even more universal usage of private cars with the well-known negative social and environmental externalities that they generate.

A further critique that applies to some aspects of applied geography in the commercial sector concerns invasions of privacy (Goss, 1995). Geodemographics and related marketing tools increasingly require large databases, including information about many individuals and local geographical areas. Some of this is collected in unobtrusive ways, such as recording address and purchasing information from store loyalty cards, or from participants in competitions and promotional activities. Apart from the invasion of privacy such activities can be regarded as constituting, there may be practical consequences, such as the refusal of credit, or the insistence on high insurance premiums, to somebody purely on the basis of where they live.

In terms of the concerns of contemporary geography, moreover, it can be argued that commercial applications lead to a focus on the tangible and perhaps on the economic. There may be little scope for studies in this area to engage with questions about values, interpretations and the taken-for-granted world. Arguably, a concern with the mappable and the unambiguously locatable tends to distract our attention from the ambiguous perceptions, interpretations and projects of those we are studying. The concern with the interests of clients, often largely their commercial interests, detracts from the full humanity of the people who are reduced to a geocode and a record of purchasing behaviour.

Other problems with a concentration on applied geography within an academic setting are institutional rather than ethical. Many of these have already been mentioned in the previous sub-section, such as the funding treadmill which requires a constant need to bring in more funding to keep the organisation going and to keep research staff in post. University administrations are not designed to promote commercial activities, however much current funding systems have forced them to try. Academic goals and commercial goals are not the same. Universities cannot ignore their finances and their need to bring in revenue, but to an important extent they still value truth and the search for knowledge. Commercial companies and government, quite rightly, have their own different goals and values, and an academic trying to combine both sets of goals must inevitably struggle to satisfy the latter without abandoning the former.

19.5 The Future for Applied Research

In conclusion, and despite the comments in the preceding section, our experience makes us generally optimistic about the future of applied geography and the use of the computer-based systems and methods that are now emerging for spatial analysis and forecasting. The geographical perspective, geographical data and geographical skills are highly relevant to many aspects of contemporary society. The activities of the RRLs, ReRO and companies like GMAP have done an enormous amount to make more people aware of how quantitative geography can be applied, and have helped some geographers understand better how to market their services and to meet the needs of outside clients. There is plenty of scope for applying geography.

Geographers may be in competition with other academics, especially economists, many of whom may be better integrated into the commercial and policy worlds. Certainly potential clients have a clear idea of how their organisations may benefit from applied economics, while geography still suffers from a ‘capes and bays’ or ‘geomorphology and climate’ image. Fiercer competition may come from consultancy firms employing the most recent generations of graduates or postgraduates. These firms have the economies of scale and the knowledge of client expectations to attract a lot of business. In contrast, academics may offer lower rates and more specialised knowledge, but often may be perceived as less likely to deliver on time, and more likely to produce work that is unhelpful to the client because it does not offer clear-cut solutions and recommendations. Our experience suggests that geographers in

commercial work in order to compete effectively must learn the importance of presentation, the importance of meeting deadlines, and the importance of presenting their work in a manner which the clients can most easily use.

Quantitative skills are necessary for much of what is done in the name of applied research. The quantitative revolution left geographers in the 1970s with degree courses that included statistical analysis and some practical work with maps and numerical data. Compared with many other social scientists, they could be effective in applying geography because they possessed practical and quantitative skills as well as the ability with write coherently about their results. In the current intellectual environment, some of these gains may be threatened. It is becoming easier to graduate in human geography without acquiring much quantitative expertise, and at the dissertation and postgraduate levels, there may now be more prestige in conducting a qualitative survey or a discourse analysis than doing quantitative analysis or modelling work. The discipline may lose out if we do not continue to teach and to apply quantitative skills. We must also be prepared to work to real-world rather than academic standards. The work must recognise the importance of punctuality rather than perfectionism. A willingness must be shown to commit to ideas and policies without having all the evidence.

We do not wish to argue that all geography should be applied, but there are huge opportunities for applying the work we do and it would be a great shame if we do not take further advantage of them, especially those that make use of GIS, analysis methods and modelling techniques. The examples contained in the earlier chapters of this book are testament to what has been achieved. Much more is possible in the future.

Acknowledgements

This chapter represents the fusion of ideas presented in earlier papers by Robin Flowerdew at the Annual Conference of the Royal Geographical Society with the Institute of British Geographers in Plymouth in January 2001 and by John Stillwell at a seminar on regional observatories at the University of Complutense in Madrid in November 2000.

References

Birkin, M., Clarke, G., Clarke, M. and Wilson, A. (1996) *Intelligent GIS: Location Decisions and Strategic Planning*, GeoInformation International, Cambridge.

Bright, M. (1998) Boys performing badly, *The Observer*, 4 January, p. 3.

Chorley, R. (1987) *Handling Geographic Information. Report to the Committee of Enquiry*, chaired by Lord Chorley, HMSO, London.

Clarke, G. and Langley, R. (1996) A review of the potential of GIS and spatial modelling in the new education market, *Environment and Planning C*, **14**: 301–23.

Clarke, G., Longley, P. and Masser, I. (1995) Business, geography and academia in the UK, in Longley, P. and Clarke, G. (eds) *GIS for Business and Service Planning*, GeoInformation International, Cambridge, pp. 271–83.

DoE (1991) *Household Projections England 1989–2011: 1989-based Estimates of the Numbers of Households for Regions, Counties, Metropolitan Districts and London Boroughs*, Housing Data and Statistics Division, DoE, London.

Flowerdew, R., Francis, B. and Lucas, S. (1994) The Standard Spending Assessment as a measure of spending needs in nonmetropolitan districts, *Environment and Planning C: Government and Policy*, **12**(1): 1–13.

Gordon, I. (1996) Family structure, educational achievement and the inner city, *Urban Studies*, **33**(3): 407–23.

Gore, A. and Stillwell, J. (1994) Updating population and household projections for Yorkshire and Humberside: an input into the strategic planning process, *Planning Practice and Research*, **9**(4), 381–93.

Goss, J.D. (1995) We know where you are and we know where you live, *Economic Geography*, **71**(2): 171–98.

Masser, I. (1988) The regional research laboratories initiative: a progress report, *International Journal of GIS*, **2**, 11–22.

OPCS (1991) *Subnational Population Projections: Population Projections by Sex and Age for Standard Regions, Counties, London Boroughs, Metropolitan Districts and Regional and District Health Authorities of England from Mid-1989*, Series PP3 no 8, HMSO, London.

Onions, C.T. (1972) *The Shorter Oxford English Dictionary on Historical Principles*, Third edition, revised with addenda, Clarendon Press, Oxford.

ReRO (1993) *An Analysis of the 1991 Census Results for Population and Household Projections in Yorkshire and Humberside*, Final Report for the Yorkshire and Humberside regional Planning Conference and the Department of the Environment, ReRO, Leeds, p. 259.

Simpson, S. (1995) Using the special migration statistics: examining ward migration and migrants' ethnic group, in Simpson, S. (ed.) *1991 Census Special Statistics on Migration, Workplace and Students*, Local Authorities Research and Intelligence Association, London, pp. 49–55.

Stillwell, J. and Langley, R. (1999) Information and planning in the education sector, Chapter 17 in Stillwell, J., Geertman, S. and Openshaw, S. (eds) (1999) *Geographical Information and Planning*, Springer, Heidelberg, pp. 316–33.

Stillwell, J. and Rees, P. (2001) Applied population projection for regional and local planning, Chapter 7 in Clarke, G. and Madden, M. (eds) *Regional Science in Business*, Springer, Heidelberg, pp. 115–36.

Author Index

Abowd G. 56
Abrahart R. 354, 359
Acevedo W. 29
Ahn H. 50
Aitken M. 287
Ake K. 251
Alexander S. 5, 234, 235, 236, 237, 238, 244, 251
Allen P. 2
Alvanides S. 4, 96
Amin A. 95
Anderson D. 237
Anderson M. 354
Andren H. 234
Andreussi P. 354
Andrienko G. 29
Andrienko N. 29
Anselin L. 161
Armstrong H. 297
Arulmani C. 355
Atkeson C. 56
Azzalini A. 164

Bailey A. 276
Bailey T. 163
Baldwin J. 140
Baldwin P. 99
Baran P. 355
Barmby T. 266, 290, 291, 292
Barnsley M. 97
Bartel A. 238, 239
Bascompte J. 234
Bates P. 354
Batey P. 94
Batty M. 2
Beadle H. 56
Beale R. 357, 360
Beattie A. 138, 322
Beishon S. 320
Bell W. 76
Bennison D. 37
Berrington A. 320
Berry B. 76
Berry J. 234, 235
Bertollini R. 152
Besag J. 147
Best N. 152
Betrancourt M. 29
Beven K. 362
Biasiotto M. 12, 23
Bibby P. 343
Biggeri A. 152
Birkin M. 4, 36, 38, 47, 51, 94, 381
Bishop C. 356, 358
Bishop I. 29
Blakemore M. 379
Blok C. 28, 29
Boden P. 51, 267
Böheim R. 320
Bohning D. 152
Borriello G. 56
Bosveld W. 219
Bottoms A. 140
Bowman A. 164
Boyce D. 37
Boyle P. 4, 134, 269, 274
Bracken I. 101
Bramley G. 264, 288, 290, 291, 292
Bright M. 388
Briggs D. 152
Brignone C. 56
Brion G. 354
Brocke R. 241
Brotherton J. 56
Brown J. 28
Brown P. 94
Bruna E. 251
Brunsdon C. 286, 290, 291, 292

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

Burch M. 355
Burnham K. 237, 238
Burrough P. 215, 216
Burt T. 354
Butts M. 355

Cadman D. 355
Câmara A. 4, 56
Cameron D. 354, 359, 362
Campolo M. 354
Cancilla D. 354
Canter P. 165
Card S. 29
Carree M. 302
Carroll C. 237
Carroll G. 298
Case B. 74
Casey J. 74
Cerioli A. 156
Champion A. 93, 259, 264, 269, 274, 288, 290, 291, 292, 342, 346
Chang F. 354
Chang K. 243
Charles A. 23
Charlton M. 286, 290, 291, 292
Cheng T. 29
Cheverst K. 56
Chorley R.
Clapp J. 74
Clarke G. 3, 4, 36, 37, 38, 44, 45, 47, 94, 381, 388
Clarke L. 320
Clarke M. 3, 4, 36, 38, 47, 51, 94, 381
Congdon P. 342, 346
Conroy M. 239, 251
Coombes M. 6, 288, 320
Cooper L. 218
Copley G. 197
Copp G. 355
Cormen T. 61
Corne S. 354
Craglia M. 5, 138, 140, 147
Crook A. 139
Cross A. 29
Cui J. 56
Culf R. 4, 36, 51
Curry M. 107
Curtis A. 218, 317

Dandy G. 355
Dauba F. 355
Davey-Smith G. 92
Davidson I. 172, 190
Davies N. 56
Dawson C. 354
Day N. 37
de Jong T. 216, 218, 219, 221
Dendrinos D. 2
Densham P. 218, 317, 327
Dey A. 56
Dias A. 4, 56
DiBiase D. 29
Diggle P. 164
Dijst M. 218
Dolde W. 74
Donnay J-P. 90, 97, 98
Dorling D. 29, 92
Douglas L 36, 45
Dragicevic S. 237, 239
Duke-Williams O. 134

Earnshaw R. 28
Eastman J. 239, 241, 242, 244, 245, 246
Efrat A. 56
Egbert S. 28, 29
Elliot C. 37
Elliot P. 152
Ellison A. 237, 238
Elstad J. 134
Ermisch J. 320
Errington A. 342
Evans K. 139
Evink G. 234, 235, 236
Eyre H. 44, 266, 288, 290, 291, 292

Fa J. 236, 237
Fahlman S. 358, 360
Fahrig L. 234, 239, 251
Fang X. 354
Ferrier S. 238
Feser E. 311
Fessant F. 354
Fingleton B. 156
Fischer M. 216
Floor H. 218
Florian M. 192
Flowerdew R. 3, 6, 12, 287, 383
Flynn A. 342
Foot D. 1, 37
Forrest R. 92, 94
Forman R. 234, 235
Fotheringham A. 6, 42, 218, 259, 264, 267, 269, 274, 286, 287, 288, 290, 317
Francis B. 383
Fraser P. 139
Freeman J. 298, 299, 301
French M. 355
Frisk C. 56
Frissell C. 234, 235, 236

Frost M. 6, 91, 99, 100, 104
Fuchs C. 160
Fuhrmann S. 28

Garrett P. 234, 235
Gatrell A. 135, 163
Geertman S. 3, 5, 216, 219, 220
George F. 51
Gerrard R. 328
Getis A. 161
Ghosh A. 317
Gibson C. 320
Gille J. 247
Goldstein W. 12
Golob R. 354
Goodchild M. 71, 99, 107, 317
Gordon D. 92, 94
Gordon J. 239, 246
Gore A. 387
Goss J. 393
Gouveia C. 56
Gozlan R. 355
Green A. 98, 99
Grgic D. 314
Gudgin G. 298
Gupta H. 354
Gurney K. 356
Guy C. 44, 45

Haeberling C. 28
Haggett P. 1
Haining R. 147
Hall P. 92
Hannan M. 298, 299, 301, 306
Hansen W. 219
Harche F. 317
Harkonen P. 355
Harper B. 56
Harris B. 1
Harris R. 4, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104
Harrison S. 251
Hartshorn T. 76
Harvey D. 2
Haykin S. 356
Hedley N. 28
Henig M. 145
Herbert D. 1
Hernandez T. 4, 11, 12, 30, 37
Hightower J. 56
Hilberts H. 218
Hill M. 145
Hodge I. 342
Hodgson M. 328
Holling C. 250
Hollis J. 288, 290, 291, 292
Hong J. 57
Horn M. 5, 172, 175, 181, 184, 187, 190, 191, 192
Hornhocker M. 234
Horritt M. 354
Howard A. 355
Hoyt H. 73
Hsu K. 354
Huband M. 355
Huff D. 21
Hwang Y. 354

Jackson S. 235, 236
Jackson T. 357, 360
Jern M. 28
Jia X. 244
Jiang B. 28
Johnston R. 2, 99
Jones K. 4, 23, 37
Joyce E. 29
Judge J. 56

Kalogiru S. 290
Kelsall J. 164
Kemper J. 311
Kennett A. 160
Kerr P. 359
Khattatov B. 247
Kiernan K. 320
Kloek W. 311
Kneale P. 6, 354, 359
Knudsen D. 287
Kobben B. 29
Kooper R. 57
Koussoulakou A. 29
Kraak M. 27, 28
Krugman P. 299
Krygier J. 29
Kulldorf M. 164
Kuterema A. 29

Laidler J. 197
Lamarque J. 247
Langley R. 388
Laposa S. 4, 72
Laws M. 234
Lawson A. 152
Ledbetter M. 28
Lee D. 2
Lee P. 92, 93, 94, 99
Legates D. 362
Lehmkuhl J. 236
Leiserson C. 61
Lek S. 355

Lenton S. 236, 237
Lesaffre E. 152
Levelt P. 247
Levine N. 165
Leyshon A. 50
Lingireddy S. 354
Little M. 176
Lloyd R. 28
Long S. 57
Longley P. 4, 90, 91, 92, 93, 94, 95, 97, 98, 99, 100, 107, 381
Lowe M. 342
Lucas S. 383
Lyjak L. 247

Mace R. 252
MacEachren A. 27, 28, 29
Macgill J. 264, 266, 288, 290, 291, 292
Mackinlay J. 29
Madden M. 3
Maguire D. 107
Maguire G. 56
Maier H. 355
Man H. 332
Manley T. 251
Maranzana F. 219
Marceau D. 237, 239
March J. 30
Margetts B. 45
Maritz J. 219
Marsden T. 342
Marshall C. 355
Martin D. 101
Masser I. 379, 380, 381
Massey D. 2, 95
Masuoka P. 29
Mastrorillo S. 355
Matilla K. 66
Maurer M. 236
May A. 2
McCabe G. 362
McCarthy P. 319
McCue D. 176
McDonald C. 37
McDonnell R. 216
McReynolds A. 56
Mellor D. 354
Menard R. 247
Merriam G. 234
Merrill T. 234
Merrix R. 359
Mesev T. 97
Metcalfe A. 354
Mitchel K. 56
Mitchell R. 92, 93
Moatar F. 354
Modood T. 320
Monk S. 342
Montello D. 28
Montgomery D. 165
Moore L. 98
Morrison A. 193
Morrison J. 29
Morse D. 56
Moseley M. 217, 355
Mueller G. 320
Munton R. 342
Murdoch P. 342
Murray A. 328

Naden P. 362
Nagarwalla N. 164
Nash J. 362
Neary J. 299
Neelakantan T. 354
Nelson R. 298
Newby H. 342
Newell J. 147
Noon B. 239, 251
Noss R. 234, 237
Nutter R. 316

O'Connell P. 354
O'Kelly M. 286
O'Loughlin J. 161
O'Sullivan D. 193
Omernik J. 354
Onions C. 391
Openshaw C. 51
Openshaw S. 29, 51, 92, 93, 96, 216, 354
Ord J. 161
Orford S. 92
Orr R. 56
Ottens H. 215
Ottenshaw J. 45
Ottensmann J. 45

Pacione M. 3
Pantazis C. 92, 94
Paquet P. 234, 236, 237, 238, 248
Pascoe J. 56
Pearce J. 238
Pearce M. 23
Pearl J. 237
Pedrosa P. 56
Peek J. 237
Perez de Val J. 236, 237
Pfeiffer U. 92
Phillimore P. 138, 322
Pimentao P. 56
Pinkerton M. 57
Plane D. 2

Poirel A. 354
Pollakowski H. 74
Pooler J. 45
Pradhan S. 56
Price D. 354

Quigley H. 234
Quintana A. 74

Ramasamy S. 29
Raper J. 56, 99, 347
Rawling M. 172, 190
Raybould S. 6, 320
Recknagel F. 355
Rees P. 6, 96, 264, 266, 267, 269, 274, 276, 288, 290, 291, 292, 387
Reeves C. 29
Reid H. 12
Remedio J. 56
Reynolds J. 37
Rhind D. 197, 347
Richards J. 244, 245
Ritsema van Eck J. 218, 219, 220, 221
Rivest R. 61
Roberts M. 197
Robertson G. 29
Robeson S. 29
Robinson B. 181, 184, 187, 191
Robinson G. 217
Robson B. 316
Rogerson P. 5, 156, 157, 164, 165
Romao T. 56
Ruggiero L. 234, 248
Rushton G. 327
Ruuska S. 66
Ryan N. 56

Sanglier M. 2
Santos E. 56
Sayer R. 2
Scholten H. 56
See L. 6, 354, 359
Shaw H. 241
Shaw M. 92
Shearer J. 193
Sheffield J. 354
Shepherd J. 6, 342, 343, 346, 347
Sheerwood N. 11
Shevky E. 76
Signoretta P. 5, 138, 140, 147
Simmonds D. 5, 197
Simpson B. 319
Simpson L. 384
Singleton P. 236
Skinner A. 5, 197
Sleight P. 94
Slocum T. 28, 29
Smith D. 236
Smith J. 172, 181, 184, 187, 190, 191
Smith M. 56
Smith R. 98
Soldati A. 354
Sole R. 234
Sorooshian S. 354
Spellerberg I. 234, 235
Spiess H. 192
Srinivasan A. 244, 245
Srivastava M. 162, 163
Steinitz C. 215
Stephens P. 29
Stevens B. 1
Stewart M. 354
Stillwell J. 3, 6, 37, 202, 266, 267, 269, 274, 288, 290, 291, 292, 387, 388
Stokelj T. 354
Stylianidis E. 29
Sun Y. 165
Sutcliffe J. 239, 246, 362
Swart J. 234
Sweeney S. 311

Tango T. 158, 165
Tawn J. 362
Taylor I. 139
Taylor J. 297
Thomas R. 316
Thompson W. 237, 238
Thrall G. 4, 72, 73, 74, 76, 77, 82, 83
Thrall S. 72, 73, 74, 76, 77, 82, 83
Thrift N. 50, 95
Thurik A. 302
Tie X. 247
Tirtiroglu D. 74
Tischendorf L. 239, 251
Toppen F. 221
Tomlinson R. 215
Tong C. 192
Tornqvist G. 218, 219
Townsend P. 92, 138, 322
Tremayne A. 266, 290, 291, 292
Trombulak 234, 235, 236
Tuma N. 306
Tversky B. 29

Unwin D. 97
Urbanski J. 237

van Dijk J. 298
Van Dyke F. 241
Van Kalken T. 355
van Wissen L. 6, 298, 311, 312
Verbyla D. 243

Verroen E. 218
Vidacovic V. 218
Viel J. 152
Vince J. 28
Virdee S. 320
Voas D. 94, 95

Wachter S. 74
Wade P. 237
Wakefield J. 152
Walker J. 315
Waller J. 251
Warm D. 45
Warntz W. 244
Waters N. 236, 237, 244, 251
Waugh D. 29
Weaver J. 234, 248
Weiss M. 76
Wessels C. 216
Whitehead P. 355
Wilby R. 354
Williams J. 51
Williams M. 76
Williamson P. 94, 95
Wilson A. 2, 3, 36, 47, 49, 306, 380, 381
Wilson T. 266, 288
Winter S. 216, 298
Wittinger W. 251
Wong S. 192
Woodhouse P. 139
Worsley K. 162, 163, 164
Wrigley N. 45

Yapa L. 92
Yeates M. 12, 23
Yeh A. 332
Yabunaka K. 355

Zadeh L. 239
Zeigler D. 234, 235

Subject Index

Accessibility (models) 6, 74, 217–218, 215–232, 383–384, 393
ACORN 76–85, 94, 99
Address-Point 99–100
Age Concern 384
Agglomeration (spatial) 305, 312
Aggregation effects 96, 100, 114–122, 141, 147, 218, 227
Akaike's information criterion 237
Alliance and Leicester 37
Ambulance planning 5, 56, 215–232
ARC/INFO 37, (ArcView) 152, 222, 224–225, (AMLs) 119–120, (ArcPlot) 327
Armax model 354, 355, 368, 369, 370
Arts Council 386
ASDA 36
Autocorrelation (see Spatial autocorrelation)
Automall 42
Automated banks 16, 36, 45, 46, 340

Backpropogation 358–369
Barclay's Bank 381, 386
Barrier effect (ecology) 234
Baynesian modelling 235, 237–239
Benefit agencies 325, 330–331, 348
Besag-Newall Test 147
Biased data 75
Big-box retailing 14, 15–21, 31
'Booz, Allen & Hamilton' 172
Bonferroni Adjustment 160, 163, 165
Breadline Britain index 92, 94, 95, 107
BP 37
Buffalo Police Department 152
Buffer and overlay analysis 21, 37, 193, 219, 388 (see also GIS)
Business geomatics 12

CACI 76, 336, 339 (see also, ACORN)
Calibration 153, 163, 265–266, 268, 277, 284–285, 287–288, 354, 355
Carthago 216
Carrying capacity 6, 298–312
Catchment area analysis 11–34, 37, 48
Central Scotland Transport Corridor 213
'Chapters' 16
Child Protection Register 143, 144
Child Service Planning Guidelines 142
Children and Young People Service Plan 137–138, 140, 141
Chi-square tests 153, 163
Claritas 12, 76
Cluster analysis 73, 92, 101–107, 147, 164, 165, 311
Community Health Sheffield 138
Competing destinations model 42, 267
Comprehensive models 2, 5, 37, 195–214
Compusearch 12
Coopers & Lybrand Deloitte 381
Countryside Agencies 335–340, 347
Credit scoring 50
Crime analysis (models) 4, 71, 73, 139–143, 151–168
Critical human geography 2
Critiques of models 2, 393
CSCA 16, 20, 29
CSIRO 172, 191
CUDEM 381
CURDS 378
Customer spotting 12

Dasymetric mapping 97
Data animation 28, 29
Data fusion 94
Data Protection Act 140, 148
Data mining 4, 27–29
Data sharing 143

Applied GIS and Spatial Analysis. Edited by J. Stillwell and G. Clarke
Copyright © 2004 John Wiley & Sons, Ltd. ISBN: 0-470-84409-4

Data warehouse 28
DEFRA 354
DELTA model 5, 197
Demography of firms 298, 299
Dempster-Shafer technique 235, 237, 239, 244–247
Density dependence model 298
Department of Employment 386 (see also Index of Local Conditions)
DETR 195, 336, 341, 383 (see also Index of Local Deprivation)
Digital elevation models 244
Digital mapping 12, 15, 22, 27–30, 57
Dijsktra alogorithm 61, 180
Direct marketing 12
Divorce courts 330
DiWay 55–70
Dixons 36
DTI 386
DTLR 6, 138, 142, 259, 262, 264
Dynamics 28–29

Ecological fallacy 243
Ecology models 299
E-Commerce 56
Economic base models 297
Education (models) 3, 141–143, 137–150, 267, 277, 283, 337, 339, 349, 385, 387–388
Education Action Zones 141
Elastic demand 45–46
Environment Agencies 6, 353–359
Environmental models 233–256, 278
Equilibrium 298, 302, 303, 305–312
ESRC 3, 378, 380, 384
ESRI 76, 77, 84
EuroDirect 99
European Environment Agency 383
European Structural Funds 111–116, 128, 148, 196
EuroStat 383
Exxon/Mobil 36, 37, 48–52

Family HAS 265, 269, 273–275, 280–286
Family Law Act 315
Firm demography 298, 299
Flood forecasting 6, 353–374
Flowmap 216–217, 224, 230
Fluid mapping 62–64
Food deserts 44
Ford 37
Forecasting 199–200, 354
Fuzzy set theory 239

GDP 275
Genetic algorithms 51 (see also Neural nets)
Geobusiness Solutions 37
Geocoding (or georeferencing) 57, 64, 388
Geodemographics 3, 12, 71, 71–86, 89–110, 393
Geometric growth model 300
GeoICT 215–217, 226, 230
GeoTools 288
Geovisualisation 27–29
GIS 3–6, 11, 12, 21, 22, 25, 28–31, 37, 57, 71–73, 76, 119–121, 139–143, 148, 151, 152, 172, 180, 192, 193, 197, 219, 220, 224, 227, 230, 233–256, 317, 327–330, 336, 338, 339–342, 344, 346–352, 371, 377–396
GMAP 3, 36–40, 42, 51, 380, 381, 390, 394
Gold Coast City Council 172, 174, 182, 183
Gold Coast City Transport Plan 172
Government Office for North West 384
Government office for Yorkshire & Humberside 196, 386–387
GPS 31, 58, 241
Gravity model 37, 42, 220 (see also Spatial interaction models)

Habitat selection modelling 236–237
Halifax Building Society 37, 381
Health (models) 3, 138, 137–150, 337, 339
Hedonic (price) models 72–85
Hierarchical choice models 42
Home Depot 15–21
Home Office 383
Hot spots 14, 160, 161, 236
Household projection models 262
Housing models 277–278
Humanism 2

Iceland 36
IDRISI 241–248
Ikea 36
Income 5, 76, 93, 98–110, 277, 287
Index of Local Conditions 388
Index of Local Deprivation 114, 116–134, 138, 142, 341
Industrial demography 298
Information meetings (divorce) 315–316, 317–334
Input-output models 6, 202, 297, 299, 302–306, 310, 311
Inter-industry linkages 301–302, 303, 305
Internet GIS 216
Interpolation 73

Jarman scores 92
Job Centres 316, 348

Joseph Rowntree Foundation 393
Journey to work models 47, 202, 216

Kernel smoothing 163, 163

Labour market modelling 277
Lagged variables 285
Lancaster Urban Archaeology Database 383
Landsat TM imagery 244
Leeds Development Corporation 386
Leeds TEC 386
Lifestyle databases 5, 73–84, 89–110
LISA 311
LITRES-2 172–193
Local spatial association 161–166
Location-allocation models 3, 6, 316–334, 393
Logistic growth model 300, 301, 303, 305–306
Lord Chancellor's Department 316, 317, 318, 327

'M' test 160–161, 163, 164
Mall management 23–27
MapInfo 64, 197
Market share analysis 12, 21–23, 28, 29, 31, 39–40
Marks and Spencers 37
Marxist geography 2
Mass appraisal models 71–86
MAUP (see aggregation)
Mazda 37
Mergers and acquisitions 36
MFI 36
Microsimulation 175–193
Midlands RRL 379
MIGMOD 259–260, 284, 288–292
Migration (models) 6, 260–296
Mobile phones 4, 55–70
Model accuracy 48–52
Model calibration (see Calibration)
Model disaggregation 38
Moran's I Test 156–158
MOSAIC 76
Multicollinearity 277
Multivariate analysis 36, 49–52, 145
MVA 197

Nearest neighbour analysis 101
'Needs Analysis and Mapping Project' 137–150
Neighbourhood Statistic Service 91, 141
Neural nets 6, 51, 353–374
NERC 380
NERRL 379
Network analysis 3–4, 60–62, 222, 215–232
Nexpri 216–217, 377
NHS 384
NHSCR 265–267, 269, 273, 274, 276, 280
N. Ireland RRL 379
Northern Foods 381
NUTS regions 114
NWRRL 378, 379, 382–385, 386, 389–395

'Office Place' 16
OLS 284, 287
ONS 269, 273, 384
OPCS 386
Optimisation 4, 51, 218–219, 310
Organisational ecology 298
Oscillatory behaviour 308–309, 312
OS Master Map 90, 100

PAT18 91, 107, 141, 144, 145
Performance analysis (retail) 12, 13, 23–27, 48–52
Personal computers 3, 37
Petrol retail modelling 48–52
'PETsMART' 16, 20
Phonemarks 66–68
Poisson regression 287
Population projections 384–385, 387
Portbury Dock Study 1–2
Postal geography 12, 13, 142–143
Postcode Address File 339
Potential values 219–220
Power centres 15–21, 31
'Price Costco' 16
Price (modelling) 49, 72–85
PriceWaterhouseCoopers 4, 72, 377
Prince's Trust 140–142, 386
Public policy models (especially tax) 71, 278–279

Q2 statistics 162–163
Queensland Transport Department 171
Quickprop 358–369

'Regent Taxis' 184
Regional Assemblies 384, 386
Regional development agencies 374, 386
Regional growth models 6, 205, 297–314
Regional multiplier model 297–314 (see also Input-output models, CGE models)
Regional Planning Guidance 387
Regional Review 385, 391
Regional science 2

Regression 6, 49, 50, 72, 74–84, 95, 238, 280–281, 283, 285, 287, 320, 321, 349, 368–370
Remote sensing 5, 31, 90, 97, 98,
Residential delivery points 346–348
Retail (models) 2–4, 11–54 (see also Performance analysis)
Revenue Canada 13
RRLs 3, 7, 377, 378–380, 382–385, 386, 389–395
Rotterdam Municipal Health Authority 217, 221, 226
Royal Mail Address Manager 340, 346
Rural Consultative Councils 337
Rural Development Commission 337
Rural service provision 335–352
Rural Services Survey 338–340
Rural White Paper (UK) 336, 337, 342–343
Ryerson University 16

SAC-SMA model 354
Sainsbury's 37
SARTRE 13
Satellite imagery (see remote sensing)
Scottish RRL 379
SERRL 377, 379
Sheffield Centre for Geographical Information 138
Sheffield DSS 138
Sheffield Health Authority 138
Silicon Graphics Ltd 15, 27
Simulation models 5, 164, 175–193
Site ratings 49–52
Social area analysis 76
Social deprivation 3, 4, 91–110, 111–136, 141, 142, 143–150 (see also Breadline Britain, Townsend Index, Index of Local Deprivation, Income)
SONAE 4, 56, 57
Source–sink models 176, 180
South Yorkshire Police 138
Spatial agglomeration (see Agglomeration)
Spatial autocorrelation 75, 76, 161
Spatial barcodes 59
Spatial decision support systems 28, 30
Spatial demand field 303–304
Spatial interaction models 5, 6, 35–54, 286–287, 303–306, 310, 381, 393
Spatial laws 1
Spatial multimedia systems 55–70
Spatial outliers 163–165
Spatial scan statistics 164
Spatial smoothing 83, 163, 164
Special Migration Statistics 273–274
S-PLUS 152,165
SPSSX 323
Standard spending assessment 383–384
START 197, 199–214
State Bank of South Australia 36, 38–40
Statistics Canada 13
'Surfside Bus Company' 183
Sustainable development 174, 386

Tasselated Cap Analysis 244
Tesco 37, 44–45
Tessellation 227
Thomas Cook 37
TIGER/Line files 76–85
Townsend Index 92, 138, 139, 141, 143, 322, 324
Toyota 37
Transport models 1, 3, 5, 37, 48, 172–193, 195–214
Two speed cities 4, 112–113, 115

University of Bradford 381
University of Buffalo 152
University of Huddersfield 381
University of Hull 381
University of Lancaster 7 (see also NWRRL)
University of Leeds 3, 7, 36, 259, 288, 359, 380, 382
University of Lincoln 381
University of Newcastle 259
University of Pennsylvania 1
University of Sheffield 381
University of Utrecht 216, 217
University of York 381
Urban and Regional Policy RRL 379
Urban gatekeepers 2
US Postal Service 82

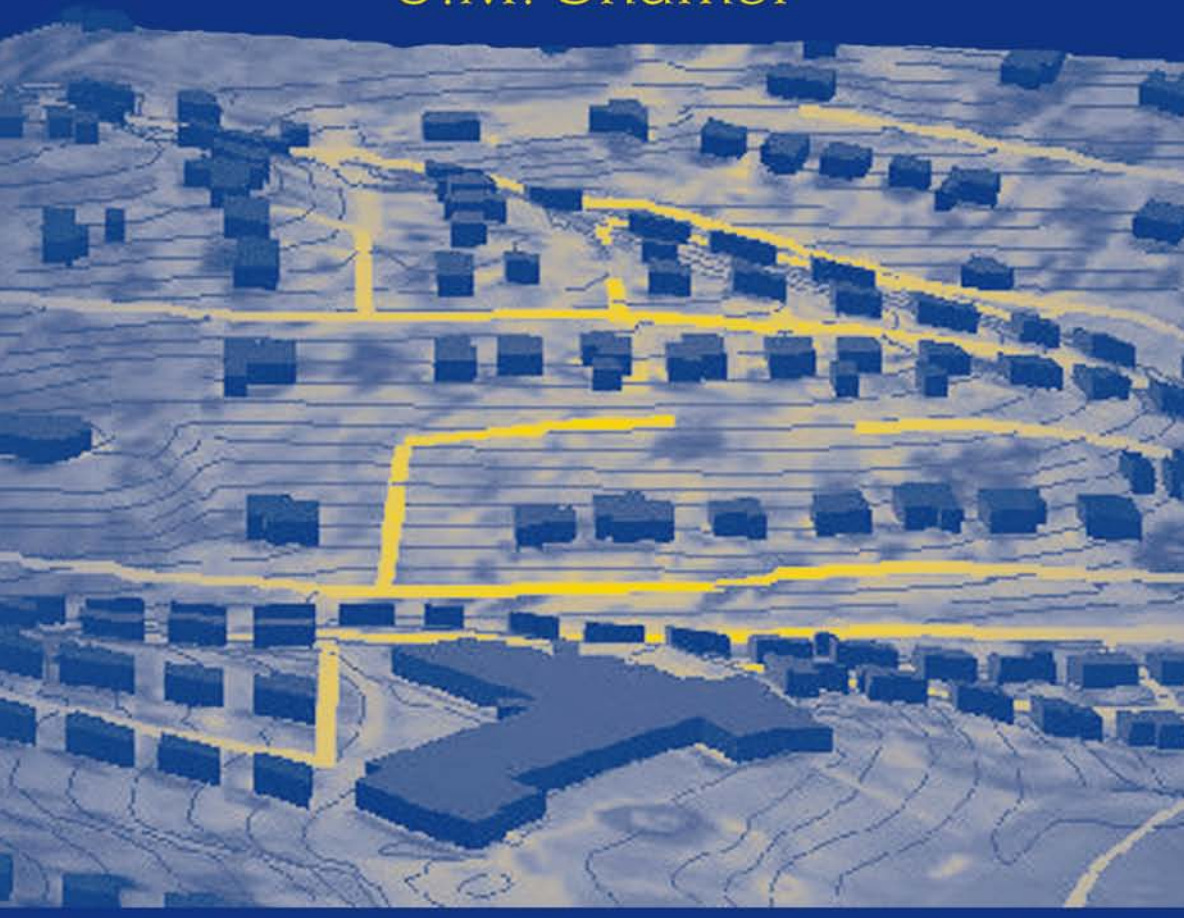
Vacancy rates 16
VisualBASIC 359
Volvo 37

Wales & SWRRL 379
Wal-Mart 15–21, 36
WAP (see Mobile phones)
Whitbread 37
W.H. Smith 36, 47
WLS 284
Work-based trips 46–47, 218

YHDA 386
YHRRO 7, 378, 380–382, 385–395
YHUA 382, 386

GIS Applications for Water, Wastewater, and Stormwater Systems

U.M. Shamsi



A CRC PRESS BOOK

GIS Applications for Water, Wastewater, and Stormwater Systems

U.M. Shamsi

Taylor & Francis
Taylor & Francis Group

Boca Raton London New York Singapore

A CRC title, part of the Taylor & Francis imprint, a member of the
Taylor & Francis Group, the academic division of T&F Informa plc.

Library of Congress Cataloging-in-Publication Data

Shamsi, U. M. (Uzair M.)
GIS applications for water, wastewater, and stormwater systems / U.M. Shamsi.
p. cm.
Includes bibliographical references and index.
ISBN 0-8493-2097-6 (alk. paper)
1. Water—Distribution. 2. Sewage disposal. 3. Runoff—Management. 4. Geographic information systems. I. Title.

TD482.S53 2005
628.1—dc22 2004057108

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

The consent of CRC Press does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press for such copying.

Direct all inquiries to CRC Press, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

Visit the CRC Press Web site at www.crcpress.com

© 2005 by CRC Press

No claim to original U.S. Government works
International Standard Book Number 0-8493-2097-6
Library of Congress Card Number 2004057108
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Dedication

Dedicated to my beloved wife, Roshi, and my children, Maria, Adam, and Harris

Preface

To fully appreciate the benefits of GIS applications consider the following hypothetical scenario. On March 10, 2004, following a heavy storm event, a sewer customer calls the Sewer Authority of the City of Cleanwater to report minor basement flooding without any property damage. An Authority operator immediately starts the GIS and enters the customer address. GIS zooms to the resident property and shows all the sewers and manholes in the area. The operator queries the inspection data for a sewer segment adjacent to the customer property and finds that a mini movie of the closed-circuit television (CCTV) inspection dated July 10, 1998, is available. The operator plays the movie and sees light root growth in the segment. A query of the maintenance history for that segment indicates that it has not been cleaned since April 5, 1997. This information indicates that the roots were never cleaned and have probably grown to "heavy" status. The operator highlights the sewer segment, launches the work order module, and completes a work order form for CCTV inspection and root removal, if necessary. The export button saves the work order form and a map of the property and adjacent sewers in a PDF file. The operator immediately sends the PDF file by e-mail to the Authority's sewer cleaning contractor. The entire session from the time the customer called the Authority office took about 30 min. The operator does not forget to call the customer to tell him that a work order has been issued to study the problem. This book presents the methods and examples required to develop applications such as this.

The days of the slide rule are long gone. Word processors are no longer considered cutting-edge technology. We are living in an information age that requires us to be more than visionaries who can sketch an efficient infrastructure plan. This tech-heavy society expects us to be excellent communicators who can keep all the stakeholders — the public, the regulators, or the clients — "informed." New information and decision support systems have been developed to help us to be good communicators. GIS is one such tool that helps us to communicate geographic or spatial information. The real strength of GIS is its ability to integrate information. GIS helps decision makers by pulling together crucial bits and pieces of information as a whole and showing them the "big picture." In the past 10 years, the number of GIS users has increased substantially. Many of us are using GIS applications on the Internet and on wireless devices without even knowing that we are using a GIS. Experts believe that in the near future, most water, wastewater, and stormwater system professionals will be using the GIS in the same way they are now using a word processor or spreadsheet. Except for the computer itself, no technology has so revolutionized the water industry. The time has come for all the professionals involved in the planning, design, construction, and operation of water, wastewater, and stormwater systems to enter one of the most promising and exciting technologies of the millennium in their profession — GIS applications.

According to some estimates, more than 80% of all the information used by water and sewer utilities is geographically referenced.

This book was inspired from a continuing education course that the author has been teaching since 1998 for the American Society of Civil Engineers (ASCE). Entitled ''GIS Applications in Water, Wastewater and Stormwater Systems,'' the seminar course has been attended by hundreds of water, wastewater, and stormwater professionals in major cities of the United States. Many models, software, examples, and case studies described in the book (especially those from Pennsylvania) are based on the GIS projects worked on or managed by the author himself.

This is my second GIS book for water, wastewater, and stormwater systems. The first book, *GIS Tools for Water, Wastewater, and Stormwater Systems*, published by American Society of Civil Engineers (ASCE) Press in 2002, was a huge success. The first printing was sold out, and the book achieved ASCE Press's best-seller status within months of publication. Whereas the first book focused on GIS basics and software and data tools to develop GIS applications, this second book focuses on the practical applications of those tools. Despite the similarity of the titles, both books cover different topics and can be read independent of each other.

STYLE OF THE BOOK

This book has been written using the recommendations of the Accreditation Board for Engineering and Technology (ABET) of the U.S. and the American Society of Civil Engineers' (ASCE) Excellence in Civil Engineering Education (ExCEEd) program. Both of these organizations recommend performance- (or outcome-) based learning in which the learning objectives of each lecture (or chapter) are clearly stated up front, and the learning is measured in terms of achieving these learning objectives. Each chapter of this book accordingly starts with learning objectives for that chapter and ends with a chapter summary and questions. Most technical books are written using the natural human teaching style called *deductive*, in which principles are presented before the applications. In this book, an attempt has been made to organize the material in the natural human learning style called *inductive*, in which examples are presented before the principles. For example, in most chapters, case studies are presented before the procedures are explained. The book has numerous maps and illustrations that should cater well to the learning styles of "visual learners" — GIS, after all is regarded as a visual language.

The primary learning objective of this book is to document GIS applications for water, wastewater, and stormwater systems. This book will show you how to use GIS to make tasks easier to do and increase productivity, and hence, save time and money in your business.

ORGANIZATION OF THE BOOK

There are 17 chapters in this book, organized as follows:

- Chapter 1, GIS Applications: Describes why GIS applications are important and how they are created

- Chapter 2, Needs Analysis: Explains how to avoid potential pitfalls of GIS implementation by starting with a needs analysis study

The next five chapters describe four GIS-related technologies that are very beneficial in developing GIS applications:

- Chapter 3, Remote Sensing Applications: Shows how to use satellite imagery in GIS applications
- Chapter 4, DEM Applications: Describes the methods of incorporating digital elevation model (DEM) data
- Chapter 5, GPS Applications: Discusses how to benefit from global positioning system (GPS) technology
- Chapter 6, Internet Applications: Explains the applications of Internet technology in serving GIS maps on the Internet
- Chapter 7, Mobile GIS: Provides information on using GIS in the field for inspection and maintenance work

The GIS applications that are of particular importance to water industry professionals are: Mapping, Monitoring, Modeling, and Maintenance. These four *M*s define some of the most important activities for efficient management of water, wastewater, and stormwater systems, and are referred to as the "4M applications" in this book. The next ten chapters focus on these four *M*s.

- Chapter 8, Mapping: Describes how to create the first *M* of the 4M applications
- Chapter 9, Mapping Applications: Describes examples of the first *M* of the 4M applications
- Chapter 10, Monitoring Applications: Describes the applications of the second *M* of the 4M applications
- Chapter 11, Modeling Applications: Describes the applications of the third *M* of the 4M applications
- Chapter 12, Water Models: Describes examples of the third *M* of the 4M applications for modeling water distribution systems
- Chapter 13, Sewer Models: Describes examples of the third *M* of the 4M applications for modeling sewage collection systems
- Chapter 14, AM/FM/GIS Applications: Describes automated mapping/facilities management/geographic information system (AM/FM/GIS) software tools for implementing the fourth *M* of the 4M applications
- Chapter 15, Maintenance Applications: Describes the applications of the fourth *M* of the 4M applications
- Chapter 16, Security Planning and Vulnerability Assessment: Discusses GIS applications for protecting water and wastewater systems against potential terrorist attacks
- Chapter 17, Applications Sampler: Presents a collection of recent case studies from around the world

Acknowledgments

Case studies presented in Chapter 17, Applications Sampler, were written specially for publication in this book by 18 GIS and water industry experts from 6 countries (Belgium, Bulgaria, Czech Republic, Denmark, Spain and the United States) in response to my call for case studies distributed to various Internet discussion forums. I thank these case study authors for their contributions to this book:

- Bart Reynaert, Rene Horemans, and Patrick Vercruyssen of Pidpa, Belgium
- Carl W. Chen and Curtis Loeb of Systech Engineering, Inc., San Ramon, California
- Dean Trammel, Tucson Water, Tucson, Arizona
- Ed Bradford, Roger Watson, Eric Mann, Jenny Konwinski of Metropolitan Sewerage District of Buncombe County, North Carolina
- Eric Fontenot of DHI, Inc., Hørsholm, Denmark
- Milan Suchanek and Tomas Metelka of Sofiyska Voda A.D., Sofia, Bulgaria
- Peter Ingeduld, Zdenek Svitak, and Josef Drbohlav of Praûská vodohospodáská spolenost a.s. (Prague stockholding company), Prague, Czech Republic
- Hugo Bartolin and Fernando Martinez of Polytechnic University of Valencia, Spain

I also thank the following organizations and companies for providing information for this book: American Society of Civil Engineers, American Water Works Association, Azteca Systems, *CE Magazine*, CH2M Hill, Chester Engineers, Computational Hydraulics International, Danish Hydraulic Institute (DHI), Environmental Systems Research Institute, *Geospatial Solutions Magazine*, *GEOWorld Magazine*, Haestad Methods, Hansen Information Technology, *Journal of the American Water Resources Association*, *Journal of the American Water Works Association*, MWH Soft, *Professional Surveyor Magazine*, USFilter, *Water Environment Federation, and Water Environment & Technology Magazine*. Some information presented in this book is based on my collection of papers and articles published in peer-reviewed journals, trade magazines, conference proceedings, and the Internet. The authors and organizations of these publications are too numerous to be thanked individually, so I thank them all collectively without mentioning their names. Their names are, of course, included in the Reference section.

Finally, I would like to thank you for buying the book. I hope you will find the book useful in maximizing the use of GIS in your organization to make things easier to do, increase productivity, and save time and money.

About the Author

Uzair (Sam) M. Shamsi, Ph.D., P.E., DEE is director of the GIS and Information Management Technology division of Chester Engineers, Pittsburgh, Pennsylvania, and an adjunct assistant professor at the University of Pittsburgh, where he teaches GIS and hydrology courses. His areas of specialization include GIS applications and hydrologic and hydraulic (H&H) modeling. He has been continuing education instructor for the American Society of Civil Engineers (ASCE) and an Environmental Systems Research Institute (ESRI)-authorized ArcView® GIS instructor since 1998. He has taught GIS courses to more than 500 professionals throughout the United States, including a course on "GIS Applications in Water, Wastewater, and Stormwater Systems" for ASCE. Sam earned his Ph.D. in civil engineering from the University of Pittsburgh in 1988. He has 20 years of GIS and water and wastewater engineering experience in teaching, research, and consulting. His accomplishments include more than 120 projects and over 100 lectures and publications, mostly in GIS applications. His previous book, *GIS Tools for Water, Wastewater, and Stormwater Systems*, was an ASCE Press best seller. He is the recipient of the ASCE's Excellence in Civil Engineering Education (EXCEED) training and is a licensed professional engineer in Pennsylvania, Ohio, and West Virginia. In addition to ASCE, he is a member of the American Water Resources Association, the Water Environment Foundation, and the American Water Works Association.

E-mail: ushamsi@GISApplications.com
Web site: www.GISApplications.com

GIS is an instrument for implementing geographic thinking.

Jack Dangermond (1998)

Iron rusts from disuse; water loses its purity from stagnation and in cold weather becomes frozen; even so does inaction sap the vigors of the mind.

Leonardo da Vinci (1452–1519)

Life is like a sewer… what you get out of it depends on what you put into it.

Tom Lehrer (1928–)

Times of general calamity and confusion create great minds. The purest ore is produced from the hottest furnace, and the brightest thunderbolt is elicited from the darkest storms.

Charles Caleb Colton (1780–1832)

Contents

Chapter 1 GIS Applications
Learning Objective 2
Major Topics 2
List of Chapter Acronyms 2
Introduction 2
What Are GIS Applications? 3
History of GIS Applications 4
4M Applications 6
Advantages and Disadvantages of GIS Applications 6
Advantages 7
GIS Applications Save Time and Money 7
GIS Applications Are Critical to Sustaining GIS Departments 7
GIS Applications Provide the Power of Integration 8
GIS Applications Offer a Decision Support Framework 8
GIS Applications Provide Effective Communication Tools 9
GIS Applications Are Numerous 9
Disadvantages 12
Success Stories 13
San Diego 13
Boston 13
Cincinnati 13
Knoxville 14
Dover 14
Charlotte 14
Albany County 14
GIS Applications Around the World 15
Evolving GIS Applications and Trends 15
Future Applications and Trends 16
GIS Application Development Procedure 19
Application Programming 20
GIS-Based Approach 20
GIS Customization 20
Scripting 20
Extensions 21
External Programs 23
Application-Based Approach 24
Useful Web Sites 24
Chapter Summary 24
Chapter Questions 25

Chapter 2 Needs Analysis
Learning Objective 28
Major Topics 28

List of Chapter Acronyms 28
Ocean County's Strategic Plan 28
Introduction 28
Needs Analysis Steps 29
Step 1. Stakeholder Identification 30
Step 2. Stakeholder Communication 30
Introductory Seminar 31
Work Sessions and Focus Groups 31
Interviews 31
Step 3. Resource Inventory 32
Step 4. Need Priorities 33
Step 5. System Design 33
Data Conversion (Mapping) 33
Database 34
Software Selection 36
Hardware Selection 37
User Interface 38
Step 6. Pilot Project 40
Step 7. Implementation Plan 41
Step 8. Final Presentation 43
Needs Analysis Examples 43
Pittsburgh, Pennsylvania 43
Borough of Ramsey, New Jersey 44
The City of Bloomington, Indiana 45
San Mateo County, California 45
Chapter Summary 45
Chapter Questions 46

Chapter 3 Remote Sensing Applications
Learning Objective 48
Major Topics 48
List of Chapter Acronyms 48
Albany County's Remote Sensing Application 48
Introduction 49
Remote Sensing Applications 51
Remote Sensing Satellites 52
Spatial Resolution 53
Low-Resolution Satellite Data 53
Medium-Resolution Satellite Data 54
High-Resolution Satellite Data 56
High-Resolution Satellites 56
High-Resolution Imagery Applications 58
Data Sources 59
Digital Orthophotos 59
USGS Digital Orthophotos 60
Case Study: Draping DOQQ Imagery on DEM Data 62

Examples of Remote Sensing Applications 62
LULC Classification 62
Soil Moisture Mapping 65
Estimating Meteorological Data 66
Geographic Imaging and Image Processing Software 66
ERDAS Software Products 66
ERDAS Software Application Example 68
ArcView Image Analysis Extension 69
MrSID 69
PCI Geomatics 70
Blue Marble Geographics 71
Future Directions 72
Useful Web Sites 73
Chapter Summary 73
Chapter Questions 73

Chapter 4 DEM Applications
Learning Objective 76
Major Topics 76
List of Chapter Acronyms 76
Hydrologic Modeling of the Buffalo Bayou Using GIS and DEM Data 76
DEM Basics 77
DEM Applications 79
Three-Dimensional (3D) Visualization 79
DEM Resolution and Accuracy 80
USGS DEMs 81
USGS DEM Formats 82
National Elevation Dataset (NED) 83
DEM Data Availability 83
DEM Data Creation from Remote Sensing 84
Image Processing Method 84
Data Collection Method 84
LIDAR 85
IFSAR 85
DEM Analysis 86
Cell Threshold for Defining Streams 86
The D-8 Model 86
DEM Sinks 87
Stream Burning 88
DEM Aggregation 88
Slope Calculations 88
Software Tools 88
Spatial Analyst and Hydro Extension 90
ARC GRID Extension 93
IDRISI 94
TOPAZ 95

Case Studies and Examples 95
Watershed Delineation 95
Sewershed Delineation 101
Water Distribution System Modeling 103
WaterCAD Example 104
Useful Web Sites 105
Chapter Summary 105
Chapter Questions 106

Chapter 5 GPS Applications
Learning Objective 108
Major Topics 108
List of Chapter Acronyms 108
Stream Mapping in Iowa 108
GPS Basics 109
GPS Applications in the Water Industry 110
Surveying 111
Fleet Management 111
GPS Applications in GIS 111
GPS Survey Steps 112
GPS Equipment 113
Recreational GPS Equipment 113
Basic GPS Equipment 114
Advanced GPS Equipment 115
Survey Grade GPS Equipment 116
Useful Web Sites 117
Chapter Summary 117
Chapter Questions 118

Chapter 6 Internet Applications
Learning Objective 120
Major Topics 120
List of Chapter Acronyms 120
Dublin's Web Map 120
Internet GIS 122
Internet Security 123
Internet GIS Software 124
Internet GIS Applications 124
Data Integration 124
Project Management 124
3D Visualization Applications 126
Case Studies 126
Tacoma's Intranet and Mobile GIS 126
Montana's Watershed Data Information Management System 127
Useful Web Sites 128

Chapter Summary 128
Chapter Questions 128

Chapter 7 Mobile GIS
Learning Objective 130
Major Topics 130
List of Chapter Acronyms 130
Mobile GIS Basics 130
Mobile GIS Applications 131
Wireless Internet Technology 133
GPS Integration 133
Useful Web Sites 134
Chapter Summary 135
Chapter Questions 135

Chapter 8 Mapping
Learning Objective 138
Major Topics 138
List of Chapter Acronyms 138
Los Angeles County's Sewer Mapping Program 138
Mapping Basics 139
- Map Types 139
- Topology 139
- Map Projections and Coordinate Systems 140
- Map Scale 140
- Data Quality 140
- Data Errors 141
- Map Accuracy 141

Map Types 142
- Base Map 142
  - Digital Orthophotos 143
  - Planimetric Maps 143
  - Small-Scale Maps 144

Advantages of GIS Maps 145
GIS Mapping Steps 147
- Needs Analysis 147
- Data Collection 148
- Data Conversion 148
  - Capturing Attributes 148
  - Capturing Graphics 149
    - Digitization 149
    - Scanning 150
    - Data Conversion Software 150
- Data Processing 153
  - Data Preparation 153
  - Topological Structuring 153

Data Management 154
Quality Control 155
Map Production 155
Case Studies 156
Borough of Ramsey, New Jersey 156
City of Lake Elsinore, California 158
Allegheny County, Pennsylvania 159
Useful Web Sites 159
Chapter Summary 160
Chapter Questions 160

Chapter 9 Mapping Applications
Learning Objective 162
Major Topics 162
List of Chapter Acronyms 162
Customer Service Application in Gurnee 162
Common Mapping Functions 164
Thematic Mapping 164
Spatial Analysis 164
Buffers 164
Hyperlinks 167
Water System Mapping Applications 167
MWRA Water System Mapping Project 167
Service Shutoff Application 167
Generating Meter-Reading Routes 169
Map Maintenance Application 169
Wastewater System Mapping Applications 169
Public Participation with 3D GIS 169
Mapping the Service Laterals 170
Stormwater System Mapping Applications 173
Stormwater Permits 173
Chapter Summary 175
Chapter Questions 175

Chapter 10 Monitoring Applications
Learning Objective 178
Major Topics 178
List of Chapter Acronyms 178
Monitoring Real Time Rainfall and Stream-Flow Data in Aurora 178
Monitoring Basics 179
Remotely Sensed Rainfall Data 179
Satellite Rainfall Data 180
Radar Rainfall Data 181
NEXRAD Rainfall Data 181
NEXRAD Level III Data 181
Estimating Rainfall Using GIS 183

Radar Rainfall Application: Virtual Rain-Gauge Case Study 184
Flow-Monitoring Applications 187
SCADA Integration 187
NPDES-Permit Reporting Applications 188
Monitoring via Internet 189
Monitoring the Infrastructure 190
Useful Web Sites 190
Chapter Summary 191
Chapter Questions 191

Chapter 11 Modeling Applications
Learning Objectives 194
Major Topics 194
List of Chapter Acronyms 194
Temporal-Spatial Modeling in Westchester County 194
H&H Modeling 195
Application Methods 196
Interchange Method 197
Subbasin Parameter Estimation 198
Runoff Curve Number Estimation 199
Water Quality Modeling Data Estimation 200
Demographic Data Estimation 202
Land-Use Data Estimation 204
Interface Method 205
HEC-GEO Interface 207
HEC-GeoHMS 207
HEC-GeoRAS 207
Watershed Modeling System 208
GISHydro Modules 208
GISHydro Prepro 209
GISHydro Runoff 210
ArcInfo Interface with HEC Programs 210
Intermediate Data Management Programs 211
Interface Method Case Study 212
Integration Method 212
EPA's BASINS Program 213
BASINS Examples 217
MIKE BASIN 218
Geo-STORM Integration 219
ARC/HEC-2 Integration 219
Integration Method Case Study 220
Which Linkage Method to Use? 221
Useful Web Sites 222
Chapter Summary 222
Chapter Questions 223

Chapter 12 Water Models
Learning Objective 226
Major Topics 226
List of Chapter Acronyms 226
City of Germantown's Water Model 226
GIS Applications for Water Distribution Systems 227
Development of Hydraulic Models 229
Software Examples 231
EPANET 231
H_2ONET^{TM} and H_2OMAP^{TM} 232
Demand Allocator 235
Skeletonizer 235
Tracer 235
WaterCAD™ and WaterGEMS™ 235
MIKE NET™ 236
Other Programs 237
EPANET and ArcView Integration in Harrisburg 237
Mapping the Model Output Results 242
Network Skeletonization 243
Estimation of Node Demands 249
Demand-Estimation Case Studies 252
Newport News, Virginia 252
Round Rock, Texas 252
Lower Colorado River Authority, Texas 253
Estimation of Node Elevations 253
Pressure Zone Trace 255
Chapter Summary 255
Chapter Questions 255

Chapter 13 Sewer Models
Learning Objectives 258
Major Topics 258
List of Chapter Acronyms 258
MapInfo™ and SWMM Interchange 258
GIS Applications for Sewer Systems 259
Sewer System Modeling Integration 260
Software Examples 261
SWMM 261
Useful SWMM Web Sites 264
SWMM Graphical User Interface 264
XP-SWMM and XP-GIS 266
GIS Data for SWMM 267
Estimating Green-Ampt Parameters Using STATSGO/SSURGO GIS Files 267
GIS Applications for SWMM 270

AVSWMM 270
AVSWMM RUNOFF Extension 271
AVSWMM EXTRAN Extension 274
Task 1: Create EXTRAN input file 274
Task 2: Create SWMM EXTRAN output layers in ArcViewGIS 277
SWMMTools 278
AGSWMM 280
PCSWMM GIS™ 281
SWMM and BASINS 282
SWMMDUET 283
AVsand™ 284
Other Sewer Models 284
DHI Models 284
MOUSE™ 284
MIKE SWMM™ 285
MOUSE GIS™ 285
MOUSE GM™ 286
InfoWorks™ 287
SewerCAD™ and StormCAD™ 289
Sewer Modeling Case Studies 289
XP-SWMM and ArcInfo Application for CSO Modeling 289
AM/FM/GIS and SWMM Integration 290
SWMM and ArcInfo™ Interface 290
Hydra™ and ArcInfo™ Interface 291
Useful Web Sites 291
Chapter Summary 291
Chapter Questions 292

Chapter 14 AM/FM/GIS Applications
Learning Objective 294
Major Topics 294
List of Chapter Acronyms 294
Hampton's Wastewater Maintenance Management 294
Infrastructure Problem 295
AM/FM/GIS Basics 297
Automated Mapping (AM) 298
Facilities Management (FM) 300
Automated Mapping (AM)/Facilities Management (FM) 300
AM/FM/GIS Systems 300
AM/FM/GIS Software 300
ArcFM 302
Cityworks 304
Chapter Summary 305
Chapter Questions 305

Chapter 15 Maintenance Applications
Learning Objective 308
Major Topics 308
List of Chapter Acronyms 308
Buncombe County's Sewer System Inspection and Maintenance 309
Asset Management 310
GASB 34 Applications 312
Wet Weather Overflow Management Applications 312
AutoCAD Map GIS Application for CMOM 313
CCTV Inspection of Sewers 314
Convert Existing Video Tapes to Digital Files 315
Digitize Existing VHS Tapes 316
WinCan 317
Retrofit Tape Systems with Digital Systems 317
Record Directly in Digital Format 319
Linking Digital Movies to GIS 319
Video Mapping 321
Thematic Mapping of Inspection Data 322
Work Order Management 325
Water Main Isolation Trace 327
Case Studies 328
Isolation Trace Case Studies 328
Sewer System Inspections in Washington County 328
Sewer Rehabilitation in Baldwin 330
Useful Web Sites 333
Chapter Summary 333
Chapter Questions 333

Chapter 16 Security Planning and Vulnerability Assessment
Learning Objective 336
Major Topics 336
List of Chapter Acronyms 336
GIS Applications in Planning 336
Security Planning 337
Vulnerability of Water Systems 338
Vulnerability of Sewer Systems 338
GIS Applications in Vulnerability Assessment 338
Security Modeling Software 340
H_2OMAP^{TM} Protector 340
WaterSAFE™ 340
VSAT™ 342
Security Planning Data Issues 342
Useful Web Sites 343
Chapter Summary 343
Chapter Questions 343

Chapter 17 Applications Sampler
Learning Objective 346
Major Topics 346
List of Chapter Acronyms 346
Drainage Area Planning in Sofia 346
Pipe Rating Program in Buncombe County 347
Water System Modeling in Tucson 352
Water System Modeling in the City of Truth or Consequences 353
    Background 355
    Building the MIKE NET Model from Various Data Sources 355
ArcGIS and ArcFM Integration in Belgium 356
Water System Master Planning in Prague 358
Water Quality Management in Mecklenburg County 360
Water Master Planning in Sueca, Spain 362
Chapter Summary 364
Chapter Questions 364

Appendix A Acronyms 365
Appendix B Conversion Factors 371
References 373
Index 389

CHAPTER 1

GIS Applications

Geographic Information System (GIS) is one of the most promising and exciting technology of the decade in our profession. This book will show you that with GIS the possibilities to manage your water, wastewater, and stormwater systems are almost endless.

GIS applications can take you from work frustration to job satisfaction.

LEARNING OBJECTIVE

The learning objective of this chapter is to understand the importance and scope of geographic information system (GIS) applications for water, wastewater, and stormwater systems.

MAJOR TOPICS

- Definition of GIS applications
- History of GIS applications
- Advantages and disadvantages of GIS applications
- Evolving and future GIS applications and trends
- Methods of developing GIS applications

LIST OF CHAPTER ACRONYMS*

CAD Computer-Aided Drafting/Computer-Aided Design
ESRI Environmental Systems Research Institute
GIS Geographic Information Systems
GPS Global Positioning System
GUI Graphical User Interface
H&H Hydrologic and Hydraulic
LBS Location-Based Services
PC Personal Computer
PDA Personal Digital Assistant

GIS Project Nominated for OCEA Award

American Society of Civil Engineers (ASCE) awards Outstanding Civil Engineering Achievement (OCEA) awards to projects based on their contribution to the well-being of people and communities; resourcefulness in planning and solving design challenges; pioneering in use of materials and methods; innovations in construction; impact on physical environment; and beneficial effects including aesthetic value. The Adam County (Illinois) 2002 GIS Pilot Project was a nominee for the 1997 awards. This project was a 10-year, multiparticipant (Adams County, City of Quincy, Two Rivers Regional Planning Council, and a number of state and local agencies) project to develop an accurate, updated GIS designed to create a more efficient local government.

INTRODUCTION

The water industry** business is growing throughout the world. For example, the U.S. market for water quality systems and services had a total value of $103 billion in 2000. The two largest components of this business are the $31-billion

* Each chapter of this book begins with a list of frequently used acronyms in the chapter. Appendix A provides a complete list of acronyms used in the book.
** In this book, the term *water industry* refers to water, wastewater, and stormwater systems.

public wastewater treatment market and the $29-billion water supply market (Farkas and Berkowitz, 2001).

One of the biggest challenges in the big cities with aging water, wastewater, and stormwater infrastructures is managing information about maintenance of existing infrastructure and construction of new infrastructure. Many utilities tackle infrastructure problems on a react-to-crisis basis that, despite its conventional wisdom, may not be the best strategy. Making informed infrastructure improvement decisions requires a large amount of diverse information on a continuing basis. If information is the key to fixing infrastructure problems, the first step of any infrastructure improvement project should be the development of an information system.

An information system is a framework that provides answers to questions, from a data resource. A GIS is a special type of information system in which the data source is a database of spatially distributed features and procedures to collect, store, retrieve, analyze, and display geographic data (Shamsi, 2002).

More than 80% of all the information used by water and wastewater utilities is geographically referenced.

In other words, a key element of the information used by utilities is its location relative to other geographic features and objects. GIS technology that offers the combined power of both geography and information systems is an ideal solution for effective management of water industry infrastructure. Geotechnology and geospatial technology are alias names of GIS technology.

The days of the slide rule are long gone. Word processors are no longer considered cutting-edge technology. We are living in the information age, which requires us to be more than visionaries who can sketch an efficient infrastructure plan. Today's tech-savvy society expects us to be excellent communicators who can keep all the stakeholders — the public, the regulators, or the clients — "informed." New information and decision support systems have been developed to help us become good communicators. GIS is one such tool that helps us to communicate geographic or spatial information. In fact, a carefully designed GIS map can be worth more than a thousand words. Sometimes the visual language of GIS allows us to communicate without saying a single word, which is the essence of effective communication.

WHAT ARE GIS APPLICATIONS?

An application is an applied use of a technology. For example, online shopping is an application of Internet technology, automobile navigation is an application of GPS technology, and printing driving-direction maps is an application of GIS technology. No matter how noble a technology is, without applied use it is just a theoretical development. Applications bridge the gap between pure science and applied use. Highly effective water and wastewater utilities strive for continuous operational improvements and service excellence. GIS applications have the potential to enhance

the management of our water, wastewater, and stormwater systems and prepare them for the operational challenges of the 21st century.

HISTORY OF GIS APPLICATIONS

GIS technology was conceived in the 1960s as a digital layering system for coregistered overlays. Started in the mid-1960s and still operating today, Canadian GIS is an example of one of the earliest GIS developments. Civilian GIS in the U.S. got a jump start from the military and intelligence imagery programs of the 1960s. The Internet was started in the 1970s by the U.S. Department of Defense to enable computers and researchers at universities to work together. GIS technology was conceived even before the birth of the Internet.

Just as technology has changed our lifestyles and work habits, it has also changed GIS. Though the art of GIS has been in existence since the 1960s, the science was restricted to skilled GIS professionals. The mid-1990s witnessed the inception of a new generation of user-friendly desktop GIS software packages that transferred the power of GIS technology to the average personal computer (PC) user with entry-level computing skills. In the past decade, powerful workstations and sophisticated software brought GIS capability to off-the-shelf PCs. Today, PC-based GIS implementations are much more affordable and have greatly reduced the cost of GIS applications. Today's GIS users are enjoying faster, cheaper, and easier products than ever before, mainly because of the advent of powerful and affordable hardware and software.

There were only a few dozen GIS software vendors before 1988 (Kindleberger, 1992); in 2001, the number had grown to more than 500. This revolution rightfully steered the GIS industry from a focus on the technology itself toward the applications of the technology (Jenkins, 2002). The strength of GIS software is increasing while its learning curve is decreasing. At this time, GIS is one of the fastest growing market sectors of the software industry and for a good reason: GIS applications are valuable for a wide range of users, from city planners to property tax assessors, law enforcement agencies, and utilities. Once the exclusive territory of cartographers and computer-aided drafting (CAD) technicians, today's GIS is infiltrating almost all areas of the water industry.

A GIS article published in *American City and County* in 1992 predicted faster computers and networks and that efficient database management and software will move GIS applications from property recording, assessing, and taxing functions to much more diverse applications during the 1990s (Kindleberger, 1992). This article anticipated future GIS applications to be rich in their use of multimedia, images, and sound. It expected GIS applications to become more closely linked to the 3D world of CAD as used by architects and engineers. Almost all of the GIS applications predicted in 1992 are now available except interacting with GIS data in a "virtual reality" medium wearing helmets and data gloves.

GIS literature is broad due to the wide variety of areas that utilize geographic data. Likewise, the literature describing GIS applications in the water industry is itself very

broad. However, much of this work has been in the area of natural hydrology and large-scale, river-basin hydrology. A recent literature review conducted by Heaney et al. (1999) concluded that GIS applications literature exists in several distinct fields. In the field of water resources, recent conferences focusing on urban stormwater have several papers on GIS. Proceedings from two European conferences on urban stormwater by Butler and Maksimovic (1998), and Seiker and Verworn (1996), have a wealth of current information on GIS. The American Water Resources Association (AWRA) has sponsored specialty conferences on GIS applications in water resources, such as Harlin and Lanfear (1993) and Hallam et al. (1996). These reports have sections devoted to urban stormwater, of which modeling is a recurring theme. The International Association of Hydrological Sciences (IAHS) publishes the proceedings from its many conferences, some of which have dealt specifically with the integration and application of GIS and water resources management (e.g., Kovar and Nachtnebel, 1996).

In the early 1990s, not too many people were very optimistic about the future of GIS applications. This perception was based, in part, on geographic information technologies being relatively new at that time and still near the lower end of the growth curve in terms of (1) applications and (2) their influence as tools on the ways in which scientific inquiries and assessments were conducted (Goodchild, 1996). It was felt that several challenges related to our knowledge of specific processes and scale effects must be overcome to brighten the future of GIS applications (Wilson et al., 2000).

GIS applications for the water industry started evolving in the late 1980s. In the early 1990s, the water industry had started to use GIS in mapping, modeling, facilities management, and work-order management for developing capital improvement programs and operations and maintenance plans (Morgan and Polcari, 1991). In the mid-1990s, GIS started to see wide applicability to drinking water studies. Potential applications identified at that time included (Schock and Clement, 1995):

- GIS can provide the basis for investigating the occurrence of regulated contaminants for estimating the compliance cost or evaluating human health impacts.
- Mapping can be used to investigate process changes for a water utility or to determine the effectiveness of some existing treatment such as corrosion control or chlorination.
- GIS can assist in assessing the feasibility and impact of system expansion.
- GIS can assist in developing wellhead protection plans.

According to the American Water Works Association (AWWA), approximately 90% of the water utilities in the U.S. were using GIS technology by the end of the year 2000.

The use of GIS as a management tool has grown since the late 20th century. In the past 10 years, the number of GIS users has increased substantially. GIS technology has eased previously laborious procedures. Exchange of data between GIS, CAD, supervisory control and data acquisition (SCADA), and hydrologic and hydraulic (H&H) models is becoming much simpler. For example, delineating watersheds and stream networks has been simplified and the difficulty of conducting spatial data management and model parameterization reduced (Miller et al., 2004).

Today GIS is being used in concert with applications such as maintenance management, capital planning, and customer service. Many of us are using GIS applications on the Internet and on wireless devices without even knowing that we are using a GIS. These developments make GIS an excellent tool for managing water, wastewater, and stormwater utility information and for improving the operation of these utilities. Experts believe that in the near future, most water industry professionals will be using GIS in the same way they are now using a word processor or spreadsheet. Except for the computer itself, no technology has so revolutionized the field of water resources (Lanfear, 2000). In the early 1990s, GIS was being debated as the most controversial automation technology for the water industry (Lang, 1992). However, the time has come for all the professionals involved in the planning, design, construction, and operation of water, wastewater, and stormwater systems to enter one of the most promising and exciting technologies of the decade in their profession — GIS applications.

The Environmental Systems Research Institute (ESRI), the leading GIS software company in the world, has been a significant contributor to GIS applications in the water industry. ESRI hosts a large annual international user conference. The proceedings archives from these conferences are available at the ESRI Web site. This Web site also has a homepage for water and wastewater applications.

More information about GIS application books, periodicals, and Internet resources is provided in the author's first GIS book (Shamsi, 2002).

4M APPLICATIONS

Representation and analysis of water-related phenomena by GIS facilitates their management. GIS applications that are of particular importance to water industry professionals are: mapping, monitoring, modeling, and maintenance. These four *M*s define some of the most important activities for efficient management of water, wastewater, and stormwater systems, and are referred to as the "4M applications" in this book. With the help of new methods and case studies, the following chapters will show you how a GIS can be used to implement the 4M applications in the water industry. This book will demonstrate that with GIS the possibilities to map, monitor, model, and maintain your water, wastewater, and stormwater systems are almost endless. It will teach you how to apply the power of GIS and how to realize the full potential of GIS technology in solving water-related problems. This book does not train you in the use of a particular GIS software. It is not intended to help you run a GIS map production shop. Simply stated, this book will enable you to identify and apply GIS applications in your day-to-day operations.

ADVANTAGES AND DISADVANTAGES OF GIS APPLICATIONS

As described in the following subsections, GIS applications offer numerous advantages and a few drawbacks.

Advantages

Thanks to recent advances in GIS applications, we are finally within reach of organizing and applying our knowledge of the Earth in our daily lives. Typical advantages of GIS applications are described in the following subsections.

GIS Applications Save Time and Money

The foremost benefit of GIS technology is increased productivity and quicker turnaround. Increased efficiency saves time, which translates into saving money. GIS applications improve the quality of life because they make things easier to do. GIS allows us to perform routine work, such as keeping records of maintenance work or customer complaints, more efficiently. GIS tools have become user-friendly and easier to use. Local governments, utilities, and their consultants are using GIS to analyze problems and recommend solutions in a fraction of the time previously required.

GIS provides a spatial approach to organizing information about customers and the assets of a water or sewer utility, such as pipes, hydrants, pumps, and treatment equipment. GIS applications help a utility to analyze the spatial information about its customers and assets to improve planning, management, operation, and maintenance of its facilities. Municipalities and utilities that have successfully implemented GIS have seen dramatic improvements in the way in which data are retrieved, analyzed, and maintained. These improvements are allowing municipal and utility personnel to collect information more efficiently, better perform routine activities, and make more informed decisions.

GIS applications can significantly reduce time and costs associated with conventional analysis and evaluation methodologies (EPA, 2000).

GIS Applications Are Critical to Sustaining GIS Departments

Continued development of new applications is critical to sustaining the growth of a new technology. GIS, being a new technology itself, might not survive unless people use it in their routine business operations to make things easier to do, to enhance productivity, and to save both time and money. To justify their existence in an organization, GIS departments should use GIS to develop cost-effective solutions that make people's life easier. Simply stated, GIS applications are the key to garnering the management's financial support for GIS departments. This book shows how to put GIS technology to productive use in the water industry.

Although GIS applications in the water industry are not new, getting beyond basic inventory and mapping functions is often challenging. Unfortunately, mapping efforts alone do not always justify the financial support for a GIS group. Unless a GIS is taken to the operational level, it is nothing but a pretty map. That is why GIS emphasis is now shifting from producing high-quality maps to

enterprise-wide mission-critical applications. The benefit–cost ratio of GIS increases with its functionality and applications. GIS applications of automated mapping return a 1:1 benefit–cost ratio. Benefit–cost ratios of 4:1 can be attained when GIS use expands to all the departments of an organization (Alston and Donelan, 1993).

GIS Applications Provide the Power of Integration

The typical local government office contains hundreds of maps displaying such information as municipal boundaries, property lines, streets, sewer pipes, water mains, voting district boundaries, zoning areas, flood plains, school bus routes, land use, streams, watersheds, wetlands, topography, geology, and soil types, to name a few. Paper maps, after all, have been the traditional method of storing and retrieving geographically referenced information. The sheer number, range of types, and diversity of maps used by municipalities are evidence of the importance geographically referenced information plays in our day-to-day operations. Unfortunately, the wide variety of maps and diversity of their scales and designs at our disposal make it extremely difficult to access, use, and maximize the value of the information they contain. GIS *integrates* all kinds of information and applications with a geographic component into one manageable system.

The real strength of GIS is its ability to integrate information. This integration power makes the scope of GIS applications almost infinite. A GIS can be whatever we want it to be. GIS can organize the geographic information of a municipality or utility into one seamless environment. The unique integration capability of GIS allows disparate data sets to be brought together (*integrated*) to create a complete picture of a situation. GIS technology illustrates relationships, patterns, and connections that are not necessarily obvious in any one data set but are amazingly apparent once the data sets are integrated. The integration capability of GIS technology empowers organizations to make better and informed decisions based on all relevant factors (ESRI, 2003). GIS offers integrated solutions in the areas of planning and engineering, operation and maintenance, and even finance and administration.

GIS Applications Offer a Decision Support Framework

GIS helps decision makers by pulling together crucial bits and pieces of information as a "whole" and showing them the "big" picture. In this regard, GIS can be used as a consensus-building and decision-making tool. By using geography as the common denominator, GIS permits data from a wide range of disparate sources to be combined and analyzed. Therefore, an important benefit of GIS applications is their inherent ability to integrate and analyze all spatial data to support a decision-making process. GIS provides uniformity of data usage and the flexibility to test and evaluate multiple scenarios. The use of a common database eliminates the differences in presentation, evaluation, and decision making based on using different forms and types of data. For example, civil engineers who need

to know the lay of the land to design, build, and maintain projects can learn from the ways in which utilities and municipalities are linking GIS data to every aspect of their computing enterprise (Goldstein, 1997). A GIS provides the opportunity to conduct sensitivity analyses appropriate for the level of accuracy of the input data. This allows engineers, planners, elected officials, and the public to focus on the impacts and analysis of alternatives rather than the accuracy of data. After the planning and decision-making phase has been completed, GIS can continue to support the implementation phase of a project by tracking the success and failures of alternative approaches. Plan performance tracking and testing of new approaches is based on new parameters, new information, and new conditions within or outside the study area (EPA, 2000).

GIS Applications Provide Effective Communication Tools

GIS fosters better communication and cooperation among various stakeholders (e.g., community leaders and the public) of a water industry construction or improvement project. Many people learn better with maps than they do with words or numbers. GIS can be used to communicate with different audiences using visually different views of the same data. For instance, 3D plan views of a water or sewer system improvement project can be used for presentations at town meetings to graphically illustrate necessary improvements. Because GIS is a visual language, it is an excellent communication tool for visual learners. A picture is only worth a thousand words. A map may be worth a thousand numbers. But a GIS is worth a thousand tables.

In the late 1990s, GIS advocates noted that robust citizen participation in ongoing policy making was limited because many groups lacked access to the GIS environment (Obermeyer, 1998). This has started to change over the last 5 years mainly due to availability of browser-based "Internet GIS" technology described in Chapter 6 (Internet Applications). However, the challenge remains to provide citizens with GIS applications that until recently were only available to professionals.

If bottom–up decision making is to succeed, all stakeholders must have access to GIS technology in various forms (Miller et al., 2004).

GIS Applications Are Numerous

GIS is as ubiquitous as it is today because it is a very effective tool for an incredible number of applications (Zimmer, 2001). Today's GIS is limitless! The number of GIS applications is limited only by our own imagination and the availability of data. For instance, the municipal applications alone are numerous, including:

- Water, wastewater, and stormwater operations
- Comprehensive planning

- Vulnerability assessment and security planning
- Permitting and code enforcement
- Building inspections
- Zoning
- Parcel mapping
- Pavement management
- Tracking customer complaints
- Grant applications
- Routing

The magnitude of GIS applications can be appreciated by the fact that this entire book is devoted to one application from the preceding list.

A recent survey conducted by Geospatial Information and Technology Association (GITA) indicates the following top ten applications in the utility industry (Engelhardt, 2001; GITA, 2001):

- Landbase model
- Work management
- Facility model analysis/planning
- Operations and maintenance
- Document management
- Customer information systems
- Workforce automation
- Regulatory reporting
- Environmental testing
- Marketing

The same survey lists the following technologies driving GIS implementation for utilities:

- Nonproprietary programming
- Pen/mobile computing/field data capture
- Internet/intranet
- Data exchange/open GIS
- Document management/workflow
- GPS
- Digital orthophotography
- Satellite imagery

The water industry was found to be focusing on the following applications:

- Work management
- Facility model analysis/planning
- Pen/mobile computing

Water utilities were found to be implementing more pen/mobile systems than any other market segment. It was also found that water utilities were seeking higher landbase accuracy of 5 ft compared with other utilities, for instance, the 50-ft accuracy sought by gas companies.

The major GIS applications for the water industry are summarized in the following list:

- GIS provides the ideal means of describing water and sewer infrastructure facilities, identifying problems and recommending solutions, scheduling and recording maintenance activities, and supporting technical analysis (e.g., hydraulic modeling) of the facilities. For example, GIS can be used for mapping the water mains and identifying water main breaks in terms of location, pressure, soil type, pipe size, pipe material, or pipe age. Such applications are described in Chapter 9 (Mapping Applications), Chapter 10 (Monitoring Applications), and Chapter 15 (Maintenance Applications).
- Various spatial data layers can be combined and manipulated in a GIS to address planning, operation, and management issues. For example, water and sewer line information can be combined with population statistics and ground elevation data to assess the adequacy of water and sewer utilities. The Metropolitan District of Greater Cincinnati (Ohio) uses GIS to locate storm sewer system problem areas. The trouble spots are identified and targeted for preventive maintenance by mapping the relationship between customer complaints and amount of rainfall (Mitchell, 1997). Such applications are described in Chapter 9 (Mapping Applications), and Chapter 15 (Maintenance Applications).
- GIS topology provides information about how the network elements are connected with each other and what is the direction of flow. This capability makes GIS ideally suitable for identifying customers of a utility network affected by service interruption, such as water main leaks and breaks. For instance, the Cherokee Metropolitan District (Colorado) uses a "Water/Wastewater" option on the district's GIS menu to display, plot, and identify the valves to be shut off to repair water system leaks and identify and notify the customers who will be out of water due to the valve closure (Mitchell, 1997). Such applications are described in Chapter 9 (Mapping Applications), and Chapter 15 (Maintenance Applications).
- GIS can be used to satisfy regulatory requirements that are increasingly reliant on computer-generated data and maps. For example, GIS can be used to develop water/sewer system inventory reports and watershed protection/management plans. Such applications are described in Chapter 10 (Monitoring Applications).
- GIS can be used to develop hydrologic and hydraulic (H&H) computer models for water and sewer systems, watersheds, and floodplains. These applications are described in Chapter 11 (Modeling Applications), Chapter 12 (Water Models), and Chapter 13 (Sewer Models).
- GIS can be used to design efficient meter-reading routes. This can be accomplished by linking the customer account database to the streets GIS layer. This application is described in Chapter 9 (Mapping Applications).
- GIS topology can help us to simulate the route of materials along a linear network. For example, we can assign direction and speed to a streams layer to simulate the fate of an accidental contaminant release by a factory through the stream network.
- GIS can be integrated with automated mapping/facilities management (AM/FM) systems to automate inspection, maintenance, and monitoring of water and sewer systems. Sample applications include:
  - Preparing work orders for inspection and maintenance activities
  - Scheduling TV inspection and cleaning of sewers
  - Identifying the valves that must be closed to repair a broken water pipe
  - Keeping track of leak detection survey for a water system

  - Creating a map of customer complaints, pipe breaks, and basement flooding and identifying the reasons
  - Improving management of labor resources through more efficient deployment of field crews. AM/FM applications are described in Chapter 14 (AM/FM/GIS Applications), and Chapter 15 (Maintenance Applications).
- GIS analyses can be used to develop a decision support system for the efficient operation and management of stormwater, best management practices (BMPs), floodplains, combined sewer overflows (CSOs), and sanitary sewer overflows (SSOs). Such applications are described in various chapters of this book.
- GIS can be used to integrate related technologies, such as relational database management systems (RDBMSs), the Internet, wireless communications, CAD, GPS, and remote sensing (satellite imagery). The integrated platform provides the best of all worlds. Such applications are described in various chapters of this book.

Disadvantages

As with any new technology, GIS has some drawbacks. The first issue is the substantial time and cost required to compile and analyze the necessary data. High initial costs are generally incurred in purchasing the necessary hardware, software, and for ongoing maintenance. Though the advantages of GIS applications are dramatic, the failure to effectively implement GIS can lead to disappointment and disillusionment with the technology. Improperly designed and planned GIS applications can result in costly and time-consuming efforts. GIS applications are disadvantageous when one fails to define a vision, understand the vision requirements, define the tools needed to attain the vision, and select appropriate technology to integrate those tools. Only when a GIS is fully understood with proper training and education should one expect its applications to be limitless. Needs analysis, described in Chapter 2, can be used to avoid this pitfall.

Another common pitfall in GIS application development is capturing more data than required by the application. This approach is called a "data-driven" or "bottom–up" approach. For example, a nonpoint source modeling project spent substantial time and effort to capture detailed soils series survey data from the U.S. Natural Resources Conservation Service (EPA, 2000). However, the final application ended up using the Universal Soil Loss Equation for erosion modeling, which required the simpler soil associations data instead of the detailed soils series data. Considerable money could have been saved if an appropriate model and its data requirements had been identified before starting the data conversion process. The use of inappropriate data in GIS applications may lead to misleading results. The data must be of appropriate scale and resolution and highly documented to be useful in GIS applications. Users must be extremely conscious of the nature of the source information to avoid abusive extrapolations and generalizations.

The GIS learning curve, privacy issues, and periodic shortage of skilled personnel are some other challenges of GIS implementation. However, as the *San Diego Union-Tribune* (1998) reports, "Those who overcome such hurdles soon find GIS applications

breathtaking in scope and far-reaching in the potential to affect, if not shape and change, everyday lives."

SUCCESS STORIES

Some compelling examples demonstrating the benefits of GIS applications are described in the following subsections.

San Diego

The City of San Diego was an early convert to GIS technology and is considered a leader in GIS implementation. Having the motto, "We have San Diego covered," SanGIS is a joint agency of the City and the County of San Diego, responsible for maintenance of and access to regional geographic databases for one of the nation's largest county jurisdictions covering more than 4200 mi^2. SanGIS spent approximately $12 million during a 14-year period from 1984 to 1998 to collect GIS data. The conventional surveying approach would have cost them about $50 million (the *San Diego Union-Tribune*, 1998). The GIS/GPS approach has saved the City and the County of San Diego millions of dollars.

If we define a "savings factor" as the ratio of conventional approach (non-GIS) cost (or time) to GIS approach cost (or time), San Diego's success story resulted in a savings factor of 4.2 (50/12) or 420% cost savings.

Boston

The Massachusetts Water Resources Authority (MWRA) provides water and wastewater services to 2.5 million people in 60 municipalities of the Greater Boston area. The MWRA service area spans more than 800 mi^2 that contains several treatment plants, 780 mi of large-size pipelines, and dozens of pumping stations and tunnels. The MWRA recognized the potential for GIS to save ratepayers' money and initiated a GIS program in 1989. As the GIS data have been used repeatedly by individual communities to protect their water resources, the investment continues to pay off years later. For example, their geologic database "gBase" contains information on deep rock borings for the many tunnel and dam projects designed over the past 100 years. Ready access to this information guides current geologic exploration and enables MWRA to better locate new borings, which can cost $20,000 to $50,000 each (Estes-Smargiassi, 1998).

Cincinnati

Faced with a $10-billion network of aging infrastructure that included a water system with a capacity of 50 billion gallons a year, the City of Cincinnati, Ohio, conducted an in-depth feasibility study. The study showed that a GIS would save the City $11 million over a 15-year period, with payback anticipated within 8 years of implementation (American City & Country, 1993).

Knoxville

A $5.5-million GIS project that was started in 1986 jointly by the City of Knoxville, Knox County, and the Knoxville Utility Board (Tennessee) expected to pay for itself within 8 years (Dorris, 1989). In 1992, General Waterworks (King of Prussia, Pennsylvania) owned and managed 34 small- and medium-size water utility companies. The company's GIS experience indicated that a GIS at a medium-size water company can pay for itself in as little as 3 years (Goubert and Newton, 1992).

Dover

The planners in the City of Dover (New Hampshire) used GIS from Intergraph to determine where and how much solid waste needed to be collected and removed. Without GIS, collecting this information would have taken approximately 6 months. Using GIS, it only took a few weeks. This information was provided to bidders to help them base their bids on facts rather than assumptions. This application resulted in lower-priced bids (Thompson, 1991). *This success story resulted in a savings factor of 12 or 1200% cost savings.*

Charlotte

The City of Charlotte and Mecklenburg County of North Carolina, which are among the fastest growing metropolitan areas of the U.S., have developed a Watershed Information System called WISE that integrates data management, GIS, and standard hydrologic and hydraulic (H&H) programs such as HEC-1, HEC2, HEC-HMS, and HEC-RAS. Using this information technology (IT)–based integration method, the existing H&H models can be updated at a fraction (less than $100,000) of the cost of developing a new model (more than $1 million) (Edelman et al., 2001). *This success story resulted in a savings factor of 10 or 1000% cost savings.*

Elsewhere, an ArcInfo GIS software interface with H&H modeling packages HEC-1 and HEC-2 was found to offer substantial cost savings. It was determined that if the H&H models were not linked to the GIS, manual calculations would be required that would take at least five times longer than using the GIS linkage (Phipps, 1995). *This success story resulted in a savings factor of 5 or 500% cost savings.*

Thanks to a GIS approach, planning tasks previously requiring months to complete now take only days in Albany County, Wyoming.

Albany County

Albany County, located in southeastern Wyoming, covers 4,400 mi^2, a student-based population of 30,000, and 1,600 mi of roads. For rural communities like Albany County, building a GIS from scratch can be an expensive endeavor due to a lack of resources. For example, the cost associated with updating the county's existing digital aerial photography was estimated to exceed $100,000. A geographic-imaging

approach consisting of GIS and satellite imagery, on the other hand, allowed the county to have high-resolution and up-to-date views of the entire county for $32,000. Thanks to the GIS approach, planning tasks previously requiring months to complete now take only days (Frank, 2001). *This success story resulted in a savings factor of over 3 or 300% cost savings.*

GIS Applications Around the World

Although most of the GIS application examples in this book are from the U.S., GIS applications for water and wastewater systems are growing throughout the world. In the early 1990s, the United Nations Institute for Training and Research (UNITAR) started to provide GIS training to specialists from developing countries with the broader goal that each participant would eventually establish a foundation for GIS technology in his or her own country. Established in 1992, the Center for Environment and Development for the Arab Region and Europe (CEDARE) recognized that GIS was a key tool for efficient collection, management, and analysis of environmental data in the Middle East. Recognized as one of the best GIS implementations in the world, the Environment Department of the State of Qatar developed a GIS in the mid-1990s and integrated it with their environment management tasks. Fueled by high-resolution satellite imagery from the Indian Remote Sensing (IRS) satellite IRS-1C, the GIS/GPS technology saw a boom in the late 1990s.

The Sistema de Agua Potabley Alcantarillado de León (SAPAL) in León, Mexico, began its GIS implementation in 1993 and built a complete inventory of their water system, including fittings, pipes, and pumps, using ESRI's ArcView and ArcInfo GIS software packages. The Center for Preparation and Implementation of International Projects, Moscow, implemented several environmental management GISs in the Yaroslavl, Vologda, Kostroma, and Ivanovo regions of Russia. The South Australian Water Corporation (SA Water) developed the Digitized Facilities Information System (DFIS) to extract digital data for conceptual design of more than 32,000 km of water and wastewater mains. SA Water's DFIS was used to develop a Water Master Plan for the Indonesian province of West Java for the design, construction, and operation of wastewater treatment, water supply, water storage, and flood control facilities. South Africa's Working for Water organization developed a prototype ArcView GIS-based project information management system to track more than 250 water resources management projects across the country (ArcNews, 2001).

EVOLVING GIS APPLICATIONS AND TRENDS

GIS technology is changing rapidly. New GIS applications are evolving frequently mainly due to the successful marriage of GIS and the Internet. GIS applications are being fueled by the recent advances in wireless, the Internet, networking, and satellite technologies. The cost of spatial data is falling rapidly due to competition in data acquisition, processing, and distribution. More intuitive and simpler interfaces are taking GIS beyond the world of the computer geeks, techno wizards, and GIS gurus. User interfaces are becoming friendlier, wizards are replacing obscure command

lines, and the use of GIS by semiskilled end users is growing (Limp, 2001). These factors are resulting in the evolution of new GIS applications at unprecedented speed. Major evolving GIS applications and trends are listed in the following list:

- Storing both the geographic and the attribute data in one database
- Representing real-world features as objects rather than as geometric entities (e.g., points, lines, and polygons)
- Convergence of image processing and GIS into the large umbrella of information technologies (Limp, 2001)
- Using high-resolution submeter satellite imagery as base maps
- Capturing high-resolution images suitable for photogrammetric analysis using inexpensive and powerful digital cameras
- Using GPS technology for better locations of mobile sensor platforms and ground control points
- Accessing geographic data through wireless application protocol (WAP)–enabled devices, such as cellular phones, personal digital assistants (PDAs), mobile terminals, and pocket personal computers
- Using GIS and hydraulic modeling for vulnerability assessment and security planning to safeguard the water industry against sabotage and potential terrorist attacks
- Integration of computer, video camera, and GPS referred to as "video mapping" to document smoke tests and TV inspection of pipes
- Creation of an accurate virtual reality representation of landscape and infrastructure with the help of stereo imagery and automatic extraction of 3D information
- Rapid emergence of Extensible Markup Language (XML) as the *lingua franca* for data encoding and wireless Web applications
- Evolution of Geography Markup Language (GML) from XML for accurate and effective representation of GIS data in Web browsers (Waters, 2001)
- Evolution of a spatial Network Query Language (NQL) from the structured query language (SQL) for serving GIS data over the Internet, wireless, and PDAs (Waters, 2001)
- Migration from Unix workstations to Microsoft Windows desktop environment

FUTURE APPLICATIONS AND TRENDS

It is a rainy day on May 1, 2010, and it is Chrysa's first day on her new job. Chrysa is a newly hired sewer system maintenance person for a large city. She has a few years of prior sewer maintenance experience, but she is not thoroughly familiar with the sewer system of this city. As she is arranging her modular work space, the customer service department forwards her a basement flooding complaint. She talks to the customer, hangs up the phone, and slips on a headset as she leaves the office. The headset provides an enhanced reality system that combines glasses, earphones, and a tiny microphone, but weighs a little more than a pair of current sunglasses (Turner, 2001). When Chrysa reaches the subject property, she issues a simple voice command to view the buildings and sewer lines superimposed on a 3D wire-frame display of the landscape around her. As she walks around the property, her virtual reality display changes to show what is in front of her. From the 3D displays Chrysa notices that the basement of the subject property is only

slightly higher than the sanitary sewer serving the property. She then uses her wireless and GPS-enabled PDA to connect to and query the City's centralized enterprise database. The GPS automatically identifies the Lot ID, and the database tells her that the subject property's service lateral does not have a backflow prevention valve.

Chrysa's display also appears simultaneously on the office computer and PDA of her supervisor, Brian. Brian sees a shopping center upstream of the property and clicks on it. The computer indicates that the shopping center construction was completed last month. Brian concludes that the new flow from the shopping center is overloading the sanitary sewer and causing the basement flooding. Brian also authorizes Chrysa to make arrangements for the installation of a backflow prevention valve. Chrysa clicks on the service lateral of the property and creates a work order with a few clicks on her PDA. Before leaving the property, she e-mails the work order to the City's sewer maintenance contractor and copies the same to Mr. Jones, the owner of the property.

Such scenarios may seem like science fiction, but much of the technology to support them already is under development or available in prototype form, including high-speed wireless, GPS, and mobile GIS.

Experts believe that in the near future most water industry professionals will be using GIS in the same way they have used a word processor or spreadsheet.

Around the mid-1990s, not too many people were too optimistic about the future of GIS applications. Today, the future of GIS looks bright. GIS is experiencing rapid growth as an information management tool for local and regional governments and utilities because of its powerful productivity and communication capabilities. Many current and evolving GIS applications have been described in the preceding text. Other potential applications will be discovered in the future, as GIS technology and regulatory requirements continue to evolve. All sectors of business and government that depend on accurate geographical information will continue to benefit from GIS applications (Robertson, 2001). In 1999, core worldwide GIS business (hardware, software, services) revenue grew by 10.6% to an estimated $1.5 billion (Daratech, 2000). In 2001, this revenue jumped to $7.3 billion. The world will be very different a decade from now. Advances already underway in IT, communications infrastructure, microelectronics, and related technologies will provide unprecedented opportunities for information discovery and management (Turner, 2001). Looking beyond 2005, GIS-related technologies (e.g., wireless, Internet, networking, GPS, and remote sensing) are expected to do even better. Integration of photogrammetry with other geospatial technologies is opening new doors for exploring and developing alternative spatial applications. The commercial remote sensing market is flourishing and is poised for revenues of $2.7 billion within 3 to 4 years. Location-based wireless services revenues in North America alone are expected to reach $3.9 billion by 2004 (Geospatial Solutions, 2001). A growing GIS market has resulted in lower GIS implementation costs, increased data accuracy, and continued improvement in GIS hardware and software. These developments will help us to develop broader GIS applications (Robertson, 2001).

Future prognosis is a popular hobby of GIS gurus. Most GIS experts believe that the future of GIS looks bright thanks to recent advances in GIS-related technologies. Due to GIS-related innovations, future GIS applications are expected to be centered around commercial remote sensing, location-based services (LBS), mobile GIS, and the Internet. The early 21st century will bring Web-based client/server solutions, interoperability, on-demand data, and software download via the Internet on a pay-per-use basis, and an abundance of online data that can be processed in the field using wireless Web-savvy devices (Waters, 2001). Instead of purchasing GIS data permanently, future GIS users will be able to subscribe to pay-per-view type plans and pay based on the extent and duration of their usage. The technology is moving so fast that you may be using some of these applications by the time you read this book.

Though GIS has come far, big changes are still in the works (Lanfear, 2000).

To some it may appear that the future of GIS is already here; but it is not. Though GIS has come far, big changes are still in the works (Lanfear, 2000). Based on the predictions of some GIS industry experts (Lanfear, 2000; Geospatial, 2001; Engelbrecht, 2001), the following trends are expected in future GIS applications:

- Spatial data will become more widely accepted, unit costs of spatial data will fall, and GIS applications will become more affordable.
- The explosion of spatially enabled consumer products will benefit the entire GIS applications industry. Soon GPS will reside inside our wrist watches. Location information via GPS will become as common as timekeeping on a wrist watch. The new generation of low-cost consumer GPS products and spatially enabled mobile phones and PDAs will provide consistent accuracies of 10 m or better — sufficient to navigate people to their destination.
- As the bandwidth of the Internet improves, it will enable us to deliver information-rich geospatial content directly to end users.
- Wireless Internet use will surpass wired use. By 2005, over 60% of world population will have wireless Internet connections.
- Billions of people will use a Web-enabled mobile device, and millions of those devices will be equipped with LBS capabilities, either through GPS, a wireless network, or a hybrid solution (Barnes, 2001). LBS will combine GIS applications with user-friendly mobile devices to provide needed information at any time or place. LBS devices paired with wireless technologies will provide on-demand geospatial information.
- High-resolution local GIS data will be merged with federal data inventories.
- In addition to GIS input (raw data), the World Wide Web clearinghouses will share the GIS output (processed data) and the processing tools themselves.
- The new generation of raster imaging tools will allow processing of satellite imagery inside spreadsheets and word processors.
- Commercial remote sensing companies will be selling more than satellite images; they will be offering subscription-based monitoring services for change detection and vegetation indices, etc.
- Maps of dubious accuracy will not be digitized into a GIS. Contemporary imagery will be obtained for each unique application or solution.

- There will be a digital database of land-cover imagery and vectors that will include the major landmasses of the entire world.
- Laser imaging detection and ranging (LIDAR) will become a true GIS tool and produce new varieties of GIS data products. Enhanced interpretation of the strength of airborne laser signals will enable automatic creation of low-grade digital orthophotos.
- GIS databases will be connected with real time sensor data inputs, which will provide new opportunities for mapmakers and resellers of GIS products and services. New markets will be developed to expand our selection of data sets.
- Last but not least, the future may create a new set of degree, licensing, and certification programs in the field of GIS.

GIS APPLICATION DEVELOPMENT PROCEDURE

As shown in Figure 1.1, developing GIS applications generally requires six typical steps:

1. Needs analysis (strategic planning)
2. Specifications (system design)
3. Application programming
4. Testing (pilot project)
5. Installation (hardware, software, and data)
6. Ongoing operation and maintenance (includes training)

Most of these activities are generally conducted in a needs analysis study as described in Chapter 2 (Needs Analysis). Step No. 3 requires software development as described in the following text. Step No. 4 is highly recommended because it provides an opportunity to test the system design for a small pilot project area and fine-tune the design and computer programs, if necessary. Pilot testing allows system designers and programmers to tweak the design to meet user needs. Pilot testing can even save money because the data or software can be changed before full

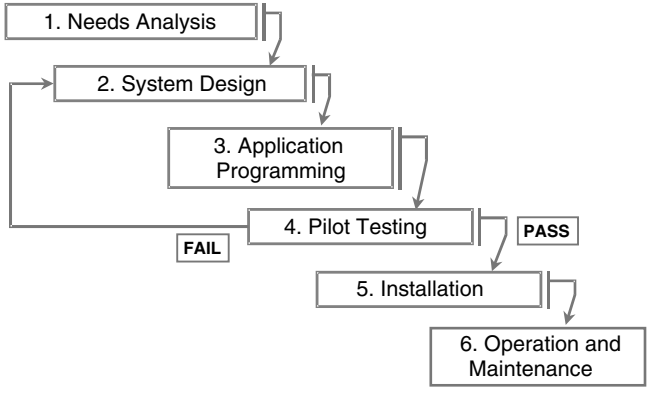


Figure 1.1 Six typical steps for developing GIS applications.

systemwide implementation, at which time it might be too expensive to make major changes.

APPLICATION PROGRAMMING

Application programming means writing the computer programs for GIS applications. Creative application developers strive to develop GIS applications to streamline the operations of their water industry business. They can find something useful in a GIS package and turn it into an innovative application. There are two methods of application programming:

1. GIS-based approach
2. Application-based approach

GIS-Based Approach

The easiest way to develop GIS applications is by extending the core capabilities of a GIS software. In this method, application functions are incorporated in a GIS. Application programs can be developed in a GIS or called from a GIS. This approach provides more GIS capability than the application-based approach. Application capability depends on the GIS software and the method of developing and integrating the application. The flexibility, adaptability, and openness of a GIS software are key to developing applications for that GIS platform. Depending on the complexity of the application, the GIS-based approach can be implemented using one or more of the following four methods:

1. GIS customization
2. Scripting
3. Extensions
4. External programs

GIS Customization

When supported by a GIS package, this method is implemented by customizing the GIS software's graphical user interface (GUI). Examples include adding or deleting a button or pull-down menu. For example, the buttons or menus that are irrelevant to a particular application can be deleted to simplify the learning process and prevent unwanted edits. New buttons or menus can be added to display helpful tips and messages. This method is more suitable for supporting applications rather than developing them. For example, it can be used to add a button that runs another program or script. This method is not very useful when used alone; it is therefore used in conjunction with the other methods described below. When used alone, it does not require computer programming.

Scripting

Scripting is suitable for simple applications such as creating a link to an H&H model. This method requires programming using a scripting language. Scripts are

small computer programs written in a scripting language. A scripting language is a programming language that is (usually) embedded in another product, such as Microsoft's Visual Basic or Autodesk's AutoLISP. ESRI's ArcGIS software (Version 8 and higher) allows programming using Visual Basic for Applications (VBA). Older versions of ESRI's ArcInfo and ArcView software allow programming using Arc Macro Language (AML) and Avenue scripts, respectively. Avenue is ArcView's native scripting language built into ArcView. Avenue's full integration with ArcView benefits the user in two ways: first, by eliminating the need to learn a new interface; second, by letting the user work with Avenue without exiting ArcView (ESRI, 1995). A sample Avenue script for displaying sewer TV inspection video movies in ArcView GIS is given in the following text.

```
theVal = SELF
if (not (theVal.IsNull)) then
if (File.Exists(theVal.AsFileName)) then
 System.Execute("C:\Program Files\Windows Media
Player\MPLAYER2.EXE "+theVal)
else
 MsgBox.Warning("File "+theVal+" not found.,""Hot Link")
end
end
```

This "Movie&ImageHotLink" script enables users to click on a sewer line in the GIS map to play the digital movie of the sewer defect. It receives the movie filename, checks to see if the file exists, and then opens it with a user-specified movie player such as Windows Media Player. The reason for the small size of the script is the fact that it takes advantage of an existing program (i.e., it uses Media Player). Figure 1.2 shows an ArcView screenshot displaying the results of running this script. As the user click on a collapsed pipe, its video file starts playing in the Media Player window.

Figure 1.3 shows the integrated VBA programming interface of ArcGIS 8.3. The left window shows VBA code to automatically assign polygon feature IDs to point features (Lundeen, 2003). For example, this script can be used to create parcel IDs for water meters or to create basin names for manholes. The right window shows the water system layers and parcels in ArcMap. ESRI's ArcScripts Web site is a good place to find downloadable scripts for ESRI GIS software.

Extensions

Extensions are suitable for developing complex applications such as adding raster GIS capability to support hydrologic modeling. Scripts should be compiled and linked to GIS software before they can be executed, which is cumbersome. A set of scripts can be converted to an extension for faster and user-friendly installation and execution. Extensions are special scripts that are loaded into the GIS software to

Figure 1.2 Sample application that plays sewer TV inspection videos in ArcView GIS.

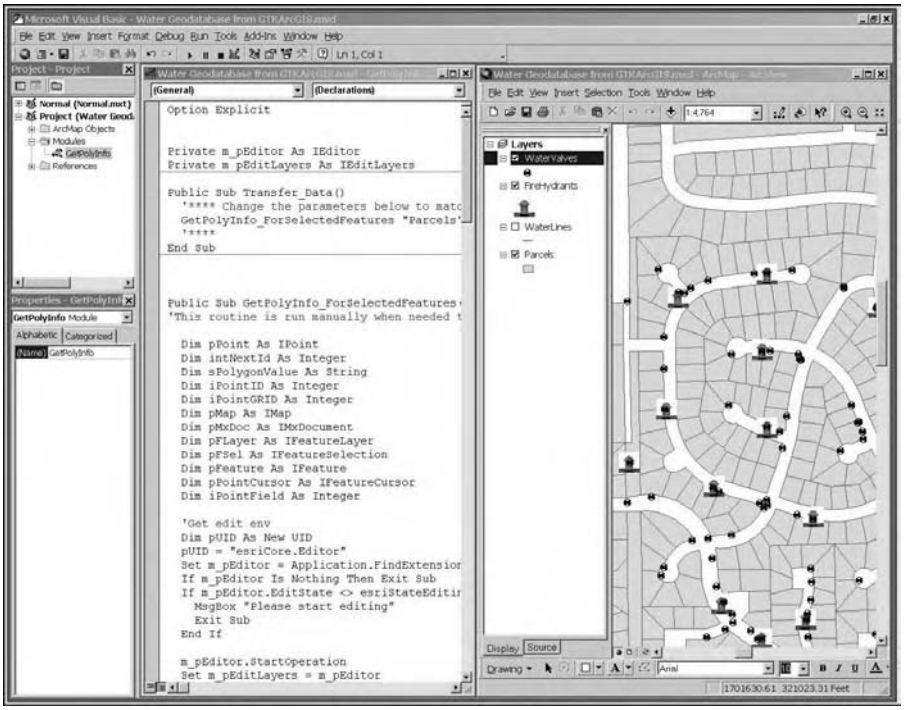


Figure 1.3 Integrated application development environment of ArcGIS (Version 8.3).

increase its capabilities. For example, ArcView 3.2 extensions can be loaded (or unloaded) by selecting them from the File\Extensions pull-down menu. Small basic extensions, such as the "hydro" (hydrology) extension are included with the ArcView software. Hundreds of free basic extensions written by the ESRI software user community can be downloaded from the ArcScripts Web site. Complex special-purpose extensions, such as Spatial Analyst Extension (adds raster GIS capability to ArcGIS) or 3D Analyst Extension (adds 3D capability to ArcGIS) should be purchased separately.

External Programs

Some scripting languages are not suitable for programming mathematically complex algorithms. For many applications (e.g., H&H modeling), the computer code is written and maintained by government agencies. Converting an existing code to a script might be error prone and cumbersome. In these situations, external programs can be called in from inside a GIS. The Movie&ImageHotLink script described earlier uses this method when it uses the external Media Player program. ArcGIS allows developing cross-platform applications with C++ or Java languages.

A dynamic link library (DLL) is an executable module that contains functions or procedures that can be called from an external program. For example, each major

component of the Windows operating system is a DLL. Custom DLLs can be created using languages such as C and C++. Many GIS packages such as ArcView 3.x support DLLs that can be used to provide efficient interfacing with third-party external programs. Functions in DLLs can be called directly from Avenue, and data can be passed back and forth.

Application-Based Approach

In this method, GIS (mapping) functions are incorporated in existing applications. GIS functions are developed in, or are called from, an existing application. This method offers limited GIS and complete application functions. For instance, ESRI MapObjects and ArcObjects software provide programmable objects that can be plugged into VBA applications. RJN Group's CASSView module is an example of this approach. CASSView integrates an infrastructure maintenance management system in MapObjects environment. It requires the RJN CASS WORKS system and a minimum of a geo-coded street centerline to operate. CASSView provides CASS WORKS users with access to maintenance information, as-built records, maps, and GIS data. The end result is a database-driven desktop mapping application that provides visual data on demand.

In-house development of applications requires computer programming skills. It is often difficult to find people skilled in GIS-specific programming languages such as Avenue and AML. For advanced applications, the purchase price of commercial packages is usually less than the labor cost of in-house software development. A large variety of third-party applications from facilities management to H&H modeling is available for purchase from commercial vendors. Most commercially available software is listed in the "Business Partner" page of the GIS software Web site or advertised in the periodicals of the GIS software company.

USEFUL WEB SITES

ESRI user conference proceedings	www.esri.com/library/userconf/archive.html
ESRI water and wastewater industry	www.esri.com/industries/water/water.html
ESRI ArcScripts	arcscripts.esri.com
RJN Group (CASSView)	www.rjn.com
San Diego GIS home page	www.sangis.org

CHAPTER SUMMARY

This chapter showed that with GIS the possibilities to manage your water, wastewater, and stormwater systems are almost endless. It provides an ideal platform to integrate various business operations and technologies. GIS applications increase worker productivity and save time and money, and are critical to garnering financial support for the GIS departments of an organization. They equip us with better communication and decision-making tools. GIS applications are so numerous that the possibilities can be limited only by our own imagination. The following chapters

will describe the procedures for and examples of developing GIS applications for the typical needs in the water industry. In particular, we will focus on the four areas of applications: mapping, monitoring, modeling, and maintenance. Referred to as the 4M applications, these are most critical to the effective operation and management of water, wastewater, and stormwater systems. Due to recent advances in Internet and wireless technologies, new GIS applications are evolving at an unprecedented rate. It is expected that future GIS applications will be even more exciting than those that have been listed in this chapter. The effect of these new applications will be profound. They will be inexpensive, user friendly, and ubiquitous, and will support the integration of GIS and related technologies in ways unlike anything current water industry professionals have ever envisioned.

CHAPTER QUESTIONS

1. What are GIS applications and how are they developed?
2. What are the pros and cons of GIS applications? How can you avoid the potential pitfalls of GIS applications?
3. What technologies are fueling the popularity of GIS applications?
4. List your ten favorite GIS applications.
5. How will your current GIS needs benefit from potential future GIS applications?

CHAPTER 2

Needs Analysis

A careful needs analysis is critical to successful GIS implementation and should be the first task of any GIS project.

Needs analysis requires a thorough analysis of existing data sources.

LEARNING OBJECTIVE

The learning objective of this chapter is to understand the value and methodology of needs analysis for GIS application projects.

MAJOR TOPICS

- Definition of needs analysis
- Needs analysis steps
- Needs analysis case studies

LIST OF CHAPTER ACRONYMS

COTS Commercial Off-the-Shelf
DBMS Database Management System
GUI Graphical User Interface
IT Information Technology
RDBMS Relational Database Management System

OCEAN COUNTY'S STRATEGIC PLAN

Ocean County Utilities Authority, located in central New Jersey, serves 36 municipalities through three wastewater treatment plants with a combined capacity of 80 MGD. The sewage collection system consists of 200 mi of sewer lines and 40 pumping stations. The Authority developed a GIS program as an integral part of its information technology (IT) strategy. Instead of treating the GIS as a stand-alone system, the Authority envisioned using it as integrated part of its other enterprise systems such as the SAP Maintenance Management System. Authority's IT-based approach included the following steps (Stupar et al., 2002):

- Needs analysis: This step identified goals of the GIS as they related to the Authority's business needs.
- Spatial database design: The objective of this step was to treat GIS data as an enterprise asset. ESRI's geodatabase data model was used to design the database.
- Data conversion: This step involved collecting the needed data and converting it to a geodatabase.
- Application development: This step designed and implemented applications for the GIS.

INTRODUCTION

Needs analysis (or needs assessment) identifies and quantifies the GIS needs of an organization and its stakeholders. Performing a needs analysis is the crucial first step to a successful GIS project. Needs analysis is analogous to strategic planning; it is a blueprint for funding, implementing, and managing a GIS. Like strategic planning, a careful needs analysis is critical to a successful GIS implementation and

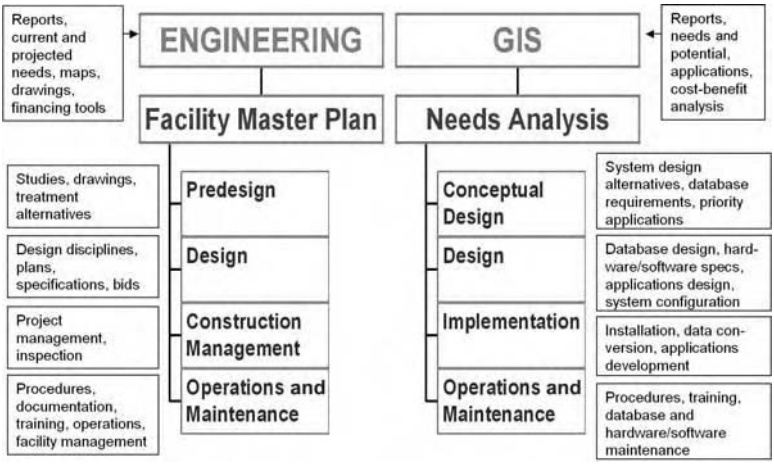


Figure 2.1 Comparison of needs-analysis study and facility master plan.

should be the first task of any GIS project. Needs analysis clarifies the project's specific needs and defines how a GIS will benefit an organization by relating specific organizational resources and needs to specific GIS capabilities (Wells, 1991). There are three major goals of needs analysis:

1. To define current work processes, data, and IT resources in place
2. To determine how the organization hopes to use GIS to streamline its operations
3. To make recommendations defining the path to meet these hopes

Needs analysis identifies potential applications that can be performed more efficiently using GIS technology. The applications are identified through stakeholder interviews and through an inventory of potential data resources.

The needs analysis findings are presented in a needs analysis study or needs assessment report. Depending on the size of the city or the organization, these studies can be generally completed in 6 to 12 months for $25,000 to $100,000. In certain ways, a GIS needs analysis study is analogous to a facility master plan in engineering. Figure 2.1 shows a comparison of the two documents. A facility plan provides information on the design, construction, and operations and maintenance of a structure. A needs analysis study provides the same information for a GIS application.

Table 2.1 lists the potential applications for the Borough of Ramsey, NJ, identified in a 1994 needs analysis study completed by Chester Engineers (Pittsburgh, Pennsylvania). This needs analysis study was completed in 1 year at a cost of $40,000.

NEEDS ANALYSIS STEPS

There are eight typical needs analysis steps:

1. Identify stakeholders
2. Talk to stakeholders

Table 2.1 Potential Applications of GIS Technology

Department	Representative Applications
Public works (responsible for operation and maintenance of the sewer system)	Emergency repair Utility markout Asset management Maintenance tracking and scheduling One-call support Critical water user location Wellhead protection
Engineering and planning	Generation of 200-ft notification lists Demographic analysis Planning review Permit approval Green space analysis Floodplain mapping Drainage management (storm sewer maps)
Tax office	Parcel location and data retrieval Tax-map production Property photo storage and retrieval Automated public access
Public safety	Fire hazard locations Hazardous-materials inventory Hydrant flow data retrieval Preferred hydrant location Crime location analysis

3. Inventory resources
4. Establish need priorities
5. Create system design
6. Conduct a pilot project
7. Prepare an implementation plan
8. Conduct the final presentation

Step 1. Stakeholder Identification

Stakeholders include all the people who can affect or can be affected by a GIS project, such as users, executives, and GIS team members. A broader classification can include users, managers, IT personnel, customers, clients, policymakers, elected officials, politicians, regulators, partners, and funding experts.

Step 2. Stakeholder Communication

This step requires talking with and listening to stakeholders to identify their GIS needs. Some or all of the following formal communication methods can be employed:

1. Introductory seminar
2. Work sessions
3. Focus groups
4. Interviews

The first three methods are optional and can be used depending on the size, scope, and budget of the project. The last method is almost always necessary to perform a comprehensive needs analysis study.

Introductory Seminar

An introductory seminar is warranted if most of the stakeholders are new to GIS. It should focus on three critical questions:

1. What is GIS?
2. Why use GIS?
3. Who else is using GIS?

The first item defines GIS for those new to it. This part should be done in layperson's language. Avoid technical terms and acronyms typically found in a college lecture on GIS-101. The second item answers the question of why GIS is important. It should emphasize that GIS benefits people, it makes things easier to do, and it increases productivity and reduces the cost of business functions. The third item presents compelling examples and success stories from other communities and organizations, such as:

- By using GIS, Charlotte, North Carolina, completed an H&H modeling project initially estimated at $1 million for less than $100,000.
- By using GIS and LIDAR elevation data, Chatham County, Georgia, saved $7 million in construction cost.
- By using GIS, Cincinnati, Ohio, expects to save $11 million in 15 years on a water infrastructure improvement project.

Success stories and associated savings factors presented in Chapter 1 (GIS Applications) can be used to answer questions 2 and 3.

Work Sessions and Focus Groups

These 1- to 2-h management level focus group sessions are conducted to gain a top–down view of the issues confronting the organization. They also allow capture of the business processes that can be enhanced using GIS applications. The discussions lead to establishing goals that can be forged into benchmarks for GIS implementation, such as:

- Creation of a map showing sewer pipes and manholes on an aerial photograph
- Creation of a map showing water mains, fire hydrants, and valves on city streets
- An application running on field (mobile) computers capable of identifying valves that should be closed to isolate a broken water main for repair

Interviews

This is often considered the single most important step in the development of a GIS plan (Henstridge, 1998). A number of selected stakeholders should be interviewed to identify the needs, uses, problems, and requirements the organization has

for GIS and its applications. The interviews should attempt to quantify the cost and effort that can be reduced by implementing GIS. People with good communication skills should conduct this step. You should listen, learn, and encourage audience participation. Prepare a list of questions but avoid distributing a long questionnaire. Do not organize the interviews like a pop-up quiz. The typical interview duration is 1 to 2 h. The key questions that should be explored are:

- Do they want a GIS?
- Do they need a GIS?
- What is their wish list?
- What functions (operations) need to be improved?
- Do these functions need GIS?
- Can GIS improve these functions?
- What are the need priorities?

A GIS vision can evolve during the above communication steps. For large GIS projects, articulating a vision statement is highly recommended because, in the famous words of Yogi Berra, "If you don't know where you're going, you might not get there." The vision should reflect how users envision the business processes working in an ideal way. Most vision statements reflect easy access to accurate information (Zimmer, 2001).

When communicating with stakeholders, remember the famous metaphor, "to a man with a hammer, every problem looks like a nail." GIS cannot and will not solve every problem in an organization.

We sometimes set ourselves up for failure by creating unrealistic expectations of what GIS will do for us. Before the commencement of a needs analysis, some unscrupulous or unknowledgeable salespeople might make unsuspecting users believe that all applications will function flawlessly regardless of the quality or even existence of data. A GIS project's success is measured against expectations. Managing these expectations helps to achieve project goals and claim success. A mismatch between expectations and resources is often a source of GIS failure. Some GIS experts believe that an ill-defined scope and unrealistic expectations are the most common causes of GIS project failures (Somers, 1999; Davis, 2000; Meyers, 2002). Stakeholder communications provide an excellent opportunity to clearly define the scope of a GIS project and manage end-user expectations.

Step 3. Resource Inventory

This step requires inventorying GIS-related resources, such as maps and drawings, spatial and tabular data, databases, IT resources (hardware, software, and network), and the staff skills and interests. This step describes the as-is condition of the organization.

Lehigh County Authority located in Allentown, Pennsylvania, provides water and wastewater services to 12,000 customers in a service area of approximately 50 mi^2. In 1999, the Authority hired a consultant to complete a combined needs assessment and GIS strategic implementation plan. The needs analysis study included a complete

inventory of the Authority's digital and hard-copy data sources and compiled the following list (Babbitt, 2002):

- A 1-in. = 800-ft schematic map
- A series of valve maps for each development
- Some CAD files
- Approximately 900 as-built drawings
- A valve and hydrant database in Microsoft Access
- An in-house customer information system in the ADMINS language

Step 4. Need Priorities

Because it is not feasible to implement a large number of applications simultaneously, the needs identified in Step 1 should be prioritized. This can be done by assigning a priority ranking to needs, such as mission critical, most urgent, very high, high, medium, and low. For example, Lehigh County Authority's needs analysis study identified 34 potential applications. Of these, five were considered high priority based on organization-wide use and business process improvements (Babbitt, 2002).

Remember that the people who are paying for your GIS project would be anxious to see the return on their investment. Applications that can be developed quickly (e.g., digital orthophotos) should, therefore, be given a higher priority to help show early progress during the implementation phase. The applications that are least costly or those that give the "biggest bang for the buck" (e.g., downloadable digital elevation models) can also be assigned a higher priority.

A common approach to implementing this step is to summarize and correlate the interview results into a matrix with priorities assigned to each need. Table 2.2 shows a popular method to quantify the needs by preparing a matrix of needs (applications) and the departments that need those applications. The last column shows the total score of each identified need as the sum of departments requiring that need.

Step 5. System Design

This step recommends a GIS to support the applications identified in the previous steps. It should evaluate the as-is state of the organization relative to key business functions and identify the required elements needed to create the desired improvements. System design includes determining specifications for the following components:

- Data conversion (mapping)
- Database
- Software
- Hardware
- User interface

Data Conversion (Mapping)

Approximately 75% of typical GIS costs are related to data conversion and creation. This component includes data conversion methods (scanning, digitization, etc.)

Table 2.2 Sample Departmental Needs Matrix

	Departments									
Needs (Applications)	**Administration**	**Plants**	**Customer Service**	**Lab**	**Engineering**	**Collection System**	**Shop/Purchasing**	**Pretreatment**	**Solids**	**Total Score**
Public access/general GIS data viewing	☑	☑	☑	☑	☑	☑	☑	☑	☑	9
Emergency management response	☑	☑	☑	—	☑	—	—	—	—	4
Special notification	☑	☑	☑	—	☑	—	—	—	—	4
Customer location and service verification	—	—	☑	—	☑	—	—	—	—	2
Customer service/complaint tracking	☑	—	☑	—	☑	☑	—	—	—	4
Sample point monitoring and tracking	—	☑	—	☑	☑	—	—	—	—	3
As-built indexing	—	—	—	—	☑	☑	—	—	—	2
Work order management	☑	☑	☑	—	☑	☑	☑	☑	☑	8
Project management tracking	☑	—	—	—	☑	—	—	—	—	2
Link TV inspections to GIS/work order management	—	—	—	—	☑	☑	—	—	—	2
Pretreatment tracking	—	—	☑	—	☑	☑	—	☑	—	4
Biosolids tracking	—	—	☑	—	☑	—	—	—	☑	3

and the source, resolution, and scale of the maps or map layers. Needs-driven design is particularly critical for data conversion and database components. The design and content of the mapping and database components determine the fundamental capabilities of the entire GIS. These components can be considered to be the foundation of the system. Once actual implementation has begun, extensive changes to either the mapping or the database design can be very costly. Additional information for this component is provided in Chapter 8 (Mapping).

Database

A GIS database stores descriptive information about map features as attributes. For example, a water system database includes attributes for pipes, valves, meters, hydrants, and so on; and a sewer system database contains attributes for pipes, manholes, catch basins, outfalls, and so on. The creation of an appropriate GIS database is the most difficult and expensive part of developing GIS applications. Successful GIS applications require a database that provides appropriate information in a useful and accessible form. The design of the database is, therefore, driven by application needs.

For data storage and manipulation, a database management system (DBMS) uses a data model, such as a hierarchical, network, or relational data model. Conventional GIS databases consist of graphic features with links or pointers (usually facility identification numbers, or IDs) to related attribute or tabular data. In the late 1990s, a new object-oriented (rule-based) data model was introduced that stores attributes as an intrinsic part of the graphics feature. These modern databases do not require links between features and attributes because they can store both the graphic features and attributes inside a single relational DBMS (RDBMS).

Database design involves three steps: (1) conceptual design, (2) logical design, and (3) physical design (Shamsi, 2002). Conceptual design does not depend on hardware or software. Logical design depends on software only. Physical design depends on hardware and provides a detailed definition of the structure and content of the database. Database design is presented with the help of a data dictionary, which documents the logical and physical structure of the layers of the GIS. Table 2.3 shows a sample data dictionary for a 2000 sewer mapping project in the city of Monongahela, Pennsylvania.

Table 2.3 Sample Data Dictionary

Layer	Name	Class	Attribute	Description
Study area	STUDY	Lines	none	Defines general study area and clips other layers
City boundary	BOUNDARY	Lines	none	Municipal boundary
Roads	ROADS	Lines	none	City streets from CAD file
Railroads	RAILS	Lines	none	Railroads from TIGER file
Surface water	RIVERS	Lines	none	River and creeks from CAD file
Subarea	SUBAREA	Lines	none	Digitized sewershed boundaries
		Polygons	Subarea_id	Subarea identification (ID) number
			Acres	Area in acres
			Mn_percimp	Mean percentage imperviousness
			Pop	Population
			Mn_popden	Mean population density
			Houses	Number of houses
			Mn_famsize	Mean family size
			Mn_mktval	Mean market value
			Mn_medinc	Mean median family income
Sewers	SEWERS	Lines	None	Collection system sewers
Interceptor	INTERCEP	Lines	None	Interceptor sewers
Treatment plant	TREATMNT	Lines	None	Treatment plant footprint
CSO discharge point	CSO	Points	Cso_id	ID Number of combined sewer overflow (CSO) discharge points

For detailed information on water, wastewater, and wastewater system database design, please refer to the author's companion book *GIS Tools for Water, Wastewater, and Stormwater Systems* (Shamsi, 2002).

Software Selection

GIS software represents less than 10% of the total GIS cost in most cases, yet people spend a lot of time in selecting the best software for their GIS applications. Many users are so proficient (and sometimes loyal) in using a given software that they will lose interest in GIS if they are not allowed to use their favorite GIS software. Because user needs are key to the success of a GIS project, their preferences should be taken seriously. Thus, despite its relatively lower cost, GIS software can have a significant impact on the success of a GIS project.

GIS software should have an "open" architecture. An open GIS allows for sharing of geographic data and integration among various GIS, GIS-related (e.g., remote sensing), and even non-GIS (e.g., hydraulic modeling) technologies. An open GIS can operate on different operating systems, platforms, and database management systems. It can be scaled to support a wide range of application needs, from an office engineer using GIS on a desktop, to a mobile field technician using a handheld device, to hundreds of users working on an enterprise system across multiple departments.

Today, GIS development software used for data conversion work is almost always a commercial off-the-shelf (COTS) program because custom development would be simply too expensive. Many commercial packages are available for implementing common GIS applications, such as work order management. Thus, no software development is necessary for data conversion or application software because reinventing the wheel is "penny-wise and pound-foolish." Some U.S. government agencies that invested in developing their own home-grown GIS packages did not see a return on their investment (Henstridge, 1998a). Custom development is warranted if COTS packages are not available for an application or if they are too expensive. As described in the following text, creation of a custom user interface also requires software development.

Custom software can be developed in-house if computer programmers skilled in the GIS-specific programming languages are available. If this is not possible, consultants can be hired. Development and (often overlooked) maintenance costs are the major disadvantages of in-house development. For some organizations (e.g., consultants) an advantage of the in-house software is the competitive edge of proprietary packages.

A needs analysis must carefully select an appropriate GIS software for the GIS needs of an organization. Fortunately, the field has narrowed and there are only a few major vendors today (Henstridge, 1998a). Some needs analysis studies utilize a decision matrix of user needs (applications) and available software capabilities to help select the most appropriate software. For example, Table 2.4 shows a sample decision matrix for various applications of a typical water utility that uses an ESRI GIS platform. This table indicates that ArcView, ArcInfo, and Spatial Database Engine (SDE) should be given high priority because they are required to implement all the GIS applications of the organization.

Table 2.4 Software Selection Matrix Example

Application	AV	SA	NA	3DA	AI	SDE	AO
Planning and engineering	🖫	—	—	🖫	🖫	🖫	🖫
Operations and maintenance	🖫	—	🖫	—	🖫	🖫	🖫
Construction	🖫	—	🖫	—	🖫	🖫	🖫
Infrastructure management	🖫	—	🖫	—	🖫	🖫	🖫
Water resources and hydrology	🖫	🖫	—	—	🖫	🖫	—
Finance and administration	🖫	—	—	—	🖫	🖫	—

Notes: AV = ArcView; SA = Spatial Analyst; NA = Network Analyst; 3DA = 3D Analyst; AI = ArcInfo; SDE = Spatial Database Engine; and AO = ArcObjects.

Hardware Selection

The selection of appropriate GIS hardware depends on the scope of the GIS application project and the available resources. GIS hardware can be set up as a stand-alone workstation or in a network configuration. Typical hardware requirements are listed in the following text:

- Computer with fast processor, large memory, and extensive disk space
- DVD drive
- Large-screen, high-resolution, color monitor (21 in. minimum)
- Backup and storage device, such as a CD or DVD burner
- Color printer
- Large-format (wide-format) color plotter
- Digitizer
- Large-format scanner
- High-speed Internet connection (T1, cable, or DSL)
- GPS receiver
- Computer server (for networked configurations)

Hardware can be expensive, but like the GIS software, the hardware for a typical municipal GIS accounts for less than 10% of the total GIS investment. Thus, cost alone should not be used in hardware selection. The following additional factors should also be considered: service and support, capacity for growth (storage and memory), connectivity with peripheral devices (plotters, scanners, digitizers, GPS equipment, and field computers), and connectivity with Internet and networks (intranets and extranets). When comparing prices and hardware features, remember the golden rule: You get what you pay for. Another valuable advice for hardware procurement is: Try before you buy.

Today's laptops are just as powerful as desktops. For certain users, like a GIS manager who must commute between different offices, a laptop computer might be more suitable than a desktop computer. The laptops cost more but are worth the extra money for mobile users. As of 2003, the minimum requirements for a laptop to be able to run GIS software are: a 1.7 GHz processor; 512 MB memory; 60 GB hard drive; 1280 x 1024 resolution, 32-bit color XGA screen; DVD R/RW drive; Internal 56K modem; and an Ethernet or wireless Local Area Network (LAN) (Thrall, 2003).

Client/server architecture is a common computer system configuration in which one main server stores and distributes all data. Workstation users, called "clients," access and manipulate server data using client software installed on the workstation computer. A network-based design is necessary if more than one person will need to access the same GIS data, i.e., when a team rather than an individual is working on a project. For an individual researcher doing a thesis or term project, a personal workstation is sufficient. However, in real-world projects the entire project team must share the data, which requires a client/server architecture. Server maintenance is absolutely necessary. Without a dedicated technician, the client/server architecture is difficult to maintain.

Wide- or large-format plotters allow printing on 24- to 60-in. rolls. Today's wide-format plotters are available in both ink and laser technology. The cost depends on resolution, color, and speed. Common inkjet plotters range from $2,500 to $25,000. Common laser plotters range from $15,000 to $175,000. When budgeting for hardware cost, do not forget to include the maintenance cost; for example, supplies for printer and scanner such as toner, ink cartridge, paper rolls, etc. Compare the up-front cost of a laser printer to the savings of not replacing cartridges (approximately $40 each). The maintenance cost can range $100 to $500 per month based on usage. Thus, when pricing a plotter, do not forget the cost of media and ink. Compare the plotters using a cost-per-page figure for different media types. Make sure that the plotter you choose can handle a large number of plots per day at the quality you prefer. Request for a demonstration and plot your own maps before making the final selection. Some manufacturers like Hewlett Packard allow trading in the old plotter toward the purchase of a new one and have lease options starting at less than $100 per month.

As of 2000, ESRI software runs on many platforms including Digital Tru64 UNIX and Alpha Windows NT, IBM, AIX, HP-UX, Microsoft Windows 95/98/NT/2000, NEC EWS-UX, SGI IRIX, and Sun Solaris. Linux is being evaluated and tested as a viable platform (ESRI, 2000).

Orange County Geomatics/Land Information Office started to provide enterprise-wide access to GIS data in 1997 using hardware and software from Intergraph. Their Oracle database containing 7 GB of attribute information resided on a 200 MHz Quad Pentium Pro with 256 MB of RAM. The data could be viewed using Intergraph's MGE product suite (Goldstein, 1997).

Table 2.5 provides sample hardware and software specifications for a single-user installation specified in a 1994 needs analysis study for the Borough of Ramsey, New Jersey. The 1994 cost for this setup was approximately $3000. Table 2.6 provides hardware and software specifications and the approximate cost specified in 2000 for a four-seat, server-based GIS laboratory. The 2000 cost for this setup was approximately $145,000. A comparison of Table 2.5 and Table 2.6 indicates that computer hardware has changed substantially within the period of 6 years from 1994 to 2000.

User Interface

A user interface is a computer program that acts as an interpreter between the user and a computer. It generally operates on top of a GIS package as a menu-driven shell to the software's commands. A graphical user interface (GUI) is a popular

Table 2.5 Single-User Hardware and Software Specifications (Year 1994 Example)

Component	Recommended Specifications
CPU	Intel Pentium, 256 KB RAM cache
Video card	VL-bus Windows accelerator capable of 72 Hz vertical refresh rate at 1024 x 768 resolution with 256 simultaneous colors, noninterlaced
Monitor	17 in. size, performance to match video card, 0.28 mm dot pitch
RAM	16 MB, expandable to 32 MB, 36-bit SIMMs
Hard drive	500 MB IDE
Floppy drives	One 3.5 in. (1.44 MB), one 5.25 in (1.2 MB)
Expansion slots	Eight ISA slots, including two VL-bus slots
Drive bays	Total of five half-height bays, three with front access
Fax modem	14,400 bps Hayes compatible, internal
Printer/plotter	HP 650C DesignJet
Software	DOS 6.0, Windows 3.1, ArcView

type of user interface that replaces difficult-to-remember text commands by interactive computer graphics consisting of menus, dialogue boxes, input and output windows, and icons (Shamsi, 1997). For example, Internet Explorer™ is a GUI for the World Wide Web.

Creation of a custom GUI is optional. It is recommended when the GIS functions are very complex and/or the end users are complete novices. A custom GUI provides access (links) to all the GIS functions of an organization from a single program (screen) that bears the organization name and corporate logo. A custom GUI can also be used for user authentication using passwords.

Table 2.6 Hardware and Software Specifications, and Cost of a Four-User GIS Laboratory (Year 2000 Example)

Item	Year 2000 Cost ($)
One 400 MHz Pentium II NT server with 256 MB 100 MHz SDRAM, 3 x 4 GB hot swap hard drives, and Uninterruptible Power Supply (UPS)	9,000
Four Pentium II personal workstations (400 MHz, 128 MB RAM, 4.3 GB hard disk, writeable CD-ROM) with 21 in. monitors ($4,000 per workstation)	16,000
UPS backup for all personal workstations ($300 each)	1,200
Large-format color scanner with floor stand Calcomp ScanPlus III 800C (A–E size, 800 dpi max resolution)	20,000
Flatbed desktop scanner	100
Digitizer (42 x 60 in.) with floor stand. Must be WinTab compliant. Calcomp Summagrid V or GTCO Accutab.	10,000
Color plotter. Must be E size (36 in. wide). HP DesignJet 750C Plus (36-in.-wide sheet roll)	7,500
B&W LaserJet printer (HP LaserJet 5 or better)	500
Color inkjet desktop printer (HP DeskJet 1600C or better)	500
Iomega 2 GB Jaz drive for backup	500
ARC/INFO NT bundle 4 seats (includes ARC/INFO, ArcView, and ArcView extensions) ($20,000 each)	80,000
Total Cost	**145,300**

Some GIS packages allow customization of their default interfaces with little or no programming to meet user-specific needs. When available, this feature can be used to create simple GUIs. Custom GUIs for complex applications require computer programming or scripting. Custom GUI development needs will be gradually reduced and eventually eliminated as GIS packages complete their migration from a command-based to an interface-based environment.

Step 6. Pilot Project

This is a very important step. Inclusion of a pilot project in a GIS implementation plan allows the project team to test an initial strategy before committing to a specific course of action. For example, database design determines the fundamental capabilities of the entire GIS. Once actual implementation has begun, extensive changes to database design are very costly. Thus, before a systemwide implementation, the initial database design and data conversion methodology should be evaluated for a small portion of the project area (pilot area). The pilot project allows you to test the system design, evaluate the results, and modify the system design if necessary. It provides an opportunity to refine the proposed approach to full-scale data conversion and implementation. This step should also produce a working demonstration of the intended applications for the pilot area.

The Washington Suburban Sanitary Authority (WSSC) is among the ten largest water and wastewater utilities in the U.S. According to the 2003 records, it serves 420,000 customer accounts, provides water and sewer services to 1.6 million residents in a nearly 1000 mi^2 area, and employs more than 1500 people. In the early 1990s, WSSC undertook GIS implementation as a cooperative effort with five other regional agencies. The interagency organization called Geographic Information System for Montgomery and Prince George's Counties (GeoMaP) proceeded with studies to assess needs, analyze requirements, and evaluate alternatives. A detailed GIS system design was then completed for specifications of database, data conversion, hardware, software, and communications network. A pilot project was completed to evaluate the system design prior to full-scale implementation. The pilot was intended to be the first phase of GIS implementation rather than a separate experimental effort or prototype. The pilot project developed all components of the proposed database design for sample areas and designed and developed a sample set of GIS applications. The water and sewer facility layers were structured as ARC/INFO network coverages to support network analysis applications as well as to provide standard water and sewer reference maps. The sewer layer was found much easier to develop than the water layer (Cannistra et al., 1992).

Mohawk Valley Water (MVW) is a water utility serving more than 125,000 people in approximately 150 mi^2 area in central New York. MVW completed a 6-month pilot project in 2002 to develop a water distribution system geodatabase using ESRI's ArcGIS, ArcIMS, and ArcSDE software and a hydraulic model using Haestad Method's WaterCAD software. The pilot project covered approximately 5% of the representative distribution system area including three storage tanks and two pumping stations. The database design used ESRI's Water Distribution Data Model. A major goal of the pilot project was to estimate the time and money required to

implement GIS and hydraulic modeling in the entire system (DeGironimo and Schoenberg, 2002). Another pilot project example, for the Borough of Ramsey, is described later in this chapter.

Step 7. Implementation Plan

An implementation plan identifies specific actions and resources required to implement a GIS. It provides information on schedule, cost, staffing, training, and operation and maintenance. It is used by a project or program manager to build the GIS, so it should be as detailed as possible. Phased implementation is preferable. The phases should be designed to distribute the cost evenly and to allow a comfortable pace over the project duration. Schedule success into your project schedule. Create your implementation plan with a few small but early successes to help you sustain the momentum over potential rough spots. Plan for early success by first implementing quick and high-impact applications that can be put to use right away. GIS integration with other information systems and applications (e.g., hydraulic modeling) is beneficial but technically complex. Because clearly identifiable benefits should be achieved in the earlier phases of an implementation plan, integration tasks should be scheduled for later phases where possible.

A successful implementation plan should be based on a holistic approach that achieves balance among stakeholders, technology, time, and budget.

Define key success factors before starting. Treat the success factors — such as acquiring hardware, installing and testing software, conducting training and workshops, populating the database, testing the database, and developing applications — as milestones. Project milestones provide a ruler against which success can be measured. Hitting these milestones will show the success of your GIS in clearly demonstrable ways and will allow you to measure success in small steps along the way. The milestones should specify the data sets, applications, and user access. A milestone like "Hydraulic Modeling by January 2004" leaves too much room for interpretation and, thus, dissatisfaction. A more useful milestone would be "Linking the water layer to EPANET model for the Engineering Department by January 2004."

User involvement is a key factor in the success of a GIS project. They should be incorporated in all stages of GIS development, rather than only in the needs analysis interviews. User training is paramount to the success of a GIS application project after it has been implemented.

The implementation plan should identify GIS staffing needs. This requires understanding the composition of the user community. When considering the best approach to GIS staffing and expertise investments, four factors should be evaluated: GIS expertise requirements, GIS project situation (e.g., early or later project phases), internal environment (e.g., hiring policies and procedures), and external environment (e.g., job market) (Somers, 1999a). All alternatives to fulfill GIS staffing needs including training existing staff, hiring new personnel, and using contractors and consultants should be evaluated. In-house projects do not require capital funds, and internal staff is more

familiar with source documents. Contractors can usually complete the projects faster because they have specialized experience, more workers, and higher productivity.

The long-term viability of a GIS requires ongoing maintenance. Maintenance activities include updating the GIS data layers to reflect changes such as new streets, subdivisions, or sewer lines. Such routine maintenance activities can be contracted out on an hourly basis. The typical data maintenance cost for medium-size communities is $3000 to $5000 per year. The implementation plan should provide the maintenance procedure and responsibility for each layer in the GIS. For example, standards for site plans from developers and engineers should be developed so that changes can be easily incorporated into the GIS. Today, most plans are available in AutoCAD DXF or DWG file format and might require a data conversion effort for incorporating them into a GIS. For ESRI users, the Shapefile will be a more suitable format for accepting the site plans. If required, multiple-user editing of GIS layers should be provided by selecting appropriate GIS software. Separate operations and maintenance (O&M) manuals should be developed for large installations.

Procedures for obtaining ongoing operations support should be recommended. Today, annual software support can be easily purchased with the software. On-site

Table 2.7 Implementation Priorities and Project Phases

Needs (Applications)	Total Score	Cost to Implement	Difficulty to Implement	Mission Critical	Project Phase	Data Type
Public access/general GIS data viewing	9	L	L	VH	1	Core, facility, department
Work order management	8	M	L	VH	1	Core, facility
Emergency management response	4	M	L	H	1	Core, facility
Special notification	4	L	L	M	1	Core, facility
Customer service/complaint tracking	4	M/L	L	H	1	Core, facility
Pretreatment tracking	4	L	L	M	2	Core, facility, department
Sample point monitoring and tracking	3	L	L	L	2	Core, facility, department
Biosolids tracking	3	L	L	L	2	Core, facility, department
Customer location and service verification	2	M/L	M	H	1	Core, facility
As-built indexing	2	M	M	M	1	Core, facility
Project management tracking	2	M	M/L	M/L	2	Core, facility, department
Link TV inspections to GIS/work order management	2	M/H	M/H	M	1	Core, facility, department

Notes: L = low; M = medium; H = high; VH = very high; M/L = medium/low; and M/H = medium/high.

troubleshooting support can be obtained from consultants on an hourly or incident basis.

Finally, perceptions and expectations and hence the needs of people often change with time. A GIS implementation plan, especially if it is a long one, should be flexible to accommodate changes in organizations, people, and technology.

Table 2.7 shows a sample implementation plan summary. The first two columns show the needs and their scores from Table 2.2. Columns 3 to 5 assign rankings based on cost, implementation difficulty, and mission criticality, respectively. The sixth column assigns a project phase based on columns 2 to 5. The last column shows the data type required for each identified need.

Step 8. Final Presentation

This presentation should be conducted for the same audience that attended the introductory seminar. It should summarize the findings of the needs analysis study. The presentation should also include a live demonstration of the pilot project.

NEEDS ANALYSIS EXAMPLES

Some needs analysis examples are given in the following subsections that demonstrate the results of conducting various needs analysis steps described in the preceding text.

Pittsburgh, Pennsylvania

The city of Pittsburgh performed an in-house needs analysis in 1984 that spanned a period of 4 years and took 18 months of staff time (Wells, 1991). The heart of this needs analysis was a comprehensive inventory of existing files and maps. The needs analysis listed major functions and responsibilities of each city department. The major GIS applications included:

- Maintaining infrastructure maps and inventories
- Allowing network tracing and analysis for water, sewer, and stormwater systems
- A water and sewer facilities management system
- Computing shortest routes for emergency response and work crews
- Permit processing
- Address matching
- Producing maps
- Enhancing property and business tax administration
- Automating acquisition, maintenance, and disposal of tax-delinquent properties
- Integrating census and municipal data for community planning and budget analysis
- Facilitating production of routine reports, tables, form letters, and mailing lists
- Creating and maintaining disaster preparedness map layers (hazardous-material sites, hospitals, group care facilities, etc.)

Once the needs analysis was completed, the city hired a consultant (PlanGraphics, Inc.) to procure a GIS implementation project. The consultant translated the city's GIS needs into the following technical specifications:

- Database structure
- Data entry
- Graphic data manipulation
- Database query
- Data output and display
- User-directed software development
- System operation
- Applications
- System configuration and hardware

The consultant prepared a Request for Qualifications (RFQ) and Request for Proposals (RFP). The RFQ was sent to 80 firms and the RFP was sent to 11 firms. The proposals were submitted by five vendors.

Borough of Ramsey, New Jersey

The Ramsey Board of Public Works is responsible for the operation and maintenance of water distribution and sanitary sewer collection systems throughout the Borough of Ramsey. These operations necessitate the daily use of a variety of map products and associated geographically referenced (or "georeferenced") information resources, which describe or are related to a specific location, such as a land parcel, manhole, sewer segment, or building. In an effort to more efficiently manage its geographically referenced data, the board started to explore the benefits and applications of GIS technology and computer mapping in early 1993. As a first step, with the assistance of Chester Engineers, Inc. (Pittsburgh, Pennsylvania), they started a GIS pilot project (Shamsi et al., 1996). The goals of the pilot project were to:

- Thoroughly evaluate the benefits and costs of a GIS
- Develop specifications for GIS implementation
- Confirm the suitability of tasks selected for GIS automation
- Demonstrate GIS functional capabilities
- Firmly quantify unit costs of GIS implementation
- Demonstrate GIS benefits to the departments not participating in the pilot project
- Identify any technical problems
- Provide immediate tangible benefits
- Assess the quality of existing records and procedures

The GIS pilot project produced a GIS Needs Assessment Report, a GIS Implementation Plan, and a functioning GIS demonstration system for a selected portion of the borough. With the knowledge and experience gained from the pilot project, the board was prepared to pursue a broader implementation of GIS technology. The geographic limits of the pilot project area were selected on the basis of two criteria:

- Size: large enough to allow realistic evaluation, small enough to be affordable
- Location: encompassing a portion of the borough where activities supported by priority applications are likely to occur

The selected pilot area measuring approximately 2600 ft × 2600 ft was located in the central portion of Ramsey. The pilot project was completed in 1 year for approximately $40,000 (in 1994 dollars).

From the applications listed in Table 2.1, the following two priority applications were selected for pilot testing:

1. Public Works Department: Management of infrastructure asset location information to support emergency repair and utility markout functions. The GIS will automate the storage and retrieval of location and maintenance data for water and sewer infrastructure. Expected benefits include more rapid emergency repair response, more efficient utility markouts, and more efficient scheduling of routine maintenance.
2. Planning Department: Generation of 200-ft notification lists as part of plan review process. GIS automation will generate a 200-ft buffer, identify affected parcels, retrieve lot and block numbers, retrieve property owner names and addresses, and print a notification list. GIS benefits are expected to dramatically reduce the time required to complete these tasks.

The City of Bloomington, Indiana

The City of Bloomington (with a population of approximately 60,000) started their GIS strategy with a needs analysis in 1989. They completed their base map by the end of 1993, and got the system fully functional in the next year. They stored the GIS map and Oracle RDBMS data on workstations located at Bloomington's Utilities Service Center and distributed them through a fiber-optic network to city hall and to police and fire departments. The GIS software was loaded on ten Unix workstation host machines, most of which supported a number of X-terminal and P users across the network (Goldstein, 1997).

San Mateo County, California

San Mateo County, also referred to as the "Gateway to Silicon Valley," has approximately 200,000 property parcels and covers 552 mi^2. In 2001, the county started development of an enterprise-wide GIS base map. A success factor for the project was that all aspects of its development were planned and agreed on by all the stakeholders before the project started. A GIS Steering Committee, in consultation with a larger group of county departments, hired a consultant (GIS Consultants, Inc.) to conduct a needs analysis, which resulted in a strategic plan. Thereafter, a specific implementation plan was designed for base-map construction, followed by detailed and complete specifications and preliminary cost estimates. Formal needs analysis efforts were credited for the timely completion of base-map creation, hardware and software purchase, and training (Joffe et al., 2001).

CHAPTER SUMMARY

This chapter described the importance and benefits of a needs analysis prior to implementing a full-scale GIS project. Various steps involved in needs analysis were explained. When carefully done, needs analysis clarifies the project's specific needs,

provides a process for arriving at interdepartmental understandings about the objectives and limits of a project, and establishes a detailed basis for developing GIS maps and applications.

CHAPTER QUESTIONS

1. What is needs analysis? How does it compare to strategic planning and facility master planning?
2. What steps are recommended for needs analysis? Which step is most critical?
3. What is a pilot project? What are the benefits of conducting a pilot project?
4. How would you prepare a needs analysis study for a water system mapping project?
5. You are tasked with integrating a large city's H&H model with the city's GIS. What will be your first step and why?

CHAPTER 3

Remote Sensing Applications

Can a satellite 400 miles above the ground surface help you locate a leaking pipe? Read this chapter to find out.

The Landsat 7 Enhanced Thematic Mapper (ETM+) scene of the lower Chesapeake Bay region acquired on July 5, 1999 (Image courtesy of USGS).

LEARNING OBJECTIVE

The learning objective of this chapter is to comprehend the applications of remote sensing technology in the water industry.

MAJOR TOPICS

- Remote sensing satellites
- Applications of satellite imagery
- Types of remote sensing data
- Digital orthophotos
- Using remote sensing for land-use classification
- Image processing software
- Anticipated future trends

LIST OF CHAPTER ACRONYMS

DEM Digital Elevation Model
DOP Digital Orthophoto
DOQ Digital Orthophoto Quadrangle
DOQQ Digital Orthophoto Quarter Quadrangle
LIDAR Laser Imaging Detection and Ranging
LULC Land Use/Land Cover
TM Thematic Mapper (onboard Landsat satellite)
USGS United States Geological Survey

ALBANY COUNTY'S REMOTE SENSING APPLICATION

Public-domain digital aerial photography data, such as USGS digital orthophoto quadrangles (DOQs) and digital orthophoto quarter quadrangles (DOQQs), usually become outdated in rapidly developing areas. For such areas, high-resolution satellite imagery may be a cost-effective source of more recent overhead images.

Albany County, located in southeastern Wyoming, covers 4,400 mi^2, has a student-based population of 30,000, and has 1,600 mi of roads. For rural communities such as Albany County, building a GIS from scratch can be an expensive endeavor due to lack of resources. The County's day-to-day mapping functions required a data layer of imagery for the entire county. Various data options were reviewed, including aerial flights, existing DOQs, and satellite imagery. New aerial imagery was eliminated because it was too expensive. Existing DOQs were not suitable because they were 7 years old and did not reflect recent county growth trends. In addition, costs associated with updating the County's existing digital aerial photography exceeded $100,000. High-resolution satellite imagery, on the other hand, allowed the County to have high-resolution up-to-date views of the entire county for $32,000. For 85 mi^2 of populated areas, the County selected 1-m pan-sharpened IKONOS satellite imagery (described later in this chapter). For the rest of the county, 90 quads of

CARTERRA DOQ 5-m black and white (B&W) imagery was selected. Both products were produced by Space Imaging (Thornton, Colorado). Thanks to this geographic-imaging approach, planning tasks previously requiring months to complete took only days after the County implemented this project (Frank, 2001).

In the Albany County of Wyoming, addition of high-resolution up-to-date imagery to GIS data reduced the completion of typical planning tasks from months to a few days.

INTRODUCTION

The technologies that are commonly used in conjunction with GIS are commonly referred to as GIS-related technologies. Examples include remote sensing, global positioning system (GPS) surveying, the Internet, and wireless technologies. This chapter will focus on remote sensing, one of the most successful GIS-related-technologies. Other related technologies are described elsewhere in the book.

Remote sensing allows obtaining data of a process from a location far away from the user. Remote sensing can, therefore, be defined as a data collection method that does not require direct observation by people. Remote sensing is the process of detection, identification, and analysis of objects through the use of sensors located remotely from the object. Three types of remote sensing systems are useful in the water industry:

1. Aerial photographs
2. Satellite imagery
3. Radar imagery

The data from these systems are commonly referred to as remote sensing or remotely sensed data. Sometimes, remote sensing data are incorrectly confused with supervisory control and data acquisition (SCADA) data used to operate water and wastewater treatment plants. Remote sensing data collected using airplanes are called aerial photographs or aerial photos. Digital remote sensing data collected from satellites are called satellite imagery or images. Digital pictures of the Earth are taken by satellites from 400 to 500 mi above the ground compared with aerial photographs that are taken by aircraft from 1 mi above the ground (for low-altitude photography) to 7 to 8 mi above the ground (for high-altitude photography). The chart in Figure 3.1 shows the altitude difference between the aircraft- and satellite-type remote sensing systems. Radar imagery or images are another type of remote sensing data but their usage is not widespread in the water industry. Although the definition of remote sensing includes aerial photos and radar data, remote sensing is often considered synonymous with satellite imagery.

The American Society for Photogrammetry and Remote Sensing (ASPRS) values the U.S. remote sensing industry at about $1.3 billion (as of 2001) and forecasts 13% annual growth, giving values of $3.4 billion by 2005 and $6 billion by 2010. The industry currently consists of about 220 core companies employing about 200,000 employees in the areas of remote sensing, photogrammetry, and GIS imaging. A 2001

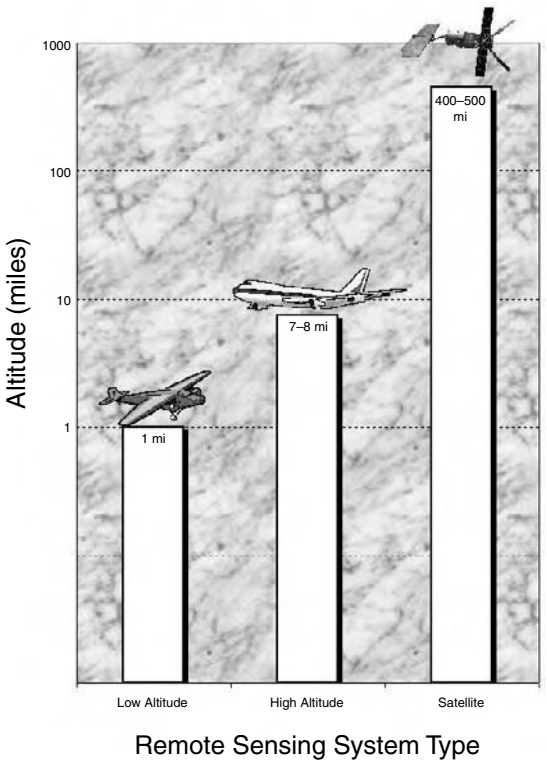


Figure 3.1 Altitude difference in aerial photography and satellite imagery.

ASPRS study concludes that utilities are one of the greatest untapped potential markets and that a shortage of trained workers is one of the greatest challenges to the growth of the remote sensing industry (Barnes, 2001a).

Although vector GIS data are still an important and vital tool for many water industry applications, the newer raster GIS applications of satellite imagery are beginning to make a major move into the GIS and mapping market. The benefits of satellite imagery are (Schultz, 1988):

1. They enable aerial measurements in place of point measurements.
2. They offer high spatial and/or temporal resolution.
3. All information is collected and stored at one place.
4. Data are available in digital form.
5. Data acquisition does not interfere with data observation.
6. Data can be gathered for remote areas that are otherwise inaccessible.
7. Once the remote sensing networks are installed, data measurement is relatively inexpensive.

Satellite imagery is stored in a pixel (raster) format that makes it ideally suited for incorporation into a GIS (Engman, 1993). Thus, satellite imagery can be treated as raster-type GIS data. Image processing equipment and methods can be used to

extract useful information from hard copy and digital images and combine it with other data layers in a GIS. Image data sources including scanned paper maps, aerial photographs, and satellite imagery can be used in a GIS when reprojected as image maps. Projected images can be used as a background or as a base map upon which other vector layers are overlaid.

Casual GIS users can easily import remote sensing imagery into their GIS projects as an image theme (or layer). However, advanced remote sensing applications and image analyses require formal remote sensing training and digital image processing skills. The incorporation of remote sensing data in a GIS requires a digital image processing software such as ERDAS IMAGINE, Geomatica, ER Mapper, or ENVI, or a raster GIS software with image processing capability, such as ArcGRID or IDRISI. Such programs are described later in this chapter.

These are exciting times both for the GIS and the remote sensing industries, thanks to dramatic price and performance breakthroughs in GIS hardware and software. The increasing use of GIS is contributing to a renewed interest in satellite imagery by nongeographers, such as civil and environmental engineers. Although GIS technology is promoting the use of satellite imagery, satellite imagery is also in turn advancing the use of GIS. Although non-GIS stand-alone image processing software can be used for exploring satellite imagery, those with GIS capabilities are more suitable because they can combine imagery with additional information, such as demographic and topographic data (Corbley, 2000).

REMOTE SENSING APPLICATIONS

Satellite imagery is not restricted to the visible (0.4 to 0.7 μm wavelength) part of the electromagnetic spectrum. Satellite sensors can record Earth images at wavelengths not visible to the human eye, such as near-infrared and thermal-infrared bands. Different satellite bands provide information about different objects and conditions of the Earth. For example, thermal-infrared band (10.4 to 12.5 μm wavelength) data are useful for soil–moisture discrimination. These bands of satellite data can be used as different data layers in a GIS for further analysis.

Remote sensing applications in the water industry are as diverse and numerous as the GIS applications themselves. Typical examples are listed below:

1. Satellite remote sensing has contributed to water resources applications and research for three decades (Jackson, 2000). Remote sensing data are especially useful in watershed hydrologic modeling. Satellite imagery can be used to estimate input parameters for both the lumped-parameter and distributed-type hydrologic models.
2. Satellite imagery can be used for delineating watersheds and streams. For example, SPOT satellite's stereographic capability can generate topographic data. Terra satellite can provide digital elevation models (DEMs) from stereo images. (These and other satellites are discussed later in this chapter.) Topographic and DEM data collected by satellites can be processed in GIS for automatic delineation of watershed boundaries and streams.
3. Remote sensing data are used for land-use classification. GIS can help to refine or verify the imagery-based land-use classes.

4. Satellite and radar data can be used to estimate the area and intensity of rainfall.
5. Remote sensing can produce surface temperature data through thermal-infrared images.
6. Microwave remote sensing can produce soil-moisture data.
7. Remotely sensed temperature and moisture data can be combined to estimate evaporation and evapotranspiration rates.
8. Remote sensing data are used to estimate vegetation indices and the leaf area index. These parameters can be combined to delineate areas where a subsurface supply of water is available for vegetation.
9. Remote sensing can be used in real time flood forecasting with a distributed hydrologic model into which radar rainfall data can be input.
10. Other applications (Singh, 1995; ASCE, 1999) are:
    - Utility routing
    - Weather forecasting
    - Environmental impact assessment of large water resources projects
    - Snow and ice conditions (microwave region)
    - Forecasting seasonal and short-term snowmelt runoff
    - Evaluation of watershed management strategies for conservation planning
    - Inventory surface water, such as rivers, lakes, reservoirs, swamps, and flooded areas
    - Water quality parameters such as algae, chlorophyll, and aquatic life
    - Thermal and chemical pollution and oil spills
    - Drought assessment and forecasting
    - Geologic and geomorphologic information
    - Groundwater mapping

REMOTE SENSING SATELLITES

Satellite data became available to water industry professionals in 1972 when the U.S. government launched the first Landsat satellite, which was specifically designed to provide imagery of the Earth (Miotto, 2000). In the late 1970s and early 1980s, a second generation of Landsat satellites was developed. Landsats 4 and 5 were launched in July 1982 and March 1984, respectively. They were equipped with two instruments:

- Multispectral scanner (MSS) having 80-m resolution and 4 spectral bands
- Thematic mapper (TM) having 30-m resolution and 7 spectral bands

MSS sensors capture imagery at different wavelengths of light to produce color images. Landsat 4 was retired in 1991, and Landsat 5's MSS sensor failed in October 1993. The successor satellite, Landsat 6, failed to achieve orbit in 1993. To keep the imagery flowing, Landsat 7 was launched on April 15, 1999.

Popular satellite-based sensors and platforms include Landsat MSS and TM, AVHRR, AVIRIS, SPOT XS, GOES, SEASAT, SIR, RADARSAT, SRTM, TOPSAT, ERS-1 and 2, and JERS-1 (Luce, 2001; Lunetta and Elvidge, 1998). The remote sensors that provide hydrologically useful data include aerial photographs, scanning radiometers, spectrometers, and microwave radars. The satellites that provide hydrologically useful data are the NOAA series, TIROS N, SPOT, Landsat, and the

geostationary satellites GOES, GMS, and Meteosat. Satellites can capture imagery in areas where conventional aircraft cannot fly. However, bad weather, especially cloud cover, can prevent satellites from capturing imagery (Robertson, 2001).

SPATIAL RESOLUTION

The spatial resolution of an image is defined as the size of the smallest feature that can be discerned on the image. Spatial coverage is defined as the area of the Earth's surface captured by the image. In general, the higher the spatial resolution, the smaller is the spatial coverage. For example, NASA's Terra Satellite MODIS sensor has 36 spectral channels at 250 m, 500 m, and 1 km. A standard MODIS image covers 1200 km × 1200 km, whereas a standard IKONOS satellite image covers 11 km × 11 km. At such a large spatial coverage, MODIS spatial resolution is more than 50 times coarser than the IKONOS imagery (Space Imaging, 2001). In 2001, the approximate number of 30-m (or better) resolution satellites in the world was 30, and the number of 10-m (or better) resolution satellites was 14 (Limp, 2001).

Figure 3.2 provides a comparison of image resolution. It shows five images at various resolutions for the same geographic area (Gish, 2001). The top-left image, with the highest resolution, is a 0.15-m (0.5-ft) B&W orthophoto taken in 1993. The top-right image is a 0.6-m (2-ft) 1998 B&W orthophoto. The center-left image is a 1-m (3.28-ft) 1999 color-infrared orthophoto taken in invisible light in the infrared bands. The center-right image is a simulated B&W SPOT image with a 10-m (32.8-ft) resolution. Finally, the bottom image has the lowest resolution of 30 m (98.4 ft) and consists of Landsat 7 TM color imagery taken in 2000.

In remote sensing, B&W or gray-scale imagery is called *panchromatic* and color imagery is called *multispectral*. Panchromatic satellite-imagery resolution varies from 15 m (49 ft) for the Landsat 7 satellite, 10 m (33 ft) for the French SPOT satellite series, 5 m (16 ft) for the Indian Remote Sensing series, 1 m (3.2 ft) for the IKONOS satellite (Gilbrook, 1999), to 60 cm (2 ft) for the QuickBird-22 satellite. Until recently, satellite images tended to have very low resolutions. In January 2000, IKONOS high-resolution satellite imagery became available in the commercial marketplace for the first time.

Based on their spatial resolution, remote sensing data can be divided into three categories:

1. Low-resolution data corresponding to imagery with a resolution greater than 30 m
2. Medium-resolution data corresponding to imagery with a resolution between 5 and 30 m
3. High-resolution data corresponding to imagery with a resolution less than 5 m

Low-Resolution Satellite Data

The United States Earth Observing System (EOS) satellites are an excellent source of low- and medium-resolution satellite data. There are four EOS satellites currently in orbit: Landsat 7, QuickSAT, ACRIMSAT, and Terra. Terra, launched by NASA in December 1999, has three remote sensing instruments that could be

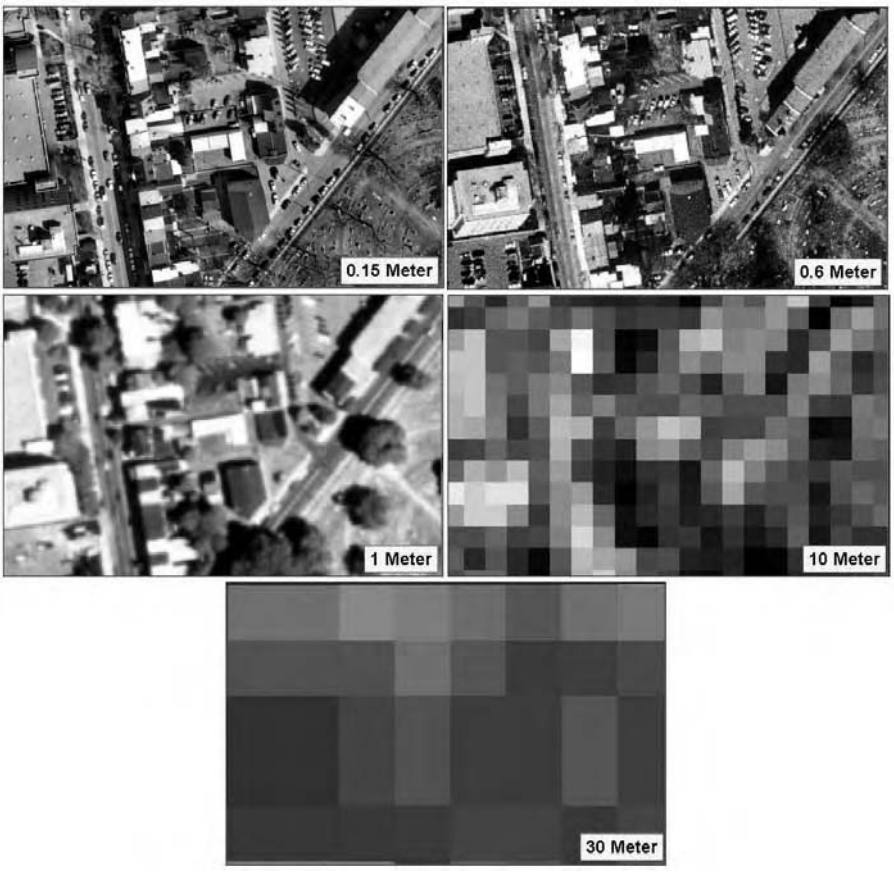


Figure 3.2 Image resolution comparison. Top left: 0.15-m B&W orthophoto (1993); top right: 0.6-m B&W orthophoto (1998); center left: 1-m color infrared orthophoto (1999); center right: 10-m simulated SPOT; bottom: 30-m Landsat TM (2000).

useful for certain water resources applications: Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Moderate Resolution Imaging Spectroradiometer (MODIS), and Multiangle Imaging Spectroradiometer (MISR). ASTER provides digital elevation maps prepared from stereo images. MODIS provides data on cloud- and snow-cover characteristics. MISR data can distinguish different types of clouds, land cover, and vegetation canopy. Although low-resolution satellite imagery works well for regional level studies, it is not very useful in water industry applications.

Medium-Resolution Satellite Data

Table 3.1 provides a summary of medium-resolution satellites. Landsat 7 is the most recent satellite in the Landsat series. By May 2001, Landsat 7 had captured more than 200,000 15-m scenes throughout the world. The Enhanced Thematic

Table 3.1 List of Major Medium-Resolution Satellites

Feature	Landsat 7	SPOT 4	IRS-1C
Company	USGS	French Government	Indian Government
Launch date	April 15, 1999	1998	December 28, 1995
B&W resolution	15	10	5
Color resolution	30	20	23
Swath width (km)	185	120	70
Global cover repeat days	16	26	24

Mapper Plus (ETM+) sensor onboard Landsat 7 provides 15-m panchromatic and 30-m multispectral resolutions. Landsat 7 offers imagery of the highest resolution and lowest price of any Landsat. The USGS ground-receiving station in Sioux Falls, South Dakota, records 250 Landsat scenes a day that are available online within 24 hours. Landsat 7 is expected to have a design life of 5 years. Each Landsat image covers about 10,000 mi^2. Landsat 7 is very useful in water resources applications.

Figure 3.3 shows a modified Landsat TM image for southwestern Pennsylvania, which can be used in a GIS to consistently map land use/land cover (LULC) throughout the state. These images, called Terrabyte, are extracted from the 30-m resolution TM data using an extractive process based on a research trust at Penn State University under cooperation between the Office for Remote Sensing of Earth Resources in the Environmental Resources Research Institute and the Center for

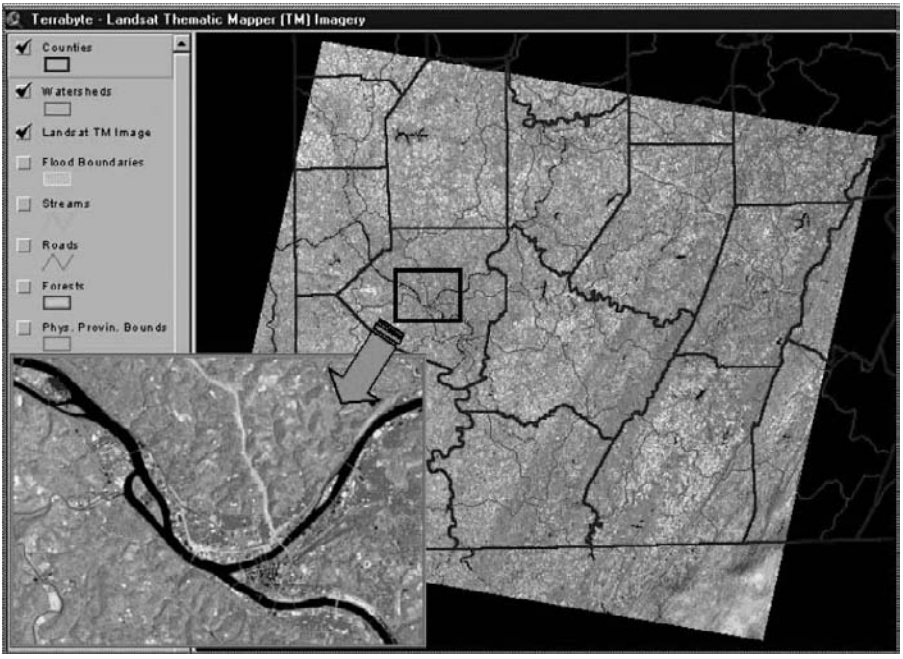


Figure 3.3 Terrabyte Landsat Thematic Mapper image for southwestern Pennsylvania.

Statistical Ecology and Environmental Statistics in the Department of Statistics, with sponsorship from the National Science Foundation and the Environmental Protection Agency. The Terrabyte images are not intended to provide fine detail such as individual buildings at the site level, but rather to convey a sense of landscape organization. Each pixel record occupies one byte; hence the name Terrabyte. Terrabyte condensations for ten satellite scenes will fit on one CD-ROM, whereas two scenes of original satellite data would more than fill one CD-ROM. Terrabyte CD-ROMs of Pennsylvania data have been distributed by Pennsylvania Mapping and Geographic Information Consortium (PaMAGIC) (www.pamagic.org).

In addition to the EOS satellites, France's SPOT 4 satellite provides 10-m panchromatic and 20-m color imagery. In the U.S., 60 × 60 km SPOT scenes cost $750 (pre-1998) to $1500 (post-1998). India's Indian Remote Sensing (IRS) satellite IRS-1C provides 5-m panchromatic and 23-m or 188-m color imagery. Commercial companies and government agencies around the world had plans to launch more than 25 medium-resolution (30 m or better) satellites by the end of 2003.

High-Resolution Satellite Data

High-resolution data correspond to imagery whose resolution is less than 5 m. Traditionally, water industry professionals have purchased aerial photography services on an as-needed basis, which is costly and time-consuming. Now, thousands of square miles of GIS-ready seamless imagery is available in various formats with the promise to bring remote sensing data to any desktop (Robertson, 2001). Until recently, some water industry professionals used 5-m panchromatic imagery from India's IRS-1C satellite or 10-m panchromatic imagery from France's SPOT 4 satellite for their high-resolution data needs. The recent launches of IKONOS (1-m) and QuickBird-2 (60-cm) satellites have changed this by starting to provide high-resolution panchromatic imagery, which meets the U.S. National Map Accuracy Standards for 1:5000-scale maps. 1-m imagery represents an accuracy level commensurate with 1:2400 mapping, which is more than adequate for many planning and H&H modeling applications.

GIS applications are poised to bring the recently available high-resolution satellite imagery directly to the dispatch office of a water, wastewater, or stormwater utility.

High-Resolution Satellites

There are three major satellites that are providing high-resolution satellite imagery today: IKONOS, OrbView, and QuickBird. Table 3.2 provides more information about high-resolution satellites.

High-resolution imagery shows detailed features, such as houses, schools, street centerlines, rights-of-way, trees, parks, highways, and building facilities. They can be used for base-map and land-registry updates, infrastructure mapping analysis and management, natural resource inventories, ecological assessments, transportation mapping, and planning the construction of new highways, bridges, and buildings (Murphy, 2000).

Table 3.2 List of Major High-Resolution Satellites

Feature	QuickBird-2	IKONOS	OrbView-3
Company	DigitalGlobe	Space Imaging	ORBIMAGE
Launch date	October 18, 2001	September 24, 1999	June 26, 2003
B&W resolution (m)	0.61	1	1
Color resolution (m)	2.5	4	4
Swath width (km)	16.5	12	8
Global cover repeat days	148	247	3
Standard scene size	40 km × 40 km	13 km ×13 km	User defined
Web site	digitalglobe.com	spaceimaging.com	orbimage.com

DigitalGlobe's (formerly Earth Watch) QuickBird-1, designed for 1-m panchromatic and 4-m color resolution, failed to achieve its proper orbit after being launched from Plesetsk, Russia, on November 20, 2000. QuickBird-1's unfortunate failure is a good example of a "beneficial loss." QuickBird-2 was launched on October 18, 2001, on a Boeing Delta II rocket from Vanderberg Air Force Base in California. At the time of this writing, QuickBird-2 provides the only commercial satellite imagery of resolution less than 1 m. Figure 3.4 shows the sample QuickBird imagery taken in 2002 and a photograph of the San Diego Convention Center area in California that hosts the world's largest GIS conference (ESRI Annual User Conference) every year. Note that the boats at the marina and the north and south towers of the Marriot

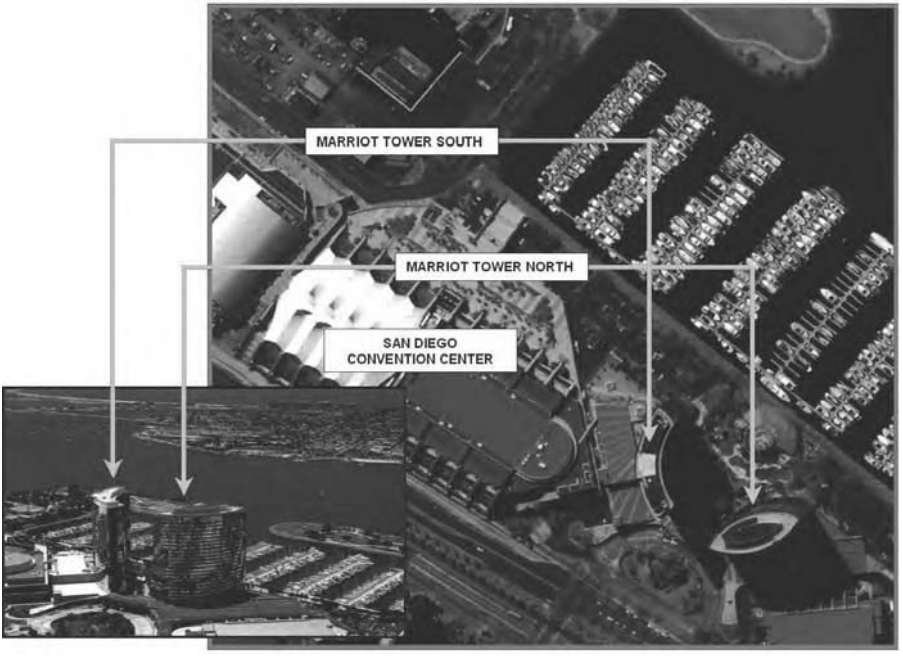


Figure 3.4 Sample QuickBird imagery for San Diego, California. (Image courtesy of DigitalGlobe.)

Hotel adjacent to the convention center are clearly visible on the imagery. DigitalGlobe is planning to launch another satellite dubbed WorldView in 2006, which will be capable of collecting 50-cm panchromatic and 2-m multispectral imagery. WorldView's 800-km-high orbit will allow the satellite to visit imagery collection sites more frequently, letting users repeat their image acquisition about once a day.

IKONOS provides 1-m panchromatic and 4-m multispectral (color) imagery. The satellite weighs 1600 lb and orbits 438 mi above the ground surface. IKONOS products are available under the CARTERRA™ brand name in TIFF and GeoTIFF format. CARTERRA also provides DOQ — B&W, color, or false color IR 5-m imagery, cut into a convenient 7.5-min USGS quadrangle format. Orthorectified CARTERRA DOQs provide an image map suitable for water resources management, urban and rural planning, change detection, and map creation and revision.

High-Resolution Imagery Applications

GIS applications are poised to bring the recently available high-resolution satellite imagery directly to the dispatch office of a water, wastewater, or stormwater utility. For years, aerial photography has been used in many utility GISs, and that use will continue because of its submeter resolution. High-resolution satellite imagery is now available commercially at a reasonable cost. Therefore, when a 1-m resolution is sufficient, satellite imagery can be used as a base map instead of orthorectified digital orthophotos. High-resolution satellite imagery provides digital data at a fraction of the cost people pay for aerial photographs of the same level of accuracy. There is no question that the launch of high-resolution satellites marks a new era in the remote sensing industry.

Typical applications of high-resolution satellite imagery for the water industry are described in the following list:

- High-resolution satellite imagery can enable the water and wastewater system utilities to gather information quickly and inexpensively, allowing them to perform daily operations more efficiently.
- Multispectral imagery can detect vegetation stress before it is visible to the naked eye. Watermain leaks and manhole overflows can impact the soil and vegetation conditions around them. These potential indicators may be used to identify leaks and overflows. *Thus, although satellites cannot directly locate leaking pipes, they can provide the surrogate data that can lead to locating them.*
- High-resolution imagery is especially useful in remote areas of the world where there are no governments and commercial archives and where cost and regulatory hurdles preclude aerial missions.
- Medium-resolution imagery cannot capture some landscape characteristics, such as distribution of shrubs vs. bare ground or gaps in the forest crown (Space Imaging, 2001). High-resolution imagery bridges the gap between field measurements and medium-resolution imagery, providing a continuum from point measurements to medium resolution. High-resolution imagery can also be used to "ground truth" the low- and medium-resolution imagery.
- High-resolution imagery can be used to study urban growth and detailed urban land-use mapping. It can be used to identify growth trends in order to develop the necessary infrastructure in advance.

- Damage from natural disasters (earthquakes, hurricanes, fires, and floods) can be analyzed and response plans prepared. High-resolution imagery can be used to prepare more accurate flood-prediction computer models, monitor stormwater runoff, and study erosion and sedimentation.
- Right-of-way encroachments can be identified by periodically running automated change-detection routines on new imagery.
- Inexpensive pipeline siting and corridor selection can be performed using least-cost path analysis.
- Ecological assessments can be conducted.

Data Sources

TerraServer was started as a joint research project by Aerial Images, Microsoft, USGS, and Compaq. It is considered one of the world's largest online atlases of high-resolution satellite imagery and aerial photography (Thoen, 2001).

In 2001, USGS, PCI Geomatics, Oracle, and Sun Microsystems teamed up to provide a new data delivery service called Real-time Acquisition and Processing of Imagery Data (RAPID). It provides same day service for conversion of TM imagery into easy-to-use data that can be downloaded using an Internet connection. Using RAPID, users can have georeferenced, GIS-ready processed imagery within 10 min of receipt from USGS.

Some consumer-oriented companies are also selling high-resolution satellite imagery. For example, Eastman Kodak Company's CITIPIX imagery database consists of 95 major North American metropolitan areas, including 7000 cities and towns and 600 U.S. and Canadian counties. This ready-to-use "Earth Imaging Products" consist of orthorectified imagery in 6-in., 1-ft, 2-ft, and 1-m resolutions. Kodak's 24-bit color images exceed National Map Accuracy Standards' accuracy requirement at 1:1200. These products are intended for applications in architecture, engineering, construction, telecommunication, utilities, insurance, and real-estate industries as well as local, state, and provincial governments.

The cost of spatial data is falling rapidly due to competition in data acquisition, processing, and distribution. As satellite imagery has become more widely accepted, its unit cost has started to decline. For example, Landsat-4 and -5 imagery used to cost $4400 per scene; now the same scene costs $600. After the launch of QuickBird-2 and OrbView-3 satellites, the price of IKONOS imagery has come down from $62/km^2 to $29/km^2 (with a 100 km^2 minimum order) — a decrease of over 50%.

DIGITAL ORTHOPHOTOS

Digital orthophotos are a special type of high-resolution remote sensing imagery. Traditional aerial photos contain image displacements caused by camera lens distortion, camera tip and tilt, terrain (topographic) relief, and scale (Michael, 1994). Because of these problems, an aerial photograph does not have a uniform scale, and therefore, it is not a map. The distortions are removed through a rectification process to create a computer file referred to as a digital orthophoto (DOP). The image rectification is done with the help of geodetic surveying and photogrammetry. A DOP

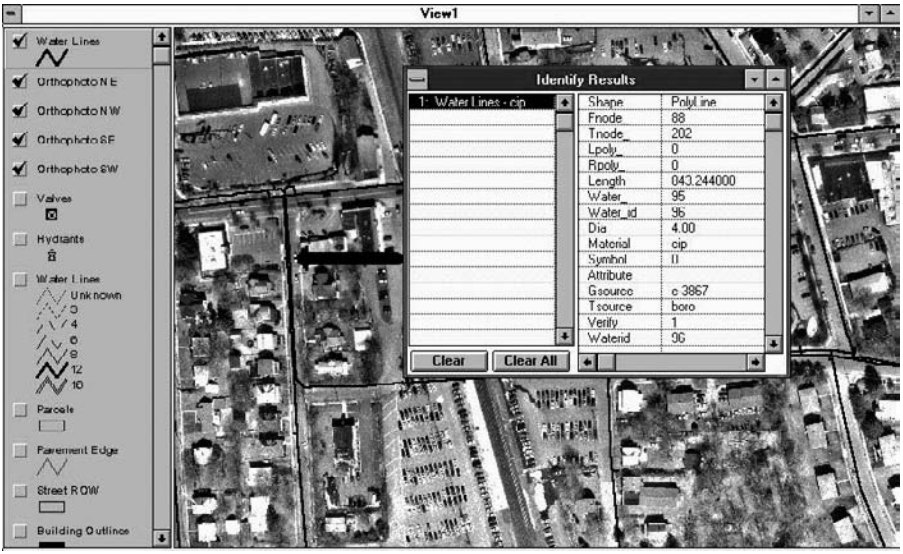


Figure 3.5 Digital orthophoto base map for water system mapping.

is a uniform-scale photographic image and can be considered a photographic map. As the name implies, orthophotos are orthographic photographs or simply photo maps.

DOPs are very detailed, can be easily interpreted, and provide excellent accuracy that can be easily quantified and verified. DOPs are a valuable source for developing an accurate landbase in a GIS mapping project. Because it is a photographic image, the DOP displays features that may be omitted or generalized on other cartographic maps. This makes the digital orthophoto valuable as a base map layer in a GIS. Figure 3.5 shows a DOP base map for a water distribution system mapping project. The DOP has been processed to remove scale distortion and create an accurate and true-to-scale base map with an accuracy of ±1.25 ft. Waterlines have also been plotted on the digital orthophoto.

USGS Digital Orthophotos

A digital orthophoto quadrangle (DOQ) is an orthorectified raster image of a low-altitude USGS aerial photograph in Universal Transverse Mercator (UTM) projection in North American Datum of 1983 (NAD83). DOQs have the geometric properties of a map and meet the National Map Accuracy Standards. DOQs are based on the 1:80,000-scale National High Altitude Photography (NHAP) aerial photos with an altitude of 40,000 ft. They cover an area equal to 7.5 min (1:24,000) USGS quads (quadrangles); hence their name. They have a resolution of 2 m and an accuracy of 40 ft. They are available as 5-MB compressed JPEG files that can be purchased for $35.5 per county CD.

Digital orthophoto quarter quadrangles (DOQQs) are based on the 1:40,000-scale National Aerial Photography Program (NAPP) aerial photos with an altitude

Figure 3.6 USGS digital orthophoto quarter quadrangle (DOQQ) for Washington, D.C. (Image courtesy of USGS.)

of 20,000 ft. They cover an area equal to a quarter of a 7.5-min USGS quads (3.75 min of latitude by 3.75 min of longitude in geographic extent); hence their name. They have a resolution of 1 m and an accuracy of 33 ft. DOQQs also have a UTM map projection system and NAD83 datum. A sample DOQQ image for Washington, D.C., is shown in Figure 3.6.

DOQQs are available as 40 to 50 MB uncompressed files that can be purchased for $60 per file on CD-ROM from the USGS EarthExplorer Web site. State or regional mapping and spatial data clearinghouse Web sites are the most valuable source of free local spatial data. For example, Pennsylvania Spatial Data Access system (PASDA), Pennsylvania's official geospatial information clearinghouse and Pennsylvania's node on the National Spatial Data Infrastructure (NSDI), provides free downloads of DOQQ and other spatial data.

Because DOQQs are based on the NAD83 datum, most U.S. ArcView 3.x users who want to use DOQQs in their GIS projects will have to reproject their NAD27 vector themes into NAD83. NAD83 is an Earth-centered datum (GRS 80 ellipsoid) unlike NAD27, which is based on an arbitrary starting point in Meades Ranch (Kansas). The positions or points or features based on these reference datums will differ considerably. Though reprojected themes will line up reasonably well with features in DOQQs, a more precise alignment can be achieved by adjusting the false easting and northing values to "tweak" the theme (Miller, 2000).

Like DOPs, DOQs and DOQQs also can be used as GIS base maps to overlay other thematic data layers. DOQs can be used in land management, habitat analysis, evacuation planning, and in many other areas (Miller, 2000).

Case Study: Draping DOQQ Imagery on DEM Data

The city of Ventura, California, is located 70 mi north of Los Angeles, covers almost 21 mi^2, and is home to 105,000 people. In an effort to redefine its GIS, the City acquired 11 USGS DOQQs in 1999. The DOQQs were mosaicked and reprojected to California State Plane with ERDAS IMAGINE software and compressed with LizardTech's MrSID software to create a seamless image. They also created a panoramic view of the city using IMAGINE VirtualGIS software by draping the DOQQ mosaic over 10-m DEMs obtained from USGS and Intermap Technologies (Ottawa, Ontario, Canada). The image draping demonstrated the value of orthorectified imagery and helped the City to derive other GIS data layers, such as slope, aspect, and shaded relief (ERDAS, 2001a).

EXAMPLES OF REMOTE SENSING APPLICATIONS

A USGS project helping to restore the original water flow of Everglades is a good example of blending various types of remote sensing data. This project uses aerial videography, laser imaging detection and ranging (LIDAR)*-based topography, and Landsat-based vegetation to create hydrodynamic models of water flow (Miotto, 2000).

The Arizona Department of Water Resources (ADWR) uses satellite data (Landsat and SPOT) and GIS software (ERDAS IMAGINE and ArcInfo) to monitor water use, regulate water rights, and enforce limits on irrigated acreage expansion. ADWR purchases growing-season imagery at an annual cost of $15,000. An investigation is started if the images show crop growth that does not correspond to water rights (McKinnon and Souby, 1999).

A 1997 USGS study for the California Gulch Superfund Site in Leadville, Colorado, demonstrated an application of remote sensing in locating and pinpointing sources of acid mine drainage and a tremendous variety of other surface materials. Mineral maps produced by the study helped expedite the cleaning of hazardous material and reduced the cleanup cost. These maps were produced using a new tool called imaging spectroscopy, the latest in remote sensing science.

Satellite imagery can be used for irrigation monitoring and assessment, or to evaluate erosion potential or landslides. Other remote sensing applications for the water industry include land-use classification, soil moisture mapping, and estimation of meteorological data. These applications are described in the following subsections.

LULC Classification

The San Diego Association of Governments (SANDAG) deals with one of the nation's largest county jurisdictions covering more than 4200 mi^2. Before 1988, SANDAG used costly and time-consuming aerial-photography and photo-interpretation techniques to create LULC maps and updated them only once every 5 years. To meet

* LIDAR is a new remote sensing technology that measures ground surface elevation from an airplane to create DEM data. Additional information is provided in Chapter 4 (DEM Applications).

the special challenge of keeping track of this rapidly changing area in a cost-effective manner, SANDAG turned to GIS. It used raster GIS and image processing software ERDAS, vector GIS software ArcInfo from ESRI, color-infrared aerial photographs, and satellite imagery. Switching to satellite imagery and GIS as a land inventory tool allowed SANDAG to see the region in a new way and permitted rapid change detection. The GIS-based LULC-mapping approach provided SANDAG with current and verified LULC data for modeling transportation, infrastructure, and water needs (Kindleberger, 1992).

In 2003, a digital database of land-cover imagery and vectors was created that includes the major landmasses of the entire world. Called GeoCover LC (Land Cover), this was the first worldwide LULC database prepared using medium-resolution Landsat 7 TM imagery.

LULC classification is one of the most common applications of satellite imagery. Remote sensing data are a valuable source of information for land-use modeling. Derivation of LULC classes from low-level aerial photography is referred to as the *conventional* method compared with the remote sensing techniques that employ satellite imagery. Studies have shown that the remote sensing techniques for LULC classification are more cost-effective. The cost benefits have been estimated on the order of 6 to 1 in favor of the satellite imagery approach. Although remote sensing-based land-use statistics may not be as detailed as those derived using the conventional manual method, many computed parameters such as runoff curve numbers and discharges are nearly the same (Engman, 1993).

The choice of sensors for LULC applications is dictated by the time and space resolution needed for interpretation. The sensor chosen must permit a suitable classification of LULC. The quality of this classification will depend on the LULC types, the quantity of images available, and the dates on which they were taken as well as the classification technique used. Ambiguities in interpretation can be reduced by using field information (ground truthing) to improve classifications. The number of LULC categories depends on the intended application. For example, in watershed nutrient modeling this number must permit the estimation of total nutrient export from a subwatershed and to compare the possible effects of land-use changes. At a minimum, urban, agricultural, grasslands, and forested areas must be clearly distinguished. A distinction between types of farming will provide better nutrient estimates (Payraudeau et al., 2000).

When collecting LULC data, keep in mind the four "Cs": currency, construction, categories, and consistency. Currency determines how current your LULC data are. Construction pertains to new developments that can significantly alter land use and affect soil erosion, sedimentation, and even water quality. Categories represent the LULC classes (polygons). One may not want to have an "agricultural" class created by lumping together cropland, pasture, and dairy farms. And, finally, consistency determines how consistent in time, scale, resolution, and classification various sources of LULC data are (Slawecki et al., 2001).

Vegetation is an important part of a watershed's ecosystem. To preserve and maintain the watershed ecosystems, resource managers need high-quality vegetation maps to monitor temporal vegetative changes and to pinpoint habitats and species likely to be affected by management decisions. Based on the spectral return from

the infrared imagery and using classification techniques of a remote sensing software package, people can identify trees and other vegetation growing in a watershed and even determine the total irrigated area. California Department of Forestry and Fire Protection (Sacramento, California) updates California's LULC mapping and monitoring program on a 5-year cycle. The program utilizes image classification and GIS modeling to generate data that describe the condition and extent of various land-cover types, as well as the magnitude and cause of such changes. Data sources include DOQs, aerial photographs, and satellite images from Landsat TM, Indian Remote Sensing, and SPOT. LULC layers are produced using automated image classification of satellite imagery. This approach efficiently and consistently maps large areas at a low cost. Next, the LULC layer is converted to vector format and hand-edited with high-resolution DOQ as a backdrop. GIS models are run in ESRI's ArcInfo GRID program to label each LULC polygon with a vegetation type. The updated LULC map is created by adding the new 5-year LULC change layer to the old LULC map (Rosenberg et al., 2001).

Dry- and wet-weather flows from sewersheds depend on land use. The hydrologic and environmental effects of land do not depend on the amount of land-use change alone. The location of land-use change is equally important. For example, a development adjacent to a water body might produce a greater impact on water quantity and quality. GIS and remote sensing are ideally suited to studying the temporal and spatial variation of land use. Remote sensing has been recently recognized as a tool that holds great promise for water quality monitoring activities. In 2003, three U.S. agencies, the Environmental Protection Agency (EPA), National Aeronautic and Space Administration (NASA), and the National Oceanic and Atmospheric Administration (NOAA) entered into a partnership to more fully explore how remote sensing may support state water quality management activities (Mehan, 2003).

Digital LULC data are used to derive several watershed parameters, such as the percent imperviousness and runoff curve number, as described in Chapter 11 (Modeling Applications) and Chapter 13 (Sewer Models). This runoff-coefficient estimation technique is a good example of maximizing the capabilities of remote sensing. The LULC layer can be developed from supervised (interactive) or unsupervised (automatic) classification of satellite imagery. GIS can be used to reclassify or aggregate the usually large number of imagery-based LULC classes into a small number of user-specified LULC classes. GIS can also help to refine or verify the imagery-based LULC classes. For example, population-density and unit-type (single or multiple family) attributes of census GIS data (e.g., census blocks) can be used to reclassify the typical residential land-use classes (e.g., low-, medium-, and high-density residential) into single-family residential and multifamily residential land-use classes.

LULC changes observed in remote sensing imagery are of interest to planners, ecologists, hydrologists, and atmospheric scientists. Remote sensing imagery taken at different times can be compared to detect changes in LULC using a process called *change detection*. Remote sensing change detection can also be used for near-real-time detection and mapping of fires, power outages, and floods. Remote sensing change detection techniques can also be adapted for assessing the differences between modeled and observed images for validation of distributed hydrologic and ecosystem models (Luce, 2001; Lunetta and Elvidge, 1998).

Standard multispectral data provide only a grid of numbers that literally reflect the amount of electromagnetic energy measured at a specified small square on the Earth's surface. The most common and pervasive problem in any spectral classification of these data is the presence of mixed cells called "mixels." A 10 m × 10 m location on the ground may be half lawn and shrub and half rooftop. Obviously, because the spectral reflectance value for this cell cannot show two half-values, the value becomes the average of the two surface types, statistically dissimilar from both vegetation and urban areas. Taking into consideration the fact that cells are not spectrally divided into only two distinct surface types but every possible combination, the difficulties in spectrally based classification procedures become apparent. A key challenge to using multispectral satellite imagery data is to "unmix" (usually urban) land-use areas that contain numerous mixed pixels due to a highly variable landscape. High-spectral (or hyperspectral) imagery that divides the electromagnetic spectrum into a large number of segments for a finer classification of the Earth's surface is an important evolving trend. The mixel problem can be eliminated by using hyperspectral data from new satellites, such as NASA's Earth Observing-1 (EO-1) and Orbimage's OrbView 3. One-and-a-half months after its launch aboard the EO-1 spacecraft Hyperion, NASA's first hyperspectral imager was transmitting 30-m resolution images of the Earth in 220 spectral bands from the visible to shortwave infrared. Hyperion captures 7.5 km × 180 km images with high radiometric accuracy. NASA's ASTER is capable of collecting 14 bands of data at 15- to 19-m resolutions. The next OrbView launch is expected to provide a 200-band hyperspectral capability with 8-m resolution. The hyperspectral analysis software packages such as ENVI use special techniques like spectral angle mapping or linear spectral unmixing to isolate different contributions to a single pixel. The IMAGINE software from ERDAS provides a subpixel classifier to handle the mixed cells (ERDAS, 2001).

Note that local governments are usually the best sources of LULC data because they are involved most directly in land-planning activities. Many cities maintain accurate, aerial photo-based parcel maps for tax assessment. Local LULC data can include outlines of buildings, driveways, and pavements, which are excellent sources of data for applications that require highly accurate parameters, such as estimation of percent imperviousness by LULC classes and urban runoff modeling (Slawecki et al., 2001). Before using GIS and remote sensing techniques to create LULC data, one should check with local governments and regional organizations in the study area to determine if existing LULC maps or data are available.

Soil Moisture Mapping

Recent developments have opened the doors for exploring remotely sensed data in the microwave region. The strength of the microwave signal is directly related to the amount of water present below the target surface. This feature makes microwave remote sensing particularly attractive in water resources studies. Microwave sensors can produce soil-moisture data that measure the surface dielectric properties. Daily soil-moisture maps can be composed to establish antecedent conditions for runoff modeling. NASA's Tropical Rainfall Mapping Mission (TRMM) Microwave Imager has been used to establish soil moisture over a 140 km × 280 km region in Oklahoma.

Estimating Meteorological Data

Many watersheds, especially those smaller than 1000 km^2, do not have recording rain gauges. These data gaps can be filled by the rainfall data provided by weather satellites and radars. Detailed applications of remote sensing data in rainfall measurement are provided in Chapter 10 (Monitoring Applications).

Evapotranspiration data are critical in water budget computations of a watershed. Evapotranspiration measurements require expensive instruments for an area no larger than a field. Estimation of evapotranspiration for large areas using computer models may be difficult and time consuming. Research hydrologists at the University of Nebraska-Lincoln have developed a way to measure watershed evapotranspiration using vegetative productivity (greenness) extracted from satellite imagery. Because producing green matter involves using a specific amount of water, evapotranspiration can be calculated for large areas. Evapotranspiration estimates are used to determine irrigation-water and groundwater recharge, two of the most difficult-to-estimate parameters in most water budget computations.

GEOGRAPHIC IMAGING AND IMAGE PROCESSING SOFTWARE

In recent days, geographic imaging has become an essential part of GIS applications. Without geographic imaging capability, a GIS software is like a car without a spare wheel, or a computer without a modem, or a pizza without toppings. Imagine tens or hundreds of image tiles that have unbalanced color, brightness, or contrast, which must be manually adjusted to plot nicely. Imagine a GIS project involving 6-in. resolution images for an entire county and the user waiting for the computer to redraw the screen as he or she zooms in and out. Geographic-imaging and image processing products solve these problems by helping users to visualize, manipulate, analyze, measure, and integrate geographic imagery and geospatial information. These programs can be classified in two categories:

1. Preprocessors: These programs (such as MrSID described in the following text), help to preprocess raster data for GIS applications. Some preprocessing programs can join multiple images to provide georeferenced color-balanced mosaics. They combine multiple frames of scanned film, digital photos, or satellite images into a single picture and apply mapping-coordinates information to the mosaicked image for spectral and spatial analysis.
2. GIS Extensions: These programs add image processing capability to a GIS software. For example, Image Analysis is an extension that adds image processing capability to ArcView GIS software.

Representative software are listed in Table 3.3 and described in the following subsections.

ERDAS Software Products

ERDAS, Inc. (Atlanta, Georgia), is a leading geographic imaging and image processing software company. The company's software products help organizations

Table 3.3 Geographic-Imaging and Image Processing Software

Software	Version	Vendor	Cost ($)	Web site
ERDAS IMAGINE	8.7	Leica Geosystems, Atlanta	>5000	gis.leica-geosystems.com www.erdas.com
ER Mapper	6.1	Earth Resources Mapping, West Perth, Australia	5000	www.ermapper.com
Image Analysis	1.1	ERDAS and ESRI	2500	
Geomatica EASI/PACE	— 7.0	PCI Geomatics, Ontario, Canada	N/A	www.pcigeomatics.com
Image Analyst for MicroStation	—	Z/I Imaging Corp.	N/A	www.ziimaging.com
Geographic Transformer	4.2	Blue Marble Geographics	800	www.bluemarblegeo.com
MrSID	1.4	LizardTech	1000–3500	www.lizardtech.com
ENVI	3.4	Research Systems, Inc.	4000	www.rsinc.com

visualize, manipulate, analyze, measure, and integrate geographic imagery and geospatial information into 2D and 3D environments. In July 2001, ERDAS's geographic imaging software division was acquired by Leica Geosystems (Heerbrugg, Switzerland) to form a new GIS and mapping division. In the GIS community, Leica Geosystems is known for its GPS and field data collection related equipment.

The ERDAS flagship product is called IMAGINE (previously known as ERDAS; the latest version is 8.7); it provides remote sensing capabilities and a broad range of geographic imaging tools. It contains tools to make data production faster and easier, such as on-the-fly reprojection, a batch wizard to automate routine procedures, the ability to create and edit ESRI Shapefiles, and faster and easier map production capabilities. It also provides enhanced mosaicking capabilities for creating seamless output images. Using specialized color-balancing procedures to remove "hot spots" from aerial photographs and satellite imagery, IMAGINE's simple but powerful mosaicking tool can run in an automated mode or allow users to intervene for quality assurance. For example, its cropping feature removes an image's rough edges, and the Exclude Areas tool can be used to define pixel groups likely to skew the image histogram and cause erroneous mosaicking artifacts.

The ERDAS IMAGINE product suite consists of three components that can be combined to create a scaleable solution for project-specific needs:

- IMAGINE Essentials: A mapping and visualization tool that allows different types of geographic data to be combined with imagery and organized for a mapping project.
- IMAGINE Advantage: This component builds upon the geographic imaging capabilities of IMAGINE Essentials by adding more precise mapping and image processing capabilities. It analyzes data from imagery via image mosaicking, surface interpolation, and advanced image interpretation and orthorectification tools.
- IMAGINE Professional: This is a suite of sophisticated tools for remote sensing and complex image analysis. It contains all the capabilities of IMAGINE Essentials and IMAGINE Advantage and adds radar analysis and advanced classification tools like the IMAGINE Expert Classifier. It also includes graphical spatial data modeling, an advanced feature for geographic data analysis.

ERDAS MapSheets is a mapping and geographic presentation software package. Because of it's compatibility with Microsoft Office, MapSheets allows using Object Linking and Embedding (OLE) technology to incorporate maps and images into reports, presentations, and spreadsheets. Reportedly, it is as easy to use as a word processor or a spreadsheet because it works directly with Microsoft Office software. It allows adding a map to a Microsoft Word report, using Excel to query the map attributes, using corporate data with Microsoft Access, and making presentations in PowerPoint. MapSheets allows reshaping images and drawings that have different projections. Its change-detection feature allows viewing changes from one image or drawing to another.

ERDAS Software Application Example

With a population of over 500,000 people, Colorado Springs, Colorado, is the U.S.' 18th-fastest growing city. In 1999, the Water Resources Department of Colorado

Springs Utilities (CSU) embarked on a pipeline-mapping project using high-resolution aerial photographs. CSU used 6-in. resolution DOPs of the Rockies to accurately, quickly, and inexpensively create a base map of two water pipelines that stretch approximately 130 mi. The resulting data were combined with GPS and ArcInfo GIS data to create an accurate set of detailed pipeline maps at a scale of 1 in. = 200 ft. CSU utilized ERDAS' Stereo Analyst, IMAGINE Virtual GIS, and IMAGINE OrthoBASE geographic imaging software products. The compatibility between ERDAS and ESRI products allowed CSU to leverage its existing ArcInfo data to analyze pipeline accessibility, maintenance history and schedules, and future water needs. CSU also used high-resolution aerial photographs to create a detailed map of its water treatment plant in about 2 weeks. The ability to process more high-resolution imagery in-house in less time helped CSU to cost-effectively and efficiently maintain its GIS and save an estimated $1.8 million in mapping cost (ArcNews, 2001a).

ArcView Image Analysis Extension

ArcView Image Analysis Extension was developed as a collaborative effort between ESRI and ERDAS. It provides a direct path from IMAGINE to ArcView for users with complex geographic imaging and processing needs and provides readily available image and remote sensing data. It allows georeferencing imagery to Shapefiles, coverages, global positioning system points, or reference images; image enhancement; automatic mapping of feature boundaries; change detection for continuous and thematic imagery; multispectral categorizations for LULC mapping and data extraction; vegetation greenness mapping; and mosaicking imagery from different sources and different resolutions. One of the most useful tools in Image Analysis Extension is the Image Align tool designed to coregister image data to vector layers. The intuitive point-marking scheme is designed to make image rectification simple for novices. In addition to the standard rectification process from user-specified control points and GPS-collected points, the software also displays selected satellite data types in proper map position automatically. This capability is called image calibration, and is based on positional information (ephemeris) provided in commercial data sources. Image Analysis' spectral-categorization capabilities include unsupervised image classification with ISODATA classifier and finding like areas for single-class identification. These capabilities can be used for automatic LULC classification from satellite imagery.

ERDAS Stereo Analyst is another ArcView extension. It allows users to collect and visualize spatial data in true stereo and to roam with real-time pan and zoom. Features include the ability to collect and edit 3D Shapefiles and visualization of terrain information, tree stands, and watersheds.

MrSID

Water system and sewer system GIS data generally have DOP base maps that are stored in extremely large files. For example, the City of Loveland, Colorado, had four aerial photo images of 1.3 gigabytes (GB) each (Murphy, 2000a). The recent explosion of high-resolution imagery has dramatically increased the size of

raster data. Raster images are storage-hungry but compressed images lose resolution because there is a trade-off between image size and resolution. Large image files take a long time to display on the computer. Fortunately, the data-reading capacity of image processing software has increased from a few gigabytes to the terabyte range. Image-compression techniques have played a critical role in storing large raster data sets in compressed file formats, such as Earth Resource Mapping's ECW format and LizardTech's MrSID format.

Multiresolution Seamless Image Database (MrSID) is a relatively new image file type (SID) from LizardTech. It encodes large, high-resolution images to a fraction of their original file size while maintaining the original image quality. Images become scalable and can be reduced, enlarged, zoomed, panned, or printed without compromising integrity. MrSID's selective decompression and bandwidth-optimization capabilities also increase the file transfer speed. MrSID provides the world's highest compression ratios that average at about 40 but can be as high as 100. For example, Mecklenburg County, North Carolina, had 708 sheets of 1-in. = 1000-ft B&W digital orthophotos for 538 mi^2 of the entire county. The scanned 9 in. × 9 in. films at a resolution of 1-ft to 1-pixel created approximately 23 MB georeferenced TIFF files. This procedure created 16 GB of imagery stored on 27 CD-ROMs. Delivering a compression ratio of 1:28, MrSID took 14 hours on a Pentium PC with 512 MB of RAM to compress 16 GB of imagery to a single 608 MB MrSID Portable Image Format that could be stored on a single 650 MB CD-ROM (Kuppe, 1999). Similarly, MrSID was able to compress 18 GB of Washington, D.C., DOPs onto one CD-ROM.

MrSID allows fast viewing of massive images. For instance, displaying a 50 MB TIFF image can take several minutes, whereas it requires only a few seconds to open the same image in the MrSID format. Reportedly, MrSID software can automatically mosaic hundreds of image tiles of virtually any size into a single, seamless image that is geometrically and geographically accurate with all georeferencing data intact. MrSID images can be viewed in most popular GIS software packages, such as ESRI's ArcInfo, ArcView, MapObjects, ArcIMS, ArcExplorer, and ArcPad; Autodesk, Intergraph, and GE Smallworld. MrSID ArcView GIS extension allows MrSID images to be instantly decompressed and displayed within ArcView GIS. It takes advantage of MrSID's image-compression and retrieval capability and offsets the problems of working with large images in ArcView GIS. MrSID's ArcView GIS extension gives users the ability to work with raster images of any size while providing instantaneous, seamless, multiresolution browsing of large raster images in ArcView.

PCI Geomatics

PCI Geomatics (Ontario, Canada) provides a suite of image processing, remote sensing, orthophoto, and GIS software tools. In 2001, PCI Geomatics software was used to produce "ImageMap USA," the first seamless color satellite image of the continental U.S. at 15-m spatial resolution. This mosaic is composed of approximately 450 individual Landsat 7 satellite scenes. PCI Geomatics' software products are described below.

Geomatica™ unites previously separate technologies that were dedicated to remote sensing, image processing, GIS (both vector and raster), cartography, and

desktop photogrammetry into a single integrated environment. Geomatica represents the most aggressive movement toward the integration of GIS and image processing functions in one software package (Limp, 2001). Although Geomatica is available in several configurations, all have a consistent user interface and data structure.

EASI/PACE desktop remote sensing software allows working with Landsat, SPOT, air photos, or radar data in one image processing package. Users can use data from GIS databases and extract information for resource analysis, mapping and environmental applications. It can also be used to make image maps and update GIS data.

ImageWorks is a GUI-based software that incorporates the most frequently needed image-display, enhancement, and data management tools in a user-friendly interface. The software provides integrated raster and vector display capability.

OrthoEngine Suite provides capabilities for producing map-accurate imagery. It can orthorectify many types of image data, including aerial photos, digital camera frames, and Landsat, SPOT, IRS, IKONOS, RADARSAT, ERS, and JERS imagery. It allows DEM generations from SPOT, IRS, or RADARSAT imagery. It also provides manual and automatic mosaicking tools to create seamless mosaics of multiple images.

FLY! is a terrain-visualization tool that drapes imagery and vectors over DEM data to create 3D perspective scenes in near real-time. An intuitive point-and-click user interface enables users to control flight speed, direction, elevation, and perspective parameters interactively during flight.

Radar imagery from satellites and aircraft has become a significant tool for a wide range of remote sensing applications. Data from sources such as RADARSAT, ERS, and JERS provide timely and consistent sources of information, regardless of weather conditions or illumination. With the expanding use of radar-imaging systems comes the need for new processing tools to extract useful information. Imaging radar data requires special handling and analysis techniques. PCI Geomatics has developed RADARSOFT to meet these needs.

PAMAP is a full-featured GIS software that combines an easy-to-use interface with increased interactivity and dynamic links to many industry-standard external databases for efficient storage of attribute data. It uses PCI Geomatics' Generic Database (GDB) technology, which supports over 50 different raster and vector data types facilitating easy exchange of spatial and attribute data.

GeoGateway allows users to visualize large data files, reproject them into any of the over 25 supported projection systems (or define a new projection system), subset the files into smaller windows of data, and write the resulting data out to a different, more distributable format, or even to an enterprise relational database such as Oracle.

Blue Marble Geographics

Blue Marble Geographics provides a host of geographic imaging, image processing, and map-projection software tools and utilities. The Geographic Transformer allows users to establish an "image-to-world" relationship between image and map coordinates and reproject an image into a georeferenced image map. It is a simple software to georeference, reproject, tile, and mosaic images. The Geographic Transformer AVX is an ArcView extension that integrates these functions directly into the ArcView GIS environment. Geographic Calculator is a stand-alone

utility than can convert coordinates from one system to another. It offers more than 12,000 predefined coordinate systems as well as user-defined systems, datums, and units. Geographic Calculator supports DXF, DWG, SHP, MIF, and TAB files.

FUTURE DIRECTIONS

Remote sensing has dramatically improved our ability to forecast weather, monitor pollution, and respond to natural disasters. As we continue to transition into the information age, we will uncover even more remote sensing applications. According to industry experts, QuickBird and LIDAR are only the tip of the iceberg. There are several new developments in the remote sensing field that will have a significantly positive impact on future applications in the water industry (Merry, 2000). These include the launch of the EOS series of satellites, the availability of satellite hyperspectral remote sensing data, and high-resolution commercial remote sensing data. Hyperspectral sensors examine hundreds of individual bands of the electromagnetic spectrum to reveal extremely subtle characteristics of the Earth's surface. The EOS time-series data is organized in an information system that will allow direct assimilation of remote sensing data into computer models for calibration purposes. New sensors, particularly in the microwave region, promise great potential for hydrologic applications. The U.S. Navy's scheduled launch of the Naval Earth Map Observer (NEMO) will provide data on shallow-water bathymetry, bottom type composition, currents, oil slicks, atmospheric visibility, beach characteristics, and near-shore soil and vegetation (Lillesand and Kiefer, 1999).

In the future, maps of dubious accuracy will not be digitized into a GIS. Contemporary imagery will be obtained for each unique application or solution. The new technology will enable improved analysis of reflected laser signals of LIDAR systems. Enhanced interpretation of the strength of the laser signals will enable automatic creation of low-grade DOPs. LIDAR technology one day will become so accurate that it will be possible to distinguish between different types of trees. As our ability to develop DEMs increases, the potential uses of LIDAR technology will multiply.

The human eye is accustomed to seeing objects such as tanks and ponds, whereas remote sensing data recognize nothing but pixels. This limitation creates the pixel problem described earlier in the chapter, which can be eliminated using the latest hyperspectral imagery. In addition to this solution, new remote sensing techniques are being developed to enable identification of objects. In such applications, the spatial and textural properties of the image are used along with the spectral properties of a single pixel to extract real-world objects from remotely sensed data (Limp, 2001). Geographic imaging software companies like ERDAS are developing hyperspectral analysis tools that will provide task-oriented, wizard-based capabilities to simplify and automate numerous preprocessing steps for analyzing hyperspectral imagery.

Autonomous aerial imaging combines the advantages of both aerial and satellite imaging. Autonomous imaging is already available for military missions, and it is anticipated that it will soon become available for commercial applications. On April 23, 2001, the U.S. Air Force's Global Hawk became the first autonomous powered aircraft to cross the Pacific Ocean. The result of a collaboration between the U.S. Air

Force and the Australian Defence Organization, Global Hawk can fly at altitudes of 65,000 ft and image 40,000 mi^2 in less than 24 hours. For commercial applications, the aircraft could be quickly deployed to an area of interest (or crisis) and data could be downloaded and processed in near real-time-like satellite imagery, but with a higher pixel resolution. This technology can reduce overhead cost on pilot and technician hours and make real-time high-resolution imagery affordable to water industry professionals.

USEFUL WEB SITES

IKONOS satellite	www.spaceimaging.com
Landsat 7 satellite	http://landsat7.usgs.gov
OrbView satellite	www.orbimage.com
Pennsylvania Spatial Data Access system (PASDA)	www.pasda.psu.edu
QuickBird satellite	www.digitalglobe.com
Terra satellite	http://terra.nasa.gov
TerraServer	www.terraserver.com
USGS DOQ	www-wmc.wr.usgs.gov/doq/
USGS EarthExplorer	http://earthexplorer.usgs.gov

CHAPTER SUMMARY

The increasing use of GIS is contributing to a renewed interest in remote sensing — the process of observing and mapping from a distance. GIS technology is promoting the use of remote sensing data such as aerial photographs, satellite imagery, and radar data. Remote sensing technology offers numerous applications in the water industry. Space-based satellite imagery in GIS-ready format can be used as cost-effective base maps for mapping water industry systems. It can be used for land-use classification, watershed delineation, and measuring rainfall. Remote sensing data are available in various resolutions; not all of which are suitable for water industry applications. The recent availability of high-resolution (0.5 to 5 m) satellite imagery is poised to bring remote sensing technology to a water utility near you. For additional information, the famous Lillesand and Kiefer (1999) textbook *Remote Sensing and Image Interpretation* is suggested.

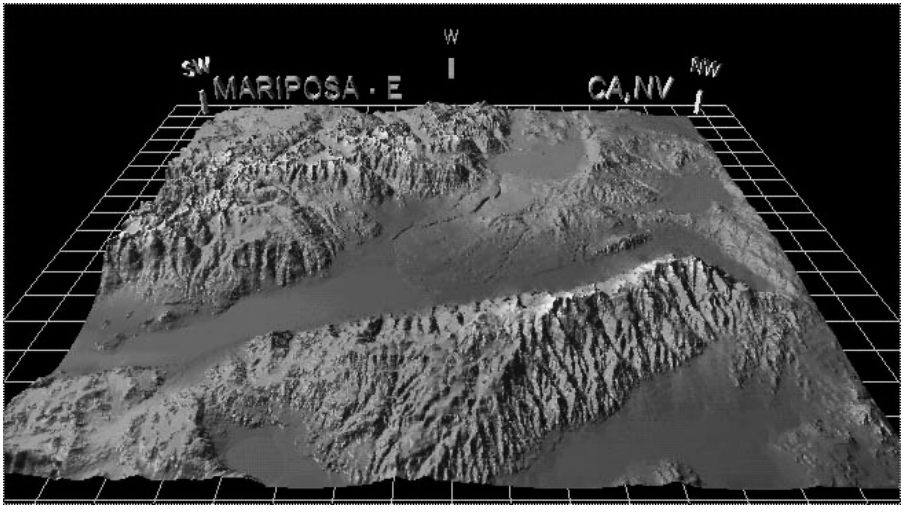
CHAPTER QUESTIONS

1. What is remote sensing and how is it related to satellite imagery?
2. How is remote sensing related to GIS and how is it used in GIS applications?
3. What is a digital orthophoto and how is it used in GIS?
4. How are remote sensing data used for land-use classification?
5. What type of software is required for using remote sensing data?

CHAPTER 4

DEM Applications

Can a laser device mounted in an airplane create a GIS-ready ground surface elevation map of your study area or measure the elevation of your manholes? Read this chapter to find out.



1:250,000 USGS DEM for Mariposa East, California (plotted using DEM3D viewer software from USGS).

LEARNING OBJECTIVE

The learning objective of this chapter is to learn how to use digital elevation models (DEM) in GIS for water industry applications.

MAJOR TOPICS

- DEM basics
- DEM data resolution and accuracy
- USGS DEM data
- DEM data from remote sensing technology
- DEM data from LIDAR and IFSAR technologies
- DEM analysis techniques and software packages
- DEM application case studies and examples

LIST OF CHAPTER ACRONYMS

3-D Three-Dimensional
DEM Digital Elevation Model
DTM Digital Terrain Model
ERDAS Earth Resource Data Analysis System
IFSAR Interferometric Synthetic Aperture Radar
LIDAR Laser Imaging Detection and Ranging/Light Imaging Detection and Ranging
NED National Elevation Detection and Ranging
TIN Triangular Irregular Network

HYDROLOGIC MODELING OF THE BUFFALO BAYOU USING GIS AND DEM DATA

In the 1970s, the Hydrologic Engineering Center (HEC) of the U.S. Army Corps of Engineers participated in developing some of the earliest GIS applications to meet the H&H modeling needs in water resources. In the 1990s, HEC became aware of the phenomenal growth and advancement in GIS. The capability of obtaining spatial data from the Internet coupled with powerful algorithms in software and hardware made GIS an attractive tool for water resources projects. The Buffalo Bayou Watershed covers most of the Houston metropolitan area in Texas. The first recorded flood in 1929 in the watershed devastated the city of Houston. Since then, other flooding events of similar vigor and intensity have occurred. During 1998 to 1999, the hydrologic modeling of this watershed was conducted using the Hydrologic Modeling System (HMS) with inputs derived from GIS. The watersheds and streams were delineated from the USGS DEM data at 30-m cell resolution, stream data from USGS digital line graph (DLG), and EPA river reach file (RF1). When used separately, software packages such as ArcInfo, ArcView, and Data Storage System (DSS)

were found to be time consuming, requiring the combined efforts of many people. HEC integrated these existing software tools with new programs developed in this project into a comprehensive GIS software package called HEC-GeoHMS. The low-relief terrain of the study area required human interpretation of drainage paths, urban drainage facilities, and man-made hydraulic structures (e.g., culverts and storm drains), which dictated flow patterns that could not be derived from DEM terrain representation. To resolve this issue, the project team took advantage of the flexibility in HMS to correct drainage patterns according to human interpretations and local knowledge (Doan, 1999).

DEM BASICS

Topography influences many processes associated with the geography of the Earth, such as temperature and precipitation. GIS application professionals must be able to represent the Earth's surface accurately because any inaccuracies can lead to poor decisions that may adversely impact the Earth's environment. A DEM is a numerical representation of terrain elevation. It stores terrain data in a grid format for coordinates and corresponding elevation values. DEM data files contain information for the digital representation of elevation values in a raster form. Cell-based raster data sets, or grids, are very suitable for representing geographic phenomena that vary continuously over space such as elevation, slope, precipitation, etc. Grids are also ideal for spatial modeling and analysis of data trends that can be represented by continuous surfaces, such as rainfall and stormwater runoff.

DEM data are generally stored using one of the following three data structures:

- Grid structures
- Triangular irregular network (TIN) structures
- Contour-based structures

Regardless of the underlying data structure, most DEMs can be defined in terms of (x,y,z) data values, where x and y represent the location coordinates and z represents the elevation values. Grid DEMs consist of a sampled array of elevations for a number of ground positions at regularly spaced intervals. This data structure creates a square grid matrix with the elevation of each grid square, called a pixel, stored in a matrix format. Figure 4.1 shows a 3D plot of grid-type DEM data. As shown in Figure 4.2, TINs represent a surface as a set of nonoverlapping contiguous triangular facets, of irregular size and shape. Digital terrain models (DTMs) and digital surface models (DSMs) are different varieties of DEM. The focus of this chapter is on grid-type DEMs.

Usually, some interpolation is required to determine the elevation value from a DEM for a given point. The DEM-based point elevations are most accurate in relatively flat areas with smooth slopes. DEMs produce low-accuracy point elevation values in areas with large and abrupt changes in elevation, such as cliffs and road cuts (Walski et al., 2001).

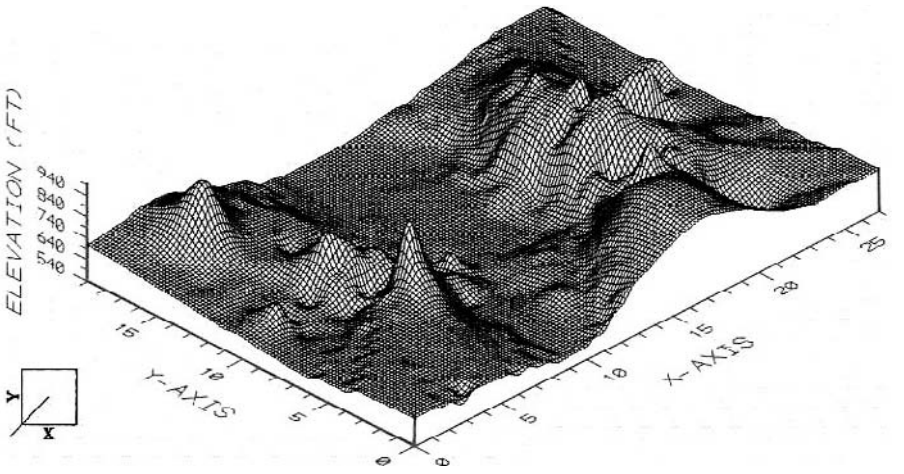


Figure 4.1 Grid-type DEM.

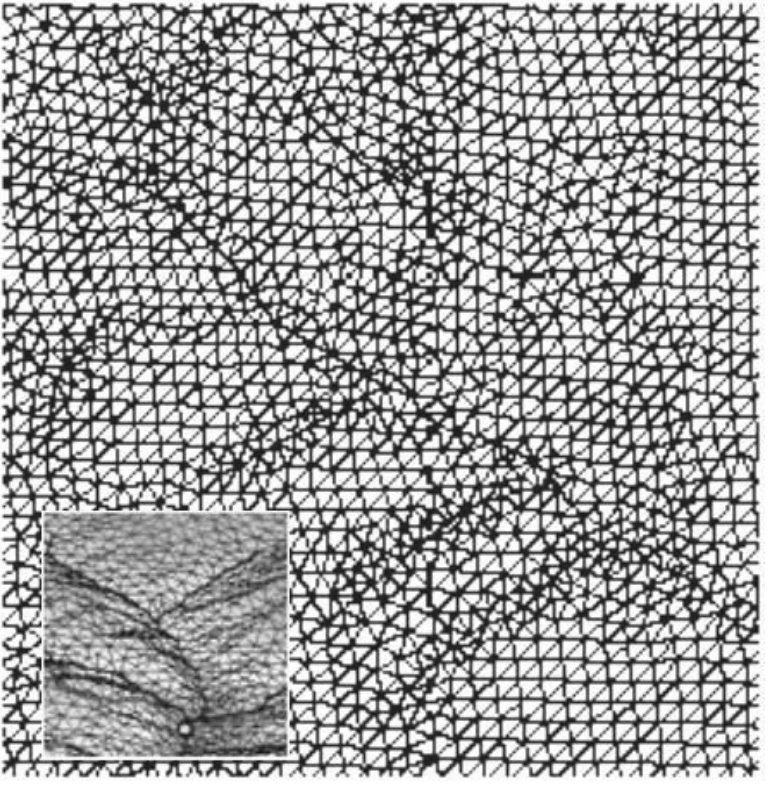


Figure 4.2 TIN-type DEM.

DEM APPLICATIONS

Major DEM applications include (USGS, 2000):

- Delineating watershed boundaries and streams
- Developing parameters for hydrologic models
- Modeling terrain gravity data for use in locating energy resources
- Determining the volume of proposed reservoirs
- Calculating the amount of material removed during strip mining
- Determining landslide probability

Jenson and Dominique (1988) demonstrated that drainage characteristics could be defined from a DEM. DEMs can be used for automatic delineation of watershed and sewershed boundaries. DEM data can be processed to calculate various watershed and sewershed characteristics that are used for H&H modeling of watersheds and sewersheds. DEMs can create shaded relief maps that can be used as base maps in a GIS for overlaying vector layers such as water and sewer lines. DEM files may be used in the generation of graphics such as isometric projections displaying slope, direction of slope (aspect), and terrain profiles between designated points. This aspect identifies the steepest downslope direction from each cell to its neighbors.

Raster GIS software packages can convert the DEMs into image maps for visual display as layers in a GIS. DEMs can be used as source data for digital orthophotos. They can be used to create digital orthophotos by orthorectification of aerial photos, as described in Chapter 3 (Remote Sensing Applications). DEMs can also serve as tools for many activities including volumetric analysis and site location of towers. DEM data may also be combined with other data types such as stream locations and weather data to assist in forest fire control, or they may be combined with remote sensing data to aid in the classification of vegetation.

Three-Dimensional (3D) Visualization

Over the past decade, 3D computer modeling has evolved in most of the engineering disciplines including, but not limited to, layout, design, and construction of industrial and commercial facilities; landscaping; highway, bridge, and embankment design; geotechnical engineering; earthquake analysis; site planning; hazardous-waste management; and digital terrain modeling. The 3D visualization can be used for landscape visualizing or fly-through animation movies of the project area. 3D animations are highly effective tools for public- and town-meeting presentations. GIS can be used to create accurate topographic elevation models and generate precise 3D data. A DEM is a powerful tool and is usually as close as most GISs get to 3D modeling.

3D graphics are commonly used as a visual communication tool to display a 3D view of an object on two-dimensional (2D) media (e.g., a paper map). Until the early 1980s, a large mainframe computer was needed to view, analyze, and print objects in 3D graphics format. Hardware and software are now available for 3D modeling of terrain and utility networks on personal computers. Although DEMs are raster images, they can be imported into 3D visualizations packages. Affordable and user-friendly software tools are bringing more users into the world of GIS.

These software tools and 3D data can be used to create accurate virtual reality representations of landscape and infrastructure with the help of stereo imagery and automatic extraction of 3D information. For example, Skyline Software System's (www.skylinesoft.com) TerraExplorer provides realistic, interactive, photo-based 3D maps of many locations and cities of the world on the Internet.

Satellite imagery is also driving new 3D GIS applications. GIS can be used to precisely identify a geographic location in 3D space and link that location and its attributes through the integration of photogrammetry, remote sensing, GIS, and 3D visualization. 3D geographic imaging is being used to create orthorectified imagery, DEMs, stereo models, and 3D features.

DEM RESOLUTION AND ACCURACY

The accuracy of a DEM is dependent upon its source and the spatial resolution (grid spacing). DEMs are classified by the method with which they were prepared and the corresponding accuracy standard. Accuracy is measured as the root mean square error (RMSE) of linearly interpolated elevations from the DEM, compared with known elevations. According to RMSE classification, there are three levels of DEM accuracy (Walski et al., 2001):

- Level 1: Based on high-altitude photography, these DEMs have the lowest accuracy. The vertical RMSE is 7 m and the maximum permitted RMSE is 15 m.
- Level 2: These are based on hypsographic and hydrographic digitization, followed by editing to remove obvious errors. These DEMs have medium accuracy. The maximum permitted RMSE is one half of the contour interval.
- Level 3: These are based on USGS digital line graph (DLGs) data (Shamsi, 2002). The maximum permitted RMSE is one third of the contour interval.

The vertical accuracy of 7.5-min DEMs is greater than or equal to 15 m. Thus, the 7.5-min DEMs are suitable for projects at 1:24,000 scale or smaller (Zimmer, 2001a). A minimum of 28 test points per DEM are required (20 interior points and 8 edge points). The accuracy of the 7.5-min DEM data, together with the data spacing, adequately support computer applications that analyze hypsographic features to a level of detail similar to manual interpretations of information as printed at map scales not larger than 1:24,000. Early DEMs derived from USGS quadrangles suffered from mismatches at boundaries (Lanfear, 2000).

DEM selection for a particular application is generally driven by data availability, judgment, experience, and test applications (ASCE, 1999). For example, because no firm guidelines are available for selection of DEM characteristics for hydrologic modeling, a hydrologic model might need 30-m resolution DEM data but might have to be run with 100-m data if that is the best available data for the study area. In the U.S., regional-scale models have been developed at scales of 1:250,000 to 1:2,000,000 (Laurent et al., 1998). Seybert (1996) concluded that modeled watershed runoff peak flow values are more sensitive to changes in spatial resolution than modeled runoff volumes. An overall subbasin area to grid–cell area ratio of 10^2 was found to produce reasonable model results.

Table 4.1 DEM Applications

DEM Resolution	Approximate Cell Size	Watershed Area (km²)	Typical Application
1 sec	30 m	5	Urban watersheds
3 sec	100 m	40	Rural watersheds
15 sec	500 m	1,000	River basins, States
30 sec	1 km	4,000	Nations
3 min	5 km	150,000	Continents
5 min	10 km	400,000	World

The grid size and time resolution used for developing distributed hydrologic models for large watersheds is a compromise between the required accuracy, available data accuracy, and computer run-time. Finer grid size requires more computing time, more extensive data, and more detailed boundary conditions. Chang et al. (2000) conducted numerical experiments to determine an adequate grid size for modeling large watersheds in Taiwan where 40 m × 40 m resolution DEM data are available. They investigated the effect of grid size on the relative error of peak discharge and computing time. Simulated outlet hydrographs showed higher peak discharge as the computational grid size was increased. In a study, for a watershed of 526 km² located in Taiwan, a grid resolution of 200 m × 200 m was determined to be adequate.

Table 4.1 shows suggested DEM resolutions for various applications (Maidment, 1998). Large (30-m) DEMs are recommended for water distribution modeling (Walski et al., 2001).

The size of a DEM file depends on the DEM resolution, i.e., the finer the DEM resolution, the smaller the grid, and the larger the DEM file. For example, if the grid size is reduced by one third, the file size will increase nine times. Plotting and analysis of high-resolution DEM files are slower because of their large file sizes.

USGS DEMS

In the U.S., the USGS provides DEM data for the entire country as part of the National Mapping Program. The National Mapping Division of USGS has scanned all its paper maps into digital files, and all 1:24,000-scale quadrangle maps now have DEMs (Limp, 2001).

USGS DEMs are the (x,y,z) triplets of terrain elevations at the grid nodes of the Universal Transverse Mercator (UTM) coordinate system referenced to the North American Datum of 1927 (NAD27) or 1983 (NAD83) (Shamsi, 1991). USGS DEMs provide distance in meters, and elevation values are given in meters or feet relative to the National Geodetic Vertical Datum (NGVD) of 1929. The USGS DEMs are available in 7.5-min, 15-min, 2-arc-sec (also known as 30-min), and 1° units. The 7.5- and 15-min DEMs are included in the large-scale category, whereas 2-arc-sec DEMs fall within the intermediate-scale category and 1° DEMs fall within the small-scale category. Table 4.2 summarizes the USGS DEM data types.

This chapter is mostly based on applications of 7.5-min USGS DEMs. The DEM data for 7.5-min units correspond to the USGS 1:24,000-scale topographic quadrangle map series for all of the U.S. and its territories. Thus, each 7.5-min

Table 4.2 USGS DEM Data Formats

DEM Type	Scale	Block Size	Grid Spacing
Large	1:24,000	7.5 ft × 7.5 ft	30 m
Intermediate	Between large and small	30 ft × 30 ft	2 sec
Small	1:250,000	1° × 1°	3 sec

by 7.5-min block provides the same coverage as the standard USGS 7.5-min map series. Each 7.5-min DEM is based on 30-m by 30-m data spacing; therefore, the raster grid for the 7.5-min USGS quads are 30 m by 30 m. That is, each 900 m^2 of land surface is represented by a single elevation value. USGS is now moving toward acquisition of 10-m accuracy (Murphy, 2000).

USGS DEM Formats

USGS DEMs are available in two formats:

1. DEM file format: This older file format stores DEM data as ASCII text, as shown in Figure 4.3. These files have a file extension of dem (e.g., lewisburg_PA.dem). These files have three types of records (Walski et al., 2001):
   - Type A: This record contains information about the DEM, including name, boundaries, and units of measurements.

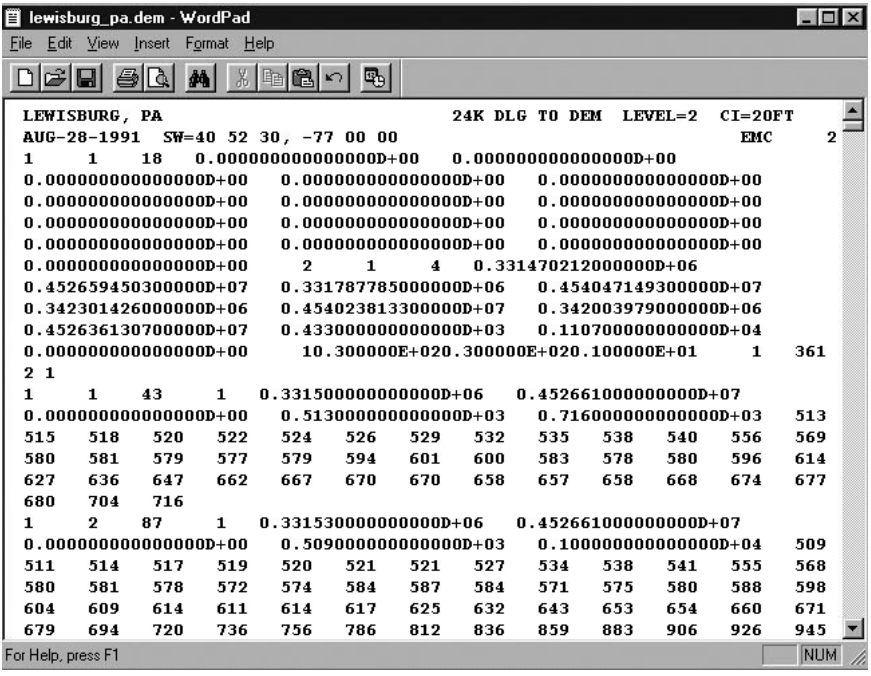


Figure 4.3 USGS DEM file.

- Type B: These records contain elevation data arranged in "profiles" from south to north, with the profiles organized from west to east. There is one Type-B record for each south–north profile.
- Type C: This record contains statistical information on the accuracy of DEM.

2. Spatial Data Transfer Standard (SDTS): This is the latest DEM file format that has compressed data for faster downloads. SDTS is a robust way of transferring georeferenced spatial data between dissimilar computer systems and has the potential for transfer with no information loss. It is a transfer standard that embraces the philosophy of self-contained transfers, i.e., spatial data, attribute, georeferencing, data quality report, data dictionary, and other supporting metadata; all are included in the transfer. SDTS DEM data are available as tar.gz compressed files. Each compressed file contains 18 ddf files and two readme text files. For further analysis, the compressed SDTS files should be unzipped (uncompressed). Standard zip programs, such as PKZIP, can be used for this purpose.

Some DEM analysis software may not read the new SDTS data. For such programs, the user should translate the SDTS data to a DEM file format. SDTS translator utilities, like SDTS2DEM or MicroDEM, are available from the GeoCommunity's SDTS Web site to convert the SDTS data to other file formats.

National Elevation Dataset (NED)

Early DEMs were derived from USGS quadrangles, and mismatches at boundaries continued to plague the use of derived drainage networks for larger areas (Lanfear, 2000). The NED produced by USGS in 1999 is the new generation of seamless DEM that largely eliminates problems of quadrangle boundaries and other artifacts. Users can now select DEM data for their area of interest.

The NED has been developed by merging the highest resolution, best-quality elevation data available across the U.S. into a seamless raster format. NED is designed to provide the U.S. with elevation data in a seamless form, with a consistent datum, elevation unit, and projection. Data corrections were made in the NED assembly process to minimize artifacts, perform edge matching, and fill sliver areas of missing data. NED is the result of the maturation of the USGS effort to provide 1:24,000-scale DEM data for the conterminous U.S. and 1:63,360-scale DEM data for Alaska. NED has a resolution of 1 arc-sec (approximately 30 m) for the conterminous U.S., Hawaii, and Puerto Rico and a resolution of 2 arc-sec for Alaska. Using a hill-shade technique, USGS has also derived a shaded relief coverage that can be used as a base map for vector themes. Other themes, such as land use or land cover, can be draped on the NED-shaded relief maps to enhance the topographic display of themes. The NED store offers seamless data for sale, by user-defined area, in a variety of formats.

DEM DATA AVAILABILITY

USGS DEMs can be downloaded for free from the USGS geographic data download Web site. DEM data on CD-ROM can also be purchased from the USGS EarthExplorer Web site for an entire county or state for a small fee to cover the shipping and handling cost. DEM data for other parts of the world are also available.

The 30 arc-sec DEMs (approximately 1 km^2 square cells) for the entire world have been developed by the USGS Earth Resources Observation Systems (EROS) Data Center and can be downloaded from the USGS Web site. More information can be found on the Web site of the USGS node of the National Geospatial Data Clearinghouse. State or regional mapping and spatial data clearinghouse Web sites are the most valuable source of free local spatial data. For example, the Pennsylvania Spatial Data Access system (PASDA), Pennsylvania's official geospatial information clearinghouse and its node on the National Spatial Data Infrastructure (NSDI), provides free downloads of DEM and other spatial data.

DEM DATA CREATION FROM REMOTE SENSING

In February 2000, NASA flew one of its most ambitious missions, using the space shuttle *Endeavor* to map the entire Earth from 60° north to 55° south of the equator. Mapping at a speed of 1747 km^2 every second, the equivalent of mapping the state of Florida in 97.5 sec, the Shuttle Radar Topography Mission (SRTM) provided 3D data of more than 80% of Earth's surface in about 10 days. The SRTM data will provide a 30-m DEM coverage for the entire world (Chien, 2000).

Topographic elevation information can be automatically extracted from remote sensing imagery to create highly accurate DEMs. There are two ways in which DEM data can be created using remote sensing methods: image processing and data collection.

Image Processing Method

The first method uses artificial intelligence techniques to automatically extract elevation information from the existing imagery. Digital image-matching methods commonly used for machine vision automatically identify and match image point locations of a ground point appearing on overlapping areas of a stereo pair (i.e., left- and right-overlapping images). Once the correct image positions are identified and matched, the ground point elevation is computed automatically. For example, the French satellite SPOT's stereographic capability can generate topographic data. USGS Earth Observing System's (EOS) Terra satellite can provide DEMs from stereo images.

Off-the-shelf image processing software products are available for automatic extraction of DEM data from remote sensing imagery. For instance, Leica Geosystems' IMAGINE OrthoBASE Pro software can be used to automatically extract DEMs from aerial photography, satellite imagery (IKONOS, SPOT, IRS-1C), and digital video and 35-mm camera imagery. It can also subset and mosaic 500 or more individual DEMs. The extracted DEM data can be saved as raster DEMs, TINs, ESRI 3D Shapefiles, or ASCII output (ERDAS, 2001b).

Data Collection Method

In this method, actual elevation data are collected directly using lasers. This method uses laser-based LIDAR and radar-based IFSAR systems described in the following text.

LIDAR

Unlike photogrammetric techniques, which can be time consuming and expensive for large areas, this method is a cost-effective alternative to conventional technologies. It can create DEMs with accuracy levels ranging from 20 to 100 cm, which are suitable for many engineering applications. This remote-sensing technology does not even involve an image. Laser imaging detection and ranging (LIDAR) is a new system for measuring ground surface elevation from an airplane. LIDAR can collect 3D digital data on the fly. LIDAR sensors provide some of the most accurate elevation data in the shortest time ever by bouncing laser beams off the ground. LIDAR technology, developed in the mid-1990s, combines global positioning system (GPS), precision aircraft guidance, laser range finding, and high-speed computer processing to collect ground elevation data. Mounted on an aircraft, a high-accuracy scanner sweeps the laser pulses across the flight path and collects reflected light. A laser range-finder measures the time between sending and receiving each laser pulse to determine the ground elevation below. The LIDAR system can survey up to 10,000 acres per day and provide horizontal and vertical accuracies up to 12 and 6 in., respectively. Chatham County, home of Savannah, Georgia, used the LIDAR approach to collect 1-ft interval contour data for the entire 250,000 acre county in less than a year. The cost of conventional topographic survey for this data would be over $20 million. The County saved $7 million in construction cost by using data from Airborne Laser Terrain Mapping (ALTM) technology, a LIDAR system manufactured by Optech, Canada. The new ALTM data were used to develop an accurate hydraulic model of the Hardin basin (Stones, 1999).

Chatham County, Georgia, saved $7 million in construction cost by using LIDAR data.

Boise-based Idaho Power Company spent $273,000 on LIDAR data for a 290 km stretch of the rugged Hell's Canyon, through which the Snake River runs. The cost of LIDAR data was found to be less than aerial data and expensive ground-surveying. The company used LIDAR data to define the channel geometry, combined it with bathymetry data, and created digital terrain files containing ten cross sections of the canyon per mile. The cross-section data were input to a hydraulic model that determined the effect of power plants' releases on vegetation and wildlife habitats (Miotto, 2000).

IFSAR

Interferometric Synthetic Aperture Radar (IFSAR) is an aircraft-mounted radar system for quick and accurate mapping of large areas in most weather conditions without ground control. Because it is an airborne radar, IFSAR collects elevation data on the first try in any weather (regardless of fog, clouds, or rain), day or night, significantly below the cost of satellite-derived DEM. The IFSAR process measures elevation data at a much denser grid than photogrammetric techniques, using overlapping stereo images. A denser DEM provides a more detailed terrain surface in

an image. IFSAR is efficient because it derives the DEM data by digital processing of a single radar image. This allows elevation product delivery within days of data collection. A DEM with a minimum vertical accuracy of 2 m is necessary to achieve the precision level orthorectification for IKONOS imagery. DEMs generated from the IFSAR data have been found to have the adequate vertical accuracy to orthorectify IKONOS imagery to the precision level (Corbley, 2001).

Intermap Technologies' (Englewood, Colorado, www.intermaptechnologies.com) Lear jet-mounted STAR-3i system, an airborne mapping system, has been reported to provide simultaneous high-accuracy DEMs and high-resolution orthorectified imagery without ground control. STAR-3i IFSAR system typically acquires elevation points at 5-m intervals, whereas photogrammetric sources use a spacing of 30 to 50 m. STAR-3i can provide DEMs with a vertical accuracy of 30 cm to 3 m and an orthoimage resolution of 2.5 m.

DEM ANALYSIS

Cell Threshold for Defining Streams

Before starting DEM analysis, users must define the minimum number of upstream cells contributing flow into a cell to classify that cell as the origin of a stream. This number, referred to as the cell "threshold," defines the minimum upstream drainage area necessary to start and maintain a stream. For example, if a stream definition value of ten cells is specified, then for a single grid location of the DEM to be in a stream, it must drain at least ten cells. It is assumed that there is flow in a stream if its upstream area exceeds the critical threshold value. In this case, the cell is considered to be a part of the stream. The threshold value can be estimated from existing topographic maps or from the hydrographic layer of the real stream network. Selection of an appropriate cell threshold size requires some user judgment. Users may start the analysis with an assumed or estimated value and adjust the initial value by comparing the delineation results with existing topographic maps or hydrographic layers. The cell threshold value directly affects the number of subbasins (subwatersheds or subareas). A smaller threshold results in smaller subbasin size, larger number of subbasins, and slower computation speed for the DEM analysis.

The D-8 Model

The 8-direction pour point model, also known as the D-8 model, is a commonly used algorithm for delineating a stream network from DEMs. As shown in Figure 4.4, it identifies the steepest downslope flow path between each cell and its eight neighboring cells. This path is the only flow path leaving the cell. Watershed area is accumulated downslope along the flow paths connecting adjacent cells. The drainage network is identified from the user-specified threshold area at the bottom of which a source channel originates and classifies all cells with a greater watershed area as part of the drainage network. Figure 4.4 shows stream delineation steps using the D-8 model with a cell threshold value of ten cells. Grid A shows the cell elevation

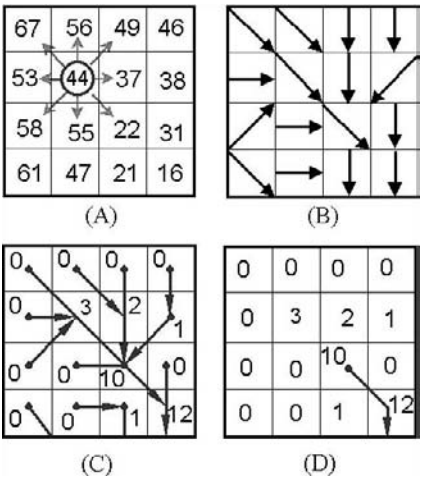


Figure 4.4 Figure 11-4. D-8 Model for DEM-based stream delineation (A) DEM elevation grid, (B) flow direction grid, (C) flow accumulation grid, and (D) delineated streams for cell threshold of ten.

values. Grid B shows flow direction arrows based on calculated cell slopes. Grid C shows the number of accumulated upstream cells draining to each cell. Grid D shows the delineated stream segment based on the cells with flow accumulation values greater than or equal to ten.

DEM Sinks

The D-8 and many other models do not work well in the presence of depressions, sinks, and flat areas. Some sinks are caused by the actual conditions, such as the Great Salt Lake in Utah where no watershed precipitation travels through a river network toward the ocean. The sinks are most often caused by data noise and errors in elevation data. The computation problems arise because cells in depressions, sinks, and flat areas do not have any neighboring cells at a lower elevation. Under these conditions, the flow might accumulate in a cell and the resulting flow network may not necessarily extend to the edge of the grid. Unwanted sinks must be removed prior to starting the stream or watershed delineation process by raising the elevation of the cells within the sink to the elevation of its lowest outlet. Most raster GIS software programs provide a FILL function for this purpose. For example, ArcInfo's GRID extension provides a FILL function that raises the elevation of the sink cells until the water is able to flow out of it.

The FILL approach assumes that all sinks are caused by underestimated elevation values. However, the sinks can also be created by overestimated elevation values, in which case breaching of the obstruction is more appropriate than filling the sink created by the obstruction. Obstruction breaching is particularly effective in flat or low-relief areas (ASCE, 1999).

Stream Burning

DEM-based stream or watershed delineations may not be accurate in flat areas or if the DEM resolution failed to capture important topographic information. This problem can be solved by "burning in" the streams using known stream locations from the existing stream layers. This process modifies the DEM grid so that the flow of water is forced into the known stream locations. The cell elevations are artificially lowered along the known stream locations or the entire DEM is raised except along known stream paths. The phrase *burning in* indicates that the streams have been forced, or "burned" into the DEM topography (Maidment, 2000). This method must be used with caution because it may produce flow paths that are not consistent with the digital topography (ASCE, 1999).

DEM Aggregation

Distributed hydrologic models based on high-resolution DEMs may require extensive computational and memory resources that may not be available. In this case, high-resolution DEMs can be aggregated into low-resolution DEMs. For example, it was found that the 30-m USGS DEM would create 80,000 cells for the 72.6 km^2 Goodwater Creek watershed located in central Missouri. Distributed modeling of 80,000 cells was considered time consuming and impractical (Wang et al., 2000). The 30 m $\times$ 30 m cells were, therefore, aggregated into 150 m x 150 m (2.25 ha) cells. In other words, 25 smaller cells were aggregated into one large cell, which reduced the number of cells from 80,000 to approximately 3,000. Best of all, the aggregated DEM produced the same drainage network as the original DEM. The aggregation method computes the flow directions of the coarse-resolution cells based on the flow paths defined by the fine-resolution cells. It uses three steps: (1) determine the flow direction of the fine-resolution DEM, (2) determine outlets of coarse-resolution DEM, and (3) approximate the flow direction of coarse-resolution DEM, based on the flow direction of the fine-resolution DEM.

Slope Calculations

Subbasin slope is an input parameter in many hydrologic models. Most raster GIS packages provide a SLOPE function for estimating slope from a DEM. For example, ERDAS IMAGINE software uses its SLOPE function to compute percent slope by fitting a plane to a pixel elevation and its eight neighboring pixel elevations. The difference in elevation between the low and the high points is divided by the horizontal distance and multiplied by 100 to compute percent slope for the pixel. Pixel slope values are averaged to compute the mean percent slope of each subbasin.

SOFTWARE TOOLS

The DEM analysis functions described in the preceding subsections require appropriate software. Representative DEM analysis software tools and utilities are listed in Table 4.3.

Table 4.3 Sample DEM Analysis Software Tools

Software	Vendor and Web site	Notes
Spatial Analyst and Hydro extension	ESRI, Redlands, California www.esri.com	ArcGIS 8.x and ArcView 3.x extension
ARC GRID extension		ArcInfo 7.x extension
Analyst		ArcGIS 8.x and ArcView 3.x extension
IDRISI	Clark University Worcester, Massachusetts www.clarklabs.org	
ERDAS IMAGINE	Leica Geosystems, Atlanta, Georgia gis.leica-geosystems.com www.erdas.com	Formerly, Earth Resource Data Analysis System (ERDAS) software
TOPAZ	U.S. Department of Agriculture, Agricultural Research Service, El Reno, Oklahoma grl.ars.usda.gov/topaz/TOPAZ1.HTM	
MicroDEM	U.S. Naval Academy www.usna.edu/Users/oceano/pguth/website/microdem.htm	Software developed by Peter Guth of the Oceanography Department
DEM3D viewer	USGS, Western Mapping Center, Menlo Park, California craterlake.wr.usgs.gov/dem3d.html	Free download, allows viewing of DEM files through a 3D perspective

Some programs such as Spatial Analyst provide both the DEM analysis and hydrologic modeling capabilities. ASCE (1999) has compiled a review of hydrologic modeling systems that use DEMs. Major DEM software programs are discussed in the following text.

Spatial Analyst and Hydro Extension

Spatial Analyst is an optional extension (separately purchased add-on program) for ESRI's ArcView 3.x and ArcGIS 8.x software packages. The Spatial Analyst Extension adds raster GIS capability to the ArcView and ArcGIS vector GIS software. Spatial Analyst allows for use of raster and vector data in an integrated environment and enables desktop GIS users to create, query, and analyze cell-based raster maps; derive new information from existing data; query information across multiple data layers; and integrate cell-based raster data with the traditional vector data sources. It can be used for slope and aspect mapping and for several other hydrologic analyses, such as delineating watershed boundaries, modeling stream flow, and investigating accumulation. Spatial Analyst for ArcView 3.x has most, but not all, of the functionality of the ARC GRID extension for ArcInfo 7.x software package described below.

Spatial Analyst for ArcView 3.x is supplied with a Hydro (or hydrology) extension that further extends the Spatial Analyst user interface for creating input data for hydrologic models. This extension provides functionality to create watersheds and stream networks from a DEM, calculate physical and geometric properties of the watersheds, and aggregate these properties into a single-attribute table that can be attached to a grid or Shapefile. Hydro extension requires that Spatial Analyst be already installed. Hydro automatically loads the Spatial Analyst if it is not loaded.

Depending upon the user needs, there are two approaches to using the Hydro extension:

1. Hydro pull-down menu options: If users only want to create watershed subbasins or the stream network, they should work directly with the Hydro pull-down menu options (Figure 4.5). Table 4.4 provides a brief description of each of these menu options. "Fill Sinks" works off an active elevation grid theme. "Flow Direction" works off an active elevation grid theme that has been filled. "Flow Accumulation" works off an active flow direction grid theme. "Flow Length" works off an active flow direction grid theme. "Watershed" works off an active flow accumulation grid theme and finds all basins in the data set based on a minimum number of cells in each basin. The following steps should be performed to create watersheds using the Hydro pull-down menu options, with the output grid from each step serving as the input grid for the next step:
   - Import the raw USGS DEM.
   - Fill the sinks using the "Fill Sinks" menu option (input = raw USGS DEM). This is a very important intermediate step. Though some sinks are valid, most are data errors and should be filled.
   - Compute flow directions using the "Flow Direction" menu option (input = filled DEM grid).
   - Compute flow accumulation using the "Flow Accumulation" menu option (input = flow directions grid).

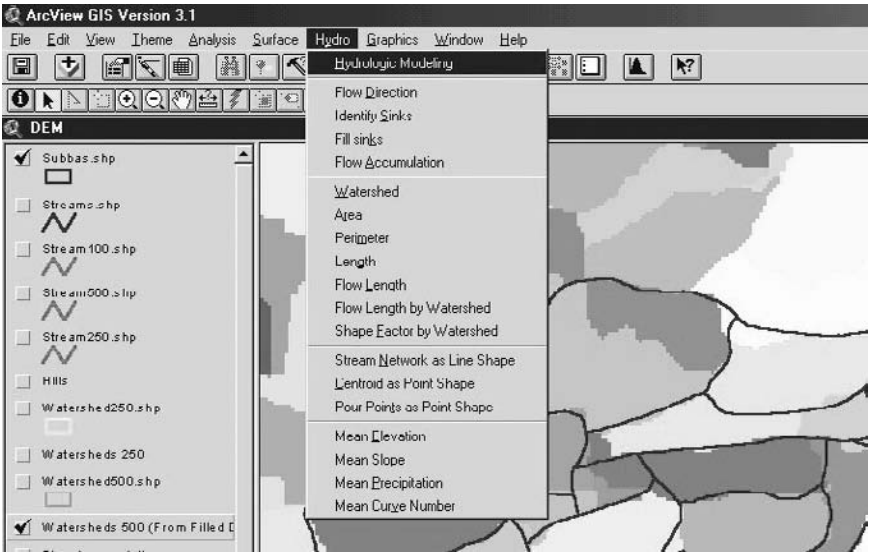


Figure 4.5 Hydro extension pull-down menu.

- Delineate streams using the "Stream Network" menu option (input = flow accumulation grid).
- Delineate watersheds using the "Watershed" menu option.

2. Hydrologic Modeling Dialogue: If users want to create subbasins and calculate many additional attributes for them, they should use the Hydrologic Modeling Dialogue (Figure 4.6), which is the first choice under the Hydro pull-down menu. The Hydrologic Modeling Dialogue is designed to be a quick one-step method for calculating and then aggregating a set of watershed attributes to a single file. This file can then be used in a hydrologic model, such as the Watershed Modeling System (WMS) (discussed in Chapter 11 [Modeling Applications]), or it can be reformatted for input into HEC's HMS model, or others. The following steps should be performed to create watersheds using the Hydrologic Modeling Dialogue:
   - Choose "Delineate" from DEM and select an elevation surface.
   - Fill the sinks when prompted.
   - Specify the cell threshold value when prompted. This will create watersheds based on the number of cells or up-slope area defined by the user as the smallest watershed wanted.

Additional DEM analysis resources (tutorials, exercises, sample data, software utilities, reports, papers, etc.) are provided at the following Web sites:

- ESRI Web site at www.esri.com/arcuser/ (do a search for "Terrain Modeling")
- University of Texas at Austin (Center for Research in Water Resources) Web site at www.crwr.utexas.edu/archive.shtml

The last four Hydro options (Table 4.4) work with existing data layers. They do not create elevation, slope, precipitation, and runoff curve number layers. They

Table 4.4 Hydro Extension Menu Options

Hydro Menu Option	Function
Hydrologic modeling	Creates watersheds and calculates their attributes
Flow direction	Computes the direction of flow for each cell in a DEM
Identify sinks	Creates a grid showing the location of sinks or areas of internal drainage in a DEM
Fill sinks	Fills the sinks in a DEM, creating a new DEM
Flow accumulation	Calculates the accumulated flow or number of up-slope cells, based on a flow direction grid
Watershed	Creates watersheds based upon a user-specified flow accumulation threshold
Area	Calculates the area of each watershed in a watershed grid
Perimeter	Calculates the perimeter of each watershed in a watershed grid
Length	Calculates the straight-line distance from the pour point to the furthest perimeter point for each watershed
Flow length	Calculates the length of flow path for each cell to the pour point for each watershed
Flow length by watershed	Calculates the maximum distance along the flow path within each watershed
Shape factor by watershed	Calculates a shape factor (watershed length squared and then divided by watershed area) for each watershed
Stream network as line shape	Creates a vector stream network from a flow accumulation grid, based on a user-specified threshold
Centroid as point shape	Creates a point shape file of watershed centroids
Pour points as point shape	Creates a point shape file of watershed pour points
Mean elevation	Calculates the mean elevation within each watershed
Mean slope	Calculates the mean slope within each watershed
Mean precipitation	Calculates the mean precipitation in each watershed
Mean curve number	Calculates the mean curve number for each watershed

simply compute mean areal values of these four parameters for the subbasins, using the existing GIS layers of these parameters. Thus, the GIS layers of elevation, slope, precipitation, and runoff curve number must be available to use the mean functions of the Hydro extension.

Figure 4.7 shows Hydro's raindrop or pour point feature. Using this capability, the user can trace the flow path from a specified point to the watershed outlet. Hydro also calculates a flow length as the maximum distance along the flow path within each watershed. The flow path can be divided by the measured or estimated velocity to estimate the time of concentration or travel time that are used to estimate runoff hydrographs. Travel time can also be used to estimate the time taken by a hazardous waste spill to reach a sensitive area or water body of the watershed. Laurent et al. (1998) used this approach to estimate travel time between any point of a watershed and a water resource (river or well). This information was further used to create a map of water resources vulnerability to dissolved pollution in an area in Massif Central, France. Subbasin area can be divided by flow length to estimate the overland flow width for input to a rainfall-runoff model such as EPA's Storm Water Management Model (SWMM).

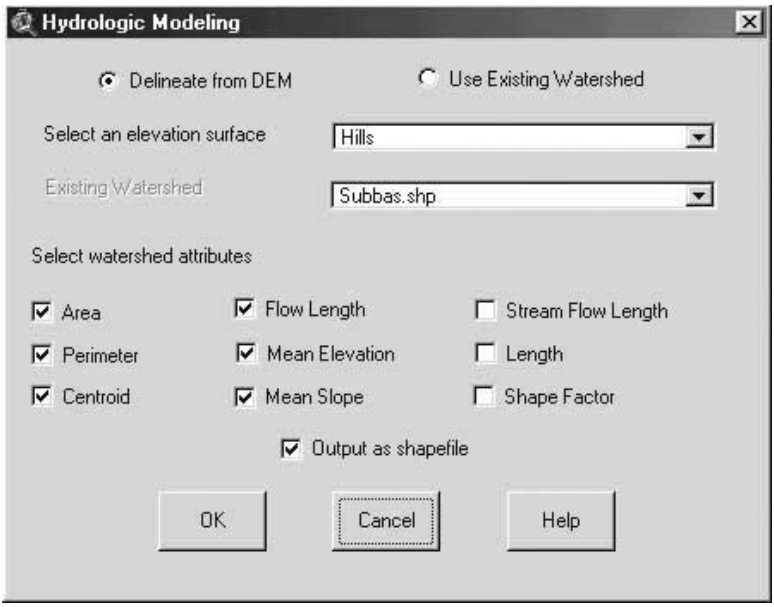


Figure 4.6 Hydro extension Hydrologic Modeling Dialogue.

ARC GRID Extension

ARC GRID is an optional extension for ESRI's ArcInfo® 7.x GIS software package. GRID adds raster geoprocessing and hydrologic modeling capability to the vector-based ArcInfo GIS. For hydrologic modeling, the extension offers a Hydrologic Tool System and several hydrologically relevant functions for watershed and stream network delineation.

The FLOWDIRECTION function creates a grid of flow directions from each cell to the steepest downslope neighbor. The results of FLOWDIRECTION are used in many subsequent functions such as stream delineation. The FLOWACCUMULATION function calculates upstream area or cell-weighted flow draining into each cell. The WATERSHED function delineates upstream tributary area at any user-specified point, channel junction, or basin outlet cell. This function requires step-by-step calculations. Arc Macro Language (AML) programs can be written to automate this function for delineating subbasins at all the stream nodes.

GRID can find upstream or downstream flow paths from any cell and determine their lengths. GRID can perform stream ordering and assign unique identifiers to the links of a stream network delineated by GRID. Spatial intersection between streams and subbasins can define the links between the subbasins and streams. This method relates areal attributes such as subbasin nutrient load to linear objects such as streams. The NETWORK function can then compute the upstream accumulated nutrient load for each stream reach (Payraudeau et al., 2000). This approach is also useful in DEM-based runoff quality modeling.

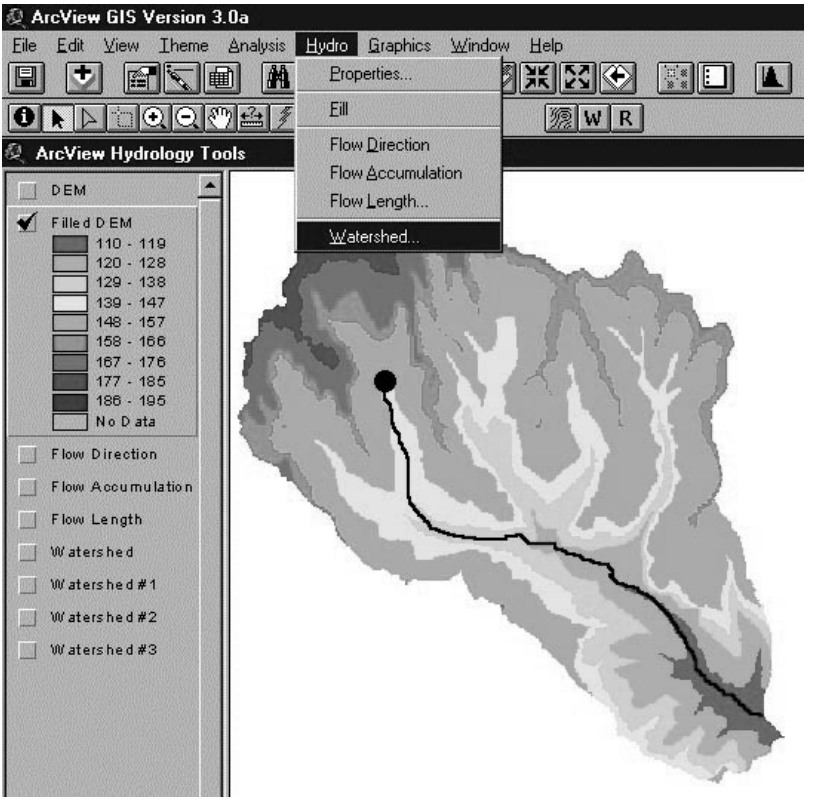


Figure 4.7 Hydro extension's pour point feature.

IDRISI

IDRISI is not an acronym; it is named after a cartographer born in 1099 A.D. in Morocco, North Africa. IDRISI was developed by the Graduate School of Geography at Clark University. IDRISI provides GIS and remote sensing software functions, from database query through spatial modeling to image enhancement and classification. Special facilities are included for environmental monitoring and natural resource management, including change and time-series analysis, multicriteria and multiobjective decision support, uncertainty analysis (including Bayesian and Fuzzy Set analysis), and simulation modeling (including force modeling and anisotropic friction analysis). TIN interpolation, Kriging, and conditional simulation are also offered.

IDRISI is basically a raster GIS. IDRISI includes tools for manipulating DEM data to extract streams and watershed boundaries. IDRISI GIS data has an open format and can be manipulated by external computer programs written by users. This capability makes IDRISI a suitable tool for developing hydrologic modeling applications. For example, Quimpo and Al-Medeij (1998) developed a FORTRAN

program to model surface runoff using IDRISI. Their approach consisted of delineating watershed subbasins from DEM data and estimating subbasin runoff curve numbers from soils and land-use data.

Figure 4.8 shows IDRISI's DEM analysis capabilities. The upper-left window shows a TIN model created from digital contour data. The upper-right window shows a DEM created from the TIN with original contours overlayed. The lower-right window shows an illuminated DEM emphasizing relief. The lower-left window shows a false color composite image (Landsat TM bands 2, 3, and 4) draped over the DEM (IDRISI, 2000).

TOPAZ

TOPAZ is a software system for automated analysis of landscape topography from DEMs (Topaz, 2000). The primary objective of TOPAZ is the systematic identification and quantification of topographic features in support of investigations related to land-surface processes, H&H modeling, assessment of land resources, and management of watersheds and ecosystems. Typical examples of topographic features that are evaluated by TOPAZ include terrain slope and aspect, drainage patterns and divides, channel network, watershed segmentation, subcatchment identification, geometric and topologic properties of channel links, drainage distances, representative subcatchment properties, and channel network analysis (Garbrecht and Martz, 2000). The FILL Function of TOPAZ recognizes depressions created by embankments and provides outlets for these without filling, a better approach than the fill-only approach in other programs (e.g., IDRISI or Spatial Analyst).

CASE STUDIES AND EXAMPLES

Representative applications of using DEM data in GIS are described in this section.

Watershed Delineation

A concern with streams extracted from DEMs is the precise location of streams. Comparisons with actual maps or aerial photos often show discrepancies, especially in low-relief landscapes (ASCE, 1999). A drainage network obtained from a DEM must be comparable to the actual hydrologic network. Thus, it is worthwhile to check the accuracy of DEM-based delineations. This can be done by comparing the DEM delineations with manual delineations. Jenson (1991) found approximately 97% similarity between automatic and manual delineations from 1:50,000-scale topographic maps.

The objective of this case study was to test the efficacy of DEM-based automatic delineation of watershed subbasins and streams. It was assumed that manual delineations are more accurate than DEM delineations. Thus, a comparison of manual and DEM delineations was made to test the accuracy of DEM delineations.

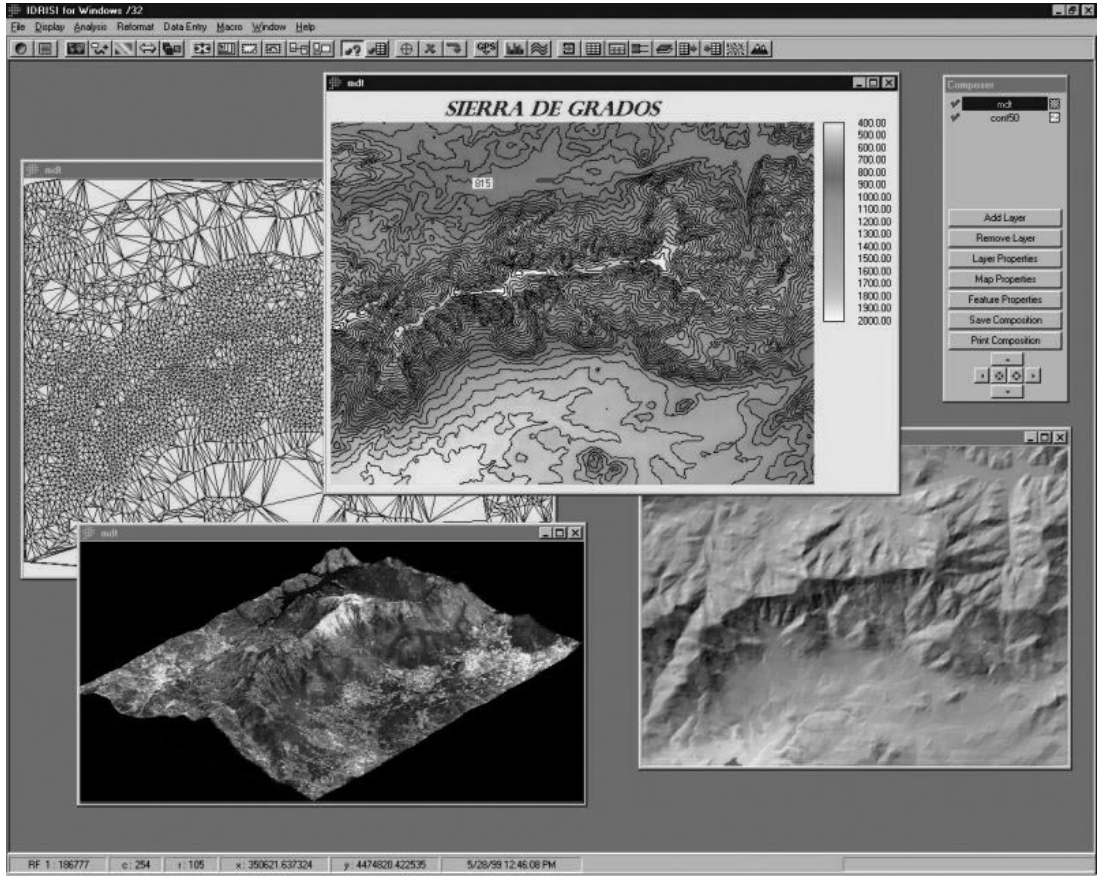


Figure 4.8 IDRISI's DEM analysis features.

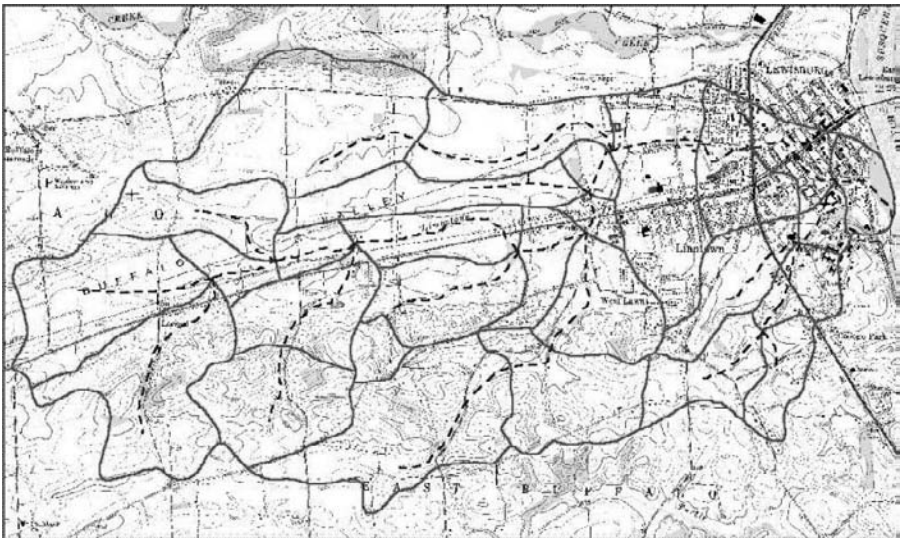


Figure 4.9 Bull Run Watershed showing manual subbasins and streams.

The case study watershed is the Bull Run Watershed located in Union County in north-central Pennsylvania (Shamsi, 1996). This watershed was selected because of its small size so that readable report-size GIS maps can be printed. The proposed technique has also been successfully applied to large watersheds with areas of several hundred square miles. Bull Run Watershed's 8.4 mi^2 (21.8 km^2) drainage area is tributary to the West Branch Susquehanna River at the eastern boundary of Lewisburg Borough. The 7.5-min USGS topographic map of the watershed is shown in Figure 4.9. The predominant land use in the watershed is open space and agricultural. Only 20% of the watershed has residential, commercial, and industrial land uses.

Manual watershed subdivision was the first step of the case study. The 7.5-min USGS topographic map of the study area was used for manual subbasin delineation, which resulted in the 28 subbasins shown in Figure 4.9. This figure also shows the manually delineated streams (dashed lines).

Next, ArcView Spatial Analyst and Hydro extension were used to delineate subbasins and streams using the 7.5-min USGS DEM data. Many cell threshold values (50, 100, 150, …, 1000) were used repeatedly to determine which DEM delineations agreed with manual delineations. Figure 4.10, Figure 4.11, and Figure 4.12 show the DEM subbasins for cell thresholds of 100, 250, and 500. These figures also show the manual subbasins for comparison. It can be seen that the 100 threshold creates too many subbasins. The 500 threshold provides the best agreement between manual and DEM delineations.

Figure 4.13, Figure 4.14, and Figure 4.15 show the DEM streams for cell threshold values of 100, 250, and 500. These figures also show the manually delineated streams for comparison purposes. It can be seen that the 100 threshold creates too many streams (Figure 4.13); the 500 threshold looks best (Figure 4.15) and

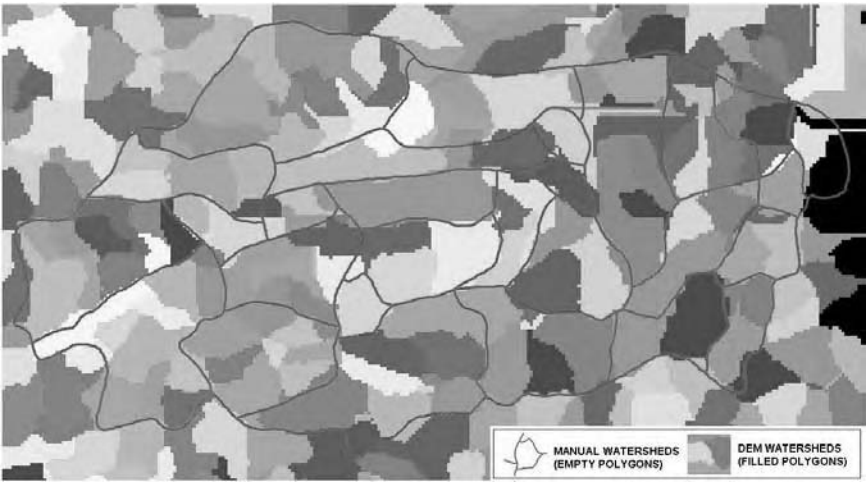


Figure 4.10 Manual vs. DEM subbasins for cell threshold of 100 (too many subbasins).

provides the best agreement between the manual and DEM streams. The upper-right boundary of the watershed in Figure 4.15 shows that one of the DEM streams crosses the watershed boundary. This problem is referred to as the boundary "cross-over" problem, which is not resolved by altering threshold values. It must be corrected by manual editing of DEM subbasins or using DEM preprocessing methods such as the stream burning method described earlier.

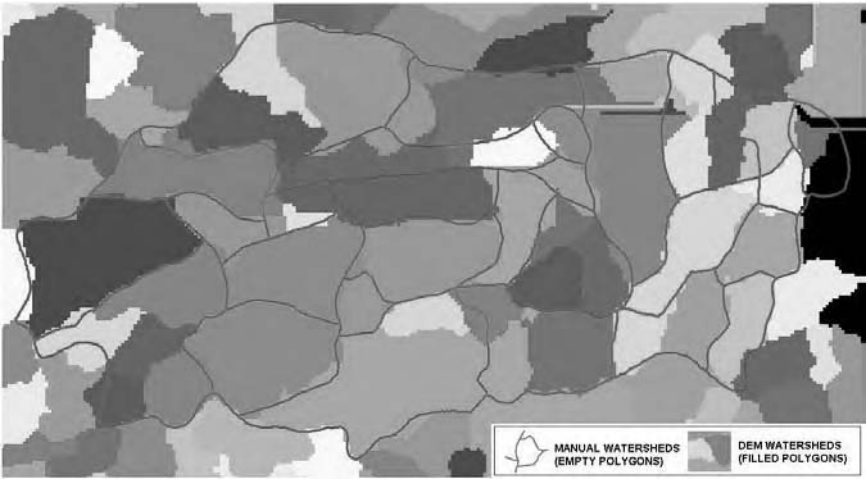


Figure 4.11 Manual vs. DEM subbasins for cell threshold of 250 (better).

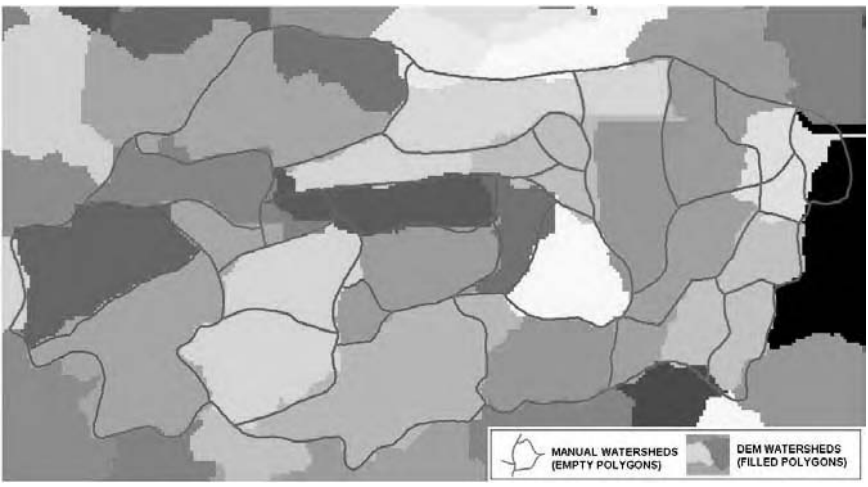


Figure 4.12 Manual vs. DEM subbasins for cell threshold of 500 (best).

Figure 4.16 shows DEM-derived subbasin and stream maps for a portion of the very large Monongahela River Basin located in south western Pennsylvania, using the 30-m USGS DEM data and a cell threshold value of 500 cells.

From the Bull Run watershed case study, it can be concluded that for rural and moderately hilly watersheds, 30-m resolution DEMs are appropriate for automatic delineation of watershed subbasins and streams. The 30-m DEMs work well for the mountainous watersheds like those located in Pennsylvania where subbasin boundaries

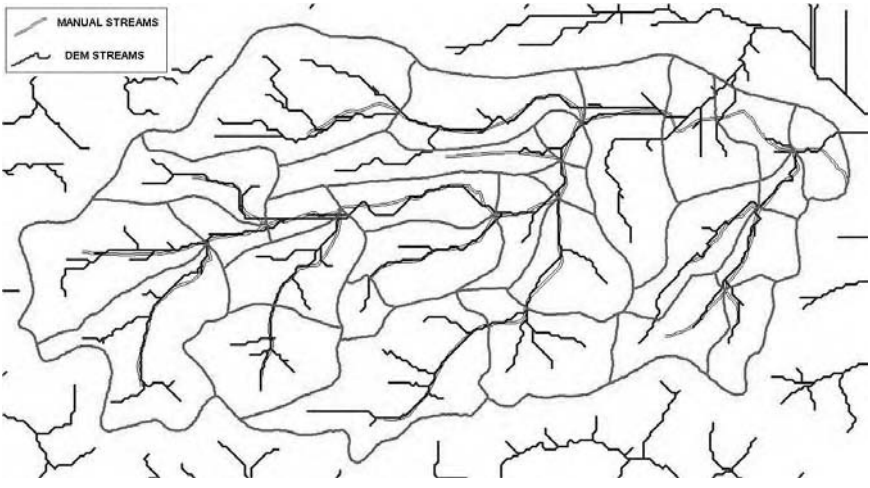


Figure 4.13 Manual vs. DEM streams for cell threshold of 100 (too many streams).

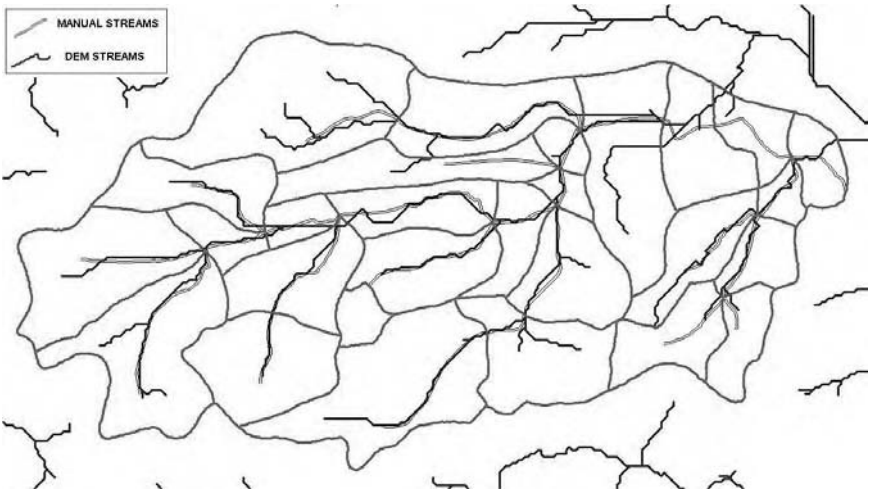


Figure 4.14 Manual vs. DEM streams for cell threshold of 250 (better).

are well defined and distinct due to highly variable topography. For 30-m USGS DEMs, a cell threshold of 500 is appropriate. Stream networks generated from the 30-m USGS DEMs at a cell threshold value of 500 were satisfactory, with a minor watershed boundary crossover problem. Correction of boundary crossover problems might require some manual intervention to ensure the efficacy of the DEM-based automatic

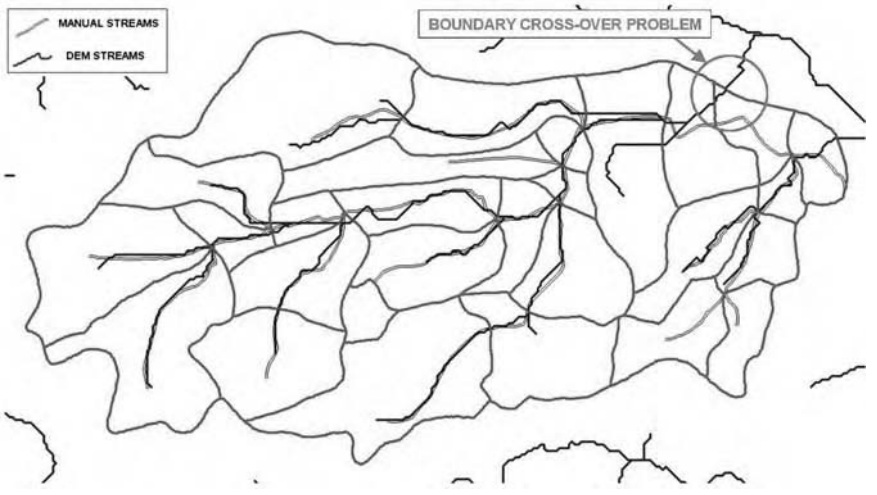


Figure 4.15 Manual vs. DEM streams for cell threshold of 500 (best).

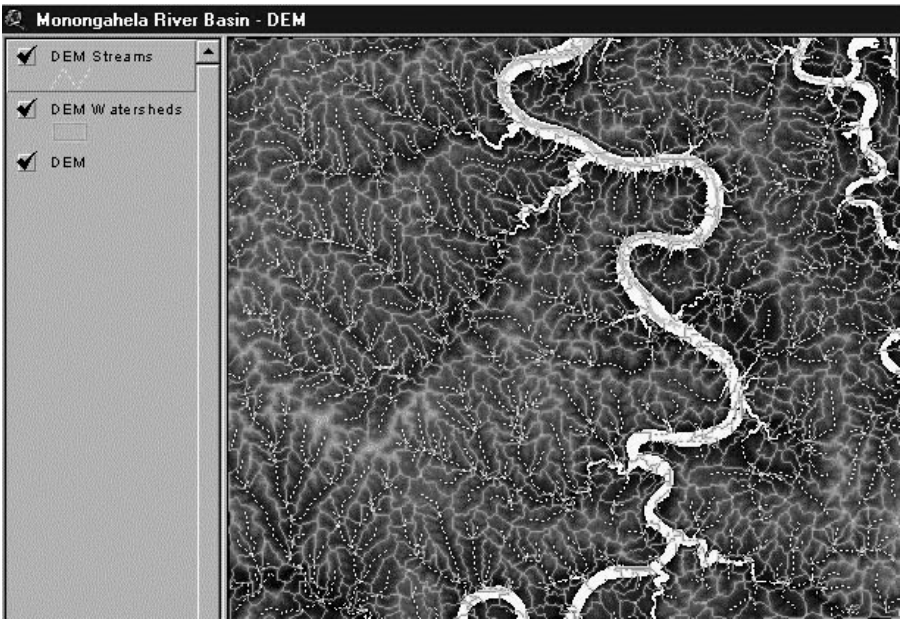


Figure 4.16 DEM-derived watersheds and streams for a large watershed.

watershed delineation approach. Compared with the manual watershed delineation approach, the DEM approach offers significant time savings, eliminates user subjectivity, and produces consistent results.

Sewershed Delineation

Watersheds define the drainage areas of natural streams and rivers. Sewersheds define the drainage areas of man-made sewers and storm drains. The conventional manual method of sewershed delineation is cumbersome, and an automated method is desirable. Can we use DEMs for automatic delineation of sewersheds? Unfortunately, the current research indicates that DEMs are generally applicable to natural landscapes and may not apply to urbanized settings (ASCE, 1999). This limitation is expected to be more pronounced in low-relief or flat areas. DEM delineations are based on the premise that water flows in the downslope direction, i.e., that all the sewers are gravity pipes. Therefore, DEM-based sewershed delineations will not be accurate for a pumped system with force mains and will require manual editing of the sewershed boundaries.

Figure 4.17 shows a comparison of DEM watersheds (filled polygons) and manually delineated sewersheds for the gravity sewers area of the Borough of Charleroi located near Pittsburgh, Pennsylvania. The watersheds were created from the 30-m resolution USGS DEM of the study area using a cell threshold of 100. The mean land slope of the study area is approximately 15%. The comparison shows that some sewersheds (for example, Sewershed Nos. 3, 4, 5, and 6) are reasonably

Figure 4.17 Comparison of DEM watersheds (filled polygons) and manually drawn sewersheds.

close to the watershed boundaries and some sewersheds (for example, Sewershed No. 2) are quite different from the watershed boundaries.

The Massachusetts Water Resources Authority (MWRA) recently determined that DEMs were not too helpful in delineating sewershed boundaries (Herrlin, 2000). MWRA, instead, used a "Euclidean Distance" approach consisting of the following steps:

1. All the sewers tributary to a pour point (point where a trunk sewer connects an interceptor) are identified using a custom network-tracing utility (upstream trace). The group of pipes identified like this defines a tributary area.
2. A new attribute with the name of the tributary area is manually added to all the sewers. This step is quite laborious but needs to be done only once.
3. Vector sewer layers are converted to a grid format. This is necessary because the Euclidean distance approach works only on raster layers.
4. The Euclidean distance allocation function is applied to all the sewer cells. The Euclidean distance is calculated in the same manner as finding the hypotenuse of a triangle. This generates a sewershed boundary around sewers.
5. The raster Euclidean distance boundaries are converted back to vector (polygon) format.

Some techniques that may be used for improving the accuracy of DEM-based sewershed delineations are listed below:

1. Use large-scale, high-resolution DEMs derived from digital orthophotos. These relatively dense DEMs have break lines at breaks in grades, which may provide valuable information for delineating sewersheds in flat urban areas. LIDAR technology can also provide highly accurate and cost-effective DEMs and 3D break lines that can be used for delineating sewersheds.
2. Burning in the DEMs with streams improves the accuracy of delineated watersheds in flat areas. Burning in the DEMs with sewers may improve the delineated sewersheds.
3. Adding manhole rim elevations to DEMs may be helpful in delineating sewersheds. Alternatively, DEMs may be burned in with manhole rim elevations.

Water Distribution System Modeling

The performance of a water distribution system depends on variations in ground surface elevation. For example, high-elevation areas may experience inadequate pressure if appropriate pumping is not provided. Thus, all distribution system models require input data for the elevation of each model node. DEMs can be used for 3D representation of water systems and estimating elevation at the nodes of a water distribution network model.

Shamsi (1991) showed applications of personal computer-based 3D graphics in network and reliability modeling of water distribution systems. Optional 3D extensions of GIS software, such as ESRI's 3D Analyst, can create 3D maps of terrain, water demand, and modeled pressures. Figure 4.18 shows three stacked surfaces: ground surface elevations, demands, and modeled average daily pressures for the Borough of White Haven, Pennsylvania. The land area of the borough is 773 acres, the population is 1,091 persons, and the average daily demand is 197,000 gal/d. The distribution system consists of water mains ranging in size from 2 to 8 in. It is interesting to note the high-pressure areas coinciding with the low-elevation areas, and vice versa. It also quickly becomes apparent that the high-demand areas predominate in the lower elevations. A display such as Figure 4.18 is an effective visual aid to displaying the hydraulic model results. Most importantly, such graphics can be easily understood by system operators without the extensive training that is usually required to understand the model output in tabular format.

Node elevation is a critical input parameter for all water distribution system models. Traditionally, the node elevations have been determined manually by overlaying the water distribution network over a topographic map of the study area. Unfortunately, due to its tedious nature, this approach is cumbersome. GIS is now being used to automatically extract the node elevations from a DEM of the study area. This approach requires a vector layer of the network nodes, a DEM grid, and a user script. The script appends a new "Elevation" field to the "Node" layer table, queries the DEM grid at node locations, and writes the elevation values in the Elevation field.

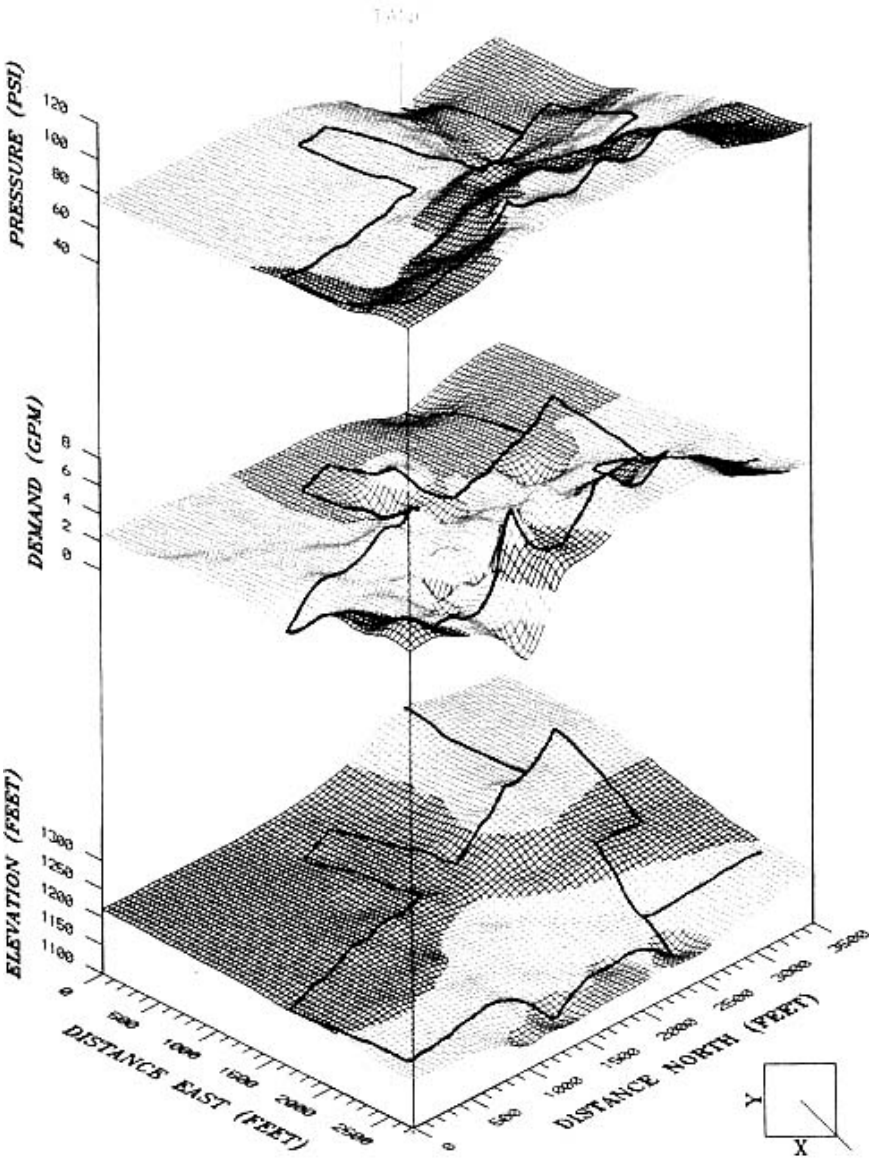


Figure 4.18 Hydraulic modeling results as 3D surfaces.

WaterCAD Example

If node locations are available in an ArcView Shapefile, ArcView 3D Analyst extension (Version 3.x) can be used to calculate node elevation from a DEM grid. This extension uses bilinear interpolation to calculate the node elevation from the

grid cell elevation data. The elevation data are pulled from the DEM overlay, and the node file is converted to a 3D file. The steps for this procedure for Haestad Method's WaterCAD modeling software are listed below (Walski et al., 2001):

1. Start ArcView and load the 3D Analyst extension.
2. Add both the DEM grid and the node point themes to the same view.
3. Select "Theme\Convert to 3D Shapefile." Select "Surface" for the elevation file type and enter the name of the DEM grid (input). Also provide a name for the 3D Shapefile (output).
4. Use an Avenue script to extract elevation data from the 3D Shapefile and place them in the node theme table. A sample script called addxyz.ave (Hsu, 1999; Shamsi, 2002) can be downloaded from the ESRI ArcScripts Web site (www.esri.com/arcscripts/scripts.cfm).
5. Select "File\Export" from the ArcView pull-down menus and enter the export filename for a dBase IV (DBF) or delimited text (TXT) file.
6. Select "File\Synchronize\database Connections" from the WaterCAD's pull-down menus and set up a link between the export file (created in the previous step) and WaterCAD.
7. Select "Elevation" as the WaterCAD field into which the elevation data should be imported and the name of the corresponding field in the export file (e.g., "Z-coord").
8. Select the link in the Database Connection Manager and click the "Synchronize In" button.
9. Finally, validate the success of elevation data import process by checking the elevations at a number of nodes against known elevations.

USEFUL WEB SITES

Center for Research in Water Resources, University of Texas at Austin	www.crwr.utexas.edu/archive.shtml
DEM translators	software.geocomm.com/translators/sdts/
Pennsylvania Spatial Data Access system (PASDA)	www.pasda.psu.edu
ESRI DEM analysis software	www.esri.com
Leica geosystems software	gis.leica-geosystems.com
USGS DEM download	edc.usgs.gov/geodata/
USGS EartExplorer site for buying DEM data	earthexplorer.usgs.gov
USGS NED home page	edcnts12.cr.usgs.gov/ned/default.asp
USGS node of the National Geospatial Data Clearinghouse	nsdi.usgs.gov

CHAPTER SUMMARY

GIS provides an integrated platform for using DEMs. DEM data can be used for automatic watershed and stream delineation and computation of watershed hydrologic parameters. DEMs can be used as source data for digital orthophotos and as base maps in a GIS. DEM data can be created from topographic maps and remote

sensing data. DEMs are not intended to replace elevation data obtained from surveys, high-accuracy GPS, or even well-calibrated altimeters. Rather, DEMs should be used as a labor-saving process to relieve the user of the tedium and human error involved in the conventional manual interpolation method from paper maps. Public-domain DEMs generally do not have adequate resolution and accuracy for engineering design. They are suitable for applications where a high degree of accuracy is not required, such as H&H modeling.

CHAPTER QUESTIONS

1. What are DEMs and how are they used in GIS?
2. List at least ten applications of DEMs?
3. What are the limitations of DEM data? How are they overcome?
4. How are DEMs used for delineating (1) streams, (2) watersheds, and (3) sewersheds?
5. How are DEMs created using the remote sensing technology?

CHAPTER 5

GPS Applications

GPS technology is no longer limited to determining coordinates. It has become an efficient and increasingly popular way for collecting GIS attribute data for water and wastewater infrastructure.

GIS data collection for sewer system manholes using GPS.

LEARNING OBJECTIVE

The learning objective of this chapter is to discover the applications of global positioning system (GPS) technology in water industry GIS projects.

MAJOR TOPICS

- GPS basics
- GPS applications in water industry
- GPS applications in GIS
- GPS survey steps
- GPS equipment
- GPS software

LIST OF CHAPTER ACRONYMS

DGPS Differential Global Positioning System
GPS Global Positioning System
GLONASS Global Navigation Satellite System (Russian System)
NAVSTAR Navigation System by Timing and Ranging
PDOP Positional Dilution of Precision
RTK Real Time Kinematic
SA Selective Availability
WGS84 World Geodetic System of 1984

STREAM MAPPING IN IOWA

Application	Stream traversing and mapping channel features using GPS
GPS software	Trimble's Pathfinder Office
GPS equipment	Trimble Pathfinder Pro XR
GPS accuracy	1 m for discrete data and 5 m for continuous data
GPS data	Streambank conditions, bottom sediment material and thickness, channel cross sections, debris dams, tile lines, tributary creeks, and cattle access points
Study area	Neal Smith National Wildlife Refuge, Jesper County, Iowa
Organization	Iowa Department of Natural Resources

In 2000, GIS and GPS were used together to map a 12-km portion of Walnut Creek located near Prairie City in Jesper County, Iowa (Schilling and Wolter, 2000). The objective of the mapping project was to locate channel features and identify spatial trends among alluvial system variables that could be used to identify and prioritize portions of the stream channel and watershed in need of further investigation or restoration. Using Trimble Pathfinder Pro XR GPS equipment, discrete locations (channel cross sections, debris dams, tile lines, tributary creeks, and cattle access points) were mapped to an accuracy of 1 m, whereas the continuous data (bank erosion rates, streambed materials, and thickness) were recorded to an accuracy of

5 m. To record continuous conditions, the GPS equipment was operated in continuous line mode with location recorded every 5 sec. GPS data were exported into a GIS format (ESRI Shapefile), using Pathfinder Office software. Field descriptions of the continuous line segments and discrete features were added to the GPS location information to create various GIS layers. Segment lengths varied from 10 to 50 m. Discrete channel features were located by pausing the continuous line mode of the GPS and taking points at feature locations. Stream survey data were used to model watershed conditions, identify water sampling points, and evaluate and select appropriate channel-rehabilitation measures.

This chapter is intended for professionals in the geographic "positioning" field. It presents applications of GIS and GPS for water industry infrastructure management. The use of GPS in collecting attributes data is discussed, and methods of data attribution are described. A review of GPS equipment and software is presented. GPS accuracy issues are also discussed.

GPS BASICS

GPS, also referred to as Navigation System by Timing and Ranging (NAVSTAR), is a satellite-based radio navigation system developed and operated by the U.S. Department of Defense. The Russian government operates a similar system called Global Navigation Satellite System (GLONASS). At the present time, GPS includes 29 active satellites located in 6 orbital planes.

GPS systems utilize a constellation of satellites orbiting the earth twice daily (i.e., passing over approximately the same world location every 12 hours) and transmitting precise time and position signals. GPS receivers read signals from orbiting satellites to calculate the exact spot of the receiver on Earth as geographic coordinates (latitude and longitude) referenced to the World Geodetic System of 1984 (WGS84) datum. The signals from at least four satellites should be available to determine the coordinates of a position. Physical obstructions such as mountains, trees, and buildings, and other factors such as satellite malfunction and rephasing operations can restrict GPS signals and degrade GPS accuracy. Ideal GPS operating conditions that provide the best accuracy are listed below (Lyman, 2001):

- Low positional dilution of precision (PDOP), a measure of best geometrical configuration of satellites
- Good signal strength
- Little or no multipath (reflection of GPS signals off distant reflective environment such as mountains and buildings)
- Little or no signal degradation because of geomagnetic storms and ionospheric or atmospheric effects

The accuracy of GPS coordinates can be increased by applying differential corrections. Differential corrections move user points closer to their "actual" location. This is done by comparing the user's new data on unknown locations with the data collected at the same time on a point with known coordinate values (Zimmer, 2001b). The GPS receivers that can receive and apply the corrections in real time

are called real time kinematic (RTK) receivers. Non-RTK receivers require postprocessing of raw GPS data in the office. In the U.S., federal and state agencies are cooperating to make differential GPS readily available to all users.

GPS precision can vary from a few millimeters to hundreds of meters. The required precision depends on the project-specific requirements. The available precision varies with GPS mode (static, RTK, or kinematic), GPS equipment, time of occupation, and location (vegetation, reflection, and buildings). GPS survey cost increases with the accuracy requirements. The typical utility precision standard is 3 to 5 cm.

To prevent misuse by hostile forces, the U.S. Department of Defense had introduced an intentional error called the Selective Availability (SA) error in their GPS signals. The recent advances in GPS technology had reduced the effectiveness of the SA error. On midnight May 2, 2000, the SA error was removed 6 years ahead of schedule by a presidential order. This event marked an important day in the history of GPS because it increased the GPS accuracy up to ten times. The SA removal has improved the accuracy of inexpensive GPS receivers. SA removal has also increased the accuracy of GPS receivers operating in an autonomous (unassisted) mode without a base station. No significant impact has been noted in the performance of survey grade receivers (Murphy, 2001).

GPS technology is making major progress in improving the speed, reliability, and accuracy of the mathematical processes by which coordinates are calculated from the satellite beams (ASCE, 2001).

GPS APPLICATIONS IN THE WATER INDUSTRY

At the present time, the GPS revolution is well under way. For example, a Mercedes-Benz driver equipped with TeleAid system can press an SOS button to summon a tow truck, police, or ambulance. This button uses GPS technology to transmit the specific location, model, and color of the vehicle. The GPS applications for the water, wastewater, and stormwater systems, though not as dramatic as TeleAid, are revolutionizing the way these systems are designed, constructed, operated, and maintained. Representative GPS applications for the management of water, wastewater, and stormwater systems are:

1. GPS can be used to increase the accuracy of existing system maps by verifying and correcting locations of the system components. Frequent field changes often mean utility lines can be several feet off horizontally and/or vertically from where they appear on the plans. Thus, unless updated frequently, most utility plans, especially in growing cities, are outdated frequently. GPS data collection is no longer limited to collecting coordinates of point features. Now users can bike along a channel to map line features, or walk around a detention pond to map polygon features.
2. New water system or sewer system maps can be created if they do not exist.
3. Water system or sewer system attributes can be collected for populating the GIS database.

Surveying

Traditionally, people have used geodetic surveying to locate and map utility infrastructure components. Such transit survey required traversing between a known point to the point of interest, which often took several hours per point. GPS has been found to be up to 50% faster than the traditional methods (Anderson, 1998). GPS survey takes only a few minutes or seconds per point. For example, up to 300 points can be surveyed in 1 day, using a GPS survey. A two-man crew with bicycle-mounted equipment can survey up to 500 points per day.

The familiar total station is still the backbone of engineering survey work. Although GPS is not a replacement for optical surveying, interoperability between optical equipment and GPS is growing, and GPS is gaining ground slowly among engineers. GPS is an ideal addition to the surveying toolbox for a variety of applications, such as locating the starting point for a new stakeout. The conventional total station survey will require traversing a large distance for this work. GPS can do this work much faster by navigating a person right to the point where the stake should be placed (ASCE, 2001).

Fleet Management

An efficient fleet management system is essential to improving customer service. A wireless system that uses GIS and GPS technologies can substantially and economically improve the efficiency of fleet management. An integrated GIS/GPS procedure can be used to track multiple moving vehicles from a command center. It can show the location, speed, and movement of each vehicle on the tracking display. Off-the-shelf mobile devices such as Web-enabled cell phones and personal digital assistants (PDAs) can be used in conjunction with a GIS/GPS to provide the information needed for fleet management, such as dispatching and tracking the maintenance vehicles, generating driving directions, and trip routing. For example, Gearworks customized MapInfo's MapMarker J Server and Routing J Server to calculate driving directions and travel statistics. A client-side custom mapping application was created, which communicates with a MapXtreme server to create the data requested by the dispatcher and to deliver the information to the drivers mobile device. This application allows the dispatcher to view a map of the entire fleet to better assign work orders and deliveries, perform real time tracking, and deliver accurate status updates through the Web interface (GEOWorld, 2001).

GIS/GPS applications have consistently lowered the cost of fleet management by 10 to 15%.

GPS APPLICATIONS IN GIS

GPS technology represents a space-age revolution in GIS data collection. It is providing an efficient and increasingly popular way for collecting both the location (coordinates) and the attributes data in the field. The new line of GPS receivers brings

technology to water- and wastewater system operators and managers, who can populate their existing maps with precise location of features such as manhole covers, catch basins, overflow points, hydrants, valves, pumps, flow meters, and rain gauges.

The most basic application of GPS is collection of (x,y) coordinates for the GIS features. Initially, these coordinates were manually entered into the GIS database. Most GPS receivers record their data in an American Standard Code for International Interchange (ASCII) format that can be imported into a GIS without having to type the coordinates' values. Recent GPS equipment can provide data in a GIS-compatible format, such as the ESRI Shapefile format.

Attributes are usually collected in two phases:

1. Initial attribution phase (Phase I) or first pass: Limited attributes can be collected by the GPS survey crew when a structure is visited or surveyed for the first time. These attributes — such as manhole cover type, catch basin condition (clean, debris, etc.), or outfall condition (submerged, flowing, dry, etc.) — should be visible from outside and should not require confined space entry or opening of structures. This phase may also include ID-marking of structures for the second phase or future visits. In this phase, typically 250 to 300 points can be surveyed per day.
2. Final attribution phase (Phase II) or second pass: Additional attributes can be collected when a structure is revisited by the attribute crew. These attributes may require field measurements, detailed inspections, confined space entry, or opening of structures. Examples include manhole depth, pipe size, and structural or hydraulic deficiencies. The final attributes depend on the mapping application. For example, if the maps will be used to develop an H&H model, weir height in an overflow structure might be measured. If the GIS will be used for NPDES permit reporting, the condition of the sanitary outlet (clean, clogged, surcharged, etc.) in the overflow structure might be a critical attribute. In this phase, final attributes for 50 to 75 structures can be surveyed per day. New real time differential GPS (DGPS) receivers can navigate the attribute crew back to the structures visited in Phase 1.

GPS SURVEY STEPS

The data conversion process for municipal utilities often begins with GPS data. Features such as manholes and hydrants are collected using a GPS unit. These point features are combined into a GIS layer and then attributed. At the conclusion of each day, GPS corrections are applied if necessary. The data set is massaged to ensure its stability including calculating the IDs and building the data set. The data set is then exported and saved as a backup file. After the process is completed for an entire area, check plots are created. These check plots show the lines and points along with their attributes. The flow direction of the lines is also displayed to ensure that proper connectivity has been established for the system. The plots are then delivered to the municipal engineer for review.

Generally, the following steps are required to collect GIS data, using GPS (Zimmer, 2001b):

1. Prepare a project plan to define the scope, resource inventory (paper maps, databases, etc.), and data dictionary. The data dictionary is a template for field data

collection. It specifies the features to be surveyed and their attributes. A simple data dictionary for Phase-1 GPS survey of a sewer system is given below:

- Sanitary features
  - San_mh
    - Lid_size
    - Lid_type
    - Condition
    - ID
- Storm features
  - Catchbasin
    - Size
    - Type
    - Condition
    - ID
- Endwall
  - Condition
    - ID

2. Conduct field work.
3. Download data from GPS receiver to computer.
4. Perform quality assurance/quality control (QA/QC) for completeness and attribute information.
5. Apply differential corrections if not using RTK equipment.
6. Calculate the spatial error.
7. Export GPS data and attributes to GIS.
8. Create a metadata file to document GPS survey information, such as survey date, equipment used, horizontal and vertical accuracy, etc.

GPS EQUIPMENT

GPS equipment selection depends on the purpose or application of GPS data. For example, for utility mapping projects, mapping-grade GPS equipment equipped with code-based receivers is suitable. For engineering design and projects that require highly accurate survey data, survey grade GPS equipment equipped with carrier-phase receivers is needed. Table 5.1 provides a summary of GPS receiver types. Accuracy specifications are based on the ideal operating conditions described earlier.

Table 5.1 GPS Receivers

Receiver Type	Horizontal Accuracy (m)	Price (US$)
Recreation	5–10	100–300
Basic	1–5	3,000–5,000
Advance	<1	8,000–10,000
Survey and geodetic	<0.2	25,000 and upward

Recreational GPS Equipment

These types of receivers are intended to support recreational activities such as hiking and camping. Commonly referred to as the "*$200 receivers*," they provide

low accuracy and are, therefore, not suitable for most engineering, mapping, or surveying work. They can be used as basic navigation and reconnaissance tools.

Basic GPS Equipment

These lightweight and handheld systems are about the size of a calculator. They are less expensive ($4000) and less accurate (1 to 5 m; least accurate in woods). Differential corrections are applied in a postprocessing step after the collected field data have been downloaded to a computer running a GPS-processing software package. Because of their low accuracy, they are not suitable for utility surveys. They are more suitable for educational use. Trimble's Geoexplorer III handheld GIS data collection system shown in Figure 5.1 is an example of this category. Geoexplorer

Figure 5.1 Basic GPS equipment example (Trimble Geoexplorer III).

provides support for popular GIS and CAD software packages such as ArcInfo, Intergraph, MGE, MapInfo, ERDAS, and AutoCAD. Geoexplorer provides for GIS export of GPS data, using Trimble's Pathfinder Office software. The steps for using Geoexplorer III GPS data in GIS are given below:

- Install Pathfinder Office software on your PC
- Create a data dictionary, i.e., define attributes for each feature type
- Transfer data dictionary from PC to receiver
- Enter attributes during GPS survey
- After the GPS survey, transfer the location and attribute data from receiver back to Pathfinder Office on your PC
- From Pathfinder Office, export data to your GIS in ArcView Shapefile or MapInfo format

Advanced GPS Equipment

Advanced-grade GPS receiver types include the so-called professional, mapping-grade, and resource-grade products. These systems are more expensive ($10,000) but provide the submeter horizontal accuracy generally required in utility mapping. Differential corrections, which are especially applicable for accurate navigation, can be applied in real time if the GPS receiver is linked to a radio receiver designed to receive broadcast corrections; Trimble's GPS Pathfinder Pro XR/XRS GIS data collection system is an example. In Pathfinder, attributes can be entered using Trimble's handheld Asset Surveyor, or a pen-based notebook or Tablet PC running Trimble's ASPEN software. As with Geoexplorer, GIS export is accomplished using Pathfinder Office software. The XRS models have an integrated differential correction receiver and can provide real time submeter accuracy. Reportedly, Pathfinder also features patented Everest multipath rejection technology to reduce or eliminate multipath signals in reflective environments. Figure 5.2 shows Trimble's Pathfinder Pro XR system, which is also referred to as the "Backpack" because it stores the system components in a convenient backpack. Figure 5.3 shows the components stored inside the Backpack.

Leica Geosystems (Heerbrugg, Switzerland) is a surveying and geomatics company that provides systems for accurate 3D data capturing, visualization, and modeling of space-related data to customers in surveying, engineering, construction, GIS, and mapping. In the GIS community, Leica Geosystems is known for its GPS and field data collection related equipment. ERDAS, Inc. (Atlanta, GA), is the world's leading provider of geographic-imaging solutions. The company's products and related services help organizations visualize, manipulate, analyze, measure, and integrate geographic imagery and geospatial information into 2D and 3D environments. In July 2001, ERDAS's geographic-imaging software company was acquired by Leica Geosystems to form a new GIS and Mapping Division for growth in the GIS and remote sensing markets. Congestion builds up as too many users crowd the same radio frequencies. Leica's new GPS equipment, Leica System 500 (GS50), provides uninterrupted RTK data link through cell phones. Also referred to as the *GIS Receiver*, GS50 is Shapefile-compatible and boasts a new "maximum tracking technology" that allows users to work under trees and in deep urban or natural canyons.

Figure 5.2 Advanced GPS equipment example (Trimble Pathfinder Pro XR Backpack).

Mapping-grade GPS equipment vertical accuracy of ± m is not adequate if accurate elevation data (e.g., manhole rim elevations) are required to support specific applications, such as H&H modeling and engineering design that require precise elevation data.

SURVEY GRADE GPS EQUIPMENT

Survey grade equipment features RTK technology, which can deliver ±1 cm horizontal and vertical accuracy in real time, using dual frequency receivers. Due to their high accuracy, they are expensive ($25,000). Trimble's GPS Total Station 4800 integrated real time kinematic survey system is an example of this category. Due to their more-than-needed accuracy and high cost, they are not essential for wastewater system mapping applications. However, survey grade GPS can save substantial amount of time compared with the conventional ground survey method when high accuracy is required. For example, the RTK survey for a sewer construction project in Fitchburg (Massachusetts) was completed in a quarter of the time that a ground survey would have taken (Mische, 2003). Survey grade GPS equipment's centimeter-level vertical accuracy is adequate for collecting accurate elevation data.

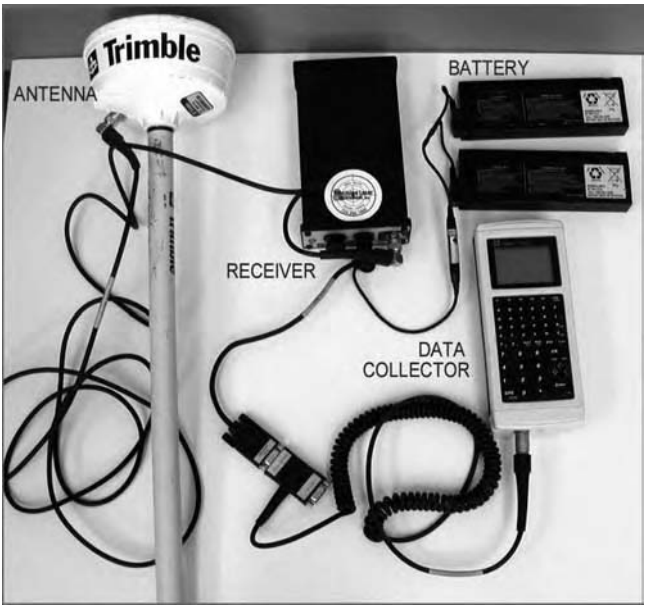


Figure 5.3 Components of Trimble Pathfinder Pro XR Backpack.

USEFUL WEB SITES

Professional Surveyor Magazine	www.profsurv.com
Trimble Navigation	www.trimble.com
Leica Geosystems Software	www.leica-geosystems.com

CHAPTER SUMMARY

Today's GPS equipment is like a GIS mapping device. GPS receivers are being referred to as the GIS receivers because they can collect both the geographic data (coordinates) and the attribute data for the water industry infrastructure. The water system and sewer system feature attributes can be captured from a GPS survey by using field computers, handheld computers, and field data collectors. The GPS equipment should be selected carefully to meet both the horizontal and vertical accuracy requirements of the utility mapping projects. The so-called $200 receivers designed for recreation activities that cannot provide accuracy of 5 m or better are not suitable for GPS mapping of water and sewer systems. A differential GPS that can provide real time submeter horizontal accuracy is suitable if accurate elevations are not necessary. Survey sgrade RTK GPS equipment that provides centimeter-level horizontal and vertical accuracy is suitable if accurate elevations are important.

CHAPTER QUESTIONS

1. What is GPS accuracy, and what factors can affect it?
2. List the typical applications of GPS technology in the water industry.
3. How are GIS and GPS related? How are they helping each other?
4. Describe the typical steps for conducting a GPS survey.
5. List the various types of GPS equipment and their suitability for different applications.

CHAPTER 6

Internet Applications

Internet-based mapping applications are changing the way people use and manage geographic information for their water, wastewater, and stormwater systems.

Johnson County (Kansas) Web-based wastewater mapping application.

LEARNING OBJECTIVE

The learning objective of this chapter is to familiarize ourselves with the applications of Internet technology in GIS for the water industry.

MAJOR TOPICS

- Internet GIS applications
- Internet GIS software
- Internet GIS case studies

LIST OF CHAPTER ACRONYMS

AM/FM/GIS Automated Mapping/Facilities Management/Geographic Information System
DXF Drawing Exchange Format
PDA Personal Digital Assistant
RDBMS Relational Database Management System
WWW World Wide Web

DUBLIN'S WEB MAP

Application	Web-based mapping
GIS software	AutoCAD, AutoCAD Map, and MapGuide
GIS data	1-m, natural color, orthorectified, precision pan-sharpened imagery of the entire city from Space Imaging (Thornton, Colorado) and vector layers for school districts, corporate limits, county boundaries, subdivisions, streets, parcels, and water (mains and hydrants), stormwater (sewers, structures, and streams), and sanitary sewer (sewers and manholes) systems
Study area	Dublin, Ohio

One of the fastest growing communities in the Midwest, Dublin's 22 mi^2 area is home to about 30,000 people and 2,500 businesses. The city received the year 2000 Best Practices Award from the Ohio Geographically Referenced Information Program, a state agency responsible for educating local government about GIS technology. The city is constantly enhancing its GIS capabilities and serves as a model medium-sized municipal GIS for the state. This application was developed after City personnel started to realize that retrieving data from hard-copy records was inefficient because it reduced staff time to perform engineering review and analysis. The city needed an automated system that could recall, evaluate, and reproduce data more quickly. The City implemented a Web-based GIS data server using Autodesk's (San Rafael, California) Web-based MapGuide technology

Figure 6.1 Screenshot of Dublin (Ohio) Internet GIS Web Site.

(Cerasini and Herskowitz, 2001). Figure 6.1 shows a screenshot of Dublin's Internet GIS Web site.

In addition to serving the development department, Dublin's GIS serves all other city departments and divisions, with data distributed to more than 200 users. Dublin's GIS Web site (gis.dublin.oh.us) provides quick access to utility data for maintenance, planning, and emergency response activities. The users range from planning and engineering professionals to grounds maintenance staff. A major challenge for any GIS is deciding who will maintain the data and how. Dublin's management structure encourages departments and divisions to exchange information freely, allowing departments to maintain their own data within the enterprise framework while feeding the GIS. Such an approach benefits everyone.

In 2001, the City started to implement GIS based asset management systems for the water, wastewater, and stormwater systems. Dublin also moved much of the GIS mapping and city assets to a mobile environment with wireless technology and GPS to allow data products to be distributed to a mobile workforce with personal digital assistants (PDAs) and wireless notebooks. The future is expected to bring even more integration between traditional municipal applications and GIS because the City also plans to integrate other applications, including work orders, asset management, crime analysis, public safety, and emergency dispatch.

Determining how much money Internet GIS technology is saving the City of Dublin is difficult to assess. However, the City feels that in most cases the time and effort required to respond to utility data requests have decreased from several hours to minutes, which translates to increased staff productivity and enhanced responsiveness to customers.

This chapter provides information on the GIS applications for serving water, wastewater, and stormwater system maps on the Internet.

INTERNET GIS

The Internet is the place where an increasing number of people are going for their mapping needs. In fact, the word *map* is one of the most commonly searched keywords on the Web (ESRI, 2002).

Imagine going to a GIS Web site, zooming to a piece of property, defining the building footprint, and displaying various cut and fill configurations. You could then simulate a 10-year, 24-hour design storm to model where the runoff will go and how you would control it to meet the local stormwater management ordinance.

The Internet is the fastest growing and most efficient technology for distributing GIS data to the public. The Internet and the World Wide Web (WWW or Web) are simplifying how the maps are created and maintained. The Internet is facilitating GIS data sharing between different organizations. Intranets are making it easy to share GIS data between coworkers of the same organization. Until the mid-1990s spatial information was generally distributed via the Internet as static map images. In the late 1990s dynamic GIS Web pages started allowing the public to interactively display and query GIS databases, using what is now referred to as Internet GIS technology (Wild and Holm, 1998).

Web browsers now provide both experts and novices with a common, powerful, inexpensive, and intuitive interface for accessing a GIS. Because casual public users can now access GISs through Web browsers, the water industry GIS staff is free to spend more time on improving the database for providing better customer service (Irrinki, 2000). Many large utilities are now using the Internet and GIS to improve the operation and management of their water and sewer systems. They are using Internet and GIS-based work order management and customer service software to accomplish this.

Figure 6.2 shows a screenshot of an Internet GIS Web site for the municipality of Penn Hills, Pennsylvania, created using ESRI's ArcIMS software. The tools shown at the top of the screen can be used to navigate (pan, zoom, etc.) and query the maps. Users can also apply the measure tool to determine distance and area from the map. The right side of the screen helps the users select the active layer for analysis (e.g., query) and turn it on (visible) or off (invisible). The bottom of the screen shows the query results.

Cincinnati Water Works reported a $2-million cost savings in the first year of using Internet GIS (Anderson, 2000).

GIS and enterprise database management systems can be combined to create geospatial data warehouses. For example, using the power of Oracle's spatial extensions, a data mining software (e.g., KnowledgeMiner from SpatialAge Solutions, Atlanta, Georgia, www.byers.com) can mine a network inventory model from an automated mapping/facilities management/geographic information system (AM/FM/GIS) and make it readily available to the enterprise using the Web. The object

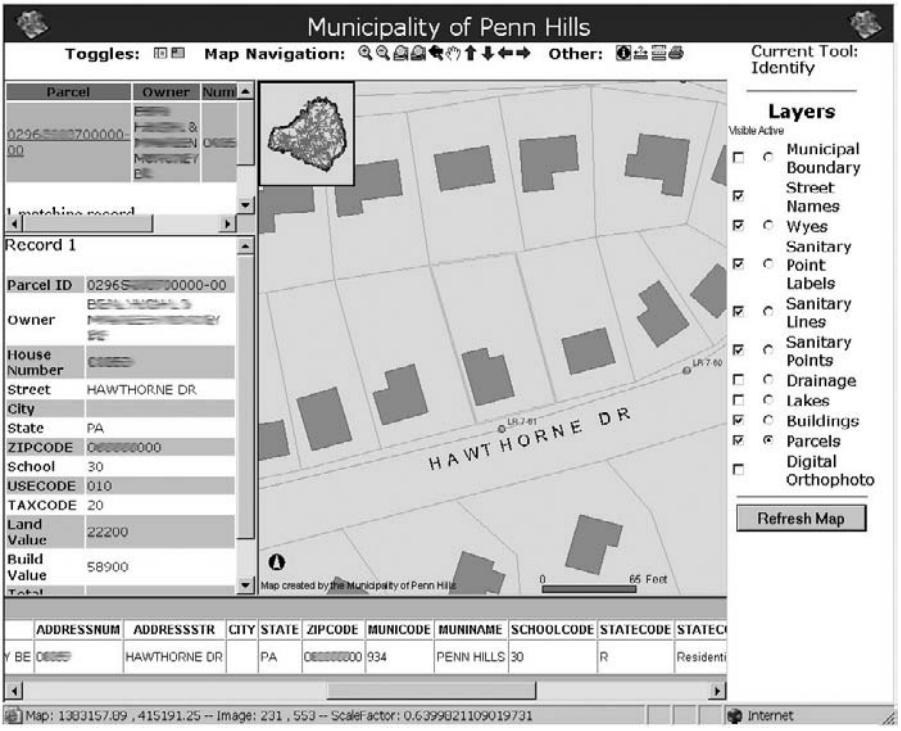


Figure 6.2 Screenshot of Penn Hills, Pennsylvania ArcIMS Web site.

relational database management system (RDBMS) spatial extensions, such as Oracle Spatial, facilitate linking data warehouses with GIS software packages. RDBMS is especially well suited for Internet and intranet data sharing. A data mining application can translate Internet requests into RDBMS queries and send RDBMS responses back to the Internet as Web pages. Utilities that rely on large and complex spatial data sets could greatly benefit from the recent fusion of object RDBMS and Internet technologies (Lowe, 2000a).

Internet Security

Security issues and concerns should be evaluated when posting GIS maps on the Internet. Precautions need to be taken to assure application security and data integrity (Miller et al., 2004). Appropriate security measures such as firewalls, data backups, user authentication, and password access can be used to reduce the chances of security breaches by hackers to harm the GIS data or corrupt the application software. Sensitive information that could help terrorists should not be posted or should be restricted to authorized users only. According to a Rand Corporation (www.rand.org) study released in April 2004, less than 1% of publicly available federal Web sites and databases in the U.S. contain geospatial information that could

help terrorists and other hostile forces mount attacks in the U.S. (Geospatial Solutions, 2004). In the wake of the September 11, 2001, terrorist attacks in the U.S., many Internet GIS sites were removed. Others began restricting public access to some of their sensitive information through the Internet.

INTERNET GIS SOFTWARE

Internet GIS uses a client/server architecture. Therefore two types of software are required to Web-enable a GIS: client-side and server-side (Shamsi, 2003). Client-side software is very easy to use but does not provide data creation or editing capability. It is usually available for free. Examples include:

- ArcExplorer, ESRI
- GeoMedia Viewer, Intergraph
- MapGuide Viewer, Autodesk

Server-side software, accessible through browsers, makes GIS usable over the Web. Based on the capabilities, the cost ranges from $5,000 to $25,000. Examples include:

- ArcIMS, ESRI
- GeoMedia Web Map, Intergraph
- MapGuide, Autodesk
- MapXtreme, MapInfo

INTERNET GIS APPLICATIONS

Representative applications of Internet GIS technology are described in the following subsections.

Data Integration

Imagine an organization with watershed data in ESRI Shapefiles, parcel data in an Oracle Universal Data Server with Spatial Cartridge, and water main data in an AutoCAD file. Imagine further that different data are stored on different network drives in different projections. Software is now available to integrate different data formats and convert them into a common projection system on the fly. Intergraph's GeoMedia Web Map is an example of such a software (Lowe, 2000).

Project Management

A project Web site promotes collaboration and efficient completion of utility construction or rehabilitation projects by providing all involved parties instant access to current information and allowing them to disseminate information automatically and immediately. A Web site also eliminates software and hardware compatibility problems among project team organizations. Redline (edit) capability can be added to a project Web site to expedite the design review process (Irrinki, 2000).

These possibilities are the result of a new generation of software from the leading GIS and CAD software companies. For example, in addition to the traditional CAD work, Autodesk's AutoCAD 2000i enables users to collaborate with design professionals anywhere in the world and move the designs and processes to their companies' intranets and the Web. Similarly, Intergraph's GeoMedia, in conjunction with Oracle Spatial, enables all the users to see the latest engineering project information against the base-map images. The various projects are represented as polygons overlayed on the base-map image. A user can simply click on one of the project polygons to obtain the most recent information about that project. After a project has been completed, the engineering information (including redlines) is used to update the spatial data stored in Oracle Spatial. This approach was successfully implemented by the Public Works Department of the City of Minneapolis. The City used GeoMedia and an Oracle Spatial data warehouse to provide future users the most current information about the city's infrastructure (Murphy, 2000). Figure 6.3 shows a project portal designed using this approach for a wastewater improvement project for the Borough of California (Pennsylvania), created by Chester Engineers, Pittsburgh, Pennsylvania.

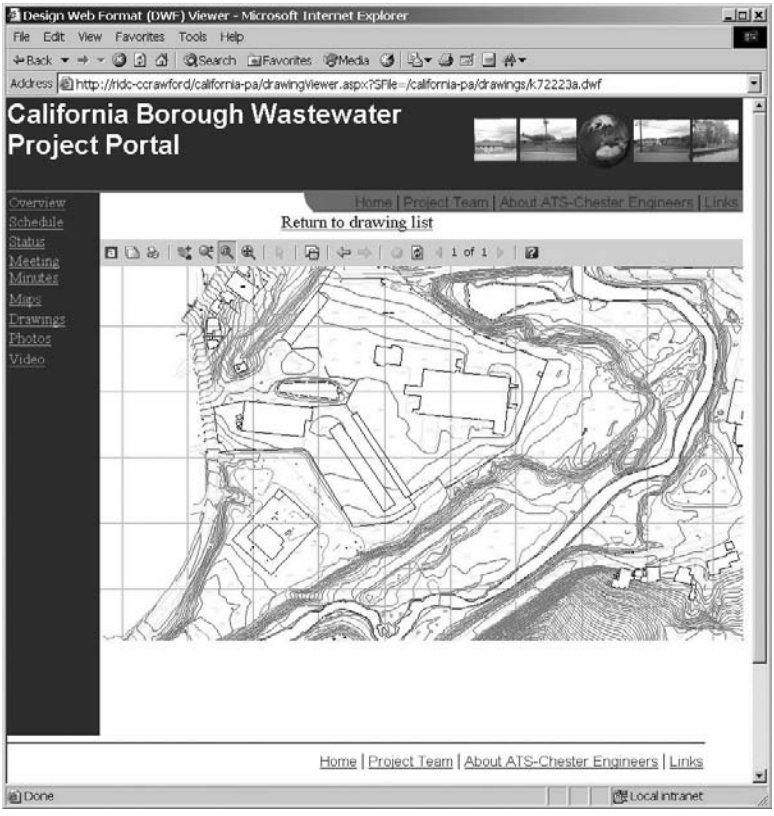


Figure 6.3 Screenshot of the Borough of California (Pennsylvania) project portal.

3D Visualization Applications

Internet GIS applications can also provide 3D visualization capability that converts 2D aerial photographs and satellite images into 3D terrain. The 3D terrain can be used as a base map for draping other information such as realistically textured buildings. For example, Skyline Software Systems' (www.skylinesoft.com) TerraExplorer provides realistic, interactive, photo-based, 3D maps of many locations and cities of the world on the Internet. Skyline's Internet technology enables creation, delivery, and interaction with content-rich, 3D geographic environments. Skyline enhances the Web experience for a variety of location-specific applications such as travel and tourism, real estate, entertainment, mapping, tracking, planning, and engineering. Standard CAD or GIS data layers such as DXF or Shapefiles can be easily incorporated to create dramatic 3D movies. Such Internet-based mapping applications are designed for a layperson and do not require GIS expertise.

CASE STUDIES

Two case studies illustrating the application of Internet GIS technology are described in the following subsections.

Tacoma's Intranet and Mobile GIS

Application	Intranet and Wireless GIS
GIS software	AutoCAD, AutoCAD Map, MapGuide, and GE Smallworld
Other software	Microsoft Access and SQL Server databases
Hardware	Laptop computers from Dell Computer Corp. and Gateway, Inc., 14.4K CDPD wireless PC card modems from Sierra Wireless
GIS data	Black and white orthophotos and planimetric base map layers from Bentley ADR, 3-ft and 6-ft color digital orthophotos from Triathlon, Inc.
Study area	Tacoma, Washington
Organization	Tacoma Public Utilities and Tacoma Water

Tacoma Water, a division of Tacoma Public Utilities (TPU) located in the state of Washington, serves 146,000 customers in the city of Tacoma's 52 mi^2 area and surrounding parts of Pierce and King counties. Tired of flipping through outdated map books and clumsy blueprints to locate water and sewer lines, Tacoma Water implemented intranet and wireless GIS applications.

Because of their familiarity with CAD, which was used in their engineering projects, Tacoma Water began a $750,000 multiyear project in 1990 to create CAD-based digital base maps and facility layers. The base map consisted of digital orthophotos and planimetric layers, such as building footprints, pavement edges, rights-of-way, lots and blocks, and easements, in a CAD format. Data conversion was completed in 1997 producing 6.5 GB of data with a horizontal accuracy of 6 in. Initially these data were distributed on CD-ROMs. In 1998, Tacoma Water

implemented an intranet GIS using Autodesk's intranet and Internet software suite consisting of AutoCAD, AutoCAD Map, and MapGuide for an initial investment of $12,000. Using these packages, they integrated their CAD data in a GIS and hosted it on a central server from where it was accessible to any computer using a standard Internet browser. The intranet GIS enabled the users to query the map features by simply placing the mouse cursor over that feature. By putting the cursor over a valve, for instance, the user will see a pop-up window showing the size, type, and number of turns required to open or close the valve.

According to Tacoma Public Utilities, "Having multiple layers of information available on mobile computing devices is like a holy grail to the water field crews."

In order to provide the available intranet GIS data to crews in the field, Tacoma Water implemented a wireless GIS in 1999. Consisting of off-the-shelf laptop computers and wireless modems, the truck-mounted wireless GIS was initially used by 18 field crews. The cost of unlimited wireless airtime was $50 per month. The mobility allowed by such a system puts field crews light years ahead of the time when finding a water main meant rummaging through a map book or reading microfiche (Teo, 2001).

Montana's Watershed Data Information Management System

Application	GIS-based watershed data management system
GIS software	ArcView
Organization	Montana Department of Environmental Quality

Montana Department of Environmental Quality (DEQ) is responsible for developing integrated water, air, waste management, and energy plans to protect Montana's environmental resources. To comply with the Montana Water Quality Act and the Federal Clean Water Act, an ArcView-based watershed data management system called MontanaView was developed in 2000. The system was designed to meet the specific information management needs of the state of Montana, focusing on water resource management issues. The project goals were: (1) integrating extensive and diverse data into one system, covering demographic, land-use and land-cover, infrastructure environmental, and natural resource information and (2) orienting programs and organizations around watersheds as the primary unit for assessing and managing resources.

These goals were accomplished by developing MontanaView, a PC-based application that runs under Windows operating systems, and linked to distributed databases over the Internet. The system provided online access to U.S. EPA mainframe databases and United States Geological Survey (USGS) real time and historical stream-flow data. For example, it provided access to EPA's national water quality database, STORET, to identify water quality monitoring stations that met user-defined requirements for selected parameters. It could then go to the Montana USGS Internet site (montana.usgs.gov) and import stream-flow data for the user-specified

station. The maps generated by MontanaView proved to be excellent visual aids during public meetings (Samuels et al., 2000).

USEFUL WEB SITES

Dublin, Ohio, Internet GIS Web site	gis.dublin.oh.us
Johnson County, Kansas, Internet GIS Web site	aims.jocogov.org
Town of Blacksburg, Virginia, Internet GIS Web site	arcims2.webgis.net/blacksburg/default.asp

CHAPTER SUMMARY

This chapter provided information on the GIS applications for serving water, wastewater, and stormwater system maps on the Internet. The Internet is the fastest growing and most efficient technology for distributing GIS data. Internet-based mapping applications are changing the way people use and manage geographic information in the water industry. The Internet can be used to both obtain and distribute GIS data. Two types of software are required for creating Internet GIS Web sites: client-side software and server-side software. Many cities and counties in the U.S. have developed Internet GIS sites for sharing their GIS data and maps.

CHAPTER QUESTIONS

1. What is Internet GIS?
2. What are the applications of Internet GIS technology?
3. What kind of software is required for developing Internet GIS applications?
4. Search the Internet to describe three Internet GIS Web sites that provide water, wastewater, and stormwater system maps on the Internet.

CHAPTER 7

Mobile GIS

Mobile GIS is helping the field workers by providing them with GIS maps on their hand-held computing devices. Field workers equipped with mobile GIS can update their enterprise GIS database from the field.

Mobile GIS brings GIS maps to field workers using mobile computers and PDAs.

LEARNING OBJECTIVE

The learning objective of this chapter is to determine the applications of mobile GIS technology in water industry.

MAJOR TOPICS

- Mobile GIS basics
- Mobile GIS applications
- Wireless Internet technology
- GPS integration

LIST OF CHAPTER ACRONYMS

LBS Location Based Services
PDA Personal Digital Assistant
RTK Real Time Kinematic

MOBILE GIS BASICS

Integration of GIS with consumer electronics such as cell phones and automobile navigation systems has spawned a whole new industry called location-based services (LBS). LBS services are used for managing dispatch and routing, fleets, work orders, and field crews — all of which improve customer service. Telegeoinformatics is also a new discipline resulting from the integration of mobile computing with wired and wireless communications, GIS, GPS, and remote sensing technologies (Karimi and Hammad, 2004). GIS lies at the heart of most mobile information resource systems. For example, it is now possible to wirelessly transmit a GPS location from the field, which can then be viewed on a GIS map via the Internet.

Using GIS directly in the field for data collection into a GIS database is referred to as mobile or field GIS. According to a recent study, about 40% of all mobile data applications in utilities are using a GIS (Wilson, 2001). With mobile GIS, feature location and attribute data can be verified, augmented, and immediately updated. New features can be added, existing features moved to their correct location, and nonexisting features deleted in the field. The ultimate goal of a mobile GIS is to link mobile workers with data to make their jobs easier and more efficient.

Mobile computers are the field devices used by mobile workers and field personnel to run GIS software (or its field version) and to collect location and attribute data in the field. Examples include notebooks, laptops, and pocket computers; tablet PCs and slate tablets; pen and touch-screen computers; handheld devices and PDAs; and smart phones (a combination of a PDA and a cell phone). Due to their small screen; limited memory, storage, processing speed, and battery life; and lack of a keyboard for field data entry, PDAs may not be suitable for large inspection projects. Currently, in the U.S., most PDAs use three operating systems: Linux, Palm OS, and Windows

Figure 7.1 Tablet PC being used for GPS and field data collection.

CE (Lowe, 2002). Basically a hybrid between a laptop and a PDA, the tablet PC combines the mobility and handwriting recognition of PDAs with the larger screen size and full-featured operating systems provided by laptops. Tablet PC screens can rotate 180° to create digital writing pads that lie flat against the keyboard. This feature allows field workers to sketch notes and diagrams or redline existing maps using a digital pen. Figure 7.1 shows a tablet PC connected to a mapping-grade GPS being used for field inspections. Figure 7.2 is a tablet PC screenshot showing notes and sketches taken in the field.

This field editing feature requires taking appropriate GIS layers and base maps in the field on a mobile computer. Custom forms can be created to mimic the paper forms that field crews are accustomed to using. The main advantage of a mobile GIS is that it eliminates postprocessing of paper data in the office.

Mobile GIS is the result of advances in four key areas: (1) field GIS software (e.g., ESRI's ArcPad); (2) lightweight, reliable, and affordable field computing devices; (3) more accurate and affordable GPS receivers; and (4) wireless communications (ESRI, 2004).

MOBILE GIS APPLICATIONS

Mobile GIS is helping cities in their field data collection and asset inventory activities. For example, many local governments are using ESRI's ArcPad field GIS software to improve speed and accuracy and lower the costs associated with the field inventory of their infrastructure assets (ESRI, 2004a). ArcPad, now a mobile client

Figure 7.2 Tablet PC screenshot showing field notes and sketches.

for ArcGIS, allows data collectors to collect asset locations and attributes in the field as ESRI-compatible GIS files. ArcPad Application Builder allows creation of custom field inspection forms using Visual Basic Scripting (VBScript). OnSite from Autodesk, InService from Intergraph, and MapXtreme from MapInfo are other examples of mobile GIS software.

Mobile GIS applications can benefit from wireless and Internet technologies for live import or export to office or enterprise data. Wireless connections allow real time access to office GIS data in the field and real time updates to office data. When a persistent network connection is not feasible, mobile GIS can transport the field data back to the parent software in an office for automatic validation and delivery to the target system at the end of each inspection day. Intergraph's IntelliWhere OnDemand vector-based mapping product for PDAs also runs on some smart phones. Among other features, this product allows workers to red-line maps on their mobile

devices and then wirelessly send the edits back to GIS servers in the office for immediate inclusion into the system. It also has an "occasionally connected" mode that allows field workers to extract data from the office GIS, view it, and disconnect until it is time to synchronize updates back to the office GIS (Lowe, 2002a).

WIRELESS INTERNET TECHNOLOGY

Wireless Internet solutions consist of four key technology components: (1) applications software running on the Internet and mobile devices; (2) mobile middleware software that resides on a server and a mobile device; (3) wireless data network to deliver data to and from mobile devices; and (4) mobile devices. In 2000, one of Southern California's largest wholesale water providers, the Calleguas Municipal Water District began a program to improve its infrastructure and operations management. It used Fieldport™ mobile operations management solution from Spacient Technologies, Inc. (Los Angeles, California). The system provides both Web-based and wireless asset management as either a traditional on-site software installation or as an Internet-hosted service. Fieldport allows GIS integration using its GIS/Mapping Module with an industry-standard database design powered by ESRI's ArcIMS software (Stern and Kendall, 2001).

GPS INTEGRATION

Mobile GIS applications typically use a GPS receiver; however, using a GPS is optional. Tiny GPS receivers are now available from various vendors that clip into most mobile devices and enable real time positioning on the map display. Without GPS integration, mobile GIS collects only attribute data; with a GPS receiver, data such as coordinates, elevation, and attributes can be collected. Inexpensive recreation-grade GPS receivers that provide typical 10-m horizontal accuracies can be used for navigation purpose. This arrangement would eliminate the need to pan and zoom around the map file to find the inspection location; the GPS connection can automatically center the map on the inspection location. Note that recreation-grade GPS receivers cannot and should not be used for collecting accurate horizontal and vertical positions. Figure 7.3 shows a sample GPS integration with a mobile device. The top part of the figure shows Hewlett Packard's iPAQ Pocket PC and Navman's recreation-grade GPS receiver for iPAQ. The GPS unit comes in the form of a sleeve that the iPAQ can slide into, as shown in the bottom part of the figure. According to the manufacturer, this GPS receiver typically provides better than 5 m (16 ft) horizontal accuracy for 95% of the time under optimal conditions.

Another technology that is helping mobile GIS is the Bluetooth wireless technology. Bluetooth is an international specification for a compact, low-cost radio solution that links mobile devices without using cables. Bluetooth devices can find each other when they are within their working range, which is typically 10 m (33 ft). For example, an RTK GPS rover on a pole can be wirelessly connected to the GPS receiver mounted inside a vehicle, or RTK GPS corrections can be received via a Bluetooth-enabled cell phone.

Figure 7.3 Sample mobile GIS device. Above: Navman GPS sleeve and iPAQ Pocket PC shown separately; below: iPAQ inserted into the sleeve.

USEFUL WEB SITES

Hewlett Packard's iPAQ Pocket PC	www.hp.com
Navman's recreation grade GPS for iPAQ	www.navman.com

CHAPTER SUMMARY

Mobile GIS plays an important role in simplifying the jobs of field workers and reducing the cost of field inspections and maintenance. For example, GIS allows users to identify valves to be closed for repairing or replacing broken water mains by simply clicking on a mobile field device such as a tablet PC.

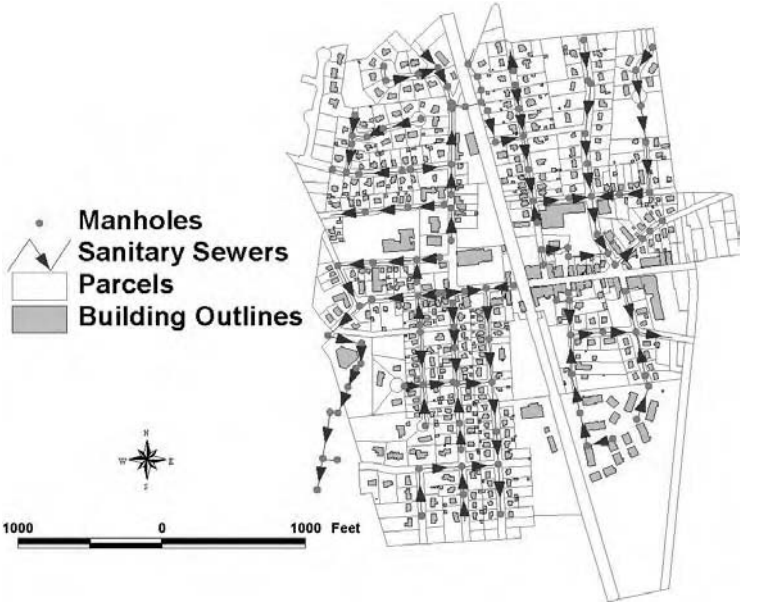
CHAPTER QUESTIONS

1. What is mobile GIS?
2. What is the main application of mobile GIS in the water industry?
3. How is mobile GIS related to Internet, GPS, and wireless technologies?

CHAPTER 8

Mapping

GIS provides powerful and cost-effective tools for creating intelligent maps for water, wastewater, and stormwater systems.



A sewer system map created by GIS (Borough of Ramsey, New Jersey).

LEARNING OBJECTIVE

The learning objective of this chapter is to understand how to create GIS maps for water, wastewater, and stormwater systems.

MAJOR TOPICS

- Mapping basics
- Map types
- Advantages of GIS maps
- GIS mapping steps
- Mapping case studies

LIST OF CHAPTER ACRONYMS

AM/FM Automated Mapping/Facilities Management
AML Arc Macro Language
DRG Digital Raster Graphics (USGS topographic maps)
NAD-27 North American Datum of 1927
NAD-83 North American Datum of 1983
QA/QC Quality Assurance/Quality Control
SPC State Plane Coordinate (Map Projection System)
TIGER Topologically Integrated Geographic Encoding and Referencing System (U.S. Census Bureau Mapping System)
UTM Universal Transverse Mercator (Map Projection System)
VBA Visual Basic for Applications

This book focuses on the four main applications of GIS, which are mapping, monitoring, maintenance, and modeling and are referred to as the "4M applications." In this chapter we will learn how to implement the first *m* (mapping).

LOS ANGELES COUNTY'S SEWER MAPPING PROGRAM

In the 1980s, the Sanitation Districts of Los Angeles County, California, envisioned a computerized maintenance management system that would provide decision makers with essential information about the condition of the collection system. A sewer and manhole database was subsequently developed, but investments in GIS technology were deferred until the early 1990s when desktop PCs became powerful enough to run sophisticated GIS applications. In 1993, a GIS needs analysis study was performed, which recommended implementation of a large-scale enterprise-wide GIS. An in-house effort was started to implement the recommendations of the study. Several sections in the Districts formed a project committee to pilot test GIS technology that could be duplicated in all 25 districts. The pilot project developed a mapping application for a small district that acted as a front-end to the large database of sewerage

facilities developed in the 1980s and 1990s. At this point, much of the information was nonspatial, including multiple databases in a variety of formats and paper maps. Converting this information into GIS proved to be the most time consuming and costly operation. Creation of the layers for sewers and manholes was the most laborious of all the layers that had to be created. Manhole data were represented by approximately 24,000 points or nodes digitized from the paper maps, using a base map. The GIS layers were created in CAD software and linked to the legacy databases. Once linked, detailed data such as sewer pipe and manhole construction material, size, condition, flow, capacity, and inspection data were available for query and analysis through an intuitive map-driven interface. By 2003, the pilot project had grown to become the first enterprise-wide solution deployed by the Districts. Called "Sewerage Facilities GIS," the system allowed users to access and view data from the legacy databases by selecting sewers and manholes on a map or using standard queries. The mapping application also provided sewer tracing functionality that proved helpful in delineating study area boundaries for design projects, annotating sewers with flow direction, and tracking potential discharge violations (Christian and Yoshida, 2003).

MAPPING BASICS

The basic concepts essential for understanding GIS mapping are summarized in the following subsections.

Map Types

There are two major types of GIS maps: vector and raster. In *vector* format, objects are represented as points, lines, and polygons. Examples of the vector format are maps of water mains, hydrants, and valves. Scanned maps, images, or aerial photographs are examples of *raster* format. Raster data are also referred to as grid, cell, or grid–cell data. In raster format, objects are represented as an image consisting of a regular grid of uniform size cells called *pixels*, each with an associated data value. Many complex spatial analyses, such as automatic land-use change detection, require raster maps. Raster maps are also commonly used as base maps (described later in this chapter). Existing paper maps that are used to create GIS maps are called *source maps.*

Topology

Topology is defined as a mathematical procedure for explicitly defining spatial relationships between features. Spatial relationships between connecting or adjacent features, such as a sewer tributary to an outfall or the pipes connected to a valve, which are so obvious to the human eye, must be explicitly defined to make the maps "intelligent." A topological GIS can determine conditions of adjacency (what is next to what), containment (what is enclosed by what), and proximity (how near something is to something else). Topological relationships allow spatial analysis functions, such as network tracing, that can be used to facilitate development of hydraulic models for water and sewer systems.

Map Projections and Coordinate Systems

Because the Earth is round and maps are flat, transferring locations from a curved surface to a flat surface requires some coordinate conversion. A *map projection* is a mathematical model that transforms (or projects) locations from the curved surface of the Earth onto a flat sheet or 2D surface in accordance with certain rules. Mercator, Robinson, and Azimuthal are some commonly used projection systems. Small-scale (1:24,000 to 1:250,000) GIS data intended for use at the state or national level are projected using a projection system appropriate for large areas, such as the Universal Transverse Mercator (UTM) projection. The UTM system divides the globe into 60 zones, each spanning 6° of longitude. The origin of each zone is the equator and its central meridian. X and Y coordinates are stored in meters. Large-scale local GIS data are usually projected using a State Plane Coordinate (SPC) projection in the United States.

A *datum* is a set of parameters defining a coordinate system and a set of control points with geometric properties known either through measurement or calculation. Every datum is based on a spheroid that approximates the shape of Earth. The North American Datum of 1927 (NAD27) uses the Clarke spheroid of 1866 to represent the shape of the Earth. Many technological advances, such as the global positioning system (GPS), revealed problems in NAD27, and the North American Datum of 1983 (NAD83) was created to correct those deficiencies. NAD83 is based on the GRS80 spheroid, whose origin is located at the Earth's center of mass. The NAD27 and NAD83 datum control points can be up to 500 ft apart.

Coordinates are used to represent locations on the Earth's surface relative to other locations. A *coordinate system* is a reference system used to measure horizontal and vertical distances on a map. A coordinate system is usually defined by a map projection. The GIS and mapping industries use either latitude/longitude- or geodetic-based coordinate grid projections. Because much of the information in a GIS comes from existing maps, a GIS must transform the information gathered from sources with different projections to produce a common projection.

Map Scale

Map design addresses two fundamental map characteristics: accuracy and depicted feature types. Both characteristics vary with map scale. Generally, larger scale maps are more accurate and depict more detailed feature types. Smaller scale maps, such as U.S. Geographical Survey (USGS) quadrangle maps, generally show only selected or generalized features. Table 8.1 summarizes the relationships among map scale, accuracy, and feature detail.

Data Quality

The famous computer industry proverb "garbage in, garbage out" conveys very well the importance of GIS data quality. A GIS map is only as good as the data used to create it. Data quality roughly means how good the data are for a given application. Data quality is important because it determines the maximum potential reliability of the GIS application results. Use of inappropriate data in a GIS map may lead to

Table 8.1 Relationships among Map Scale, Accuracy, and Feature Detail

Map Scale	Minimum Horizontal Accuracy, per National Map Accuracy Standards	Examples of Smallest Features Depicted
1 in. = 50 ft	±1.25 ft	Manholes, catch basins
1 in. = 100 ft	±2.50 ft	Utility poles, fence lines
1 in. = 200 ft	±5.00 ft	Buildings, edge of pavement
1 in. = 2000 ft	±40.00 ft	Transportation, developed areas, watersheds

misleading results and erroneous decisions, which may erode public confidence or create liability.

Data Errors

There are two types of data errors: inherent errors embedded in the source of data and operational errors introduced by users during data input, storage, analysis, and output. Inherent errors can be avoided by using the right kind of data. Operational errors can be prevented by quality control and training.

A data conversion team should be aware of sources and magnitudes of data error. For example, spatial information in USGS 1:24,000-scale (7.5-min) topographic maps is certified to have 90% of its features within 50 ft (15 m) of their correct location. 50 ft is large enough to underestimate the runoff from a new development and undersize a detention pond for adequate stormwater management.

Map Accuracy

A primary factor in the cost of data conversion is the level of positional accuracy. Required map accuracy and resolution depend on the application in which the maps will be used. A 2000 survey conducted by the Geospatial Information and Technology Association (GITA) indicated that the water utilities were seeking more landbase accuracy of 5-ft compared with other utilities, such as the 50-ft accuracy sought by the gas companies (Engelhardt, 2001; GITA, 2001). The same survey for 2002 indicated that the water industry required the highest accuracy in their GIS projects. Among the water organizations, 27% were using a 6-in. landbase accuracy compared to electric (12%), gas (17%), pipeline (17%), and telecom (0%) organizations (GITA, 2003). These data reveal that a trend toward increasing accuracy may be emerging in the water industry.

Engineering applications usually require ±1 to 2 ft accuracy. For planning and regional analysis applications, ±5 to 10 ft accuracy is generally appropriate (Cannistra, 1999). Sometimes relative accuracy (e.g., ±1 ft from the right-of-way line) is more important than an absolute level of accuracy (e.g., ±1 ft from the correct location). For the applications where positional accuracy is less important, supposedly low-resolution data, such as USGS digital orthophoto quadrangles (DOQs), may be acceptable. In other applications where features must be positioned within a foot of their actual position, even the presumably high-resolution data, such as 1-m IKONOS imagery,

may not be accurate enough. As a rule of thumb, a database built from a map will have positional inaccuracies of about 0.5 mm at the scale of the map because this is the typical line width of the drawing instrument. This can cause inaccuracies of up to 12 m in a database built from 1:24,000 mapping, such as USGS DRGs (Goodchild, 1998).

Precision and accuracy are two entirely different measures of data quality and should not be confused. A GIS can determine the location of a point feature precisely as coordinates with several significant decimal places. However, many decimal places in coordinates do not necessarily mean that the feature location is accurate to a 100th or 1000th of a distance unit. Once map data are converted into a GIS environment, the data are no longer scaled, as the data can be scaled as desired to create any output map scale. However, the spatial data can never be any more accurate than the original source from which the data were acquired. GIS data are typically less accurate than the source, depending on the method of data conversion. Therefore, if data were captured from a source map scale of 1 in. = 2000 ft, and a map was created at 1 in. = 100 ft, the map accuracy of features shown would still be 1 in. = 2000 ft (PaMAGIC, 2001).

MAP TYPES

Various map types used in GIS are discussed in the following subsections.

Base Map

The map layers are registered to a coordinate system geodetic control framework and a set of base information, often referred to as a base map. The foundation for a successful GIS mapping project is an appropriately designed base map. The base map is the underlying common geographic reference for all other map layers. The common reference provides registration between various layers and allows them to be overlayed, analyzed, and plotted together. Because the base map serves as the reference layer for other layers, its accuracy can affect the accuracy of other layers. This is especially true if the base map is used to create other layers by on-screen (heads-up) digitization.

Selection of an appropriate base-map scale is largely determined by the earlier choice of GIS applications. Each application inherently requires a certain minimum base-map accuracy and certain map features. For engineering and public-works applications, the required map accuracy is in the range of ±1 ft, as dictated by the need to accurately locate specific physical features, such as manholes and catch basins. Planning applications, which most often deal with areawide themes, do not generally require precise positioning. Accuracies of ±5 ft, or perhaps as much as ±40 ft, are often acceptable. Less detailed maps, showing nothing smaller than roads and buildings, for example, may be adequate for many planning applications.

Whatever the range of mapping requirements, the base map must be accurate and detailed enough to support applications with the most demanding map accuracies of better than ± 2 ft. Utility asset location also requires mapping that depicts specific small features such as manholes and catch basins. As shown in Table 8.1, these requirements are met by a map scale of 1 in. = 50 ft.

There are three common types of base maps: digital orthophotos, planimetric maps, and small-scale maps.

Digital Orthophotos

For laypersons, digital orthophotos (or orthophotographs) are scanned aerial photos. For GIS professionals, they are orthorectified raster images of diapositive transparencies of aerial photographs. Creation of a digital orthophoto requires more than a photo and a scanner, and includes surveyed ground control points, stereo plotters, and a digital elevation model. In fact, the digital orthophoto creation process involves many steps, which are listed below:

- Establish ground control
- Conduct aerial photography
- Perform analytical aerotriangulation
- Set stereo models in stereo plotters
- Capture digital elevation models
- Scan aerial photographs
- Digitally rectify the scanned photographs to an orthographic projection
- Produce digital orthophotos

Digital orthophotos are popularly used as base maps that lie beneath other GIS layers and provide real-life perspectives of terrain and surroundings that are not available in the vector GIS layers. Typical vector data do not show vegetation. The vector layers can show the manhole location but may not include the vegetation hiding the manhole. High-resolution orthophotos with submeter accuracy can guide the public-works crews directly to a manhole hidden behind bushes. Knowing the land-cover characteristics before leaving for an emergency repair of a broken water main will allow the crews to bring the appropriate tools and equipment. Knowing whether the job will be on a busy intersection or in somebody's backyard will determine the kind of equipment, material, and personnel required for the job. Figure 8.1 shows a water system map overlayed on a digital orthophoto base map with an accuracy of ±1.25 ft. Typical digital orthophotos cost $800 to $1600 per mi^2.

Planimetric Maps

Like digital orthophotos, planimetric base maps are also created from aerial photographs. However, instead of scanning the aerial photos, the features are digitized from them. Thus, whereas digital orthophotographs are raster files, planimetric maps are vector files. Planimetric maps generally show building footprints, pavement edges, railroads, and hydrography. Parcels digitized from existing maps are often added to the mix. Figure 8.2 shows a sample planimetric map for the Borough of Munhall, Pennsylvania, extracted from the Allegheny County land base. The borough's sewer lines and manholes are overlayed on the planimetric base map. The cost of planimetric maps depends on the level of detail and, therefore, varies significantly from project to project. The typical cost range is $2,500 to $10,000 per mi^2.

Figure 8.1 A water distribution system overlayed on a digital orthophoto base map.

Small-Scale Maps

Small or rural systems often use small-scale street maps or topographic maps as base maps. Street maps can be created by digitizing the existing maps, obtained from a government agency (e.g., U.S. Census Bureau, USGS, or state department of transportation) or purchased from commercial data vendors. In the U.S., 1:24,000-scale raster topographic map layers called digital raster graphics (DRG) are provided by USGS. Shamsi (2002) provides detailed information about the sources of small-scale maps. Users should be aware of the scale, resolution, accuracy, quality, and intended use of small-scale base maps before using them. Most maps at scales of 1:100,000 and smaller are not detailed enough to be used as site maps or engineering drawings, but they can be used for preliminary studies and planning projects. Figure 8.3 shows interceptor sewers and pumping stations for the Kiski Valley Water Pollution Control Authority in Pennsylvania, on a base map

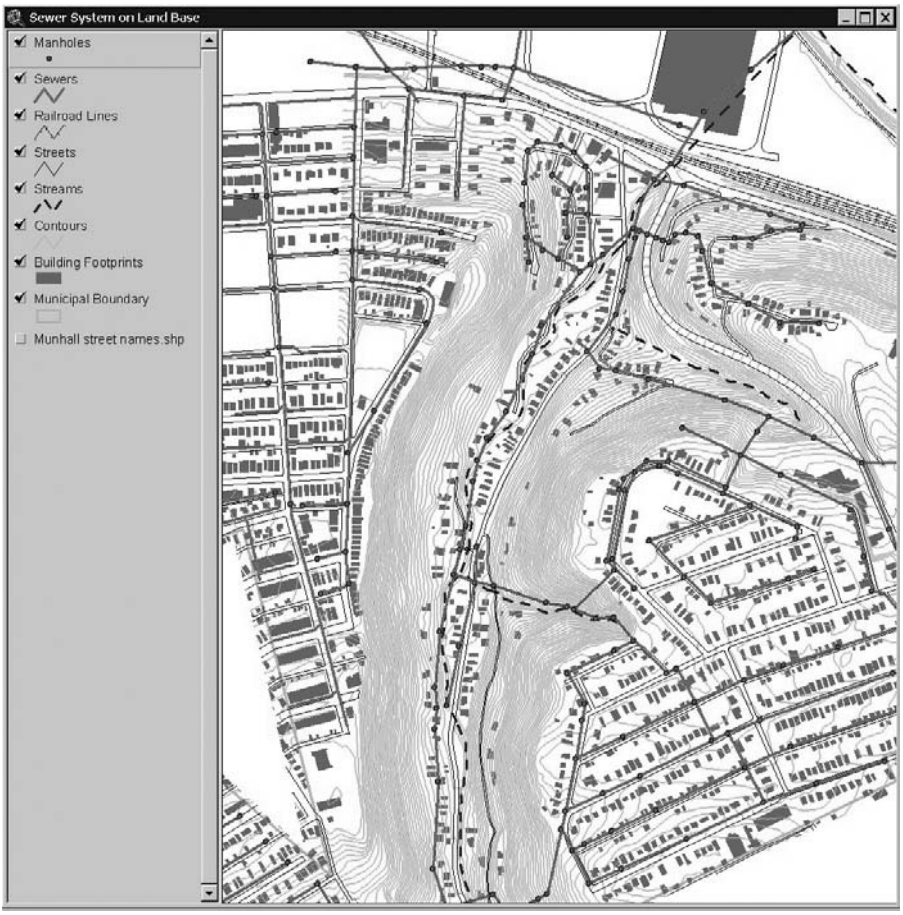


Figure 8.2 A sewer system overlayed on a planimetric base map.

of streets. The 1:100,000-scale solid roads are from the U.S. Census Bureau's 1990 Topologically Integrated Geographic Encoding and Referencing System (TIGER) data. The 1:24,000-scale dashed roads are from the Pennsylvania Department of Transportation. Unlike double-line pavement edges shown on the planimetric maps, these road layers show the single-line street center lines. The difference in the position of the roads in the two layers can be attributed to the resolution, scale, and accuracy of the two layers.

ADVANTAGES OF GIS MAPS

The most challenging part of a GIS application project is to obtain the right kind of maps in the right format at the right time. Therefore, maps are the most important component of a GIS. Without maps, you simply have a computer program, not a GIS.

Figure 8.3 A sewer system overlayed on a streets base map.

In many water and wastewater systems, there is a backlog of revisions that are not shown on the maps and the critical information is recorded only in the memories of employees. However, there is no longer any excuse to procrastinate because GIS-based mapping is easy and affordable.

In the past, users have selected computer-aided drafting (CAD) and automated mapping and facilities management (AM/FM) systems to map their water and sewer systems, CAD being the most common method. Although a map printed from CAD or AM/FM might look like a GIS map on paper, it does not have the intelligence of a GIS map. GIS maps are intelligent because they have attributes and topology. Most conventional CAD maps do not have attributes; they simply print data as labels or annotations. For example, the mapmaker must manually write the pipe diameter next to a pipe or must manually change the pipe color or line type to create a legend for pipe size. This is a very cumbersome process. On the other hand, GIS stores the attributes in a database and links them to each feature on the map. This capability allows automatic creation of labels and legends at the click of a mouse button. In GIS, map labels and legends are changed automatically if an attribute changes. In CAD, the mapmaker must manually delete the old label and retype the new label. Only a GIS map knows the spatial relationships among its features. Called *topology*, this capability makes the GIS maps intelligent. For example, a GIS map is intelligent enough to know which watershed is adjacent to which. Although both the CAD and AM/FM offer map layers to store different types of objects, only a GIS map has the capability to relate data across layers. The spatial relations among layers allow spatial

analysis operations such as identifying the gate valves that must be closed to isolate a broken water main for repair.

A commercial map atlas company may use a CAD system because its applications are primarily for cartographic products. A telephone company will use an AM/FM system to support its telephone system operations and maintenance, because it must be able to quickly trace a cable network and retrieve its attributes (Korte, 1994). For a water or sewer system, a GIS map is most suitable because the system must conduct many types of spatial analyses, asking questions such as how many customers by type (residential, commercial, industrial) are located within 1000 ft of a proposed water or sewer line.

In addition to water and sewer system management, GIS-based maps can also support other needs of the municipality. For example, a planning department can generate 200-ft notification lists as part of its plan review process. A public works department can conduct maintenance tracking and scheduling. A public-safety department can perform crime location analysis. GIS-based mapping is therefore the most appropriate mapping technology to meet all the mapping needs of a municipality.

GIS MAPPING STEPS

GIS mapping consists of five typical steps:

1. Needs analysis
2. Data collection
3. Data conversion
4. Data processing
5. Map production

These steps are intended for GIS technicians who work in a GIS lab (or production shop) equipped with map-making equipment such as digitizers, scanners, and plotters. Some professionals, such as civil and environmental engineers who are users rather than creators of GIS maps, generally do not perform all of these steps.

Needs Analysis

Mapping work should begin with needs analysis as described in Chapter 2. The needs analysis study describes the features that should be captured during the data conversion step, mapping specifications (accuracy, resolution, scale, etc.), and source documents. For example, the needs analysis determines whether or not the customer meters should be mapped. If the answer is yes, then how and to what accuracy should they be mapped. Are precise coordinates needed or can they be drawn at the end of service line? For large systems (populations greater than 50,000), a pilot project should be conducted as described in Chapter 2 (Needs Analysis). The pilot project allows potential fine-tuning of mapping specifications in a two- to four-sheet area before starting the map production phase.

Data Collection

When existing maps do not exist or are inadequate, mapping data should be collected using a field survey with or without GPS. Required data are usually scattered among a multitude of different organizations and agencies. For instance, it is claimed that 800 worker-years would be needed to convert England's sewer system to digital format, including field verification work (Bernhardsen, 1999). Field data collection using mobile GIS and GPS technology that employs handheld devices and tablet PCs is becoming common for collecting attributes (Chapter 7, Mobile GIS). The question of quantity should be evaluated carefully before the data conversion is started. Too little data limits the mapping applications. Too much data typical of a "data-driven" approach might be wasteful.

Data Conversion

Also referred to as the production work, this step converts hard-copy maps into digital files using digitization with or without scanning. The data conversion process is typically the most expensive and time-consuming part of a GIS mapping project. Depending on the level of detail, data conversion can cost $500 to $5000 per map sheet. Approximately 75% of typical GIS costs are related to data conversion and creation. The data conversion work, therefore, should be structured to ensure consistency, integrity, and accuracy.

Captured data are stored in *layers* also referred to as coverages or themes. For example, manholes are captured as a point layer, sewer pipes are captured as a line layer, and sewersheds are captured as a polygon layer. The layers that support development of other layers should be given a higher development priority. For example, geodetic control and base-map layers should be developed first because most other layers depend on them.

Data conversion includes capturing both the graphic (geometry and coordinates of features) and nongraphic (attributes) data. Graphic and nongraphic data may be entered simultaneously or separately.

Capturing Attributes

The annotations (labels) shown on the source map are the most common source of attributes. The source document ID (e.g., drawing number) is one of the most important attributes that should always be captured during data conversion. Features such as valves and hydrants have unique IDs, which can be easily captured from labels (annotations) on source maps. The features without IDs should be assigned new IDs during data conversion. Some data conversion application software allow automatic ID creation during data conversion.

Most source maps show pipe sizes as labels or legends, which can also be easily captured during data conversion. Geometric properties like length, area, and perimeter of features can be internally calculated by GIS and entered as an attribute. For example, if pipe length is labeled, it can be entered as a Length_Map (source map)

attribute. If pipe length is not shown, it can be automatically calculated by GIS and entered as Length_GIS (GIS-calculated) attribute.

Additional attributes are collected from other sources such as legacy databases, tables, and field and GPS surveys. Some existing databases and tables can be linked to GIS databases to avoid manual entry of attributes. Once the linkage has been established, attributes from the linked documents are available in the GIS database for query, analysis, and thematic mapping. For example, unique hydrant IDs captured during data conversion can be used to link to a hydrant-testing database. Applications that require locating the facilities by customer address such as work orders and customer complaints should also capture the address attribute in at least one layer (buildings, parcels, or meters). GIS attributes can also be found in some unexpected places, such as customer billing records. With a technique called address geocoding, GIS can convert the billing records or any postal address to a point on a map. Alternatively, facilities can be located by customer address using a geocoded streets layer.

Capturing Graphics

The data conversion methodology for capturing graphics depends on database design, source materials, and project budget. Digitization and scanning are the two most common data conversion methods.

Digitization

Digitization is a process of converting a paper map into a vector file by a computer using a digitizer or digitizing table (or tablet). The source map is placed on the usually backlit surface of the digitizer, and map features are traced using the digitizer puck that looks like a computer mouse with a crosshair. The source map should be registered (or calibrated) to control points. Figure 8.4 shows map conversion work using a digitizer. Digitizers cost approximately $10,000. Digitized files, when properly prepared, are ready for immediate use in a GIS.

Conventional table digitization is a laborious process. New "heads-up" or "on-screen" digitization is a less cumbersome alternative in which the visible features from scanned maps, digital aerial photos, or satellite imagery can be traced with a standard computer mouse on the computer monitor. If a digitizer is not available but a scanner is, data conversion can be done by scanning the source maps followed by heads-up digitization. The use of this process for certain specialized mapping projects, such as land-use/land-cover mapping, requires familiarity with photo interpretation techniques. For easy identification, certain utility assets that are difficult to see on the aerial photographs (e.g., valves, hydrants, manholes) can be premarked. Premarking is done by placing (or painting) targets with special symbols and colors over or adjacent to the asset to be captured. Premarking cost is usually $2 to $10 per target.

As-built or construction drawings usually have dimensional or offset information on utility assets (e.g., a sewer pipe 10 ft from the street right-of-way). This information can be used to position the utility assets using the automated drafting capabilities of modern GIS packages. This method is more accurate than digitization (Cannistra, 1999).

Figure 8.4 Data conversion using digitization.

Scanning

Scanning is a process of converting a paper map into a raster file (or image) by a computer using a map-size scanner. Black-and-white scanners are most common because they cost about the same as a digitizer. Color scanners cost twice as much. Figure 8.5 shows map-conversion work using a scanner. Scanning is generally most efficient when converting maps for archival purposes. Sometimes, as-built drawings are scanned and linked to vector features. Because many applications (e.g., hydraulic modeling and field inspections) require the utility features in vector format, scanned images must be converted to a vector format. This can be done by heads-up digitization of the scanned maps or by using a process called *vectorization* that automatically converts the raster data to vector format. Vectorization is a cost-effective data conversion method but it is based on new technology that is still improving. Vectorization may not always produce reliable results, especially in the unsupervised (unmanned) mode. Linework and symbols vectorize better than annotations. Complex, faded, or unclear maps may require extensive postprocessing and manual touch-up.

Data Conversion Software

Data conversion requires GIS development software that draws objects as points, lines, or polygons or represents them as pixels. ESRI's ArcGIS® and Intergraph's GeoMedia Pro® are examples of GIS development software (Shamsi, 2002). For large projects, data conversion application programs are very useful. These programs

Figure 8.5 Data conversion using scanning.

are usually developed as extensions to standard GIS packages to supplement specific mapping needs. For example, ESRI's ArcFM water data model supports water and sewer system mapping in a geodatabase. Chester Engineers' (Pittsburgh, Pennsylvania) water editing tool is an application that allows user-friendly mapping of water and sewer systems for ESRI software users. It eliminates attribute entry and editing via command prompts or database tables, which require data conversion skills. Users without prior data conversion experience can enter attributes from sources documents (e.g., as-built drawings) in a menu environment simply by clicking on points (e.g., manholes) and lines (e.g., sewers). The water editing tool was developed in several

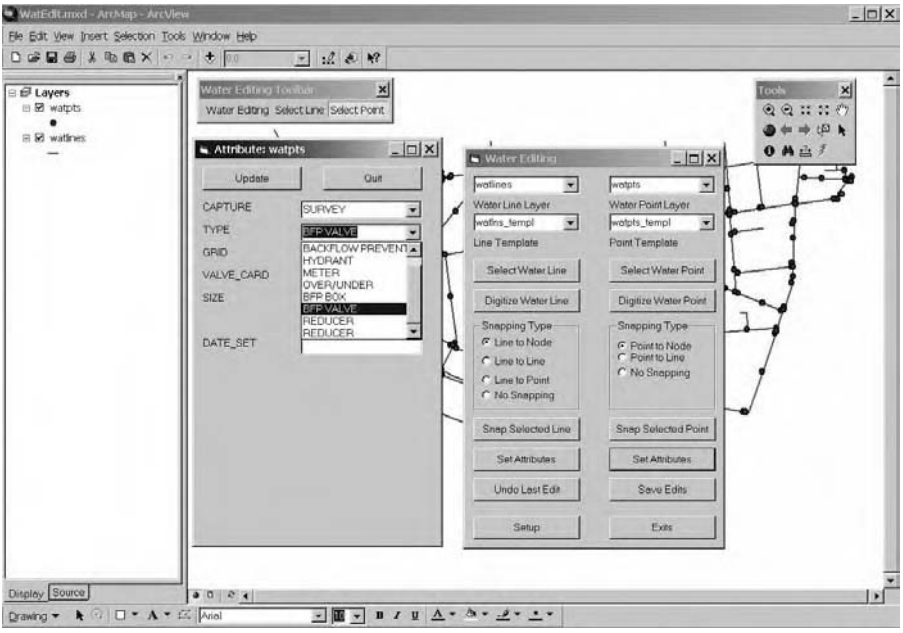


Figure 8.6 Sample water system data conversion application for ArcMap.

versions for different users. The ArcView version was developed using Avenue and ESRI's Dialog Designer Extension to support water system and sewer system mapping needs of the ArcView 3.x users. The ArcInfo 7.x version was written in Arc Macro Language (AML) to support water system and sewer system mapping in an Arcedit session. The latest ArcMap version that adds a new water editing toolbar in ArcMap was developed using Visual Basic for Applications (VBA) for ArcGIS users. Figure 8.6 shows a screenshot of the ArcMap version of the water editing tool. The right window shows the form for adding and editing point and line features. The left window shows the form for editing point attributes.

Data conversion is started by drawing and attributing the points. After the points are attributed, the connecting lines are added from point to point. The application prompts the user to select the upstream and the downstream points. After both are selected, the line is added to the data set. By selecting upstream and downstream manholes to add the lines, the flow direction of a gravity sewer is incorporated into the data set, ensuring the proper connectivity of the system. Additionally, the lines are snapped to the point features, ensuring that the points of connection correspond to manholes.

Data conversion applications like the water editing tool increase the speed, efficiency, and accuracy of data conversion for water, wastewater, and stormwater systems. They reduce data-entry errors and can even be used to validate the input data. If desired, the applications can be modified to meet the project-specific mapping requirements.

Data Processing

Data processing includes some or all of the following tasks:

1. Data preparation
2. Topological structuring
3. Data management
4. Quality control

Data Preparation

This task makes the raw data from source maps or field/GPS surveys available for GIS use or GIS-ready. Typical activities include postprocessing of capture data, changing data formats (e.g., from DXF to Shapefile), applying map projections, georeferencing the image data, and/or clipping or mosaicking the aerial photos.

Postprocessing of captured data is warranted if the source maps and the base map have different accuracy, and captured data do not align with the base map. In this case, captured data is edited (moved, stretched, or resized) to fit the base map.

GIS data are stored in various file formats. The number of data formats has increased exponentially with the growth in the GIS industry (Goodchild, 2002). According to some estimates, there might be more than 80 proprietary geographic data formats (Lowe, 2002b). Why are there so many geographic data formats? One reason is that a single format is not appropriate for all applications. For example, a single format cannot support both fast rendering in a command and control system and sophisticated hydraulic analysis in a water distribution system. Different data formats have evolved in response to diverse user requirements.

GIS software cannot read all the data formats simply because there are so many formats. Disparate data formats should be converted to one of the formats compatible with a particular GIS software. Although many GIS packages provide data conversion for the most common data formats, no GIS can support all possible conversions. Many government agencies and data and software companies provide data translators for this purpose (Shamsi, 2002).

Data translation from one format to another can potentially lead to the loss or alteration of data. For instance some platforms can support better numeric precision than others. Depending on the mapping scale and the coordinate system, this could seriously affect data quality. In one system, curves representing a water main might be defined mathematically, but in another described as a series of straight line segments.

Topological Structuring

How much topology should be captured in a GIS map depends on the intended applications of the map. For example, in a sewer system modeling application, correct topology (e.g., flow direction) is generally more important than the correct pipe location on the street (e.g., left, right or center of the street). As much topology as possible should be captured during data conversion. For example, flow direction

Figure 8.7 Common digitizing errors: undershoot (left) and overshoot (right).

should be captured if flow direction arrows are shown on source maps. Flow direction can be automatically captured if sewer pipes are digitized from upstream to downstream direction. In ArcInfo 7.x "coverage" model, the topology of an arc includes its upstream and downstream nodes (fnode, tnode). ArcInfo 7.x performs computationally intensive topological relationship calculations up front and stores that information as attributes (fnode, tnode). These attributes are used to determine flow direction of arcs. If source maps do not have topological relationships or if the data conversion technicians are not adequately skilled to understand them, topology can be added after the data conversion by knowledgeable staff. For example, in a water distribution system mapping project, valves and hydrants are logically linked to their appropriate pipe segments by the people who are knowledgeable about system characteristics (Cannistra, 1999).

Conventional GIS packages, such as ArcInfo 7.x and ArcCAD, provide the legacy *BUILD* and *CLEAN* commands to build topology and eliminate topological digitizing errors such as undershoots and overshoots. As shown in Figure 8.7, an undershoot occurs when a line stops short of an intended intersection point and an overshoot occurs when a line extends past an intended intersection point. Traditionally, these "cleaning" operations were done after the data conversion step. Today's modern GIS packages allow enforcing the topological rules (spatial constraints) during data conversion and editing. Thus, when a manhole is moved, the topological rules ensure that the sewers connected to that manhole move simultaneously. For example, ArcInfo 8.3 provides new tools to define, validate, and maintain topology in a geodatabase. Users define which layers participate in a topology and which rules are appropriate. Topology is stored in a commercial off-the-shelf DBMS.

Data Management

This step organizes attribute data in a database management system (DBMS) format. A GIS database, though not as visible as a GIS map, is an equally critical GIS component. The most difficult job in creating a GIS is the enormous effort required to enter the large amount of data and to ensure both its accuracy and proper maintenance (Walski and Male, 2000). Conventional file-based GIS databases (e.g., ArcInfo 7.x) store geographic data in vendor-specific proprietary format files. Attribute

data are stored in a separate file, which generally has a common nonproprietary format (e.g., a .dbf file). The two separate files are linked using pointers. Modern spatial or object-oriented databases do not require links between features and attributes because they store both the features and attributes in the same file. Incorporating spatial data with attribute data in a relational database provides a platform for efficient management of huge data resources, as well as fast and sophisticated spatial query and retrieval methods (Dalton, 2001).

Quality Control

A comprehensive and well-defined quality assurance/quality control (QA/QC) program should be implemented for large data conversion projects. Data conversion QA/QC can be ensured by checking data completeness, content, symbology, format, structure, and compliance with database specifications during data production. Positional accuracy can be checked using field trips, survey, GPS, photographs, and video tapes (Cannistra, 1999).

QA/AC can employ both manual checks and automatic methods. For example, the water editing application shown in Figure 8.6 provides an automatic QA/QC capability. It displays the points that are completely defined using a white box. Points that have missing attributes are displayed as a box with an X through it. Points that are added from source drawings are shown as a green box with an X through it. After each point is defined, the list of attributes for each point is displayed on the screen, allowing the user to double-check the attributes entered for each point.

Data-entry QA/QC can be accomplished by restricting the user input to valid data ranges and rejecting the data entries outside the valid ranges. This capability generally requires development of custom data-entry forms using computer programming.

Map Production

The previous steps enable display of GIS layers on the computer screen. The map production step allows plotting of GIS data on paper and other print media. It includes creation of thematic maps using appropriate symbols, colors, patterns, and legends. For example, water main sizes are shown by different line widths, pressure zones by different colors, and valve types by different symbols. Map production also involves printing the scale bar, north arrow, and organization logo. Sometimes tables, charts, and photographs are also inserted. Every map should have a title. Other information such as filename, map number, project name and number, plot date, mapmaker, and map source should also be printed on each important map.

This last step is important because it shows the final result of all the previous steps. Plotting of GIS maps in the correct style, orientation, scale, size, and format is a time-consuming process. Plotter speed and network speed are often the main bottlenecks. Templates for symbology, legends, and layouts should be developed in-house or obtained from other organizations to expedite map production. Some vendors package templates and symbol libraries specific to water, wastewater, and stormwater system with their software products.

Maps are needed for map books, wall mounting, archives, presentations, reports, etc. The purpose of the map and the target audience should be considered to design an appropriate map layout. For example, a map for an engineering application should be plotted differently from a map for a water utility newsletter. This step also includes distribution of the maps to the staff of an organization and optionally to the general public using the Internet, intranet, or wireless technology.

CASE STUDIES

Borough of Ramsey, New Jersey

The Ramsey Board of Public Works is responsible for the operation and maintenance of water distribution and sanitary sewer collection systems throughout the Borough of Ramsey, New Jersey. In an effort to more efficiently manage its geographically referenced data, the board started to explore the benefits and applications of GIS technology and computer mapping in early 1993. As a first step, with the assistance of Chester Engineers, Inc. (Pittsburgh, Pennsylvania), the board completed a needs analysis and a GIS pilot project described in Chapter 2 (Shamsi et al., 1995; Shamsi et al., 1996). The Borough did not have any complete large-scale mapping resources suitable for utility asset location. System design recommended the following mapping specifications:

- A new digital orthophoto base map should be developed accurate to the tolerances normally associated with mapping at a scale of 1 in. = 50 ft.
- Orthophoto base mapping should be established on the New Jersey State Plane Coordinate System. Survey monuments should be established to enable subsequent mapping efforts to be tied to the basemap coordinate system.
- Lot and block boundaries should be included as a unique layer in the base map.
- Additional map layers for the water distribution system, the sanitary sewer collection system, the stormwater collection system, and census geography should be developed to support other needs of the Borough.
- Use GPS to verify the locations of valves, hydrants, and manholes.

Table 8.2 summarizes the required map layers and the features depicted in each layer. Table 8.3 presents the data dictionary for Ramsey's pilot project.

Table 8.2 Map Layers and Depicted Features

Layer	Features
New Jersey State Plane Grid System	Grid tics, survey monuments
Photogrammetric base map	Building outlines, street outlines and centerlines, vegetation, water courses, topography
Parcels	Lot and block boundaries
Water distribution system	Mains, valves, hydrants
Sanitary sewer collection system	Sanitary sewers, manholes, cleanouts
Stormwater collection system	Storm sewers, manholes, catch basins
Census geography	Blocks, block groups, tracts

Table 8.3 Data Dictionary

Feature Type	Required Attributes
Water mains	Segment ID Pipe material Pipe size Repair date(s) Repair type(s)
Water valves	Valve ID Installation date Type Date last cycled Normal position
Fire hydrants	Hydrant ID Associated valves Type Installation date Flow data Preferred status
Sewers	Sewer ID Pipe material Pipe size From manhole ID To manhole ID
Manholes	Manhole ID Manhole type Lid type Depth Inflow problem status
Catch basins	Catch basin ID

All of the borough's sewer system data were stored in a nondigital form. This information was manually entered into the GIS database. Map features and attribute information were gleaned from a variety of sources. Table 8.4 lists recommended sources for specific map features and attributes. After reviewing the Borough's available maps and considering the mapping requirements of the priority applications, the use of digitization was recommended for data conversion.

ESRI's ArcInfo (Version 6) and ArcView (Version 2) software products were the two main software programs used in this project. Because the Borough did not have the technical staff to maintain an ArcInfo GIS, data conversion work was done by the consultant using ArcInfo. Due to its low cost, ease of use, and compatibility

Table 8.4 Source Documents for Map Features and Attributes

	Source Documents			
Types of Features	Base Map	Utility Markouts	Water System as Built, 1970, 1:50-Scale	Sewer System as Built, 1970, 1:100-Scale
Water mains		☑	☑	
Fire hydrants	☑	☑	☑	
Water valves		☑	☑	
Sewer pipes		☑		☑
Manholes	☑	☑		☑
Catch basins	☑	☑		☑

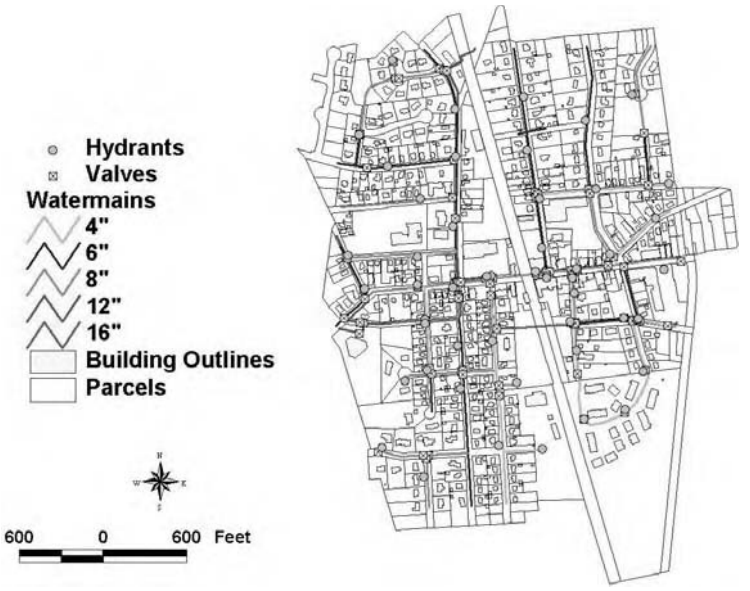


Figure 8.8 Water distribution system map showing pipes classified by diameter.

with the ArcInfo file format, ArcView was installed in the Borough for routine display and plotting of maps and querying the GIS database.

Figure 8.1 shows the Ramsey digital orthophoto base map with an accuracy of ± 1.25 ft. The figure on the first page of this chapter shows a map of sanitary sewers and manholes overlayed on the parcel and buildings layers. The sewer features were taken from as-built drawings and adjusted to locations of surface features that were field verified. Figure 8.8 shows a water system map with water mains, hydrants, and valves overlayed on the parcel and buildings layers. Water mains have been classified by the diameter attribute of the water main layer.

City of Lake Elsinore, California

Elsinore Valley Municipal Water District located in the city of Lake Elsinore, CA, provides water, wastewater, and agricultural services to approximately 25,000 customers. It maintains an infrastructure consisting of 305 mi of water mains, 290 mi of sewer pipes, and 246 mi of irrigation pipelines. An ESRI software user since 1992, the District migrated to ESRI's new ArcGIS geodatabase platform in 2000. The migration data conversion work, which included 5000 as-built drawings and 300 atlas pages, was completed in 8 months. The source documents were scanned and linked to GIS features. The system has 35 users and employs SQL Server 7 relational database management system software, Gateway 7400 and XI NetRAIDer dual 800 MHz servers, and Pentium III workstations. The system maintains approximately 75,000 maps in an enterprise GIS environment. GIS allows operations

personnel to create atlas maps and perform asset maintenance, a customer service department to analyze meter-reading routes and identify illegal connections, and an engineering department to generate as-built construction drawings (ESRI, 2002a).

Allegheny County, Pennsylvania

Each year, sewage overflows affect Pittsburgh's rivers up to 70 days during the boating season, making water unacceptable for recreational contact. In addition, the overflows affect Allegheny County's primary source of drinking water. Municipalities in the region could face a potential $2-billion investment in order to correct the problem. In 2003, the U.S. Environmental Protection Agency (USEPA) issued administrative consent orders to 83 municipalities of Allegheny County for controlling the sewage overflows. The orders required communities to address the sewage overflow issue through mapping, cleaning, and televising the system, and flow monitoring. This information will help communities develop a long-term wet-weather control plan. The orders required the following GIS mapping requirements:

- Create an ESRI-compatible base map that includes streets, street names, municipal boundaries, and streams.
- Create an updated ESRI-compatible comprehensive sewer map of the sanitary sewers with the following content:
  - Location of the sewer lines
  - Direction of flow
  - Size of the sewer lines
  - Sewer line material
  - Locations of interconnections with other municipal sewer systems
  - Field-verified location of manholes and catch basins (identified by a comprehensive numbering or lettering system)
  - Location of pump stations, force mains, and siphons
  - Location of streams or drainage ways tributary to the sewers
  - Location of overflow (combined sewer overflow [CSO] and sanitary sewer overflow [SSO]) structures
- Locate all significant structures to a minimum horizontal accuracy of 3 ft.
- Use as-built drawings, GPS, or traditional land surveying methods to determine structure locations.
- Use State Plane Pennsylvania South NAD83 projection system for coordinates and NAVD88 datum for elevations.
- Manhole inverts and rim elevations to a minimum vertical accuracy of 0.10 ft for overflow structures, sewers with greater than 10-in. diameters, and other critical sewers.
- Incorporate field inspection data into comprehensive sewer maps.

USEFUL WEB SITES

Geospatial Information and Technology Association (GITA)	www.gita.org
USGS GIS Maps for Water Resources	water.usgs.gov/maps.html
USEPA GIS Information	www.epa.gov/epahome/gis.htm

CHAPTER SUMMARY

GIS provides powerful capabilities for creating cost-effective and intelligent maps for water, wastewater, and stormwater systems. The quality of maps depends on data quality, errors, accuracy, scale, and resolution. GIS maps can be developed in many different ways. A mapping approach that best meets the application needs of an organization should be selected. GIS mapping requires five steps: needs analysis, data collection, data conversion, data processing, and map production. Needs analysis and pilot project testing is highly recommended, as explained in Chapter 2 (Needs Analysis). Data conversion is the most time-consuming and costly part of mapping projects and should be performed carefully. Data conversion application software can be used to reduce data conversion cost. Digitization and scanning are the two popular methods of data conversion but users can benefit from innovative variations of these methods. GIS maps should always be constructed in reference to a base map that has the required accuracy and resolution to support the intended applications of GIS maps.

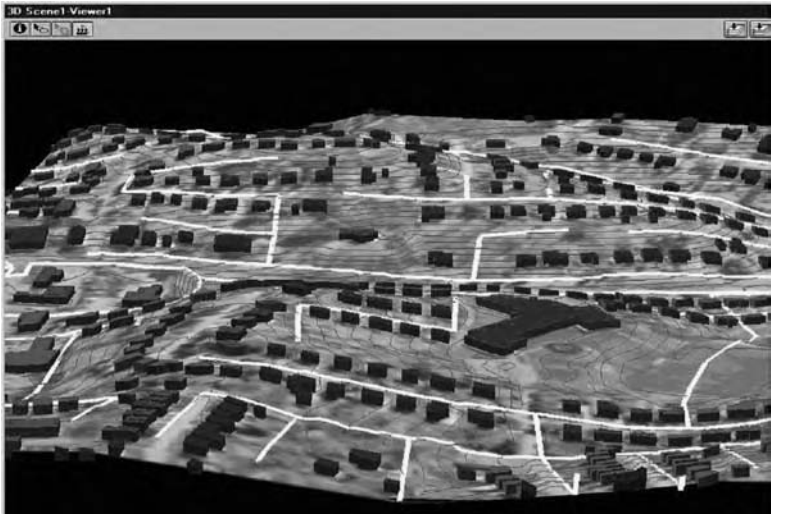
CHAPTER QUESTIONS

1. What are the different types of maps?
2. What are the different types of base maps and their characteristics?
3. How is a GIS map different from a CAD drawing?
4. What is the relationship between map accuracy, scale, and resolution?
5. What are the various steps for making GIS maps? Which step is most critical and why?
6. What is data conversion and how is it done?
7. What is topology, and how does it affect GIS maps?
8. Identify a GIS application in your organization and recommend the appropriate mapping layers and source documents to implement that application.

CHAPTER 9

Mapping Applications

A map is worth a thousand numbers but you must go beyond mapping to sustain your GIS program. This requires putting your GIS maps to beneficial use.



3D display of sewer system for the Municipality of Bethel Park, Pennsylvania.

LEARNING OBJECTIVE

The learning objective of this chapter is to understand the applications of GIS maps in the water industry.

MAJOR TOPICS

- Common mapping functions
- Mapping application examples for water systems
- Mapping application examples for wastewater systems
- Mapping application examples for stormwater systems

LIST OF CHAPTER ACRONYMS

2-DTwo Two-Dimensional
3-DThree Three-Dimensional
COGO Coordinate Geometry
CSO Combined Sewer Overflow
DEM Digital Elevation Model
DTM Digital Terrain Model
I/I Inflow/Infiltration
MS4 Municipal Separate Storm Sewer System
NPDES National Pollution Discharge Elimination System
SQL Structured Query Language
TIN Triangular Irregular Network/Triangulated Irregular Network

---

This book focuses on the four main applications of GIS, which are mapping, monitoring, modeling, and maintenance and are referred to as the "4M applications." In this chapter we will learn the applications of the first *M* (mapping).

CUSTOMER SERVICE APPLICATION IN GURNEE

The Village of Gurnee, located near Chicago (Illinois), has seen explosive growth since the 1970s. Population more than doubled from 13,701 in 1990 to 28,834 in 2000. This growth bogged down Gurnee personnel with public meetings, rezoning hearings, code violations, road construction, and water line and sewer line repairs. Several departments were legally required to notify property owners of meetings, pending changes, or future disruptions. The Village used GeoMedia GIS software from Intergraph to automate the resident notification process (Venden and Horbinski, 2003). The automated system — known as Notifast™ — allows staff members to enter a street address or select a parcel, specify a distance radius, and identify all property owners within that distance. Notifast then generates a list of mailing labels

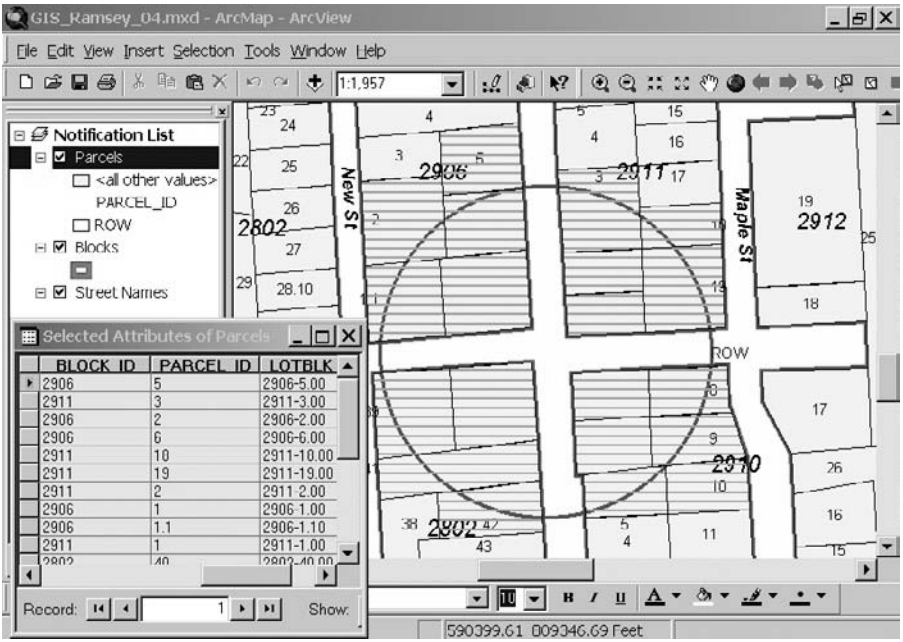


Figure 9.1 Selecting notification parcels in ArcGIS.

and mail-merged form letters. Each letter includes an embedded map that clearly shows the area in question. This application of existing GIS maps has improved efficiency and customer service. What used to take hours is now completed in minutes.

Figure 9.1 shows a screenshot of how parcels can be selected in ESRI's ArcGIS software by drawing a circle. The 200-ft radius circle is drawn with its center at the project site. All the parcels that are intersected by the circle are selected as shown in the hatched pattern. The attribute data for the selected parcels are shown in the table to the left of the circle. These data can be copied in a word processor's mail-merge feature to generate notification letters.

As explained in Chapter 8 (Mapping), creation of GIS maps is a huge undertaking. Unfortunately, in today's competitive world, despite the enormous effort required to create them, colorful maps alone can no longer justify the existence of a GIS program. This is because, after spending large sums of money on a mapping project, many water and sewerage agencies are wondering what else to do with their GIS maps. To garner continued financial support for a GIS program, GIS programs must go beyond conventional map printing activities and demonstrate the usefulness (applications) of GIS maps. This chapter shows how to do that.

It does not occur to most people that a good map raises more questions than it answers — that the question of why things are located where they are raises important intellectual issues, with immediate serious implications. (Lewis, 1985).

A paper map provides the power of place but a map in GIS can automate "the why of where." Most GIS applications originate from GIS maps. This chapter describes simple mapping applications. Advanced applications like hydrologic and hydraulic modeling, facilities management, inspection and maintenance, and planning are described in other chapters.

COMMON MAPPING FUNCTIONS

Most GIS packages provide simple data analysis capabilities that can be used to create basic mapping applications without custom programming. Typical examples of these GIS functions are given in the following subsections.

Thematic Mapping

Thematic mapping is one of the easiest ways to develop mapping applications. Map components can be logically organized into sets of layers called *themes*. A theme is a group of similar geographic features, such as hydrants (points), water mains (lines), or pressure zones (polygons). These themes can be manipulated using classification and legend (symbology) to create different thematic maps. Classification places attribute values into groups. For example, water mains for a city can be classified by diameter to create groups such as pipes smaller than 6 in., pipes 8 to 12 in., and pipes greater than 12 in. These groups can then be displayed using different legends reflecting different colors, line thicknesses, and symbols (e.g., solid, dashed, or dotted lines). Each combination of a classification method and legend will produce a different thematic map. Figure 9.2 shows the difference between an unclassified and nonsymbolized sewer map (top) and the same map classified and symbolized by pipe diameters (bottom). Obviously, the bottom map will be more useful in applications that require pipe diameters, such as flow monitoring and hydraulic modeling.

Spatial Analysis

Many GIS packages also allow users to select the features of one or more themes using the features of another theme. Referred to as spatial or overlay analyses, these *theme-on-theme* selection functions identify whether one feature lies within another, whether it completely contains another, or whether it is within a specified distance of another, and so on. For example, these mapping operations can identify the manholes in a given drainage basin, or census blocks within a watershed, or the properties within a certain distance of a water main.

Buffers

A buffer can be created around one or many features by using the buffer function typically found in most GIS packages. Creating a buffer provides a visual representation on a map of the area within a certain distance of one or more features. Buffer can also be used to select features in other layers that fall within the buffered area.

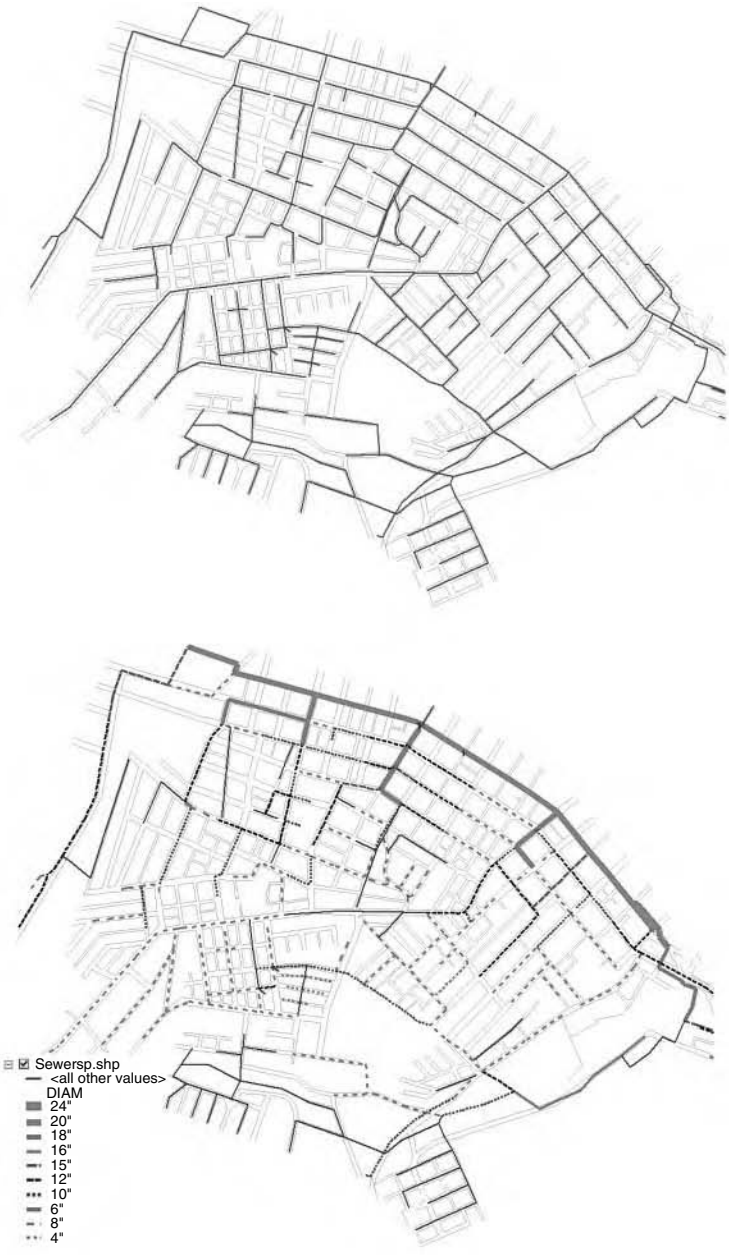


Figure 9.2 A sewer system map. Above: without attribute classification and symbology; below: with attribute classification and symbology for pipe diameter.

Figure 9.3 Buffer for a sewer separation project.

For instance, a buffer can be drawn to show the area around a well that is contaminated or to represent a floodplain around a river. Figure 9.3 shows a buffer created around existing combined sewers in the Borough of Kittanning, Pennsylvania. It identifies the area targeted for sewer separation in the Borough's combined sewer overflow (CSO) control plan.

Hyperlinks

Some GIS packages such as ArcView and ArcGIS have Hotlink or Hyperlink functions respectively to provide additional information about the features. For example, when a feature is clicked with the Hyperlink tool, a document or file is launched using the application associated with that file type. This function can be used as a document management system from within a GIS. For example, users can click on a valve to display the valve card, on a sewer pipe to launch a televised video file, or on a water main to display its as-built drawing. Figure 9.4 shows a display of a scanned image of a water distribution system valve parts list using ArcView's Hotlink function. Additional information on hyperlinks is provided in Chapter 15 (Maintenance Applications).

WATER SYSTEM MAPPING APPLICATIONS

Representative mapping applications for water distribution systems are given in the following subsections.

MWRA Water System Mapping Project

In the late 1990s the Massachusetts Water Resources Authority (MWRA) conducted a major in-house effort to develop a linked system schematic set. These overview maps did not have a scale or base, but showed how the elements of the system connected hydraulically. In addition to developing the digital version of these maps, the GIS staff created links to a number of real time databases and the location plans for each valve. Thus, the schematic showed the status of all mainline valves for emergency response planning. If a valve was inoperable or had been opened or closed, the GIS updated the color of the pipe line on the schematic, alerting the users to its condition.

MWRA then consolidated their water distribution facility data and source document references into a manageable set of data tables and digital maps that were easily accessible to MWRA staff. They also linked the records of every pipe, valve, meter, and fitting in the distribution system to existing databases of record plans, land takings, hydraulic conditions, leaks, maintenance activities, and schematic representations in the GIS. This effort allowed MWRA users to click on a pipe drawing, and instead of being able to access just the record drawing, choose from a menu of options regarding the land records and model scenarios, as well as the physical condition and maintenance history of the pipe. This provided access to a variety of maintenance and data development activities throughout the water division (Estes-Smargiassi, 1998).

Service Shutoff Application

To strengthen enforcement activities on delinquent accounts, the Philadelphia Water Department (PWD) developed an ArcView GIS tool to support its shutoff program. The shutoff tool identifies priority sites and automates crew assignments

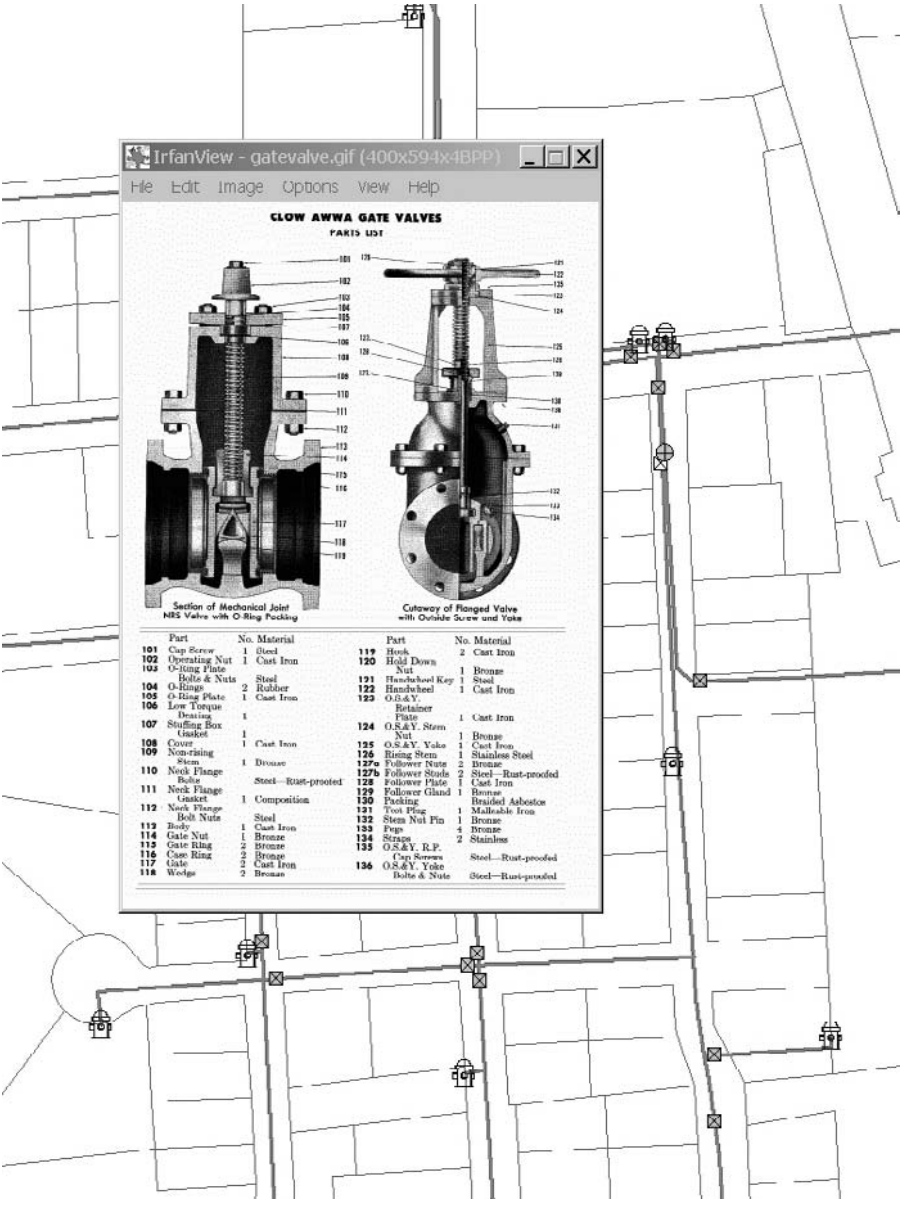


Figure 9.4 Hotlink function displaying valve information for a water distribution system.

to increase crew productivity. A prioritizing algorithm identifies delinquent customers based on age and amount in arrears. These priority accounts identify clusters of shutoff sites. Clusters are formed using ArcView Network Analyst extension to locate nearby sites reachable by the street network. The application generates maps and tables for each crew's daily schedule. With the help of this

application and other measures, PWD improved enforcement efficiency by 60% (Lior et al., 1998).

Generating Meter-Reading Routes

Seattle Public Utilities (SPU) has more than 174,000 customers. In an effort to save meter-reading cost and improve customer service, GIS was used to create more efficient meter-reading routes. In 1995, SPU's Metering and GIS departments joined forces to explore the feasibility of rerouting meter-reading routes. In 1997, a pilot project was completed using software from ESRI and RouteSmart Technologies, which convinced the SPU management to implement a citywide rerouting project. The project implementation resulted in cost savings for the ratepayer, especially because of SPU's prior investments in GIS data conversion (Lee, 1998).

Map Maintenance Application

In Ohio, Cincinnati Water Works (CWW) developed a GIS application for water distribution system maintenance, hydraulic modeling, and customer service. A major component of this application was the development of a Map Maintenance Application (MMA), which is an update developed using ArcInfo and Oracle. It allows simultaneous maintenance of water system layers and related RDBMS tables. The MMA features COGO data entry, scroll lists for table inserts, and feature abandonment. It also utilizes the integrity constraint checking capability of the RDBMS (Guetti, 1998).

WASTEWATER SYSTEM MAPPING APPLICATIONS

Representative mapping applications for wastewater systems are given in the following subsections.

Public Participation with 3D GIS

At the present time many cities in the U.S. are planning to spend billions of dollars to meet the regulatory requirements for correcting their wastewater problems, such as wet-weather overflows from the sewer systems. Given the potential for significant expenditures of public funds, a key to the future success of capital projects is public participation. The regulatory agencies are mandating public participation as a required element of these projects. Therefore, establishing early communication with both the public and the regulatory agencies should be an important first step in a long-term planning approach, and crucial to the success of a wastewater capital improvement program. By informing the public early on in the planning process about the scope and goals of the program, and continuing this public involvement during development, evaluation, and selection of alternatives, issues and potential conflicts can be identified and addressed more expeditiously, minimizing the potential for prolonged delay or additional cost.

Three-dimensional (3D) maps are an effective means of public education, notification, and communication. They provide impressive "real-life" visions of proposed

projects. GIS, CAD, and multimedia technologies can be combined to produce 3D models and 3D animations.

3D maps can be created by adding elevation data to existing 2D vector layers. Elevation data can be obtained from contour maps, digital elevation models (DEM), digital terrain models (DTM), aerial surveys, engineering drawings, or GPS surveys. The quality of 3D analysis largely depends on the accuracy and resolution of elevation data. For engineering design, highly accurate data are required. For presentation mapping, less accurate data can be used. For example, in-lieu of accurate heights, buildings can be elevated in proportion to the number of floors.

GIS fly-through scenes allow the public to experience a 3D walk-through of a proposed project. Animation can animate objects within a model, e.g., cars moving down the road or people walking on the curbs. Realistic materials and textures can be added to depict the actual field conditions, e.g., wet streets. Climatic conditions, e.g., snow cover or sunlight can be added to portray the specific design conditions.

3D capability is usually added to standard GIS packages by installing additional 3D extensions or modules. For example, 3D Analyst Extension enables ESRI's ArcGIS users to visualize and analyze surface data. Using 3D Analyst, one can view a surface from multiple viewpoints, query a surface, determine what is visible from a selected location on a surface, and create a realistic perspective image draping raster and vector data over a surface.

The figure on the first page of this chapter shows a 3D map of a sewer system for the Municipality of Bethel Park, Pennsylvania. Created using 3D Analyst Extension, the map shows buildings, contours, and sewer lines overlaid on an aerial photograph. Because wastewater must flow down the hill, design of gravity sewers depends on topography. GIS can help create 3D maps to facilitate efficient engineering of a new sewer system, quick assessment of an existing sewer system, or presentation of the project layout to the general public.

Figure 9.5 shows a 3D fly-through animation of an Allegheny County Sanitary Authority (ALCOSAN) sewer construction project in the City of Pittsburgh, Pennsylvania. Completed in 1999, this $24-million project involved construction of a 5.2-mi-long parallel relief interceptor to relieve sewer overflows and basement flooding in the Saw Mill Run Watershed of the city. The pipe sizes for the relief interceptor ranged from 36 to 54 in. in diameter. The parallel interceptor was designed to follow the alignment of the Saw Mill Run Creek because the right-of-way corridor along the main street (Route 51) was not conducive to a major interceptor construction project. Animation was used to present the project design in a general public meeting. Figure 9.5 shows the sewers and manholes exposed above the ground surface to help the general public envision the proposed interceptor layout. The animation clearly indicated that the new sewers will mostly avoid backyards, parking lots, and roads — an issue that was very important to the residents of the community. This approach helped ALCOSAN to win the public support for the project.

Mapping the Service Laterals

Recent inflow and infiltration (I/I) studies have shown that service laterals are one of the most significant I/I contributors, generating as much as 60% of the total

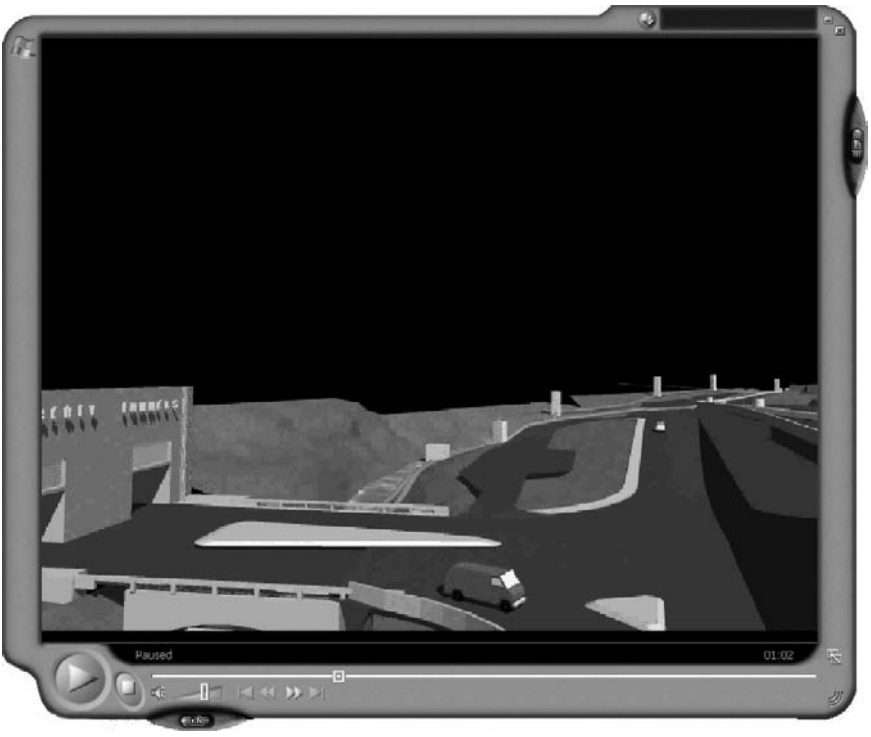
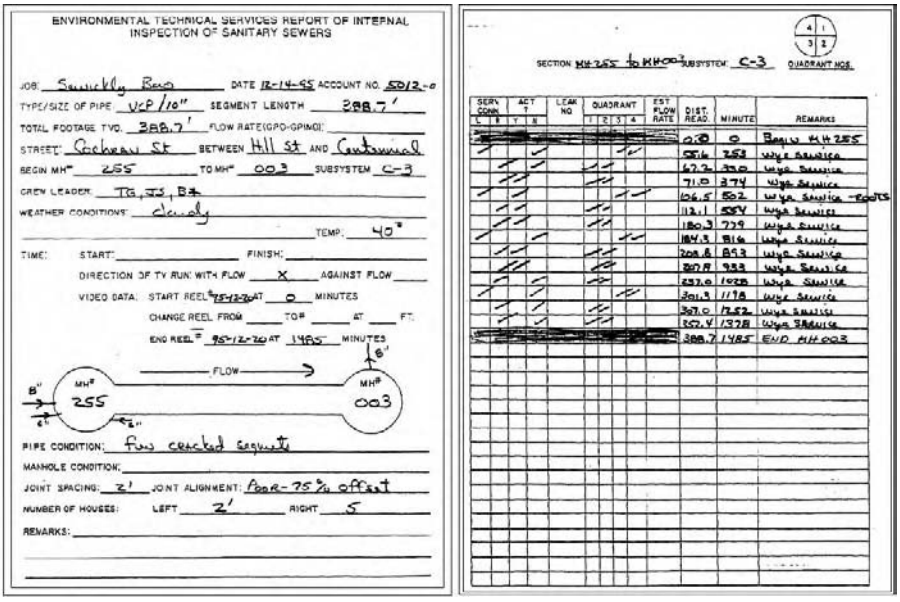


Figure 9.5 3D fly-through animation of a sewer construction project.

I/I quantity. Unfortunately, most collection systems do not have maps of their service laterals. TV inspection logs provide accurate lateral location information, which can be used to prepare service lateral GIS maps. For example, Chester Engineers (Pittsburgh, Pennsylvania) developed a Wyemaker ArcView extension to add service wyes (Ys) to sewer lines. Developed using ArcView Dialog Designer extension, this utility is a good example of GIS customization.

Figure 9.6 shows a typical two-page TV inspection log. The first page provides general information about a sewer pipe. The second page details the specific pipe defects and their location. This page also includes the location of service laterals measured in feet from one manhole to the next. The Wyemaker extension will first create a database containing the wye information for each sewer main. The wye information is entered in a wye input form shown in Figure 9.7 that basically replicates the data fields on the log sheets. The upstream and downstream manhole identification numbers are entered. In this case, the sewer segment ID is created by concatenating the upstream and downstream manholes. The sewer length from the inspection log is entered as well as the beginning distance (usually zero). The user then indicates the direction of the measurements, the wye measurement and the side of the line in which the wye is located. The direction of the measurements is the direction in which the line was inspected. The direction is either "against the flow" (from the downstream manhole to the upstream manhole) or "with the flow" (from



ENVIRONMENTAL TECHNICAL SERVICES REPORT OF INTERNAL INSPECTION OF SANITARY SEWERS

JOB: Sawickly Road DATE: 12-14-95 ACCOUNT NO. 5012-0
TYPE/SIZE OF PIPE: VCP / 10" SEGMENT LENGTH: 388.7'
TOTAL FOOTAGE TVD: 388.7' FLOW RATE(GPD-GPM):
STREET: Cochran St BETWEEN: Hill St AND: Centennial
BEGIN MH#: 255 TO MH#: 003 SUBSYSTEM: C-3
CREW LEADER: TG, JS, BA
WEATHER CONDITIONS: cloudy TEMP: 40°
TIME: START: FINISH:
DIRECTION OF TV RUN: WITH FLOW: X AGAINST FLOW:
VIDEO DATA: START REEL# 95-12-20 AT 0 MINUTES
CHANGE REEL FROM: TO#: AT: FT:
END REEL # 95-12-20 AT 1485' MINUTES
FLOW
MH# 255
MH# 003
PIPE CONDITION: Few cracked seguents
MANHOLE CONDITION:
JOINT SPACING: 2' JOINT ALIGNMENT: Poor - 75% offset
NUMBER OF HOUSES: LEFT: 2' RIGHT: 5
REMARKS:

SECTION MH 255 to MH 003 SUBSYSTEM: C-3 QUADRANT NOS.

SERV CONN L	SERV CONN R	ACT ? Y	ACT ? N	LEAK NO	QUADRANT 1	QUADRANT 2	QUADRANT 3	QUADRANT 4	EST FLOW RATE	DIST. READ.	MINUTE	REMARKS
										0.0	0	Begin MH 255
										56.6	258	Wye Service
										67.2	330	Wye Service
										71.0	374	Wye Service
										106.5	502	Wye Service - Roots
										112.1	584	Wye Service
										180.3	729	Wye Service
										184.3	816	Wye Service
										208.8	893	Wye Service
										207.8	933	Wye Service
										237.0	1028	Wye Service
										301.9	1178	Wye Service
										367.0	1352	Wye Service
										357.4	1378	Wye Service
										388.7	1485	END MH 003

Figure 9.6 Sample TV inspection log.

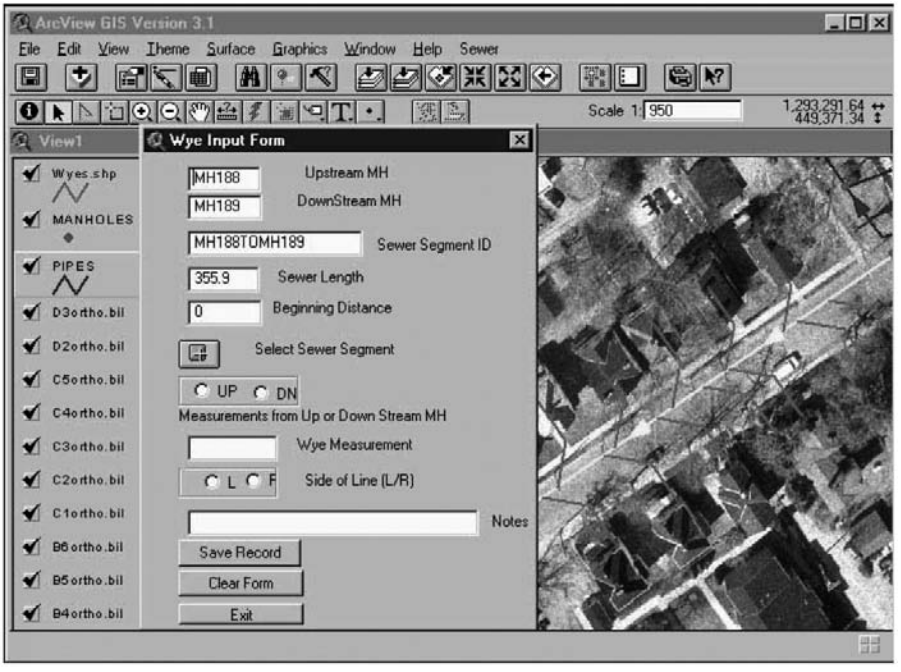


Figure 9.7 Service laterals view in ArcView.

the upstream manhole to the downstream manhole). Entering the direction of measurement and the side of the line eliminates the need for the user to recalculate distances if the line was measured against the flow. The calculations are automatically done by the extension and the position of the wye (right or left) is automatically switched if necessary. Once the form is completed, the user simply clicks on "Save Record" and the wye record is automatically saved to the wye table.

When all records are entered, the database table is used to create a Shapefile of the wyes. All the wyes in the Shapefile are exactly the same length and at the exact same angle from the sewer line, creating an aesthetically appealing data set. The wyes are kept in a separate Shapefile rather than being added to the existing sewer theme in order to maintain the connectivity of the main lines. A sample lateral map created using this method is shown in Figure 9.7.

Additional examples of wastewater system mapping applications are included in Chapter 15 (Maintenance Applications).

STORMWATER SYSTEM MAPPING APPLICATIONS

A representative mapping application for stormwater systems is presented below.

Stormwater Permits

In the U.S., many cities with storm sewers and outfalls must comply with a National Pollution Discharge Elimination System (NPDES) Stormwater Permit under the U.S. Environmental Protection Agency's Municipal Separate Storm Sewer System (MS4) permit program. These permits require U.S. municipalities to develop stormwater pollution prevention programs, track the compliance schedules, and monitor stormwater system maintenance and water quality sampling.

Some cities finance the operation of a stormwater utility that pays for permit compliance by assessing a stormwater fee (drainage tax) on the properties. Initially, some U.S. municipalities chose to implement a flat tax for all property owners, based on the tax value of their property. Because only the impervious area of a property contributes to stormwater runoff, many cities used a more equitable approach that assessed the tax in proportion to the impervious area of each property. Accurate calculation and tracking of impervious areas from conventional methods (e.g., from survey or paper maps) is cumbersome and costly. GIS allows accurate and inexpensive capture of impervious surfaces from aerial photographs, orthophotos, scanned site plans, and parcel maps.

GIS can be used to develop stormwater user fees using the following steps:

1. Delineate property land covers using digital orthophotos.
2. Assign a percent imperviousness factor to each land cover. For example:
   - Roofs, 100%
   - Concrete driveway, 100%
   - Gravel driveway, 75%
   - Parking lot, 90%

3. Estimate land-cover area-weighted mean percent imperviousness of each property.
4. Determine stormwater rates based on mean percent imperviousness of a property rather than the total impervious area.

For example, the Texas legislature passed a law allowing municipalities to create drainage utilities that assessed utility fees associated with each land parcel based on the amount of stormwater runoff it contributed. The City of Irving, Texas, started using GIS and color-infrared aerial photography in 1993 to examine and map the sources of the city's stormwater runoff. The City assumed that a simple distinction between pervious surfaces such as vegetative cover and impervious surfaces such as buildings and parking lots would be sufficient to determine the stormwater runoff of each parcel with acceptable accuracy. To determine the area of each surface type with individual parcels, the tax-parcel layer was overlaid with the pervious/impervious layer. This operation created a list of each parcel's tax identification number and the associated pervious/impervious area. This table was then linked to the City's water utility billing system to calculate the amount due from each property owner and include that amount in the monthly utility bill (*American City&County*, 1994).

The City and County of Denver, Colorado, has a storm drainage utility that assesses a fee on all properties in Denver based on the amount of impervious surface present on each property, as well as on the percentage of the total area of each property that the impervious surface represents. Initially, Denver used field measurements but did not find them cost-effective. In 2002, they hired the Denver firm of Elroi Consulting, Inc., to implement an impervious surface GIS for the County. Initially, the system was implemented using ArcGIS 8.1 and a personal geodatabase that required only ArcView licenses. The system was later upgraded to work with an ArcSDE geodatabase, complete with versioning and multiuser editing. The system builds on the ArcMap component of ArcGIS, both removing unnecessary toolbars from the user interface and adding several custom toolbars. The user-friendly interface enables even the non-GIS users (e.g., field inspectors) to use the system. Impervious surfaces are digitized directly into the GIS, thus making impervious surface fee assessments defensible. The impervious surface GIS was subsequently linked to Denver's stormwater billing system for rapid database update. With this system, Denver now collects over $12 million annually in impervious surface fees (Elroi, 2002).

Oakland County, Michigan, which has more than 440,000 parcels, uses high-density spot elevation and break-line data in GIS to create triangular irregular network (TIN) surface models. These models help the county drain commissioner's office to visualize the terrain and delineate catchment areas. This approach identifies precise drainage patterns and the land that contributes runoff to a storm drain. The catchments define the drainage district boundaries that are used to assess the appropriate drainage tax in an equitable manner (Gay, 2003).

The conventional paper-based permit compliance approach can be very hectic for large cities. GIS is being used in many U.S. cities to meet the stormwater permit requirements system in a user-friendly environment. For example, SQL, Avenue, and Dialog Designer have been utilized to develop an interface between ArcView GIS and Microsoft Access for developing an NPDES compliance application (Huey, 1998). This application is helping both the management of stormwater features such

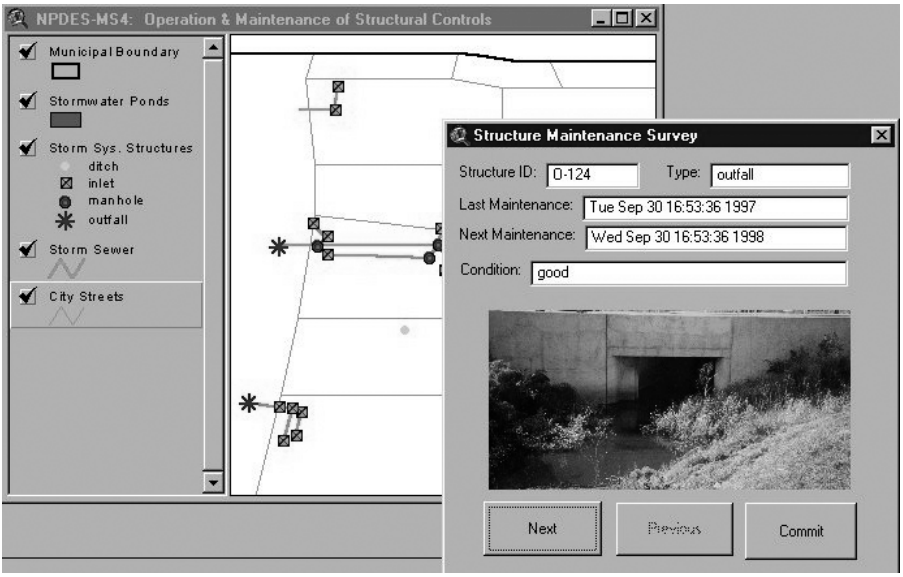


Figure 9.8 Utilization of dialogs to monitor and update structure-maintenance activities.

as sewer lines and outfalls, and preparation of complex NPDES Annual Reports. The report requires summarizing all stormwater management program goals and efforts toward meeting those goals. An example of the use of dialogs and custom-programmed icon buttons for query is shown in Figure 9.8.

CHAPTER SUMMARY

Paper maps are useful but their applications are limited. In most cases, mapmaking alone cannot justify the continued financial support for a GIS department. GIS maps and associated databases can be manipulated in a variety of ways to develop applications that automate and simplify the routine tasks. GIS maps are the starting point for developing GIS applications: no maps, no applications. In this chapter we learned about some basic applications of GIS maps. Advanced applications will be presented in subsequent chapters.

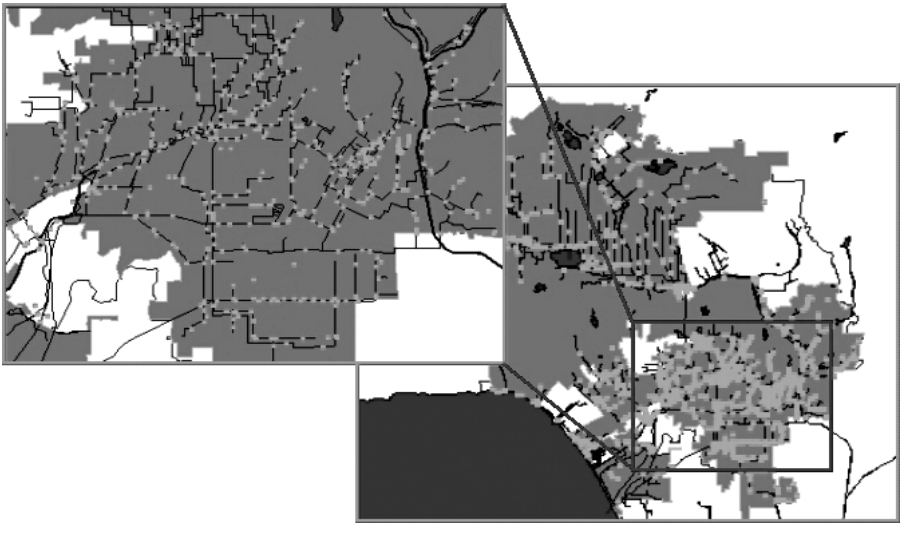
CHAPTER QUESTIONS

1. What is the meaning of "beyond mapping?" Why is it important to go beyond mapping?
2. Compile a list of ten mapping applications.
3. Using literature review or the Internet, write a 200-word example of a mapping application for a water distribution system not described in this chapter.
4. Repeat Question 3 for a wastewater collection system.
5. Repeat Question 3 for a storm drain system.

CHAPTER 10

Monitoring Applications

GIS is ideally suited to install, maintain, and query monitoring equipment such as, rain gauges, flow meters, and water quality samplers for system physical and hydraulic characterization. GIS allows display and analysis of monitoring data simply by clicking on a map of monitoring sites.



Gauging and inspection stations in the city of Los Angeles, California.

LEARNING OBJECTIVE

The learning objective of this chapter is to familiarize ourselves with GIS applications in monitoring data for effective operation and management of water, wastewater, and stormwater systems.

MAJOR TOPICS

- Rainfall monitoring
- Satellite and radar rainfall data
- Flow-monitoring applications
- Permit reporting applications
- Internet monitoring applications
- Infrastructure monitoring applications

LIST OF CHAPTER ACRONYMS

CCTV Closed Circuit Television
GML Geography Markup Language
HRAP Hydrologic Rainfall Analysis Project (a map projection system)
NCDC National Climatic Data Center (U.S.)
NEXRAD Next Generation Weather Radar
NOAA National Oceanic and Atmospheric Administration (U.S.)
NPDES National Pollution Discharge Elimination System (U.S.)
NWS National Weather Service
OGC Open GIS Consortium
SCADA Supervisory Control and Data Acquisition Systems
WSR Weather Surveillance Radar

This book focuses on the four main applications of GIS, which are mapping, monitoring, modeling, and maintenance and are referred to as the "4M applications." In this chapter we will learn about the applications of the second *M* (monitoring).

MONITORING REAL TIME RAINFALL AND STREAM-FLOW DATA IN AURORA

The city of Aurora, Colorado, is a community of approximately 247,000 people located in the southeast Denver metro area. The City's Utilities Department used ESRI's ArcInfo and ArcView software to analyze and display real time rain-gauge and stream-flow data from the City's ALERT Flood Warning System. Data were obtained by querying a remote real time data-collection database. The Utilities Department was able to access and analyze both historical and current real time rainfall and stream-flow data from an easy-to-use graphical interface. ArcView's Spatial Analyst Extension was used to develop a continuous areal rainfall surface from the point rainfall data. This allowed a better and clearer understanding of a particular storm and allowed a

more complex analysis of the true impact of the storm event. Applications of this program included studies in emergency flood response, location of flood reports, routine maintenance for storm sewers, and National Pollution Discharge Elimination System (NPDES) compliance for water quality (Rindahl, 1996).

MONITORING BASICS

Monitoring of various types of data is essential to the effective management of water, wastewater, and stormwater systems. Generally, two types of data are required: (1) physical characterization data and (2) hydraulic characterization data. Physical characterization data describe the physical condition of the infrastructure, such as pipe and manhole conditions. Examples of physical characterization data sources include closed-circuit television (CCTV) inspection of pipes, manhole inspections, and smoke-testing of buildings. GIS applications for these types of data are described in Chapter 15 (Maintenance Applications).

Hydraulic characterization data describe the quantity and quality of flow through pipes and open channels as well as meteorological factors impacting the flow, such as precipitation.

Wastewater and stormwater systems typically require data on flow quantity (depth, velocity, flow rate, and volume), quality (e.g., suspended solids and bacteria), and rainfall. Figure 10.1 shows a flowmeter and weir installation in a combined sewer system overflow manhole. The flowmeter (Flo-Dar from Marsh-McBirney, Inc.) shown on the left records incoming combined sewage depth, velocity, and flow data. The weir shown on the right collects outgoing overflows. Some wastewater and stormwater hydrologic and hydraulic (H&H) models require data on additional meteorological parameters such as ambient temperature, evaporation rate, and wind speed. Water distribution systems typically require water pressure and water quality data. In many projects, monitoring tasks make up a significant portion of the scope of work and could cost 20 to 30% of the total budget. Careful installation of monitors and effective management of monitoring data is, therefore, highly desirable for the on-time and on-budget completion of monitoring projects.

GIS is ideally suited for selecting the best sites for installing various hydraulic characterization monitors. Once the monitors have been installed, GIS can be used to query the monitored data simply by clicking on a map of monitoring sites. GIS can also be used to study the spatial trends in the monitored data. GIS is especially useful in processing and integrating radar rainfall data with H&H models of sewage collection systems and watersheds. This chapter will present the methods and examples of how to use GIS for installing and maintaining the monitors, and for querying and analyzing the monitoring data.

REMOTELY SENSED RAINFALL DATA

Many watersheds, especially those smaller than 1000 km^2, do not have recording rain gauges capable of recording at hourly or subhourly intervals. Sometimes rain gauges exist, but data are found missing due to equipment malfunction. Quite often,

Figure 10.1 Flowmeter and weir installation in a manhole for monitoring incoming and outgoing flows.

rain-gauge density is not adequate to accurately capture the spatial distribution of storm events. Such data gaps can be filled by the rainfall data provided by weather satellites and radars, as described in the following subsections.

Satellite Rainfall Data

Direct measurement of rainfall from satellites is not feasible because satellites cannot penetrate the cloud cover. However, improved analysis of rainfall can be achieved by combining satellite and conventional rain-gauge data. Meteorological satellites such as the NOAA-N series, those of the Defense Meteorological Satellite Program, and U.S. geostationary satellites can observe the characteristics of clouds with precipitation-producing potential and the rates of changes in cloud area and shape. Rainfall data can now be estimated by relating these cloud characteristics to instantaneous rainfall rates and cumulative rainfall over time. Cloud area and temperature can be used to develop a temperature-weighted cloud cover index. This index can be transformed to estimate mean monthly runoff values. Satellite rainfall estimating methods are especially valuable when few or no rain gauges are available (ASCE, 1999).

Radar Rainfall Data

Weather radars provide quantitative estimates of precipitation, which can be used as input to H&H models. Radar rainfall estimates augmented with data from sparse rain-gauge networks are useful in H&H modeling. Weather radars provide real time, spatially distributed rainfall data that can be extremely valuable for flood forecasting and flood warning.

NEXRAD Rainfall Data

The U.S. National Weather Service (NWS) has a group of weather radars called the Next Generation Weather Radar (NEXRAD) system. NEXRAD comprises approximately 160 Weather Surveillance Radar–1988 Doppler (WSR-88D) sites throughout the U.S. and selected overseas locations. This system is a joint effort of the U.S. Departments of Commerce (DOC), Defense (DOD), and Transportation (DOT). The controlling agencies are the NWS, Air Weather Service (AWS), and Federal Aviation Administration (FAA), respectively. Level II data provide three meteorological base data quantities: reflectivity, mean radial velocity, and spectrum width. These quantities are processed to generate numerous meteorological analysis products known as Level III data. Level II data are recorded at all NWS and most AWS and FAA WSR-88D sites. Level III products are recorded at the 120 NWS sites. The data are sent to the National Climatic Data Center (NCDC) for archiving and dissemination.

NEXRAD Level III Data

There are a total of 24 Level III products routinely available from NCDC, including 7 graphic products in clear-air mode, 11 in precipitation mode, 5 graphic overlays, and 1 alphanumeric product. Each product includes state, county, and city background maps. Level III graphic products are available as color hard copy, grayscale hard copy, or acetate overlay copies. A brief description and possible uses of these products are given below:

- Base Reflectivity (R): A display of echo intensity measured in dBZ (decibels of Z, where Z represents the energy reflected back to the radar). This product is used to detect precipitation, evaluate storm structure, locate boundaries, and determine hail potential.
- Base Spectrum Width (SW): A measure of velocity dispersion within the radar sample volume. The primary use of this product is to estimate the turbulence associated with mesocyclones and boundaries.
- Base Velocity (V): A measure of the radial component of the wind either toward the radar (negative values) or away from the radar (positive values). Negative values are represented by cool colors (green), whereas positive values are represented by warm colors (red). This product is used to estimate wind speed and direction, locate boundaries, locate severe weather signatures, and identify suspected areas of turbulence.

- Composite Reflectivity (CR): A display of maximum reflectivity for the total volume within the range of the radar. This product is used to reveal the highest reflectivities in all echoes, examine storm structure features, and determine the intensity of storms.
- Echo Tops (ET): An image of the echo top heights color-coded in user-defined increments. This product is used for a quick estimation of the most intense convection and higher echo tops, as an aid in identification of storm structure features, and for pilot briefing purposes.
- Hail Index Overlay (HI): A product designed to locate storms that have the potential to produce hail. Hail potential is labeled as either probable (hollow green triangle) or positive (filled green triangle). Probable means the storm is probably producing hail and positive means the storm is producing hail.
- Mesocyclone Overlay (M): This product is designed to display information regarding the existence and nature of rotations associated with thunderstorms. Numerical output includes azimuth, range, and height of the mesocyclone.
- One-Hour Precipitation (OHP): A map of estimated 1-h precipitation accumulation on a 1.1 nmi × 1.1 nmi grid. This product is used to assess rainfall intensities for flash flood warnings, urban flood statements, and special weather statements.
- Severe Weather Probability Overlay (SWP): A measure of a storm's relative severity as compared with those around it. The values are directly related to the horizontal extent of vertically integrated liquid (VIL) values greater than a specified threshold. This product is used for quick identification of the strongest storms.
- Storm Structure (SS) (alphanumeric product): A table displaying information on storm attributes that include maximum reflectivity, maximum velocity at lowest elevation angle, storm overhang, mass-weighted storm volume, storm area base and top, storm position, and storm tilt.
- Storm Total Precipitation (STP): A map of estimated storm total-precipitation accumulation continuously updated since the last 1-h break over the entire scope. This product is used to locate flood potential over urban or rural areas, estimate total basin runoff, and provide rainfall data 24 h a day.
- Storm Tracking Information Overlay (STI): A product that shows a plot of the past hour's movement, current location, and forecast movement for the next hour or less for each identified thunderstorm cell. This product is used to determine reliable storm movement.
- Tornadic Vortex Signature Overlay (TVS): A product that shows an intense gate-to-gate azimuthal shear associated with tornadic-scale rotation. It is depicted by a red triangle with numerical output of location and height.
- VAD Wind Profile (VWP): A graphic display of wind barbs plotted on a height staff in 500-ft or 1000-ft increments. The current (far right) and up to ten previous plots may be displayed simultaneously. This product is an excellent tool for meteorologists in weather forecasting, severe weather, and aviation.
- Vertically Integrated Liquid (VIL): The water content of a 2.2 nmi × 2.2 nmi column of air, which is color-coded and plotted on a 124 nmi map. This product is used as an effective hail indicator to locate most significant storms and to identify areas of heavy rainfall.

NEXRAD data and various visualization and analysis software tools are available from NCDC and commercial vendors.

Estimating Rainfall Using GIS

Because rainfall is a critical component in conducting H&H analyses, the quality of rainfall data is critical for accurate system hydraulic characterization. It is often the case that the spatial distribution of rain-gauges over a collection system is too sparse to accurately estimate the rainfall over a given basin. Radar rainfall technology can be used to obtain high-resolution rainfall data over an approximately 2 km $\times$ 2 km area (called a pixel). GIS can be used to display and process the radar rainfall data.

NEXRAD Level III data from Doppler radar measurements provide spatially dense rainfall data. These data are similar to those commonly seen on weather maps. Such data do not require interpolation between point data from widely scattered rain-gauges because they provide continuous rainfall measurements throughout a watershed or sewershed (Slawecki et al., 2001).

WRS-88D radar images have a mean resolution of 4 km $\times$ 4 km. They are registered to the Hydrologic Rainfall Analysis Project (HRAP) map projection system. The radar rainfall data can be incorporated in a GIS-based distributed hydrologic model by importing it in a raster grid format. However, this may require reprojecting the rainfall grid from the HRAP coordinate system to the coordinate system being used by the model.

The accuracy of radar rainfall estimates can be improved substantially by calibrating the radars using the point rain-gauge observations. Generating NEXRAD rainfall estimates requires extensive expertise and computational resources (ASCE, 1999). Hourly NEXRAD rainfall data called hourly digital precipitation array (DPA) are available from several NWS-authorized commercial data vendors, such as WSI Corporation and Paramax Systems Corporation. Weather data vendors such as DTN Weather Services (now part of Meteorlogix, Inc.) provide continuously updated, real time, GIS-ready, georeferenced weather data in ESRI GRID (raster) and Shapefile (vector) formats and georegistered TIFF format to weather-enable the GIS applications for water and wastewater utilities. The DTN data include 5-min NEXRAD updates and storm cell type, severity, speed, and direction. Meteorlogix provides weather extensions for ESRI's ArcGIS software package. NEXRAIN Corporation (Orangevale, California) provides georeferenced polygon Shapefiles of radar pixels over a given study area. Each radar pixel is given a unique ID field, which can be used to visualize and analyze rainfall data from a separate text or database file. Shapefiles are projected into the user-specified coordinate system. NEXRAIN-2k product is extracted from a national mosaicked data set of 2-km radar data, updated every 15 min. Figure 10.2 shows a Shapefile of the NEXRAIN-2k radar pixels in Colorado State Plane Central (NAD83 ft) over Colorado Springs, Colorado, for a storm event on August 31, 2001. The top-left window shows rainfall distribution at 2 A.M.; the top-right window shows rainfall distribution at 5 P.M.; the bottom-left window shows rainfall distribution at 8 P.M.; the bottom-right window shows the attribute table; the top-center window shows the results of a pixel query; and the bottom-center insert shows a rainfall map overlaid on a street map.

WSI Corporation, a supplier of weather data in the U.S., provides the WSI PRECIP data over these pixels in 15-min time increments. These data can be obtained

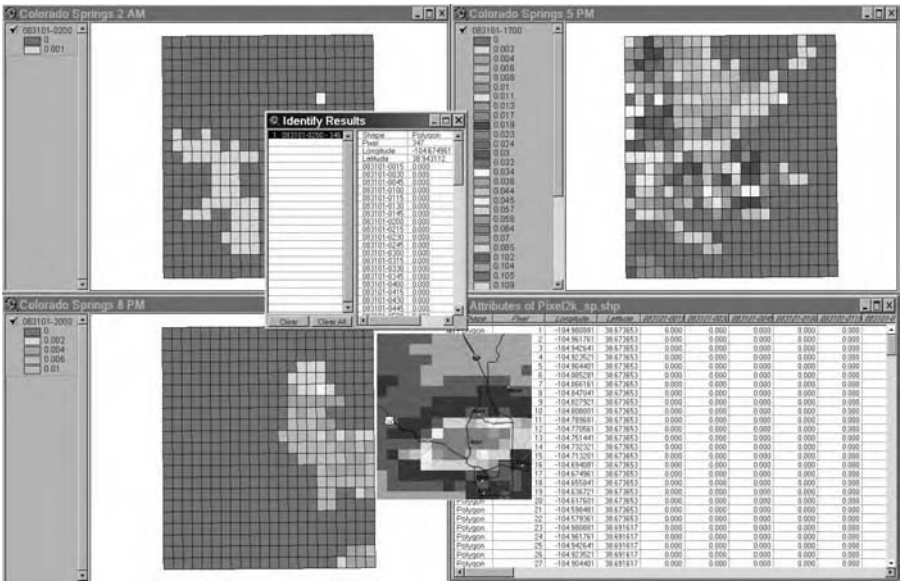


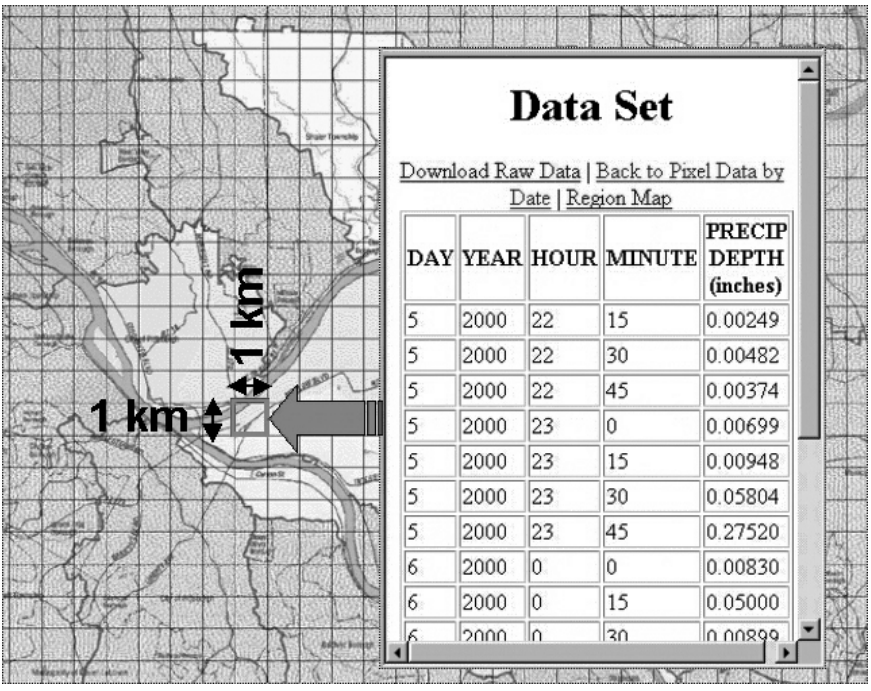
Figure 10.2 Shapefile of the NEXRAIN-2k radar pixels for Colorado Springs, CO, on August 31, 2001. Top left: rainfall at 2 A.M.; top right: rainfall at 5 P.M.; bottom left: rainfall at 8 P.M.; bottom right: attribute table; top center: pixel query window; bottom center: rainfall overlaid on street map.

through commercial vendors and adjusted to reliable ground gauges, if necessary. Hence, the radar rainfall volume adjusted through calibration with ground gauge, combined with high-resolution spatial distribution, provides modelers with a very accurate local estimate of rainfall. These estimates can be used with their flowmetering to characterize the basin of interest (Hamid and Nelson, 2001).

Radar Rainfall Application: Virtual Rain-Gauge Case Study

Precipitation varies with time and across a watershed. GIS can easily create a rainfall contour map from point rainfall values. Unfortunately, point rainfall values are generally scarce due to inadequate rain-gauge density. Many watersheds have few or no rain gauges to accurately measure the spatial rainfall distribution. In this case, a hydrologist may have to use the precipitation data from a nearby watershed. Many watersheds do not have recording-type rain gauges to measure the hourly variation of precipitation. In this case, a hydrologist may need to apply the temporal distribution of a nearby watershed. The recent availability of radar precipitation data via the Internet eliminates these problems by providing subhourly precipitation data anywhere in the watershed.

Three Rivers Wet Weather Demonstration Program (3RWWDP) is a nonprofit organization located in Pittsburgh, Pennsylvania. It was created in 1998 to help Allegheny County (Pennsylvania) municipalities address the region's aging and deteriorating sewer infrastructure to meet the requirements of the federal Clean Water Act.



Data Set

Download Raw Data | Back to Pixel Data by Date | Region Map

DAY	YEAR	HOUR	MINUTE	PRECIP DEPTH (inches)
5	2000	22	15	0.00249
5	2000	22	30	0.00482
5	2000	22	45	0.00374
5	2000	23	0	0.00699
5	2000	23	15	0.00948
5	2000	23	30	0.05804
5	2000	23	45	0.27520
6	2000	0	0	0.00830
6	2000	0	15	0.05000
6	2000	0	30	0.00899

Figure 10.3 Radar-based virtual rain-gauge map and retrieved rainfall data.

Funded by federal grants and local matching funds, 3RWWDP strives to assist communities through education, financial grants, and outreach efforts. In 1999, 3RWWDP developed an Internet-based precipitation atlas for Allegheny County. The atlas consisted of clickable maps that provided 15-min precipitation data every 1 km^2 from the CALAMAR system. CALAMAR is a software and NEXRAD radar service package that integrates high-resolution radar data with precipitation measurements collected from a network of 21 rain-gauges to produce accurate, dependable precipitation maps. Each 1 km $\times$ 1 km pixel of the rainfall atlas is equivalent to a virtual rain-gauge. The users can click on any of the 2276 pixels to retrieve the 15-min rainfall data in HTML or spreadsheet (Excel) format. This is equivalent to having 2276 virtual rain gauges in the County. This powerful tool revolutionizes collection system planning, modeling, and management by providing the critical missing link in urban hydrology — accurate local precipitation data. The rainfall atlas for a watershed near the city of Pittsburgh is shown in Figure 10.3. The figure also shows the clickable pixel cells and retrieved 15-min-interval rainfall data for November 2000 for one pixel.

Vieux & Associates (Norman, Oklahoma) provides basin-averaged radar hyetographs at user-defined time increments (5, 15, 30, 60 min, or daily). Spatial resolutions smaller than 1 km with data precision at 256 levels can be achieved. A real time archive provides access to the radar rainfall estimates. Data are provided in

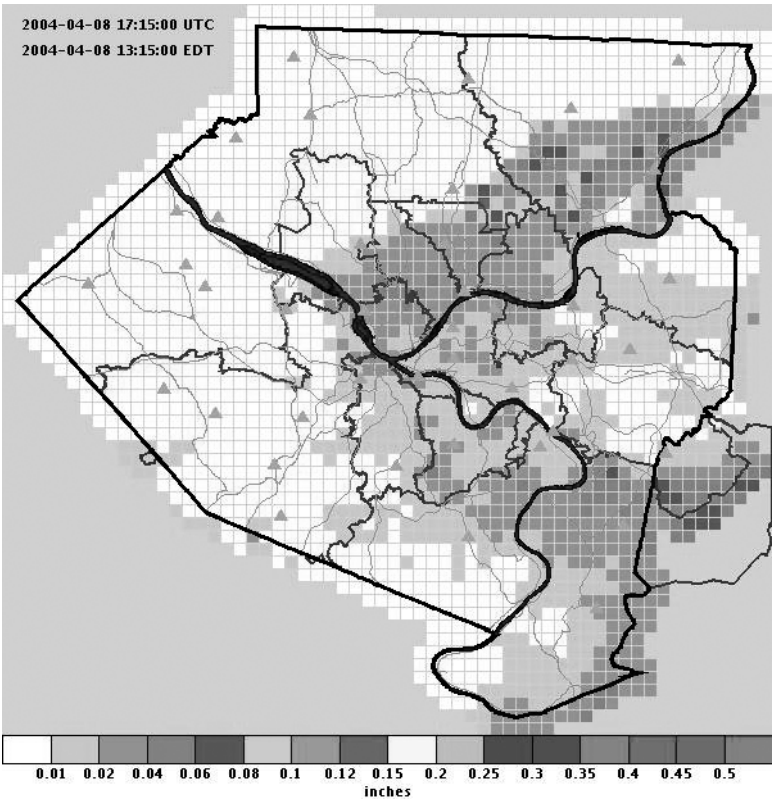


Figure 10.4 RainVieux radar rainfall data for Pittsburgh.

GIS-ready (Shapefile) and hydrologic-model-ready formats. Figure 10.4 shows a sample image from the RainVieux Web site tailored for 3RWDDP. It provides historic and real time radar rainfall measurements that can be queried from a Web site. Radar and rain-gauge measurements are used for cost-effective design of sewer rehabilitation projects in the Pittsburgh Metro area.

Mandated by an EPA consent decree, Miami-Dade Water and Sewer Department (MDWASD), which operates a 317-MGD wastewater treatment plant, implemented a hydraulic computer model to reduce sanitary sewer overflows (SSOs). In order to provide rainfall input to the computer model, MDWASD implemented a virtual rain-gauge system similar to 3RWWDP. A grid-system GIS coverage was created, which yielded 256 virtual rain gauges of 2 km $\times$ 2 km size each. The rainfall data are captured every 15 min and downloaded to a database. The grid is joined with the rainfall database, which allows quantification of impacts of individual rainfall events on specific regions within the system. This approach improves the accuracy of modeled flow estimates that are used to design appropriate SSO control measures (Day, 1998).

FLOW-MONITORING APPLICATIONS

GIS is ideally suited to selecting the best sites for installing flowmeters, water quality samplers, and rain gauges. The following case study illustrates a GIS application for selecting appropriate flow-monitoring sites.

The city of Huntington, WV, has a combined sewer system with 23 combined sewer overflow (CSO) discharge points permitted under the NPDES program. The sewer system is operated by the Huntington Sanitary Board (HSB). One of the CSO requirements specified in the NPDES permit was to monitor each CSO event for cause, frequency, duration, quantity, and quality of flow. As a first alternative to complying with this requirement, the feasibility of monitoring all the 23 CSO locations was studied in 1993. This option required purchase, installation, monitoring, and maintenance of flow monitors, water quality samplers, and rain gauges for each CSO site. This option was ruled out because of its excessive cost, estimated at over a million dollars, and other problems such as adverse environmental conditions in the CSO structures, difficult access issues, and new federal requirements for confined space entry. The second option consisted of a combination of monitoring and modeling. In this option, a representative subset of CSOs can be monitored temporarily to collect sufficient calibration data and develop a calibrated model for each monitored CSO area. Calibrated model parameters could subsequently be applied to the models of unmonitored CSO areas. CSO models can eventually be used to predict the quantity and quality of CSO discharges from observed rainfall data. This option, costing approximately one third of the first option, was preferred by both the HSB and the EPA (Region V), and was selected for implementation.

In 1994, ArcView was used to select a representative subset of the 23 CSO sites to be monitored. Selecting the representative sites is a complex process. The objective is to maximize the extent and value of the data for use in model calibration and provide a uniform distribution of flowmeters, samplers, and rain gauges throughout the system so that all parts of the service area are equally covered. Other factors such as providing coverage to all land uses, known problem CSO areas, and all types of diversion chambers, as well as access for installation and maintenance of monitoring equipment should be thoroughly studied. Using these selection criteria, a total of six sites were finally selected for sampling and monitoring. Figure 10.5 presents an ArcView plot showing the monitored vs. unmonitored CSOs and CSO areas. The map indicates that by monitoring only 25% of the total number of CSOs, approximately 70% of the study area was covered.

SCADA Integration

Real time monitoring systems called supervisory control and data acquisition systems (SCADA) monitor field operations or remote sites for water and wastewater utilities. Also referred to as telemetry, SCADA is a sophisticated means of monitoring, reporting, and controlling the operation of modern water and wastewater systems and treatment plants.

GIS stores the system design and operation data. A link between SCADA and the GIS database can be established to use the GIS for real time decision making. For

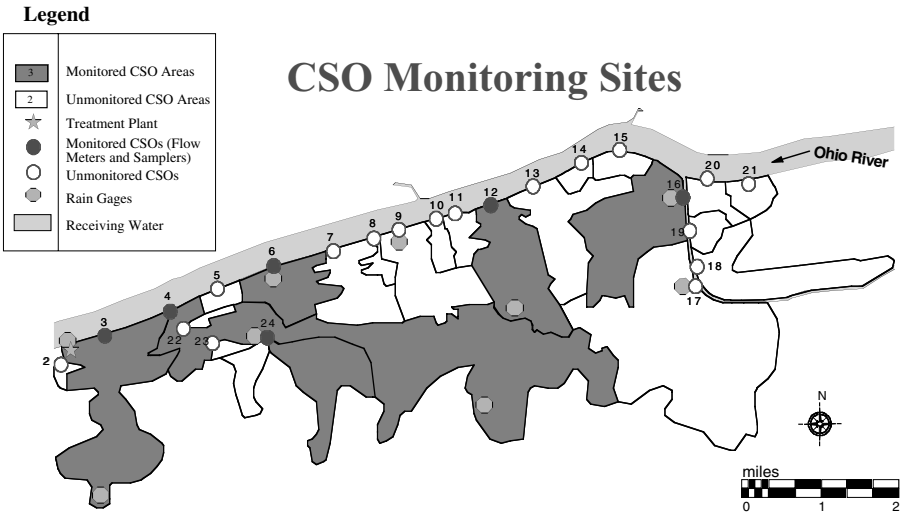


Figure 10.5 Flow-monitoring sites selected by using GIS.

example, real time or near-real time rainfall data can be used in a raster GIS to estimate the subbasin mean aerial rainfall. Real time subbasin rainfall data can be used to estimate real time surface runoff, for flood forecasting, or to estimate real time sewer flows to optimize storage and pumping operations for wet-weather overflow control.

Pidpa is a European company that produces and distributes drinking water to more than 1.1 million people throughout the Antwerp province in Belgium (Flanders). Pidpa used ESRI's ArcIMS technology to create a Web site named GeoLink as the central application to retrieve GIS data linked to corporate data. In the GIS viewer of GeoLink, water storage tanks have a hyperlink that opens the SCADA Web interface. The interface shows the real time water level or 24-hour history of the water level for the selected tank (Horemans and Reynaert, 2003).

SCADA Connect from Haestad Methods (acquired by Bentley Systems in August 2004) is a commercially available software product for dynamically connecting a wide range of SCADA systems with Haestad Methods' WaterCAD, WaterGEMS, and SewerCAD software programs. Using SCADA Connect, information can be automatically queried from the SCADA system and passed to the WaterGEMS database, eliminating manual processes. For example, tank levels and pump and valve settings can be updated in WaterGEMS based on the SCADA database, so that the next time WaterGEMS is run, it will reflect the current field conditions. SCADA Connect has been deployed for the City of Bethlehem, Pennsylvania's Bristol Babcock SCADA system and City of Casselberry, Florida's Citect SCADA system (Haestad Methods, 2004).

NPDES-PERMIT REPORTING APPLICATIONS

The well-known GIS magazine *Geospatial Solutions* holds an "Applications Contest" each year (Geospatial Solutions, 2001a). The entries are judged based on

interest, technology, and importance. The first place winner for the 2001 contest was a stormwater pollutant load modeling application for Los Angeles, California.

As a condition of its NPDES Municipal Stormwater permit, the Los Angeles County Department of Public Works is required to calculate and report pollutant load for more than 256 contaminants. The County maintains a network of stormwater monitoring devices to determine and report pollutant load for more than 256 contaminants. But, in addition to providing data about monitoring system locations and readings, the County also has to estimate the pollutant load to the Pacific Ocean from unmonitored drainage areas. The manual calculation of this information took engineers 250 working hours for an area one quarter of the required study area. The County, therefore, turned to GIS to create an automated model for calculating pollutant load. Relying on an existing public-works GIS, a collaborative team added several data layers to the system to integrate information about rainfall depth, drainage area, land use, imperviousness, and mean countywide concentrations for all the 256 pollutants based on monitoring-station data.

With that information, the team developed a modeling algorithm and refined its operations (the weighting and correlations of different variables) by running models for areas with known pollutant loads and comparing the results with monitoring-station readings. After several iterations, the calibrated model estimated pollutant loads with 95% accuracy. The team then developed an integrated menu-driven interface that walks the user through the process of creating and setting parameters and generating reports of pollutant load.

After specifying that he or she wants to perform a new load run, the user identifies a study area by interactively defining a polygon on the screen. The program then processes information in the background, extracting land-use information for the study area stored in a land-use/imperviousness table. Next, the user selects a rainfall year and event, which specifies the tabular database to use for the modeling, to identify the inches of rainfall at each monitoring station. The application then generates a rainfall grid using the tabular data for the specified rainfall year and event to calculate the pollutant load. The final parameter the user specifies is the pollutants. A dialog box enables interactive selection of the 256 pollutants (the program automatically selects a default set of 64 pollutants). Once all the parameters are defined, the user can calculate the pollutant load and generate maps and reports for print using built-in templates. The program also automatically saves the selected parameters and model results in a database table so that others can retrieve the load run without having to reset all the parameters.

The project team used ESRI's ArcView and Spatial Analyst for GIS analysis and an automated report generation process from Crystal Decisions' Crystal Reports.

MONITORING VIA INTERNET

Integration of GIS and Internet technologies is allowing users to click on a GIS map and see the real time data from any type of Web-connected sensor, from stream gauges to space and airborne Earth-imaging devices. Referred to as the Sensor Web, the sharing of sensor data via the Internet represents a revolution in the discovery

and use of live monitoring data. For example, data from the SAIC (a research and engineering company in San Diego, California) Weather Monitoring System is supplied from an Open GIS Consortium (OGC) Web Collection Service to a mapping client, overlaying geographic information coming from an OGC Web Map Service. The weather data are encoded in OGC Observation and Measurements Language, which is an extension of Geography Markup Language (GML) (Reichardt, 2004).

MONITORING THE INFRASTRUCTURE

From 1998 to 2000, the City of Toronto (Ontario, Canada) collected water pipe and soil samples to investigate the condition of its water distribution network. The City hoped that this effort would explain the causes of water main deterioration and guide the City in its water main rehabilitation efforts. The GIS data used in this project consisted of three layers: (1) Geocoded street center lines, (2) sample locations, and (3) surface soils. Eighty-four pipe and associated soil samples (resistivity, pH, and sulfide content) were collected, tested, and plotted on the street and soil maps using the GIS' address-matching capability. Using the GIS, a spatial analysis of the data collected was performed to identify any trends between the data samples and their locations. For example, water mains located in the clay-silt deposit near lakeshores were suspected to be at greater risk for corrosion, and there was no evident relationship between pH and soil formation (Doyle and Grabinsky, 2003).

Address geocoding is a GIS process that locates street addresses as points on a map, using a reference layer (usually a geocoded street layer). *Address matching* is the comparison of a street address to the address ranges in the reference layer. When an address matches the address range of a street segment, an interpolation is performed to locate and assign real-world coordinates to the address.

A Massachusetts water utility, puzzled by abnormally high lead and copper levels in its distribution system, found that interpretations of water quality problems based on aggregate monitoring data can be very misleading unless analysis is performed at the appropriate scale. Incorporation of system and monitoring-site physical characteristics data as well as the monitoring results into a GIS could have saved them considerable investigatory effort (Schock and Clement, 1995).

USEFUL WEB SITES

Crystal Decisions	www.crystaldecisions.net
Geospatial Solutions	www.geospatial-online.com
Meteorlogix	www.meteorlogix.com/GIS
NCDC Radar Products	www.ncdc.noaa.gov/ol/radar/radarproducts.html
NEXRAIN Corporation	www.nexrain.com
Vieux & Associates	www.vieuxinc.com
WSI Corporation	www.wsi.com

CHAPTER SUMMARY

This chapter presented the methods and examples of GIS applications in installing and maintaining various types of hydraulic characterization monitors (e.g., flowmeters and rain gauges) and querying and analyzing the monitoring data. The examples and case studies presented in this paper indicate that GIS is ideally suited for selecting the best sites for installing the monitors. Once the monitors have been installed, GIS allows user-friendly query of the monitoring data simply by clicking on a map of monitoring sites. GIS facilitates studying the spatial trends in the monitoring data. GIS is especially useful in processing and integrating radar rainfall data with H&H models of sewage collection systems and watersheds.

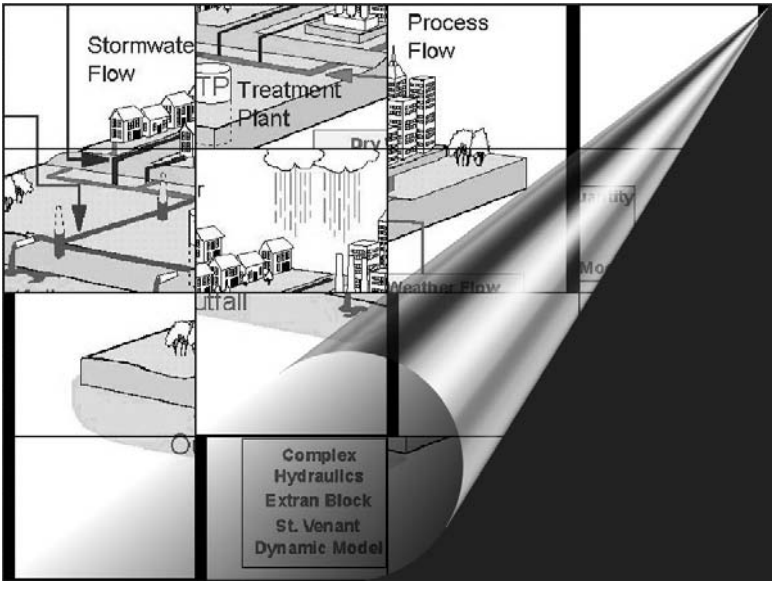
CHAPTER QUESTIONS

1. What are radar rainfall data? How are they used in H&H modeling? How does GIS help in processing them for H&H models?
2. How does GIS help in selecting the feasible flow-monitoring sites?
3. How does GIS allow a trend analysis of spatial patterns in monitoring data?

CHAPTER 11

Modeling Applications

GIS and H&H model integration allows the users to be more productive. Users can devote more time to solving the problems and less time on the mechanical tasks of inputting data and interpreting reams of model output. More than just text outputs, models become automated system-evaluation tools. GIS integration saves time and money.



GIS integration is ideally suited to solve the computer modeling puzzle.

LEARNING OBJECTIVES

The learning objectives of this chapter are to classify the methods of linking water, wastewater, and stormwater system computer models with GIS, and to understand the differences between various linkage methods.

MAJOR TOPICS

- GIS applications in H&H modeling
- Modeling application methods
  - Interchange method
  - Interface method
  - Integration method
- Examples and case studies of the preceding methods

LIST OF CHAPTER ACRONYMS

AML Arc Macro Language
BASINS Better Assessment Science Integrating Point and Nonpoint Sources
COE Corps of Engineers (U.S. Army)
DEM Digital Elevation Model
GUI Graphical User Interface
H&H Hydrologic and Hydraulic
HEC Hydrologic Engineering Center (U.S. Army Corps of Engineers)
HSG Hydrologic Soil Group
HSPF Hydrologic Simulation Program — FORTRAN
NPS Nonpoint Source
NRCS Natural Resources Conservation Service (U.S.)
SCS Soil Conservation Service (U.S.) (Now NRCS)
SWAT Soil and Water Assessment Tool
SWMM Storm Water Management Model
TMDL Total Maximum Daily Load
TR-20 Technical Release 20
VBA Visual Basic for Applications
WMS Watershed Modeling System

This book focuses on the four main applications of GIS, which are mapping, monitoring, modeling, and maintenance and are referred to as the "4M applications." In this chapter we will learn about the applications of the third *M* (modeling).

TEMPORAL-SPATIAL MODELING IN WESTCHESTER COUNTY

GIS software	ArcInfo
Modeling software	Rational method and kinematic wave model

GIS data	Land use (from low-altitude infrared color photography), USGS DEM (1:24,000 scale), NRCS soils (1:12,000 and 1:24,000 scales), streams, and watersheds
Study area	0.14 mi^2 (0.36 km^2) Malcolm Brook watershed, Westchester County, New York
Organization	New York City Department of Environmental Protection

Lateral flow is an explicit component of the kinematic wave routing equation, yet some kinematic wave models such as HEC-HMS do not include lateral flow. The lateral flow component can be accounted for by using GIS. Until recently, it was considered impossible to conduct time-varying computations within a GIS. In this case study, lateral flows were derived from GIS output for each segment of the stream and at each time interval of the rain storm and were routed using the kinematic wave routing method. Rather than a raster GIS that uses a constant cell size, a vector GIS was used to define hydrologic response units that divide the stream channel into segments that vary in size according to the combined characteristics for land use, slope, and soil type. This approach permitted vector-based spatially distributed modeling of stream flow during storm events. GIS was used to map and visualize contributing areas around a stream channel. During each calculation of the discharge, a graphical image of the watershed and contributing areas was captured as a Graphics Interchange Format (GIF) image. A series of these images were displayed in sequence to produce a continuous animation (Gorokhovich et al., 2000).

H&H MODELING

No one said mathematical modeling would be easy, but the preparation of input data and interpretation of output results required by the ever-changing complexities of sophisticated hydrologic and hydraulic (H&H) models have been mind boggling. The good news is that model building and interpretation of results are now easier than ever before thanks to recent advances in computer hardware and software. The assembly of model input data by traditional, manual map measurements is just too time consuming and difficult to justify, now that personal computers are affordable and powerful. It is much easier to see a color-coded map of surcharged sewers to pinpoint the areas of flooding and surcharging than to digest reams of computer output, especially for nonmodelers. Nonmodelers are those people who are not expert modelers but who need to know the modeling results, such as clients of consulting firms, project managers, elected officials, and regulatory agencies. After all, as advocated in Chapter 1 (GIS Applications), GIS technology is an effective means of bridging the gap between information and its recipients.

Rapidly developing computer technology has continued to improve modeling methods for water, wastewater, and stormwater systems. GIS applications provide an accurate and manageable way of estimating model input parameters such as node demands, sewage flows, and runoff curve numbers. GIS-based modeling, as a side benefit, also provides an updated database that can be used for nonmodeling activities such as planning and facilities management.

There are two fundamental requirements in most H&H modeling projects: a suitable model and input data for the model. It is often difficult to select a modeling approach because of trade-offs between models and data. For instance, a detailed model requires a large amount of input data that is often too difficult to obtain or is too expensive. On the other hand, a simple model that requires little data may not provide a detailed insight into the problem at hand. Modelers must, therefore, use an optimal combination of model complexity (or simplicity) and available data. The recent growth in computational hydraulics has made it increasingly difficult for practitioners to choose the most effective computational tool from among a variety of very simple to very complex H&H models. Fortunately, thanks to the advances in GIS applications, creation of input data sets is easier than ever before.

This chapter serves as a guide to help professionals select the most appropriate GIS applications for their modeling needs. It presents an overview of the GIS and computer modeling integration approaches and software. The chapter also shows how to estimate the physical input parameters of H&H models using GIS. The chapter largely uses watershed hydrologic modeling examples to explain modeling integration concepts. However, the integration methods presented here are equally applicable to modeling of water and wastewater systems. Water system modeling applications and examples are presented in Chapter 12 (Water Models). Sewer system modeling applications and examples are presented separately in Chapter 13 (Sewer Models).

APPLICATION METHODS

There are two types of hydrologic models: lumped and distributed. Lumped-parameter models lump the input parameters of a study area over polygons and use vector GIS applications. Distributed models distribute the input parameters of a study area over grid cells and use raster GIS applications. Application of GIS technology to H&H modeling requires careful planning and extensive data manipulation work. In general, the following three major steps are required:

1. Development of spatial database
2. Extraction of model layers
3. Linkage to computer models

H&H models, databases, and GIS applications are critical in efficiently and effectively completing large modeling studies. The models and GIS can be linked to other databases for data sharing purposes. For example, data can be imported from other sources like AutoCAD, edited and modified within the application, and exported or linked to other databases. Database files can be saved in .dbf format and linked or imported into Microsoft Access for further data manipulation. For example, ArcView GIS provides Open Database Connectivity (ODBC) features that can be used to link ArcView tables with other tables and queries in Microsoft Access or other database programs, without actually going through any import and export exercises. Such procedures eliminate the data redundancies and user errors typically associated

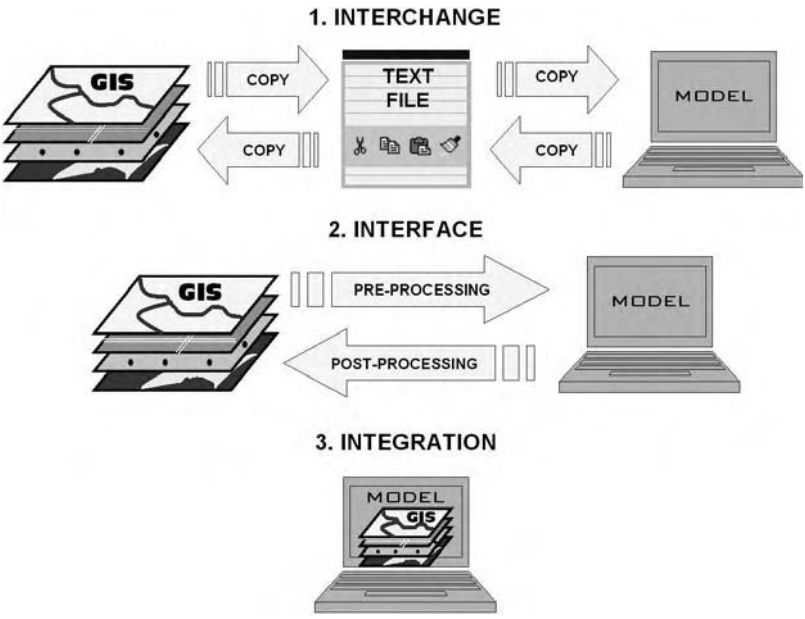


Figure 11.1 Three methods of GIS applications in H&H modeling.

with such cumbersome tasks. This has proven to be beneficial and time-saving during alternative analysis of systems where a large number of scenarios are modeled and reviewed using the same data connectivity and templates for maps (Hamid and Nelson, 2001).

A useful taxonomy to define the different ways a GIS can be linked to computer models was developed by Shamsi (1998, 1999). The three methods of GIS linkage defined by Shamsi are:

1. Interchange method
2. Interface method
3. Integration method

Figure 11.1 shows the differences between the three methods. Each method is discussed in the following text with the help of examples and case studies. As described in Chapter 1 (GIS Applications), advanced applications such as the capability to automatically create a model input file always require writing a computer program using a scripting language like Visual Basic for Applications (VBA).

INTERCHANGE METHOD

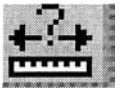
Preprocessing is defined as the transfer of data from the GIS to the model. Postprocessing is defined as the transfer of data from the model to the GIS. The interchange method employs a batch-process approach to interchange (transfer) data

between a GIS and a computer model. In this method, there is no direct link between the GIS and the model. Both are run separately and independently. The GIS database is preprocessed to extract model input parameters, which are manually copied into a model input file. Similarly, model output data are manually copied in a GIS as a new spatial layer for presentation-mapping purposes. Script programming is not necessary for this method, but it may be done to automate some manual operations such as derivation of runoff curve numbers (described in the following text). This is the easiest method of using GIS in computer models and is the most commonly utilized method at the present time.

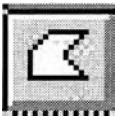
In this method GIS is essentially used to generate model input files and display model output data. Any GIS software can be used in the interchange method. A GIS with both vector and raster capabilities provides more interchange options. Representative examples of the interchange method are described in the following subsections.

Subbasin Parameter Estimation

Most H&H models need input data for subbasin parameters such as area, overland flow width, and slope. If subbasins are represented as polygons, GIS can automatically calculate the area as a layer attribute. Overland flow width can be measured interactively in any direction (e.g., along a stream or a sewer) by using the measurement tools available in most GIS packages. For example, ArcView 3.x provides a Measure Tool



that can be used for on-screen measurement of distance. It also provides a Drawing Tool



that can be used for on-screen measurement of polygon area. Subbasin slope can be estimated from digital elevation models (DEMs) as described in Chapter 4 (DEM Applications).

GIS estimation of subbasin runoff curve number, a critical input parameter in many rainfall-runoff models, is perhaps the best example of the interchange method.

Runoff Curve Number Estimation

Runoff curve numbers are a set of standard empirical curves that are used to estimate stormwater runoff. Accurate estimation of curve number values is critical to accurate runoff modeling because the quantity of runoff is very sensitive to runoff curve number values. GIS estimation of subbasin runoff curve number, a critical input parameter in many rainfall-runoff models, is perhaps the best example of the interchange method. The estimation approach is based on the land use, hydrologic soil group (HSG), and runoff curve number relationships developed by U.S. Natural Resources Conservation Service (NRCS), formerly known as the Soil Conservation Service (SCS). These relationships are available in the form of runoff curve number tables (U.S. Department of Agriculture, 1986). These tables provide runoff curve numbers for a large number of land uses and four hydrologic soil groups: A, B, C, and D. They also list average percent imperviousness values for various land-use classes. Table 11.1 shows selected data from the NRCS tables for percent imperviousness and runoff curve number for typical land-use classes. If existing land-use classifications are not consistent with SCS taxonomy, they may be replaced with an equivalent SCS land-use class shown in Table 11.2.

Table 11.1 SCS Runoff Curve Numbers

Land-Use Class	Average Percent Imperviousness	Runoff Curve Number for Hydrologic Soil Group			
		A	B	C	D
Fully Developed Urban Areas (Vegetation Established)					
Open space (lawns, parks, golf courses, etc.)	—	49	69	79	84
Paved parking lots, roofs, driveways, etc.	—	98	98	98	98
Paved streets and roads	—	98	98	98	98
Dirt roads	—	72	82	87	89
Commercial and business	85	89	92	94	95
Industrial	72	81	88	91	93
Residential: lots 1/8 acres or less (town houses)	65	77	85	90	92
Residential: lots 1/4 acres	38	61	75	83	87
Residential: lots 1/2 acres	25	54	70	80	85
Residential: lots 1 acres	20	51	68	79	84
Residential: lots 2 acres	12	46	65	77	82
Developing Urban Areas					
Newly graded areas		77	86	91	94
Idle lands		77	86	91	94
Nonurban Areas					
Row crops (straight row, poor condition)	—	72	81	88	91
Row crops (straight row, good condition)	—	67	78	85	89
Meadow (continuous grass, no grazing)	—	30	58	71	78
Brush (fair condition)	—	35	56	70	77
Woods (fair condition)	—	36	60	73	79
Farmsteads (buildings, driveways, and lots)	—	59	74	82	86

Table 11.2 Equivalent SCS Land-Use Classes

Land-Use Class	SCS Equivalent Land-Use Class	Average Percent Imperviousness	Runoff Curve Number for Hydrologic Soil Group			
			A	B	C	D
Parks, cemeteries, ball fields, etc.	Open space	0	39	61	74	80
High-density residential	Average lot 506–1012 m^2 (0.125–0.25 acres)	51	69	80	87	90
Medium-density residential	Average lot 1349–2024 m^2 (1/3–1/2 acres)	28	56	71	81	86
Low-density residential	Average lot 4047–8094 m^2 (1–2 acres)	16	49	66	78	83
Schools	Average lot 1349–2024 m^2 (1/3–1/2 acres)	28	56	71	81	86
Agricultural	Row crops average	0	64	75	83	87
Disturbed	Newly graded areas	35	77	86	91	94

A vector layer for subbasin runoff curve numbers is created by overlaying the layers for subbasins, soils, and land use to delineate the runoff curve number polygons. Each resulting polygon should have at least three attributes: subbasin ID, land use, and HSG. The NRCS landuse-HSG-curve number matrix (Table 11.1 or Table 11.2) can now be used to assign runoff curve numbers to each polygon according to its land use and HSG. Polygon runoff curve number values are then area-weighted to compute the mean runoff curve number for each subbasin. These subbasin-runoff curve numbers can now be entered into the model input file. The runoff curve number estimation technique is shown in Figure 11.2 to Figure 11.5. Figure 11.2 shows the layers for subbasins and land use. Figure 11.3 shows the layers for subbasins and HSGs. Figure 11.4 shows the computed runoff curve number polygons. Figure 11.5 shows the average runoff curve numbers for the subbasins.

Some H&H models also need an input for the subbasin percent imperviousness, which can be estimated in GIS using the NRCS runoff curve number tables. A layer for the subbasin percent imperviousness can be created by overlaying the layers for subbasins and land use to delineate the percent imperviousness polygons. The land-use percent imperviousness matrix can then be used to assign percent impervious values to the polygons. Finally, polygon percent imperviousness values are area-weighted to compute the mean percent imperviousness value for each subbasin.

Water Quality Modeling Data Estimation

Some urban-runoff quality models such as the TRANSPORT Block of SWMM need curb length as a model input parameter, which can be estimated from a GIS layer of road or street centerlines. Subbasin curb length can be estimated by performing

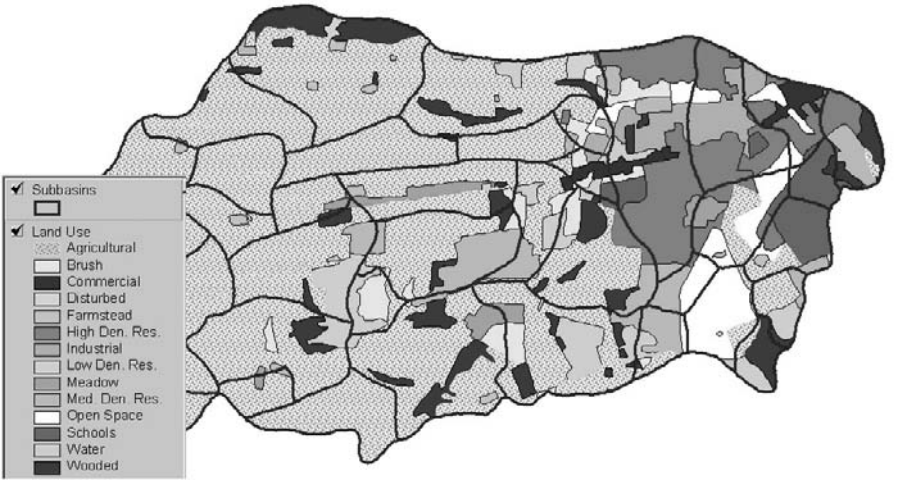


Figure 11.2 Layers for subbasins and land use for runoff curve number estimation.

an overlay analysis of street and subbasin layers. For example, in ArcView 3.x, curb length can be estimated by performing a *theme-on-theme* selection using roads as the target theme and subbasins as the selector theme. Alternatively, for large areas such tasks can be handled more efficiently in ArcView's Network Analyst extension, which has been designed for transportation network applications. Network Analyst is an optional extension that must be purchased separately.

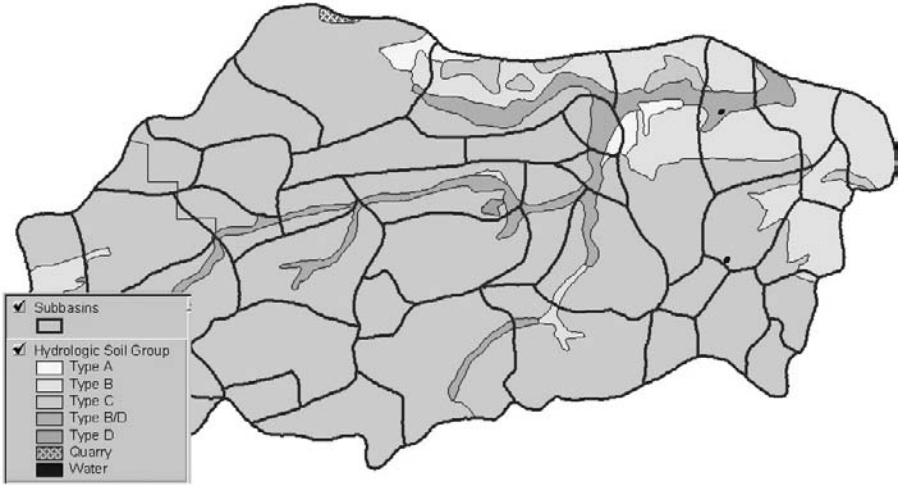


Figure 11.3 Layers for subbasins and HSGs for runoff curve number estimation.

Figure 11.4 Computed runoff curve number polygons.

Demographic Data Estimation

Demographic data can be used for the following modeling tasks:

1. Estimation of quantity and quality sanitary sewage flow
2. Estimation of present and potential future development
3. Creation of auxiliary layers for visualization and presentation purposes

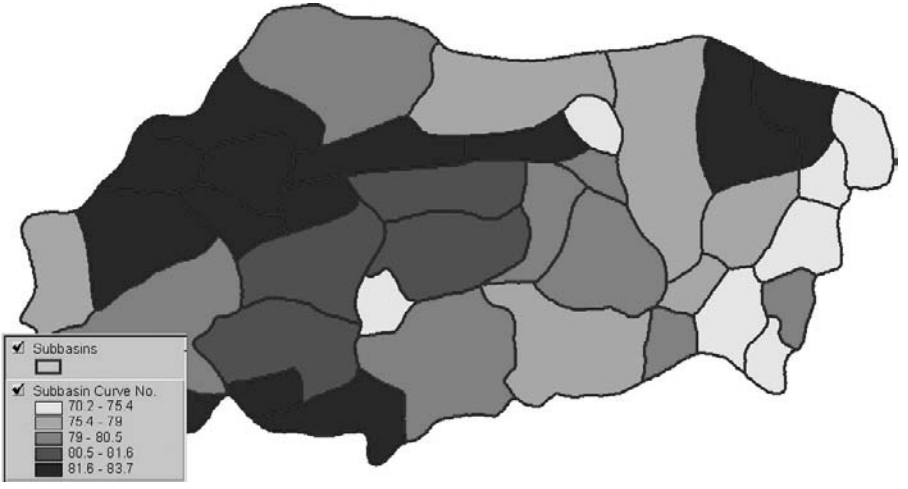


Figure 11.5 Subbasin runoff curve number map.

The first task is required when modeling both combined and sanitary sewer areas in urban sewersheds. For example, SWMM's TRANSPORT Block (Version 4.x) needs inputs for some or all of the following demographic parameters:

1. Dwelling units
2. Persons per dwelling unit
3. Market value of average dwelling unit
4. Average family income

A GIS can compute such demographic parameters from the U.S. Census Bureau's TIGER data (Shamsi, 2002). To estimate subbasin demographic parameters, census-block attribute data files from TIGER files are linked to the polygon topology of the census blocks. Only those attributes pertinent to the model are retained within the working layers. The population (or other demographic parameters) within each subbasin is estimated by determining the census blocks or their portions making up each subbasin. Block data are area-weighted to estimate the mean or total subbasin values of demographic parameters. For example, this can be done in ArcView 3.x from *polygon-on-polygon* selection using census blocks as the target theme and subbasins as the selector theme. Figure 11.6 shows an ArcView map of subbasins, census blocks, and census block groups for a SWMM-based sewer system model.

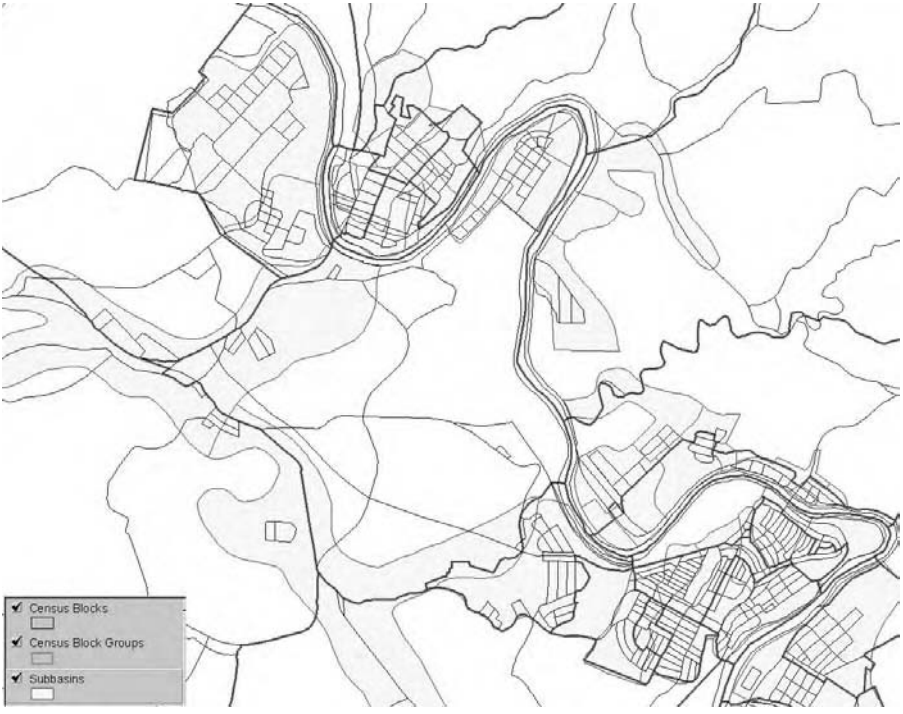


Figure 11.6 Sample census block and block group layers for estimating subbasin population.

Land-Use Data Estimation

Dry- and wet-weather flows from sewersheds depend on land use. Most rainfall-runoff models, therefore, need land-use information. In some models such as SWMM's TRANSPORT Block, land-use type is input directly. In others, such as Penn State Runoff Model (PSRM) and SWMM's RUNOFF Block, land-use type is needed to derive certain model parameters such as percent imperviousness. TRANSPORT needs input data for the the land-use designation of modeled subbasins. The subbasins should be classified as having one of the five land-use classes:

1. Single-family residential
2. Multifamily residential
3. Commercial
4. Industrial
5. Open space

Most land-use maps and layers have more than these five land-use classes. However, this is not a problem in GIS because the existing classes can be easily consolidated into the five land-use classes mentioned above. Figure 11.7 shows an ArcView 3.x screenshot for subbasin and land-use layers for a SWMM-based sewer system model. The land-use layer was prepared from classification of 30-m-resolution Landsat Thematic Mapper (TM) satellite imagery of the study area. Initially, the layer contained 19 land-use classes. Subsequently, the initial land-use classes were reclassified into the five land-use types required for TRANSPORT input. In order to refine the land-use classification, the layer was combined with the census-block layer

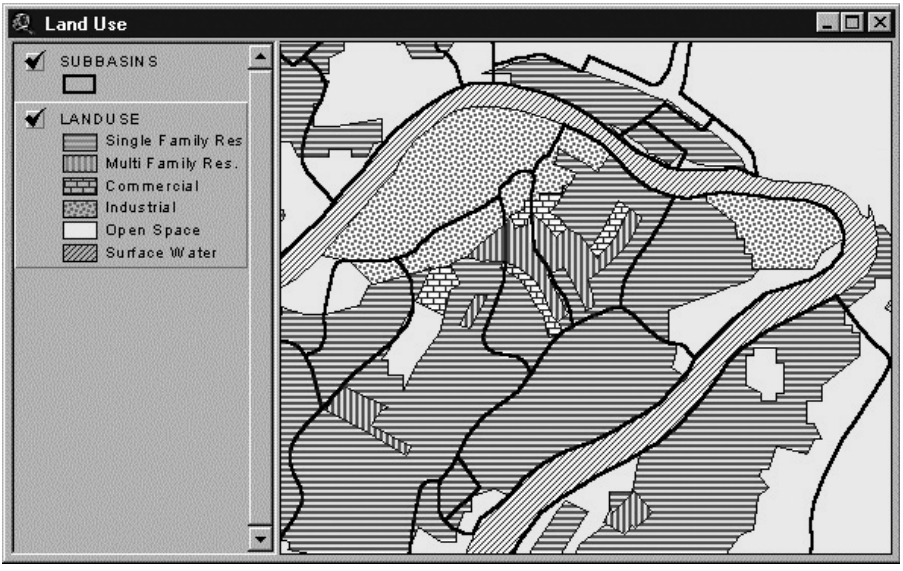


Figure 11.7 Estimating land use using interchange method.

to determine where population densities indicated residential districts. This was done in two ways. First, all nonresidential classed areas were examined to determine if population densities in the local census blocks were below or above a threshold of ten persons per acre. This threshold was chosen as a conservative breakpoint that would allow a reasonable portion of the classification to be correctly reclassified. The areas corresponding to any blocks with densities above the threshold were changed to the residential class. Second, the census blocks of all residential areas were examined to determine in which blocks single-family or multi-family residential units were the majority. On this basis polygons of the residential class were divided into single-family residential or multi-family residential classes.

INTERFACE METHOD

The interface method provides a direct link to transfer information between the GIS and a model. The interface method consists of at least the following two components: (1) a preprocessor that analyzes and exports the GIS data to create model input files and (2) a postprocessor that imports the model output and displays it as a GIS theme. Processing of the model input and output files requires computer programming using the GIS software's scripting language (such as Avenue or VBA). User routines and scripts are incorporated in the GIS by calling them through new pull-down menus and icons.

The interface method basically automates the data interchange method. The automation is accomplished by adding model-specific menus or buttons to the GIS software. The model is executed independently of the GIS; however, the input file is created in the GIS. The main difference between the interchange and interface methods is the automatic creation of the model input file. In the data interchange method, the user finds a portion of a file and copies it. In the interface method, an interface automates this process, so that the pre- and postprocessor can find the appropriate portion of the file automatically.

Learning with examples is an effective method, so let us begin with an interface example. Let us assume that we want to create an input file for a hydrologic model by exporting GIS data to an ASCII text file. Our watershed GIS has the following data:

- Subbasins with attributes: ID, area, overland flow width, slope, and percent imperviousness.
- Reaches (streams) with attributes: ID, upstream subbasin, downstream subbasin, type (natural stream, concrete channel, or concrete pipe), condition (best, good, fair, or bad), depth, width, length, upstream elevation, and downstream elevation.

Our hydrologic model requires a text file with the following input parameters:

- Subbasin data: ID, area, overland flow width, slope, and percent imperviousness.
- Conduit data: ID, upstream subbasin, downstream subbasin, type, depth, width, length, slope, and roughness coefficient.

Table 11.3 Look-up Table for Reach Roughness Coefficient

	Reach Condition			
Reach Type	**Best**	**Good**	**Fair**	**Bad**
Natural stream	0.025	0.028	0.030	0.033
Concrete channel	0.012	0.014	0.016	0.018
Concrete pipe	0.012	0.013	0.015	0.016

Note that our GIS does not have reach attributes for slope and roughnesss coefficient, but they can be estimated in the GIS from the given attributes. For example, slope can be estimated from the upstream and downstream elevations and length using the following formula:

$$slope = \frac{upstream\ elevation - downstream\ elevation}{length}$$

The roughness coefficient can be estimated using a lookup table like Table 11.3. For example, if the reach type is natural and condition is good, then GIS can estimate the roughness coefficient as 0.028. Now that all the required subbasin and conduit parameters are available in the GIS database, the required fields of subbasin and reach theme tables can be saved as text files.

Let us further assume that we want to create model output maps in a GIS by importing model output files into it. This task can be easily accomplished by linking the model output file to a GIS layer using a common attribute (subbasin and reach IDs). For example, a thematic map of stream flow can be created by linking the modeled stream flow to the streams layer. First of all, a database table of the relevant results (stream flow) should be created from the ASCII output file. The output database file can then be joined to the streams attribute table. The linked output results can now be queried or classified with a legend to make a thematic map.

Typical modeling steps involved in using the interface method are listed below:

1. Start your GIS.
2. Export model input data from GIS to a text or database file (preprocessing).
3. Exit the GIS.
4. Start your model.
5. Import GIS data (saved in Step 2) in the model.
6. Run the model.
7. Exit the model.
8. Reenter the GIS and import the model output (postprocessing).

It can be seen from these steps that the interface method does not run the model inside the GIS. The model must be run outside the GIS by the user. Some interface examples are given in the following text. Additional examples are presented in Chapter 12 and Chapter 13.

HEC-GEO Interface

The Hydrologic Engineering Center (HEC) is an office of the U.S. Army Corps of Engineers (COE) established to support the nation's hydrologic engineering and water resources planning and management needs. To accomplish this goal, HEC develops state-of-the-art, comprehensive computer programs that are also available to the public. HEC-1 and HEC-2 are COE's legacy DOS programs for hydrologic and hydraulic modeling, respectively. Recently HEC-1 and HEC-2 have been replaced with Windows programs called the Hydrologic Modeling System (HEC-HMS) and River Analysis System (HEC-RAS), respectively. HEC Geo-HMS and HEC Geo-RAS have been developed as geospatial hydrology toolkits for HEC-HMS and HEC-RAS users, respectively. They allow users to expediently create hydrologic input data for HEC-HMS and HEC-RAS models. Free downloads of these programs are available from the HEC software Web site.

HEC-GeoHMS

HEC-GeoHMS is an ArcView 3.x GIS extension specifically designed to process geospatial data for use with HEC-HMS. It allows users to visualize spatial information, delineate watersheds and streams, extract physical watershed and stream characteristics, perform spatial analyses, and create HEC-HMS model input files. HEC-GeoHMS uses ArcView's Spatial Analyst Extension to develop a number of hydrologic modeling inputs. Analyzing digital terrain information, HEC-GeoHMS transforms drainage paths and watershed boundaries into a hydrologic data structure that represents watershed's response to precipitation. In addition to the hydrologic data structure, capabilities include the development of grid-based data for linear quasi-distributed runoff transformation (ModClark), the HEC-HMS basin model, and the background map file.

HEC-GeoHMS provides an integrated work environment with data management and customized toolkit capabilities, which includes a graphical user interface (GUI) with menus, tools, and buttons. The program features terrain-preprocessing capabilities in both interactive and batch modes. Additional interactive capabilities allow users to construct a hydrologic schematic of the watershed at stream gauges, hydraulic structures, and other control points. The hydrologic results from HEC-GeoHMS are then imported by HEC-HMS, where simulation is performed. HEC-GeoHMS works with Windows 95/98/NT/2000 operating systems.

HEC-GeoRAS

HEC-GeoRAS (formerly named AV/RAS) for ArcView is an ArcView 3.x GIS extension specifically designed to process geospatial data for use with HEC-RAS. The extension allows users to create an HEC-RAS import file containing geometric attribute data from an existing digital terrain model (DTM) and complementary data sets. HEC-GeoRAS automates the extraction of spatial parameters for HEC-RAS input, primarily the 3D stream network and the 3D cross-section definition. Results exported from HEC-RAS may also be processed in HEC-GeoRAS. ArcView 3D

Analyst extension is required to use HEC-GeoRAS. Spatial Analyst is recommended. An ArcInfo 7.x version of HEC-GeoRAS is also available.

Haestad Methods' HEC-GIS system combines COE's HEC-RAS floodplain modeling program with ArcInfo. HEC-GIS is an Arc Macro Language (AML) program that allows importing floodplain cross-section data from a DEM within ArcInfo, in the form of triangular irregular networks (TINs). It also allows importing HEC-RAS water surface profiles back into HEC-GIS to perform additional spatial analyses and mapping.

Watershed Modeling System

Watershed Modeling System (WMS) is a comprehensive hydrologic modeling package developed by the Environmental Modeling Research Laboratory (EMRL) of Brigham Young University in cooperation with the U.S. Army COE Waterways Experiment Station (WES) (WMS, 2000; Nelson et al., 1993). The focus of WMS is to provide a single application that integrates digital terrain models with industry-standard lumped-parameter hydrologic models such as HEC-1 and TR-20. In this way, hydrologic data developed in ArcInfo, ArcView, or WMS can be directly linked to many common hydrologic models. WMS includes an interface to HEC-1, TR-20, and the Rational Method programs used by many hydrologic engineers to model the rainfall-runoff process.

WMS can import and export GIS data. WMS' GIS interface is accomplished through its map module. Map module data is structured after ArcInfo 7.x, which makes it easy to import and use vector data created inside ArcInfo or any other GIS that can export an ArcInfo GENERATE file. WMS includes an ArcView 3.x extension that facilitates the movement of data from GIS to WMS for hydrologic model development. The interface allows for automatic development of a topological model from TINs. Furthermore, any geometric parameters computed by WMS are supplied to the corresponding model input fields. The extension preprocesses the GIS data to export hydrologic modeling data. WMS then reads a single file (describing the relevant themes exported) to obtain data from the GIS. These data can then be used in WMS to create an initial model for any of the hydrologic programs supported by WMS. WMS itself is integrated with the hydrologic models. Data entry for the hydrologic model — including rainfall, job control, or any other parameters not defined as attributes in GIS Shapefiles — can be completed using WMS' hydrologic modeling interface. WMS can be used to postprocess and then import model results in the GIS software. A screenshot of the WMS interface is shown in Figure 11.8. Unlike some HEC programs such as HEC-HMS and HEC-RAS, WMS is not free. A demo version can be downloaded from the BYU Web site.

GISHydro Modules

GISHydro is a collection of programs, demos, exercises, data, reports, and information designed to support the use of GIS in hydrology and water resources studies. These materials are mostly drawn from the work of Professor David R. Maidment and a community of his graduate students, research scientists, and faculty

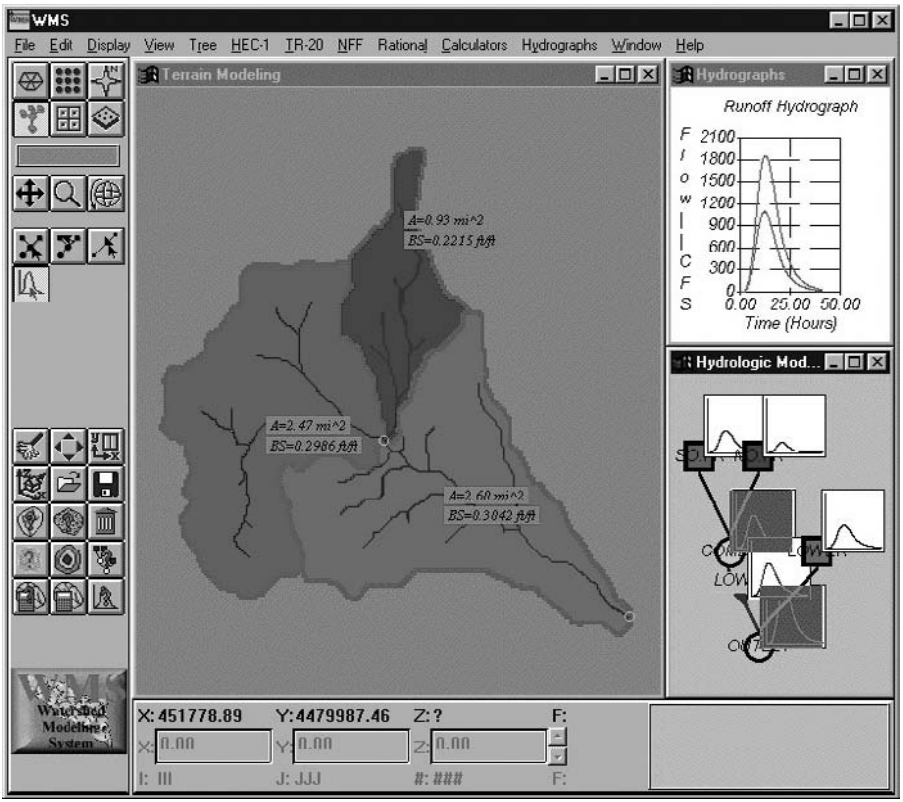


Figure 11.8 WMS interface.

in the Center for Research in Water Resources (CRWR) and the Department of Civil Engineering of the University of Texas at Austin. Initially, CRWR developed a prototype Spatial Hydrology Analysis System (Maidment, 1997) consisting of eight GISHydro modules. These modules, more than just computer programs, are a methodological approach to solving specific types of hydrologic problems using GIS. They contain ArcView and ArcInfo extensions, Avenue and AML scripts, sample data, user instructions, and example exercises. Two GISHydro modules are described in the following subsections:

GISHydro Prepro

This module is a procedure for connecting GIS to external models. Consisting of Avenue and AML scripts, it is designed to transform a subbasins-and-streams network into a schematic model of the hydrologic elements of a watershed. It was initially developed as a preprocessor for COE's new Hydrologic Modeling System (HEC-HMS), which replaced the old HEC-1 model. GH-Prepro takes stream and subbasin GIS data layers and analyzes them to produce a schematic model that

defines and spatially connects seven types of hydrologic elements in the landscape: subbasins, reaches, junctions, diversions, reservoirs, sources, and sinks.

GISHydro Runoff

This module provides articles, reports, exercises, and a sample ArcView 3.x extension to compute watershed parameters and runoff at the outlet of a watershed. The ArcView extension uses the Texas Department of Transportation (TxDOT) statewide regional rural regression equations for estimating stormwater runoff. Figure 11.9 shows a screenshot of the TxDOT extension (Olivera and Bao, 1997).

ArcInfo Interface with HEC Programs

Another example of the interface method is the ArcInfo and HEC-1 and HEC-2 interface developed by Woolpert, Inc. (Phipps, 1995). In HEC-1, the interface model uses vector layers for streams and open channels and the ArcInfo networking function to show the locations of and the relationships between different types of flows: sheet flow, shallow channel flow, and concentrated channel flow. The ArcInfo and HEC-1

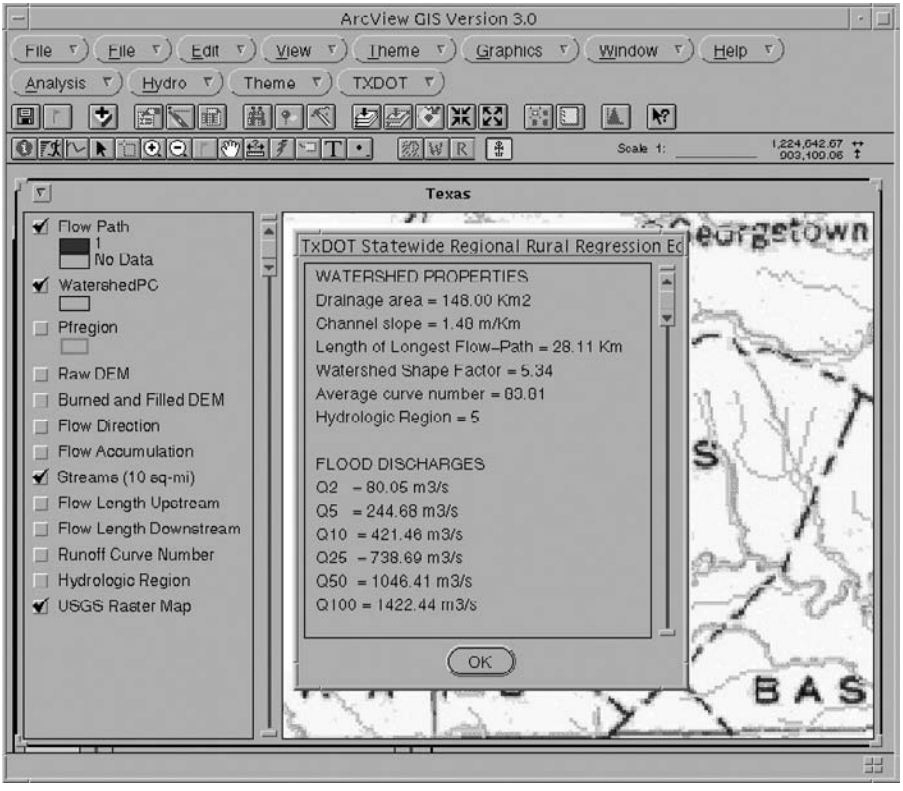


Figure 11.9 GISHydro TxDOT extension.

linkage allows users to see graphically, on a map, how fast the three flow types are moving at different locations within the model (volume/time). The unit hydrographs generated in the HEC-1 hydrologic model are then used to feed the HEC-2 hydraulic model. The ArcInfo and HEC-2 interface allows graphical display of the 2-, 10-, 25-, 50-, 100-, and 500-year frequency flood profiles for the streams and open channels within the watershed. It was found that this interface allowed for easier model development and calibration, which resulted in substantial savings in the cost of modeling tasks.

Intermediate Data Management Programs

Some commercial modeling products use an intermediate data management program to facilitate data transfer between the GIS and a model. This program is written specifically to import data from a variety of common third-party GIS software and export data to a model data set. Under certain conditions this approach could be defined as an interface, but generally it is not. In relation to the Shamsi taxonomy, this method lies between the interchange and interface methods (Heaney et al., 1999). Haestad Methods' GIS connection wizards for the company's WaterCAD, SewerCAD, and StormCAD programs (described elsewhere in this book) are typical examples of this method. Figure 11.10 shows WaterCAD's Import Shapefile wizard,

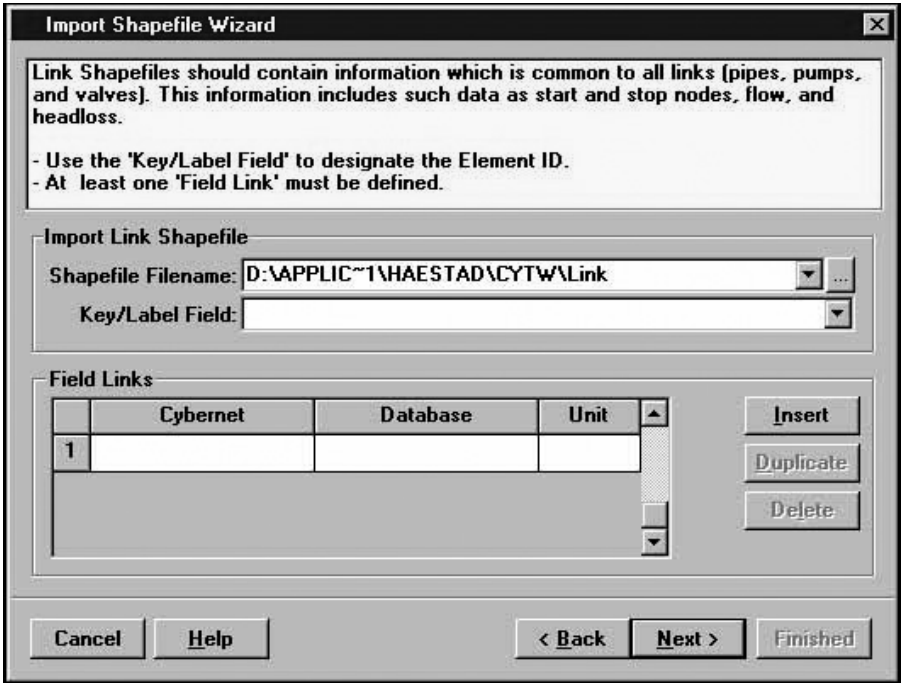


Figure 11.10 WaterCAD's Import Shapefile wizard.

which is used to import WaterCAD model input parameters from ArcView GIS Shapefiles.

Interface Method Case Study

GIS software	ArcInfo
Modeling software	Rhea, an orographic precipitation model developed at the Colorado State University (Rhea, 1977). The Rhea model simulates the interaction of air layers with underlying topography by allowing vertical displacement of the air column while keeping track of the resulting condensate or evaporation.
Other software	Data Explorer (DX), a Scientific Visualization System (SVS) software from IBM
GIS data	DEM grid, city locations, rivers, and rain-gauge locations
GIS format	Raster and vector
Study area	20,530 km^2 Gunnison River Basin in southwestern Colorado
Organization	U.S. Geological Survey

Very often, the software required to solve complex hydrologic problems is available only in separate pieces rather than as an integrated package. Simulation models and GIS and SVS packages are individually strong in specific aspects of spatial analysis but inadequate in cross-functional analysis of complex hydrologic problems. An integrated system consisting of a hydroclimatic simulation model (Rhea), a GIS (ArcInfo), and an SVS (DX) was developed to facilitate model development, calibration, and verification and visual representations of precipitation across space and time. AML tools translated ArcInfo data into a format that could be read by DX. The converted DX file was used to generate animated, interactive displays using DX software. The integrated system provided improved understanding of hydroclimatic processes in mountainous regions. An additional benefit of the integrated system, the value of which is often underestimated, is the improved ability to communicate model results (Hay and Knapp, 1996).

INTEGRATION METHOD

In the interface method, options for data editing and launching the model from within the GIS software are not available. An interface simply adds new menus or buttons to a GIS to automate the transfer of data between a computer model and a GIS. GIS integration, on the other hand, is a combination of a model and a GIS such that the combined program offers both GIS and modeling functions. This method represents the closest relationship between a GIS and a model. Two integration approaches are possible:

1. GIS-based integration: In this approach, modeling modules are developed in or are called from within a GIS. All the four tasks of creating model input, editing

data, running the model, and displaying output results are available in the GIS. There is no need to exit the GIS to edit the data file or run the model. Because development and customization tools within most GISs provide relatively simple programming capability, this approach provides limited modeling power. It offers full GIS functionality and simplistic modeling capability. It is often more popular because of the availability of computer code for a large number of public-domain legacy computer models, such as SWMM and EPANET, which can be easily customized for GIS integration.

2. Model-based integration: In this method, GIS modules are developed in or are called from a computer model. It provides GIS-like functionality to computer models. Because it is difficult to program all the GIS functions into a computer model, this approach provides limited GIS capability. In this method, GIS is essentially used as a mapping tool. Sophisticated data management and data visualization capabilities are not present. PCSWMM GIS [described in Chapter 13 (Sewer Models)] and Runoff97 (Ritter and Gallie, 1998) are examples of the model-based integration method. In this method, GIS and models are integrated through a common interface that requires extensive computer programming using conventional computer languages (e.g., FORTRAN), and/or GIS macro and scripting languages (e.g., AML, Avenue or VBA).

Representative examples of the integration method are given in the following subsections.

EPA's BASINS Program

EPA's BASINS program is the best example of GIS integration with multiple models.

The best example of GIS integration with multiple models is the Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) program. Originally released in September 1996, BASINS was developed by TetraTech, Inc., for the U.S. Environmental Protection Agency (EPA). BASINS supports the development of total maximum daily loads (TMDLs) using a watershed-based approach that integrates both point and nonpoint sources. BASINS integrates ArcView 3.x GIS with national watershed data and state-of-the-art environmental assessment and modeling tools into one convenient package. The heart of BASINS is its suite of interrelated components essential for performing watershed and water quality analysis. These components are grouped into the following five categories:

1. National databases
2. Assessment tools for evaluating water quality and point source loadings at a variety of scales
3. Utilities, such as land-use and DEM reclassification and watershed delineation
4. Watershed and water quality models including NPSM, TOXIROUTE, and QUAL2E
5. Postprocessing output tools for interpreting model results

What Is TMDL?

According to EPA, TMDL is defined as "a calculation of the maximum amount of a pollutant that a water body can receive and still meet water quality standards, and the allocation of that amount to the pollutant's sources." Technically, TMDL can be defined with the help of the following equation:

$$\text{TMDL} = \text{WLAP} + \text{LAN} + \text{RC} + \text{FOS}$$

where:

WLAP = waste load allocations for point sources
LAN = load allocations for nonpoint sources
RC = reserve capacity for future growth
FOS = factor of safety

BASINS integrates three models within an ArcView GIS environment: NPSM (HSPF), TOXIROUTE, and QUAL2E. Nonpoint Source Model (NPSM) estimates land-use-specific NPS loadings for selected pollutants at a watershed (cataloging unit or user-defined subwatershed) scale. The model uses landscape data such as watershed boundaries and land-use distribution to automatically prepare many of the required input data parameters. NPSM combines a Windows-based interface with EPA's HSPF and is linked to ArcView. HSPF is a model for developing NPS loads and is the backbone of the BASINS modeling system. TOXIROUTE is a screening-level stream-routing model that performs simple dilution and decay calculations under mean- or low-flow conditions for a stream system. QUAL2E is a one-dimensional, steady-state water quality and eutrophication model. It is integrated with ArcView through a Windows-based interface. It allows fate and transport modeling for both point and nonpoint sources. Both TOXIROUTE and QUAL2E can use NPS loadings calculated by HSPF.

Overcoming the lack of integration, limited coordination, and time-intensive execution typical of more traditional assessment tools, BASINS makes watershed and water quality studies easier by bringing key data and analytical components together under one roof. The simulation models run in a Windows environment using data input files generated in ArcView. By using GIS, a user can visualize, explore, and query to bring a watershed to life. This allows the user to assess watershed loadings and receiving-water impacts at various levels of complexity. ArcView geographic data preparation, selection routines, and visual output streamline the use of the models. A postprocessor graphically displays model results. Figure 11.11 shows a screenshot of the BASINS program showing an example of delineated watersheds.

BASINS is distributed on a CD-ROM containing the BASINS program and data for an EPA region of interest. It can also be downloaded from the BASINS Web site. BASINS Version 1.0 was released in September 1996 (Lahlou, et al., 1996) for ArcView Version 2.1. BASINS Version 2.0 was released in December 1998 for ArcView 3.0. It included a national data set on the attributes listed in Table 11.4 (Battin et al., 1998). BASINS Version 3.0 was released in 2001 for ArcView 3.2.

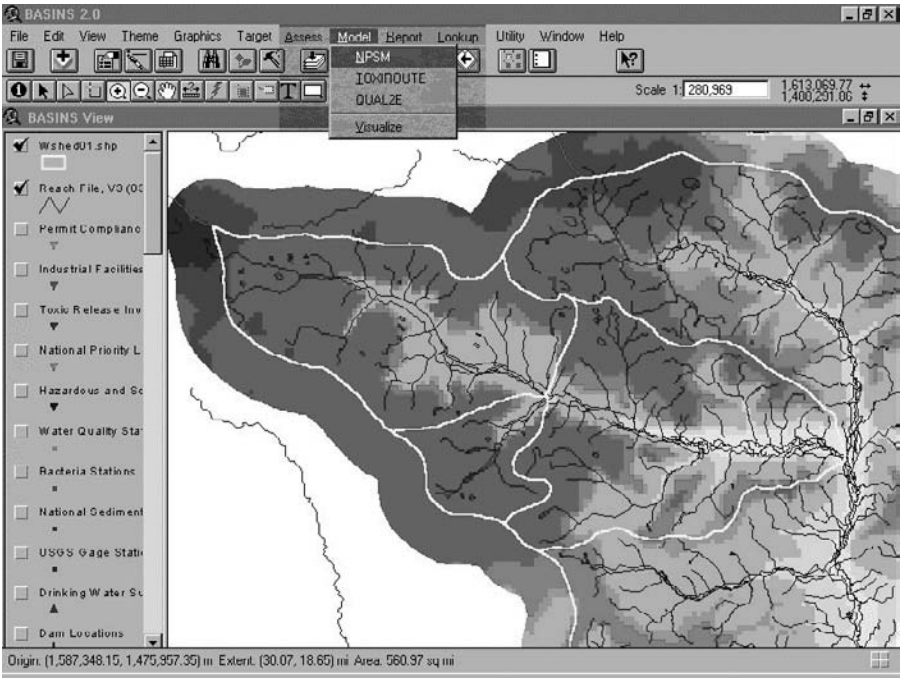


Figure 11.11 BASINS screenshot showing an example of delineated watersheds.

Table 11.4 BASINS Database Attributes

Spatially Distributed Data	
Land use/land cover	USGS hydrologic unit boundaries
Urbanized areas	Drinking water supplies
Populated place location	Dam sites
Reach File, Version 1 (RF1)	EPA region boundaries
Reach File, Version 3 (RF3)	State boundaries
Soils (STATSGO)	County boundaries
Elevation (DEM)	Federal and Indian lands
Major roads	Ecoregions
Environmental Monitoring Data	
Water quality monitoring station summaries	USGS gauging stations
Water quality observation data	Fish and wildlife advisories
Bacteria monitoring station summaries	National Sediment Inventory (NSI)
Weather station sites (477)	Shellfish contamination inventory
Clean water needs survey	—
Point-Source Data	
Permit compliance system	Industrial Facilities Discharge (IFD) sites
Toxic Release Inventory (TRI) sites	Superfund national priority list sites
Mineral availability system/mineral industry location	Resource Conservation and Recovery Act (RCRA) sites

BASINS 2.0 included a watershed delineation tool, a land-use reclassification tool, an HSPF interface, a watershed characterization report on STORET water quality data, DEM elevation data, and USEPA Reach File Version 3 Alpha (RF3 Alpha). Using the import tool, users can add their own data for watershed boundaries, streams, elevation, land use, and observed water quality into the BASINS system. Using the land-use reclassification tool, users can reclassify part of the land-use theme or the entire theme interactively to create more detailed levels of land use. This feature can be used to study the potential impact of land-use changes on water quality. The watershed-delineation tool allows users to interactively subdivide a USGS 8-digit watershed into smaller subwatersheds using mouse point-and-click inputs. Subdelineated watersheds and underlying data are then available for more detailed modeling.

BASINS 3.0 is the latest (year 2001) release at the time of this writing. This release includes additional functional capabilities as well as an updated and expanded set of national data layers. New data and functions include 1° DEM grids, an automatic delineation tool that creates watershed boundaries based on DEM grids, and a new watershed report function for land use, topography, and hydrologic response units. In addition, a new watershed model, the Soil and Water Assessment Tool (SWAT), has been added. SWAT is a watershed-scale model developed to predict the impact of land management practices on water, sediment, and agricultural chemical yields in large complex watersheds with varying soils, land use, and management conditions over long periods of time.

BASINS 3.0 also demonstrates an important change in the development philosophy of BASINS. It is distributed as a core system with several ArcView extensions. All BASINS data management utilities, assessment tools, report tools, and models were converted as extensions. This modular and open architecture offers numerous advantages both to users and developers. For instance, users may customize their BASINS project to include only extensions that are needed. Moreover, it will be easier for developers to maintain and provide updates of individual extensions because it will not involve changing and reissuing a patch or new version for the entire system. This makes it easier for users to upgrade their system. It also encourages the development of extensions for BASINS by users and developers.

BASINS 3.0 installation software contains the new BASINS Version 3.0 system and the tutorial data set (HUC-05010007). In addition to BASINS 3.0, the installation program installs WinHSPF, GenScn, WDMUtil, and PLOAD. WinHSPF is a new interface to HSPF Version 12, which is replacing NPSM from BASINS 2.0. GenScn is a model postprocessing and scenario-analysis tool that is used to analyze output from HSPF and SWAT. WDMUtil is used to manage and create watershed data management files (WDMs) that contain the meteorological data and other time-series data used by HSPF. PLOAD is a GIS and spreadsheet tool to calculate nonpoint sources of pollution in watersheds.

EPA plans to release an update (Version 3.1) in 2004 on the existing ArcView 3.x platform. EPA is currently working on pieces of BASINS for the ArcGIS environment. After the BASINS 3.1 release, EPA plans to port it onto the ArcGIS platform based on the work done by the Center for Research in Water Resources at the University of Texas at Austin, Water Resources Consortium, and ESRI. Future releases of BASINS may include EPA's SWMM program, automatic model calibration, and a reach-indexing tool

for georeferencing. A redesign of the BASINS system using RDBMS, Internet, MapObjects, and Java technologies is also anticipated. This move toward a more fully integrated GIS and database system will provide many more BASINS applications. The watershed tool will evolve into a more component-based architecture, so that users will need to install only the components required for their specific analytical needs, including their own legacy models (Whittemore and Beebe, 2000).

BASINS is a valuable tool that integrates environmental data, analytical tools, and modeling programs to support the development of cost-effective approaches to environmental protection. However, BASINS' seamless integration power is not without some disadvantages. BASINS' simplistic modeling approach and user-friendly tools may encourage inexperienced users to become instant modelers, which could be undesirable. The inexperienced users' overreliance on default model segmentation and model coefficients, without an understanding of the underlying assumptions, may result in inaccurate and unrealistic TMDLs (Thuman and Mooney, 2000). BASINS builds upon federal databases of water quality conditions and point-source loadings for numerous parameters where quality assurance is suspect in some cases (Whittemore and Beebe, 2000).

BASINS Examples

Most of the BASINS work and modeling have been on a watershed- or regional-level scale. An example for the BASINS data set for Boulder, Colorado, is shown in Figure 11.12. The size of this file relative to the area it represents reflects a scale

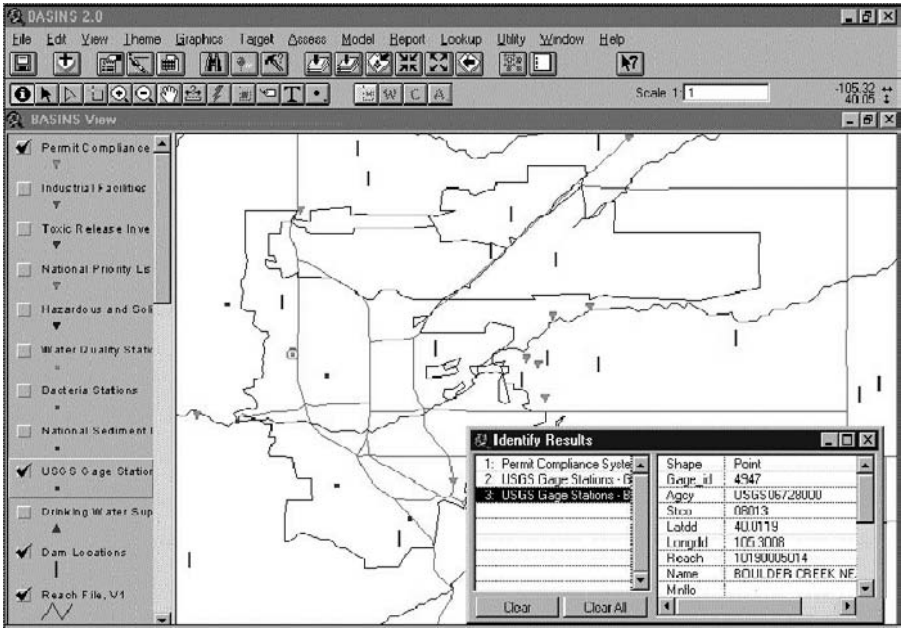


Figure 11.12 BASINS GIS layers for Boulder, Colorado.

of about 1:2000 (Heaney et al., 1999). The Columbus Water Works of Columbus, Georgia, developed a BASINS 2 model for the 2400 mi^2 (3600 km^2) Middle Chattahoochee River Watershed. The main objectives of the model were to formulate TMDL allocation requirements of the Clean Water Act and to demonstrate compliance with EPA's National Combined Sewer Overflow (CSO) Policy. They used monitoring and data management methodologies to calibrate their BASINS model and to perform source-water assessments and protection through provisions of early-warning systems for drinking-water intake operations (Turner et al., 2000).

MIKE BASIN

Danish Hydraulic Institute's (DHI) MIKE BASIN is an ArcView-based decision support system for water resources planning and management. It serves as a generalized river basin network simulation modeling system for regional water resources allocation of complex systems. MIKE BASIN can be used to analyze water supply capabilities in connection with water rights for municipal and industrial water supply, irrigation, and many multipurpose reservoir operations.

MIKE BASIN is fully integrated into ArcView 3.x. It uses a graphical user interface, which links MIKE BASIN to ArcView. It is developed using the object-oriented C++ programming language, which allows for easy implementation of new features. MIKE BASIN permits on-screen creation of models using image themes as backdrops. Figure 11.13 shows the MIKE BASIN interface.

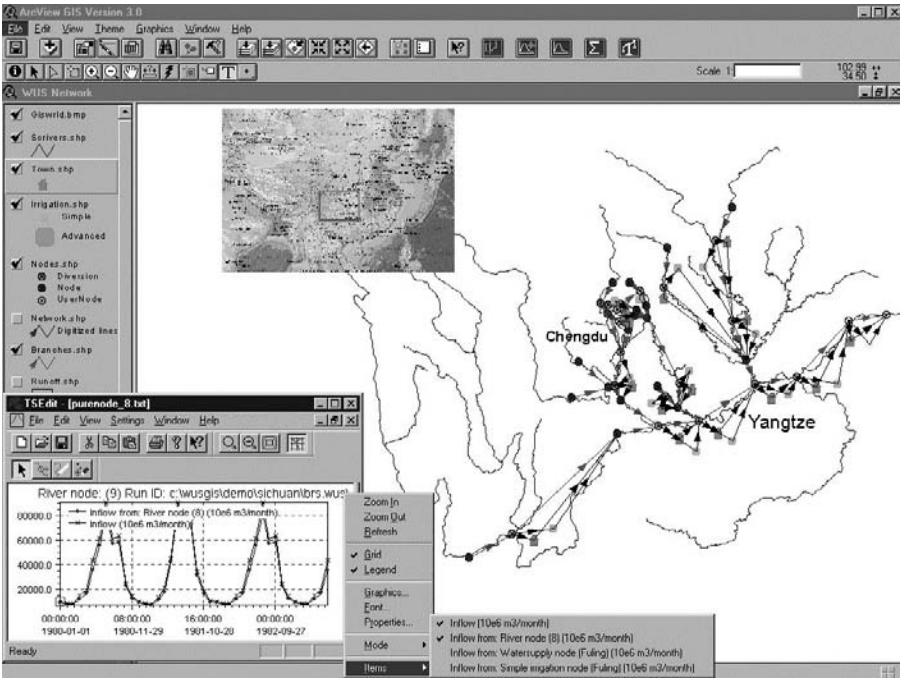


Figure 11.13 MIKE BASIN screenshot.

An important advantage of MIKE BASIN is its vertical compatibility with other DHI modeling systems, which allows the use of several modeling tools of various complexities at different stages of a project cycle. MIKE BASIN can be used in conjunction with other DHI models when detailed studies of specific problems are required. For example, MIKE SHE can be used for modeling runoff quality and quantity, including interaction with surface water systems, MIKE 11 can be used for detailed modeling of river hydrodynamics and water quality, and MIKE 21 can be used for modeling bays and estuaries.

Geo-STORM Integration

An early example of ArcInfo 7.x and modeling integration is the Geo-STORM package developed by Innovative System Developers, Inc., Columbia, Maryland. Geo-STORM computed subbasin runoff, performed flow routing through reach and reservoir networks, and analyzed river hydraulics using SCS Technical Release 20 (TR-20) and Technical Release 55 (TR-55) programs, and COE's HEC-2 computer programs. All these models were contained within a GUI environment within ArcInfo GIS. The software was written in AML, the ArcInfo scripting language. There were external calls to the operating system only to execute the computer programs. The common interface provided easy interaction between the models and switching between the models was hardly noticeable, and output from one model was used as input for another model. The TR-55 model was embedded within the TR-20 model. The output from TR-55 became input for TR-20. The TR-20 model provided discharge values for input to the HEC-2 model, and output from HEC-2 could be directly inserted into TR-20 as rating-table information. All modeling applications were integrated in Geo-GUIDE, a graphical interface that made working with ArcInfo easier. The computer models were embedded within Geo-GUIDE, which provided plotting, reporting, and querying capabilities for the stormwater modeling software suite. The GeoSTORM system was considered to be a database repository for stormwater-related geography and facilities. Using programs like Geo-STORM translated into productivity gains realized by eliminating tedious tasks such as manual map measurements, data entry, and tabular-report interpretation. Geo-STORM automated these time-consuming processes, drawing model inputs from the GIS database. This in time improved the quality and accuracy of model results. Figure 11.14 shows a screenshot of Geo-STORM (Thomas, 1995).

ARC/HEC-2 Integration

ARC/HEC-2 was one of the earliest integrations of ArcInfo and HEC-2, developed at the University of Texas at Austin. ARC/HEC-2 facilitated floodplain analysis in a GIS environment. It allowed direct usage of ArcInfo coverages for development of HEC-2 input, running HEC-2 from within ArcInfo, and for conversion of HEC-2 water-surface elevations computed at each cross section into an ArcInfo coverage representing the floodplain. There were two main data requirements for ARC/HEC-2 to operate: (1) a terrain representation in the form of constructed TIN or contours, and (2) channel centerline (thalweg). ARC/HEC-2 consisted of three separate modules: preprocessor,

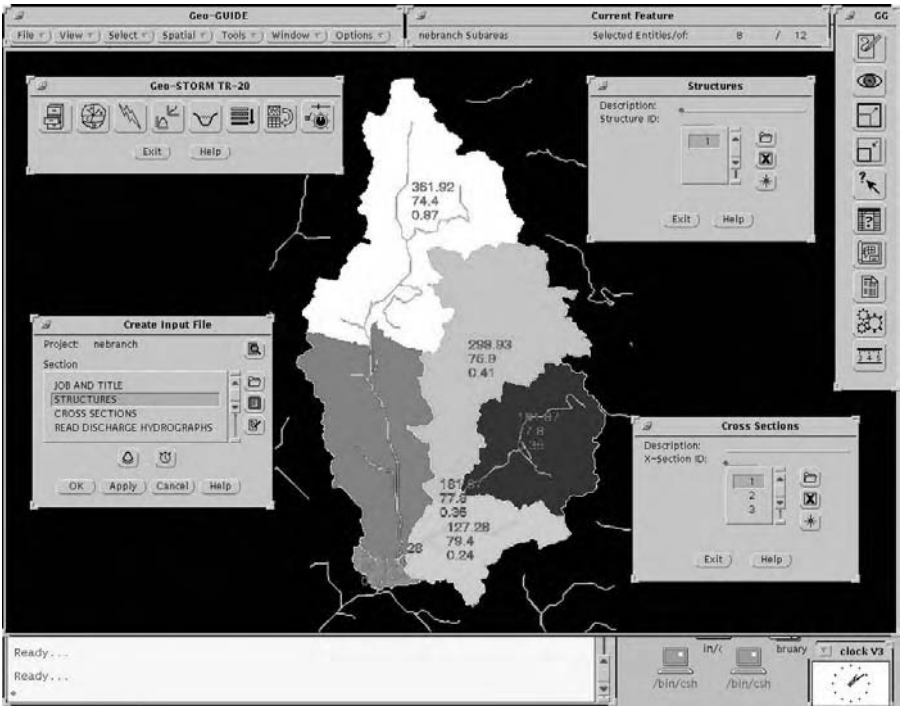


Figure 11.14 Geo-STORM screenshot showing ArcInfo and TR-20 integration.

postprocessor, and viewer. The preprocessor created a complete HEC-2 input file ready to run. For complex HEC-2 models, manual editing of the input file was required, in which case ARC/HEC-2 served as an interface rather than an integration. The postprocessor used the HEC-2 output to determine water surface elevation at each cross section and, based on that information, found the outline of the flooded area in the form of a lattice and a polygon layer. This lattice and coverage could be used in ArcInfo with all other pertinent coverages for the area. Viewer was an AML application that allowed viewing of the main input and output coverages in 2D and 3D formats (Djokic et al., 1993).

Integration Method Case Study

GIS software	ArcView and Spatial Analyst Extension
Modeling Software	Agricultural Nonpoint Source Pollution Model (AGNPS)
GIS data	Soil (STATSGO), USGS DEM (1:250,000 scale), land use, hydrography, watersheds, and crop management
Study area	280 mi^2 (810 km^2) Dowagiac River Watershed located in southwestern Michigan
Organization	Western Michigan University, Kalamazoo, Michigan

Successful implementation of environmental modeling databases involves the integration of GIS, remote sensing, multiple databases, and visualization tools for

extraction of the needed model input parameters and for analysis and visualization of the simulated results. Although it was obvious that GIS technology was ideally suited to accomplish these goals, current GIS packages were found to be difficult to use and offered limited functions to support environmental modeling. It was also noted that in current practice, GIS and hydrologic models were just used together, not really integrated. Therefore, a custom interface called ArcView Nonpoint Source Pollution Modeling (AVNPSM) was developed between ArcView GIS and AGNPS in support of agricultural watershed analysis and nonpoint source pollution management.

AGNPS (Version 5.0) is a single storm event based simulation model for evaluating sediment and nutrient (nitrogen, phosphorus, and chemical oxygen demand or COD) transport from agricultural watersheds. It represents the landscape of a watershed by discretizing the study area into individual grid cells. The PC Windows-based interface developed using Avenue scripts and Dialogue Designer consists of seven modules: the AGNPS utility, parameter generator, input file processor, model executor, output visualizer, statistical analyzer, and land-use simulator. Each module is incorporated into ArcView through pull-down menus and icons. The input file processor menu produces an input file that is compatible with AGNPS input format requirements. The AGNPS model is executed either within Windows using the model executor menu or run separately in simulated DOS mode. The output visualizer menu shows the simulated results of hydrology, sediment, and nutrients in tabular or map format. The calibrated model indicated that development of about 1240 acres of nonurban land caused little increase in the peak flow rate and sediment yield at the mouth of the river and had almost no impact on the runoff volume and nutrient yields. The case study concluded that integration of GIS and simulation models facilitates the development and application of environmental simulation models and enhances the capabilities of GIS packages (He et al., 2001).

WHICH LINKAGE METHOD TO USE?

Each method has its pros and cons. Table 11.5 lists the advantages and disadvantages of the three methods of linking H&H models with GIS. Integrated systems are easy to learn and use and make the excruciating task of H&H modeling a fun

Table 11.5 Comparison of GIS Linkage Methods

Feature	Interchange Method	Interface Method	Integration Method
Automation	Low	Medium	High
User-friendliness	None	Medium	High
Ease of use	Cumbersome	Easy	Easiest
Learning curve	Long	Medium	Quick
Data-entry error potential	High	Medium	Low
Data-error tracking	Easy	Complex	Difficult
Misuse potential by inexperienced users	Low	Medium	High
Development effort	Low	Medium	High
Computer programming	Low	Medium	Extensive

activity. Integrated systems save model development and interpretation time and make modeling cost-effective.

Seamless model integration with GIS is not without some disadvantages. The simplistic modeling approach and user-friendly tools provided by integrated systems like BASINS may encourage inexperienced users to become instant modelers, which could be undesirable (Thuman and Mooney, 2000). GIS can easily convert reams of computer output into colorful thematic maps that can both expose and hide data errors. The user-friendly interface of an integrated package may incorrectly make a complex model appear quite simple, which may lead to misapplication of the model by inexperienced users. Modeling should not be made easy at the expense of good science. Modeling, for all the new tools emerging in the 1990s, is still an art that is done well by only a select few (Whittemore and Beebe, 2000).

The inexperienced users' overreliance on default model parameters may result in inaccurate modeling results. Inexperienced H&H modelers should exercise caution when using integrated systems, make sure to browse the model input and output files for obvious errors, and avoid overreliance on thematic mapping and graphing of model output. Modification and recompilation of computer codes of existing legacy programs like SWMM or EPANET to create integrated systems should be avoided. If recoding is necessary, developers should be extremely cautious to avoid coding errors and the associated liability. Users should remember that a difficult-to-use model is much better than an inaccurate model (Shamsi, 2001).

The development of a GIS for use in urban H&H modeling is an expensive investment (Heaney et al., 1999). Typically the most advanced tools are created for advanced applications, where a full-fledge GIS is in place. For some applications, a DOS-based model may still be the most appropriate. However, as more urban areas create GIS layers, the integration of modeling software and GIS software will become more useful and prevalent.

USEFUL WEB SITES

HEC software	www.hec.usace.army.mil/
GIS Hydro	www.crwr.utexas.edu/archive.shtml
BASINS software	www.epa.gov/OST/BASINS
WMS software	http://emrl.byu.edu/wms.htm

CHAPTER SUMMARY

This chapter described GIS applications that have revolutionized the way H&H modeling is performed. It showed how GIS can be used to develop H&H models, and also provided a software review of GIS application programs. From the material presented in this chapter, it can be concluded that GIS applications allow modelers to be more productive. Users devote more time to understanding the problem and less time to mechanical tasks of data input and checking, getting the program to run, and interpreting reams of output. More than just tabular outputs, models become an automated system evaluation tool. There are three modeling application methods:

interchange, interface, and integration. Interchange is the simplest method but offers limited benefits. Interchange systems have existed since 1980s in all domains. The interface method's benefits lie between those of the interchange and the integration methods. There are a few interface systems in the public domain and many in commercial and proprietary domains. The integration method offers the maximum productivity benefits. BASINS is the only public-domain modeling system that uses the integration method. A few integrated commercial systems are available starting at $5000. There are many in-house and proprietary integrated systems. The number of public-domain and commercial integration packages is expected to grow steadily in the near future.

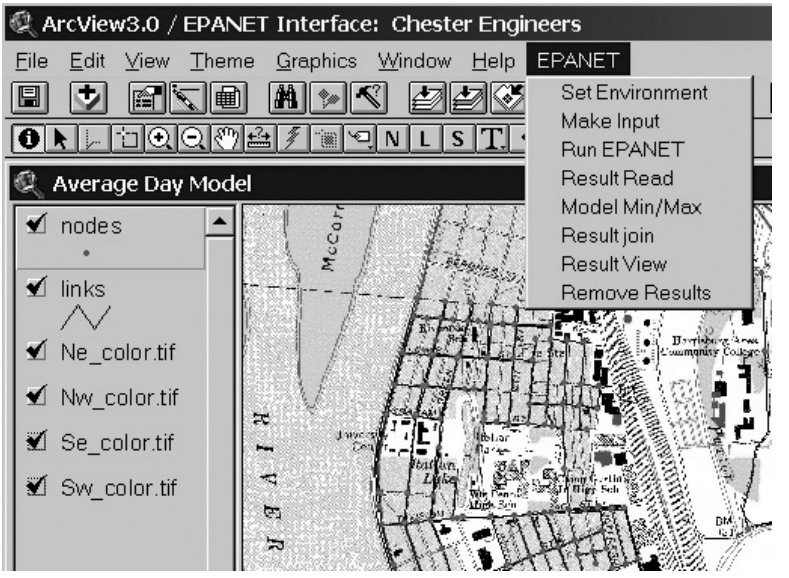
CHAPTER QUESTIONS

1. What are the GIS applications in computer modeling of water, wastewater, and stormwater systems?
2. What are computer modeling application methods?
3. Give an example of each application method using programs other than those described in this chapter.
4. Which GIS application method is the easiest to use? Why?
5. Which GIS application method is the easiest to implement? Why?
6. List the typical steps used in the interface method.
7. Make a list of the H&H models used in your organization and the type of GIS application method used by each.

CHAPTER 12

Water Models

GIS saves time and money in developing water distribution system hydraulic models for simulating flows and pressures in the system. GIS also helps in presenting the model results to non-technical audiences.



EPANET and ArcView GIS integration.

LEARNING OBJECTIVE

The learning objective of this chapter is to understand how GIS can be applied in developing water distribution system hydraulic models and presenting the results of the hydraulic models.

MAJOR TOPICS

- Development of hydraulic models
- Software examples
- EPANET model
- Network simplification (skeletonization)
- Estimation of node demands
- Estimation of node elevations
- Delineation of pressure zones
- Mapping the model output results
- GIS application examples and case studies

LIST OF CHAPTER ACRONYMS

AM/FM/GIS Automated Mapping/Facilities Management/Geographic Information System
CIS Customer Information System
COM Component Object Model
COTS Commercial Off-the-Shelf
DEM Digital Elevation Model
DHI Danish Hydraulic Institute
DLL Dynamic Link Library
DRG Digital Raster Graphic
GEMS Geographic Engineering Modeling Systems
MMS Maintenance Management System
ODBC Open Database Connectivity
PRV Pressure Regulating Valve
SCADA Supervisory Control and Data Acquisition
TIF Tag Image File

CITY OF GERMANTOWN'S WATER MODEL

Application	Water distribution system mapping and modeling
GIS software	ArcGIS, ArcGIS Water Utilities Data Model (formerly ArcFM Water Data Model), and MapObjects
Other software	WaterCAD
GIS layers	Digital orthophotos, pipes, hydrants, valves, manholes, pumps, meters, fittings, sampling stations, and monitoring wells
Hardware	Two Dell Precision 220 computers
Study area	City of Germantown, Tennessee
Organization	City of Germantown, Tennessee

The City of Germantown (Tennessee) water distribution system serves 40,000 people through 12,000 installations in a 17.5 mi^2 area. The system has approximately 95,000 ft of water pipes.

The City used ESRI's geodatabase, an object-oriented GIS data model, to design a database. The database design was done with minimal customization of ESRI's ArcGIS Water Utilities Data Model. This approach helped complete the database design in a few weeks instead of several months. ESRI's MapObjects was used to create a custom utility tool set for developing the water distribution network. Valve and hydrant locations from the digital orthophotos and planimetric mapping were used with the existing 1 in. = 500 ft-scale paper map of the water system as the basis for connecting water pipes.

The GIS layers (Figure 12.1) were initially developed using the ArcView Shapefiles. The Shapefiles were then migrated to a geodatabase. To ease the data migration, connectivity rules were stored in ArcCatalog rather than directly in the model. ArcMap "flag and solver" tracing tools were used to ensure topological integrity of the migrated data. ArcMap also provided the City with out-of-the-box network analysis functionality such as tracing paths and water main isolation.

For the development of the hydraulic model, the City did not have to create a modeling network from scratch. GIS was used to skeletonize the network for input to the Haestad Method's WaterCAD 4.1 hydraulic model. Spatial operations available in ArcGIS were used to assign customer-demand data — obtained from the City's billing database and stored in parcel centroids — to modeling nodes. Node elevations were extracted from the City's DEM.

The hydraulic model was run to assess the future system expansion and capital improvement needs of the City. The model indicated that the system needed a new elevated storage tank and a larger pipe to deliver acceptable pressure under peak demand conditions (ESRI, 2002b).

GIS APPLICATIONS FOR WATER DISTRIBUTION SYSTEMS

GIS has wide applicability for water distribution system studies. Representation and analysis of water-related phenomena by GIS facilitates their management. The GIS applications that are of particular importance for water utilities include mapping, modeling, facilities management, work-order management, and short- and long-term planning. Additional examples include (Shamsi, 2002):

- Conducting hydraulic modeling of water distribution systems.
- Estimating node demands from land use, census data, or billing records.
- Estimating node elevations from digital elevation model (DEM) data.
- Performing model simplification or skeletonization for reducing the number of nodes and links to be included in the hydraulic model.
- Conducting a water main isolation trace to identify valves that must be closed to isolate a broken water main for repair. Identifying dry pipes for locating customers or buildings that would not have any water due to a broken water main. This application is described in Chapter 15 (Maintenance Applications).
- Prepare work orders by clicking on features on a map. This application is described in Chapter 14 (AM/FM/GIS Applications).
- Identifying valves and pumps that should be closed to isolate a contaminated part of the system due to acts of terrorism. Recommending a flushing strategy to clean the contaminated parts of the system. This application is described in Chapter 16 (Security Planning Vulnerability Assessment).

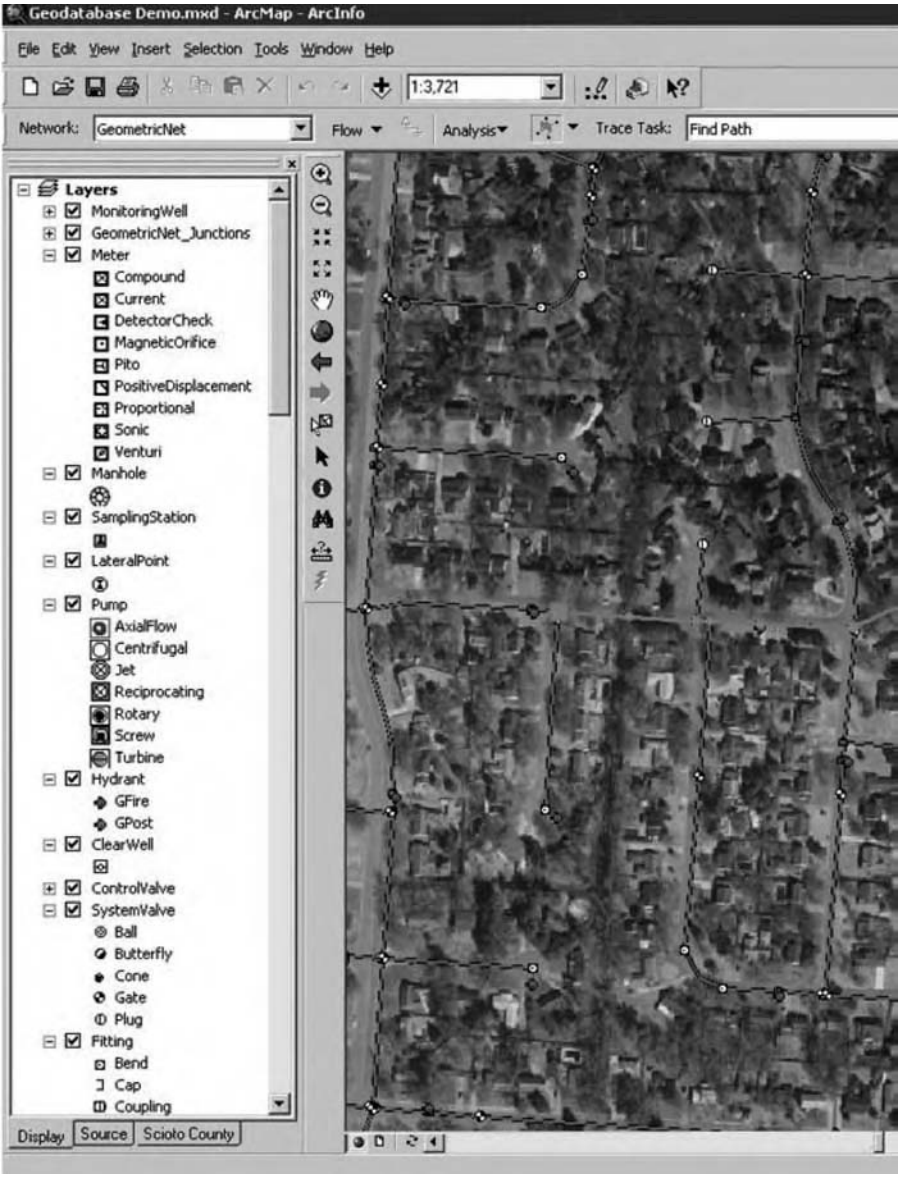


Figure 12.1 Germantown water distribution system layers in ArcGIS.

- Providing the basis for investigating the occurrence of regulated contaminants for estimating the compliance cost or evaluating human health impacts (Schock and Clement, 1995).
- Investigating process changes for a water utility to determine the effectiveness of treatment methods such as corrosion control or chlorination.
- Assessing the feasibility and impact of system expansion.
- Developing wellhead protection plans.

By using information obtained with these applications, a water system manager can develop a detailed capital improvement program or operations and maintenance plan (Morgan and Polcari, 1991). The planning activities of a water distribution system can be greatly improved through the integration of these applications.

In this chapter, we will focus on the applications related to hydraulic modeling of water distribution systems.

DEVELOPMENT OF HYDRAULIC MODELS

The most common use of a water distribution system hydraulic model is to determine pipe sizes for system improvement, expansion, and rehabilitation. Models are also used to assess water quality and age and investigate strategies for reducing detention time. Models also allow engineers to quickly assess the distribution network during critical periods such as treatment plant outages and major water main breaks.

Today, hydraulic models are also being used for vulnerability assessment and protection against terrorist attacks.

Hydraulic models that are created to tackle a specific problem gather dust on a shelf until they are needed again years later. This requires a comprehensive and often expensive update of the model to reflect what has changed within the water system. A direct connection to a GIS database allows the model to always remain live and updated.

Denver Water Department, Colorado, installed one of the first AM/FM/GIS systems in the U.S. in 1986. In the early 1990s, the department installed the $60,000 SWS water model from the Stoner Associates (now Advantica Stoner, Inc.). Around the same time, Genesee County Division of Water and Waste Services in Flint, Michigan, started developing a countywide GIS database. The division envisioned creating sewer and water layers on top of a base map and integrating the GIS layers with their water and sewer models (Lang, 1992).

GIS applications reduce modeling development and analysis time. GIS can be used to design optimal water distribution systems. An optimal design considers both cost-effectiveness and reliability (Quimpo and Shamsi, 1991; Shamsi, 2002a) simultaneously. It minimizes the cost of the system while satisfying hydraulic criteria (acceptable flow and pressures). Taher and Labadie (1996) developed a prototype model for optimal design of water distribution networks using GIS. They integrated GIS for spatial database management and analysis with optimization theory to develop a computer-aided decision support system called WADSOP (Water Distribution System Optimization Program). WADSOP employed a nonlinear programming technique as the network solver, which offers certain advantages over conventional methods such as Hardy-Cross, Newton Raphson, and linear system theory for balancing looped water supply systems. WADSOP created a linkage between a GIS and the optimization model, which provided the ability to capture model input data; build network topology; verify, modify, and update spatial data; perform spatial analysis; and provide both hard-copy reporting and graphical display of model results. The GIS analysis was conducted in PC ARC/INFO. The ROUTE and ALLOCATE programs of the NETWORK module

of PC ARC/INFO were found to be very helpful. Computational time savings of 80 to 99% were observed over the other programs.

Generally, there are four main tasks in GIS applications to water system modeling:

1. Synchronize the model's network to match the GIS network.
2. Transfer the model input data from a GIS to the model.
3. Establish the model execution conditions and run the model.
4. Transfer the model output results to GIS.

As described in Chapter 11 (Modeling Applications), GIS is extremely useful in creating the input for hydraulic models and presenting the output. A hydraulic modeler who spends hundreds of hours extracting input data from paper maps can accomplish the same task with a few mouse clicks inside a GIS (provided that the required data layers are available). GIS can be used to present hydraulic modeling results in the form of thematic maps that are easy for nontechnical audiences to understand. Additional information about GIS integration methods is provided in Chapter 11, which provided a taxonomy developed by Shamsi (1998, 1999) to define the different ways a GIS can be linked to hydraulic models. According to Shamsi's classification, the three methods of developing GIS-based modeling applications are interchange, interface, and integration. For GIS integration of water distribution system hydraulic models, one needs a hydraulic modeling software package. The software can run either inside or outside the GIS. As explained in Chapter 11, when modeling software is run inside a GIS, it is considered "seamlessly integrated" into the GIS. Most modeling programs run in stand-alone mode outside the GIS, in which case the application software simply shares GIS data. This method of running applications is called a "GIS interface."

The interchange method offers the most basic application of GIS in hydraulic modeling of water distribution systems. In 1995, the Charlotte–Mecklenburg Utility Department (CMUD) used this method to develop a KYPIPE model for an area serving 140,000 customers from 2,500 mi of water pipes ranging in size from 2 in. to 54 in. (Stalford and Townsend, 1995). The modeling was started by creating an AutoCAD drawing of all pipes 12 in. and larger. The pipes were drawn as polylines and joined at the intersections. The AutoCAD drawing was converted to a facilities management database by placing nodes at the hydraulic intersections of the pipes, breaking the polylines as needed, adding extended attributes to the polylines from the defined nodes, extracting this information into external tables, and adding the nongraphical information to the tables. Then, by using ArcCAD inside AutoCAD, coverages were created for the pipes and nodes, which could be viewed inside ArcView. Although this method was considered state-of-the-art in 1995, it is not a very efficient GIS integration method by today's standards. The EPANET integration case study presented later in the chapter exemplifies application of the more efficient integration method.

In the simplest application of the integration method, it should be possible to use a GIS to modify the configuration of the water distribution network, compile model input files reflecting those changes, run the hydraulic model from within the GIS, use the GIS to map the model results, and graphically display the results of the simulation

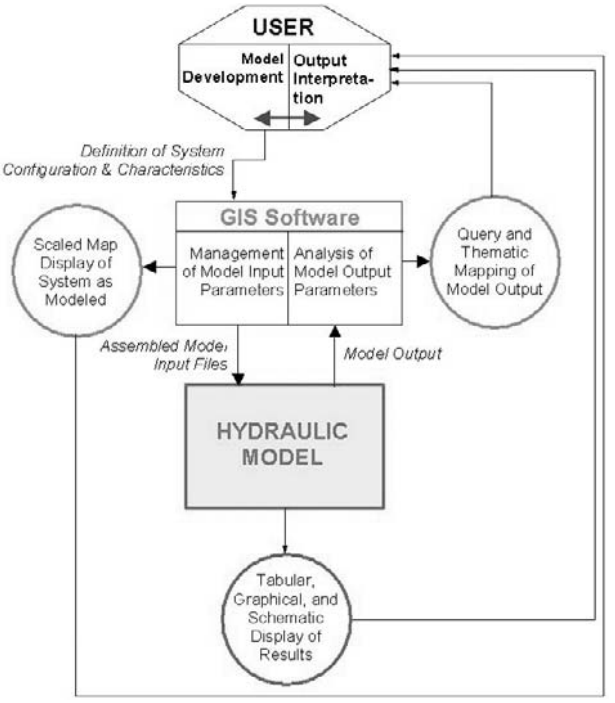


Figure 12.2 Integration method interrelationships.

on a georeferenced base map. The integration method interrelationships among the user, the hydraulic model, and the GIS software are illustrated in Figure 12.2.

SOFTWARE EXAMPLES

Table 12.1 lists representative water distribution system modeling packages and their GIS capabilities, vendors, and Web sites. The salient GIS features of these programs are described in the following subsections.

EPANET

EPANET was developed by the Water Supply and Water Resources Division (formerly the Drinking Water Research Division) of the U.S. Environmental Protection Agency's National Risk Management Research Laboratory (Rossman, 2000). It is a public-domain software that may be freely copied and distributed.

EPANET is a Windows 95/98/2K/NT program that performs extended-period simulation of hydraulic and water quality behavior within pressurized pipe networks. A network can consist of pipes, nodes (pipe junctions), pumps, valves, and storage tanks or reservoirs. EPANET tracks the flow of water in each pipe, the pressure at each node, the height of water in each tank, and the concentration of a chemical species

Table 12.1 Water Distribution System Modeling Software

Software	Method	Vendor	Web site
EPANET	Interchange	U.S. Environmental Protection Agency (EPA)	www.epa.gov/ord/NRMRL/wswrd/epanet.html
AVWater	Integration	CEDRA Corporation	www.cedra.com
H_2ONET and H_2OMAP	Interface and Integration	MWH Soft	www.mwhsoft.com
InfoWorks WS and InfoNet	Interface and Integration	Wallingford Software	www.wallingfordsoftware.com
MIKE NET	Interface	DHI Water & Environment	www.dhisoftware.com
PIPE2000 Pro	Interface	University of Kentucky	www.kypipe.com
SynerGEE Water	Interface and Integration	Advantica Stoner	www.stoner.com
WaterCAD	Interface	Haestad Methods	www.haestad.com
WaterGEMS	Integration		

throughout the network during a simulation period made up of multiple time steps. In addition to chemical species, water age and source tracing can also be simulated.

The Windows version of EPANET provides an integrated environment for editing network input data, running hydraulic and water quality simulations, and viewing the results in a variety of formats. These include color-coded network maps, data tables, time-series graphs, and contour plots. Figure 12.3 shows an EPANET screenshot displaying pipes color-coded by head-loss and nodes color-coded by pressure.

The EPANET Programmer's Toolkit is a dynamic link library (DLL) of functions that allow developers to customize EPANET's computational engine for their own specific needs. The functions can be incorporated into 32-bit Windows applications written in C/C++, Delphi Pascal, Visual Basic, or any other language that can call functions within a Windows DLL. There are over 50 functions that can be used to open a network description file, read and modify various network design and operating parameters, run multiple extended-period simulations accessing results as they are generated or saving them to file, and write selected results to file in a user-specified format. The toolkit should prove useful for developing specialized applications, such as optimization models or automated calibration models, that require running many network analyses while selected input parameters are iteratively modified. It also can simplify adding analysis capabilities to integrated network modeling environments based on CAD, GIS, and database management packages (Rossman, 2000a).

The current version of EPANET (Version 2.0) has neither a built-in GIS interface nor integration. Therefore, users must develop their own interfaces (or integration) or rely on the interchange method to manually extract model input parameters from GIS data layers.

H_2ONET™ and H_2OMAP™

MWH Soft, Inc. (Broomfield, Colorado), provides infrastructure software and professional solutions for utilities, cities, municipalities, industries, and engineering organizations. The company was founded in 1996 as a subsidiary of the global

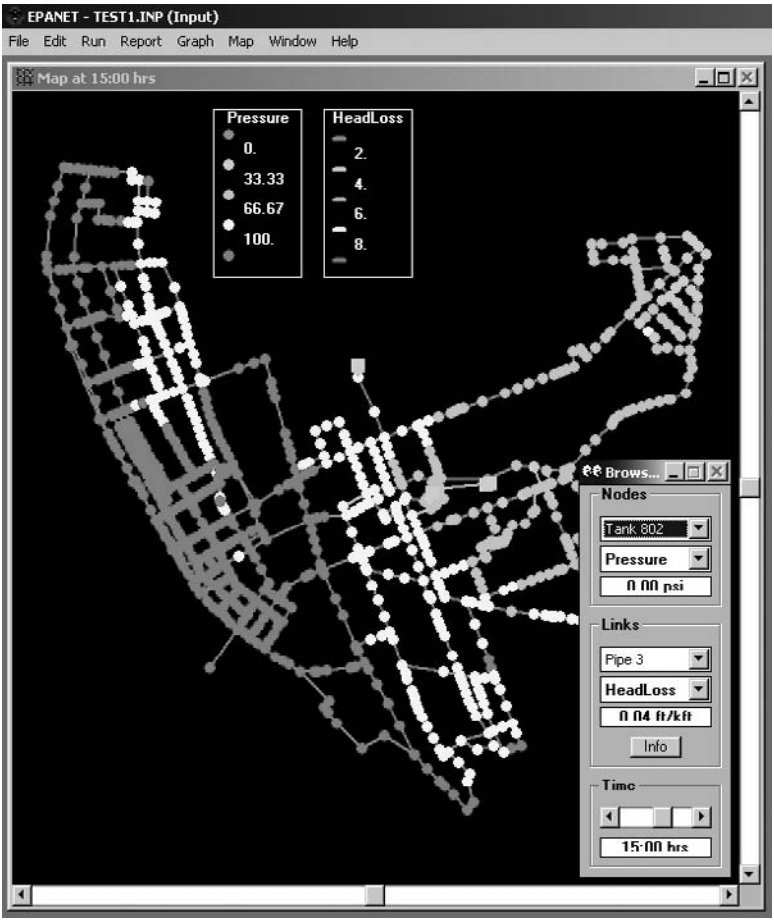


Figure 12.3 EPANET screenshot (Harrisburg, Pennsylvania, model).

environmental engineering, design, construction, technology, and investment firm MWH Global, Inc. The company's CAD- and GIS-enabled products are designed to manage, design, maintain, and operate efficient and reliable infrastructure systems. MWH Soft, provides the following water distribution system modeling software:

- H_2ONET
- H_2OMAP Water
- InfoWater
- H_2OVIEW Water

H_2ONET is a fully integrated CAD framework for hydraulic modeling, network optimization, graphical editing, results presentation, database management, and enterprise-wide data sharing and exchange. It runs from inside AutoCAD and, therefore, requires AutoCAD installation.

Built using the advanced object-oriented geospatial component model, H_2OMAP Water provides a powerful and practical GIS platform for water utility solutions. As a stand-alone GIS-based program, H_2OMAP Water combines spatial analysis tools and mapping functions with network modeling for infrastructure asset management and business planning. H_2OMAP Water supports geocoding and multiple mapping layers that can be imported from one of many data sources including CAD drawings (e.g., DWG, DGN, and DXF) and standard GIS formats (Shapefiles, Generate files, MID/MIF files, and ArcInfo coverages). The program also supports the new geodatabase standard of ArcGIS through an ArcSDE connection.

InfoWater is a GIS-integrated water distribution modeling and management software application. Built on top of ArcGIS using the latest Microsoft .NET and ESRI ArcObjects component technologies, InfoWater integrates water network modeling and optimization functionality with ArcGIS.

H_2OVIEW Water is an ArcIMS-based (Web-enabled) geospatial data viewer and data distribution software. It enables deployment and analysis of GIS data and modeling results over the Internet and intranets. Figure 12.4 shows a screenshot of H_2OVIEW Water.

Various GIS-related modules of H_2O products are described in the following subsections.

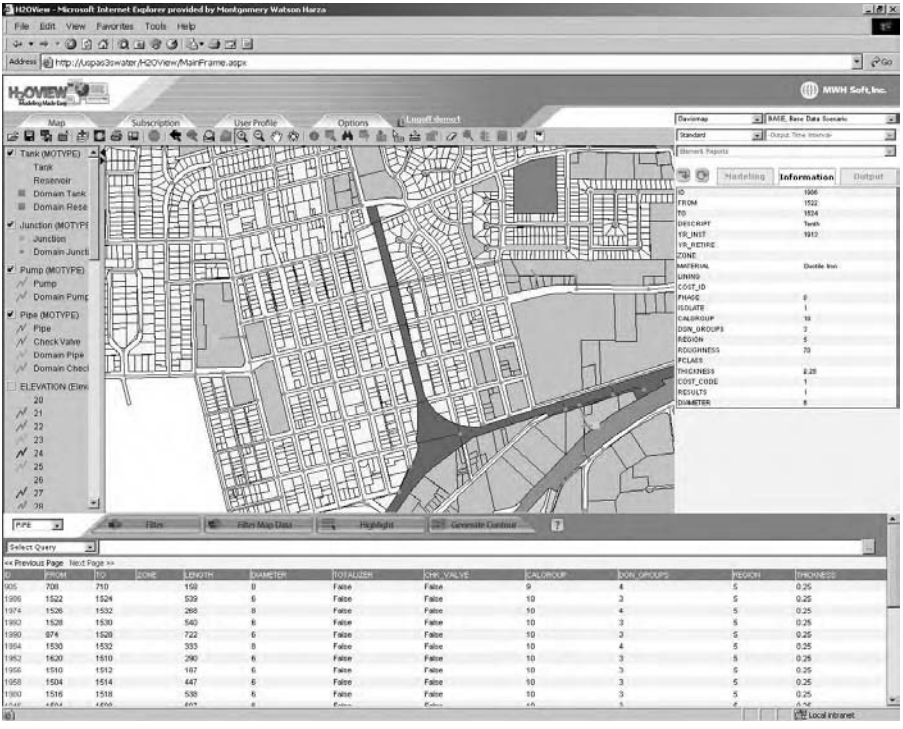


Figure 12.4 H_2OVIEW screenshot (Screenshot courtesy of MWH Soft).

Demand Allocator

Automatically computes and geographically allocates water demand data throughout the water distribution network model.

Skeletonizer

Automates GIS segmentation (data reduction, skeletonization, and trimming) to construct hydraulic network models and reallocates node demands directly from the GIS.

Tracer

H_2ONET Tracer is a network tracing tool that locates all valves which need to be closed to isolate a portion of the network, visually identify all customers affected by loss or reduction in service, and determine the best response strategy. It is a valuable tool for vulnerability analysis and security planning to protect systems against terrorist attacks.

WaterCAD™ and WaterGEMS™

Haestad Methods, Inc. (Waterbury, Connecticut), provides a wide range of H&H computer models, publishes books, and offers continuing education for the civil engineering community. Haestad Methods is a growing company with more than 100,000 users in more than 160 countries around the world. Haestad Methods was acquired by Bentley Systems (Exton, Pennsylvania) in August 2004. Haestad Methods provides the following water distribution system modeling software:

- WaterCAD Stand-alone
- WaterCAD for AutoCAD
- WaterGEMS for ArcGIS

WaterCAD is Haestad Method's water distribution analysis and design tool. It can analyze water quality, determine fire flow requirements, and calibrate large distribution networks. The stand-alone version does not require a CAD or GIS package to run. The AutoCAD version runs from inside AutoCAD and, therefore, requires AutoCAD.

Shapefile Wizard, the ArcView GIS interface module for WaterCAD, is sold separately. The wizard provides import and export capability to transfer data between GIS and WaterCAD models. For example, Import Shapefile Wizard imports WaterCAD model input parameters from ArcView GIS Shapefiles. Similarly, the Export Shapefile Wizard exports hydraulic model networks to ArcView. This feature lets users build and maintain their water and sewer networks directly inside a GIS.

Geographic Engineering Modeling Systems (GEMS) is the next generation of Haestad Methods' hydraulic network modeling products. GEMS is a new product

family that combines water, wastewater, and stormwater system modeling with an open database architecture. It was developed using ESRI's ArcObjects technology and computing tools from Microsoft's .NET initiative. GEMS technology eliminates the need for a separate model database. It streamlines the process of accessing information on, for example, facilities management, zoning, topography, and gauging stations for use with the model. GEMS-based models can be run within ESRI's ArcGIS software using the geodatabase structure (Shamsi, 2001).

WaterGEMS uses a geodatabase, Water Data Model, for integrated WaterCAD modeling within the ArcGIS platform. Because WaterGEMS runs from inside ArcGIS, it requires an installation of ArcGIS software. Although WaterGEMS accommodates Shapefiles and ArcInfo coverages, it provides users with the tools to migrate to the new geodatabase format. Users can easily save an existing hydraulic model as a geodatabase.

In an interview by Haestad Methods, Jack Dangermond, president of ESRI, stated, "GEMS bridges the gap between engineering and planning departments and allows users to further leverage their GIS investment" (Haestad Methods, 2003).

MIKE NET™

DHI Water and Environment (formerly Danish Hydraulic Institute, Hørsholm, Denmark, www.dhi.dk) is a global provider of specialized consulting and numerical modeling software products for water, wastewater, river and coastal engineering, and water resources development. DHI has developed a large number of hydraulic, hydrologic, and hydrodynamic computer models for water, wastewater, and stormwater systems. The company also specializes in linking its computer models with GIS so that modelers can use both modeling and GIS technologies within a single product. Since 1998, DHI has embarked on an ambitious program to link its models with the ESRI family of GIS products. Many of DHI's modeling systems now support GIS data transfer. For example, DHI has developed ArcView GIS extensions and interface programs for a number of its products.

MIKE NET is a software package for the simulation of steady flow and pressure distribution and extended-period simulation of hydraulic and water quality behavior within drinking water distribution systems. MIKE NET was developed in cooperation with BOSS International (Madison, Wisconsin, www.bossintl.com). MIKE NET uses the EPANET 2.0 numerical engine.

Data can be shared with any standard Windows spreadsheet (e.g., Microsoft Excel) or relational database (e.g., Oracle, Microsoft SQL Server, Informix, and Sybase) either directly or using Open Database Connectivity (ODBC) links, or by simply cutting to and pasting from the Microsoft Windows clipboard.

Important GIS features of MIKE NET are described below:

- MIKE NET can read the network models directly from an ArcInfo, ArcFM, ArcView, or MapInfo GIS database. MIKE NET can build a link to a GIS database structure, using attribute mapping and geocoding. For example, a pump or valve can be represented as either a node or an arc in the originating GIS database, while still being linked with MIKE NET.
- MIKE NET can read the node elevation field from Shapefile polygons and assign it to each junction node within the polygon.

- MIKE NET can import a fire hydrants layer to automatically determine which pipe the hydrant belongs to. It temporarily inserts the hydrant into the pipe at the hydrant location, performs the fire flow simulation, and resets the network back to the original state without the hydrant.
- MIKE NET allows the user to assign nodal demands to the entire network system using the program's Distributed Demands capability. In addition, GIS layers can be used to geocode water consumption as well as multiple demand patterns from land-use and population-density aerial coverages. Multiple overlying layers can be aggregated and assigned to the appropriate junction nodes.
- It can perform automated and/or user-assisted skeletonization of the distribution network. Node demands are automatically accumulated during the skeletonization process so that the skeletal model continues to perform and respond closely to those of the original network model.

Other Programs

CEDRA AVWater from Cedra Corporation (Rochester, New York) provides an ArcView 3.x GIS extension interface to the University of Kentucky's KYPIPE and EPA's EPANET models.

The InfoWorks WS from Wallingford Software (Wallingford, United Kingdom) imports models from any data source using the InfoWorks CSV data gateway. It exchanges data with MapInfo Professional and ArcView GIS and exports data and simulation results to MapInfo Professional and ArcView. InfoNet has a multiuser database with built-in data models for water, wastewater, and stormwater infrastructure; a GIS user-interface; and a report generator. It can be integrated with GIS, Desktop Office systems, hydraulic modeling systems, maintenance management systems (MMS), field data systems, SCADA, and corporate databases.

PIPE2000 is the hydraulic modeling software from the developers of the legacy KYPIPE program from the University of Kentucky (Lexington, Kentucky). PIPE2000 uses the EPANET simulation engine for water quality simulation. The Pro version of PIPE2000 imports and exports PIPE2000 data to and from Shapefiles.

SynerGEE Water (formerly Stoner Model) is a hydraulic modeling software from Stoner Advatica, Inc. (Carlisle, Pennsylvania). Water Solver uses SynerGEE's simulation engine as a Microsoft Component Object Model (COM)-compliant component. This capability allows integration of hydraulic simulation capability into a GIS.

EPANET AND ARCVIEW INTEGRATION IN HARRISBURG

A representative example of model integration is presented in this section. Additional water system modeling case studies are provided in Chapter 17 (Applications Sampler).

Chester Engineers (Pittsburgh, Pennsylvania) developed a GIS integrated model by combining ESRI's ArcView GIS software and EPA's EPANET hydraulic modeling program (Maslanik and Smith, 1998). EPANET provides extensive capability to support decisions in the operation, management, planning, and expansion of water distribution systems. ArcView is a user-friendly desktop

Table 12.2 AVNET Functions

Function	Description
	Menu Functions
Set Environment	Enter model execution parameters (e.g., length of simulation)
Make Input	Makes EPANET input file
Run EPANET	Launches the EPANET program
Results Read	Translates the EPANET output file into GIS database files
Model Min/Max	Finds the minimum and maximum EPANET output parameters
Results Join	Joins the EPANET results for a selected hour or minimum/maximum values to a layer
Results View	Sets the legend to an EPANET output parameter
Remove Results	Removes the join of the EPANET results from a theme
	Tool Functions
N Tool	Adds a water system node
L Tool	Adds a water system main
S Tool	Splits an existing water system line and adds a node

mapping and GIS tool, which can be learned by most users without extensive training and experience. Due to its low cost, ease of use, and customization capability, ArcView was found to be a suitable tool for developing the EPANET–GIS integration.

The custom ArcView 3.x and EPANET 2.x integration is called AVNET. AVNET performs preprocessing of GIS data to create model input files. It also conducts postprocessing of model results to create GIS maps of the model output results. Both the GIS and the modeling functions are available in ArcView. There is no need to exit ArcView to edit or run the model. AVNET provides seamless model input creation, model data editing, model execution, and model output display capabilities from within ArcView GIS. Other important features of AVNET are:

- As shown in the figure on the first page of this chapter, AVNET adds a new EPANET menu to ArcView's standard interface. The AVNET menu provides eight functions listed in Table 12.2.
- As shown in Figure 12.5, AVNET adds three new editing tools to ArcView's standard interface. These tools provide three functions listed in Table 12.2. These CAD-like editing tools allow adding nodes and pipes and on-screen input of model attributes.
- AVNET uses ArcView to modify the configuration of the water distribution system.
- Water mains and links can be added interactively while maintaining the required network topology.
- AVNET compiles new EPANET model input files reflecting the edits made by the user.
- AVNET requires ArcView Shapefiles or ArcInfo coverages for nodes and pipes. The required attributes for these layers are shown in Table 12.3.

Table 12.3 AVNET Attributes

Attribute	Type	Description
	Attributes for the Node Layer	
Node	Numeric	Node ID number
X_coord	Numeric	X coordinate value
Y_coord	Numeric	Y coordinate value
Type	Character	Node type (JUNCTION or TANK)
Elev	Numeric	Node elevation
Demand	Numeric	Node demand (water consumption)
Init_level	Numeric	Initial level of water system tank
Min_level	Numeric	Minimum level of water system tank
Max_level	Numeric	Maximum level of water system tank
Diam	Numeric	Diameter of water system tank
	Attributes for the Pipe Layer	
Link	Numeric	Line ID
Start_node	Numeric	Node ID number of starting node of pipe
End_node	Numeric	Node ID number of ending node of pipe
Type	Character	Main type (PIPE or PUMP)
Length	Numeric	Pipe length
Diam	Numeric	Pipe diameter
Rough_coef	Numeric	Pipe roughness coefficient
Pump_H	Numeric	Pump design head
Pump_Q	Numeric	Pump design flow

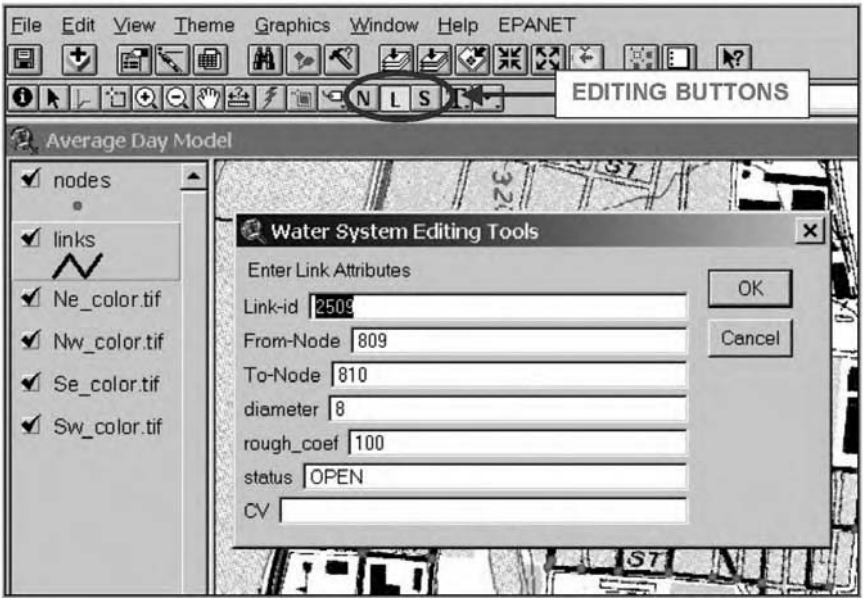


Figure 12.5 Editing pipe attributes in AVNET model.

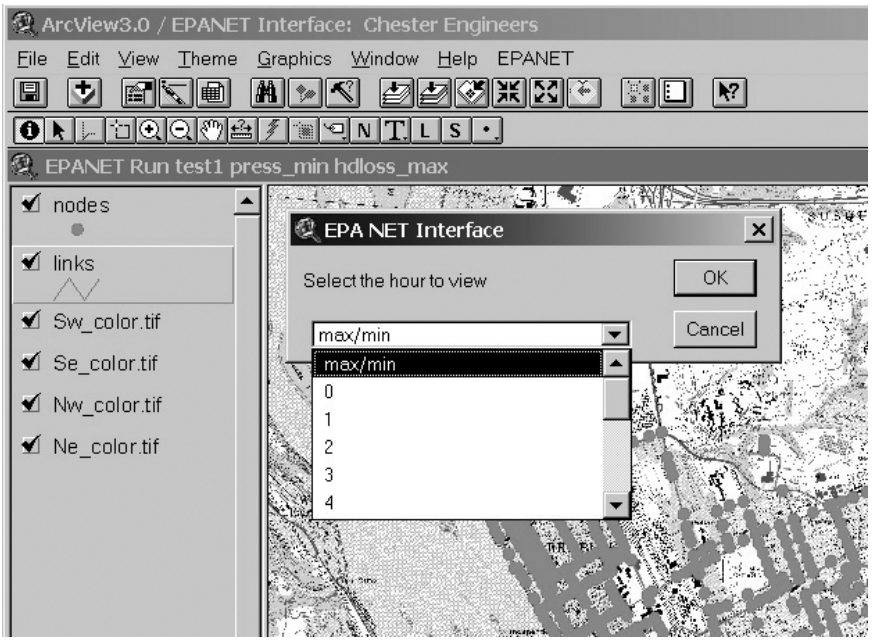


Figure 12.6 Reading the EPANET model output results in ArcView GIS.

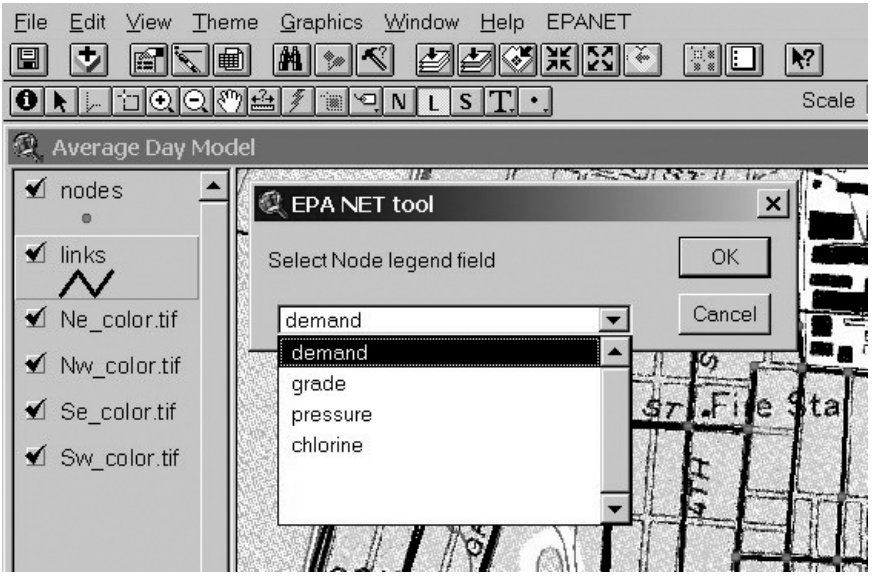


Figure 12.7 Classifying the legend to create model output thematic maps.

- AVNET creates a map file for EPANE model that contains a line for each node with the node ID and the x and y coordinates. The data are extracted from the water system node attribute table.
- AVNET runs the EPANET model from within ArcView.
- AVNET uses ArcView to map the water distribution system model and graphically display the results of the simulation on a georeferenced base map. The output file is linked to the GIS attribute files for pipes and nodes. As shown in Figure 12.6, AVNET reads the output file to extract the output parameters for minimum and maximum values or values at any user-specified simulation time step. Thematic maps of model output parameters (flows, pressures, residual chlorine, etc.) are created by classifying the ArcView theme legends (Figure 12.7).

Figure 12.8 shows the AVNET model for the city of Harrisburg, Pennsylvania. This model consists of 2500 pipes and 1650 nodes, 4 tanks, and 6 pumps. It shows layers for model nodes and links (pipes). The base-map layer consists of a USGS digital raster graphic (DRG) layer. A DRG is a TIF file of a scanned USGS 7.5-min paper topographic map.

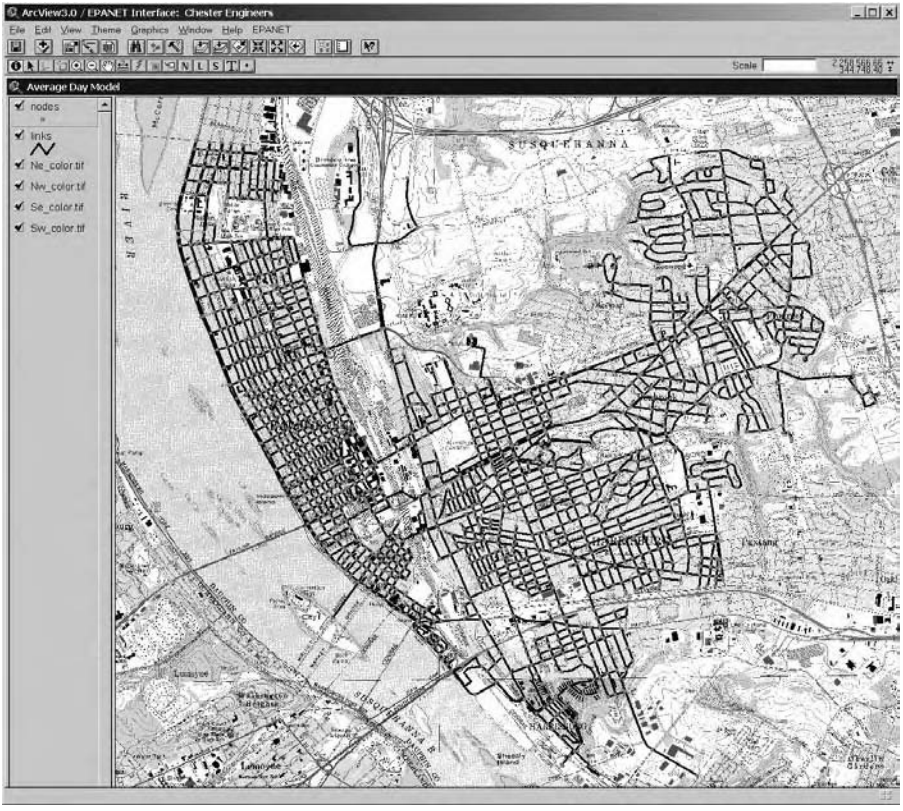


Figure 12.8 AVNET model for the City of Harrisburg, Pennsylvania.

Mapping the Model Output Results

The power of GIS gives a new meaning to the popular phrase "a picture is worth a thousand words." In fact, the thematic maps created by GIS are so effective in displaying large amounts of complex data that the popular phrase can be "plagiarized" to create a new cliché "a map is worth a thousand numbers."

The model output results in tabular or spreadsheet format are not very meaningful for nonmodelers. GIS technology is an effective means of bridging the gap between modeling information and its recipients. GIS can convert textual output results into easily interpretable thematic maps, for example, a map of pipes color-coded by flows and nodes color-coded by pressures. Figure 12.9 shows the Harrisburg model output results presented as a thematic map in ArcView GIS. The first layer (represented as dots) shows the minimum working pressures at the nodes. The second layer (represented as lines) shows the maximum head losses for the pipes. An overlay of the two layers clearly shows areas where pressures are low and head losses high. Such areas usually represent inadequate pipe sizes for delivering required flow rates and might be targeted for water main installation and replacement work. In fact, the GIS can automatically flag such hot spots for further study. The model output thematic maps are excellent for conveying modeling results to nontechnical audiences such as members of the city council or water utility board, or the general public.

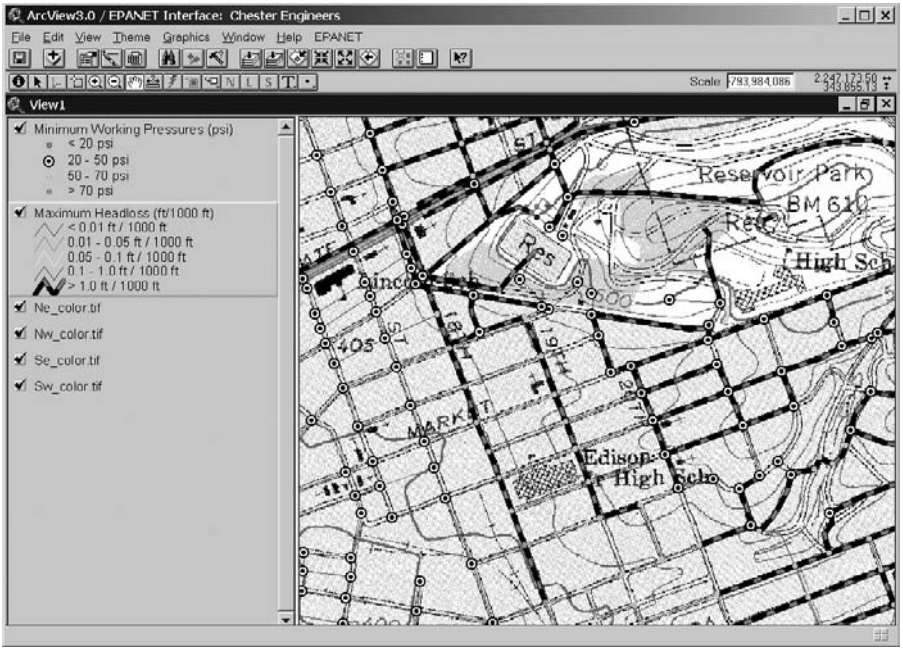


Figure 12.9 EPANET output results presented as a map in ArcView GIS.

NETWORK SKELETONIZATION

GIS and CAD layers of water distribution systems often have a level of detail that is unnecessary for analyzing system hydraulics. This can include hydrants, service connections, valves, fittings, and corner nodes. Although development of models that are an exact replica of the real system is now possible due to enormously powerful computers, the cost of purchasing, developing, and running very large models can be excessive. Attempting to include all the network elements could be a huge undertaking that may not have a significant impact on the model results (Haestad Methods, 2003). Modelers, therefore, resort to a network simplification process, which is also referred to as the "skeletonization" process.

Network simplification is needed to increase processing speed without compromising model accuracy. Manual skeletonization is a cumbersome process. Modelers, therefore, have resorted to rules-of-thumb, such as excluding all pipes smaller than a certain size. This approach requires a significant effort in locating and removing the candidate pipes. In addition, the network integrity and connectivity might be adversely impacted, with a potential to alter the output results significantly.

An example of how GIS performs skeletonization is shown in Figure 12.10 to Figure 12.16. Figure 12.10 shows a portion of the water distribution system of the city of Sycamore, Illinois. It shows the GIS layers for parcels, water mains, service lines, service valves, distribution system valves, and fire hydrants. If a modeling node is created at each pipe junction, valve, and hydrant, the resulting network model (shown in Figure 12.11) would have 93 nodes and 94 pipes (links). If this process is extended to the entire system, we could expect thousands of nodes and pipes. For large

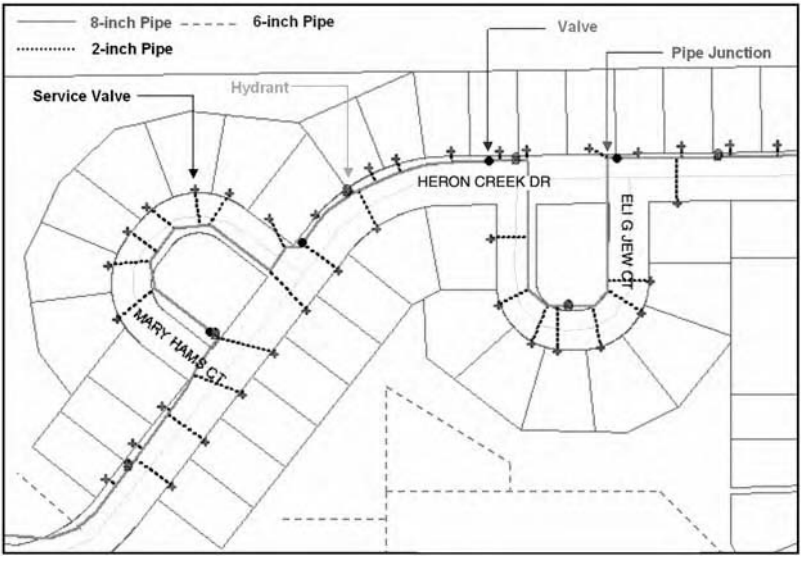


Figure 12.10 GIS layers for the Sycamore water distribution system.

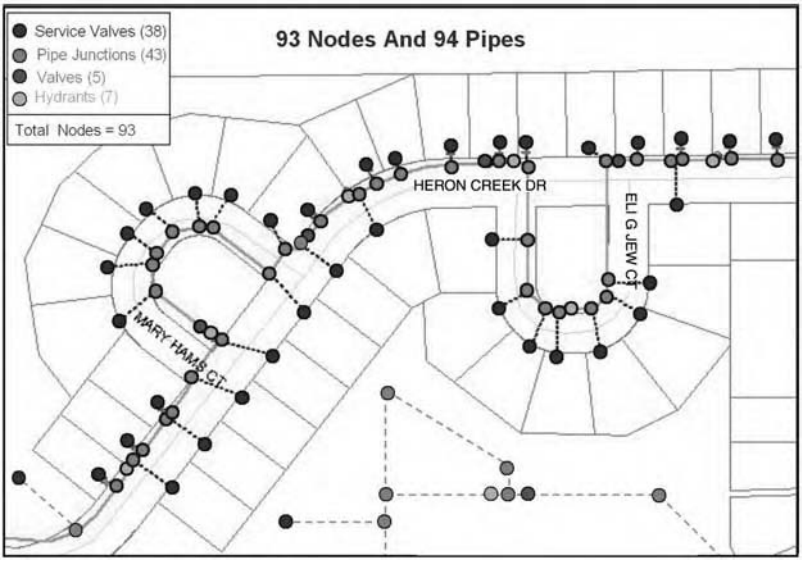


Figure 12.11 Network model without skeletonization.

systems, this process can generate several hundred thousand nodes and pipes. Skeletonization can be performed in a GIS using the following steps:

1. It is not necessary to create nodes at the fire hydrants because fire flow demand can be modeled at a nearby node. Thus, nodes corresponding to fire hydrants can be eliminated without compromising the modeling accuracy. Figure 12.12 shows the resulting skeletonized network with 86 nodes and 87 pipes. This represents an 8% reduction in the number of nodes and pipes.
2. It is not necessary to create nodes at all the distribution system valves. For example, the effect of an isolation valve (e.g., gate valves and butterfly valves) can be modeled as part of a pipe by using minor loss coefficients. In some models, closed valves and check valves can be modeled as links. Thus, nodes corresponding to most valves can be eliminated without compromising modeling accuracy. Figure 12.13 shows the skeletonized network after removing the valve nodes. It now has 81 nodes and 82 pipes representing a 13% reduction in the number of original nodes and pipes. This level of skeletonization is considered minimum.
3. It is not necessary to model corner nodes that describe actual layout of pipes. The pipe layout does not affect model output as long as the pipe length is accurate. Any potential head loss due to sharp turns and bends can be modeled using minor loss coefficients or friction factors. Thus, nodes corresponding to corner nodes can be eliminated without compromising modeling accuracy. Figure 12.14 shows the skeletonized network after removing the corner nodes. It now has 78 nodes and 79 pipes, representing a 16% reduction in the number of original nodes and pipes.

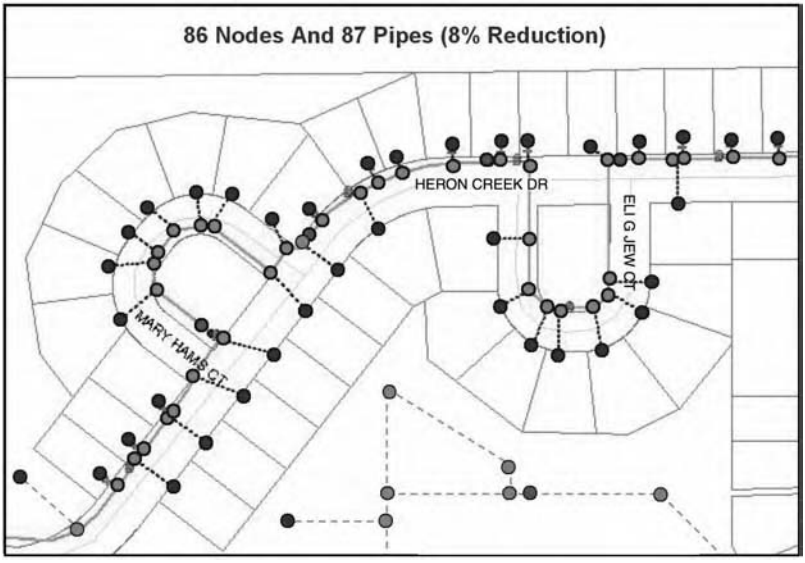


Figure 12.12 Skeletonized network model after removing nodes at fire hydrants.

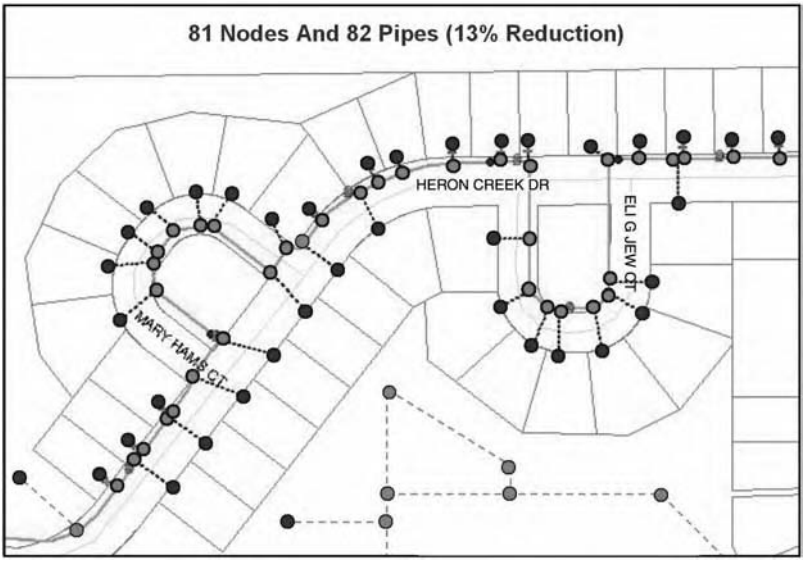


Figure 12.13 Skeletonized network model after removing nodes at valves.

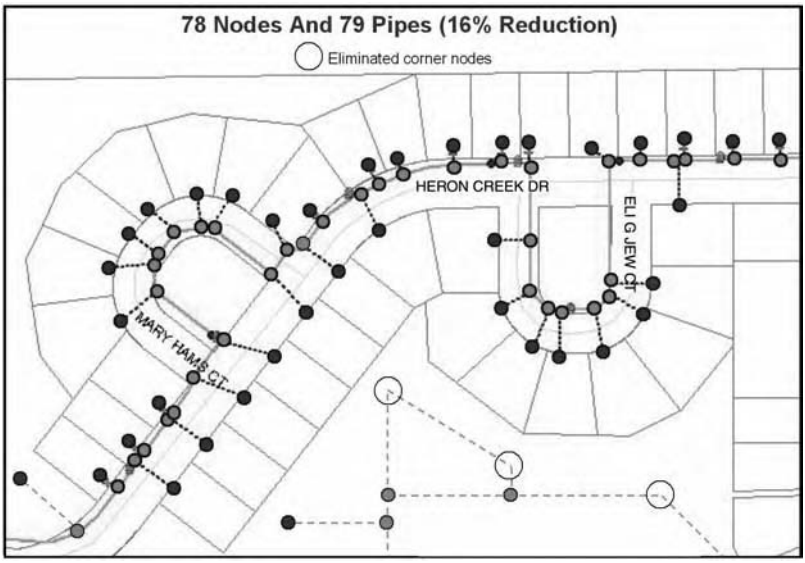


Figure 12.14 Skeletonized network model after removing corner nodes.

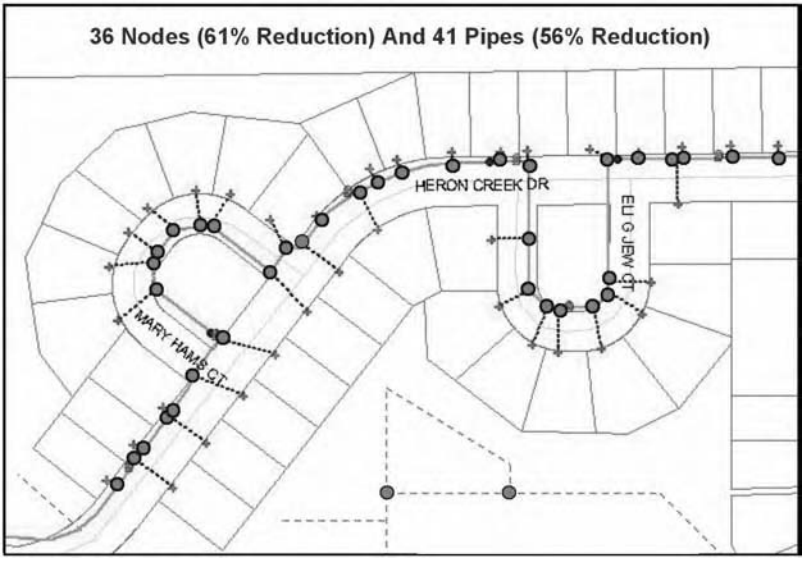


Figure 12.15 Skeletonized network model after removing pipes less than 8 in.

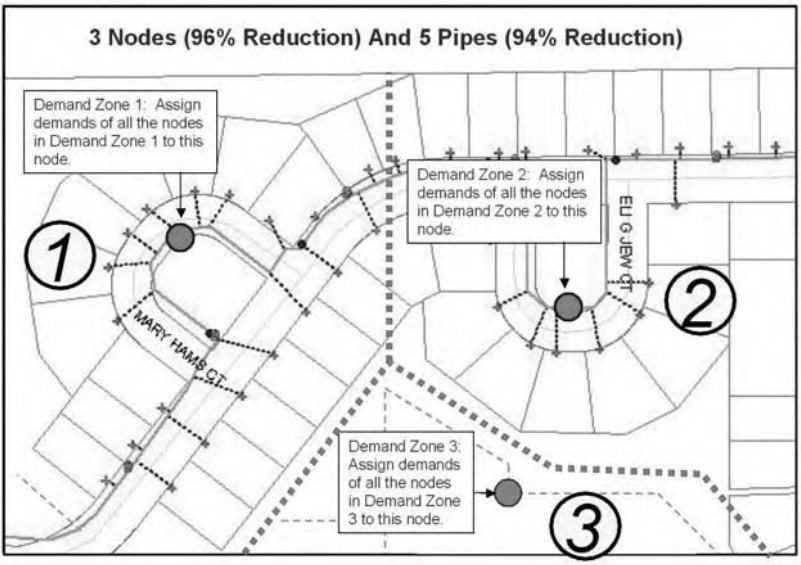


Figure 12.16 Skeletonized network model after aggregating demand nodes.

4. Small pipes with low flows can also be eliminated without significantly changing the modeling results. There are no absolute criteria for determining which pipes should be eliminated (Haestad et al., 2003). Elimination of pipes less than 6 in. or 8 in. is most common. Thus, small pipes and corresponding nodes can be eliminated for further network simplification. Figure 12.15 shows the skeletonized network after removing pipes of less than 8 in. width. The most dramatic effect of this step is the removal of all the service lines from the network model. The network now has 36 nodes and 41 pipes, representing a 56% reduction in the number of original nodes and pipes. This moderate level of skeletonization is acceptable in most cases.
5. Further reduction in the size of a network model is possible by aggregating the demands from several customers to a single node. Figure 12.16 shows the skeletonized network after dividing the network into three demand zones and assigning the total zone demand to a node inside that zone. The network now has only 3 nodes and 5 pipes, representing a 96% reduction in the number of original nodes and a 94% reduction in the number of original pipes. This high level of skeletonization can affect modeling results. For example, this model cannot model the flows and pressures at different houses within the subdivision.

Although the preceding steps can be performed manually in a GIS by editing the data layers, the process will be too cumbersome unless the system is very small. Automatic GIS-based skeletonization that can be completed quickly is now available in some software packages. Automatic skeletonization removes the complexity from the data, while maintaining the network integrity and hydraulic equivalency of the system. Reductions of 10 to 50% in the size of a network are possible. These programs allow the users to input the skeletonization criteria such as the smallest

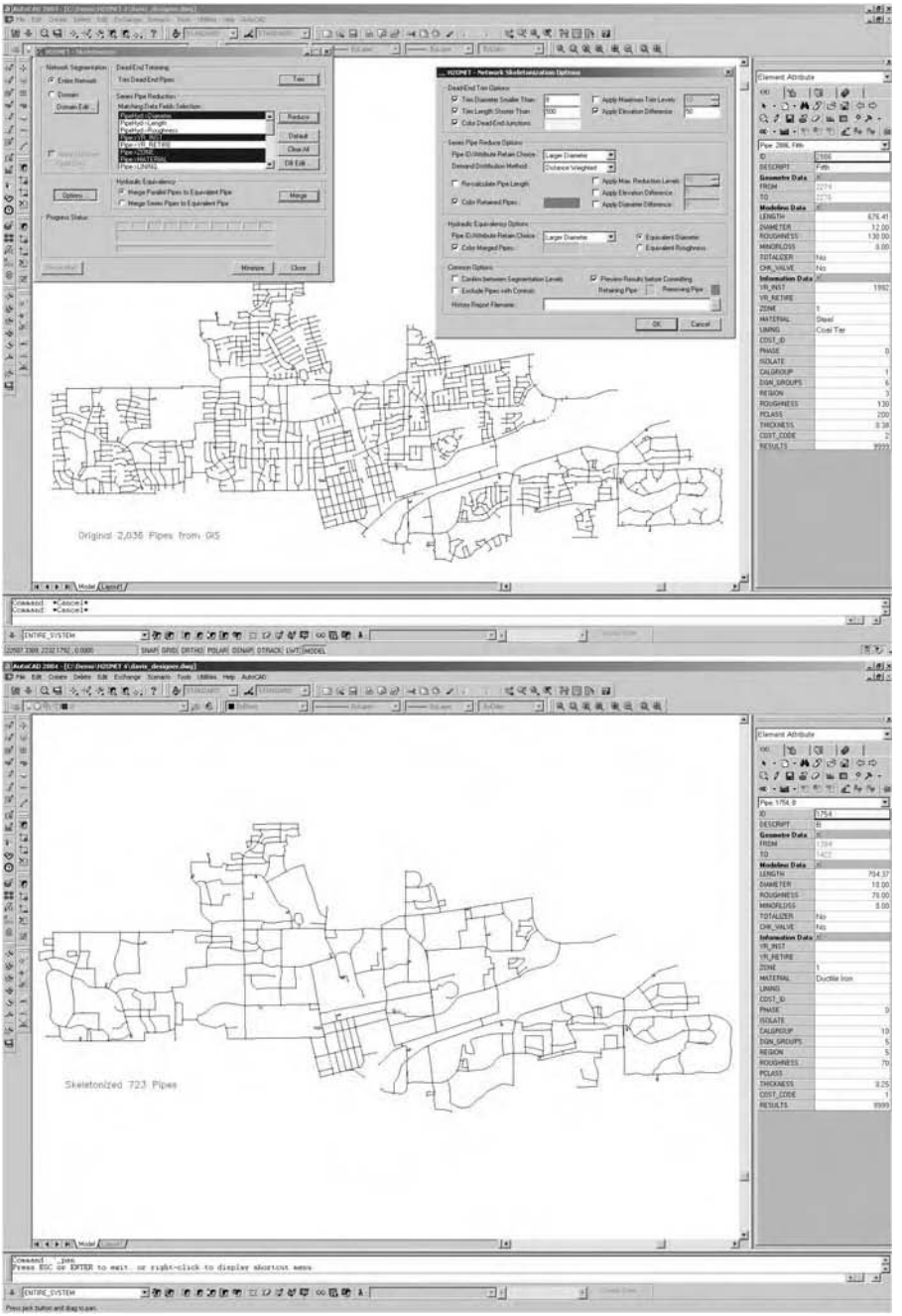


Figure 12.17 Skeletonizer's network simplification results. Top: original system with 2036 pipes. Bottom: skeletonized network with 723 pipes (65% reduction). (Maps courtesy of MWH Soft, Inc.)

pipe size that should be included in the model. The smallest-pipe-size criterion should be used carefully because both size and accuracy of the skeletonized model depend on it. A highly skeletonized model does not work well in areas where most of the water supply is through small pipes.

The Skeletonizer™ module of H_2OMap water system modeling software (MWH Soft, Inc., Broomfield, Colorado, www.mwhsoft.com) and Skelebrator™ module of WaterCAD and WaterGEMS water system modeling software (Haestad Methods, Waterbury, Connecticut, www.haestad.com) are examples of software tools that automate the skeletonization of CAD and GIS layers. Figure 12.17 shows an example of Skeletonizer skeletonization. The top screenshot shows the original network with 2036 pipes imported from a GIS. The bottom screenshot shows the skeletonized network with 723 pipes, indicating a 65% reduction in the size of the network.

Mohawk Valley Water (MVW) is a water utility serving more than 125,000 people in approximately 150 mi^2 area in central New York. In 2002, MVW created an all-pipe WaterCAD hydraulic model that was connected to a geodatabase. A reduce/skeletonize/trim (RST) process was used to eliminate hydrant laterals and private service connections to scale the hydraulic model to pipe segments 4-in. and greater. The demands, pipe attributes, and elevations were extracted from the geodatabase (DeGironimo and Schoenberg, 2002).

ESTIMATION OF NODE DEMANDS

Node demand is an important input parameter of a hydraulic model, representing the water consumption rate in units of volume/time (e.g., gallons per day or liters per second). Demands are grouped by customer type (e.g., residential, commercial, industrial, or institutional) or magnitude (e.g., average daily, maximum daily, peak hourly, or fire flow). Manual estimation of node demands from customer billing records is a cumbersome process. Modelers, therefore, have traditionally resorted to approximate methods, such as uniform distribution of the residential water use across all nodes followed by adding the nearest commercial and industrial use at various nodes. This approach requires a significant effort in fine-tuning the node demands during the model calibration phase of the project.

GIS allows implementation of more accurate demand estimation methods. One such method utilizes the GIS to geocode (or geolocate) all individual water users (defined by the address at which each meter is located) and to assign the metered water consumption rates for each meter location to the nearest node in the model. The geocoding process links a street address to a geographic location and finds the coordinates of a point location based on its street address. This method assumes that a computer file (spreadsheet or delimited text file) containing meter addresses and metered water use for specified durations (e.g., quarterly or monthly) can be produced from the utility's billing database. If this information is available, the following procedure can be used to assign demands to nodes:

1. Create a GIS layer (e.g., a Shapefile) representing all nodes in the model.
2. Obtain a street centerline layer of the system with address ranges. Such data can also be purchased from commercial GIS data vendors (Shamsi, 2002).
3. Obtain and preprocess customer meter location and consumption database.
4. Process the meter address and street address range data in the GIS to geocode the meter locations into a point layer registered to the model node layer.
5. Process the geocoded meter data to assign and aggregate individual demands to the nearest nodes.
6. Analyze the assigned metered demand to determine unaccounted-for water and metered demands that were not successfully geocoded. Uniformly distribute these remaining demands across nodes.
7. Load the node demand data into the hydraulic model.

The preceding process produces a highly accurate distribution of water use across the nodes. The intermediate product of point locations with associated demands can also be used to visualize the spatial distribution of individual demand points throughout the system. A metered demand database can also be created to include customer classes as defined in the billing system. This data can be used to assign differing demand patterns to residential, commercial, and industrial demand classes for use in extended-period simulations.

Haestad Methods' WaterGEMS hydraulic modeling software has a Loadbuilder module that allows modelers to automatically calculate and assign demands to junctions using the following three methods (Haestad Methods, 2002):

1. Customer meter data method: This method is used if geocoded customer billing data are available. It assigns the demands to nodes using any combination of the following three methods:
   - Nearest node method: This method assigns the meter demand to the nearest node.
   - Nearest pipe method: This method assigns the meter demand to the two nodes of the nearest pipe. Each node of the nearest pipe can receive equal demands or weighted demands in proportion to its distance from the meter location. Alternatively, the demand can be assigned to the pipe node closest to the meter location.
   - Meter aggregation method: This method creates *service areas* around the nodes of a hydraulic model. A service area polygon is associated with each model node, enclosing all service connections that consume their water from the node and whose consumption is aggregated to that node. Service areas can be drawn manually or using automatic methods such as the Theissen Polygon method. Programs are available for automatic delineation of Theissen Polygons in a GIS package. All the meters located within a service area are assigned to its associated node. Figure 12.18 shows the service areas created using the Thiessen Polygon method in WaterGEMS software.
2. Area water use data method: This method is used if billing meter data are not available. It assigns the demands to nodes using the following three methods:
   - Equal flow distribution: This method requires a polygon layer for various flow zones (e.g., meter routes or pressure zones) in the system. A GIS overlay of nodes and flow zone layers is conducted to divide each zone's flow equally among all the nodes within that zone.

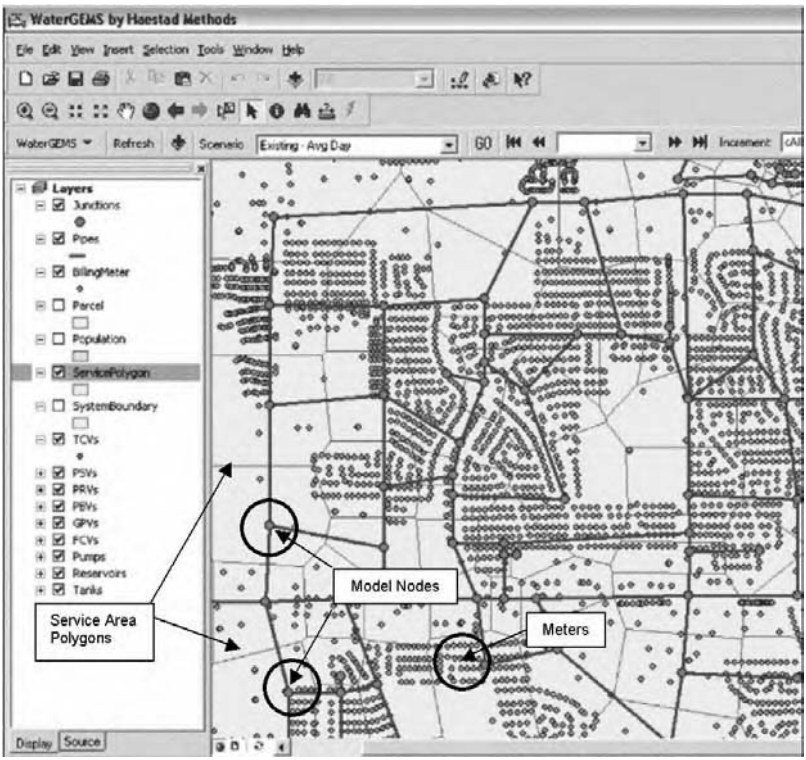


Figure 12.18 Example of meter aggregation method (WaterGEMS™ screenshot courtesy of Haestad Methods).

- Proportional distribution by area: In this method, a GIS overlay of service area (of nodes) and flow zone layers is conducted. Each zone's flow is divided among all the service area polygons within that zone in proportion to the area of service area polygons.
- Proportional distribution by population: In this method, a GIS overlay of service area (of nodes) and flow zone layers is conducted. Each zone's flow is divided among all the service area polygons within that zone in proportion to the population of service area polygons. The service area population can be calculated in GIS by conducting an overlay of service area polygons with census-block* (or census-tract) polygons.

3. Land-use/population data method: This method forecasts the future demands for nodes using the following two methods:
   - Land-use data method: This method uses future land-use polygons to forecast future demands. A GIS overlay of service area and land-use polygons is conducted to calculate the area of each land-use type within each service area. A demand factor (per unit area) is specified for each land-use type. Calculated

* Census blocks and census tracts from U.S. Bureau of Census are polygon layers of demographic data for the U.S. Additional information is available in Chapter 11.

land-use areas and demand factors are multiplied to estimate a future demand for each service area.

- Population data method: This method uses projected future population polygons to forecast future demands. A GIS overlay of service area and population polygons is conducted to calculate the future population of each service area. A demand density (gallons per capita per day) is specified by the user. Calculated future population and demand density are multiplied to estimate a future demand for each service area.

Detailed information on the preceding methods is available in *Advanced Water Distribution Modeling and Management* (Haestad et al., 2003). For future land-use mapping techniques and information on how to calculate population from U.S. Bureau of Census GIS data (e.g., Census blocks and tracts), please refer to *GIS Tools for Water, Wastewater, and Stormwater Systems* (Shamsi, 2002).

Demand-Estimation Case Studies

Newport News, Virginia

In 1996, Newport News Waterworks (Newport News, Virginia) developed a water system hydraulic model composed of 540 nodes representing approximately 110,000 service connections. The model contained all pipes 12 in. and larger and some 8 in. critical pipes. Model node demands were computed using the meter aggregation method described in the previous section. The service area polygons were drawn manually, which was found to be the largest engineering effort required by the modeling project. A GIS Demand Application was developed to create, modify, or delete model nodes and service area polygons. The application was developed using ArcTools, a collection of ArcInfo productivity tools implemented through an Arc Macro Language (AML) interface.

The demand data were extracted from a public utilities billing system. Average daily consumption of each customer was calculated by dividing the total consumption by the number of days in the billing cycle. The consumption data were loaded in a tabular GIS file. Account numbers were used to associate demands with service connections. Consumption for all service connections that geographically fell within a service area polygon were aggregated. The aggregated consumption was stored as an attribute of the model node (Sevier and Basford, 1996).

Round Rock, Texas

The City of Round Rock (Texas) used Haestad Methods' WaterCAD hydraulic modeling software, ESRI's ArcView GIS package, and Microsoft's Access billing database for estimating the node demands. Approximately 95% of the city's 22,000 water meters were assigned to precise geographic locations using geocoding. This

approach provided accurate node demand data and reduced model calibration time (Haestad Methods, 2002). The following steps were undertaken:

1. The geocoding process was performed in ArcView by linking the addresses within an existing parcels Shapefile to the billing address in the water meter database.
2. Using WaterCAD's Shapefile connection wizard, the node data from WaterCAD was synchronized to ArcView.
3. The modelers then drew polygons around the nodes and corresponding meters.
4. ArcView was used to sum the demand data for the meters in each polygon.
5. The total demand was applied to appropriate nodes.
6. The junction file with the newly calculated node demands was synced back to WaterCAD.

Lower Colorado River Authority, Texas

Pimpler and Zhan (2003) used an alternative grid-based (raster GIS) approach that worked well for estimating water demand for Lower Colorado River Authority. The Authority serves more than 1 million customers in Texas in an area spanning more than 29,800 mi^2. Their method involved four steps:

1. Divide the service area into grid–cells.
2. Calculate a score for each grid–cell based on weighted variables (e.g., land use, proximity to employment centers, etc.) considered important for water demand analysis.
3. Distribute the service area population proportionally to each grid–cell based on the score assigned to each cell.
4. Estimate water demand associated with each cell based on its allocated population.

ESTIMATION OF NODE ELEVATIONS

Node elevation is also an important input parameter for a hydraulic model. It is usually defined by the ground elevation. Because 2.31 ft of elevation corresponds to 1 psi of water pressure, modeling water pressure to 1-psi precision requires elevation data accurate to approximately 2 ft. Table 12.4 provides the accuracy of common sources of ground surface elevation. USGS topographic maps are generally acceptable because an elevation accuracy of 5 ft is adequate for most nodes (Walski, 2001).

Surveying provides the most accurate ground surface elevations. Unfortunately, surveying is very costly and time consuming and is recommended for critical model nodes such as the centerline of a pump or a model calibration node (Walski, 2001). Manual estimation of node elevations from contour maps or as-built plans is also a

Table 12.4 Accuracy of Ground Elevation Data Sources

Source	Accuracy (ft)
USGS 7.5-min 1:24,000-scale topographic maps	10
USGS 30-m DEMs	5–10
GPS survey — code phase	10
GPS survey — carrier phase	1
Digital orthophotos	1

cumbersome process. A GIS allows automatic computation of node elevations from DEM data using the following procedure:

1. Create a GIS layer (e.g., a Shapefile) for the nodes in the model.
2. Obtain a DEM layer covering the water distribution service area.
3. Georeference (georegister) the node and DEM layers (i.e., assign both layers the same map projection system).
4. Overlay the node layer over the DEM layer to assign DEM elevations to nodes. Interpolate in between the DEM grid cells (pixels) to estimate the correct elevation at each node.
5. Load the node elevation data into the model.

Figure 12.19 shows an overlay of a water distribution system layer over a DEM layer for calculating node elevations.

WaterGEMS has a TRex Wizard that automates extraction of elevation data from a DEM layer and assigns the extracted data to model nodes. TRex uses the preceding steps to calculate node elevations. It interpolates in between the DEM pixels (grid–cells) rather than using a simple spatial join. A spatial join is less accurate

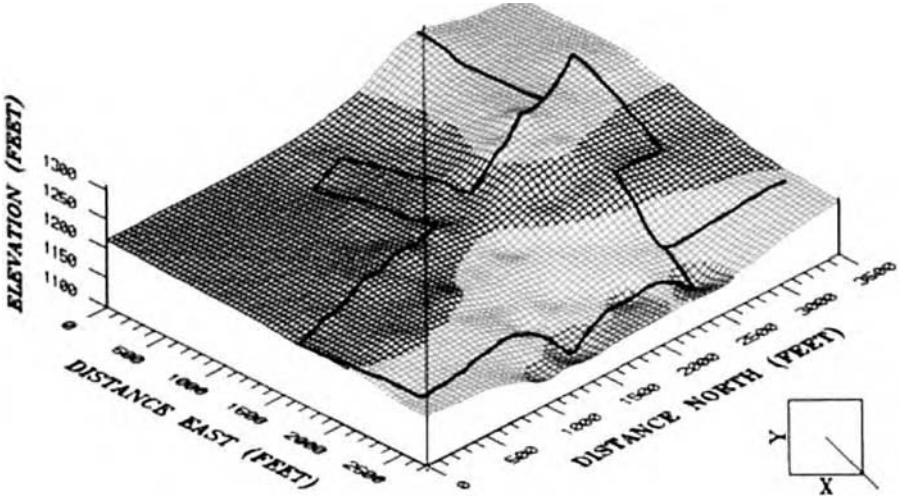


Figure 12.19 Water distribution system draped over a DEM

because nodes that fall in between the DEM pixels would be assigned the elevation of the nearest cell.

PRESSURE ZONE TRACE

A GIS also allows identification of various pressure zones in a water distribution network. This is done by selecting a network pipe or node that is part of a larger pressure zone. The GIS traces the water network in all possible directions until either sources of water or assets (e.g., pressure regulating valves — PRVs) that isolate water flow are found. All the pipes (or nodes) in the selected pressure zone are highlighted. A buffer can be created around the highlighted features to create a polygon for the pressure zone boundary. Once the trace is complete, the user can assign a pressure zone attribute to the selected pipe or node.

CHAPTER SUMMARY

This chapter shows that GIS offers many useful applications for water distribution systems. The applications described in this chapter indicate that GIS provides many powerful and efficient tools for performing various hydraulic modeling tasks such as network skeletonization, demand calculation and allocation, and elevation calculations. The case studies and examples presented in the chapter illustrate that many water utilities are successfully using GIS applications for hydraulic modeling of their water distribution systems. Various software examples presented in the chapter indicate that many commercial off-the-shelf software packages are available to help users benefit from hydraulic modeling applications of GIS.

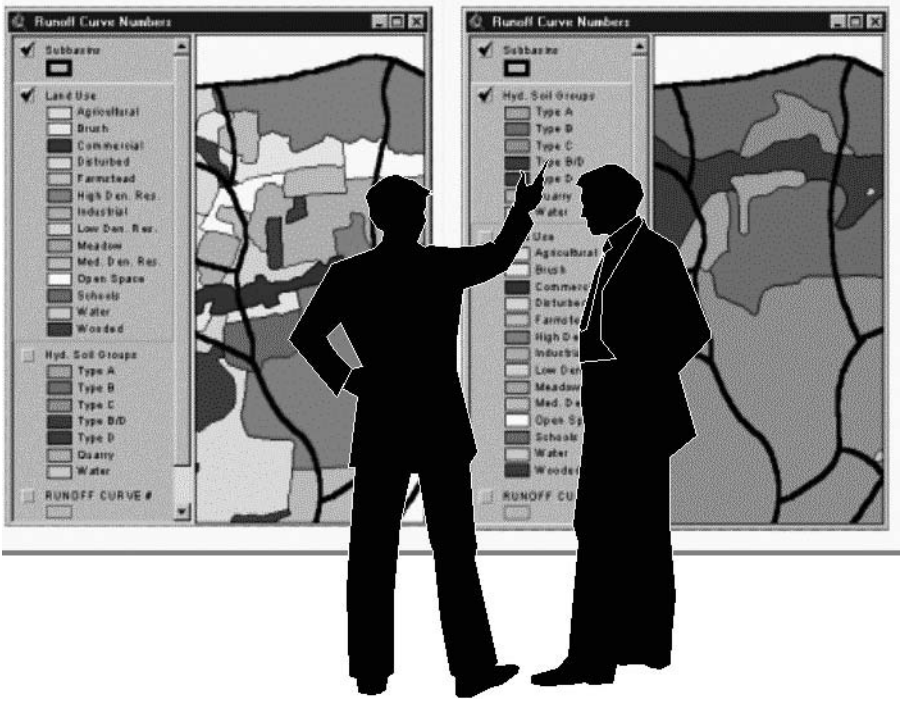
CHAPTER QUESTIONS

1. Prepare a list of GIS applications for water distribution systems.
2. Which hydraulic modeling tasks can be performed using GIS?
3. What are the typical input parameters for a water distribution system hydraulic model? Which of these parameters can be obtained from a GIS?
4. List the GIS data layers that are useful for hydraulic modeling.
5. How does a GIS help in presenting model output results?
6. What is skeletonization? How does a GIS help this task?
7. List the hydraulic models you are familiar with. Which of these models use a GIS? Which method of GIS linkage (interchange, interface, or integration) do they use?

CHAPTER 13

Sewer Models

GIS saves time and money in developing sewer system models for simulating flows and depths in the collection system. GIS also helps to prepare maps of the model results, which can be easily understood by nonmodelers.



GIS-based sewer models bridge the gap between information and its recipients.

LEARNING OBJECTIVES

The learning objectives of this chapter is to understand how GIS can be applied in developing sewer system hydraulic models, and to understand how GIS can be used to present modeling results

MAJOR TOPICS

- Representative sewer models
- Storm Water Management Model (SWMM)
- Graphical user interface (GUI)
- SWMM applications
- Data preparation
- Interface development steps
- GIS application case studies

LIST OF CHAPTER ACRONYMS

CMOM Capacity, Management, Operations, and Maintenance
CSO Combined Sewer Overflow
DSS Decision Support System
EPA (U.S.) Environmental Protection Agency
GUI Graphical User Interface
HEC Hydrologic Engineering Center
HGL Hydraulic Gradient Line
LTCP Long-Term Control Plan (for CSOs)
NMC Nine Minimum Controls (for CSOs)
NPDES National Pollution Discharge Elimination System
ODBC Open Database Connectivity
OLE Object Linking and Embedding
SHC System Hydraulic Characterization (for CSOs)
SSO Sanitary Sewer Overflow
SWMM Storm Water Management Model

MAPINFO™ AND SWMM INTERCHANGE

Application area	Basement flooding modeling using the Interchange GIS Linkage Method described in Chapter 11 (Modeling Applications)
GIS software	MapInfo
Modeling software	SWMM
Other software	Microsoft Access
GIS data	Sewer pipes, aerial photographs, buildings and parking lot polygons
GIS format	Raster and vector
Study area	City of Portland, Oregon
Organization	Bureau of Environmental Services, City of Portland, Oregon

In the late 1990s, Portland's Bureau of Environmental Services (BES) developed a suite of GIS and database tools to develop data for SWMM models. Model data were

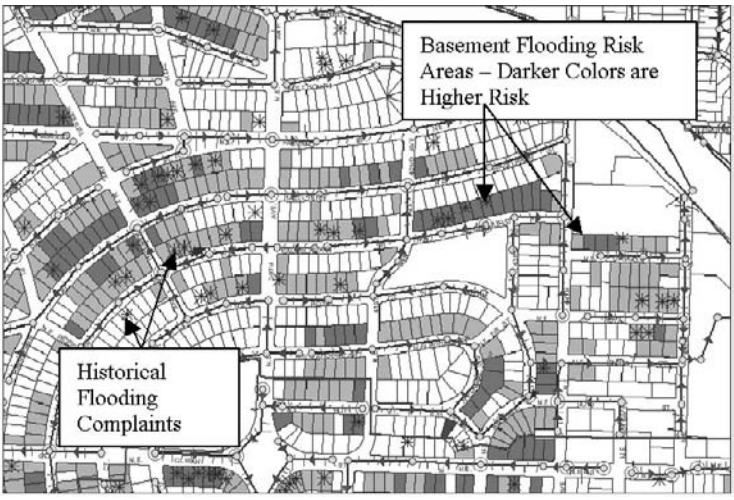


Figure 13.1 Thematic map showing flooding risk and historical complaints (map courtesy of Computational Hydraulic Int.).

developed from maps and databases with automation of many of the steps for data extraction, variable assignments, and model building. Detailed collection-system models of basins were developed to analyze local areas. These models contained essentially all of the pipes in the collection system and utilized two subcatchment definitions to place runoff into the system in the correct location. Surface water subcatchments were defined as the areas that flowed into the sewer through street inlets in the public right-of-way. Direct connection subcatchments were defined as the areas that contributed to sanitary and stormwater drainage through service laterals.

A GIS tool was developed to predict flooding of individual parcels by overlaying modeled hydraulic gradient lines (HGLs) with estimated finished floor elevations obtained from a DEM. Detailed SWMM RUNOFF and EXTRAN models were developed from facility maps and photogrammetric data with the use of a MapInfo GIS and a Microsoft Access database. The SWMM model showed that many of the system capacity problems were in the local areas at the upstream ends of the system and that the local collection system tended to hold water back from the downstream trunk sewers in the system. This better understanding of where flooding was likely to occur and where system capacity constraints actually occurred allowed BES a top-down view of the basin during alternatives analysis. Local inflow control alternatives were considered first, and pipe upsizing along with passing the problem downstream was the second alternative. Figure 13.1 shows a thematic map showing flooding risk and historical complaints (Hoffman and Crawford, 2000).

GIS APPLICATIONS FOR SEWER SYSTEMS

GIS applications for sewer systems include (Shamsi, 2002):

- Performing H&H modeling of collection systems, including:
  - Automatic delineation of watersheds, sewersheds, and tributary drainage areas

    - Model simplification or skeletonization to reduce the number of manholes and sewers to be included in the H&H model
    - Estimating surface elevation and slope from digital elevation model (DEM) data
    - Estimating dry-weather sewage flow rates from land use, census data, and billing records
    - Estimating wet-weather sewage flow rates from land use, soil, surface imperviousness, and slope
- Creating maps for wastewater National Pollution Discharge Elimination System (NPDES) permit requirements (as described in Chapter 9 [Mapping Applications]), such as:
    - U.S. Environmental Protection Agency's (EPA) combined sewer overflow (CSO) regulations, such as System Hydraulic Characterization (SHC), Nine Minimum Controls (NMC), and long-term control plan (LTCP)
    - U.S. EPA's sanitary sewer overflow (SSO) regulations, such as capacity, management, operations, and maintenance (CMOM)
- Documenting field inspection data (as described in Chapter 15 [Maintenance Applications]), including:
    - Work-order management by clicking on map features
    - Inspection and maintenance of overflow structures and manholes
    - Television (TV) inspection of sewers
    - Flow monitoring and sampling
    - Smoke-testing, dye-testing, and inflow/infiltration (I/I) investigations
- Performing and planning tasks such as assessment of the feasibility and impact of system expansion

In this chapter, we will focus on the sewer system H&H modeling applications of GIS.

SEWER SYSTEM MODELING INTEGRATION

Tight integration between a sewer model and a GIS is highly desirable.

The GIS applications in sewer system modeling are developed by using the three application methods (interchange, interface, and integration) described in Chapter 11 (Modeling Applications). These methods facilitate preparation of model-input data and mapping of model output results similar to their application for water system models described in Chapter 12 (Water Models).

Transfer of data between a GIS and a sewer model is handled differently by different software vendors. Some products offer nothing more than a manual cut-and-paste approach for transferring data between a GIS and the model. Others offer a truly integrated package. Obviously, tight integration between the sewer model and the GIS is desirable. However, it should not be assumed that a tightly integrated model is the best modeling tool. For example, the interchange or interface capability of a comprehensive sewer model might be more useful for some users compared with a tightly integrated model having limited modeling capability to simulate complex situations such as surcharged sewers.

Table 13.1 Representative Sewer System Modeling Software

Software	GIS Linkage Method	Vendor	Web Site
SWMM	Interchange	U.S. Environmental Protection Agency (EPA)	www.epa.gov/ceampubl/swater/swmm/index.htm
CEDRA AVSand	Integration	CEDRA Corporation	www.cedra.com
H_2OMAP Sewer H_2OVIEW Sewer	Interface and Integration	MWH Soft	www.mwhsoft.com
InfoWorks CS and InfoNet	Interface and Integration	Wallingford Software	www.wallingfordsoftware.com
Mouse GM and MIKE SWMM	Interface	DHI Water & Environment	www.dhisoftware.com
PCSWMM GIS	Integration	Computational Hydraulics Int.	www.computationalhydraulics.com
StormCAD and SewerCAD	Interface	Haestad Methods	www.haestad.com
XP-SWMM	Interchange and Interface	XP-Software	www.xpsoftware.com

SOFTWARE EXAMPLES

Sewer system modeling software is used to compute dry- and wet-weather flow for hydraulic analysis and design of collection systems and wet-weather control facilities such as storage tanks and equalization basins.

Table 13.1 lists representative sewer system modeling packages and their GIS capabilities, vendors, and Web sites. The salient GIS features of some programs are described in the following subsections.

What is common among the recent software developments is a transferability of fundamental database information, also referred to as a decision support system (DSS). Under a DSS framework, neither the GIS nor the model is central to the process. Both perform satellite functions for a central master database. (Heaney et al., 1999).

SWMM

Numerous hydrologic models were created in the U.S. during the 1970s, including the U.S. EPA's legacy computer program SWMM and the U.S. Army Corps of Engineers' Hydrologic Engineering Center's HEC series of models (HEC-1 through 6). Two of the most popular models, HEC-1 and HEC-2, have been updated and renamed HEC-HMS and HEC-RAS.

Collection system hydraulics can be characterized by using an H&H computer model such as SWMM. SWMM is a comprehensive computer model for analysis of quantity and quality problems associated with urban runoff. Both continuous and single-event simulation can be performed on catchments having sanitary

sewers, storm sewers, or combined sewers and natural drainage, for prediction of flows, stages, and pollutant concentrations.

SWMM is one of the most successful models produced by EPA for the wastewater environment. SWMM was developed from 1969 to 1971 as a mainframe computer program and has been continually maintained and updated. Versions 2, 3, and 4 of SWMM were distributed in 1975, 1981, and 1988, respectively. The first batch-mode microcomputer version of SWMM (Version 3.3) was released by EPA in 1983. The first conversational-mode user-friendly PC version of SWMM, known as PCSWMM, was commercially distributed in 1984 by Computational Hydraulics, Inc., of Guelph, Ontario (Computational Hydraulics, 1995). The first PC version of EPA's SWMM was Version 4, which was distributed in 1988 (Huber and Dickinson, 1988; Roesner et al., 1988). An excellent review of SWMM's development history can be found in the book by James (1993).

SWMM is regarded as the most widely used urban H&H model in the U.S. (Heaney et al., 1999). SWMM continues to be widely used throughout the world for analysis of quantity and quality problems related to stormwater runoff, combined sewers, sanitary sewers, and other drainage systems in urban areas, with many applications in nonurban areas as well. From 1988 through 1999, the EPA Center for Exposure Assessment Modeling in Athens, Georgia, distributed approximately 3600 copies of SWMM. The University of Florida distributed roughly 1000 copies of SWMM in the late 1980s. Third-party interfaces for SWMM, such as MIKE-SWMM, PCSWMM, and XP-SWMM, have several thousand users. The number of subscribers to the SWMM Users Group Internet discussion forum (Listserv) on the Internet is nearly 10,000 (EPA, 2002).

The current version (4.3) was released by EPA in May 1994 for the 16-bit MS-DOS operating system. Professor Wayne Huber of Oregon State University, one of the authors of SWMM, has developed several updated versions of SWMM 4. His latest version (4.4H) was updated in March 2002. All of these and previous SWMM versions were written in the FORTRAN programming language. None of them have a user interface or graphical capability. Users must provide ASCII text input and rely on ASCII text output. EPA is currently developing an updated version of SWMM referred to as SWMM 5. Written in the C++ programming language, SWMM 5 will incorporate modern software engineering methods as well as updated computational techniques. The modular, object-oriented computer code of SWMM 5 is designed to simplify its maintenance and updating as new and improved process submodels are developed. The new code will also make it easier for third parties to add GUIs and other enhancements to the SWMM engine. This approach follows the same approach used by the highly successful EPANET model from EPA (described in Chapter 12 [Water Models]), which analyzes hydraulic and water quality behavior in drinking water distribution systems (Rossman, 2000).

SWMM simulates dry- and wet-weather flows on the basis of land use, demographic conditions, hydrologic conditions in the drainage areas, meteorological inputs, and conveyance/treatment characteristics of the sewer system. The modeler can simulate all aspects of the urban hydrologic and quality cycles, including rainfall, snow melt, surface and subsurface runoff, flow routing through drainage

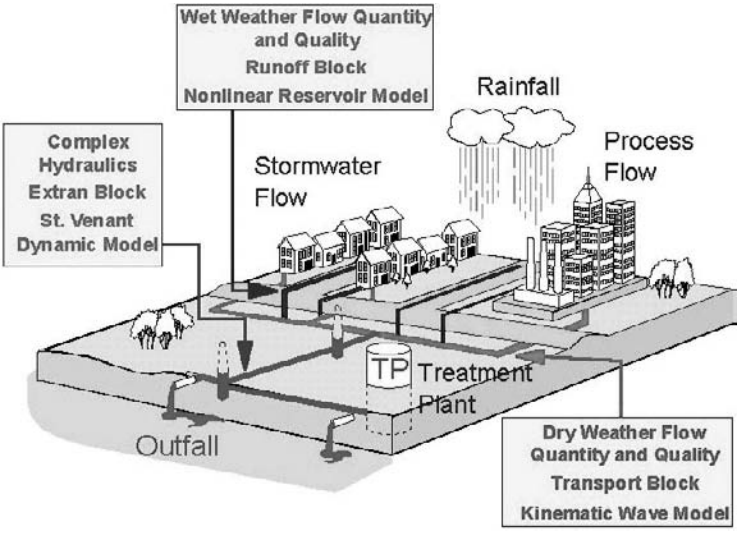


Figure 13.2 SWMM conceptual schematic.

network, storage, and treatment. Statistical analyses can be performed on long-term precipitation data and on output from continuous simulation. Figure 13.2 shows a conceptual schematic of SWMM. This figure shows that SWMM is a complex model capable of modeling various phases of the hydrologic cycle using different blocks (modules) such as RUNOFF, TRANSPORT, and EXTRAN.

SWMM can be used both for planning and design. Planning mode is used for an overall assessment of the urban runoff problem or proposed abatement options. SWMM is commonly used to perform detailed analyses of conveyance system performance under a wide range of dry- and wet-weather flow conditions. These days, SWMM is frequently used in the U.S. for modeling wet-weather overflows including CSO, SSO, and stormwater discharges from collection systems. As such, it is the model of choice for use in many collection-system modeling studies. For example, SWMM can be used to develop a CSO model to accomplish various tasks leading to the development of a CSO Plan of Actions mandated by EPA. These tasks include characterizing overflow events, developing the CSO vs. rainfall correlation, maximizing the collection-system storage, and maximizing the flow to the treatment plant. The CSO model can also be used to develop the EPA-mandated LTCPs to evaluate various CSO control options. Sizing of CSO control facilities such as a wet-weather equalization tank requires CSO volume and peak discharge, both of which can be modeled using SWMM.

The use of SWMM to model wet-weather overflows is particularly advantageous for the following reasons:

- SWMM produces estimates of present and future dry- and wet-weather flow rates. Flow estimates can be prepared based upon present and future land-use

conditions, topography, sewer characteristics, and selected meteorological conditions. The model can be calibrated against measured flow rates.
- SWMM models the performance of the conveyance system under a range of dynamic flow conditions.
- Using SWMM, it is possible to assess hydraulic capacity in response to wet-weather input. This characteristic can be very useful for analyses related to abatement of overflows.

SWMM is flexible enough to allow for different modeling approaches to the same area. An approach that adequately describes the service area and accurately computes and routes the flows at reasonable computing time and effort should be adopted. The following modeling strategy is generally used for modeling wet-weather overflows:

- Delineate sewersheds (the drainage areas tributary to overflows, also referred to as subareas or subbasins).
- Use the TRANSPORT Block to generate sewershed dry-weather flows.
- Use the RUNOFF Block to generate sewershed wet-weather flows.
- Combine dry-weather and wet-weather flows to generate combined sewershed flows.
- Use the EXTRAN Block to route the flows through the collection and interceptor system.

USEFUL SWMM WEB SITES

SWMM4	www.epa.gov/ceampubl/swater/swmm/index.htm
SWMM5	www.epa.gov/ednnrmrl/swmm/
Oregon State University	ccee.oregonstate.edu/swmm/
SWMM-Users Internet Discussion Forum	www.computationalhydraulics.com/Community/Listservers/swmm-users.html
SWMM-Online	www.swmm-online.com

SWMM Graphical User Interface

Most users have now become accustomed to modeling in a point-and-click computing environment that provides a user-friendly graphical user interface (GUI). A GUI is a computer program that acts as an interpreter between users and their computers. It is designed to minimize (but not eliminate) the need for human experts and to guide the modeler through the intricacies of a particular numerical model. A GUI provides a suite of tools to create a decision support system for the numerical model that has been adopted. It also stimulates user interest and facilitates interpretation of model results. A GUI improves productivity by increasing the efficiency of data entry, eliminating data errors through expert-checking, and the use of decision-support graphics and interpretation tools. It replaces difficult-to-remember text commands by interactive computer graphics consisting of menus, dialogue boxes, input and output windows, and icons. The main goal of using GUIs is to develop user-friendly computer applications or to

add user-friendliness to existing command-line-driven applications. For example, Microsoft Internet Explorer is a GUI for user-friendly access to the Internet.

Modern software offer pull-down menus, toolbars, icons, buttons, dialogue boxes, hypertext, context-sensitive online help, etc., which are not available in legacy computer programs including SWMM 4. Many commercial modeling packages now offer on-screen point-and-click drawing and editing of drainage network maps, but unfortunately some legacy programs still rely on ASCII text input and output files. SWMM was developed in an era when input files were created on punched cards. After 30 years, SWMM 4 now runs on personal computers, but it is still a text-based, nongraphical DOS program. It reads ASCII input to produce ASCII output, which is most suitable for mainframe line printers. SWMM's ASCII format output is long, boring, difficult to interpret, and not very useful for nonmodelers. Creating computer models and reviewing the model results is often slowed by our inability to see the system being modeled. It is up to the modeler to review SWMM's voluminous output and construct a mental image of the physical system being modeled. Often, the limitation in understanding the model output has been the modeler's own comprehension of the output, not the model itself. Quite frequently, it is impossible for the modeler to absorb the large amount of information contained in the model output (TenBroek and Roesner, 1993).

A GUI and a GIS can be employed to overcome SWMM's input/output deficiencies. Two types of GUIs can be employed for GIS applications in SWMM modeling (Shamsi, 1997; Shamsi, 1998):

- An input interface (also called a front-end interface or preprocessor) extracts SWMM input from GIS layers and creates SWMM's traditional ASCII text input file. The input GUIs can also provide graphical tools to draw a network model that is subsequently converted to SWMM's ASCII input file. For example, an input interface may extract sewer-segment lengths and manhole coordinates from the sewer and manhole layers.
- An output interface (also called a back-end interface or postprocessor) converts text to graphics. It can convert SWMM's traditional ASCII text output file to graphs, charts, plots, and thematic maps that can be easily understood by everyone.

With the help of these GUI/GIS tools, everybody can understand the model output. These tools help one to see storm-surge progress through the sewer system and pinpoint areas of flooding and surcharging. More than just reams of computer paper, such models become automated system-evaluation tools. Other benefits of GUI/GIS tools are:

- Preparation of network schematics is not essential. Digitized plots of sewers and subarea boundaries can be used to create a drainage network diagram on a computer screen.
- Zoom and pan features make it possible to display even the largest networks conveniently on the screen.
- Connectivity data errors are easily detected and can be edited while still in the program. Instabilities in the model output, often the most difficult errors to find, are also easily located.

- Flow and depth data from SWMM output can be displayed in either plan or profile view, providing an animated display of the HGL during the simulation time steps.
- Flow and HGL time-series plots can be displayed for any conduit or node in the hydraulic network.
- Field-collected flow and depth data can be displayed along with the model output for model calibration and verification.
- Network graphics and modeled hydrographs can be exported to word processors to aid in report preparation.

XP-SWMM and XP-GIS

XP-SWMM by XP Software (Belconnen, Australia and Portland, Oregon) is a full-fledged 32-bit Microsoft Windows application. The program has been enhanced by the addition of a graphics database, and an adaptive dynamic wave solution algorithm that is more stable than the matrix method used in the original SWMM. The program is divided into a stormwater layer, which includes hydrology and water quality; a wastewater layer, which includes storage treatment and water quality routing for BMP analysis; and a hydrodynamic/hydraulics layer for simulation of open or closed conduits (Heaney et al., 1999).

XP-SWMM is also included in Visual SWMM from CaiCE Software Corp. (Tampa, Florida). Basically, XP-SWMM and Visual SWMM are GUI programs for SWMM. The user-friendly GUI is based upon a graphical representation of

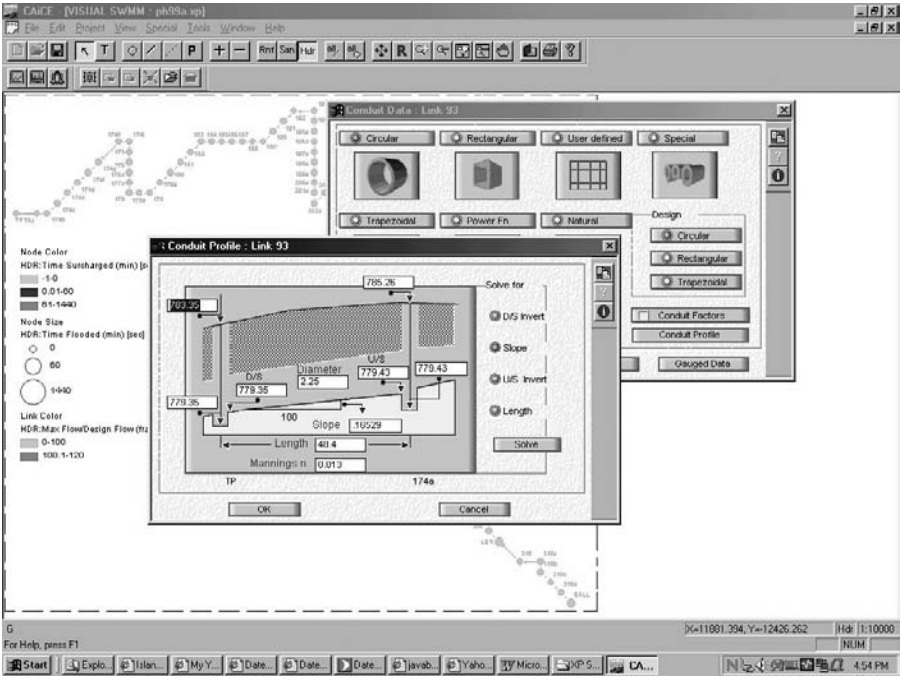


Figure 13.3 Visual SWMM screenshot.

the modeled system using a link-node architecture. It provides a shell that acts as an interpreter between the user and the model. It facilitates on-screen development of a SWMM model in the form of a link-node network. Figure 13.3 shows a Visual SWMM (Version 7) screenshot.

Because the links and nodes are set up on a coordinate-system basis, files can be translated between most CAD and GIS software systems using the interchange method. CAD or GIS files can also be used as a backdrop for the system being modeled. Although XP-SWMM is not linked or integrated to a GIS package, it provides an optional utility called XP-GIS to extract model input data from the existing GIS database tables. XP-GIS is a module for linking XP-SWMM to GIS databases. Its main purpose is to facilitate the import and export of modeling data from GIS and other data sources such as spreadsheets, asset management software, and OLE/ODBC compliant databases. It also allows for the inclusion of Shapefiles as background layers and for data to be viewed and manipulated in an XP-SWMM graphical environment.

GIS Data for SWMM

GIS layers for land use, elevation, slope, soils, and demographics can be processed to extract model parameters for input to various SWMM Blocks and other collection system models. The basic input parameters to be assembled into the RUNOFF Block input files include sewershed area, overland flow slope, overland flow width, percent imperviousness, roughness coefficients, and soil infiltration parameters. Area, slope, and percent imperviousness can be computed from the GIS layers. Table 13.2 shows sample sewershed RUNOFF data extracted from GIS layers using the interchange method for a study area in Pennsylvania.

GIS layers can also be used to estimate inputs for sewershed population, population density, number of houses, average family size, average market value of houses, average family income, and the predominant land-use type for the TRANSPORT Block. Table 13.3 shows sample sewershed TRANSPORT data extracted from GIS layers using the interchange method for a study area in Pennsylvania.

Estimating Green-Ampt Parameters Using STATSGO/SSURGO GIS Files

Based on their resolution, there are three types of U.S. Natural Resources Conservation Service (NRCS) (formerly the U.S. Soil Conservation Service or SCS) soils data that are useful in GIS applications:

- National Soil Geographic (NATSGO)
- State Soil Geographic (STATSGO)
- Soil Survey Geographic (SSURGO)

SSURGO provides the highest resolution soils data at scales ranging from 1:12,000 to 1:63,360. This resolution is appropriate for watersheds a few squares miles in area. STATSGO data are digitized at 1:250,000-scale, which is useful when

Table 13.2 Sample GIS Data for SWMM's RUNOFF Block

Sewershed	Area (Acres)	Mean PI	Mean % Slope	Land Use (Acres)					
				S.F.R.	M.F.R.	Comm.	Ind.	O.S.	S.W.
2	397.82	25.46	7.78	168.79	—	—	45.90	183.12	—
4	20.18	44.30	5.81	14.81	—	0.47	4.65	0.24	—
5	4.13	35.00	4.34	—	—	4.13	—	—	—
6	31.58	43.90	6.87	25.24	1.79	4.54	—	—	—
7	116.82	16.60	11.70	41.42	1.90	—	—	73.51	—
8	29.66	76.50	4.52	—	10.14	18.89	—	0.63	—
9	38.95	38.73	4.50	35.43	1.54	1.98	—	—	—
10	13.93	40.50	6.39	11.55	2.06	0.31	—	—	—
11	16.69	31.30	6.86	13.46	0.24	0.25	—	2.74	—
12	29.47	27.60	9.13	22.18	—	—	—	7.29	—
13	13.86	21.60	9.76	7.68	—	—	—	6.18	—
14	18.49	26.20	7.56	13.07	—	—	—	5.42	—
15	64.49	17.40	13.63	26.64	—	—	—	37.85	—
16	84.24	23.19	8.49	46.04	2.52	—	—	35.68	—
17	29.13	50.40	9.15	5.10	—	—	17.47	6.56	—
18	84.93	21.50	7.10	46.67	—	—	—	38.26	—
19	31.76	33.20	7.12	22.84	0.01	—	3.13	5.77	—
20	159.91	38.41	6.47	118.24	10.71	4.57	11.75	14.63	—
21	58.43	52.10	5.87	30.98	12.92	5.79	8.75	—	—
23	78.67	39.20	6.72	36.66	—	—	23.78	18.23	—
24	77.83	33.00	6.79	72.52	—	0.01	—	5.31	—
25	297.06	22.60	5.52	166.87	3.77	—	—	126.41	0.02
26	5.05	35.00	18.55	5.05	—	—	—	—	—
27	45.16	33.30	6.94	28.96	1.56	—	4.70	9.95	—
28	21.82	45.70	5.70	13.84	—	—	7.05	0.93	—
29	28.79	49.50	5.11	16.90	3.10	0.02	8.75	—	0.01
30	9.71	61.10	3.47	3.07	3.92	2.68	—	0.04	—
32	33.46	46.90	4.64	24.47	2.47	6.51	—	—	—
35	24.97	41.90	5.24	18.84	—	4.41	—	1.70	0.04
37	8.82	40.20	6.70	7.60	—	—	1.23	—	—
38	30.77	54.20	4.93	14.83	7.78	4.35	3.80	—	—
S21	13.43	25.33	8.88	9.10	—	—	—	4.33	—
Total/Mean	1920.01	37.24	7.26	1068.85	66.43	58.91	140.96	584.78	0.07

Notes: P.I. = Percent Imperviousness; S.F.R. = Single-Family Residential; M.F.R. = Multifamily Residential; Comm. = Commercial; Ind. = Industrial; O.S. = Open Space; and S.W. = Surface Water.

Table 13.3 Sample GIS Data for SWMM's TRANSPORT Block

Sewershed	Population	Population Density	Houses	Family Size	Family Income	Market Value
2	944	2.37	406	2.3	27,709	58,491
4	165	8.18	73	2.3	19,777	36,823
5	15	3.63	13	1.2	19,777	34,180
6	359	11.37	177	2.0	20,461	39,054
7	363	3.11	150	2.4	27,708	50,258
8	223	7.52	156	1.4	17,227	42,920
9	504	12.94	242	2.1	22,184	46,114
10	195	14.00	102	1.9	16,375	38,077
11	191	11.44	88	2.2	16,639	37,734
12	241	8.18	106	2.3	19,540	35,694
13	38	2.74	17	2.2	21,439	33,337
14	98	5.30	43	2.3	21,439	38,712
15	282	4.37	132	2.1	23,260	52,256
16	330	3.92	148	2.2	21,439	48,634
17	47	1.61	20	2.3	21,439	56,297
18	717	8.44	321	2.2	19,192	36,142
19	501	15.77	234	2.1	16,610	33,060
20	2,282	14.27	1,127	2.0	19,769	36,348
21	774	13.25	437	1.8	19,237	31,908
23	659	8.38	282	2.3	17,454	33,436
24	761	9.78	374	2.0	22,598	29,289
25	1,301	4.38	576	2.3	22,642	46,708
26	23	4.55	11	2.0	22,646	26,955
27	403	8.92	178	2.3	16,480	31,922
28	260	11.92	119	2.2	16,889	29,544
29	410	14.24	186	2.2	16,214	26,441
30	94	9.68	56	1.7	14,814	28,387
32	247	7.38	110	2.2	12,783	33,810
35	102	4.08	48	2.1	13,074	38,263
37	26	2.95	10	2.6	21,439	56,620
38	421	13.68	220	1.9	18,369	39,140
Total/Mean	12,976	8.14	6,162	2.10	19,568	38,921

analyzing large regional watersheds. NATSGO data describe variations in soil type at the multistate to regional scale, which is not suitable for wastewater and stormwater modeling applications (Moglen, 2000).

The steps for estimating Green-Ampt parameters using STATSGO GIS data are given in the following text:

1. Download the STATSGO database files from the NRCS Web site (www.ftw.nrcs.usda.gov/stat_data.html).
   - comp.dbf — Attribute HYDGRP has Hydrologic Soil Groups (A, B, C, etc.)
   - layer.dbf — Attribute TEXTURE2 has soil texture (S [sand], SIL [silt], etc.)
2. Join the soils layer to the preceding tables using the MUID attribute that provides a common link between the soils layer and the STATSGO database files.

Table 13.4 Sample EXTRAN Input Data

KS Look-Up Table	
HYDGRP	**KS**
A	0.375
B	0.225
C	0.100
D	0.025

MD Look-Up Table		
Soil	**TEXTURE2**	**KS**
Sand	S	0.31
Sandy loam	SL	0.28
Silt loam	IL	0.26
Loam	L	0.23
Sandy clay loam	SCL	0.18
Clay loam	CL	0.16
Clay	C	0.13

SU Look-Up Table		
Soil	**TEXTURE2**	**SU**
Sand	S	4
Sandy loam	SL	8
Silt loam	SIL	12
Loam	L	8
Clay loam	CL	10
Clay	C	7

3. Perform an overlay of soils with subareas layer to estimate:
   - Percentage of HYDGRP in each subarea
   - Percentage of TEXTURE2 in each subarea
4. Calculate the Green-Ampt parameters (saturated hydraulic conductivity KS, initial moisture deficit IMD, and average capillary suction SU) for each subarea using the look-up tables shown in Table 13.4. These look-up tables are based on the data provided in SWMM's users manual (Huber and Dickinson, 1988).

GIS APPLICATIONS FOR SWMM

Representative GIS applications in SWMM modeling are given in the following subsections.

AVSWMM

Shamsi (1997) developed an ArcView GIS interface called AVSWMM for collection-system and wet-weather overflow modeling. Both RUNOFF and EXTRAN Blocks were included. The interface was developed by customizing ArcView 3.2, using Avenue. Avenue is ArcView's native scripting language and is built into ArcView 3.x. Avenue's full integration with ArcView benefits the user in two ways: (1) by eliminating the need to learn a new interface and (2) by letting the user work

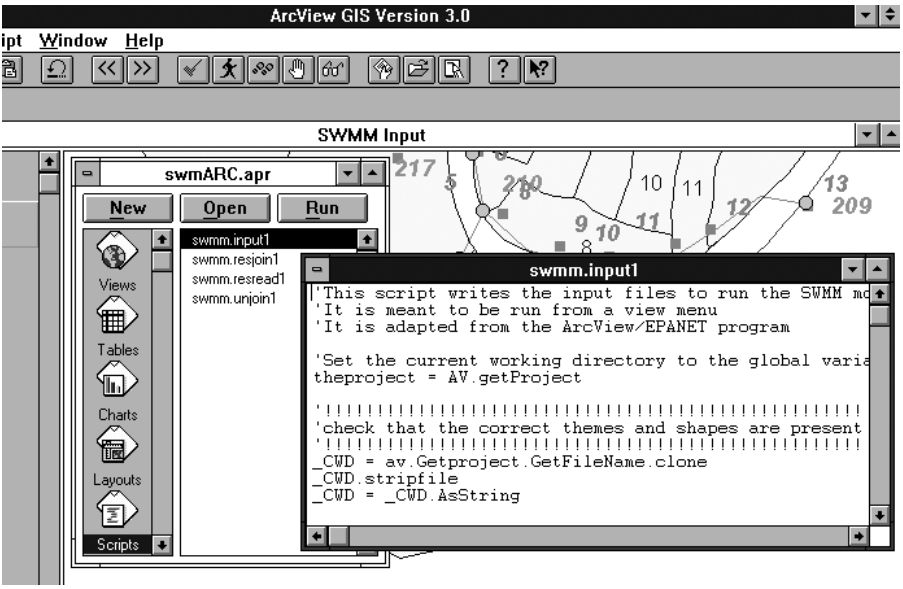


Figure 13.4 ArcView 3.x integrated programming interface for writing avenue scripts.

with Avenue without exiting ArcView (ESRI Educational Services, 1995). Figure 13.4 shows ArcView's integrated programming interface, using Avenue.

The AVSWMM interface was developed as two separate extensions, one each for the RUNOFF and the EXTRAN Blocks of SWMM.

AVSWMM RUNOFF Extension

When the extension is loaded, a menu named SWMM is added to the ArcView interface as shown in Figure 13.5. The SWMM menu has seven calculators and two functions that calculate various RUNOFF Block parameters for subareas (sewersheds). A polygon layer for subareas is required to use this extension. ArcView's Spatial Analyst Extension is required for the analysis of raster data layers. Each calculator prompts the user for the layer containing the subareas and the layer containing the data from which the model parameter is to be extracted. Each calculator adds the calculated input parameters to fields in the subarea layer. The fields are created if they do not exist. Various functions and calculators of this extension are described in the following text:

1. Average Slope Calculator

Function	Assigns average slope to subareas
Input Layers	Subarea polygon layer Slope grid
Output	Adds or updates MEANSLOPE field in subarea layer

2. Census Parameter Calculator

Function	Assigns total population and housing counts to subareas

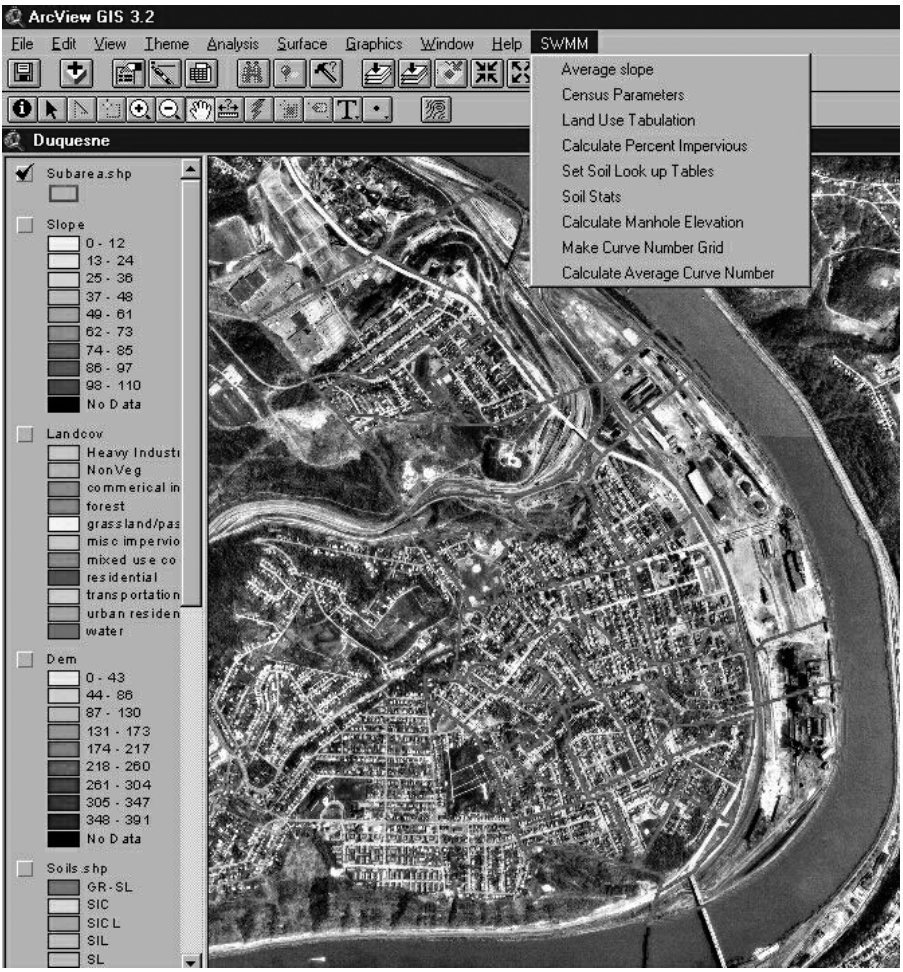


Figure 13.5 AVSWMM RUNOFF Extension.

Input Layers	Census block polygon layer	
	Subarea polygon layer	
Output	Adds or updates total population and housing fields to subareas	
	Pop100	Total Population
	Hu100	Total Households
	PopDens	Population density per acre
	HouseDens	Households density per acre
	MnFamSize	Mean family size

3. Land Use Calculator

Function	Calculates area of each land use class for subareas
Input Layers	Landuse grid layer
	Subarea polygon layer
Output	Adds fields for each land use type to subarea layer
	Calculates area for each land use type

4. Percent Imperviousness Calculator

Function	Assigns percentage of percent imperviousness for each subarea. The calculator will add a percent impervious field to the landuse polygon layer and will prompt for the percent impervious value for each land use type.
Input Layers	Landuse grid layer Subarea polygon layer
Output	Adds or updates field containing percent impervious, PERIMP, to subarea layer

5. Set Soil Look-up Tables Function

Function	Joins the look-up tables for soil infiltration parameters for the Green-Ampt method (Huber and Dickinson, 1988) IMD Initial moisture deficit KS Saturated hydraulic conductivity or permeability SU Capillary suction Additional information is presented below
Input Layers	Soils polygon layer IMD.dbf look-up table KS look-up table SU look-up table
Output	Joins the look-up tables to the soils layer

6. Soil Parameter Calculator

Function	Calculates Green-Ampt infiltration parameters, IMD, KS, and SU for subareas. The user may also select any numerical fields of the Soils layer to average.
InputLayers	Soil polygon layer Subarea polygon layer
Output	Adds or updates average soil parameters to subarea layer MNIMD Mean IMD value MNKS Mean KS value MNSU Mean SU value

7. Manhole Elevation Calculator

Function	Assigns elevation to manholes based from a DEM
Input Layers	DEM grid Manhole point layer
Output	Adds or updates ELEV field containing elevation to Manhole layer

8. Make Curve Number Grid Function

Function	Prepares a grid layer for runoff curve numbers
Input Layers	Land use layer Soil polygon layer Subarea polygon layer
Output	It overlays the land use and soil layers to make a curve number grid. The values of the grid are based on each unique combination of land use and hydrologic soil group.

9. Curve Number Calculator

Function	Calculates subarea runoff curve numbers. Although EPA's SWMM release does not use runoff curve numbers for input, many hydrologic models (e.g., HEC-HMS) and some SWMM GUI programs, such as XP-SWMM (described above) can use this parameter.

Input Layers	Curve number grid Subarea polygon layer
Output	Calculates the average curve number for each subarea. The values are added to the CURVNUM field of the Subarea layer.

AVSWMM EXTRAN Extension

Development of the AVSWMM EXTRAN Extension required the two main tasks described below:

Task 1: Create EXTRAN input file

This task requires that ArcView-compatible GIS layers (themes) exist with attribute fields for all model input parameters. The interface fetches the model input parameters from the GIS layers and exports (copies) them to an ASCII text input file specified by the user. The user should create a header file containing the model control parameters. The control parameters are entered on the A1 line (title) and on B0 to B9 lines (solution and printing instructions). The interface automatically appends this header file to the input file. A simplified example is given in the following text:

Given Data:

1. ArcView GIS layer for manholes
   Manholes layer attributes:
   - ID
   - Bottom elevation
   - Top elevation
   - Inflow rate
2. ArcView GIS layer for sewers
   Sewers layer attributes:
   - ID
   - Upstream manhole
   - Downstream manhole
   - Type (circular, rectangular, etc.)
   - Depth (diameter for circular pipes)
   - Width (not required for circular pipes)
   - Length
   - Material
   - Date installed (material and date installed can be used to automatically calculate Manning's roughness coefficient)
   - Upstream invert elevation
   - Downstream invert elevation

Data To Be Created: An ASCII text input file containing the following parameters:

- Junction (usually manhole) data:
  - Data type (D1 for junctions)
  - Manhole number
  - Top elevation (GR-EL)

  - Bottom elevation (INV-EL)
  - Inflow rate
  - Initial depth (Y0, usually 0.0)
- Conduit (usually sewer pipe) data:
  - Data type (C1 for conduits)
  - Pipe number
  - Upstream junction (FR)
  - Downstream junction (TO)
  - Initial flow (Q0, usually 0.0)
  - Type (1 for circular, 2 for rectangular, etc.)
  - Area (0.0 for circular and rectangular)
  - Depth (DEEP = diameter for Type 1, depth for Type 2)
  - Width (WIDE = width for Type 2)
  - Length (LEN)
  - ZP1 (upstream invert elevation minus upstream node bottom elevation)
  - ZP2 (downstream invert elevation minus downstream node bottom elevation)
  - N (Manning's roughness coefficient, a function of pipe material and age)
  - STHETA and SPHI (usually zero, required for open channels only)

Note that ZP1, ZP2, and N are not part of the given sewer's data, but can be estimated in ArcView from the given data. For example, if the material is concrete and the age of the pipe is less than 10 years (fair condition), then ArcView can use a look-up table to estimate N as 0.015. ZP1 and ZP2 can be similarly estimated from the upstream and downstream manhole data. Sample SWMM input lines for junctions and conduits are shown in Table 13.5.

Table 13.5 Sample EXTRAN Input Data

Sample EXTRAN Input Data Lines for Junctions

*	**JUNCTION DATA**				
*	ID	GR-EL	INV-EL	INFLOW	Y0
D1	10	554.7	522.0	0.0	0.0
D1	20	551.1	522.3	0.0	0.0
D1	30	554.7	522.7	0.0	0.0
D1	40	553.5	523.0	0.0	0.0
D1	50	550.6	524.6	0.0	0.0

Sample EXTRAN Input Data Lines for Conduits

*	**CONDUIT DATA**													
*	ID	FR	TO	Q(0)	TYPE	AREA	DEEP	WIDE	LEN	ZP1	ZP2	N	STHETA	SPHI
C1	330	50	40	0.0	1	0.0	6.0	0.0	820	0.0	0.0	0.015	0.0	0.0
C1	331	40	30	0.0	1	0.0	6.0	0.0	173	0.0	0.0	0.015	0.0	0.0
C1	332	30	20	0.0	1	0.0	6.0	0.0	200	0.0	0.0	0.015	0.0	0.0
C1	333	20	10	0.0	1	0.0	6.0	0.0	140	0.0	0.0	0.015	0.0	0.0

Table 13.6 Sample EXTRAN Output Data

Sample EXTRAN Output Data for Junctions

Junction Summary Statistics

Junction Number	Ground Elevation (FT)	Uppermost Pipe Crown Elevation (FT)	Mean Junction Elevation (FT)	Junction Average % Change	Maximum Junction Elev. (FT)	Time of Occurence		Feet of Surcharge At Max Elevation	Feet Max. Depth is Below Ground Elevation	Length of Surcharge (Min)	Length of Flooding (Min)	Maximum Junction Area (FT^2)
						Hr.	Min.					
10	554.70	528.00	527.74	3.1377	554.59	8	30	26.59	0.11	69.5	0.0	8.7171
20	551.10	528.30	528.14	2.2338	551.10	8	30	22.80	0.00	60.8	0.5	9.8781
30	554.70	528.70	528.69	1.9667	551.55	8	30	22.85	3.15	60.8	0.0	1.0841
40	553.50	529.00	529.03	1.7722	551.40	8	30	22.40	2.10	60.7	0.0	2.8831
50	550.60	530.60	530.72	1.2461	548.47	8	30	17.87	2.13	60.3	0.0	5.3431

Sample EXTRAN Output Data for Conduits

Conduit Summary Statistics

Conduit Number	Design Flow (CFS)	Design Velocity (FPS)	Conduit Vertical Depth (IN)	Maximum Computed Flow (CFS)	Time of Occurence		Maximum Computed Velocity (FPS)	Time of Occurence		Ratio of Max. to Design Flow	Maximum Inv. at Upstream (FT)	Depth Above Conduit Ends Downstream (FT)	Length of Norm Flow (Min)	Conduit Slope (FT/FT)
					Hr.	Min.		Hr.	Min.					
330	1.62E+02	5.73	72.00	2.10E+02	9	4	7.43	9	4	1.30	23.87	28.40	46.3	0.00195
331	1.53E+02	5.41	72.00	2.11E+02	9	2	7.45	9	2	1.38	28.40	28.85	22.2	0.00173
332	1.64E+02	5.81	72.00	3.21E+02	8	35	11.35	8	35	1.96	28.85	28.80	4.8	0.00200
333	1.76E+02	6.22	72.00	3.52E+02	8	35	12.45	8	35	2.00	28.80	32.46	10.3	0.00214

Task 2: Create SWMM EXTRAN output layers in ArcView GIS

This task creates model output layers in ArcView by reading EXTRAN's ASCII text output file. A database table of the relevant output results is created from the output file. The output database file is then temporarily joined to ArcView layers (usually the pipes layer). The linked output results can then be queried or used to classify a legend. A simplified example is given in the following text:
Given Data: EXTRAN's ASCII text output file. Sample SWMM output lines for junctions and conduits are shown in Table 13.6.
Data to Be Created

- An ArcView layer for manhole SWMM output
- An ArcView layer for pipe SWMM output

As shown in Figure 13.6, the AVSWMM EXTRAN interface adds a SWMM menu to ArcView's main menu. This menu has four functions: Make Input File, Read Output, Join Results, and Unjoin Results. The interface allows a user to create an EXTRAN input file, translate an output result file, join the output result file to the themes, and remove the results joined to the themes. All model runs that have been translated are eligible to be viewed. Model run data are stored in separate folders. AVSWMM EXTRAN functions are described in the following text.
Make Input File: This menu choice creates an EXTRAN input file. As shown in Figure 13.7, the user is prompted for a model input name. The input file is placed in a folder called *name*, which is under the project directory. If the folder does not exist, it is created. The function prompts for a description of the model run. It is a string that is inserted in the input file as a comment line. The function then prompts for an optional header file to attach to the input file. The header files have an extension of .hdr and are created separately by the user in a text editor. The header file contains EXTRAN input file data other than C1 and D1 records, which are already stored in the "name.inp" file.
Read Output: This menu choice translates an EXTRAN output file into GIS database files. The function prompts for an output file to translate. The program assumes the user has run the SWMM model from the input file created previously. The output file should be located in the same folder as the input file. The program translates the Conduit Summary and Junction Summary Statistics tables of the EXTRAN output into GIS databases named condsum.dbf and junctsum.dbf and stores them in the "name" folder. The function also writes a file "name.res" in the project directory. The .res file contains the absolute path to the directory where the model results are stored. This file indicates that a model has been translated. It is used to find all model runs available for linking.
Result Join: This menu choice temporarily joins the translated EXTRAN output to the ArcView attribute tables of manholes and sewers. The tool prompts the user to select a model run to link. The routine temporarily links the result databases to the attribute data. The result data can now be queried, listed, or used to classify a legend. Figure 13.8 shows ArcView database tables showing imported EXTRAN results.

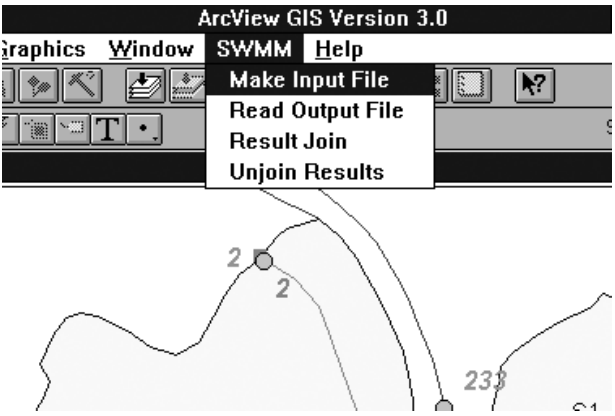


Figure 13.6 AVSWMM EXTRAN Interface.

Figure 13.9 shows a thematic map of maximum sewer flows based on classifying the Q(MAX) output variable from the EXTRAN output.

Unjoin Results: This menu choice reverses (unjoins) the link between the SWMM output and the ArcView attribute tables established by the "Results Join" menu choice. The purpose of this menu choice is to *undo* the last import of SWMM output to ArcView if the user does not like the SWMM results and wants to rerun SWMM and import the output again.

SWMMTools

SWMMTools is an ArcView extension developed by Heineman (2000). Inspired by Shamsi's work described in the preceding subsections, it allows users of SWMM 4.4 to visualize a SWMM model in conjunction with existing GIS data. SWMMTools links SWMM input and output files with ArcView GIS. It allows viewing of model input and output summary data within ArcView. The extension adds a

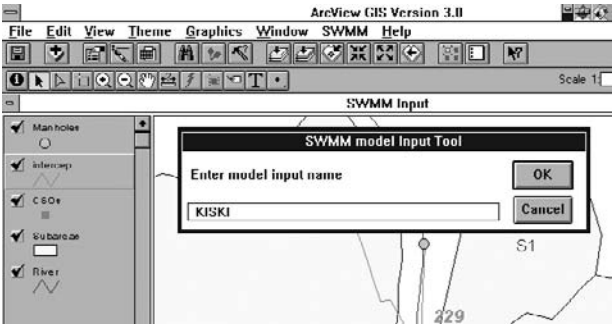


Figure 13.7 AVSWMM EXTRAN: Make Input File feature.

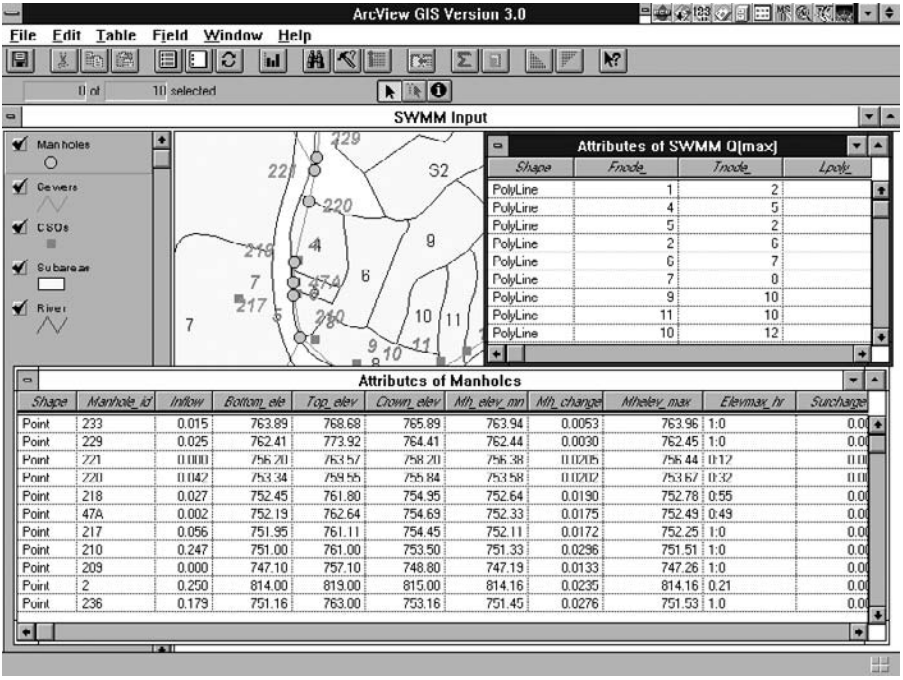


Figure 13.8 EXTRAN results in ArcView.

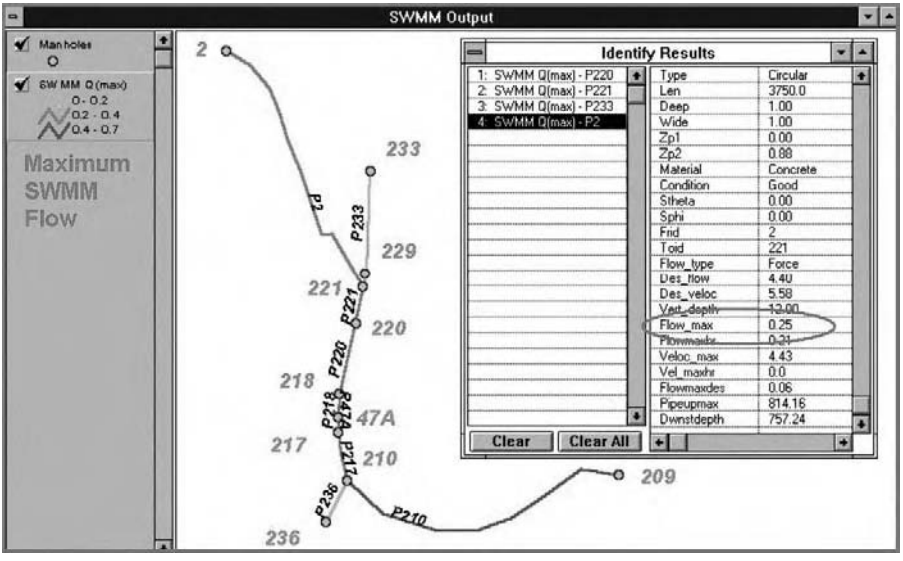


Figure 13.9 Thematic map of maximum sewer flows.

single pull-down menu to ArcView. Three scripts load plan views of RUNOFF, TRANSPORT, or EXTRAN conduit networks. Three scripts link SWMM output with existing RUNOFF, TRANSPORT, and EXTRAN GIS themes respectively. Two scripts work with a stormwater catchment theme to facilitate catchment parameterization. One estimates RUNOFF catchment widths. Another estimates RUNOFF impervious area fractions (parameter DCIA) from GIS land-use data.

AGSWMM

Existing SWMM4-ArcView3 interfaces such as AVSWMM and SWMMTools described in the preceding subsections will continue to work with SWMM4 and ArcView3 Unfortunately, these interfaces cannot work with SWMM5 or ArcView/ArcGIS8 and higher versions because of the following reasons:

1. SWMM4 and SWMM5 input file formats are different
2. SWMM4 and SWMM5 output file formats are different
3. SWMM4 and SWMM5 have different data structures (e.g., three different representations of the same pipe are not required in Runoff, Transport, and Extran Blocks)
4. ArcView3 interfaces are written in Avenue, which is not compatible with ArcView8

To benefit from new features of the new software versions, new interfaces should be developed. Shamsi and Smith (2004) developed a SWMM5-ArcView8 interface called ArcGIS SWMM (AGSWMM). The main features of AGSWMM are listed below:

- AGSWMM links SWMM5 with ArcGIS
- AGSWMM was developed using VBA
- AGSWMM setup installs a DLL file
- AGSWMM adds a new "ArcSWMM" toolbar in ArcGIS
- AGSWMM allows creation of a SWMM5 input file from GIS layers
- AGSWMM allows creation of a SWMM5 backdrop map (.emf file) in ArcGIS
- AGSWMM allows on-screen editing of input parameters in ArcGIS
- AGSWMM allows thematic mapping of SWMM5 output in ArcGIS

Figure 13.10 shows the AGSWMM toolbar. The toolbar consists of two tools: SWMM Input and SWMM Results View. The SWMM Input tool opens the input form shown in Figure 13.10. It allows users to create the SWMM5 input file. The left side of the form is used for creating input data for nodes (e.g., manholes). The right side of the form is used for creating input data for links (or conduits, e.g., pipes). Users can assign the appropriate GIS layer attributes to SWMM5 input parameters using dropdown pick lists. Users can either create a new SWMM5 input file using the "Make SWMM Input File" button at the bottom of the input form, or they can update the input parameters in an existing SWMM5 input file by selecting the "Synchronize SWMM File" button. The synchronization

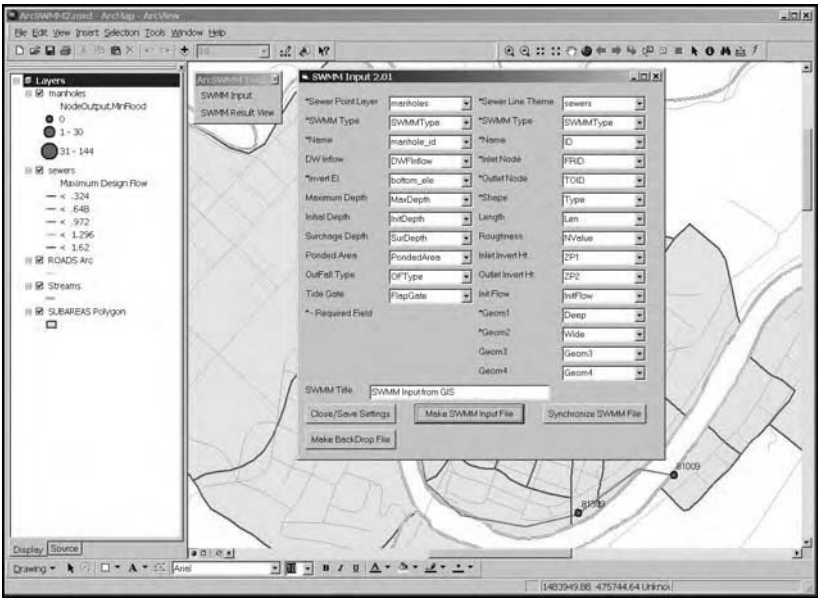


Figure 13.10 AGSWMM toolbar and input form.

capability allows users to edit the input file attributes in ArcGIS (e.g., change a sewer pipe diameter), and the changes are automatically reflected in the SWMM5 input file. This feature keeps the SWMM5 input files current. The "Make Backdrop File" button creates a backdrop map file for SWMM5 in the required .emf file format. This button automatically updates the SWMM5 input file with the appropriate scaling factor and offset distance so that it will properly align with the nodes and links in the SWMM5 map window. Figure 13.11 shows a SWMM5 map screenshot with a backdrop file creating by AGSWMM.

PCSWMM GIS™

PCSWMM GIS (Computational Hydraulics International, Guelph, Canada) is a good example of the model-based integration method described in Chapter 11 (Modeling Applications). PCSWMM GIS is a preprocessor for EPA's SWMM, and it also facilitates output visualization through a variety of plug-in tools. Model input parameters (node, conduit, and subcatchment) are extracted from an ODBC-compliant database such as Microsoft Access using SQL queries. Model input data can also be imported from an underlying dBASE-compatible GIS database. Extracted data are saved in an intermediate database (MS Access) for preprocessing into a useful model. Processed data are exported to a SWMM input file (Runoff, Transport, or Extran). The cost of PCSWMM GIS 2000 is

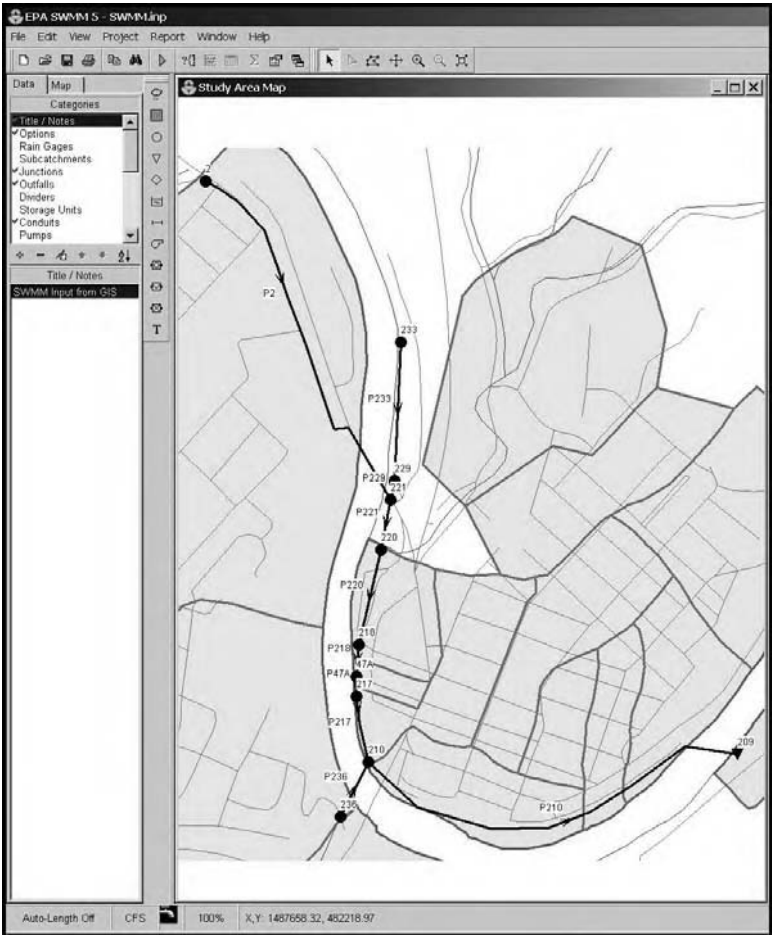


Figure 13.11 SWMM5 map window with a backdrop file created in ArcGIS.

$400. Figure 13.12 shows the PCSWMM GIS interface that has been developed using Visual Basic and resembles the ArcView GIS interface.

SWMM and BASINS

EPA has shown an interest in incorporating SWMM into its BASINS package, which is already integrated with the ArcView GIS software. EPA is now developing a SWMM linkage to its BASINS package. This link would be a valuable tool for urban planners and watershed managers to evaluate the urban pollution load in relation to other loads in the watershed and provide a more holistic view of watershed management. This project is being done through an EPA Office of Administration and Resources Management (OARM) contract with Lockheed

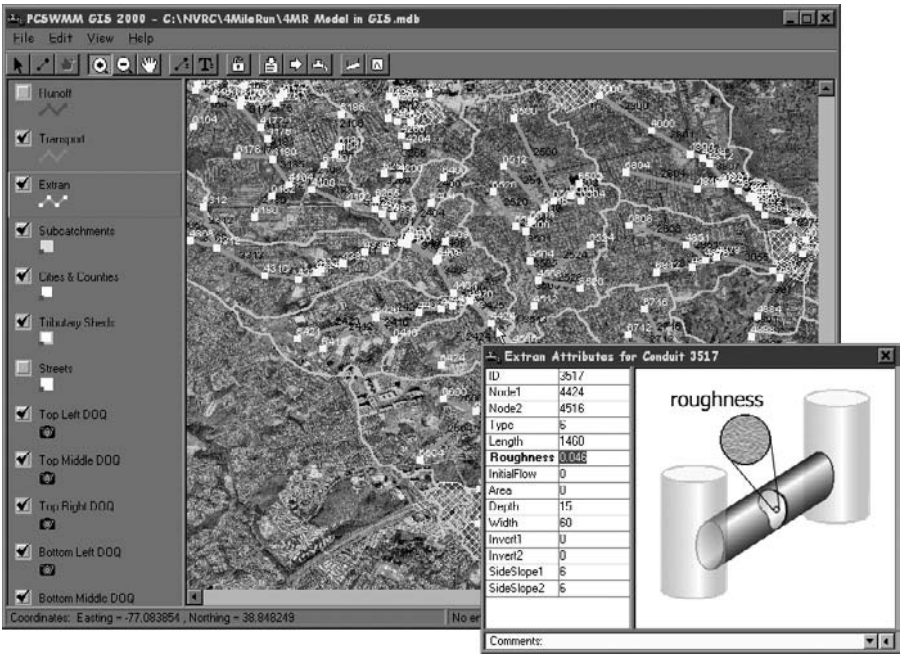


Figure 13.12 PCSWMM GIS: an example of a model-based integration method.

Martin Information Services (LMIS). When completed, this project will bundle SWMM and BASINS together with HSPF, TOXIROUTE, and QUAL2E programs that are currently part of the BASINS modeling environment. It should also allow extraction of SWMM input data from the GIS files that are part of the BASINS CD-ROMs. Detailed BASINS information is presented in Chapter 11 (Modeling Applications).

SWMMDUET

The Delaware Department of Natural Resources and Environmental Control, Division of Soil and Water, developed this program in 1994. SWMMDUET is an example of SWMM and GIS integration. This program integrates SWMM 4.2 (1993) with ArcInfo (ARC/INFO) 6.0. For this reason, it does not function with the latest versions of SWMM and ArcInfo (Schell et al., 1996). It has been written in ArcInfo's native Arc Macro Language (AML). It creates a computing environment that does not require arcane knowledge of SWMM and ArcInfo. SWMMDUET incorporates a preprocessor and a postprocessor, and therefore, provides interfaces both for creating SWMM's input file and for graphing SWMM's output file. SWMMDUET takes advantage of the graphical paradigm around which ArcInfo was constructed and utilizes relational database capabilities to organize data. Beyond data management, the program uses expert-system

logic to assemble data that define the modeling process, prepares SWMM input, executes the SWMM program, and converts the output into meaningful graphical displays. SWMMDUET, therefore, uses the integration method of GIS applications described in Chapter 11 (Modeling Applications). Data-entry sheets and forms eliminate the need for the modeler to know detailed ArcInfo processing techniques. Similarly, feature selection, spatial joins, and processing commands and options of ArcInfo are specified and executed for the user. Hyetographs are related to rain gauges where rainfall data were recorded. The storms are selected simply by a georeference to the gauges. Time-series data are stored as sequential records in the database files. SWMMDUET simplifies the management of vast amounts of hydrologic data, and lets the hydrologist concentrate on hydrologic matters rather than data input and output interpretation. The SWMMDUET software is in the public domain and can be obtained for the cost of distribution (Curtis, 1994).

AVsand™

AVsand from Cedra Corporation (Rochester, New York) is an example of ArcView and SWMM integration. This program offers only limited SWMM modeling capability because not all of the SWMM Blocks and capabilities are available. AVsand can be used for modeling sanitary sewers using a simplified implementation of SWMM's TRANSPORT Block. Storm and combined sewers are not modeled using SWMM; the Rational Method or surface hydrographs are used instead. AVsand accepts user model input via a series of dialogue boxes. AVsand also provides model development and editing tools, such as a Service Area Tool to define service areas, a Node Definition Tool to define nodes, and a Pipe Definition Tool to define pipes between nodes. A Wastewater Loads Tool enables the user to introduce wastewater loads at the selected nodes. A "Query" menu option helps the user to create GIS maps from the model output. It can also create a profile plot of the HGL, which is very effective to help pinpoint the areas of hydraulic overloads.

OTHER SEWER MODELS

The preceding section focused on SWMM models. Other sewer system models with GIS application capabilities are described in the following subsections.

DHI Models

DHI Water and Environment (formerly Danish Hydraulic Institute, Hørsholm, Denmark) and DHI, Inc. (Newton, Pennsylvania), specialize in developing GIS-based H&H computer models. For example, DHI has developed the following sewer system models for the ArcView GIS software package.

MOUSE™

Modeling of Urban SEwers (MOUSE) is a package for the simulation of surface runoff, flow, water quality, and sediment transport in urban catchments and sewer

systems. MOUSE combines complex hydrology, hydraulics, water quality modeling, and sediment transport modeling capabilities in a graphical and easy-to-use interface. MOUSE is a 32-bit Windows application specifically designed to operate within Microsoft Windows and Windows NT operating systems, and is optimized for fast simulations and graphics. Both metric (SI) and imperial units are supported. The MOUSE system is organized in several modules:

- MOUSE Runoff: Surface-runoff models for urban catchment applications
- MOUSE HD: Hydrodynamic network model with some limited RTC capabilities
- MOUSE RDII: Advanced hydrological model for continuous simulation
- MOUSE RTC: Advanced reactive RTC capabilities for MOUSE pipe models
- MOUSE LTS: Long-term hydraulic simulations with statistics
- MOUSE SRQ: Pollutant buildup and transport on catchments surfaces
- MOUSE AD: Pollutant advection-dispersion in drainage networks
- MOUSE WQ: Water quality processes in drainage networks
- MOUSE ST: Sediment transport in drainage networks

Some of these modules are marketed as optional add-on modules to the standard configuration. The various modules together with the user interface part, including the time-series and cross-section editors, advanced graphics both on the input and output sides, and the online HELP system, make the MOUSE system a desirable sewer system modeling package. MOUSE is a sophisticated hydraulic model that is commonly compared to SWMM (Heaney, 1999).

MIKE SWMM™

MIKE SWMM is DHI's implementation of SWMM. Resulting from a collaborative effort between DHI and Camp, Dresser, and McKee (CDM), MIKE SWMM combines the power of SWMM with a user-friendly interface. Data can be digitized and viewed with background images and edited with the same graphical editors that are also part of the MOUSE system. Data can also be entered and edited through database forms, which include scrollable spreadsheet-like sections for efficient editing of tabulated data. These different editors or views of the same information are dynamically linked, so that changes introduced through one editor are automatically and instantly updated in the other views. MIKE SWMM's database tool for storage of network data gives access to a range of query options.

MOUSE GIS™

MOUSE GIS is an application linking the MOUSE numerical sewer modeling system with ArcView. MOUSE GIS is an interface between ArcView and the MOUSE model. The MOUSE GIS interface is used for preprocessing external database information into MOUSE data sets. Once the link to an external database is established, manhole, link, weir, pump, and catchment data can be uploaded. The entire database can be used for the model networks, or a subset can be

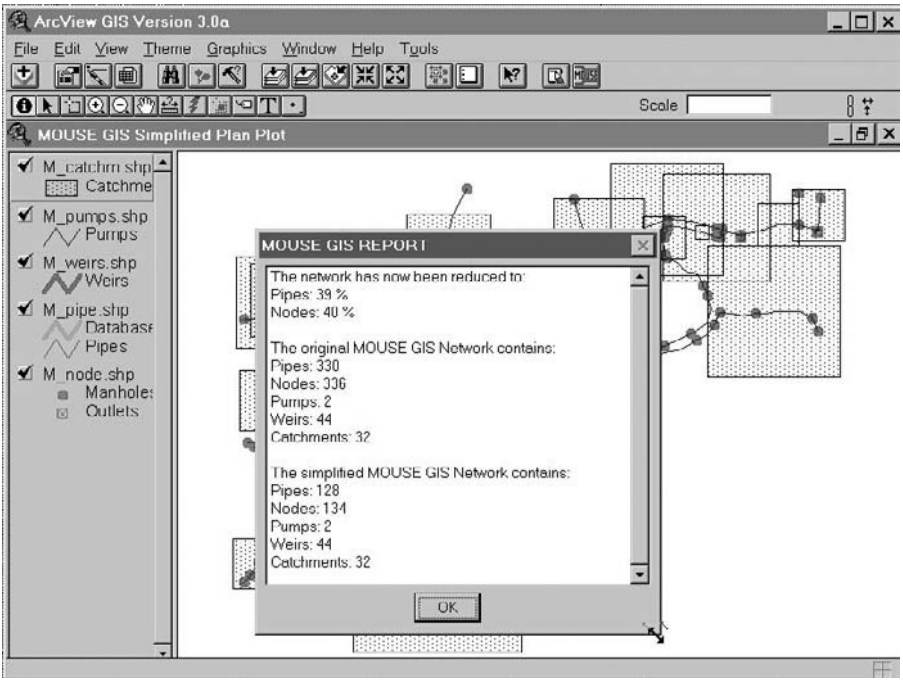


Figure 13.13 MOUSE GIS screenshot showing network skeletonization results.

selected. The selected set can subsequently be simplified (nodes excluded) to reduce the model size. MOUSE GIS also acts as a postprocessor for simulation results. Pipe flow results (maximum flow, velocity, depth and surcharge, flooding, etc.) can be displayed and overlayed with other GIS data. Mouse GIS also provides an interface to MIKE SWMM. Figure 13.13 shows a MOUSE GIS screenshot of the sewer system model for the city of Ljubljana (Slovenia) developed by DHI. The figure also shows the results of network skeletonization performed by MOUSE GIS. The automatic skeletonization has reduced the network size to 40%.

MOUSE GM™

MOUSE GM (in which MOUSE GIS is embedded), an ArcView-based module for MIKE SWMM and MOUSE users, promotes spatial and visual representation of the model and provides a direct link to GIS databases. MOUSE GM includes model-simplification tools that enable the user to create simplified (skeletonized) models from detailed data sets. This productivity tool may save days or weeks of tedious and error-prone work in the early phases of modeling projects. MOUSE GM reads and writes the SWMM file formats. It can be used as the ArcView GIS link and model-simplification tool for all SWMM modeling packages that apply the

standard SWMM data formats. With MOUSE GM, MIKE SWMM users can link to GIS databases and with any asset management system that supports ODBC connectivity.

InfoWorks™

Wallingford Software Ltd. (Wallingford, U.K.) develops and supplies data management, operation, and simulation software for the water industry worldwide. It offers two main software products for sewer system modeling:

- InfoWorks CS (formerly HydroWorks): Integrates asset and business planning with urban drainage network modeling.
- InfoNet: Provides tools, reports, displays, and data management capabilities. It integrates with GIS, desktop office systems, hydraulic modeling systems, maintenance management systems (MMS), field data systems, SCADA, and corporate databases. Data exchange interfaces have been designed not only for import and export purposes but also to connect directly with GIS (ArcView and MapInfo), hydraulic modeling systems (InfoWorks), SCADA, and logger systems. InfoNet has been designed to ensure that data exchange with Microsoft Windows applications and geographic information systems is a seamless operation.

InfoWorks CS is recognized for its fast, efficient, and stable algorithms for wastewater modeling. It models systems containing complex pipe connections, complex ancillary structures and static or variable control devices, open channels and trunk sewers, full backwater effects, and reverse flow. It can be upgraded to include one or more of the following optional modules:

- Water Quality module: Predicts quality components, volume of spillage and floods, and helps develop cost-effective solutions to pollution and sedimentation problems.
- Real Time Control (RTC) module: Allows central management of flow throughout an entire network and local flow control at individual ancillary structures. RTC helps users maximize wastewater system storage, optimize flow routing, optimize operation of ancillary structures, and design cost-effective systems.

InfoWorks products are designed to link H&H models with GIS and information management systems. InfoWorks CS combines a relational database with spatial analysis. It can be used to transfer data to and from third-party applications. One can also import asset data directly from Microsoft Access or .CSV files into the InfoWorks database. For instance, InfoWorks CS allows importing of subcatchment areas, area breakdown, and population data directly from MapInfo Professional, ArcView, or GeoMedia Professional GIS. Figure 13.14 shows the InfoWorks CS output results imported in MapInfo Professional (top) and ArcView (bottom).

In Figure 13.14, nodes are symbolized according to modeled flood volume in cubic meters, which represents the water overflowing from a manhole. Smaller (green) nodes are OK (no overflow), those in light blue and slightly larger are loosing a small amount

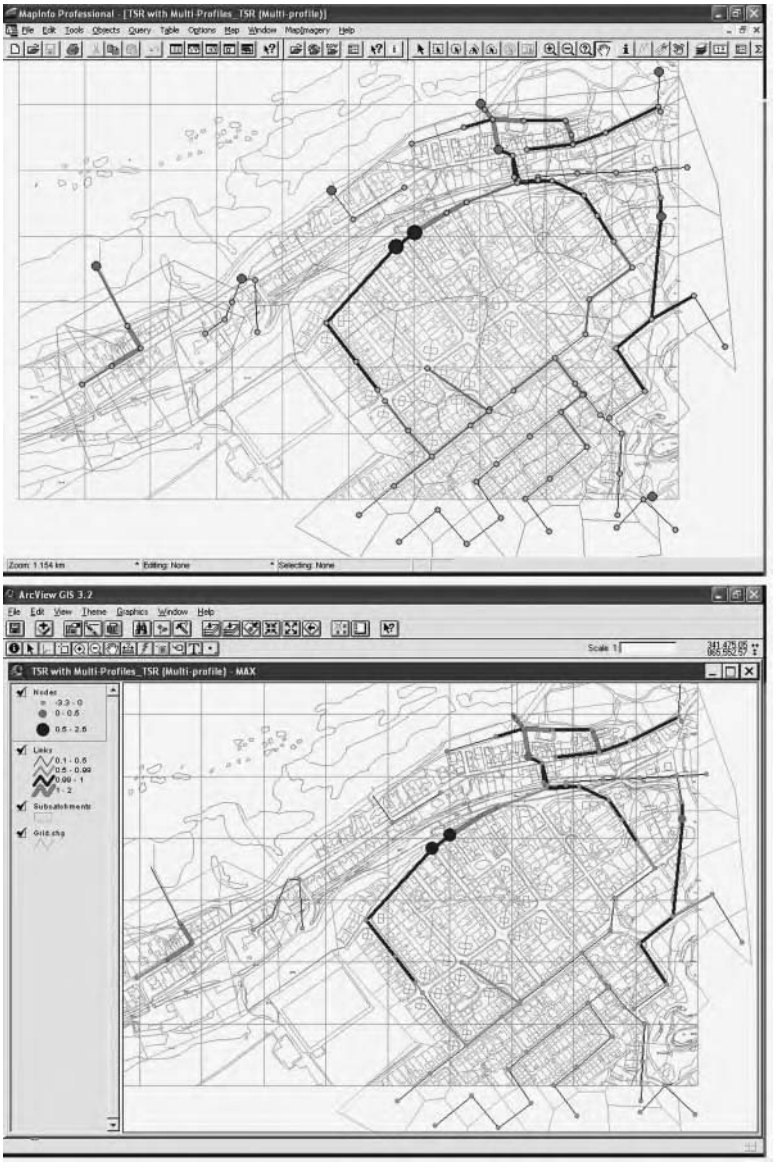


Figure 13.14 InfoWorks CS results in MapInfo Professional (top) and ArcView 3.2 (bottom).

of water, those in dark blue and larger are loosing lots of water. Link theme is showing surcharge depth. This represents 0.1 to 0.5 of pipe height (up to half full), 0.5 to 0.99 of pipe height (between half full and completely full), 1.0 of pipe height, or 100% full, and then 2.0, which is pressurized flow. The associated colors are black, light blue, dark blue and purple.

SewerCAD™ and StormCAD™

Haestad Methods, Inc.* (Waterbury, Connecticut), provides the following sewer system modeling:

- Sanitary sewers
  - SewerCAD Stand-alone
  - SewerCAD for AutoCAD
  - SewerGEMS for ArcGIS
- Storm sewers
  - StormCAD Stand-alone
  - StormCAD for AutoCAD

Stand-alone versions do not require a CAD package to run. The AutoCAD versions run from inside AutoCAD and, therefore, require AutoCAD. An ArcGIS version of SewerCAD called SewerGEMS (Geographic Engineering Modeling System) may be released in 2004.

SewerCAD is a design, analysis, and planning tool that handles both pressurized force mains and gravity hydraulics. SewerCAD performs steady-state analysis using various standard peaking factors, extended-period simulations, and automatic system design. StormCAD provides the same functionality for storm sewer systems.

GIS/database connection modules for these programs are marketed as optional modules that are sold separately. These programs establish a connection between a sewer model and relational and nonrelational database management systems, spreadsheets, and ESRI Shapefiles. For example, the Shapefile Connection Wizard provides import and export capabilities to transfer data between GIS and sewer models.

SEWER MODELING CASE STUDIES

Representative case studies of GIS applications in sewer system modeling are given in the following subsections.

XP-SWMM and ArcInfo Application for CSO Modeling

Application	Hydrodynamic sewer system modeling
Study period	1995–1998
GIS software	ArcInfo
Modeling software	XP-SWMM (sewer system), RMA-2V (river hydraulics and geometry), HEC-RAS (river hydraulics), and WASP5 (river water quality)
Other software	FoxPro and IBM Data Explorer
GIS data	Sewer components, hydrologic soil groups, land use, and curve number
Study area	Fort Wright, Kentucky
Organization	Sanitation District No. 1

* Haestad Methods was acquired by Bentley Systems (Exton, Pennsylvania) in August 2004.

The Sanitation District No. 1 (the District) located in northern Kentucky used GIS, GPS, RDBMS, and computer modeling technologies to simulate virtual flows of sewage through pipes, CSOs, and into and through local rivers (Ohio and Licking Rivers). With the concurrent tasks of updating the GIS and starting up of the facility plan, a common database was developed. The objective of the database was to provide a central location where data could be stored and used for the benefit of both tasks. An XP-SWMM Management System (XMS) was developed within FoxPro, which acts as the translator between different software packages. By including all five data sources (roads and streams, rain-gauge data, XP-SWMM output, CSO loadings, and fecal coliform response of receiving water), a more complete cause–effect relationship was developed. The visualization of the system using IBM Data Explorer provided an overall picture of the system that was not available before. This case study is a good example of GIS being used as an integrator technology or integrating platform (Martin et al., 1998).

AM/FM/GIS and SWMM Integration

Application	Sewer system modeling using the interface method
GIS software	ArcInfo
Modeling software	SWMM
Other software	CASS WORKS (RJN Environmental Associates)
Study area	Boston, Massachusetts, metropolitan area
Organization	Massachusetts Water Resources Authority (MWRA)

MWRA serves the sewerage needs of over two million people in 43 communities in the Boston metropolitan area. It was estimated that nearly half the flow in their system is a direct result of rainfall-induced inflow and infiltration (I/I) due to deteriorated sewers and manholes. In order to evaluate both the short-term and long-term needs of their sanitary sewer interceptor system, MWRA initiated and completed a sewage analysis and management system (SAMS) project, which involved linking of ArcInfo, CASS WORKS, and SWMM using the interface method of model integration. ArcInfo was used to create a link between SWMM data and computerized maps of the MWRA interceptor system and communities served. The maps were used as both a tool to input different planning scenarios into the CASS WORKS system and as a visual medium to geographically show the model results. The integrated AM/FM/GIS system was installed on a multiuser DEC VAX 4000-200. The integrated model allowed the MWRA planners to predict the I/I reduction from elimination of groundwater and rainwater from the interceptor system. Elimination of these flows resulted in significant operational cost savings to the MWRA and its ratepayers (ESRI, 1994).

SWMM and ArcInfo™ Interface

This application was developed as part of a watershed approach to assessing the ecological health of Bayou Chico, which is a subestuary of Pensacola Bay. Bayou Chico is the receiving water body of small, mostly urbanized watersheds located in

southern Escambia County, Florida. A GIS database and the use of remotely sensed satellite images were combined to determine surface characteristics, storm drainage systems, area, and slope of the watershed. These data layers were in turn linked to SWMM, which mathematically represents these physical characteristics and uses this information to determine both runoff and pollutant loading. The SWMM RUNOFF Block was broken into a hydrological component and a pollutant loading component. Both have parameters that may be retrieved from spatially based data. Pollutant loading is partially determined either by land-use class or gutter length. Gutter length can be derived from digitized road layers or are available for some metropolitan areas in the U.S. Census TIGER data files. ArcInfo tools allowed for querying of polygons for areal extent, perimeter, length, width, and other associated attributes. AML programming was used to query ArcInfo layers for SWMM parameter data and to write the information to an ASCII text file in the SWMM input file format (Schell et al., 1996).

Hydra™ and ArcInfo™ Interface

The City of Stockton, California, developed an interface between their ArcInfo GIS and Pizer's HYDRA/HydraGraphics sewer and stormwater modeling software. The GIS contained sanitary sewers and land-use layers, and was utilized as the raw data source for developing a sewer model for master planning purposes. To facilitate this process, a customized GIS data management interface was developed using ArcView GIS as the graphical interface between the City's GIS and the sewer model. This interface included several project-specific Avenue routines to import the City's GIS layers, extract the system model and land-use data, and export the model components to the sewer model (Prins et al., 1998).

USEFUL WEB SITES

Danish Hydraulic Institute	www.dhisoftware.com
Haestad Methods	www.haestad.com
MWH Soft	www.mwhsoft.com
SWMM	www.epa.gov/ceampubl/swater/swmm/index.hm www.epa.gov/ednnrmrl/swmm/
SWMM-Online	www.swmm-online.com
Wallingford Software	www.wallingfordsoftware.com

CHAPTER SUMMARY

This chapter shows that GIS offers many useful applications in sewer system H&H modeling. GIS applications are developed using the three methods of interchange, interface, and integration as described in Chapter 11 (Modeling Applications). The interface method seems to be the most prevalent method at the present time. The case studies and examples presented in the chapter illustrate that many sewer utilities are successfully using GIS applications for modeling their sewer systems. The software

examples presented in the chapter indicate that many public-domain (e.g., SWMM) and commercial-off-the-shelf software packages (e.g., MOUSE) are available to help users benefit from sewer system modeling applications of GIS.

CHAPTER QUESTIONS

1. Prepare a list of GIS applications for sewer systems.
2. What is a GUI? How does it help modelers?
3. Which sewer system modeling tasks can be performed using GIS?
4. What are the typical input parameters for a sewer system hydraulic model? Which of these parameters can be obtained from a GIS?
5. List the GIS data layers that are useful for sewer system modeling.
6. How does GIS help in presenting the output results of a sewer system model?
7. List the sewer system models you are familiar with. Which of these models use GIS? Which method of GIS linkage (interchange, interface, or integration) do they use?
8. Is a GIS-integrated sewer model always better than a stand-alone model? Provide reasons for your answer.

CHAPTER 14

AM/FM/GIS Applications

At present our water and wastewater infrastructure, especially in the older cities, is in critical stages of deterioration and has started to crumble. AM/FM/GIS technology provides a spatial infrastructure management approach that is very effective in prioritizing your infrastructure improvement and maintenance needs.

Field inspection crews inspecting a combined sewer overflow (CSO) diversion chamber located near Pittsburgh, Pennsylvania.

LEARNING OBJECTIVE

The learning objective of this chapter is to find out how GIS is integrated with automated mapping/facilities management (AM/FM) systems for efficient inspection and maintenance of water industry infrastructure.

MAJOR TOPICS

- Infrastructure maintenance issues
- AM/FM/GIS basics
- AM/FM/GIS software
- AM/FM/GIS application examples

LIST OF CHAPTER ACRONYMS

AM/FM/GIS Automated Mapping/Facilities Management/Geographic Information System
CAD Computer-Aided Drafting/Computer-Aided Design
CAFM Computer-Aided Facility Management
CCTV Closed-Circuit Television
CIFM Computer Integrated Facility Management
CMMS Computerized Maintenance Management System
GASB Government Accounting Standards Board
RBE Rule Base Engine
RDBMS Relational Database Management Systems
SDE Spatial Database Engine

HAMPTON'S WASTEWATER MAINTENANCE MANAGEMENT

GIS software	ArcView, GBA GIS Toolkit Extension for ArcView, MapObjects Internet Map Server (IMS)
Other software	GBA Master Series and WinCan TV system
GIS data	Sewers, manholes, water mains, valves, hydrants, roads, street centerlines, buildings, and digital orthophotos
Study area	City of Hampton, Virginia
Project duration	1998–2001
Organization	Public Works Department, Hampton, Virginia

Hampton, Virginia, has a population of 150,000 and has 12,000 manholes and valves. The City conducted a GPS survey to collect coordinates of more than 8000 manholes located in alleys, buried in yards, or covered with asphalt. The City utilizes more than 35 licenses of ArcView GIS. The City stored attribute data in an external infrastructure database from GBA Master Series and linked it to ArcView GIS. The GBA software performs infrastructure inventory of manholes and pipes, work-order management, and parts inventory management. It can also interface with the City's WinCan system used for CCTV inspection of sewers. The GBA GIS Toolkit, an

extension of ArcView GIS, is used to synchronize data consistency between ESRI Shapefiles, CAD, and the infrastructure management database. This application provides a GIS map linked to a comprehensive inventory, inspection, and maintenance management software program. The system also provides a graphical display of maintenance history and profile plots of user-specified sewer segments (ArcNews, 2001).

Although GIS applications in water, wastewater, and stormwater systems are not new, getting beyond the basic inventory and mapping functions is challenging. After a utility GIS database has been set up, GIS applications must be developed or else the GIS would be nothing but a pretty map. For instance, a GIS-based computerized maintenance management system (CMMS) can be implemented for more efficient maintenance of a utility because it can track problems within the utility network more accurately. The GIS and CMMS integration can facilitate proactive (preventive) maintenance, compliance with regulatory mandates, and systemwide planning activities.

According to the U.S. Government Accounting Standards Board (GASB), an infrastructure asset management system is a requirement for reporting capital asset activity. Electronic asset management systems extend the life of existing infrastructure, thereby optimizing maintenance schedules and deferring major capital expenditures until they become necessary and economically justified. Such systems can also identify risky assets and schedule rehabilitation projects that can save the cost of total replacement (Irrinki, 2000).

INFRASTRUCTURE PROBLEM

A civilization's rise and fall is linked to its ability to feed and shelter its people and defend itself. These capabilities depend on infrastructure — the underlying, often hidden foundation of a society's wealth and quality of life. Infrastructure is a Latin word meaning the structure underneath. According to a 1992 report from the Civil Infrastructure Systems Task Group of the U.S. National Science Foundation, "*A society that neglects its infrastructure loses the ability to transport people and food, provide clean air and water, control disease, and conduct commerce.*" (Tiewater, 2001)

The water, wastewater, and stormwater infrastructure is aging fast throughout the world, including in the U.S. For example, according to the 2001 counts, the U.S. has 54,000 drinking water and 16,000 wastewater systems. America's water and wastewater systems incur an expenditure of $23 billion per year on infrastructure projects. Much of the existing sewerage system in the U.S. was constructed in the 1950s and 1960s — some date back to the early 1900s and late 1800s. According to 2001 estimates, U.S. sewer systems are approaching 100 years in age. Even the newest U.S. water infrastructure systems are over 50 years old. What is worse, because of their material, they are more likely to be deteriorating than some systems dating back to 100 years. Many water systems are structurally obsolete and are now serving two to three times as many people as their design capacity. Many systems have not received the essential maintenance and repairs necessary to keep them

working properly. The American Society of Civil Engineers (ASCE) projects $1.3 trillion in infrastructure needs in the U.S. during 1999 to 2004, of which approximately one third is needed for water and wastewater improvements.

Unfortunately, adequate government funds are not available for this expensive fix. This problem can be partially blamed on the "out-of-sight, out-of-mind" mentality of the decision makers. The lack of funding for the badly needed infrastructure improvements is making people worried. This is evident in the recently conducted public opinion polls. According to a recent Water Environment Federation (WEF) survey, 31% of Americans would not swim in their own rivers. A 1998 ASCE poll showed that more than 75% of the U.S. voters were concerned about the quality of roads, drinking water, and school buildings. In Boston, more voters were concerned about roads, bridges, and drinking water than Social Security and taxes. In New York City, concerns about school buildings and drinking water outscored drug abuse worries (ASCE, 1998). Many indicated that they would vote for a candidate whose election campaign addressed waste disposal (78%), drinking water (73%), and roads and bridges (62%). These surveys indicate that water industry infrastructure is particularly important because it is so closely related to public health and safety.

In the U.S., it has been recently observed that a combination of reduced federal spending and increased federal mandates is taking its toll on the infrastructure of the country. For instance, according to American Water Works Association (AWWA), annual pipe replacement costs will jump to $3 billion by 2020 and $6 billion by 2030. Replacement expenses projected into the future indicate that sewer main replacements are more urgent than water mains and will require more money.

As of 2001, ASCE estimates that the U.S. must invest nearly $277 billion in drinking water and wastewater infrastructure repairs over the next two decades. U.S. water systems have an $11-billion annual investment shortfall and need $151 billion by 2018. U.S. sewer systems have a $12-billion annual shortfall in investment and need $126 billion by 2016. The largest need, $45 billion, is for projects to control combined sewer overflows (CSOs). The second-largest category of needs, at $27 billion, is for new or improved secondary treatment.

In a report "Water Infrastructure Now" published in February 2001, a water conservation organization called Water Infrastructure Network (WIN) estimated a $23-billion per year gap between infrastructure needs and current spending. The report indicated that U.S. water and wastewater systems faced infrastructure funding needs of nearly $1 trillion over the next 20 years and a shortfall of half a trillion dollars. The report called for a 5-year, $57-billion federal investment in water infrastructure to replace aging pipes and upgrade treatment systems. WIN estimates that household water bills must double or triple in most U.S. communities if utilities are forced to absorb the entire infrastructure bill. For instance, the City of Gloucester, Massachusetts, is charging $20,000 per home (to be paid over the next 20 years) to finance the city's conversion from septic to sewer systems (CE News, 2001).

According to a 2001 statement by ASCE President Robert W. Bein, "Something is terribly wrong … America has been seriously underinvesting in its infrastructure."

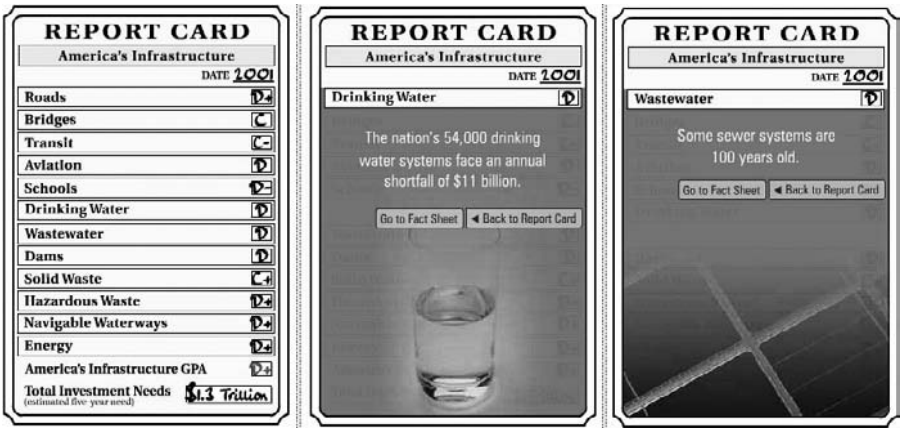


Figure 14.1 ASCE Infrastructure Report Card for the U.S.

In March 2001, ASCE released its 2001 Report Card (Figure 14.1) for America's infrastructure in which the nation's infrastructure received a cumulative grade of D+ for 12 infrastructure areas. Causes for such a dismal grade include: explosive population growth and school enrollment, which outpace the rate and impact of current investment and maintenance efforts; local political opposition and red tape, which outpace the development of effective solutions; and the growing obsolescence of an aging system — evident in the breakdown of California's electrical generation system and the nation's decaying water infrastructure. The 2001 Report Card follows the one released in 1998, at which time the ten infrastructure categories rated were given an average grade of D. The 2001 wastewater grade declined from a D+ to a D, while drinking water remained a D. Solid waste and schools scored the highest and the lowest grades, respectively (ASCE, 2001; EWRI, 2001). Despite the disturbing findings contained in the 2001 Report Card, the nation failed to heed the call to improve its deteriorating infrastructure, according to an update to the assessment released by ASCE in September 2003 (ASCE, 2003).

When the infrastructure condition is bad and funding is scarce, we need infrastructure management tools that can:

- Accurately define (or map) our infrastructure
- Identify the worst portions of the infrastructure
- Determine how to most cost-effectively improve (replace or repair) the worst parts of the infrastructure

AM/FM/GIS products provide an efficient infrastructure management system.

AM/FM/GIS BASICS

Many solutions for managing facilities have spawned from CAD-based architectural applications. Database solutions have evolved from the financial sector to

Table 14.1 Comparison of AM, FM, AM/FM, GIS, and AF/FM/GIS

Feature	AM	FM	AM/FM	GIS	AM/FM/GIS
Layers	Y	N	Y	Y	Y
Topology	N	N	N	Y	Y
Network definition	N	Y/N	Y	Y	Y
Lines	Y	N	Y	Y	Y
Nodes	N	Y/N	Y	Y	Y
Polygons (areas)	N	N	N	Y	Y
Attributes	N	Y	Y	Y	Y
Actual locations	Y/N	N	Y	Y	Y
Map intelligence	N	N	N	Y	Y
Facilities management	N	Y	Y	N	Y

Notes: Y = Yes; N = No; and Y/N = Both Yes and No.

include scheduling, inventory, and purchasing functions. Implementation of facility management automation can occur at any scale, from a small spreadsheet recording preventative maintenance items to a full corporate system with CAD capabilities managing the complete maintenance process with integrated contracting, accounting, inventory, asset, and personnel records. CAD-based products have normally been classified as computer-aided facility management (CAFM), whereas the database solutions, dealing with work-order management without a CAD component, have been termed computerized maintenance management systems (CMMS). The FM user community reflects this polarization, because the majority of current implementations are oriented either one way or the other. Computer integrated facility management (CIFM) brings both CMMS and CAFM capabilities into synergy, thereby addressing the implementation of a true facility management automation program. The core consideration is that, in practice, CMMS and CAFM capabilities augment and are essential to each other. For example, maintenance activities keep the data required for planning accurate and up-to-date (GEOTELEC, 2004).

AM/FM/GIS is a combination of four separate systems:

- Automated Mapping (AM)
- Facilities Management (FM)
- Automated Mapping/Facilities Management (AM/FM)
- Geographic Information System (GIS)

The differences between the three systems are summarized in Table 14.1 and illustrated in Figure 14.2.

Automated Mapping (AM)

Automated mapping (AM), also known as computer-aided mapping (CAM), is a CAD application for producing maps. It can be considered an alternative to the traditional manual cartographic maps. Data are organized in layers that are conceptually like registered film overlays. Layers organize data by theme (streams vs. roads) and type (linework vs. text).

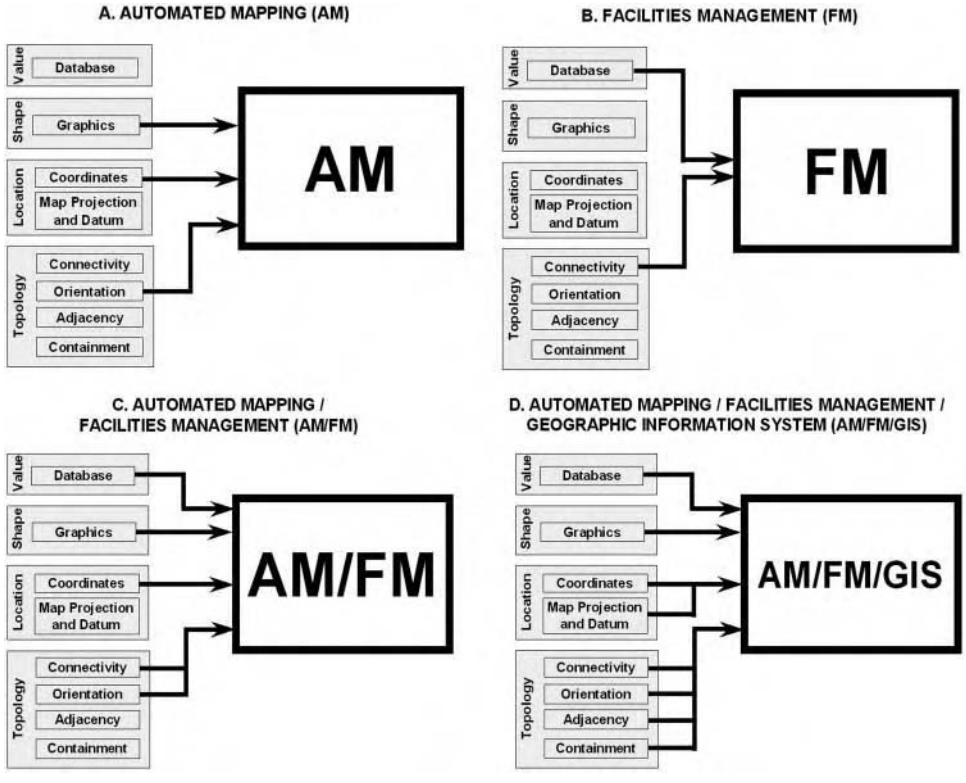


Figure 14.2 Comparison of AM, FM, AM/FM, and AM/FM/GIS.

There are no spatial relations (topology) among data elements except orientation. Figure 14.2(A) shows AM capabilities.

Facilities Management (FM)

Facilities management (FM), also referred to as asset management, is a CAD technology for managing utility system data. FM consists of such activities as inventory, inspection, and maintenance performed by cities, utilities, and government agencies. Organizations incur considerable expenses and resources as these functions are performed on a routine basis. FM includes an infrastructure management database. Compared with AM, there is less emphasis on graphical detail or precision and more emphasis on data storage, analysis, and reporting. Relationships among utility system components are defined as networks. Because FM systems have network definitions, an FM system "knows" the pipes connected upstream or downstream of a given pipe. As shown in Figure 14.2(B), FM systems generally do not have full topology; they offer connectivity and orientation only.

Automated Mapping (AM)/Facilities Management (FM)

AM/FM is a combination of AM and FM technologies. AM/FM software is used to automate maintenance. It allows the integration and automation of maintenance management. AM/FM systems have both orientation and network definitions as shown in Figure 14.2(C). The benefits of AM/FM are improved work-order management, better integrated inventory control, capture of maintenance data and costs, and allocation of costs.

AM/FM/GIS Systems

GIS and AM/FM are different systems but each has its own advantages and applications (Shamsi and Fletcher, 2000). A GIS can help locate the worst pipe. An AM/FM can help users prioritize the work required to bring their worst pipes up to a minimum operating standard. For many years, people have used the GIS and AM/FM systems separately. Developing and maintaining two different systems is expensive and inefficient. Thanks to the latest advances in computer hardware and software, integrated AM/FM and GIS systems called AM/FM/GIS systems are now available. These systems are especially useful for asset inventory, inspection and maintenance, and work management. They are especially popular among visual learners who prefer maps over tables and databases. Visual learners like to click on a manhole in a GIS map to initiate a work order rather than locate it by querying a database. As shown in Figure 14.2(D), AM/FM/GIS systems have more capabilities than AM/FM systems, such as orientation, network, database, and topology.

AM/FM/GIS SOFTWARE

There are currently a variety of computer applications on the market, which were developed for specific components of facility management. In applying automation

Table 14.2 AM/FM/GIS Software Examples

Software	Company	Web Site
	AM Focus	
ArcFM	Miner and Miner	www.miner.com
ArcFM Water	ESRI	www.esri.com
FRAMME	Intergraph	www.intergraph.com
GeoWater and GeoWasteWater	MicroStation	www.bentley.com
	FM Focus	
Cityworks	Azteca Systems	www.azteca.com
GBA Water Master and Sewer Master	GBA Master Series	www.gbamasterseries.com
GeoPlan	Regional Planning Technologies	www.rpt.com
WATERview, SEWERview, STORMview	CarteGraph Systems	www.cartegraph.com
	Proprietary Systems	
CASS WORKS	RJN Group, Inc.	www.rjn.com
IMS-AV and IMS-ARC	Hansen Information Technologies	www.hansen.com

to these diverse functions, vendors have taken approaches that originate from two different technical directions: CAD or relational database management systems (RDBMS) (GEOTELEC, 2004).

GIS-based facilities management requires AM/FM software extensions that can be run from within the GIS software. These add-on programs are also referred to as AM/FM/GIS software. Representative AM/FM/GIS software are listed in Table 14.2. According to this table, there are two types of AM/FM/GIS software:

1. AM focus: These software provide more mapping (AM) capabilities. ArcFM is an example of this type of software.
2. FM focus: These software provide more database (FM) capabilities. Cityworks is an example of this type of software.

Software with an AM focus provide better data editing and mapping capabilities and generally require an FM add-on to provide work-order management functions. Software with an FM focus provide maintenance management functions (e.g., work-order management) but may lack the map maintenance functions. Projects that require both strong AM and FM capabilities may have to use two software products. For example, ArcFM's strong suite of CAD-like map-editing capabilities can be supplemented by Cityworks' work-order management functions.

All the packages listed in Table 14.2 are available for purchase from their respective companies except CASS WORKS, IMS-AV, and IMS-ARC, which are proprietary software programs that are installed by the vendors in their turnkey projects. Figure 14.3 shows Hansen's Infrastructure Management System CCTV-inspection database. Note that each defect has been linked to its location on the sewer pipe and a picture of the defect captured from the inspection videotape.

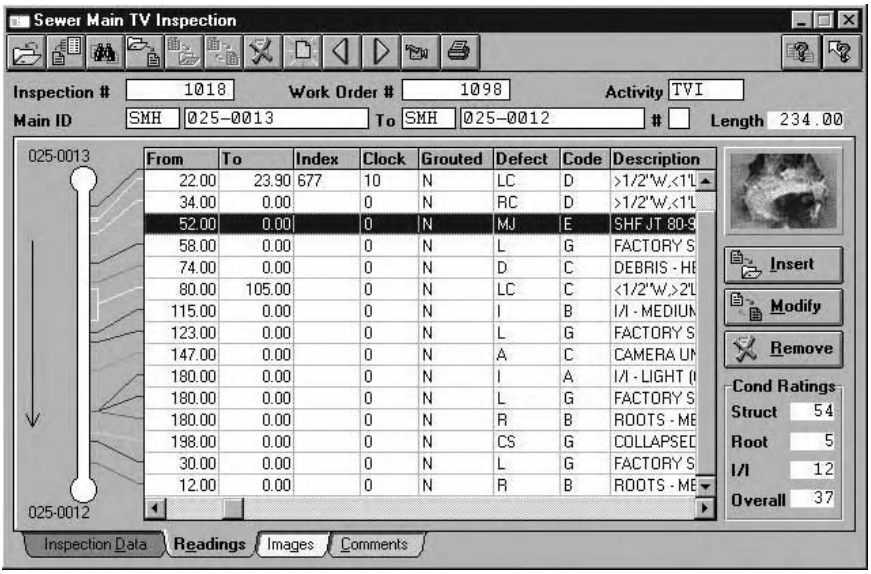


Figure 14.3 Hansen's Infrastructure Management System CCTV-Inspection Database (Photo courtesy of Hansen Information Technologies).

ArcFM

Developed jointly by ESRI and Miner & Miner, ArcFM is an off-the-shelf AM/FM/GIS application for the utility industry and for enterprise implementations. It is designed for editing, maintenance, modeling, and data management of utility information. The first version was written in Visual Basic 5.0 for the Windows NT platform. ArcFM has two components: (1) ArcFM, the main program for creating and editing maps and (2) ArcFM Viewer, the companion program for querying and viewing maps created in ArcFM. ArcFM is suitable for data conversion work by GIS technicians. ArcFM Viewer is a simple program that enables managers to review the finished maps. The relationship between ArcFM and ArcFM Viewer is similar to the relationship between ArcInfo 7.x and ArcView 3.x packages.

In ArcFM, features are modeled as objects with drag-and-drop capabilities. For example, as shown in Figure 14.4, to draw a butterfly valve, users can simply select the valve symbol from the objects library and drag-and-drop it to its correct location. In ArcFM software, GIS tools deal with features that are not simply points and lines, but valves, pipes, manholes, inlets, drainage basins, and more. ArcFM has a rule base engine (RBE) that determines how features are drawn (symbology, placement, rotation, etc.). The RBE can be individualized for different applications. For example, the stormwater RBE may define pipes, their associated features, and the methods that pertain to them. RBE can also validate user-entered data to ensure valid attribute values. With the help of these capabilities, ArcFM can perform routine editing functions, such as adding a new catch basin or service connection or splitting a pipe, with remarkable simplicity.

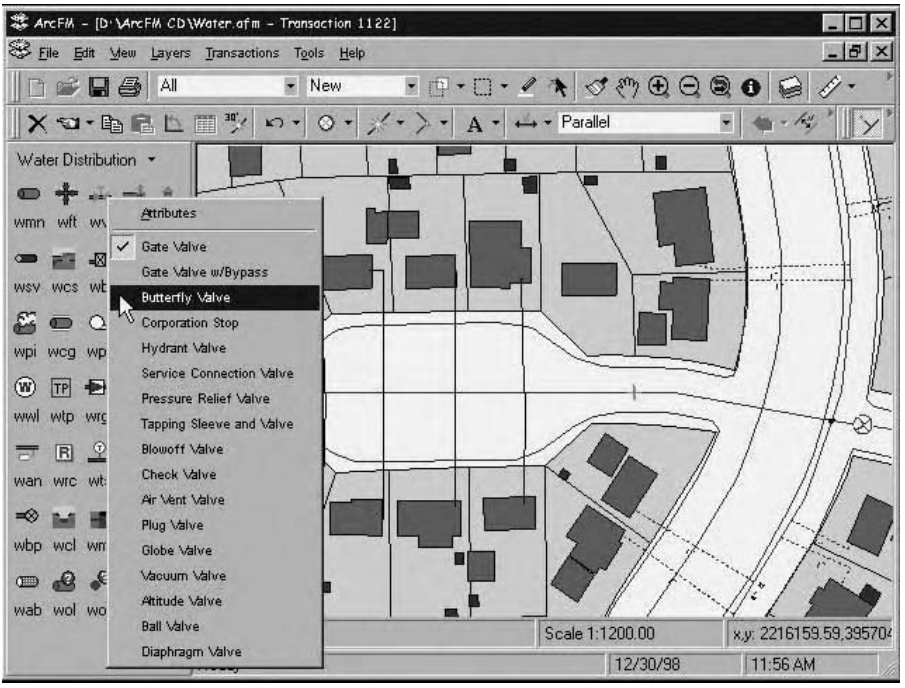


Figure 14.4 ArcFM screenshot showing water objects.

ArcFM has powerful and user-friendly data editing capabilities for routine map maintenance functions, e.g., for adding a new valve or hydrant, splitting pipes, or adding a service connection. ArcFM provides sample data models and business-rule templates for water, wastewater, and stormwater systems. These templates, called *modules,* can be modified for project-specific applications. ArcFM does not have maintenance management capability but it can be integrated with other work management or customer information systems. ArcFM is not a stand-alone program. It requires ArcInfo and Spatial Database Engine (SDE) software and an RDBMS, such as Oracle or Microsoft Access (Shamsi and Fletcher, 2000).

Greenville Utilities was chartered by the North Carolina Assembly in 1905. It is organized as an independent agency of the City of Greenville and operates electric, water, sewer, and gas utilities for the City of Greenville and a portion of Pitt County. Greenville Utilities water operations serve 25,000 customers with 475 mi of main, and sewer operations serve 20,000 customers with 333 mi of main. Greenville Utilities used ArcFM, SDE for its SQL server, and ArcInfo software for implementing an enterprise-wide GIS project for its electric, water, sewer, and gas services. ArcFM Viewer was used by casual users, customer-service representatives, field crews, and others to easily query, plot, and analyze spatial and related attribute data. ArcFM was used as an engineering tool (ESRI, 1999).

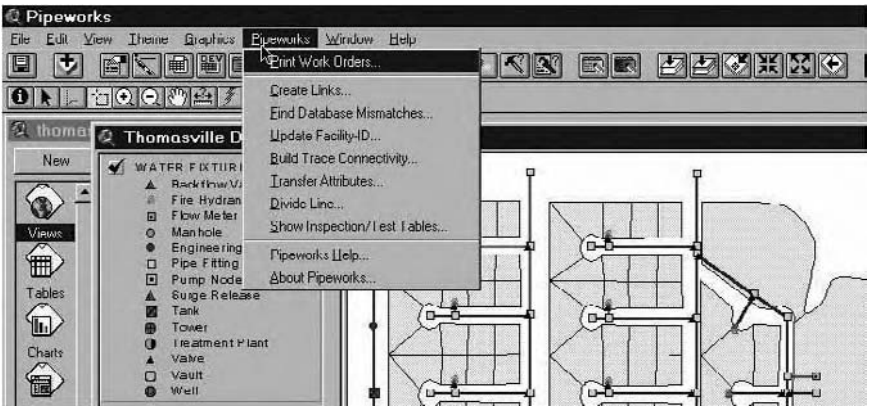


Figure 14.5 Pipeworks menu screenshot.

Cityworks

Cityworks (formerly known as Pipeworks) is available as ArcView 3.x or ArcGIS 8.x and 9.x Extension, helps users integrate their GIS and facilities management. Cityworks' water, wastewater, and stormwater module which for ArcView 3.x costs approximately $4000 per license and works with ESRI coverages and Shapefiles. Cityworks can store project data in any SQL database, such as SQL Anywhere, Oracle, Sybase, etc. Cityworks' capabilities include data inventory, data editing, work-order management, work-order scheduling, network tracing, maintenance histories, inspections, and condition ratings. Cityworks can also be used for managing and recording CCTV inspection programs for sewers. Figure 14.5 shows the Pipeworks menu in ArcView 3.x.

The Cityworks extension provides additional capabilities to make ArcView 3.x or ArcGIS 8.x a complete work-management system. Cityworks work-order modules are used to create and manage work orders, inspections and tests, and detailed inventory for water and sewer systems. The attribute data managed by Cityworks are referred to as extended attributes because these attributes are stored in the extended-coverage database or RDBMS. Examples of extended attributes are inventory data, work orders, work histories, inspections, customer complaints, etc. Cityworks for ArcView 3.x requires that all inventoried features be created and maintained as coverage data types, i.e., line, node, or point feature. The difference between Cityworks and other work-management solutions is its GIS-centric approach. Cityworks does not treat the GIS as subordinate to the work-management system. Instead of integrating the work-management system with GIS, it uses the GIS as a work-management system (Azteca, 1999). General functions of Cityworks are listed below:

- Assign work activity
- Link to service request
- Track tasks
- Track resources: labor, material, and equipment (estimated and actual)
- Link multiple inspections and tests

Table 14.3 Cityworks Inspections and Tests

Water	Wastewater	Storm
Meter inspection	Manhole inspection	Manhole inspection
Hydrant inspection	CCTV inspection	CCTV inspection
Valve inspection	Inlet inspection	Inlet inspection
Hydrant flow test	Dye test	Dye test
Meter changeout	Smoke test	Smoke test
	General test	General test

- Maintenance:
  - Corrective or reactive
  - Preventive or routine
  - Scheduled or cyclical
- Track permits
- Key dates and personnel
- Attach images
- Geocode and geolocate
- Print using Microsoft Word and ArcView templates and automated printing functions
- Search on any combination of work-order ID, work activity, selected map features, feature, type, key dates, key personnel, tasks, status, address, or geographical constraints.

Cityworks' inspections and tests are listed in Table 14.3. Another Azteca product called the Cityworks ArcGIS Interface simplifies the ArcGIS interface by providing a subset of key ArcGIS functionality (presented using large icons) commonly used by maintenance and operations staff. Cityworks Desktop ArcGIS Extension is an extension of the standard ArcGIS interface. It is accessible from a palette of tools that can be positioned anywhere on the map document.

CHAPTER SUMMARY

Deteriorating infrastructure and limited funds require that maximum system improvement be achieved at the lowest possible cost. AM/FM/GIS systems can be used to determine optimal infrastructure improvement strategies to accomplish this goal. Today, AM/FM/GIS systems are helping the water industry with their planning, design, and operational needs. AM/FM/GIS software is engineered to organize asset information, optimize maintenance activities, and prioritize necessary improvements. AM/FM/GIS software offers the integrated strength of AM/FM and GIS programs. AM/FM/GIS software is available in AM-centric and FM-centric flavors. Software selection should be made based on user-specific needs. Applications of AM/FM/GIS technology are presented in the next chapter.

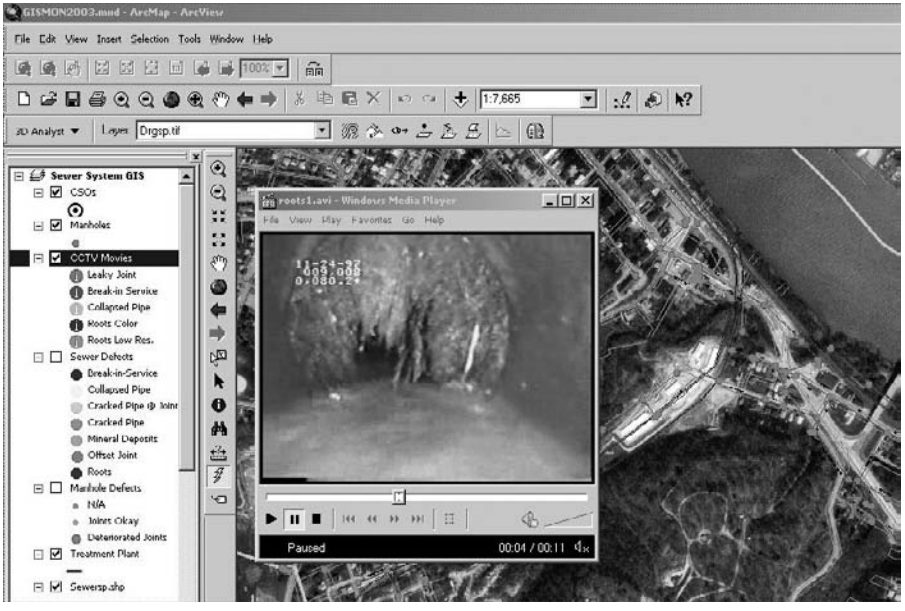
CHAPTER QUESTIONS

1. What is AM/FM/GIS, and how does it differ from AM/FM and GIS?
2. What is work-order management, and how is it done using AM/FM/GIS software?
3. List ten applications of AM/FM/GIS technology.

CHAPTER 15

Maintenance Applications

GIS can be used to prepare inspection or maintenance work orders simply by clicking on a sewer pipe or manhole. This approach simply takes just a few minutes compared to the conventional method of locating and copying maps and typing the work order forms, which usually takes several hours.



Displaying sewer CCTV inspection video in GIS. The movie shows a sewer blocked by heavy root growth.

LEARNING OBJECTIVE

The learning objective of this chapter is to understand GIS applications in inspection and maintenance of water, wastewater, and stormwater systems.

MAJOR TOPICS

- Field inspections
- Asset management
- GASB 34 applications
- Wet-weather overflow management applications
- CCTV inspections
- Video mapping
- Thematic mapping of inspection data
- Work-order management
- Identifying valves to be closed for repairing or replacing broken water mains
- System rehabilitation and repair
- Case studies

LIST OF CHAPTER ACRONYMS

AM/FM/GIS Automated Mapping/Facilities Management/Geographic Information System
AVI Audio Video Interleaved (digital movie format)
CCTV Closed-Circuit Television
CD Compact Disc
CMMS Computerized Maintenance Management System
DVD Digital Versatile Disc
GAAP Generally Accepted Accounting Principles
GASB Government Accounting Standards Board
HTML Hyper Text Markup Language (a file extension)
MPEG Moving Picture Experts Group (digital movievideo format)
O&M Operation and Maintenance
PDA Personal Digital Assistant (electronic handheld information device)
PDF Portable Document Format (Adobe Acrobat)
ROM Read Only Memory
VCR Video Cassette Recorder
VHS Video Home System (video cassette format)

This book focuses on the four main applications of GIS, which are mapping, monitoring, modeling, and maintenance and are referred to as the "4M applications." In this chapter we will learn about the applications of the last *M* (maintenance).

To fully appreciate the benefits of GIS-based inspections, consider the following hypothetical scenario. On March 10, 2004, following a heavy storm event, a sewer

customer calls the Clearwater Sewer Authority to report a minor basement flooding event without any property damage. An Authority operator immediately starts the GIS and enters the customer address. The GIS zooms to the resident property and shows all the sewers and manholes in the area. The operator queries the inspection data for a sewer segment adjacent to the customer property and finds that a mini movie of the closed-circuit television (CCTV) inspection dated July 10, 1998, is available. The operator plays the movie and sees light root growth in the segment. A query of the maintenance history for that segment shows that it has not been cleaned since April 5, 1997. This information indicates that the roots were never cleaned and have probably grown to "heavy" status. The operator highlights the sewer segment, launches the work-order module, and completes a work-order form for CCTV inspection and root removal, if necessary. The export button saves the work-order form and a map of the property and adjacent sewers in a PDF file. The operator immediately sends the PDF file by e-mail to the Authority's sewer cleaning contractor. The entire session from the time the customer called the Authority office took about 30 min. The operator does not forget to call the customer to tell him that a work order has been issued to investigate the problem.

BUNCOMBE COUNTY'S SEWER SYSTEM INSPECTION AND MAINTENANCE

During the 1990s, the Metropolitan Sewerage District (MSD) of Buncombe County (North Carolina) spent more than $111 million rehabilitating the most problematic sewer lines in its system. MSD's aggressive sewer rehabilitation program has successfully implemented numerous rehabilitation technologies. In 1999, MSD began assessing a method of comprehensive basinwide rehabilitation. In this approach, every pipe in a drainage basin was evaluated using CCTV footage and engineering analysis. After evaluating this comprehensive basinwide rehabilitation method for 4 years, MSD determined that it is overly time- and capital-intensive. Seeking a more efficient and quicker method to fix the system, MSD developed and implemented a new GIS-based rehabilitation method called "pipe rating."

The pipe rating method has five main elements: (1) CCTV information, (2) a defect scoring system, (3) GIS database software, (4) sanitary sewer overflow (SSO) history, and (5) engineering analysis. These components are combined to generate specific projects for problem lines. Each structural defect noted in the video inspection is given a defect score, in accordance with MSD's standardized scoring system. For example, a circumferential crack is given a score of 20, and a collapsed pipe is assigned a score of 100. Defects are then embedded within the GIS database, with appropriate scores attached. Upon quantifying all structural defects within a pipe segment, three defect ratings are generated in the GIS: (1) peak defect rating, (2) mean defect rating, and (3) mean pipeline rating. These ratings allow users to visualize the severity of pipe defects using GIS. Figure 15.1 shows an ArcGIS screenshot of mean pipeline ratings. Such maps are used in prioritizing pipe segments for repair work. They are also used to determine when several point repairs should be made to

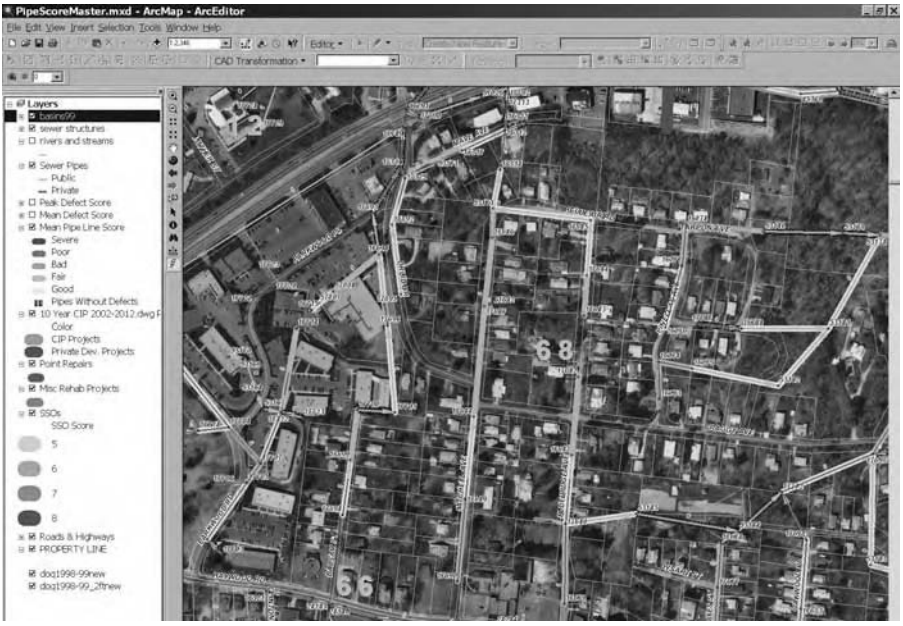


Figure 15.1 ArcGIS screenshot of mean pipeline ratings for the Metropolitan Sewerage District (MSD) of Buncombe County, North Carolina.

a line as apposed to excavating the entire line (Bradford et al., 2004). Additional information about this application is provided in Chapter 17 (Applications Sampler).

ASSET MANAGEMENT

As discussed in Chapter 14 (AM/FM/GIS Applications), at present our water and wastewater infrastructure, especially in the older cities, is in critical stages of deterioration and has started to crumble. Nationally and internationally, aging water and wastewater infrastructure is imposing enormous costs on local communities (Booth and Rogers, 2001). In the U.S., cities and utilities are demanding billions of dollars of government grants and funds for renovating their water infrastructure. Due to an "out-of-sight, out-of-mind" philosophy and the lack of funds to follow a preventive maintenance practice, the replacement is mostly performed on a react-to-crisis basis. A crisis maintenance program only corrects infrastructure problems after they have happened. Notwithstanding the conventional wisdom, this reactive approach may not be the best strategy as substantial expenditure and inconvenience can be avoided by replacing a deteriorated pipeline before it actually breaks. A preventive maintenance program is proactive because it strives to correct a problem before it occurs.

For water and wastewater systems, asset management can be defined as managing infrastructure capital assets to minimize the total cost of owning and operating them while delivering the service levels customers desire (Booth and Rogers, 2001). A typical asset management system has five components (Doyle and Rose, 2001):

1. Facilities inventory: Describes each system element in an asset group. GIS can be very useful in completing this task.
2. Condition assessment: Classifies each asset according to its capability to perform the intended function.
3. Valuation: Assigns a financial value to inventoried assets consistent with Generally Accepted Accounting Principles (GAAP).
4. Operations, maintenance, repair, and replacement management: Arguably the heart of a management system, this component tracks and records data about work orders and customer complaints, issues and tracks preventive and predictive maintenance schedules, and generates crew assignments and work-site maps. GIS has extensive capabilities to fulfill this part.
5. Analysis and evaluation: Considered as the brains of an asset management system, this component prioritizes work effort, analyzes cost-effectiveness, and optimizes asset performance.

An asset management system helps predict the future condition of assets and major rehabilitation costs for planning purposes. An effective asset management system can reduce the cost of system operation and maintenance (O&M). Every successful maintenance program should begin with an accurate system map because it is difficult to maintain a system if the users do not know where the water or sewer lines are. A well-constructed GIS should be used to create the system map. Historical maintenance data should also be linked with the GIS because it is difficult to schedule maintenance when you do not know the last time it was done (Gaines, 2001). A GIS-based asset management system can be used as a decision support system for capital improvement planning (CIP). For example, CH2M HILL (Atlanta, Georgia) used ESRI's MapObjects GIS software to develop an infrastructure capital asset management (ICAM) toolkit that can be used in a Web-based browser/server or stand-alone computing environment (Booth and Rogers, 2001).

In many cases, through more effective planning and management of infrastructure improvements and system operations, organizations can realize annual savings of 20 to 40% (Stern and Kendall, 2001). Boston Water and Sewer Commission (BWSC) has approximately 27,500 catch basins. In 2000, BWSC started enhancing its asset management program for locating, inspecting, and maintaining its catch basins. BWSC improved the productivity and efficiency of its catch basin preventive maintenance program by integrating it with GIS and a computerized maintenance management system (CMMS). In lieu of using global positioning system (GPS), the crews determined the precise location of catch basins by using a measuring wheel to measure distance from known points to the basin's center. BWSC utilized handheld touch-screen computers in the field to collect geographic location information and more than 25 attributes. GIS data in Shapefile format and a GIS interface provided the backbone of this application. The traditional paper-based data collection methods typically averaged 24 catch basin inspections per day, but the GIS-based method boosted this rate to approximately 42 per day.

Boston Water and Sewer Commission's GIS-based field-data-collection approach increased the catch basin mapping productivity by more than 40% compared with traditional, paper-based data collection methods (Lopes et al., 2002).

In the U.S., two key drivers are motivating improved asset management practices in water and wastewater utilities: (1) GASB 34 requirements, and (2) wet-weather overflow requirements.

GASB 34 APPLICATIONS

In 1999, the Government Accounting Standards Board (GASB) issued Rule 34 (known as GASB34) to govern financial reporting requirements of approximately 85,000 state and local governments in the U.S. Considered the most significant change in the history of government financing reporting in the U.S., GASB 34 requires the cities to adequately account for and report their capital asset inventory in a complete, accurate, and detailed manner (Booth and Rogers, 2001). The capital assets include infrastructure networks such as roads, bridges, and water, wastewater, and stormwater systems.

The literature indicates that a GIS-based approach saves time spent in locating, organizing, and confirming the accuracy of field inspection information (Criss, 2000). Industry experts believe that utilities can cut their maintenance costs in half by implementing GIS-based preventive maintenance programs. Thus, integration of GIS and maintenance management software is a natural progression for GIS applications in the water and wastewater industry. For instance, the Wastewater Collection Division of Fairfax County, Virginia, linked the county's sewer maps with the sanitary sewer maintenance management system database, making it easier to access maps during field activities (Fillmore et al., 2001). The City of Denver, Colorado, linked the city's sewer system GIS to an information management system to efficiently track and manage maintenance operations (Gaines, 2001).

WET WEATHER OVERFLOW MANAGEMENT APPLICATIONS

Management of wet weather overflows is a fertile field for GIS technology. By using geographic information in mapping, facilities management, and work order management, a wastewater system manager can develop a detailed capital improvement program or operations and maintenance plan for the collection system.

Broken and damaged sewers, laterals, and manholes usually contribute significant amounts of wet weather inflow and infiltration (I/I) to a wastewater collection system. This contribution often results in combined sewer overflows (CSO) from combined sewer systems and sanitary sewer overflows (SSO) from sanitary sewer systems.

In the U.S., CSO discharges are regulated by U.S. EPA's National CSO Policy. The policy requires a System Inventory and Physical Characterization report. Major portions of this report can be completed using GIS. The CSO policy's Nine Minimum Controls (NMC) mandate proper operation and regular maintenance programs for the sewer system and CSO outfalls that can also benefit from inspection and maintenance applications of GIS. SSO discharges are being regulated by U.S. EPA's SSO rule that requires implementation of a Capacity, Management, Operations, and

Maintenance (CMOM) program. CMOM requires the system owners/operators to identify and prioritize structural deficiencies and rehabilitation actions for each deficiency. CMOM requirements offer a dynamic system management framework that encourages evaluating and prioritizing efforts to identify and correct performance-limiting situations in a wastewater collection system. CMOM is a combination of planning tools and physical activities that help communities optimize the performance of their sewer systems. CMOM requirements mandate that the system owner/operator properly manage, operate, and maintain, at all times, all parts of the collection system. The owner/operator must provide adequate capacity to convey base flows and peak flows for all parts of the collection system. CMOM requirements include "maintaining a map" (Davis and Prelewicz, 2001), which is the simplest application of GIS as described in Chapter 9 (Mapping Applications).

Some wet weather overflow management experts believe that use of GIS is a must for CMOM compliance.

AutoCAD Map GIS Application for CMOM

The Stege Sanitary District (SSD) located near San Francisco Bay serves a population of approximately 40,000 in a 5.5 mi^2 area. The goal of a CMOM program is the ultimate elimination of any type of overflow from the sanitary sewer system. SSD was challenged with the goal of "no overflows" in 1996. To meet this goal, a new maintenance model was implemented for sanitary sewers that used AutoCAD Map GIS and Microsoft Access database software to identify system conditions causing overflows, and prompted immediate action to correct problems via repair or replacement. Those lines showing the greatest damage and whose repair or replacement were within the budget constraints established by the District, would be immediately set right. The ability to make effective and economic decisions regarding the capital replacement needs of the system based on the actual degradation of the system provided a degree of asset management that had not been readily available to the management. The SSD experience indicated that proactive maintenance provides the most cost-effective means of managing the system to achieve the goal of "no overflows."

During 1992 to 1994, SSD created the sewer and manhole layers in AutoCAD format and linked the capacity model output to these layers to identify line segments that were overloaded. Complete GIS capability was added later with the use of AutoCAD Map, an extension of the common AutoCAD drafting program that adds the feature of linking information from external databases. During 1994 to 2001, information was collected in a database on the physical system characteristics, system hydraulic performance under selected flow conditions, system overflows, routine maintenance activities, CCTV inspections, and repairs and replacements. All database information was linked to the mapped line segments, thereby allowing an easy evaluation of problem distribution throughout the system with a textual and graphical response from the database query. The relational database program used

was Microsoft Access, which is compatible with AutoCAD Map. All of the work was completed on a desktop PC (Rugaard, 2001).

CCTV INSPECTION OF SEWERS

Traditionally, municipalities have relied on analog video for internal inspection of sewers. In this technique, remote-controlled and self-propelled cameras move through sewer lines to record the pipe's internal conditions on a video tape. In a parallel operation, technicians create paper logs identifying the location, size, and other key information about the pipe. Referred to as CCTV inspection, this method has been considered to be the most effective and economical method of pinpointing the sources of I/I contribution. For instance, by the late 1990s the City of Boston, Massachusetts, had conducted 12 million ft of CCTV inspection on video tapes. Using an inspection cost of one dollar per foot, the cost of collecting this information could be at least $12 million. CCTV inspection videotapes contain a wealth of useful information about the state of a collection system, yet they have been often treated as single-use items. A lot of money is spent on video inspections, yet inspection data are frequently underutilized mainly because accessing the information from the conventional video tapes and paper inspection log sheets has been difficult and time consuming (Criss, 2000).

GIS can be used as a document management system for CCTV inspection data. For example, users can click on a rehabilitated sewer pipe to see the "before" and "after" movies on their computer screen. This application, however, requires converting VHS video tapes to computer files (digital movies). Once converted and stored on computers, the valuable information once hidden in videotapes can be retrieved with the click of a mouse, eliminating the need for other equipment (TV, VCR, and cables) and office space to house the equipment. CDs take up 70% less shelf space than VHS tapes. As a side benefit, digital movies can be used to create multimedia presentations for utility management when requesting maintenance and rehabilitation funds. After all, a video is worth a thousand pictures. Most video tapes have a short shelf life of approximately 10 years and their video quality and performance are inferior to digital media. Unlike VHS tapes, digital videos do not lose picture quality when copied. Digital technology also allows high-resolution snapshots of defective pipe segments taken from the video. Dynamic segmentation and image integration features allow storage and display of these images with footage reading and a description of the pipe defect. For benefits like these, experts predict that by 2008, approximately 90% of wastewater utilities will be using digital video technology (Bufe, 2003).

As stated in the preceding text, integrating CCTV videos with GIS requires migrating from video tapes to digital movies. Four migration methods are possible:

1. Convert existing video tapes to digital files
2. Digitize existing video tapes
3. Retrofit tape systems with digital systems
4. Record directly in digital format

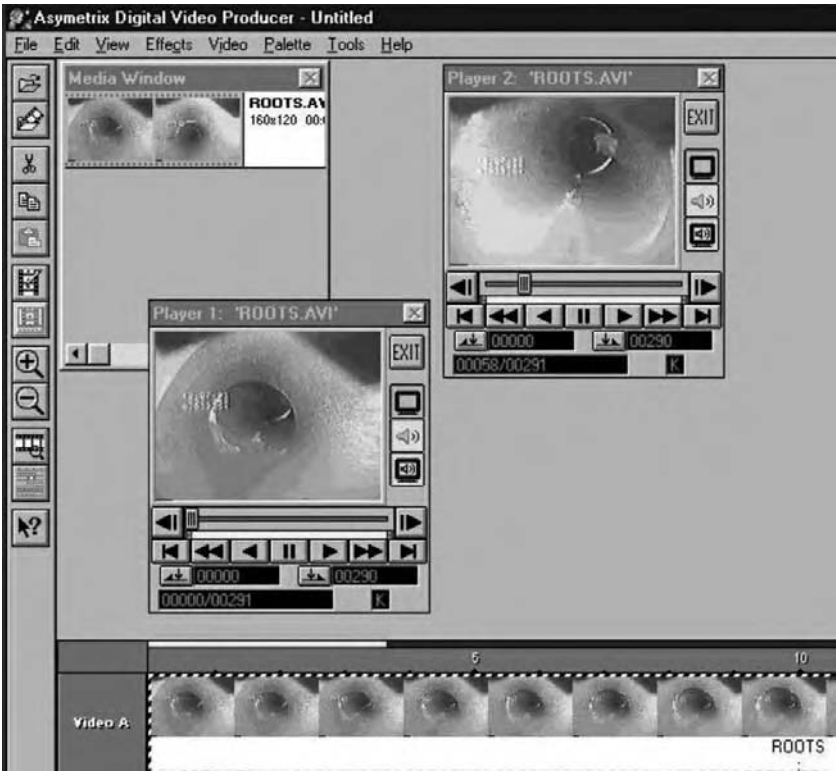


Figure 15.2 Screenshot of Digital Video Producer's video capture capability.

Convert Existing Video Tapes to Digital Files

In this method (which is the simplest), the existing video tapes are converted to multimedia computer files in the AVI or MPEG format. The hardware required for this method includes TV, VCR, multimedia computer, CD or DVD burner, and a video capture card. Video capture cards allow display of TV or VCR output on a computer screen. Some multimedia computers have built-in (internal) video capture cards. If an internal card is not installed, an external card that typically costs a few hundred dollars should be used. A video capture and editing software, which typically comes with the card, is also required. Figure 15.2 shows a screenshot of Digital Video Producer, a video capture software from Asymetrix Corporation.

The TV is connected to the VCR, and the VCR is connected to the video capture card (if external) or the computer (if internal) using AV cables. When a tape is played in the VCR, the user sees the video on the TV and on the computer screen. The video capture software has a VCR-like console with play, stop, and record buttons. When the user clicks on the software's record button, usually at the beginning of a sewer defect (e.g., collapsed pipe, root growth, break-in lateral, etc.), the computer starts to record the video in a computer file, usually in AVI or MPEG format. The

Figure 15.3 Sample mini movie in AVI format showing collapsed sewer pipe.

file recording stops when the STOP button is clicked, usually at the end of the sewer defect. Instead of converting the entire video tape, generally 5 to 30 sec mini movies showing problem areas are captured because the resulting files can be very large. For example, a 5 to 15 sec video segment recorded in color with sound can be 30 to 60 MB in size. Sound can be turned off to reduce the file size. Finally, in order to free up the computer hard disk space, the recorded digital movies are copied to digital media (e.g., CD-ROM or DVD-ROM). Raw video files can be compressed to optimize disk space. A CD cannot store more than 15 to 30 min of uncompressed data. However, in compressed format such as MPEG-1 or MPEG-2, 1 to 1.5 h of video can be stored on a CD and 6 to 12 h on a DVD. Figure 15.3 shows a sample mini movie in AVI format showing a collapsed pipe.

Digitize Existing VHS Tapes

This is basically a more efficient and automatic implementation of the previous method. Special video software is used for compressing and indexing video tapes to digital media. For example, the Tape-to-CD module of flexidata pipe survey reporting software from Pearpoint, Inc. (Thousand Palms, California), uses advanced optical character recognition (OCR) technology to read the footage count displayed on the monitor

screen. This product was designed using skills and expertise gained in the development of automatic recognition of vehicle license plates. It allows the Pearpoint system to provide indexing of sewer records against the footage shown on the original videotape. The resulting CD will not only contain the original real time video but also a report.

WinCan

WinCan from WinCan America, Inc. (Durango, Colorado), is another example of video software that allows the user to capture either videos or pictures. Pictures may be captured in several different formats such as JPG or BMP. Mini movie files can also be created in increments of 5 to 35 sec. Videos may be captured in several different formats, including the most common, MPEG and AVI. WinCan can integrate stored pictures, live video from the camera or VCR, and stored movies in the pipe inspection database. Figure 15.4 shows a screenshot of the WinCan interface. The left window lists the inspected pipes in the WinCan project database. The right windows shows the mini movies for the selected pipes. The bottom window displays the observed defects for the selected pipe. WinCan also allows personal digital assistant (PDA)-based manhole reporting. The PDA (e.g., an iPAQ or PalmPilot) is used in the field to gather data with the user standing above the structure. A customized WinCan software is added to the user's PDA. Pull-down menus and easy fill-in fields make gathering the data as easy as using the full WinCan software. Back in the office, the user HotSyncs the PDA with the computer and the information is simply added to WinCan (WinCan, 2001).

WinCan interacts with GIS packages in two ways: (1) interfaces with the GIS database and (2) updates the GIS database. In the first method, the WinCan database interfaces with the GIS database by using a metadatabase. This metadatabase contains information from WinCan projects, such as unique manhole identifiers, line section numbers, street names, and pipe sizes. When a GIS user attempts to access a particular feature (e.g., a manhole or line segment), the software searches the metadatabase to determine if the feature is available. If an associated WinCan report is available, the software opens the report and highlights the specific item. The user can then display associated pictures, videos, or inspection data. This method requires toggling between WinCan and the GIS software. In the second method, an external utility appends the WinCan data to the GIS database. GIS users can then select the locations in their maps and view WinCan database and associated graphics without starting the WinCan program (Criss, 2000).

Retrofit Tape Systems with Digital Systems

The typical cameras on board the CCTV inspection vehicles are adequate for use in digital applications. Most simply need the addition of a compression unit to convert their analog signals to a digital format. In this method, digital video processing hardware and software are installed in the CCTV inspection vehicle and linked to an existing tape system.

Pearpoint equipment users can interface the Pearpoint hardware directly with WinCan. This eliminates the need for a separate overlay system with encoder.

Figure 15.4 WinCan software screenshot.

According to Pearpoint, the installation is much easier with no hardware integration and wiring/soldering necessary. Camera inclination and other Pearpoint features function instantly with appropriate setup.

Another example of this approach is the PipeTech™ Software Suite from Peninsular Technologies, Inc. (Ada, Michigan). PipeTech is an innovative solution for simplifying sewer inspection projects. It allows collection, editing, analysis, and archiving of all inspection data digitally, eliminating the need for VHS tapes and the stacks of paper attachments that normally accompany an inspection project. PipeTech's Scan product is an on-site solution for video recording and incident reporting. Scan captures and compresses video in real time and facilitates the recording of incident locations. Incidents are entered on an electronic log sheet and are automatically cross-referenced with their occurrences in the video. Scan eliminates the need for a TV, VCR, and VHS tapes in the video vehicle. Videos captured using Scan are clearer than those captured through a VCR (PipeTech, 2000).

Video software such as WinCan and PipeTech also provide an on-screen pipeline or a video viewer for retrieval of stored images and movies. Consisting of a line segment graphic, these utilities allow the user to scroll along a line segment and display observations, photos, or movies by clicking on any desired location. An example is shown in Figure 14.3 for Hansen's Infrastructure Management System. This is possible because the defect location is cross-referenced with the video. In addition to the typical stop, play, and pause controls, video step and loop controls help the user to direct video progression. Seeking to any instant of video is also possible.

Record Directly in Digital Format

In this method, no VHS tapes are involved at all. The sewer inspection video is recorded directly in onboard computers as digital movies, using digital cameras. This method requires installation of appropriate digital recording hardware in the CCTV vehicle. For example, Pearpoint manufactures several video pipeline inspection systems with digital color video cameras.

This system offers complete integration of data and video. It enables a technician to watch the video as it is recorded and flag the video for defects in real time. Users can then jump to the flagged video frames without searching through the entire video (Bufe, 2003).

Linking Digital Movies to GIS

The digital movies created by using one of the preceding methods should be linked to the sewers layer in the GIS database. This will enable users to click on a sewer line in the GIS map to display the digital movie of the sewer defect. The following steps should be performed to accomplish the linkage in ArcView 3.x:

1. Add a new field to the sewer layer table to store the movie filename. Figure 15.5a shows an ArcView 3.2 database table for sanitary sewers containing a field named "Movie" for holding the movie filenames.
2. Enter the video filename for each sewer segment for which a movie is available.

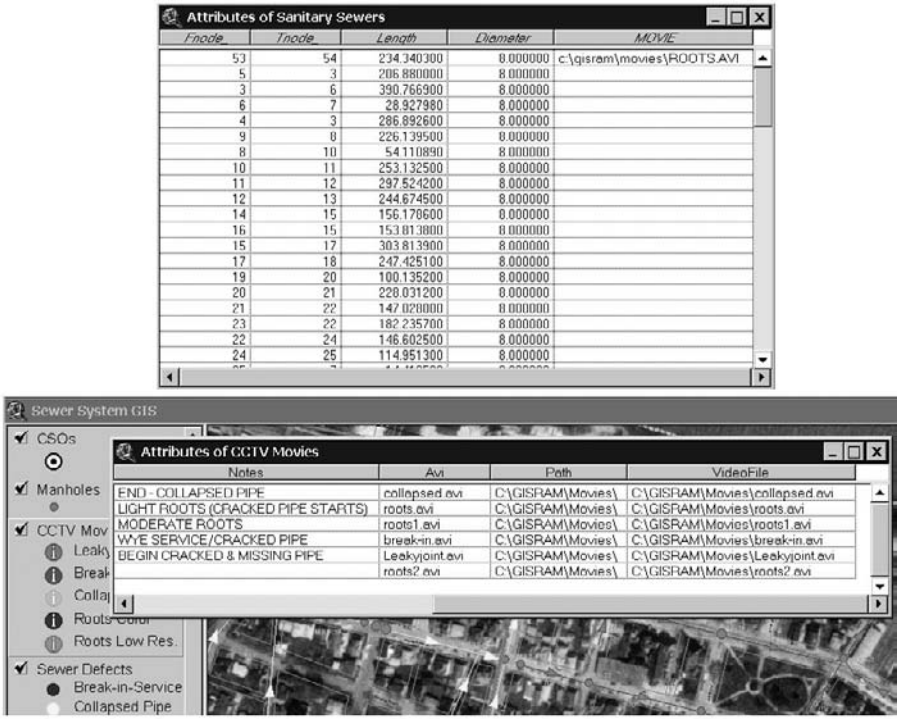


Figure 15.5 Linking digital movies to ArcView GIS. Above (a) linkage to sewers; below (b) linkage to points.

3. Write an Avenue script (shown in Figure 15.6 and described in the following text) for displaying the movie.
4. In the "Theme Properties" dialogue box for the sewers theme, click on the "Hot Link" tab as shown in Figure 15.6. Set "Field" to the field created in Step 1, set "Predefined Action" to "Link to User Script," and set "Script" to the script developed in Step 3.
5. Pick the Hot Link tool (the lightning bolt) shown in Figure 15.6 from the tool bar and click on a pipe with a video file to launch the digital movie.

Figure 15.6 shows a sample script called "Movie&ImageHotLink" for displaying movie files. The code of the script is included in Chapter 1 (GIS Applications, Applications Programming section).

The successful linkage of movies to sewers requires that both the CCTV inspection data and the sewer database table in the GIS have a common ID. This would essentially require planning ahead before conducting CCTV inspections to make sure that the inspection crews use the GIS IDs in their inspection reports. As an alternative to linking videos to sewers, a point layer can be created for video locations. Figure 15.5b shows an ArcView 3.2 point database table for CCTV movie locations. It has three fields for entering movie filenames. The users can enter the video filename and path separately in the AVI and Path fields. Or, an absolute pathname can be

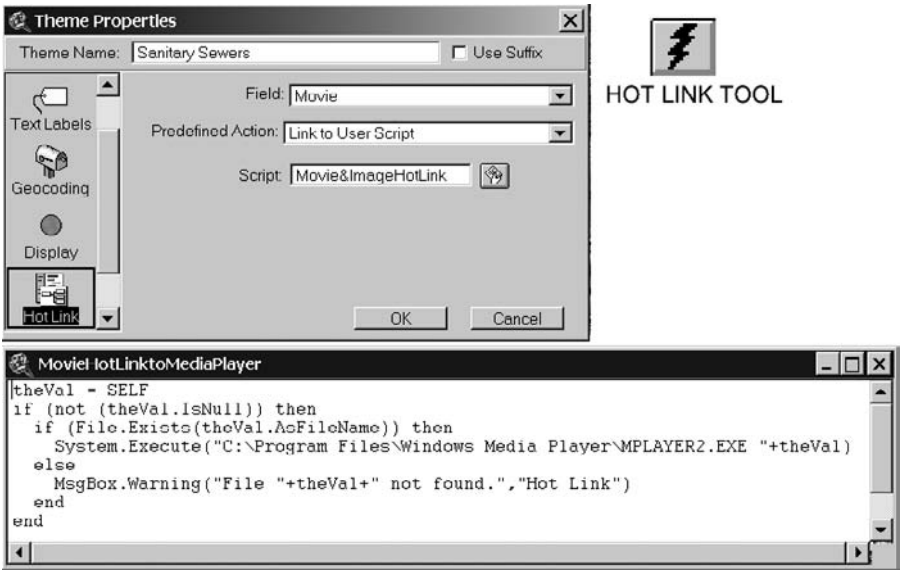


Figure 15.6 ArcView GIS Hot Link Tool for linking sewer movies to sewer pipes.

entered in the VideoFile field. The disadvantage of absolute pathnames is that the application will not work if the pathname becomes invalid when video files are moved to another location (e.g., a CD or network drive). Here the users can click on a video point to see the digital movie.

In ArcGIS, the video files can be displayed without a script using the Hyperlink tool, which also looks like a lightning bolt. As in ArcView 3.x, the movie filename is entered in a database field. When the end of the bolt is near a defective sewer, the hyperlinked filename is displayed if the file exists. Clicking on the sewer will launch the movie. The hyperlinking is configured by right-clicking on the sewer layer, selecting properties, and then the display tab. The figure on the first page of this chapter shows a digital movie being played in ArcGIS 8.2.

VIDEO MAPPING

Today, GIS map features can be linked to any kind of file to document inspection of a location through pictures, field notes, audio, streaming video, etc. When digital video, GPS, and GIS maps are combined, the possibilities are limited only by the user's imagination.

Traditionally, GIS users have employed a low-tech and cumbersome approach to associate an inspection picture (e.g., inside of a sewer manhole) or video with its location. This method involves taking pictures (or recording video), writing the pictures' (or video's) description and frame numbers, scanning the pictures (or converting the video from analog to digital format), collecting GPS coordinates of picture locations, and entering all this information in the GIS database. In today's

information age, this method is error prone and inefficient. The high-tech alternative known as video mapping (or media mapping) refers to georeferencing video footage by associating GPS data with video (Pfister et al., 2001). A GIS software can then read the GPS locations referenced to video time codes into its database. A map of video locations is created that can be clicked by the users to access the desired video footage. Video mapping technology can be applied to document sewer inspections and smoke-testing in GIS maps (Murphy, 2000).

Video mapping integrates GPS data and GIS maps to create interactive multimedia maps without manual scanning of photos or data entry into a computer.

Multimedia Mapping System (MMS) and Video Mapping System (VMS) software programs from Red Hen Systems, Inc. (Fort Collins, Colorado), are typical examples of video mapping products. VMS is a self-contained GPS encoder that records all data types — video, audio, and GPS — on a single tape. Users collect video, location, and audio information at the same time. It contains an integrated GPS receiver that receives location coordinates from satellites and then converts them to an audio signal for storage on video tape (Morrison and Ruiz, 2001). The user attaches the product to a video camera and records the field inspections. VMS sends GPS data from its internal receiver to the camera's microphone input. After recording the video, the user connects the system to a computer with Red Hen's MediaMapper software installed. VMS then sends the data to the computer to create a video map in TAB, SHP, or HTML format. These maps allow the users to click and view still images or video clips, play audio files, or display documents or data sets. MMS integrates the VMS unit and its built-in GPS receiver with a GPS antenna, headphone, and microphone.

MediaMapper software provides basic GIS functions and enables users to link georeference audio and video clips to GIS features. This functionality is essentially similar to ArcView's Hot Link tool described in the previous section but it is enabled by MediaMapper automatically. Software extensions are available for accessing MediaMapper data in MapInfo and ArcView GIS packages.

THEMATIC MAPPING OF INSPECTION DATA

Some applications use information from several different databases or tables. For example, a smoke-testing application might require information from the manhole and customer account databases. It is inefficient and cumbersome to enter the data from different tables in one table. Relational database tables allow accessing information from different tables without physically joining them.

A relational database stores information in records (rows) and fields (columns) and conducts searches by using data in specified fields of one table to find additional data in another field. This capability allows linking two tables that have a common field. The common field can be used to link the GIS feature tables to virtually any external database table. This linkage capability allows the GIS to make effective use of existing databases without requiring additional data entry in the GIS database.

Once the GIS and external database tables have been linked, the external data can be queried or mapped from within the GIS.

Typical manhole inspection reports have valuable attributes, such as:

- Structure number
- Street name
- Inspection date
- Offset from curb
- Structure type
- Cover or grate diameter
- Number of holes in the cover
- Frame adjustment type
- Corbel type
- Manhole wall type
- Manhole area
- Depth of structure
- Bricks missing (Yes/No)
- Ladder rungs (Yes/No)
- Ladder rungs condition
- Apron leaking (Yes/No)
- Pipe seals leaking (Yes/No)
- Manhole wall leaking (Yes/No)
- Catch basin connected (Yes/No)
- Root intrusion (Yes/No)
- Joint deterioration (Yes/No)

A manhole theme table can be linked to a manhole inspection table to create thematic maps of manhole conditions, which are excellent decision support tools for scheduling maintenance and inspection activities. The linkage can be created easily as long as both the inspection reports and the GIS database use the same manhole IDs.

Figure 15.7 shows the linkage of a manhole theme table in ArcView with the external manhole inspection table. The linkage is accomplished using the fields "Mhid" of the manhole theme table and "Mh_id" of the manhole inspection table. When the manhole "MH9-2" is selected in the GIS theme table, corresponding records in the manhole inspection table are also selected. Figure 15.8 shows a thematic map of manhole joint defects based on the "Joint Deterioration" attribute of the manhole inspection report created by using the table-linkage approach.

Similarly, typical CCTV inspection logs also have valuable attributes, such as:

- Date televised
- Type of pipe
- Begin and end manhole
- Total footage televised
- Overall pipe condition
- Manhole condition
- Joint spacing
- Joint alignment
- Number of houses
- Remarks

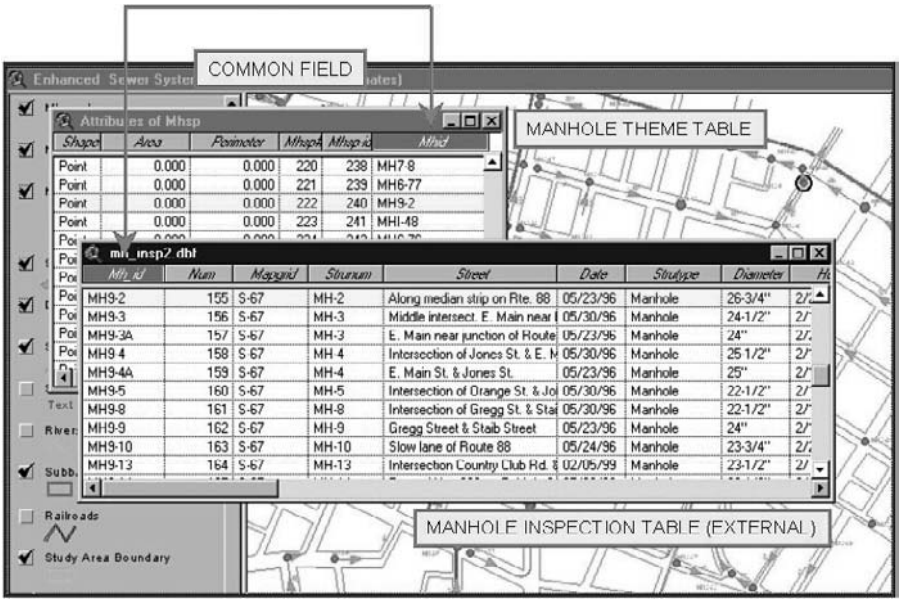


Figure 15.7 Linking manhole theme table and external manhole inspection table in ArcView.

Using the method described in the preceding text, the CCTV inspection logs can be integrated with the GIS database. This allows creation of thematic maps of sewer conditions. As described earlier, the successful linking of GIS features and inspection data requires that both the GIS database and the inspection data have a common ID.

Figure 15.8 Thematic map of manhole joint defects.

Figure 15.9 Thematic map of pipe defects.

Figure 15.9 shows a thematic map of pipe defects based on the "Remarks" attribute of the CCTV inspection logs, which contained information about sewer defects such as root growth, collapsed pipes, leaks, break-in service, etc.

WORK ORDER MANAGEMENT

Although this section focuses on wastewater system work orders, the methodology is equally applicable to water and stormwater systems.

Conventional work order management requires searching and printing the appropriate sewer system maps and manual completion of paper forms. GIS can be used to prepare work orders simply by clicking on the desired sewer or manhole in the GIS map. When the user clicks on an object, a work order inspection form is displayed that the user can fill by entering work information or by selecting predefined actions from drop-down lists. The completed work-order form and

corresponding map area can be printed for the inspection and maintenance crews. The entire process takes just a few minutes compared to the conventional method of locating and copying maps and typing the work order forms, which usually takes several hours.

GIS-based work-order management requires automated mapping/facilities management (AM/FM) software extensions that can be run from within the GIS software. These add-on programs, referred to as AM/FM/GIS software, are described in Chapter 14 (AM/FM/GIS Applications). AM/FM/GIS systems are especially useful for asset inventory, inspection and maintenance, and work management (Shamsi and Fletcher, 2000). The Cityworks (formerly Pipeworks, Azteca Systems, Sandy, Utah) AM/FM/GIS software suite includes ArcView and ArcGIS extensions, which help users integrate their GIS and work management. Its Wastewater Module can be used for inspection and maintenance of sewer systems and for managing and recording a CCTV inspection program for sewers. Cityworks Work Order modules are used to create and manage work orders for inspections and tests, and for detailed inventory for water and sewer systems. The Cityworks work management solutions are based on a GIS-centric approach. Cityworks does not treat the GIS as subordinate to the work management system. Instead of integrating the work management system with the GIS, it uses the GIS as a work management system (Azteca, 1999). Figure 15.10 shows the Cityworks work order management screen for CCTV inspection of sewers.

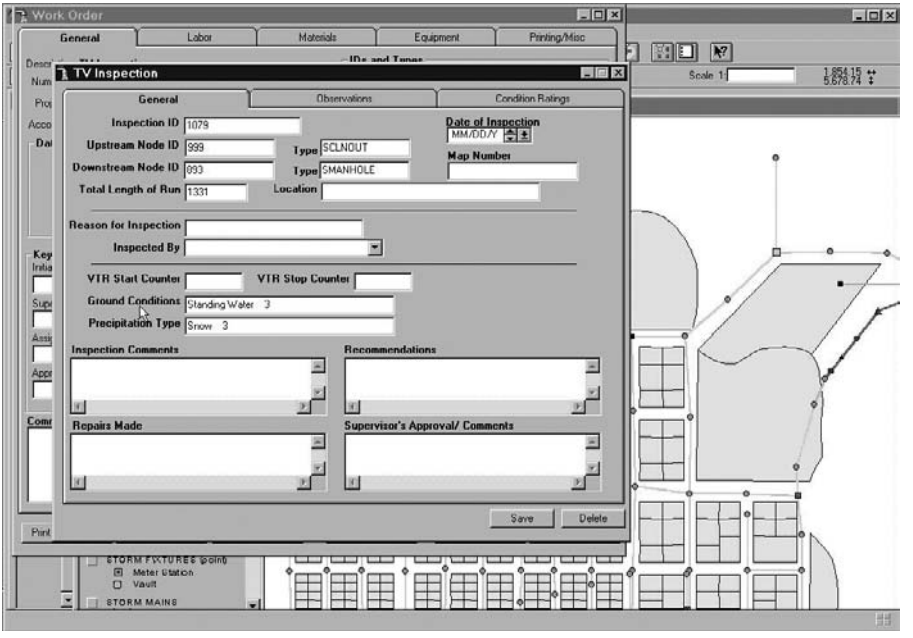


Figure 15.10 Cityworks work order management screen for CCTV inspections.

WATER MAIN ISOLATION TRACE

In many cases, the labor and material costs for fixing a broken water main are smaller than the hidden costs resulting from the pipe break. The hidden costs might be incurred due to human injury, property damage, litigation, customer inconvenience, traffic delays, police and emergency services, lost economic opportunities, and diminished fire-fighting capacity in other parts of the system. GIS offers many applications in the inspection and maintenance of water distribution systems. Water main isolation trace is one such application.

When a water main breaks, it should be isolated from the rest of the water distribution network for repairing the break or replacing the pipe. Water main isolation trace is defined as identifying those water distribution system valves that must be closed to repair or replace a broken water main. A water distribution system network layer that "knows" which links are connected to which nodes (i.e., has topology) can be used to traverse networks (Haestad et al., 2003) for solving the water main isolation problem. In addition to valve identification, GIS also allows an advance analysis of various water main shutdown scenarios, such as:

- Refining the valve search based on inoperable valves
- Identifying dry pipes (i.e., pipes without water)
- Identifying the affected customers served by dry pipes
- Identifying critical water service customers (e.g., hospitals), who need uninterrupted service
- Determining the extent of a service interruption to avoid unintentional shutdowns
- Capturing leak history data that are useful in pipeline replacement modeling

A GIS must have *topology* to benefit from the isolation trace application. Topology is defined as the spatial relationship between features. Spatial relationships between connecting or adjacent features, such as the pipes connected to a hydrant, which are so obvious to the human eye, must be explicitly defined to make the digitized maps "intelligent." For example, our eyes can instantly identify the pipe serving a house just by looking at a paper map. A GIS, on the other hand, must mathematically compare the pipe and parcel layers to complete this task. Topology is, therefore, defined as a mathematical procedure for explicitly defining spatial relationships between features. The subject of topology is so sophisticated that it is considered a branch of mathematics (Shamsi, 2002). The GIS layers that have topology can be used to trace transportation or utility networks. A water distribution system network layer that knows which links are connected to which nodes can be used to traverse networks (Haestad et al., 2003).

ESRI's ArcMap software provides many network analysis functions. ArcMap's standard interface can be customized to add network analysis functions such as network tracing. For example, a sample toolbar called the Utility Network Analyst (UNA) was developed by ESRI to help ArcGIS users develop custom applications (ESRI, 2002c). The UNA consists of a toolbar, a pull-down menu for saving and loading flags and barriers and setting options, a set of tools for specifying flags and barriers, and a combo box containing the list of available trace tasks. The UNA toolbar has a set of trace tasks (Find Connected, Trace Upstream, etc.) that can be used to solve

many problems. The Isolation Trace Task of the UNA toolbar takes a single specified edge flag and determines the valves that would need to be turned off to isolate the area. The UNA toolbar can be downloaded from http://arcobjectsonline.esri.com. The steps required to use the Isolation Trace Task are:

1. Register the trace task with Categories.exe as an ESRI Utility Network Task.
2. Close and restart ArcMap.
3. Add your network to ArcMap. Use the Add Edge Flag Tool on the Utility Network Analyst toolbar to place an edge flag at the location you wish to isolate.
4. Choose "Isolation Trace" from the list of available trace tasks and perform the trace. The result will be a display of the isolated area and a selection set of the valves to turn off.

CASE STUDIES

Representative case studies of GIS applications in maintenance are given in the following subsections.

Isolation Trace Case Studies

Figure 15.11 shows an isolation trace example from the PipeWorks example data. Pipeworks (now known as CityWorks) is an AM/FM/GIS software package from Azteca Systems (Sandy, Utah, www.azteca.com), that runs as an ArcView Extension. The trace is conducted by selecting the broken water main (midway on the pipe in the center of the map) and clicking on the "Isolation Trace" button. The top screenshot shows that four valves must be closed to isolate the broken water main for repair or replacement. Next, it was assumed that one of these four valves located at the bottom of the screen is "open inoperable," which would prevent this valve from being closed down. This scenario was modeled by changing the valve status from "open operable" to "open inoperable." The actual valve status can be made available in the GIS database if field inspections data are integrated with it. The bottom screenshot shows that due to the inoperable status of one of the valves, two additional (i.e., a total of six) valves must be closed.

In 2001, San Jose Water Company, which services the City of San Jose, California, created an ArcView 8.1 application called "Main Break" to accomplish the preceding functions using Visual Basic and ArcObjects running against a personal geodatabase. The application was installed at SJWC's main office and on seven laptop computers. The laptop computers empowered the field personnel with a mobile GIS that allowed on-site identification of the valves to be closed (Coates, 2003).

Sewer System Inspections in Washington County

A sewer system inspection GIS was implemented for a small community located in the Washington County of Pennsylvania with a population of 4900 people and a

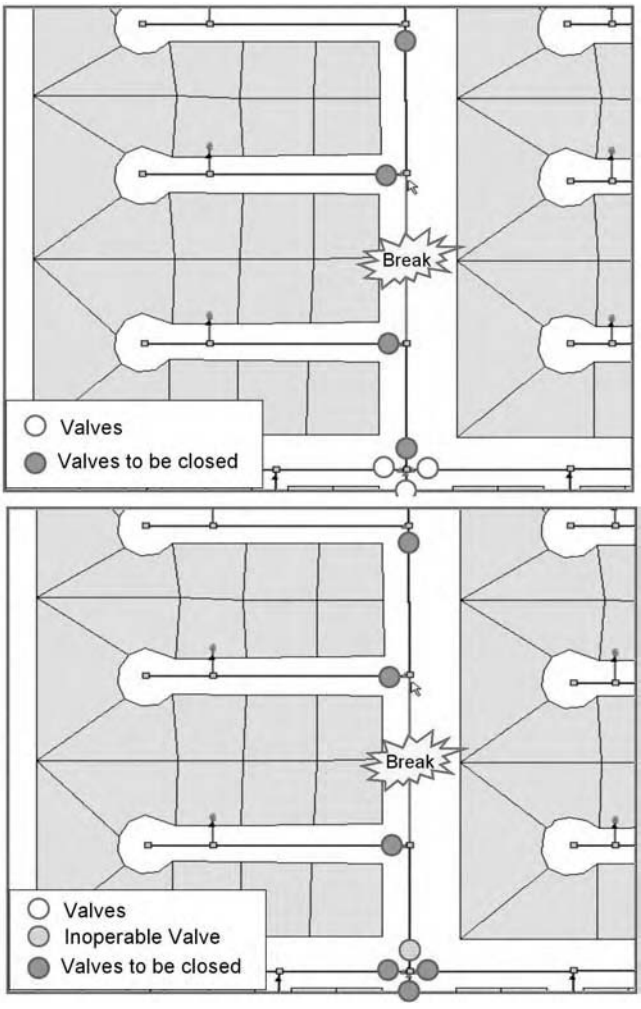


Figure 15.11 Isolation trace example. Above: all valves operable; below: one valve inoperable.

1.5 MGD (million gallons per day) sewage treatment plant. This GIS project consisted of the following tasks:

1. Develop manhole and sewer line coverages
2. Add manhole attributes from corrective action plan (CAP) manhole inspection reports
3. Add sewer attributes from CCTV inspection logs
4. Develop service laterals coverage from CCTV inspection logs
5. Develop video coverage from CCTV inspection tapes
6. Add street names to the existing roads coverage
7. Align image and feature themes
8. Package the information for delivery

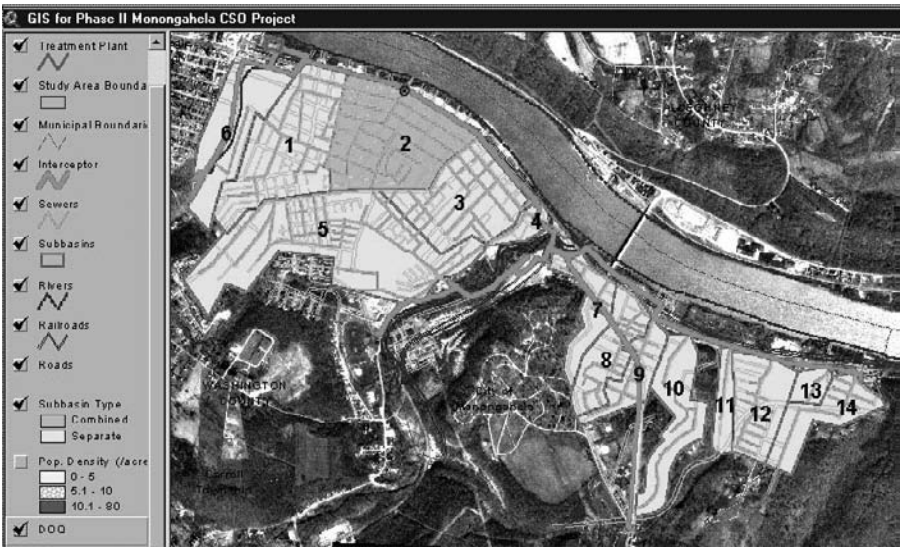


Figure 15.12 ArcView layers created from sewer system inspections data.

The project was completed in 2000 in 6 months for a cost of $15,000. Various free public-domain GIS data were used to keep the project cost low. These data included USGS digital orthophoto quarter quadrangles (DOQQ), digital raster graphics (DRG), and digital elevation models (DEM). Tasks 3, 4, and 5 were accomplished using the various techniques described in this chapter. Fifty mini movies were created from the existing VHS tapes and stored on five CD-ROMs. The movie files consisted of uncompressed AVI files of 40 to 60 MB size and resulted in approximately 3 GB of digital data.

Figure 15.12 shows various ArcView GIS themes of the project overlayed on the DOQQ image. The figure on the first page of this chapter shows an ArcView project screen displaying a mini movie of root growth for a user-specified location. Figure 15.8 and Figure 15.9 presented earlier show thematic maps of manhole and sewer defects, respectively.

Sewer Rehabilitation in Baldwin

Facing stringent regulatory requirements at the turn of the century, many communities in the Allegheny County (Pennsylvania) started undertaking comprehensive sewer rehabilitation projects to reduce systemwide inflow and infiltration. Rehabilitation activities included internal televising, pipe lining, pipe bursting, pipe replacement, manhole repair, and manhole replacement.

Baldwin Borough, a 5.5 mi^2 municipality located in the Greater Pittsburgh area with a population of 22,000, undertook a similar project. Baldwin's consultant,

(Chester Engineers, Pittsburgh, Pennsylvania) developed a sewer rehabilitation data management system (SRDMS) to assist in the planning and design phases of Baldwin's sewer rehabilitation projects. Initially, the system was successfully tested in the bidding phase for internal inspection of over 200,000 ft of sanitary sewers and 3,000 manholes. The system utilized digital map data compiled during a sewer-mapping effort initiated in 1996. The project team field-verified the sewer network data based on CCTV inspections conducted during a 3-month period in 2001. Concurrent with this effort, a Microsoft Access™ database was developed to capture conditional information for each surveyed sewer feature, apply unit costs for all necessary repair work, and generate a variety of reports for use in bid document submittals. Subsequent rehabilitation projects proved the effectiveness of SRDMS, leading to its implementation in several Allegheny County communities.

SRDMS is a GIS-based system because it reads and writes the information in a GIS database. It utilizes a relational database (e.g., Microsoft Access) to manage disparate data related to rehabilitation work, including manhole inspections, televised sewer line inspections, and flow data. This relational database is linked with the GIS database to facilitate mapping of the extent and severity of system damage and the recommended remedial activities. The required GIS layers can be created from a variety of sources, such as ArcInfo™ coverages, ArcView™ Shapefiles, and CAD files (e.g., Integraph™ or AutoCAD™). Preprocessing of CAD data to ensure the connectivity of sewer features may be necessary as the system requires that connectivity be established between all lines and manholes. SRDMS can use field inspection data from the conventional paper forms or the latest digital data collectors (e.g., PDAs, or tablet PCs). The GIS interface was initially developed as an ArcView 3.2 Extension. An ArcGIS version was created subsequently. SRDMS links inspection and repair information contained in the Access database to the corresponding GIS features. Applications of this system have resulted in a 20 to 30% reduction in project management labor effort over traditional methods. SRDMS provides the end user with the following functionalities:

- Provides the ability to visualize the overall condition of the sewage collection system using thematic maps of problematic areas.
- Identifies the precise locations of specific rehabilitation activities based upon an engineering interpretation of CCTV and manhole inspection results.
- Develops budgetary pricing for system repairs and rehabilitation work. It generates cost estimates for manhole rehabilitation and line repair work by applying multipliers to unit repair costs. Unit costs are entered by the project manager. The unit cost estimates may be modified at any time. The system will automatically update all aggregate costs based on the updated figures.
- Provides a series of reports detailing system damage, recommended remedial action, and cost estimates for the repair of damaged structures.
- Develops plans for distribution to prospective contractors during the bidding process.
- Provides a mechanism for tracking and reporting the progress of the rehabilitation activities.

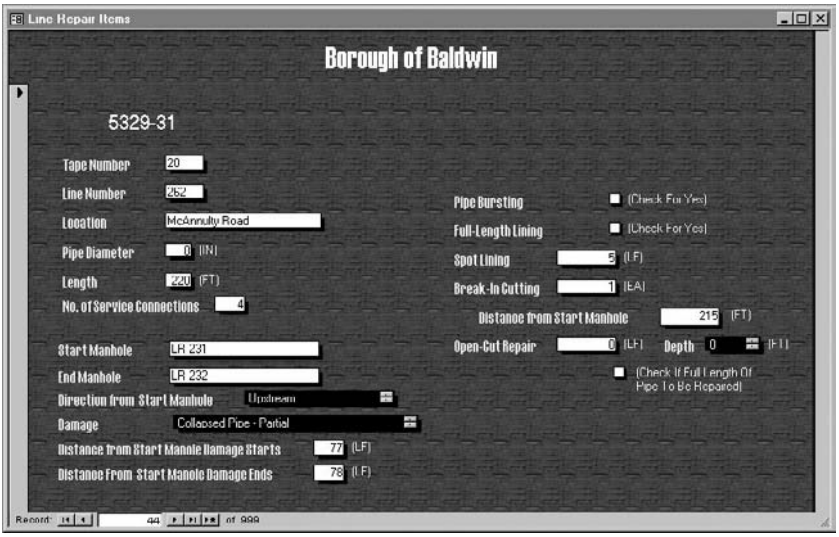


Figure 15.13 SRDMS line defect and repair entry form.

- Expands the content of the original GIS to include the location of service lateral connections identified during the televising process.
- Corrects inaccuracies in the original GIS based upon information obtained during the televising process.

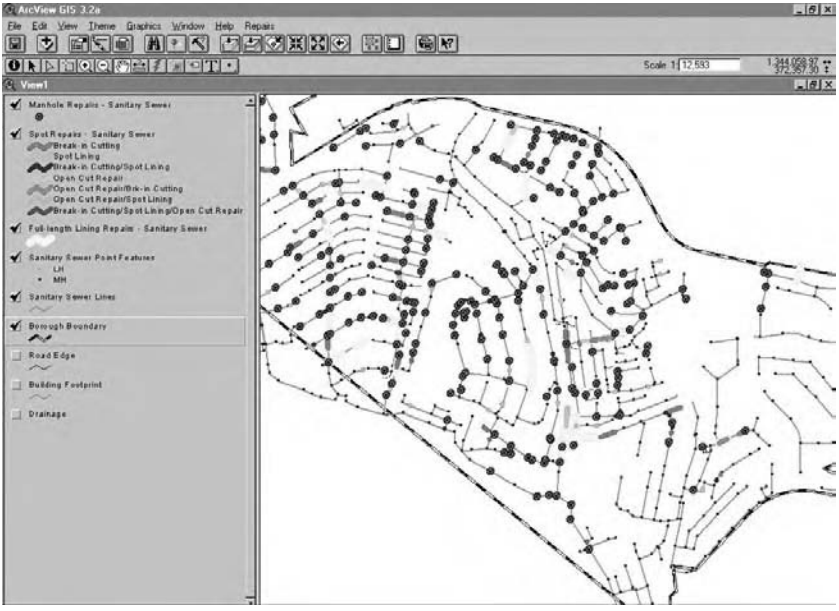


Figure 15.14 SRDMS line and manhole repair map.

Figure 15.13 shows the SRDMS line defect and repair entry form. The left side of this form receives input from the CCTV inspections database and automatically updates the GIS layer for pipes. The right side of the form, which recommends appropriate repairs, is completed by the project engineer or manager based on individual judgment and experience. Manhole and line repairs recommended by the user are added to the GIS as new layers. Figure 15.14 shows an ArcView screenshot of the line and manhole repair map. All the manholes targeted for repairs are clearly shown as large dots. All the sewers targeted for spot repairs or full-length lining repairs are shown in different colors.

USEFUL WEB SITES

Azteca Systems	www.azteca.com
Government Accounting Standards Board (GASB)	www.gasb.org
Pearpoint, Inc.	www.pearpoint.com
Peninsular Technologies, Inc.	www.pent.com
Red Hen Systems, Inc.	www.redhensystems.com
WinCan America, Inc.	www.wincanamerica.com

CHAPTER SUMMARY

This chapter provided an overview of GIS applications for the fourth *M* of the "4M approach" in the inspection and maintenance of water, wastewater, and stormwater systems. These applications include asset management and field inspections required by regulations such as GASB 34 and wet-weather overflow policies in the U.S. GIS also allows integrating work-order management and video inspections (e.g., CCTV inspections of sewers) with maps, which results in workflow efficiency improvement and cost savings for water and wastewater utilities. GIS can be used as a project management tool for system rehabilitation and repair work. Last but not least, GIS allows thematic mapping of inspection data to enable effective communication with decision makers and stakeholders. Real-world examples and case studies presented in this chapter show that GIS can be used for more efficient inspection and maintenance of water, wastewater, and stormwater systems.

CHAPTER QUESTIONS

1. What is the "infrastructure problem," and how can GIS help to solve this problem?
2. What is asset management, and how can GIS help in asset management of water, wastewater, and stormwater systems?
3. Describe two examples of GIS applications in regulatory compliance for water, wastewater, or stormwater systems.
4. What is water main isolation trace? How will you implement it on a mobile device?

CHAPTER 16

Security Planning and Vulnerability Assessment

Although water and wastewater systems are generally well prepared to deal with natural disasters and accidents, many utilities prior to September 11, 2001 had given little consideration to defending themselves against acts of terrorism. Read this chapter to see how GIS can help in safeguarding your water and sewer systems against the acts of terrorism.

Water distribution systems are especially vulnerable to acts of terror.

LEARNING OBJECTIVE

The learning objective of this chapter is to understand GIS applications in security planning and vulnerability assessment of water and sewer systems.

MAJOR TOPICS

- GIS applications in planning
- Security planning
- Vulnerability assessment
- Security modeling software

LIST OF CHAPTER ACRONYMS

AMSA Association of Metropolitan Sewerage Agencies
EPA (U.S.) Environmental Protection Agency
ERP Emergency Response Plan
OGC Open GIS Consortium
VSAT Vulnerability Self Assessment Tool

GIS APPLICATIONS IN PLANNING

GIS technologies have enabled planning professionals to make better and faster planning decisions. New software advances have revolutionized the preparation and use of master plans for water and sewer systems. GIS technologies provide various planning tools such as:

1. Remote-sensing data (described in Chapter 3)
2. Land-use/land-cover data (described in Chapter 3, Chapter 11, and Chapter 12)
3. Intelligent maps (described in Chapter 8 and Chapter 9)
4. Hydrologic and hydraulic models (described in Chapter 11 to Chapter 13)

With the help of case studies and examples, this chapter will focus on how to apply these tools for security planning purposes.

CITY OF ALBUQUERQUE'S WASTEWATER MASTER PLAN

In 1990, the City of Albuquerque, New Mexico, began a multiphase planning effort to evaluate its existing wastewater conveyance and treatment facilities, project future growth and flows, create programs for monitoring and rehabilitating existing infrastructure, and develop an organized approach to ensuring that future facilities keep pace with development. The City used GIS technology to develop a 40-year master plan for its wastewater collection and treatment system. The master plan was created by integrating the City's GIS data with a number of other computer applications including demographic modeling,

land-use mapping, hydraulic analysis, and infrastructure rehabilitation planning. The GIS approach enabled the City to create a master plan that could be easily updated in future as new development takes place (American City & County, 1991).

SECURITY PLANNING

Water, wastewater, and stormwater systems face both natural (floods, droughts, earthquakes, fires, etc.) and man-made (terrorism, vandalism, sabotage, arson, cyber attacks, etc.) threats. Although water industry utilities have learned from natural disasters, their experience with human-caused threats is small (Grigg, 2002). The basic form of protection against attacks or threats of attacks is a security system. Security planning methods include hazard assessment, vulnerability assessment, mitigation, response planning, and crisis communications (Burns et al., 2001). Vulnerability assessment is an important step in disaster preparedness. In the U.S., many utilities are performing mandatory vulnerability assessments.

When terrorists attacked the World Trade Center complex in New York and the Pentagon building in Washington, D.C., on September 11, 2001, the U.S. witnessed one of the worst days of its 225-year history. The unfortunate event also witnessed the best of the American spirit as people everywhere, including the GIS community, rushed to help. GIS supported response efforts at the World Trade Center and the Pentagon by coordinating the search for survivors and identifying hazardous areas. Planning for collection of airborne imagery over Ground Zero began within hours of the terrorist attack. Only 3 hours after the attack, France's SPOT 4 satellite captured 20-m resolution infrared images of the fires blazing in Manhattan. On September 12, Space Imaging's IKONOS satellite acquired 1-m imagery of both the World Trade Center area and the Pentagon (Barnes, 2001b). LIDAR* images collected on September 19, 2001, at 5000 ft were used to create 3D depictions of the terrain and buildings of the World Trade Center complex (Logan, 2002). The media, to cover the stories related to the terrorist attack, frequently used these 3D renderings.

Since September 11, 2001, concerns over intentional contamination have resulted in new federal legislation in the U.S. requiring community water systems to develop Emergency Response Plans (ERPs) (Haestad Methods, 2003a). In the Public Health Security and Bioterrorism Preparedness and Response Act of 2002, U.S. Congress recognized the need for drinking water system utilities to undertake a more comprehensive view of water safety and security. It amended the Safe Drinking Water Act and specified actions that community water system utilities and the U.S. Environmental Protection Agency (EPA) must take to improve the security of the nation's drinking water infrastructure. In 2002, EPA issued $90-million grant monies to help U.S. water and wastewater utilities assess their security needs. In January 2003, the U.S. Government launched the Department of Homeland Security for coordinating the work of America's security and intelligence agencies.

Although water and wastewater systems are generally well prepared to deal with natural disasters and accidents, many utilities — prior to September 11 — had given

* For additional information on LIDAR data, please refer to Chapter 3 (Remote-Sensing Applications).

little consideration to defending themselves against acts of terrorism. Immediately after the September 11 terrorist attacks, many water and wastewater utilities in the U.S. quickly adopted or expanded security measures (Landers, 2002).

Vulnerability of Water Systems

Drinking water utilities today find themselves facing new responsibilities. Although their mission has always been to deliver a dependable and safe supply of water to their customers, the challenges inherent in achieving that mission have expanded to include security and counterterrorism. Water distribution systems are especially vulnerable to acts of terror. Water systems can be contaminated by injection of poisonous substances such as cyanide. Water system operators, therefore, should be able to monitor contaminant levels and to predict contaminant movement throughout the system. These are complex tasks that require real-time monitoring and distribution-system water quality modeling. Real-time monitoring of every contaminant that could be deliberately introduced in a water system is not practical. However, a change in indicator parameters can signal the possibility of intentional contamination. The operators should be equipped with the most effective means of interpreting the monitoring data, communicating the results to decision makers and the general public, and implementing the control measures quickly (Schreppel, 2003). At the present time, most applications are focusing on the security analysis of drinking water distribution systems.

Vulnerability of Sewer Systems

Sewer systems can be used to contaminate receiving waters that are used for drinking-water supply intakes. Terrorists can pour volatile matter in the sewer system to create fire hazards. Large combined and storm sewers can also be used to gain access to restricted areas that have been blocked for security reasons. Although much of the attention regarding the security of water infrastructure has focused on the drinking water side, wastewater professionals have been quick to acknowledge and attempt to address their vulnerabilities. For example, the Metropolitan Sewer District (MSD) of Greater Cincinnati, Ohio, turned its attention to its sewage collection system, analyzing the likelihood that its sewers — particularly some of its large combined sewers — might be used as conduits in a terrorist attack. MSD used a GIS to compare maps of its collection system with the maps of Cincinnati to conclude that there were no major targets of opportunity in close proximity to any of its large sewers (Landers, 2002). Analysis of using a sewer system to attack a wastewater treatment plant or pumping station still requires some research.

GIS APPLICATIONS IN VULNERABILITY ASSESSMENT

Water and wastewater utilities must remain vigilant and develop comprehensive engineering and management measures to protect against future threats (Grigg, 2003). Implementing homeland security necessitates understanding the systems,

infrastructure, organizations, and vital interactions necessary for the well-being of communities. GIS is critical to this effort because it integrates all types of information and relates that information spatially. GIS can help in the areas of risk assessment, security planning, mitigation of the effects of attacks, preparedness activities, response measures, and recovery efforts.

GIS applications for community protection include modeling chemical releases in water and air, tracking and accessing information about hazardous storage sites, and managing hospital information. GIS applications can also reveal additional problems caused by proximity to hazards such as chemical storage. The modeling capabilities of GIS can simulate the effects of many types of attacks on critical resources, infrastructure, and populations. After identifying the hazards, assessing the risks, and prioritizing the values (i.e., assets of greatest value), both strategic and tactical plans can be formulated. GIS enhances response to acts of terror by helping managers to quickly assess the extent of damage and by speeding the decision-making process (ESRI, 2002d).

There is no one-size-fits-all security approach that will work for all water or sewer systems. Because each utility's situation is unique, security measures must be tailored to each system. In general, however, there are two basic requirements for the vulnerability assessment of water and sewer systems:

1. Network connectivity: This requirement pertains to what system components are connected to what and what is upstream (or downstream) of what. This capability would determine, for example, the area that would be impacted by a contaminated water storage tank. A GIS must have topology information to determine network connectivity. Network topology is inherent in many GIS packages and can easily provide this information. Network-tracing (described in Chapter 9 [Mapping Applications]) and isolation-tracing (described in Chapter 12 [Water Models]) functions are good examples of the network topological capabilities of GIS.
2. Contamination analysis: This requirement pertains to the strength and extent of the contaminants. GIS alone cannot provide this information. Hydraulic models capable of simulating both the quantity and the quality of water must be used in conjunction with a GIS.

GIS and hydraulic modeling can be used for infrastructure protection, vulnerability assessment, consequence management, and security planning to safeguard water, wastewater, and stormwater systems against potential sabotage activities and terrorist attacks. Lindley and Buchberger (2002) outlined a holistic approach that combined GIS and hydraulic modeling to integrate multiple risk factors so as to identify locations that may be vulnerable to contaminant intrusions. Intrusions were defined as the introduction (either accidental or deliberate) of an undesirable agent into the potable water distribution system. An intrusion pathway was defined as the connectivity route between the potable supply and the contaminant source. EPANET* software was used for hydraulic modeling. GIS was used for layering of the locations having adverse pressure, intrusion pathway, and source concerns.

* Additional information about EPANET is provided in Chapter 12 (Water Models).

SECURITY MODELING SOFTWARE

Sample security modeling software products are described in the following subsections.

H_2OMAP^{TM} Protector

H_2OMAP Protector from MWH Soft, Inc. (Pasadena, California), is an add-on module for the H_2OMAP suite* that can be used for water security planning, infrastructure protection, and vulnerability assessment. Designed with the latest geospatial modeling technology, Protector uses a geodatabase for modeling various security scenarios. It can be used for estimating the consequences of a terrorist attack or a crisis event on a drinking-water supply infrastructure as well as formulating and evaluating sound emergency response, recovery, remediation and operations plans, and security upgrades. The program can be used to identify viable solutions before an incident or disaster occurs, or to assist in responding should it occur.

Protector allows users to model the propagation and concentration of naturally disseminated, accidentally released, or intentionally introduced contaminants and chemical constituents throughout water distribution systems; assess the effects of water treatment on the contaminant; and evaluate the potential impact of unforeseen facility breakdown (e.g., significant structural damage or operational disruption or both). It enables users to locate areas affected by contamination, calculate population at risk and report customer notification information, and identify the appropriate valves to close to isolate a contamination event. Finally, it helps users track contaminants to the originating supply sources, compute required purging water volume, develop efficient flushing strategies, determine the resulting impact on fire-fighting capabilities, and prepare data for eventual prosecution.

Figure 16.1 shows a Protector screenshot showing isolation analysis. It depicts affected and closed facilities color-coded according to their vulnerability (risk factors).

WaterSAFE™

WaterSAFE from Haestad Methods, Inc. (Waterbury, Connecticut), is an add-on component for WaterCAD and WaterGEMS** software products. WaterSAFE is designed to manage and safeguard water distribution systems. It is a water system security and emergency planning tool specifically created to study infrastructure vulnerability to terrorist attacks and natural events.

Harnessing the powers of ArcGIS, WaterSAFE enables water utilities to analyze the movement of multiple constituents and track multiple sources for a given period of time. With this information, water utilities can assess the impact in real time and swiftly relay the results to customers in contaminated areas to take the necessary precautions. These water quality analysis features can enable utilities to better respond, strategize, and implement a fast and effective recovery plan in the event of

* Additional information about H_2OMAP suit is provided in Chapter 12 (Water Models)
** Additional information about WaterCAD and WaterGEMS is provided in Chapter 12 (Water Models).

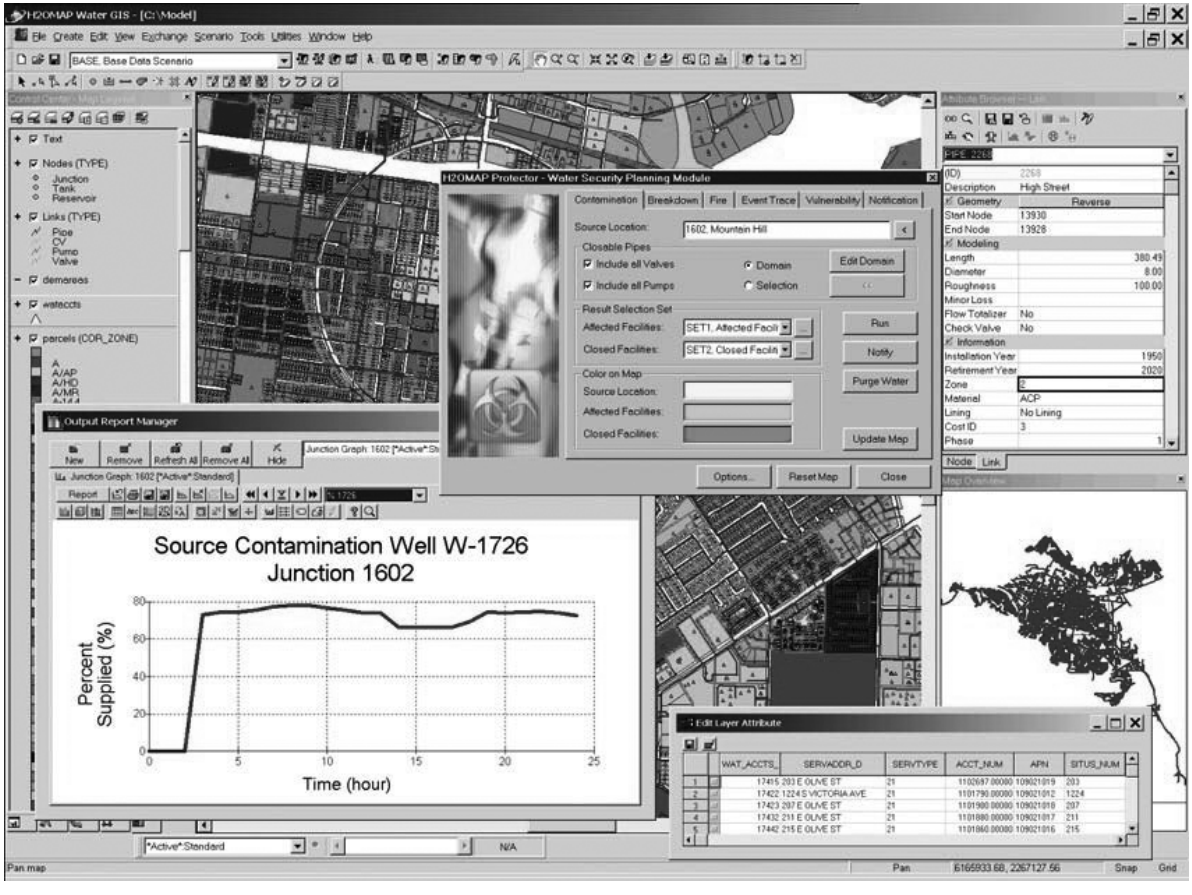


Figure 16.1 Protector screenshot showing isolation analysis.

a contamination emergency. WaterSAFE's extended water quality analysis features can also assist in conducting statistical analysis of water quality and evaluating contamination effects on a water treatment plant.

In 2003, a WaterCAD model was used to prepare an example ERP. Two scenarios were modeled: a rehearsed drill before a terrorist attack and a real-time solution. It was assumed that a contaminant was introduced into the tank of a typical midsized system and that complete mixing occurred in this tank. The vulnerability of the water system was assessed by monitoring the effects of injection location, the nature of the contaminant, the amount of contaminant added, and the period over which the contaminant was added. The modelers were able to reduce the spread of the contaminant, monitor the impact of flushing on the environment, and more quickly bring the system back to normal operation (Haestad Methods, 2003a).

VSAT™

Association of Metropolitan Sewerage Agency (AMSA) Washington, D.C., in cooperation with EPA, released a new water and wastewater system security analysis software in 2002. Known as Vulnerability Self Assessment Tool (VSAT), the software is available free of charge. Three versions, VSATWater, VSATWastewater, and VSATWater/Wastewater, are available for water, wastewater, and both water and wastewater systems, respectively. The VSAT products provide a structured, cost-effective approach for utilities to assess their vulnerabilities and to establish a risk-based approach to taking desired actions. The software allows utilities to assess the vulnerability of the complete range of utility assets including people (utility staff), physical plant, knowledge base, information-technology platform, and customers.

VSAT does not use GIS or hydraulic modeling. However, the information obtained from GIS and modeling can be used to run the VSAT program. For example, the first VSAT task is asset categorization and identification, which requires utility managers to conduct an inventory of utility assets. A GIS can be used to conduct an asset inventory for input to the VSAT program.

SECURITY PLANNING DATA ISSUES

The number of GIS data formats has increased exponentially with the growth in the GIS industry (Goodchild, 2002). According to some estimates, there might be more than 80 proprietary geographic data formats (Lowe, 2002b). Why are there so many geographic data formats? One reason is that a single format is not appropriate for all applications. For example, a single format cannot support both fast rendering in a command and control system and sophisticated topological analysis in a natural resource information system. Different data formats have evolved in response to diverse user requirements.

Users must possess considerable expertise to overlay, combine, or analyze different map layers or images. Converting from one format or type of data to another is cumbersome, time consuming, and error prone. Therefore, different GIS data formats or models used by different government organizations can lead to compatibility

problems that can hamper data sharing needs during an emergency situation such as a terrorist attack. This can be avoided by adopting compatible data models and interoperable data formats throughout a city or county. Efforts are currently underway in the GIS industry to standardize data formats and database-management systems by promoting open platforms, data, and database management systems. For example, the Open GIS Consortium (OGC) has been created with the vision of "the complete integration of geospatial data and geoprocessing resources into mainstream computing." OGC was formed in 1994 to facilitate access to and geoprocessing of data held in systems or networks. It defines an open GIS as "open and interoperable geoprocessing" or "the ability to share heterogeneous geodata and geoprocessing resources transparently in a networked environment."

USEFUL WEB SITES

Association of Metropolitan Sewerage Agencies (AMSA)	www.amsa-cleanwater.org
Haestad Methods	www.haestad.com
MWH Soft	www.mwhsoft.com
Open GIS Consortium	www.opengis.org
Vulnerability Self Assessment Tool (VSAT)	www.vsatusers.net

CHAPTER SUMMARY

Water, wastewater, and stormwater systems can be vulnerable to acts of terrorism and sabotage. This chapter shows that GIS offers many useful applications in security planning and vulnerability assessment of these systems. GIS maps and hydraulic models can be used as effective security planning tools to evaluate the system vulnerability to natural and man-made disasters, and recommend solutions. For example, GIS and hydraulic models can be used to determine the areas of a water distribution system that can be affected by a contaminated water supply source. At the present time, most applications are focusing on the security analysis of drinking water distribution systems.

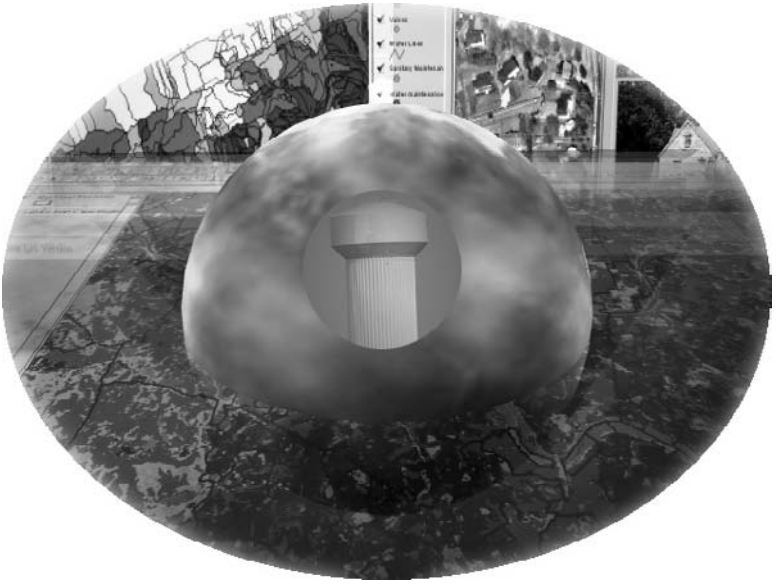
CHAPTER QUESTIONS

1. How are water and sewer systems vulnerable to acts of terrorism?
2. What can be done to protect water and sewer systems from acts of terrorism?
3. How can GIS applications help in security planning and vulnerability assessment of water and sewer systems?

CHAPTER 17

Applications Sampler

The use of GIS applications is growing throughout the world. This chapter shows how people around the world are applying GIS in their water, wastewater, and stormwater projects.



This chapter presents the latest examples of GIS applications in the water industry.

LEARNING OBJECTIVE

The learning objective of this chapter is to document GIS application projects around the world for water, wastewater, and stormwater systems.

MAJOR TOPICS

- Water system modeling
- Sewer system modeling
- Collection system rehabilitation and asset management
- Resource planning and capital improvement project (CIP) allocation
- Water quality management
- Water master planning

LIST OF CHAPTER ACRONYMS

BMP Best Management Practices
CCTV Closed-Circuit Television
CIP Capital Improvement Project
CIS Customer Information System
CSO Combined Sewer Overflow
DEM Digital Elevation Model
SCADA Supervisory Control and Data Acquisition
SSO Sanitary Sewer Overflow
SWAT Soil and Water Assessment Tool
TAZ Traffic Analysis Zone

This chapter presents a collection of recent case studies on GIS applications for water, wastewater, and stormwater systems. These case studies were written specially for publication in this book by 18 GIS and water professionals from 6 countries (Belgium, Bulgaria, Czech Republic, Denmark, Spain, and the U.S.). For the names and organizational affiliations of the case studies' authors, please see the Acknowledgments section. The case studies were submitted in response to the author's "Call for Case Studies" distributed to various Internet discussion forums.

DRAINAGE AREA PLANNING IN SOFIA

Application	Sewer system modeling
Author(s)	Milan Suchanek, Tomas Metelka
Project status	Completed in March 2003
Hardware	1000 MHz Pentium III personal computers
GIS software	Arc View 3.3
Other software	MOUSE 2002; MOUSE GM; MOUSE Gandalf; MOUSE LTS; AquaBase
GIS data	Subcatchment polygons, conduit lines, node points, CSO points, land use, population, property boundaries, buildings
Study area	City of Sofia, the capitol capital of Bulgaria, watershed area of about 400 km^2
Project duration	November 2001 to March 2003
Project budget	$460,000
Organization	Sofiyska Voda A.D.

The city of Sofia's sewerage system is a very complex combined drainage system dating back to early 1900. The drainage area is made up of about 400 km^2 of watershed area, draining wastewater from 1.2 million people. At the same time, the hydrological conditions of the city, which is located on the foothills of the Vitosha mountains, promote heavy rainfall, mainly during late spring.

By the end of March 2003, the first simulation model for the sewer drainage system was developed in Bulgaria. The skeletal planning model was built for the city of Sofia, having about 5000 manholes, 5000 pipes, 3000 catchments, and 140 CSO or diversion chambers. The study area was subdivided into seven main subcatchments and their main trunk sewers. The respective submodels were built up and calibrated based on 6 weeks of flow survey at some 75 flow and 25 raingauge sites. Finally all seven submodels were merged into one combined model covering the whole drainage area.

The local GIS system, supported by comprehensive manhole and ancillary surveys, supplied the model with the system structural data. The overall data migration process was an important task in the project execution.

The Sofia drainage area planning project brings a new experience to the general view of project management. The management project was based on British standards, and the WaPuG *Code of Practice* handbook was followed. At the same time, Danish technology was applied, along with Czech and Bulgarian know-how. Figure 17.1 shows a screenshot of the Sofia model in MOUSE (Suchanek and Metelka, 2004).

PIPE RATING PROGRAM IN BUNCOMBE COUNTY

Application	Collection system rehabilitation and asset management
Authors	Ed Bradford, Roger Watson, Eric Mann, Jenny Konwinski
Project status	Implemented in 2004 and is being used to generate rehabilitation projects
Hardware	Standard desktop PCs (connected to network)
GIS software	ArcGIS 8.x
Other software	Microsoft Access XP, Windows Media Player, Microsoft Excel, Microsoft Word, etc., as necessary to generate reports
GIS data	Sewer line polylines, sewer structure points, 6-in., 1-ft, and 2-ft resolution DOQs, county parcel polygons, drainage basin polygons, digital video hyperlinks, CCTV inspection and defect-coded Microsoft Access tables, river and stream polylines, and road centerline polylines
Study area	Entire MSD service area (180 mi^2, 920 mi of sewer)
Project duration	Began in 2001. No end date scheduled
Project budget	No defined budget *per se* for the program itself. Construction projects have generated $700,000 in actual cost so far, which could increase in future
Organization	Metropolitan Sewerage District (MSD) of Buncombe County, North Carolina
Web site	www.msdbc.org

MSD's Pipe Rating Program is a new method of generating and prioritizing sewer rehabilitation projects. The typical approach of *reactive* planning is to define, develop, and complete a rehabilitation project after problems such as sanitary sewer overflows (SSO) or structural failures occur (Bradford et al., 2004a). Pipe Rating is a *proactive* tool that utilizes CCTV information, a GIS database, and real-world maintenance history to view, score, and rate pipe segments based on a number of

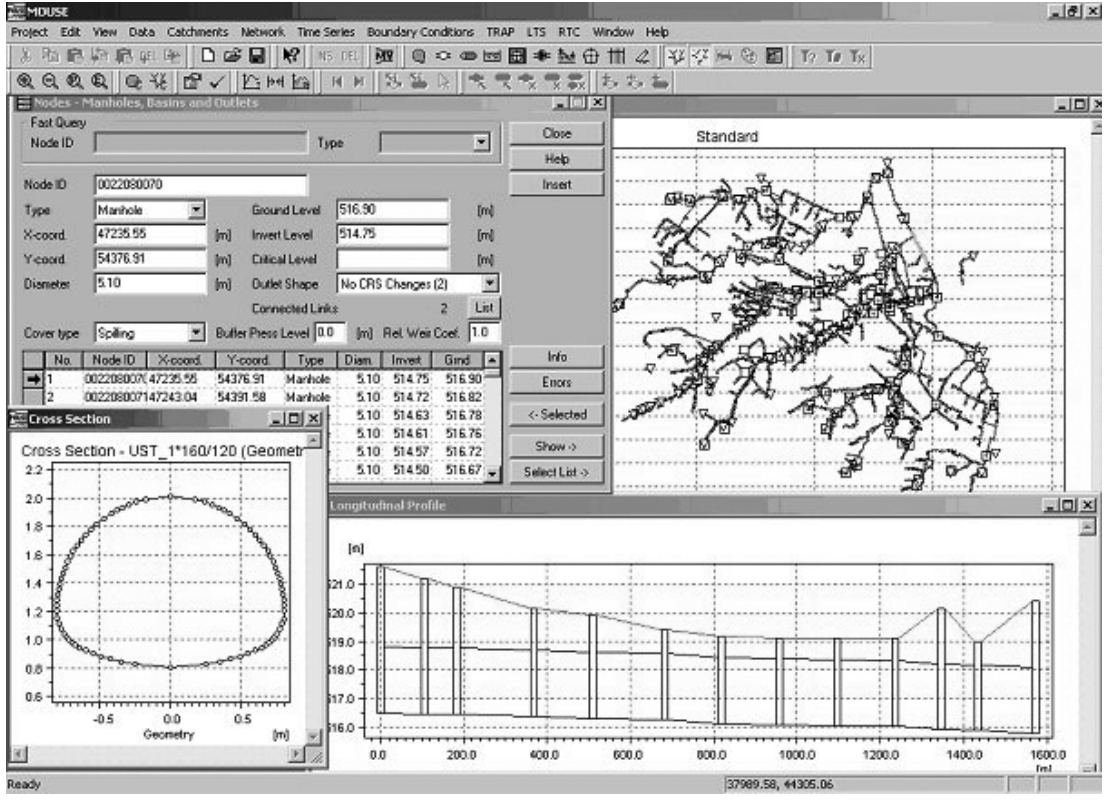


Figure 17.1 Screenshot of Sofia's MOUSE sewer system model.

factors. These factors, for any given manhole-to-manhole segment, include the number and severity of structural defects and the history of overflows on that segment. These are all combined to yield a rating, which may then flag a particular line segment for further investigation.

The data are gathered from a number of sources and incorporated into this program, which runs on the standard ArcGIS platform. CCTV video is captured by cameras traversing the pipes, and recorded on VCR tapes or in digital format (CD or DVD). First, each video is linked to its respective pipe segment within the District's GIS system so that it is immediately available to engineers and field responders. Second, a database is also created from the field data to record various features about the pipe, such as defect and structural information. This is accomplished by assigning to each defect a value pursuant to the standardized defect-rating manual developed by the District for this purpose. CCTV data are collected and embedded in a standard Microsoft Access database. When the technician identifies a defect along the segment, it is keyed in with the corresponding defect code. Each defect, and its corresponding severity score, is assigned to its pipe segment and used for future analysis performed in the GIS. Table 17.1 shows the defect structural scores used by MSD. These scores are based on MSD's standardized Sewer Condition Classification Manual. Scores are weighted according to MSD priorities.

Finally, all the information is used to calculate three pipe scores for each pipe: mean pipeline, mean defect, and peak defect scores. These scores allow users to visualize the severity in three categories for prioritization within the rehabilitation program. Once this has been accomplished, engineers can retrieve the videos from the GIS for confirmation and to evaluate them for rehabilitation procedures. Figure 15.1 shows an ArcGIS screenshot of mean pipeline ratings. Figure 17.2 and Figure 17.3

Table 17.1 Sewer Pipe Structural Defect Scores

Code	Description	Score
CC	Circumferential crack	20
CL	Longitudinal crack	20
CM	Multiple cracks	30
COH	Corrosion heavy	50
COL	Corrosion light	20
COM	Corrosion medium	30
D	Deformed sewer	65
FC	Circumferential fracture	40
FL	Longitudinal fracture	40
FM	Multiple fractures	60
HL	Hole large	50
HP	Hole patched	5
HS	Hole small	25
JDL	Joint displaced large	30
JDM	Joint displaced medium	5
OJL	Open joint large	30
OJM	Open joint medium	10
X	Collapsed pipe	100

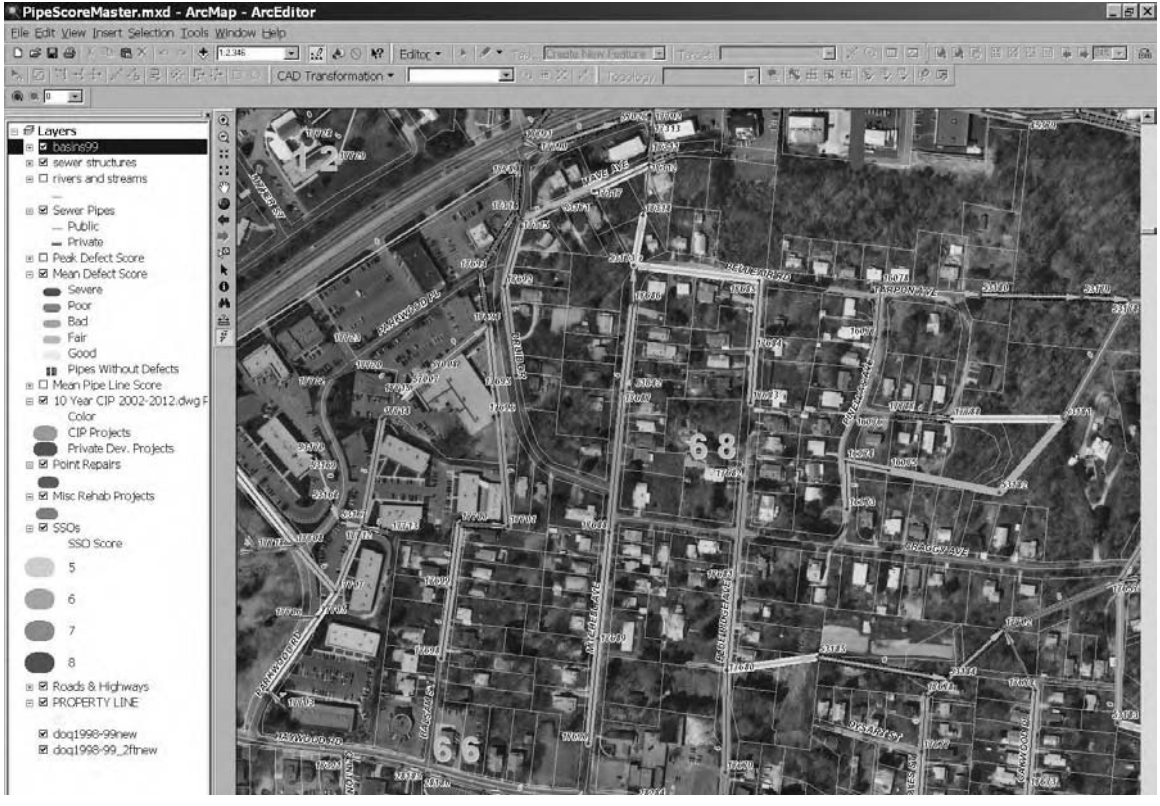


Figure 17.2 Map of mean defect ratings for sewer pipes.

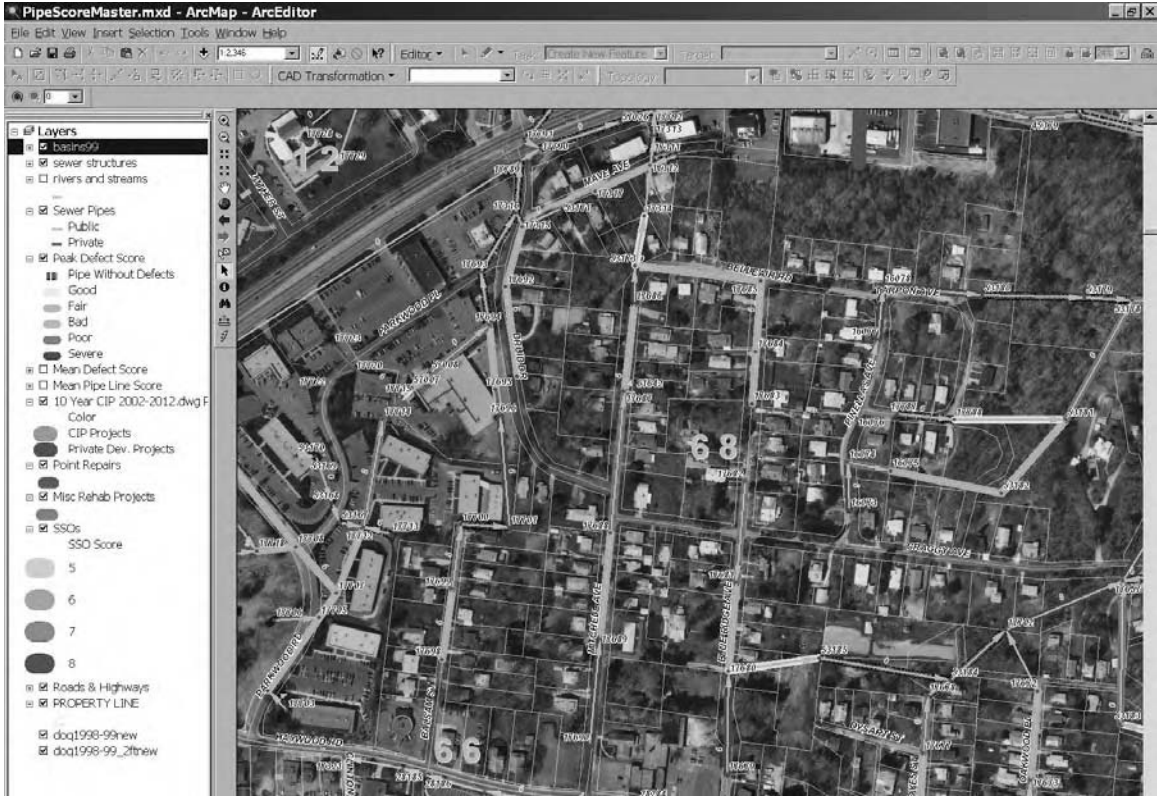


Figure 17.3 Map of peak defect ratings for sewer pipes.

show maps of mean defect and peak defect pipe ratings, respectively (Bradford et al., 2004a).

WATER SYSTEM MODELING IN TUCSON

Application	Resource planning and capital improvement project (CIP) allocation
Author	Dean Trammel
Project status	GIS and hydraulic modeling effort completed in 2003
Hardware	Dell personal computers (1495 MHz, 384 MB RAM) with Windows 2000 Professional
GIS software	ArcMap and ArcInfo
Other software	Haestad Methods' WaterGEMS, WaterCAD for AutoCAD, Microsoft Access, Microsoft Excel
GIS data	U.S. census–disaggregated population projection data in the form of polygon Shapefiles, with associated population projections for each year from 2000 through 2050
Study area	Tucson, Arizona
Project duration	Resource planning effort: March 2002 to early 2004, GIS and hydraulic modeling: about 9 months
Project budget	CIP-budgeted planning effort, part of regular business
Organization	Tucson Water, Tucson, Arizona

Tucson Water's potable system supplies over 180,000 services, maintains over 4,000 mi of pipe, and installs over 500 new meters per month. Tucson Water also serves reclaimed water to over 550 reclaimed services, maintains over 100 mi of reclaimed pipe, and installs between 75 and 100 new reclaimed meters per year. In addition to the GIS data listed above, the following data were also used:

- The Pima Association of Government (PAG) took U.S. census tracts for the year 2030 and broke them up further into smaller disaggregated population-projection polygons, or traffic analysis zones (TAZ).
- Engineers, planners, and administrators made up a Resource Planning Committee that determined a boundary for the water service provided by Tucson Water, which was a modified version of a boundary used for previous planning efforts.
- Polygon Shapefiles that describe pressure zones and water services areas (WSA) for the entire service area, based on a parcel base of Tucson Water's customers.
- Ten hydraulic models of the central water system called local area models and five smaller models that represent isolated systems outside the central system.
- Shapefiles of the proposed master plans. These files were used as a backdrop to compare population projections with prospective new developments planned for the near future, and to make sure that these areas were correctly represented in the 2030 and 2050 model.
- SCADA, billing, production records.
- Cost estimates.
- Hydrology supply scenarios.

Traffic and census tract population projections were analyzed using ArcGIS to estimate future potable water demands. The estimated demands were applied to Tucson Water's current hydraulic model using Haestad Methods' WaterGEMS software. The model was then used to determine what facility and piping modifications would be

required to meet growth within current service areas and future development in outlying areas. Using capital and energy cost estimation tools within the hydraulic modeling software, different planning scenarios can be compared as to their overall costs.

First, the WaterGEMS skeletonization tool, Skelebrator, was used to reduce the existing local area models down to a point where they could be combined into a complete system model. Second, within WaterCAD for AutoCAD, the import/export submodel tool was used to combine the ten local area models and the isolated system models into a complete water system model. Third, the new model was balanced and checked for calibration to peak-day conditions. Fourth, demands were allocated to the model using processed data to predict demands for years 2030 and 2050. The WaterGEMS demand allocation tool, LoadBuilder, was used to distribute this demand correctly to the model. Lastly, scenarios for new additions to the system were made using WaterGEMS and WaterCAD for AutoCAD, and the functionality, capital cost, energy costs, and operations and maintenance costs associated with these scenarios were evaluated using built-in tools in the hydraulic modeling software. Figure 17.4 shows a screenshot of WaterGEMS model layers in ArcMap (Trammel, 2004).

WATER SYSTEM MODELING IN THE CITY OF TRUTH OR CONSEQUENCES

Application	Development of a water distribution model in MIKE NET from GIS data
Authors	Eric Fontenot
Hardware	Standard Desktop PC, 128 MB RAM, 850 MHz Pentium III processor
GIS software	ArcView 3.2
Other software	AutoCAD, MIKE NET, Microsoft Access
GIS data	Water distribution pipes, network demands
Study area	Truth or Consequences, New Mexico
Project duration	4 months
Project budget	$42,000
Organization	DHI, Inc., Newton, Pennsylvania, and Hørsholm, Denmark
Web site	www.dhigroup.com

The City of Truth or Consequences' water supply system was strained due to population increase and real estate development. The City had concerns about the system's ability to deliver fire suppression flow to certain parts of the city at peak demand. The City, therefore, engaged in strategic planning to evaluate possible improvements.

A model of the existing water distribution system was created by DHI, Inc., from several data sources. The detailed, topologically oriented model of the water distribution network was created primarily from GIS data and updated from hardcopy maps of the water system. Water demands were determined and distributed, based on production and billing information contained in a customer information system (CIS).

The comprehensive modeling evaluated the current system, as well as those alternatives identified in a preliminary engineering report. The placement of two

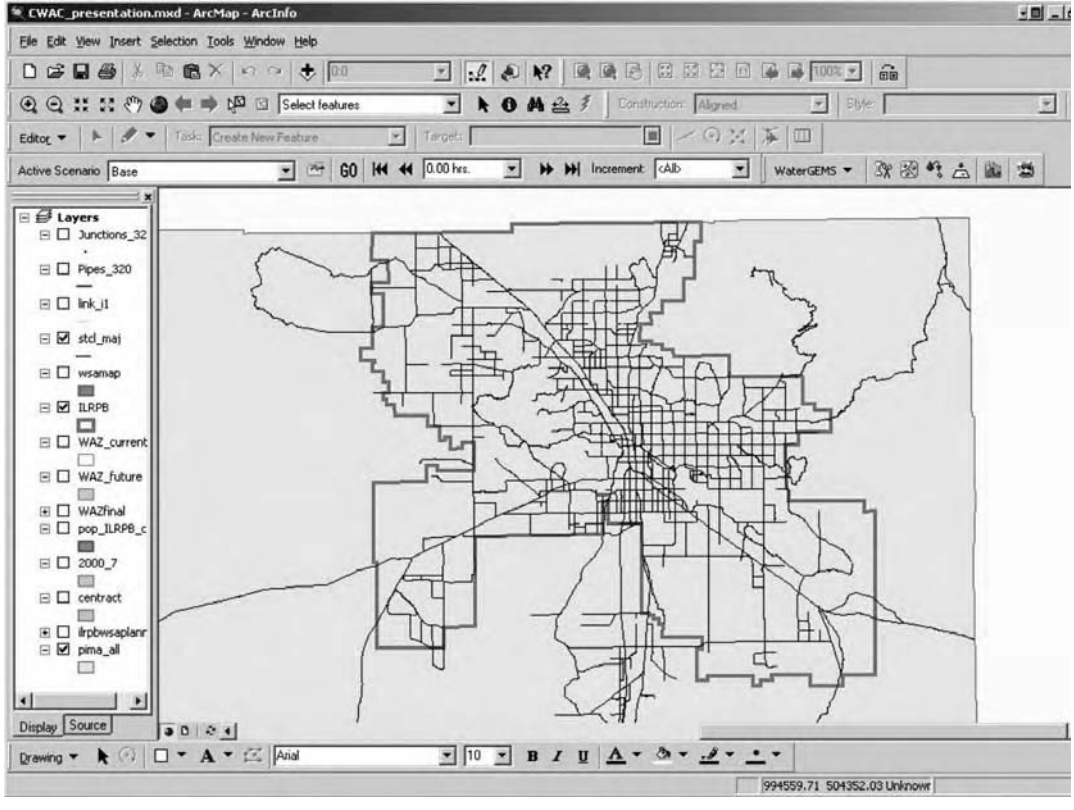


Figure 17.4 Screenshot of WaterGEMS model layers in ArcMap.

new storage tanks was evaluated, and recommendations were determined based on model results.

Background

Truth or Consequences is located east of Interstate 25 between Albuquerque and Las Cruces, adjacent to Elephant Butte Lake. The water system serves a combined population of approximately 7300 in Truth or Consequences and Williamsburg. There were 2820 residential and commercial connections, 268 of which were located in Williamsburg. The water system had 5 groundwater wells that fed a 200,000-gal clearwell, a 1,000,000-gal tank serving a low-pressure zone, a 3,000,000-gal tank serving a high-pressure zone, and 2 booster pump stations. Disinfection via chlorination was provided prior to the clearwell.

The five groundwater wells, ranging in age from 2 to 50 years, pumped into a 200,000-gal tank. Water was pumped from the 200,000-gal tank to the 1,000,000-gal storage tank, which supplied the low-pressure zone. Some of the piping in the low-pressure zone was quite old, dating back to the 1940s when the water system was originally constructed. Around 1970, the water system was divided into high- and low-pressure zones with the construction of the 3,000,000-gal tank. Water from the low-pressure zone was pumped into a storage tank to supply the high-pressure zone.

Due to increased demands and an aging system, the City was planning on constructing two additional tanks, one in each pressure zone. The current system did not maintain adequate pressure at peak demand, and the City had some concerns about the system's ability to deliver fire suppression flow to certain parts of the city. The 1,000,000-gal tank needed to be drained and rehabilitated, which could be done only after the construction of additional storage. The 3,000,000-gal tank also needed maintenance. Because it was the only tank in the high-pressure zone, an additional tank or booster station capable of providing adequate fire suppression flow was required to permit the draining and painting of the tank.

Building the MIKE NET Model from Various Data Sources

The City provided GIS data for the water distribution system, historical SCADA information for water tanks and booster pumps, billing information from the City's CIS, fire suppression flow data, and a hard copy of the water distribution system map. The GIS information included the network geometry, pipe diameter, and pipe material.

The GIS information supplied by the City was a mixture of AutoCAD files (DWG extension) and ArcView Shapefiles (SHP extension). The pipe locations and diameters were all contained in one AutoCAD drawing file. It was determined that modeling pipes 4 in. and larger would be adequate to represent the water distribution system and determine the locations of the problem areas in the system, and evaluate the proposed alternatives.

The drawing file was separated by layer and exported to the DXF format, so that it could be imported to MIKE NET. Each exported layer contained network

geometry as well as pipe material and pipe diameter information. The following nine DXF files were created and imported into separate MIKE NET project files:

- 4-, 8-, and 12-in. PVC pipes
- 4-, 8-, and 12-in. AC pipes
- 4-, 8-, and 12 in. CI pipes

These files were checked for connectivity errors and merged together to form the City's water distribution network in MIKE NET. Figure 17.5 shows the network model built from GIS data in MIKE NET.

The GIS data, water system drawings, fire suppression flow data, billing data, and SCADA data were used for preliminary model building. The GIS data were used to build the network topology and connectivity. The GIS data was found to be about 85% accurate. Errors in the GIS data included incorrect network connectivity, missing pipes, omissions in separation between the high- and low-pressure zones, incorrect pipe material labels, and incorrect pipe diameters. The network connectivity was checked against the water system drawings and updated to match the drawings in which differences occurred.

Street names, from an ArcView Shapefile, were geocoded to each pipe in the model. The billing data supplied by the city were summed by street name, and the demand for that street was distributed to those pipes in MIKE NET bearing that street name.

Elevation data were supplied in 2-ft contour intervals in the City's GIS data. These data were extracted to an X, Y, Z triplet text file using a DHI-developed ArcView extension, and the node elevations were assigned by using the Generate Node Elevation tool in MIKE NET.

Tank and pump data were added manually from the water system drawings and information obtained during interviews. Upon completion of the model, these data were exported from MIKE NET to ArcView Shapefiles and supplied to the City as an updated GIS coverage of their water distribution assets (Fontenot, 2004).

ARCGIS AND ARCFM INTEGRATION IN BELGIUM

Application	ArcGIS/ArcFM 8 for water distribution network management
Authors	Reynaert et al. (2003)
Project status	In use since January 2002
Hardware	Citrix application servers & standard Dell desktop PCs
GIS software	ArcInfo/ArcView 8 , ArcSDE 8, ArcIMS 4, ArcFM 8
Other software	SQL Server 2000, Citrix Metaframe XP
GIS data	Water distribution network data and topographic data
Study area	Antwerp province, Belgium
Project duration	1998 to 2002
Organization	Pidpa
Web site	www.pidpa.be

Pidpa is a European company that produces and distributes drinking water to more than 1.1 million people throughout Antwerp province in Belgium (Flanders).

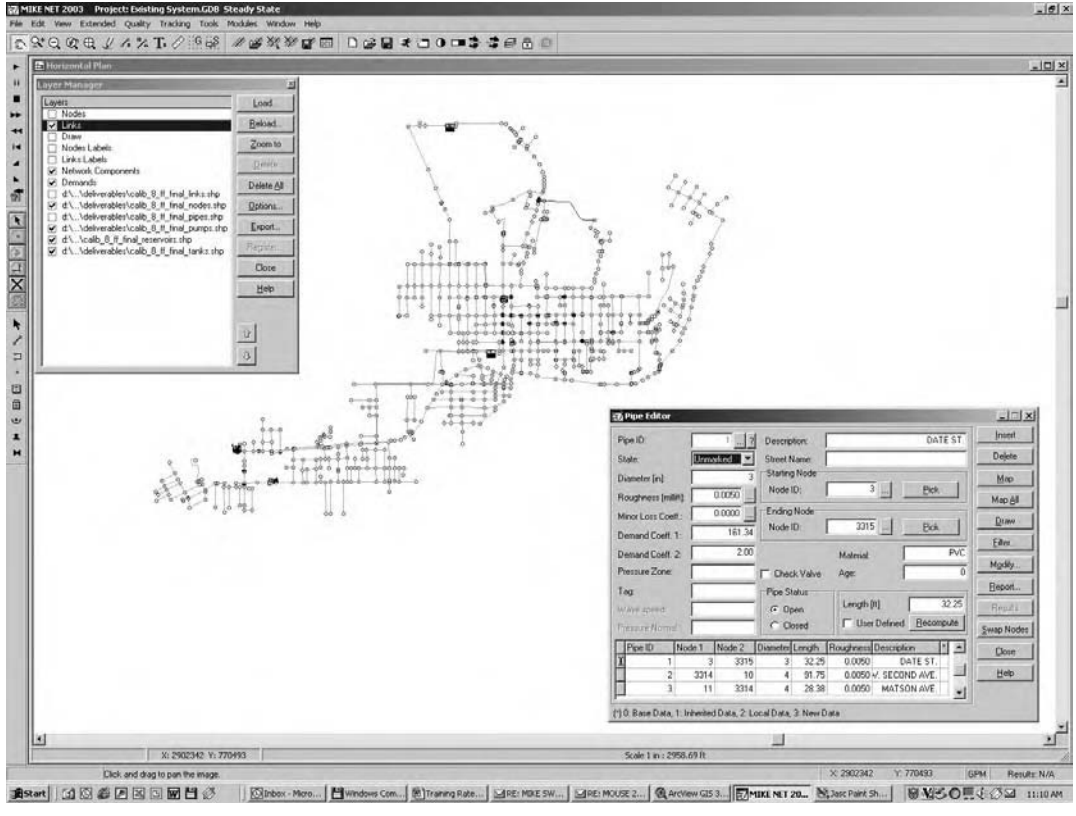


Figure 17.5 The city of Truth or Consequences' MIKE NET model developed from GIS.

Pidpa used ESRI's ArcGIS 8.2 software and Miner & Miner's ArcFM 8.1.3 software to develop a data-centric GIS solution for the management of its drinking water network, including functions for water main isolation trace and pressure zone trace (Horemans and Reynaert, 2003).

Pidpa selected ESRI as the strategic partner to implement an open and generic GIS solution for the management of its drinking water network data. The ArcGIS 8/ArcFM 8 implementation is based on a centrally located SQL Server 2000 database (with ArcSDE 8 on top) containing newly entered and converted vector data as well as large amounts of raster data at scales 1:1000 and 1:5000. All data are combined into a seamless hybrid map that is used in ArcInfo, ArcView, and ArcIMS on a daily basis by about 15 editors and a large group of viewers. Figure 17.6 shows a screenshot of ArcInfo 8 and ArcFM 8 being used on a seamless hybrid map with raster and vector data.

One of the major project goals was to create a strong generic water distribution data model that could be shared and further enriched with input from other European drinking water companies. Together with the choice for commercial off-the-shelf (COTS) software, this approach ensures a feasible future evolution of the implemented GIS and can create a large user base.

Pidpa was well aware of the fact that much of the data used throughout the company has a geographical element. With this in mind, Pidpa started integrating the GIS with many other information systems for use in the office as well as in the field. At the time of this writing, the main operational links were to CIS and to a library of 90,000 synoptic drawings. A first set of interfaces to SAP was operational, and a link to the SCADA system and hydraulic modeling software was under development. It was anticipated that by sharing data between these environments and optimizing workflow over the boundaries, users would be able to get required data in a user-friendly fashion (Reynaert et al., 2004).

WATER SYSTEM MASTER PLANNING IN PRAGUE

Application	Conceptual hydraulic modeling of water supply system
Authors	Peter Ingeduld, Zdenek Svitak, Josef Drbohlav
Project status	Conceptual model complete
Hardware	Standard Pentium III and IV desktop computers
GIS software	ArcView, LIDS (MicroStation and ORACLE based)
Other software	MIKE NET, Microsoft Access
GIS data	Water supply objects layer of Prague's water supply system
Study area	Prague, Czech Republic
Project duration	January 2001 to June 2002 (conceptual model)
Project budget	$165,000
Organization	Pražská vodohospodářská spole čnost a.s. (Prague stockholding company)

Hydraulic modeling of water supply and water distribution networks was used for planning and linking consumers to the network, evaluating the remaining capacity of the network, and planning for network breakdown. Modeling was also used to simulate various loading scenarios, such as fire suppression flow analysis and its impact on the water supply system, reconstruction of existing pipes and planning for new pipes,

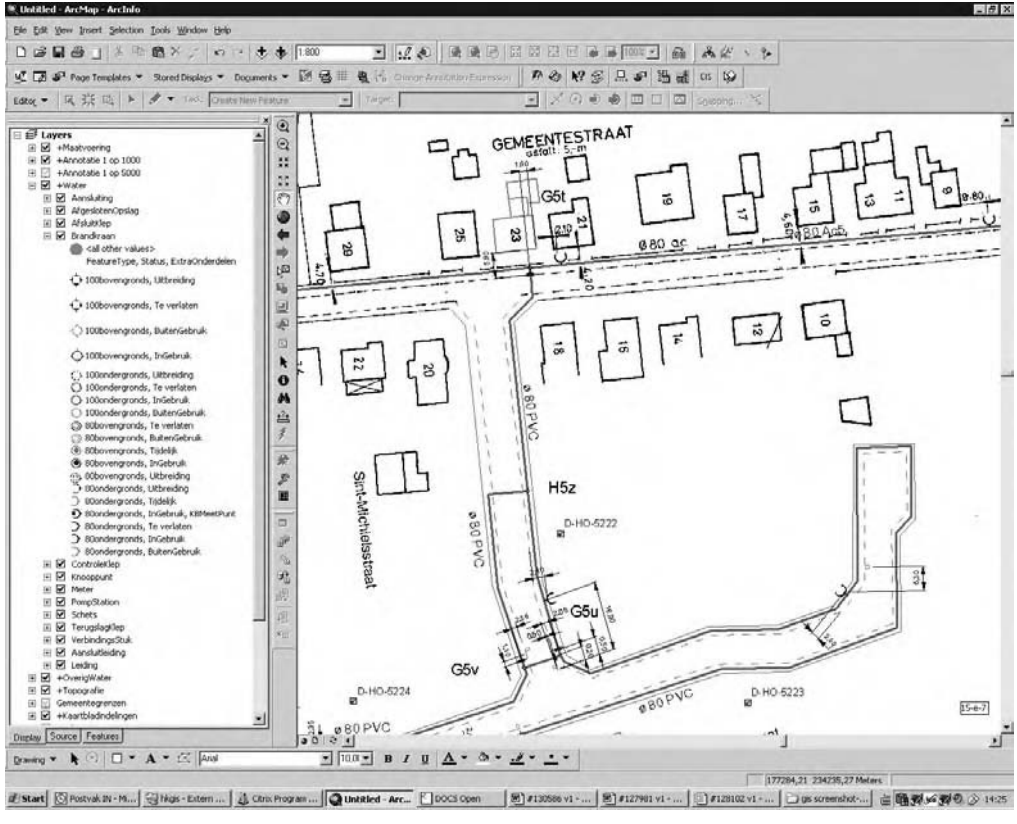


Figure 17.6 Screenshot of ArcInfo 8 and ArcFM 8.

pressure zone evaluation, and for comparing measured and simulated data. A conceptual model with all important water tanks, pumping stations, and water sources was developed for the whole Prague water supply system. This model will greatly assist in formulating and answering questions regarding supply strategies and in providing high system reliability. Hydraulic modeling was done using DHI's MIKE NET software. Figure 17.7 shows a screenshot of MIKE NET for the Prague model (Ingeduld et al., 2004).

WATER QUALITY MANAGEMENT IN MECKLENBURG COUNTY

Application	Water quality management planning
Authors	Carl W. Chen and Curtis Loeb
Project status	Completed in August 2003
Hardware	Personal computers
GIS software	Included in Watershed Analysis Risk Management Framework (WARMF)
Other software	SWAT
GIS data	10-m DEM, land-use Shapefiles for existing and future conditions
Study area	The Palisades golf course community, Mecklenburg County, North Carolina, at the eastern shore of Lake Wylie; 5.7 mi^2 for The Palisades and 1137 mi^2 of watershed surrounding Lake Wylie
Project duration	2002 to 2003 (12 months)
Project budget	$145,000
Organization	Systech Engineering, Inc., San Ramon, California

Rhein Interests, LLC, and Crescent Resources, LLC (North Carolina), proposed to develop The Palisades as a new community comprising homes, a golf course, open spaces, an equestrian center, offices, and stores. The Mecklenburg County required the project to be designed with best management practices (BMP) to meet the water quality criteria of total suspended sediment, total phosphorus, total nitrogen and algae (chlorophyll a) in the eight coves of Lake Wylie. The County approved the use of two GIS-based watershed models, SWAT and WARMF, for the project. SWAT simulated the nonpoint loads of pollutants from The Palisades, which were converted to point source loads for input to WARMF. WARMF simulated nonpoint source loads of pollutants from other areas and the water quality responses in the coves that not only receive the nonpoint source loads from The Palisades but also exchange water with the main body of Lake Wylie. Two levels of resolution were used to delineate the watershed; the finer resolution of The Palisades was nested with the coarser resolution of other subwatersheds. The lake was divided into 29 stratified sections, including coves. As a part of the project planning and design, a monitoring program was established to collect meteorology data at four stations, stream flow and water quality data at four stream stations, and water quality data at eight coves during the spring and summer months of 2002. The data were used to calibrate the models. The calibrated models were used to determine the BMP train necessary to meet the water quality criteria for coves. The water quality management plan for The Palisades was approved by the County. Figure 17.8 shows a screenshot of the WARMF application for The Palisades golf course community (Chen and Loeb, 2004).

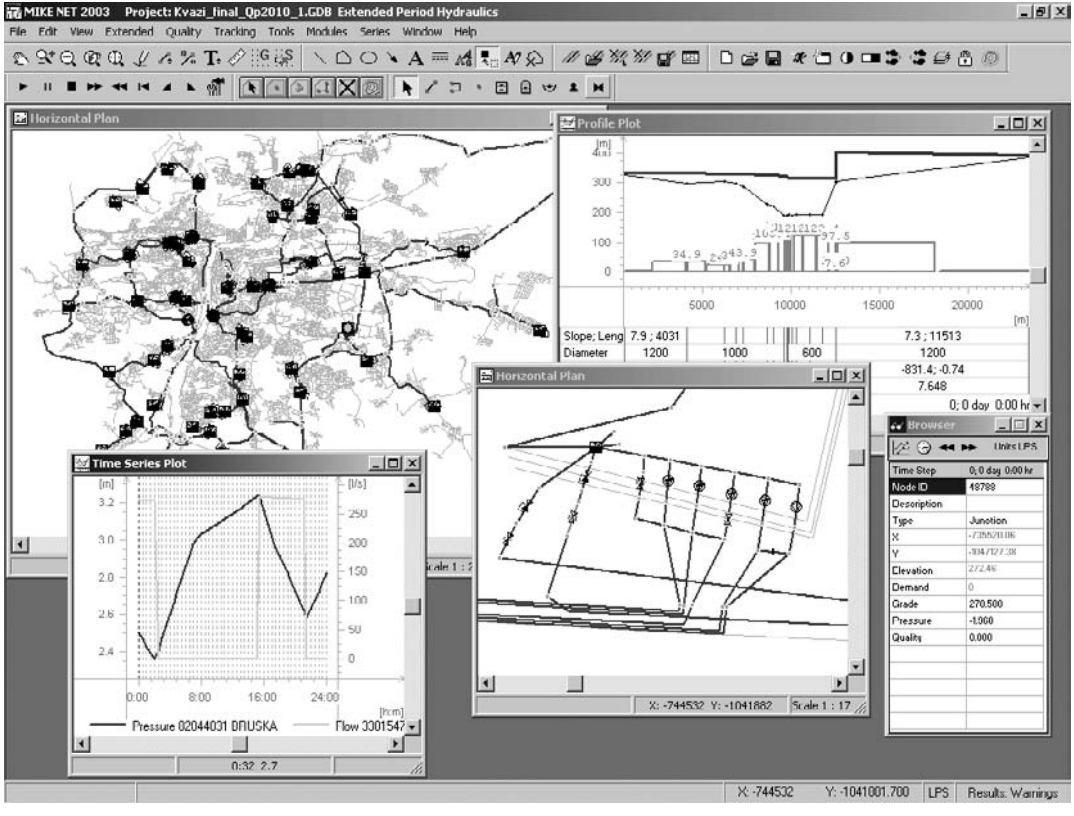


Figure 17.7 Screenshot of Prague's MIKE NET water system model.

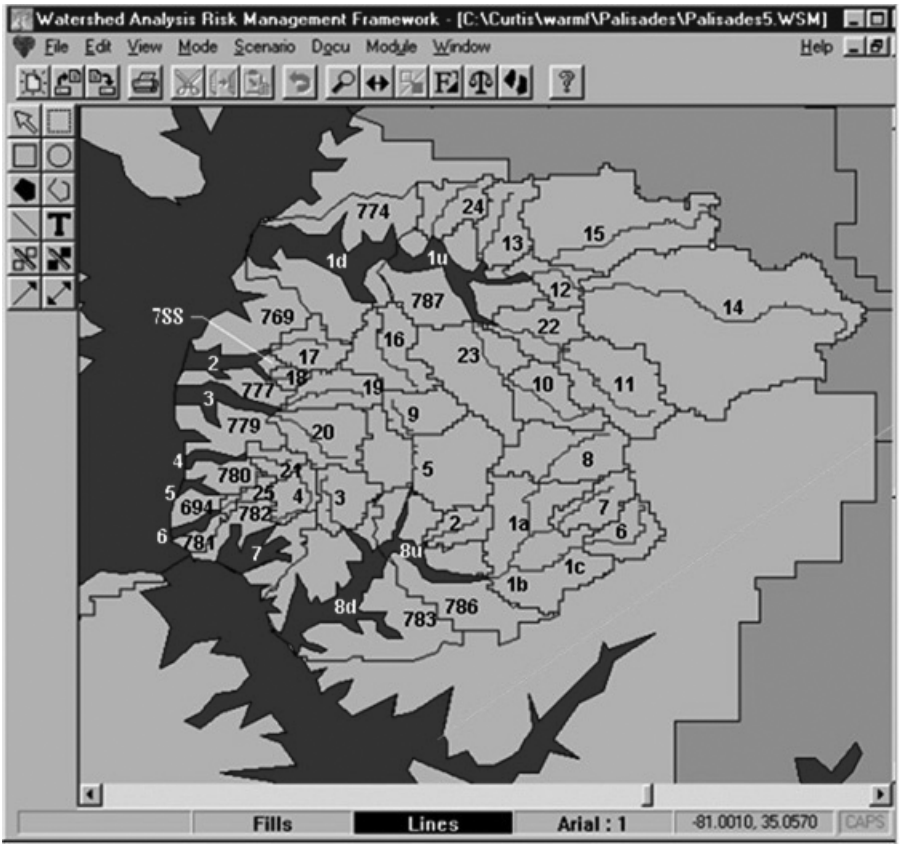


Figure 17.8 Screenshot of the WARMF application for The Palisades golf course community.

WATER MASTER PLANNING IN SUECA, SPAIN

Application	Quantity and quality modeling of water distribution systems
Authors	Hugo Bartolin and Fernando Martinez
Project status	Completed
Hardware	Standard desktop PC (Intel Pentium IV)
GIS software	ArcView GIS 3.2a, Spatial Analyst (Optional)
Other software	EPANET 2
GIS data	Three AutoCAD files with cartographic information (pipe polylines, valves, water sources, street segments, building blocks, future zones of development)
Study area	Town of Sueca (about 26,000 inhabitants) and its tourist beach area (about 30,000 in summer), both supplied by independent networks. The town is located in Valencia on the east coast of Spain.
Project duration	3 months (December 2002 to February 2003)
Project budget	$6000 (consultancy)
Organization	Polytechnic University of Valencia. Grupo REDHISP — IIAMA (Institute of Water and Environmental Engineering)
Web site	www.redhisp.upv.es www.redhisp.upv.es/software/gisred/gisred_eng.htm

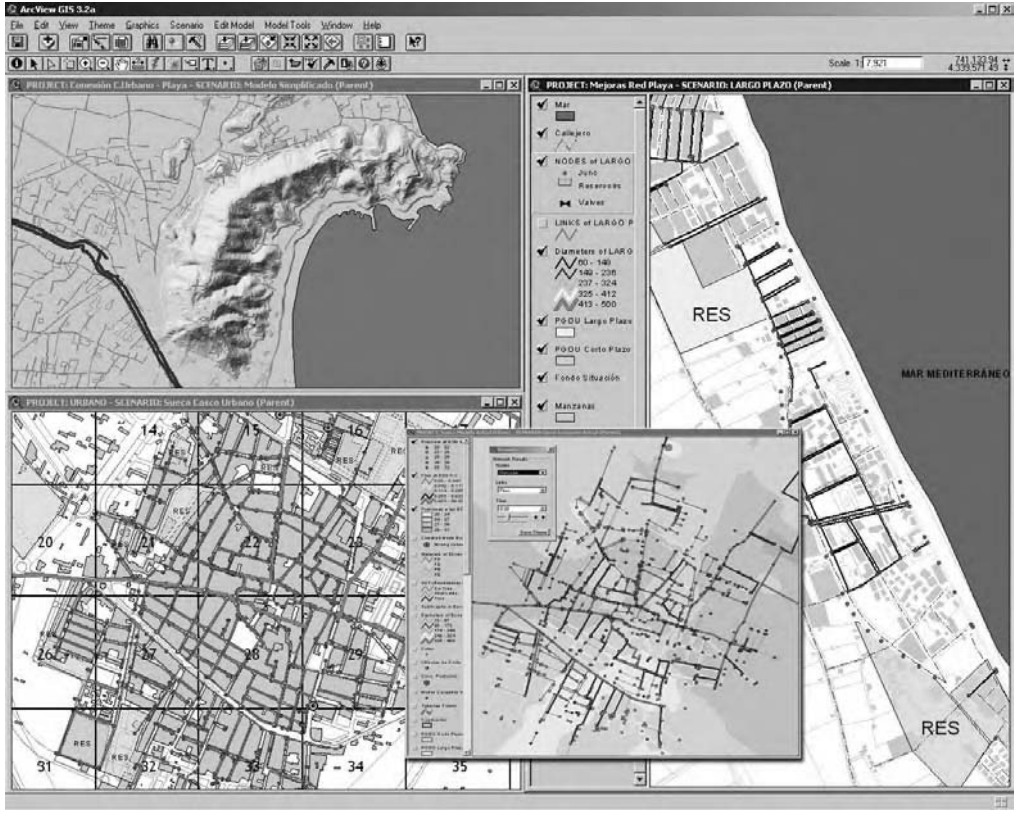


Figure 17.9 Screenshot of GISRed extension for Sueca master plan.

GISRed is a customized extension of ArcView GIS 3.2 software for modeling and calibration of water distribution networks. It basically integrates hydraulic modeling tools, the widely used hydraulic simulation software EPANET (Version 2), and a hydraulic calibration module based on genetic algorithms.

The extension is essentially a tool to assist technicians in the task of modeling water distribution networks and supporting decision-making issues based on the model, all within a GIS environment. It is useful for performing complex tasks such as importing a whole or partial network from an external data source, creating a hydraulic network model, and automatically calibrating it.

One of the tasks GISRed can assist in is master planning issues. In the Sueca master plan project, a general procedure using the GISRed extension was carried out. The main idea was to define a scenario for both independent network models and give a diagnosis of the performance of each real system. Based upon this, new short- and long-term scenarios were proposed, taking into account the variation of the population, the new growing areas, and, therefore, the future demand projection. Finally, network system improvements were planned and simulated to check the feasibility of each alternative and support all the action strategies. In addition to that, the models were connected by two mains in order to enhance the quality of service and prevent emergencies. Figure 17.9 shows a screenshot of GISRed extension for Sueca master plan (Bartolin and Martinez, 2004).

CHAPTER SUMMARY

This chapter presented eight case studies on GIS applications for water, wastewater, and stormwater systems in Belgium, Bulgaria, Czech Republic, Denmark, Spain, and the U.S. The case studies indicate that GIS is being used in all aspects of water, wastewater, and stormwater management, from planning and H&H modeling to mapping and asset management.

CHAPTER QUESTIONS

1. On the basis of the case studies presented in this chapter, which GIS application appears to be the most common?
2. On the basis of the case studies presented in this chapter, which GIS software appears to be the most commonly used?
3. Using the case study presentation format used in this chapter, write a case study summary of a GIS project completed in your organization. Optionally, submit this case study to the author (info@GISapplications.com) for potential publication at GISapplications.com Web site.

APPENDIX A

Acronyms

2-D Two-Dimensional
3-D or 3D Three-Dimensional
ACE Army Corps of Engineers
ACP Asbestos Cement Pipe
ADRG Arc Digitized Raster Graphics
AIRS Aerometric Information Retrieval System
AM/FM Automated Mapping/Facilities Management
AM/FM/GIS Automated Mapping/Facilities Management/Geographic Information System
AML Arc Macro Language
AMSA Association of Metropolitan Sewerage Agencies
API Application Programming Interface
ASCE American Society of Civil Engineers
ASCII American Standard Code for International Interchange
ASP Application Service Provider
AVI Audio Video Interleaved (digital movie video format)
AWRA American Water Resources Association
AWWA American Water Works Association
BASINS Better Assessment Science Integrating Point and Nonpoint Sources
BIL Band Interleaved by Line
BIP Band Interleaved by Pixel
BMP Best Management Practice
BRS Biennial Reporting System
BSQ Band SeQuential
BYU Brigham Young University
CAD Computer-Aided Drafting/Computer-Aided Design
CADD Computer-Aided Drafting and Design
CAFM Computer-Aided Facility Management
CAM Computer-Aided Mapping
CCTV Closed-Circuit Television
CD Compact Disc
CD-ROM Compact Disc-Read Only Memory
CGM Computer Graphic Metafile

CIFM Computer Integrated Facility Management
CIP Cast Iron Pipe
CIP Capital Improvement Project
CIS Customer Information System
CMMS Computerized Maintenance Management System
CMOM Capacity, Management, Operations, and Maintenance
CMP Corrugated Metal Pipe
COE Corps of Engineers
COGO Coordinate Geometry
COM Component Object Model
CORBA Common Object Request Broker Architecture
COTS Commercial Off-the-Shelf
CSO Combined Sewer Overflow
CTG Composite Theme Grid
DAK Data Automation Kit
DBMS Database Management System
DDE Dynamic Data Exchange
DEM Digital Elevation Model
DEP Department of Environmental Protection
DFIRM Digital Flood Insurance Rate Map
DGPS Differential Global Positioning System
DHI Danish Hydraulic Institute
DIB Device Independent Bitmap
DIME Dual Independent Map Encoding
DIP Ductile Iron Pipe
DLG Digital Line Graph
DLL Dynamic Link Library
DOP Digital Orthophoto
DOQ Digital Orthophoto Quadrangle
DOQQ Digital Orthophoto Quarter Quadrangle
DR3M Distributed Routing Rainfall Runoff Model
DRG Digital Raster Graphics (USGS topographic maps)
DSS Data Support System
DSS Decision Support System
DTED Digital Terrain Elevation Model
DTM Digital Terrain Model
DVD Digital Versatile Disc or Digital Video Disc
DXF Drawing Exchange Format
EDC EROS Data Center
EMF Enhanced Metafile Format
EOS Earth Observation System
EPA (U.S.) Environmental Protection Agency
ERDAS Earth Resource Data Analysis System
EROS Earth Resources Observation Systems
ERP Emergency Rresponse Plan
ESDLS EPA Spatial Data Library System
ESRI Environmental Systems Research Institute

FAP File of Arcs by Polygon
FAT Feature Attribute Table
FEMA Federal Emergency Management Agency
FGDC Federal Geographic Data Committee
FIRM Flood Insurance Rate Map
FIS Flood Insurance Study
FMSIS Flood Map Status Information Service
FTP File Transfer Protocol
GAAP Generally Accepted Accounting Principales
GASB Government Accounting Standards Board
GEMS Geographic Engineering Modeling Systems
GIF Graphics Interchange Format
GIRAS Geographic Information Retrieval and Analysis System
GIS Geographic Information Systems
GLONASS Global Navigation Satellite System (Russian Systemsystem)
GML Geography Markup Language
GPD Gallons Per Day
GPS Global Positioning System
GRASS Geographic Resources Analysis Support System
GUI Graphical User Interface
H&H Hydrologic and Hydraulic
HEC Hydrologic Engineering Center (U.S. Army Corps of Engineers)
HGL Hydraulic Gradient Line
HoLIS Honolulu Land Information System
HRAP Hydrologic Rainfall Analysis Project
HSG Hydrologic Soil Group
HSPF Hydrologic Simulation Program — Fortran
HTML Hypert Text Markup Language (a file extension)
HTTPS Secure Hypert Text Transmission Protocol
HUC Hydrologic Unit Code
HUD Housing and Urban Development
IAC Interapplication Communication
ICR Information Collection Rule
IFSAR Interferometric Synthetic Aperture Radar
IMS Internet Map Server
ISE Integral Square Error
IT Information Technology
JFIF JPEG File Interchange Format
JPEG Joint Photographic Experts Group
JSP Java Server Pages
LAN Local Area Network
LANDSAT Land Satellite
LBS Location-Based Services
LIDAR Laser Imaging Detection and Ranging
LOMC Letter of Map Change
LTCP Long-Term Control Plan
LULC Land Use/Land Cover

MA Metropolitan Area
MGD Million Gallons per Day
MIXEL Mixed Cells
MMS Maintenance Management System
MMS Multimedia Mapping System
MPEG Moving Picture Experts Group (digital movie video format)
MrSID Multir Resolution Seamless Image Database
MS4 Municipal Separate Storm Sewer System
MSC Map Service Center
NAD-27 North American Datum of 1927
NAD-83 North American Datum of 1983
NASA National Aeronautics & and Space Administration (U.S.)
NAVSTAR Navigation System by Timing and Ranging
NCDC National Climatic Data Center (U.S.)
NED National Elevation Dataset
NEXRAD Next Generation Weather Radar
NGDC National Geospatial Data Clearinghouse
NGVD National Geodetic Vertical Datum
NHD National Hydrography Dataset
NLCD National Land Cover Database
NMAS National Map Accuracy Standards
NMC Nine Minimum Controls (for CSOs)
NOAA National Oceanic and Atmospheric Administration (U.S.)
NPDES National Pollution Discharge Elimination System
NPS Nonp Point Source
NQL Network Query Language
NRCS Natural Resources Conservation Service
NSDI National Spatial Data Infrastructure
NSGIC National States Geographic Information Council (U.S.)
NWI National Wetlands Inventory
NWS National Weather Service
O&M Operation and Maintenance
ODBC Open Database Connectivity
OGC Open GIS Consortium
OIRM Office of Information Resources Management
OLE Object Linking and Embedding
PASDA Pennsylvania Spatial Data Access
PC Personal Computer
PCS Permit Compliance System
PDA Personal Digital Assistant (an electronic handheld information device)
PDF Portable Document Format (Adobe Acrobat)
PDOP Positional Dilution of Precision
PFPT Peak Flow Presentation Table
PIXEL Picture Element
POTW Publicly Owned Treatment Works
PPS Precise Positioning Service
PRR Practical Release Rate

PRV Pressure Regulating Valve
PVC Polyvinyl Chloride
QA/QC Quality Assurance/Quality Control
RBE Rule Base Engine
RCP Reinforced Concrete Pipe
RCRIS Resource Conservation and Recovery System
RDBMS Relational Database Management System
RF Reach File
RGB Red Green Blue (scientific hues or primary colors)
RMSE Root Mean Square Error
ROM Read Only Memory
RPC Remote Procedure Call
RR Release Rate
RTB Retention Treatment Basin
RTK Real Time Kinematic
SA Selective Availability
SAR Synthetic Aperture Radar
SCADA Supervisory Control and Data Acquisition
SCS Soil Conservation Service
SDE Spatial Database Engine
SDTS Spatial Data Transfer Standard
SHC System Hydraulic Characterization (for CSOs)
SHE System Hydrologique Europeen
SML Simple Macro Language
SMS Short Message Service
SPC State Plane Coordinate (map projection system)
SPS Standard Positioning Service
SQL Structured Query Language
SSL Secure Socket Layer
SSO Sanitary Sewer Overflow
SSURGO Soil Survey Geographic
STATSGO State Soil Geographic
STORM Storage Treatment Overflow Runoff Model
STP Sewage Treatment Plant
SVG Scalable Vector Graphics
SWAT Soil and Water Assessment Tool
SWMM Storm Water Management Model
TAZ Traffic Analysis Zone
TCP Terra-Cotta Pipe
TCP/IP Transmission Control Protocol/Internet Protocol
TIF Tag Image File
TIFF Tag Image File Format
TIGER Topologically Integrated Geographic Encoding and Referencing System (U.S. Census Bureau mapping system)
TIN Triangular Irregular Network/Triangulated Irregular Network
TM Thematic Mapper
TMDL Total Maximum Daily Load

TMS TIGER Map Service
TR Technical Release
TR-20 Technical Release 20
TRI Toxic Release Inventory
TRR Theoretical Release Rate
TVP Topological Vector Profile
UAV Unmanned Aerial Vehicles
UCGIS University Consortium of Geographic Information Science
UPS Uninterruptible Power Supply
URL Uniform (or Universal) Resource Locator
USACE United States Army Corps of Engineers
USACERL United States Army Construction Engineering Research Laboratory
USBC United States Bureau of Census
USGS United States Geological Survey
UTM Universal Transverse Mercator
VB Visual Basic
VBA Visual Basic for Applications
VCP Vitrified Clay Pipe
VCR Video Cassette Recorder
VHS Video Home System (video cassette format)
VRML Virtual Reality Modeling Language
VSAT Vulnerability Self Assessment Tool
WAAS Wide Area Augmentation System
WAP Wireless Application Protocol
WCS Web Coverage Service
WEF Water Environment Federation
WFS Web Feature Service
WGS84 World Geodetic System of 1984
WIMS Wastewater Information Management System
WISE Watershed Information System
WML Wireless Markup Language
WMS Watershed Modeling System
WMS Web Map Server
WSR Weather Surveillance Radar
WSDL Web Service Description Language
WWW World Wide Web
XLST Extensible Stylesheet Language Transformation
XML Extensible Markup Language

APPENDIX B

Conversion Factors

Conversion Factors from Customary U.S. Units to Système International (SI) Units

To Covert	To	Multiply by
Acres	Square meters (m^2)	4047
Acres	Hectares (ha)	0.4047
Cubic feet	Liters (l)	28.32
Cubic feet per second (cfs)	Cubic meters per second	0.02832
Cubic yards	Cubic meters (m^3)	0.7646
Feet (ft)	Meters (m)	0.3048
U.S. gallons (water)	Liters (l)	3.7843
U.S. gallons (water)	Cubic centimeters (cm^3)	3785
U.S. gallons (water)	Milliliters (ml)	3785
U.S. gallons (water)	Grams (g)	3785
U.S. gallons (water)	Cubic meters (m^3)	0.0038
Gallons per minute (GPM)	Liters per second	0.0631
Gallons per minute (GPM)	Liters per minute	63.08
Inches (in.)	Centimeters (mm)	2.54
Inches (in.)	Millimeters (mm)	25.4
Miles (mi)	Kilometer (km)	1.6093
Millions gallons per day (MGD)	Cubic meters per day	3785
Millions gallons per day (MGD)	Cubic meters per second	0.044
Pounds (lb)	Kilograms (kg)	0.4536
Square feet (ft^2)	Square meters (m^2)	0.0929
Square miles (mi^2)	Square kilometers (km^2)	2.5921
Yards	Meters (m)	0.9144
Square miles (mi^2)	Square kilometers (km^2)	2.5899

References

Alston, R. and Donelan, D. (1993). Weighing the benefits of GIS. *American City & County*, October 1993, Vol. 108, No. 11, 14.

American City & County (1994). GIS assists Texas town in meeting NPDES regulations. *American City & County*, March 1994, 62.

American City & County (1993). CAGIS restores order to Cincinnati infrastructure. *American City & County*, August 1993, 24.

American City & County (1991). Albuquerque integrates GIS to meet needs. *American City & County*, May 1991, 30.

Anderson, F. (2000). Water resources planning and management in the Internet era. *Water resources impact*, American Water Resources Association, Vol. 2, No. 5, September 2000, 5–8.

Anderson, G. (1998). Program yields wastewater piping hydraulic model, *Water World*, October, l998.

ArcNews (2001a). Colorado utility maps in the Rockies with imagery. *ArcNews*, Vol. 23, No. 3, Fall 2001, ESRI, 20–21.

ArcNews (2001). GIS goes with the flow. *ArcNews*, Vol. 23, No. 1, Spring 2001, ESRI, 34–35.

ArcNews (2000). ArcGIS 8.1 nears release. *ArcNews*, Vol. 22, No. 4, Winter 2000/2001, ESRI, Spring 2000, 1–2.

ASCE (2003). Little progress made on infrastructure, ASCE report finds. *ASCE News*, American Society of Civil Engineers, Vol. 28, No. 10, October 2003, 1.

ASCE (2001). New systems make surveying easier more accurate. *Civil Engineering*, May 2001, 74–76.

ASCE (1999). GIS Modules and Distributed Models of the Watershed, Task Committee Report, Paul A. DeBarry and Rafael G. Quimpo Committee Co-chairs, ASCE.

ASCE (1998). New poll shows voters worried about aging infrastructure, *ASCE News* American Society of Civil Engineers, November 1998.

Azteca Systems (1999). White Paper: Pipeworks Data Model Concepts, Azteca Systems Web site at http://www.azteca.com/data_model.htm.

Babbitt, L.M. (2002). Lehigh County Authority, Allentown, PA. *Water Writes*, ESRI, Fall 2002, 6–7.

Barnes, S. (2001a). ASPRS 2001: Remote sensing opportunities soar. *Geospatial Solutions*, Vol. 11, No. 6, June 2001, 16.

Barnes, S. (2001b). United in purpose — spatial help in the aftermath. *Geospatial Solutions*, Vol. 11, No. 11, November 2001, 34–39.

Barnes, S. (2001). ESRI and Sun Launch LBS portal. *Geospatial Solutions*, February 2001, 16.

Barnes, S. (2000). One last thing — GIS market revenues top $1.5 billion. *Geospatial Solutions*, Vol. 10, No. 6, June 2000, 58.

Bartolin, H. and Martinez, F. (2004). GISRed, an assistant tool to develop master planning projects of water distribution system. Case study submission in response to a call for case studies distributed by the author (U. M. Shamsi) for this book.

Bathurst, J.C., Wicks, J.M., and O'Connell, P.E. (1995). The SHE/SHESED Basin scale water flow and sediment transport modelling system, Chapter 16 in *Computer Models of Watershed Hydrology*, V.P. Singh (Ed.), Water Resources, 563–594.

Battin, A., Kinerson, R., and Lahlou, M. (1998). EPA's Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) — A Powerful Tool for Managing Watersheds. Internet file retrieved 11/6/98 from The Center for Research in Water Resources, The University of Texas at Austin. http://www.crwr.utexas.edu/gis/gishyd98/epa/batin/p447.htm.

Bell, J. (1999a). ArcView GIS. *Professional Surveyor*, March 1999, 48–50.

Bell, J. (1999). AutoCAD Map 3. *Professional Surveyor*, January/February 1999, 52–53.

Bernhardsen, T. (1999). *Geographic Information Systems — An Introduction*, 2nd ed., John Wiley & Sons, New York, 372.

Binford, M.W. (2000). Clark lab's Idrisi32. *Geo Info Systems*, April 2000, 40–42.

Booth, R. and Rogers, J. (2001). Using GIS technology to manage infrastructure capital assets. *Journal AWWA*, American Water Works Association, November, 2001, Vol. 93, No. 11, 62–68.

Bradford, E., Watson, R., Mann, E., and Konwinski, J. (2004a). MSD Pipe Rating Program. Case study submission in response to a call for case studies distributed by the author (U.M. Shamsi) for this book.

Bradford, E., Watson, R., Mann, E., and Konwinski, J. (2004). Pipe rating. *Water Environment & Technology*, Water Environment Federation, February, 2004, Vol. 16, No. 2, 39–43.

Bufe, M. (2003). Going digital. *Water Environment & Technology*, Water Environment Federation, October 2003, Vol. 15, No. 10, 24–27.

Burns, N.L. et al. (2001). Security analysis and response for water utilities. AWWA, Denver, CO.

Butler, D. and Maksimovic, C. (Eds.) (1998). *UDM'98 4th International Conference on Developments in Urban Drainage Modeling*. Imperial College of Science, Technology & Medicine, London.

California Water Districts Link GIS to Document Management System (1998). *ArcNews*, Spring 1998, 19.

Cannistra, J.R. (1999). Converting utility data for a GIS. *Journal AWWA*, American Water Works Association, February 1999, 55–64.

Cannistra, J.R., Leadbeater, R., and Humphries, R. (1992). Washington suburban sanitary commission implements GIS. *Journal AWWA*, AWWA, July 1992, 62–67.

CE News (2001). Group seeks Congressional aid for water infrastructure. Vol. 13, No. 5, June 2001, 13–16.

Cerasini, V. and Herskowitz, J. (2001). City streamlines customer service. *Imaging Notes*, Space Imaging, Thornton, CO, Vol. 16, No. 5, September/October 2001, 26–27.

Chang, T., Hsu, M., Teng, W., and Huang, C. (2000). A GIS-assisted dstributed watershed model for simulating flooding and inundation. *Water Resources Bulletin*, American Water Resources Association, Vol. 36, No. 5, October 2000, 975–988.

Charnock, T.W., Hedges, P.D., and Elgy, J. (1996). Linking multiple process models with GIS. In K. Kovar and H.P. Nachtnebel (Eds.), *HydroGIS'96: Application of Geographic Information Systems in Hydrology and Water Resources Managemen*t. International Association of Hydrologic Sciences Publication No. 235. IAHS, Wallingford, U.K. 29–36.

Chen, C.W. and Loeb, C. (2004). Watershed Analysis Risk Management Framework (WARMF) Case Study. Case study submission in response to a call for case studies distributed by the author (U.M. Shamsi) for this book.

Cheves, M. (2000). GIS and surveying. *Professional Surveyor*, Vol. 20, No. 9, October 2000, 24.

Chien, P. (2000). Endeavour maps the world in three dimensions. *GEOWorld*, April 2000, Vol. 13, No. 4.

Christian, S.M. and Yoshida, G.K. (2003). Mapping out success. *Water Environment & Technology*, Vol. 15, No. 8, August 2003, 50–55.

City of Glendale Database Design (1996). ESRI, Redlands, CA, July 1996.

Coates, J. (2003). Water main shutdown application prevents costly mistakes. *ArcUser*, ESRI, January–March, 2003, Vol. 6, No. 1, 20–21.

Computational Hydraulics Int. (1995). *PCSWMM*. User's manual, 36 Stuart Street, Guelph, Ontario, Canada N1E 4S5.

Corbley, K.P. (2001). Adding value to data. *Imaging Notes*, Vol. 16, No. 2, March/April 2001, 27–28.

Corbley, K.P. (2000). Plugging in. *Imaging Notes*, Space Imaging, Thornton, CO, March/April 2000, 14–15.

Criss, Cori. (2000). WinCan simplifying data gathering, *Trenchless Technology*, November 2000, 64–65.

Curtis, T.G. (1994). SWMMDUET: Enabling EPA SWMM with the ARC/INFO paradigm. *ArcNews*, Environmental Systems Research Institute, Redlands, CA, Spring 1994, 20.

Dalton, L.M. (2001). Prospecting spatial database offerings. *Geospatial Solutions*, Vol. 11, No. 10, October 2001, 40–44.

Daratech (2000). Geographic Information Systems — Markets and Opportunities, Cambridge, MA.

Davis, J. (2000). Keys to success. *Geo Info Systems*, February 2000, 32–35.

Davis, K. and Prelewicz, G. (2001). Concerned about CMOM? *Water Environment & Technology*, Vol. 13, No. 3, March 2001, 33–36.

Day, J.M. (1998). Rainfall Quantity Analysis from a Watershed Perspective: Using the Virtual Rain Gauge, International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

DeGironimo, E. and Schoenberg, J. (2002). The key to the vault. *Water Writes*, ESRI, Fall 2002, 1–3.

Denver Water Department (Colorado, USA) installed one of the first AM/FM/GIS system "Synercom" in USA in 1986 (Lang, 1992).

DeVantier, B.A. and Feldman, A.D. (1993). Review of GIS applications in hydrologic modeling, *Journal of Water Resources Planning and Managemen*t, 119 (2): 246–261.

Djokic, D., Meavers, M.A., and Deshakulakarni, C.K. (1993). *ARC/HEC2: An ArcInfo — HEC-2 Interface.* User's manual, Center for Research in Water Resources, The University of Texas at Austin, Austin, TX.

Doan, J.H. (1999). Hydrologic Model of the Buffalo Bayou Using GIS, International User Conference, ESRI, San Diego, CA, July 26–30, 1999.

Dorris, V.K. (1989). Systems link geography and data, *ENR*, June 1, 1989, 30–37.

Dougherty, L. (2000). Census 2000, census sense, part 2, geographic products for one and all. *Geospatial Solutions*, September 2000, 36–39.

Doyle, G. and Grabinsky, M. (2003). Applying GIS to a water main corrosion study. *Journal AWWA*, AWWA, May 2003, Vol. 95, No. 5, 90–104.

Doyle, M.J. and Rose, D. (2001). Protecting your assets. *Water Environment & Technology*, Water Environment Federation, July, 2001, Vol. 13, No. 7, 43–47.

Dueker, K.J. (1987). Geographic information systems and computer-aided mapping, *APA Journal*, Summer issue, 383–390.

Edelman, S., Dudley, T., and Crouch, C. (2001). Much ado about data, *CE News*, January 2001, 52–55.

Elroi Consulting, Inc. (2002). Stormwater Utility GIS. Company literature at www.elroi.com.

Engelhardt, J. (2001a). Mapping the future. *Imaging Notes*, Space Imaging, Thornton, CO, May/June 2001, 30.

Engelhardt, J. (2001). GITA's technology report. *Geospatial Solutions*, February 2001, 14.

Engman, E.T. (1993). Remote sensing. Chapter 24 in *Handbook of Hydrology*, D.R. Maidment (Ed.), McGraw Hill, NY, 24.1–24.23.

EPA (2002). *SWMM Redevelopment Project Plan.* Version 5, Water Supply and Water Resources Division, National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH.

EPA (2000). Environmental Planning for Communities — A Guide to the Environmental Visioning Process Utilizing a Geographic Information System (GIS). EPA/625/R-98/003, Technology Transfer and Support Division, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, September 2000.

ERDAS (2001a). Ventura capitalizes with imagery. *ERDAS New*s, ERDAS, April 2001, Vol. 2, No. 2, 18.

ERDAS (2001b). IMAGINE OrthoBASE Pro-Automatic DEM extraction made easy. *ERDAS New*s, ERDAS, April 2001, Vol. 2, No. 2, 5.

ERDAS (2001). There's more to a Pixel than meets the eye. *ERDAS News*, ERDAS, January 2001, Vol. 2, No. 1, 18–19.

ESRI (2004a). Inventorying Infrastructure with ArcPad. *ArcUser*, ESRI, January–March, 2004, Vol. 7, No. 1, 38–39.

ESRI (2004). Taking GIS on the road. *ArcUser*, ESRI, January–March, 2004, Vol. 7, No. 1, 30–32.

ESRI (2003). GIS standards and interoperability. *ArcNews*, ESRI, Vol. 25, No. 1, Spring 2003, 1.

ESRI (2002). AAA's TripTik/Traveler rated the best cartographer, *ArcNews*, ESRI, Vol. 24, No. 3, Fall 2002, 1.

ESRI (2002a). Elsinore Valley municipal water district goes live with ArcGIS. *ArcNews*, ESRI, Vol. 23, No. 4, Winter 2001–2002, 20.

ESRI (2002b). City of Germantown, Tennessee, implements water inventory, modeling. *ArcNews*, Vol. 24, No. 1, ESRI, Spring 2002, 16.

ESRI (2002c). Developer samples for water/wastewater, *Water Writes*, ESRI, Winter 2002, 9.

ESRI (2002d). A common vision for homeland security, *ArcUser*, ESRI, Vol. 5, No. 1, January–March 2002, 8–11.

ESRI (2000). Sharing geographic knowledge: providing access to geographic information, *ArcNews*, ESRI, Spring 2000, 6.

ESRI. (1999). Utility market embraces ArcFM GIS solution, *ArcNews*, ESRI, Spring 1999, 4 p.

ESRI (1994). Massachusetts water resources authority's sewer management project. *The GIS People*, June 1994.

ESRI (1992). ArcView User's Guide, 2nd ed., Environmental System Research Institute, Redlands, CA.

ESRI Educational Services (1995). *Introduction to Avenue*, Environmental System Research Institute, Redlands, CA.

Estes-Smargiassi, S. (1998). Massachusetts water resources authority uses GIS to meet objectives cost-effectively. *Water Writes*, ESRI.

EWRI (2001). ASCE's 2001 report card for America's infrastructure. *EWRI Currents*, Environmental and Water Resources Institute, ASCE, Vol. 3, No. 2, June 2001, 9–10.

Farkas, A.L. and Berkowitz, J.B. (2001). State-of-the-industry report. *Environmental Engineer*, American Academy of Environmental Engineers, Vol. 37, No. 4, October 2001, 22–26.

Fillmore, L., Doyle, M., Khan, I., Russell, T., and Osei-Kwadwo, G. (2001). CMOM to see efficiency. *Water Environment & Technology*, Vol. 13, No. 3, March 2001, 39–43.

Fontenot, E. (2004). Development of a Water Distribution Model in MIKE NET from GIS Data. Case study submission in response to a call for case studies distributed by the author (U.M. Shamsi) for this book.

Frank, A. (2001). Building better maps. *Imaging Notes*, Space Imaging, Thornton, CO, May/June 2001, 18–19.

Gaines, J.B. (2001). The power of preventive maintenance. *Water Environment & Technology*, Vol. 13, No. 8, August 2001, 60–65.

Garbrecht, J. and Martz, L. 2000. TOPAZ, Topographic Parameterization Software, Grazing Labs Research Laboratory Web site at http://grl.ars.usda.gov/topaz/TOPAZ1.HTM.

Gay, S. (2003). Remote sensing, GIS, and a taxing dilemma. *Geospatial Solutions*, April 2003, Vol. 13, No. 4, 34–38.

Geographic Information Systems, (2000). USGS Web site at www.usgs.gov/research/gis/title.html.

Geospatial Solutions (2004). Rand report: Federal spatial Web sites don't help terrorists. *Geospatial Solutions*, Vol. 14, No. 5, May 2004, 16.

Geospatial Solutions (2001). Gaze into the future — hot trends 2001. January 2001, 20–29.

Geospatial Solutions (2001a). Innovative, imaginative, intricate - top Apps 2001. August 2001, 16.

GEOTELEC (2004). Computerizing Facility Management Activities. Company Web site, www.geotelec.com/cfma.html.

GEOWorld (2001). Wireless fleet management improves. September 2001, Vol. 14, No. 9, 36.

Gilbrook, M.J. (1999). GIS paves the way. *Civil Engineering*, ASCE, Vol. 69, No. 11, November 1999, 34–39.

Gish, B.W. (2001). Resolution demonstration. *Local Government Handbook for GIS Implementation within the Commonwealth of Pennsylvania CD-ROM*, May 23, 2001, Pennsylvania Mapping and Geographic Information Consortium, Lewisburg, Pennsylvania, www.pamagic.org.

GITA (2001). *Geospatial Technology Report 2000.* Geospatial Information & Technology Association, www.gita.org, January 2001.

GITA (2003). *Geospatial Technology Report 2002: A Survey of Organizations Implementing Information Technologies.* Geospatial Information & Technology Association, www.gita.org.

Goldstein, H. (1997). Mapping convergence: GIS joins the enterprise, *Civil Engineering*, ASCE, June, 1997, 36–39.

Goodchild, M.F. (1996). The Application of Advanced Technology in Assessing Environmental Impacts. In D.L. Corwin and K. Loague (Eds.), *Application of GIS to the Modeling of Non-point Source Pollutants in the Vadose Zone.* Madison, WI, Soil Science Society of America, 1–17.

Goodchild, M. (1998). Uncertainty: the Achilles heels of GIS, *Geo Info Systems*, November 1998, 50.

Goodchild, M. (2002). MapFusion for GIS interoperability. *Geospatial Solutions*, Vol. 12, No. 4, April 2002, 48–51.

Gorokhovich, Y., Khanbilvardi, R., Janus, L., Goldsmith, V., and Stern, D. (2000). Spatially distributed modeling of stream flow during storm events. *Journal of the American Water Resources Association*, Vol. 36, No. 3, June 2000, 523–539.

Goubert, D. and Newton, R. (1992). Small utility GIS, *Civil Engineering*, ASCE, November 1992, 69–71.

Graybill, G.R. (1998). Database Design Issues for the Integration of ARC/INFO with Hydraulic Analysis Applications, International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

Greene, R.G. and Cruise, J.F. (1995). Urban watershed modeling using Geographic Information System, *Journal of Water Resources Planning and Managemen*t, Vol. 121, No. 4, 318–325.

Griffin, C.B. (1995). Data quality issues affecting GIS use for environmental problem-solving. *Proceedings of the National Conference on Environmental Problem-Solving with Geographic Information Systems*, EPA/625/R-95/004, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH, September 1995, 15–30.

Grigg, N.S. (2003). Surviving disasters: learning from experience. *Journal AWWA*, American Water Works Association, Vol. 95, No. 9, September 2003, 64–75.

Grigg, N.S. (2002). Surviving disasters in water utilities: learning from experience. AWWA Research Foundation, Denver, CO.

Guetti, B. (1998). Developing a Water Distribution System Maintenance Application, International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

Haestad Methods, (2004). Bethlehem SCADA connects waterGEMS. *Client Care Newsletter*, 11, 7, Haestad Press, Waterbury, CT.

Haestad Methods (2003). One-on-one with Jack Dangermond. *ClientCare Newsletter*, March/April 2003, Haestad, Waterbury, CT.

Haestad Methods (2003a). Securing water systems. *ClientCare Newsletter*, June/July 2003, Haestad, Waterbury, CT.

Haestad Methods (2002). Water, a demanding solution. *ClientCare Newsletter*, Oct/Nov 2002, Haestad, Waterbury, CT.

Haestad Methods, Walski, T., Chase, D., Savic, D., Grayman, W., Beckwith, S., and Koelle, E. (2003). *Advanced Water Distribution Modeling and Management*, Haestad, Waterbury, CT.

Hallam, C.A., Salisbury, J.M., Lanfear, K.J., and Battaglin, W.A. (Eds.) (1996). *Proceedings of the AWRA Annual Symposium: GIS and Water Resources.* American Water Resources Association, Herndon, VA.

Hamid, R. and Nelson, R.E. Models, GIS, and databases — the essential tools in wastewater planning, Collection System Odyssey, *2001 Collection Systems Specialty Conference*, WEF, Bellevue, WA, July 8–11, 2001.

Hammond, M. (1999). GIS Lands on the Map, *PCWeek*, January 4, 1999, 42–47.

Hansen, R. 1998. Importing U.S. Census TIGER Files into ArcView GIS, *ArcUser*, ESRI, July–September 1998, 46.

Harlin, J.M. and Lanfear, K.J. (Eds.) (1993). *Proceedings of the symposium on Geographic Information Systems and water resources.* American Water Resources Association, Bethesda, MD.

Hay, L.E. and Knapp, L.K. (1996). Integrating a Geographic Information System, A scientific visualization system, and a precipitation model. *Water Resources Bulletin*, American Water Resources Association, Vol. 32, No. 2, April 1996, 357–369.

He, C., Shi, C., Yang, C., and Agosti, B.P. (2001). A windows–based GIS-AGNPS interface. *Journal of American Water Resources Association*, Vol. 37, No. 2, April 2001, 395–406.

Heaney, J.P., Sample, D., and Wright, L. (1999). *Geographical Information Systems, Decisions Support Systems, and Urban Stormwater Management.* Cooperative Agreement Report No. CZ826256-01-0, U.S. Environmental Protection Agency, Edison, NJ.

Heineman, M. (2000). SWMMTools, ArcScripts Web site (arcscripts.esri.com), ESRI.

Henstridge, F. (1998). Developing the strategic plan for a GIS. *Professional Surveyor*, May/June 1998, 47–48.

Henstridge, F. (1998a). Developing the strategic plan for a GIS — Part 2. *Professional Surveyor*, November/December 1998, 48–49.

Herrlin, J. (2000). Go with the flow — delineating sewerage tributaries. *Geo Info Systems*, March 2000, 34–39.

Hoffman, J. and Crawford, D. (2000). Using comprehensive mapping and database management to improve urban flooding modeling. *Proceedings of the Conference on Stormwater and Urban Water Systems Modeling*, Computation Hydraulics International, Toronto, Canada, February 24–25, 2000.

Horemans, R. and Reynaert, B. (2003). A technical overview: How Pidpa links GIS with ERP, CIS, and SCADA. *ArcNews*, ESRI, Winter, 2003/2004.

Horemans, R. and Reynaert, B. (2002). ArcFM 8 Water: The Solution for the European Drinking Water Company. 17th ESRI European User Conference, Bruges, Belgium, October 14–16, 2002.

Hsu, Y.M. (1999). 3D.Add_XYZ Script, ArcScripts Web site, www.esri.com/arcscripts/scripts.cfm, ESRI.

Huber, W.C. and Dickinson, R.E. (1988). *Storm Water Management Model.* User's manual, Version 4, Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA.

Huey, D.M. (1998). Managing NPDES Permit Compliance: Development of an Integrated GIS Management System, International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

Hutchinson, S. and Daniel, L. (1995). *Inside ArcView.* Onward Press, 329.

Hydrosphere Data Products, 1996. *GeoSelect Pennsylvania*, CD-ROM and User's guide, NCDC Hourly Precipitation GIS, Data Vol. 6, Boulder, CO.

IDRISI 32, (2000). Image Gallery, Clark Labs Web site at http://www.clarklabs.org/03prod/idrisi.htm.

Imwalle, S. (1998). The Four Stages of Water/Sewer GIS Development, International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

Ingeduld, P., Svitak, Z., and Drbohlav, J. (2004). Prague Master Plan of Water Supply System. Case study submission in response to a call for case studies distributed by the author (U.M. Shamsi) for this book.

Irrinki, S. 2000. The digital utility, *Water Environment & Technology* Vol. 12, No. 12, December 2000, Water Environment Federation, 29–33.

Jackson, T.J. (2000). Global soil moisture monitoring with satellite microwave remote sensing, *Water Resources IMPACT*, American Water Resources Association, Vol. 2, No. 5, September 2000, 15–16.

James, W. (1993). Introduction to SWMM. Chapter 1 in *New Techniques for Modelling the Management of Stormwater Quality Impacts*. W. James (Ed.), Lewis, Boca Raton, FL, 1–28.

Jenkins, D. (2002). Making the leap to ArcView 8.1. *Geospatial Solutions*, January 2002, 46–48.

Jenson, S.K. (1991). Application of hydrologic information automatically extracted from digital elevation models. *Hydrologic Processes*, Vol. 5, 31–44.

Jenson, S.K. and Dominique, J.O. (1988). Extracting topographic structure from digital elevation data for Geographic Information System analysis. *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No.11, 1593–1600.

Joffe, B.A., Burnes, T., Jackson, M., Perry, L., Slocum, W., and Wuthrich, D. (2001). All for one and one for all — Consensus builds a successful GIS, *GEOWorld*, June, 2001, 46–49.

Karimi, H.A. and Hammad, A. (2004). *Telegeoinformatics — Location-Based Computing and Services*, CRC, Boca Raton, FL.

Kavaliunas, J. and Dougherty, L. (2000). Census 2000, census sense, part 1, what geospatial pros need to know. *Geospatial Solutions*, July 2000, 40–43.

Kindleberger, C. (1992). Tomorrow's GIS. *American City & County*, April 1992, 38–48.

Kite, G.W. (1995). The SLURP model, Chapter 15 in *Computer Models of Watershed Hydrology*, V.P. Singh (Ed.), Water Resources, 521–562.

Kopp, S.M. (1996). Linking GIS and hydrologic models: Where have we been, Where are we going? In K. Kovar and H.P. Nachtnebel (Eds.), *HydroGIS'96: Application of Geographic Information Systems in Hydrology and Water Resources Managemen*t. International Association of Hydrologic Sciences Publication No. 235. IAHS Press, Wallingford, U.K., 133–139.

Kovar, K., and Nachtnebel, H.P. (Eds.) (1996). *HydroGIS'96: Application of Geographic Information Systems in Hydrology and Water Resources Management.* International Association of Hydrologic Sciences Publication No. 235. IAHS, Wallingford, U.K.

Kuppe, J.B. (1999). Imagery technology helps county make GIS data accessible, *ArcUser*, ESRI, October–December, 1999, 10–11.

Lahlou, M., Shoemaker, L., Paquette, M., Bo, J., Choudhury, S., Elmer, R., and Xia, F. (1996). *Better Assessment Science Integrating Point and Nonpoint Sources BASINS*, User's manual, Version 1, EPA-823-R-96-001, Exposure Assessment Branch, Office of Water, U.S. Environmental Protection Agency, Washington, D.C.

Lamb, K.B., Quinn, P., Romanowicz, R., and Freer, J. (1995). TOPMODEL, Chapter 18 in *Computer Models of Watershed Hydrology*, V.P. Singh (Ed.), Water Resources Publications, 627–668.

Landers, J. (2002). Safeguarding Water Utilities, *Civil Engineering*, Vol. 72, No. 6, June 2002, 48–53.

Lanfear, K.J. (2000). The Future of GIS and Water Resources, Water Resources Impact, American Water Resources Association, Vol. 2, No. 5, September 2000, 9–11.

Lang, L. (1992). Water's New World, *Civil Engineering*, ASCE, June 1992, 48–50.

Laurent, F., Anker, W., and Graillot, D. (1998). Spatial modelling with geographic information systems for determination of water resources vulnerability application to an area in Massif Central (France). *Journal of the American Water Resources Association*, Vol. 34, No. 1, February 1998, 123–134.

Lee, S.G. (1998). Routing Meter Readers … Leveraging an Investment in Data Conversion! International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

Lewis, P. (1985). Beyond description. *Annals of the Association of American Geographers*, Vol. 75, No. 4, 465–477.

Lillesand, T.M. and Kiefer, R.W. (1999). Remote Sensing and Image Interpretation, 4th ed., John Wiley & Sons, New York, 724.

Limp, W.F. (2001). Millennium moves — raster GIS and image processing products expand functionality in 2001. *GEOWorld*, April 2001, Vol. 14, No. 4, 39–42.

Lindley, T.R. and Buchberger, S.G. (2002). Assessing intrusion susceptibility in distribution systems, *Journal AWWA*, AWWA, Vol. 94, No. 6, June 2002, 66–79.

Lior, S., Miller, K., and Thomas, C. (1998). Automating Philadelphia Water Department's Shutoff Program, International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

Logan, B. (2002). The lessons of 9/11. *Geospatial Solutions*, September 2002, 26–30.

Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W. (1999). *Geographic Information Systems*, 2nd ed., John Wiley & Sons.

Lopes, J., Bell, E., and Corriveau, A. (2002). Asset Management Rehaul. *Water Environment & Technology*, Vol. 14, No. 12, December 2002, 40–43.

Lowe, J.W. (2002). Mobile enterprise — act two: the props. *Geospatial Solutions*, April 2002, Vol. 12, No. 4, 44–47.

Lowe, J.W. (2002a). Mobile enterprise — act one: the players. *Geospatial Solutions*, March 2002, Vol. 12, No. 3, 44–47.

Lowe, J.W. (2000). Building a spatial Web site? Know your options. *Geo Info Systems*, April 2000, 44–49.

Lowe, J.W. (2000a). Utilities databases Internet. *Geo Info Systems*, March 2000, 46–48.

Lowe, J.W. (2002b). GIS meets the mapster. *Geospatial Solutions*, Vol. 12, No. 2, February 2002, 46–48.

Luce C.H. (2001). Remote sensing change detection. Book review, *Journal of the American Water Resources Association*, AWRA, Vol. 37, No. 1, 238–239.

Luijten, J.C. (2000). Dynamic hydrological modeling. *ArcUser*, ESRI, July–September 2000, 21–22.

Lundeen, M. (2003). Get Polygon Info Script, ArcScripts Web site, arcscripts.esri.com, ESRI.

Lunetta, R.S. and Elvidge, C.D. (1998). *Remote Sensing Change Detection*, Ann Arbor Press, 121 South Main Street, Chelsea, MI 48118. 318 pages, 1998, ISBN 1-57504-037-9.

Lyman, T.G. (2001). A guide to GPS — learn to obtain better accuracy. *CE News*, May 2001, Vol. 13, No. 4, 52–55.

Maguire, D.J. (1999). ARC/INFO version 8: object-component GIS, Vol. 20, No. 4, Winter 1998/1999, ESRI, 1–2.

Maidment, D.R. (2000). ArcGIS Hydro Data Model, GIS Hydro 2000 preconference seminar, 20th Annual User Conference, ESRI, San Diego, CA, June 2000. http://www. crwr. utexas.edu/giswr/.

Maidment, D.R. (1998). GIS Hydro'98 — Introduction to GIS Hydrology, GIS Hydro 1998 preconference seminar, 18th Annual User Conference, ESRI, San Diego, CA, July 1998.

Maidment, D.R. (1997). Integration of GIS and Hydrologic Modeling, preconference seminar, GIS Hydro97 CD ROM, 1997 User Conference, ESRI.

Maidment, D.R. (1993). GIS and Hydrologic Modeling. Chapter 13 in *GIS and Environmental Modeling*, M.F. Goodchild, B.O. Parks, and L.F. Steyaert (Eds.), Oxford University Press, New York, 147–167.

Martin, L., Castllo, J.G., and Ackerman, D. (1998). Innovative technology combines the components of a sewer system to create virtual computer models. *Proceedings of the 1998 National Conference on Environmental Engineering*, T.E. Wilson (Ed.), ASCE, June 7–10, 1998, Chicago, IL, 669–674.

Maslanik, J.M. and Smith, P.A. (1998). *The City of Harrisburg Water Distribution System Hydraulic Model*, USFilter Engineering and Construction, Pittsburgh, PA.

McKinnon, C. and Souby, J. (1999). Preserving the Wild West, *Imaging Notes*, Space Imaging, Thornton, CO, November/December 1999, 26–27.

Mehan, G.T., III (2003). Monitoring is the key. *Water Environment & Technology*, Vol. 15, No. 11, November 2002, Water Environment Federation, 23–27.

Merry, C.J. (2000). Role of technology in the future of water resources remote sensing developments, *Water Resources Impact*, American Water Resources Association, Vol. 2, No. 5, September 2000, 13–14.

Meyer, S.P., Salem, T.H., and Labadie, J.W. (1993). Geographic information systems in urban storm-water management, *Journal of Water Resources Planning and Managemen*t, Vol. 119, No. 2, 206–228.

Meyers, J. (2002). The worst mistakes in GIS project history (and how to avoid them). *Geospatial Solutions*, Vol. 12, No. 7, July 2002, 36–39.

Michael, J. (1994). Creating digital orthophotos requires careful consideration of project design elements. *Earth Observation Magazine*, Vol. 3 No. 2, 34–37.

Miller, D. (2000). Fine-tuning digital orthophoto quarter quads, *ArcUser*, ESRI, July–September, 2000, 45–47.

Miller, R.C., Guertin, D.P., and Heilman, P. (2004). Information technology in watershed management decision making. *Journal of the American Water Resources Association*, AWRA, April, 2004, Vol. 40, No. 2, 347–357.

Miotto, A. (2000). A better image, *Civil Engineering*, January 2000, 42–45.

Mische, M.R. (2003). Surveying with GPS. *CE News*, November 2003, Vol. 15, No. 10, 40–41.

Mitchell, A. (1997). Zeroing in, ESRI, Redlands, CA, 57–64.

Moglen, G.E. (2000). Effect of orientation of spatially distributed curve numbers in runoff calculations. *Journal of the American Water Resources Association*, AWRA, Vol. 36, No. 6, 1391–1400.

Monmonier, M. (1996). How to lie with maps, University of Chicago Press, 2nd ed.

Morgan, T.R. and Polcari, D.G. (1991). Get set! Go for mapping, modeling, and facility management. *Proceeding of Computers in the Water Industry Conference*, AWWA, Houston, TX.

Morrison, D.A. and Ruiz, M.O. (2001). Instant replay–Red hen's multimedia GIS. *Geospatial Solutions*, October 2001, Vol. 11, No. 10, 45–47.

Murphy, C. (2001). How the discontinuation of SA has affected survey-quality GPS. *CE News*, May 2001, Vol.13, No. 4, 54–55.

Murphy, C. (2000). GIS trends. *CE News*, December 2000, 62–65.

Murphy, C. (2000a). Digital cartographic data — the bottom line, *CE News*, July 2000, 52–56.

Nelson, J., Jones, N., and Miller, W. (1993). Integrated hydrologic simulation with TINs, advances in hydro-science and -engineering, *Proceedings of the 1st International Conference on in Hydro-Science and -Engineering*, Sam S.Y. Wang (Ed.), Center for Computational Hydroscience and Engineering, Washington, D.C., June 7–11, 1993, Vol. I, 571–578.

Neslon, J.E. (1997). From a grid or coverage to a hydrography: unlocking your GIS data for hydrologic applications. *Proceedings of the 1997 ESRI User Conference*, ESRI, San Diego, CA, July 8–11, 1997.

Obermeyer, N.J. (1998). The evolution of public participation GIS, *Cartrography and Geographic Information Systems*, 25(2), 65–66.

Olivera, F. and Bao, J. 1997. ArcView Extension for estimating Watershed Parameters and Peak Discharges According to the TxDOT Statewide Regional Rural Regression Equations, preconference seminar, GIS Hydro97 CD ROM, 1997 User Conference, ESRI.

Ono, T., Cote, L., Dodge, S., and Pier, E. The Wastewater Management System (WIMS), ArcUser calendar insert, ESRI, July 1998.

PaMAGIC (2001). *Local Government Handbook for GIS Implementation within the Commonwealth of Pennsylvania, Part II: GIS Data.* Interim Draft Document, March 2001, Pennsylvania Mapping and Geographic Information Consortium, Lewisburg, PA, www.pamagic.org.

Payraudeau, S., Tournoud, M.G., Cernesson, F., and Picot, B. (2000). Annual nutrients export modelling by analysis of land use and geographic information: case of a small Mediterranean catchment. *Proceedings of the First World Water Congress of the International Water Association*, Book 5, Water Resources and Waste Management, Paris, France, July 3–7, 2000, 320–327.

Peng, Z. (1998). Internet GIS, *CE News*, December 1998, 42–47.

Pfister, B., Burgess, K., and Berry, J. (2001). What's a map? *GEOWorld*, Vol. 14, No. 5, May 2001, 40–43.

Phipps, S.P. (1995). Using raster and vector GIS data for comprehensive storm water management, *Proceedings of the 1995 User Conference*, ESRI. May 22–26.

Pimpler, E. and Zhan, B. (2003). Drop by drop, modeling water demand on the lower Colorado. *Geospatial Solutions*, Vol. 13, No. 5, May 2003, 38–41.

PipeTech Brochure (2000). Peninsular Technologies, 555 Ada Dr., Ada, MI 49301-0728.

PipeTech Website (2001). Peninsular Technologies, www.pent.com.

Prins, J.G., Reil, J., Madison, M., and Swimley, C. (1998) ArcView GIS for City of Stockton's Sewer Model, International User Conference, ESRI, San Diego, CA, July 27–31, 1998.

Quimpo, R.G. and Al-Medeij, J. (1998). Automated watershed calibration with Geographic Information Systems. *Proceedings of the 1st Interagency Hydrologic Modeling Conference*, Las Vegas, NV, 765–772.

Quimpo, R.G. and Shamsi, U.M. (1991). Reliability-based distribution system maintenance, *Journal of Water Resources Planning and Management*, ASCE, Vol. 117, No. 3, May/June 1991, 321–339.

Reichardt, M. (2004). Building the sensor web. *Geospatial Solutions*, March 2004, Vol. 14, No. 3, 36–47.

Reynaert, B., Horemans, R., and Vercruyssen, P. (2004). ArcGIS/ArcFM 8 For Water Distribution Network Management. Case study submission in response to a call for case studies distributed by the author (U. M. Shamsi) for this book.

Rindahl, B. (1996). Analysis of Real-Time Raingage and Streamgage Flood Data Using ArcView 3.0, International User Conference, ESRI, Palm Springs, CA, May 20–24, 1996.

Ritter, K. and Gallie, E.A. (1998). Runoff97: a GIS-based, distributed, surface-runoff model, *Proceedings of the Conference on Stormwater and Related Modeling*: *Management and Impacts*, Computation Hydraulics International, Toronto, Canada, February 19–20, 1998.

Robertson, J.R. (2001). Feeding the flames — airborne imagery fuels GIS. *GeoWorld*, Vol. 14, No. 3, March 2001, 36–38.

Roesner, L.A., Aldrich, J.A., and Dickinson, R.E. (1988). *Storm Water Management Mode EXTRAN Addendum.* User's manual, Version 4, Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA.

Rosenberg, M., Schwind, B., and Lewis, A. (2001). Mapping and monitoring California's vegetation. *Imaging Notes*, Space Imaging, May/June 2001, 24–25.

Rossman, L.A. (2000). *EPANET 2* User's manual, EPA/600/R-00/057, National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH.

Rossman, L.A. (2000a). Computer models/EPANET, Chapter 12 in *Water Distribution Systems Handbook*, L.W. Mays (Ed.), McGraw-Hill, 12.1–12.23.

Rugaard, L.C. (2001). CMOM++ — improving collection system reliability to achieve a goal of no overflows. *Conference Proceedings, 2001 Collection System Odyssey: Combining Wet Weather and O&M Solutions*, Water Environment Federation.

Samuels, W.B., Pickus, J.M., Field, M., Ingman, G., and Welch V. (2000). Watershed data information management in Montana — finding the trouble spots, *ArcNews*, Vol. 22, No. 3, Fall 2000, ESRI, 10.

Schell, T.T., Acevedo, M.F., Bogs, F.C., Newell, J., Dickson, K.L., and Mayer, F.L. (1996). Assessing pollutant loading to Bayou Chico, Florida by integrating an urban stormwater runoff and fate model with GIS. *CD-ROM Proceedings of The 3rd International Conference/Workshop on Integrating GIS and Environmental Modeling*, U.S. National Center for Geographic Information and Analysis, Santa Fe, NM, January 21–25, 1996.

Schilling, K.E. and Wolter, C.F. (2000). Application of GPS and GIS to map channel features in Walnut Creek, Iowa. *Journal of American Water Resources Association*, AWRA, December 2000, 1423–1434.

Schock, M.R. and Clement, J.A. (1995). You can't do that with these data! or: uses and abuses of tap water monitoring analyses. *Proceedings of the National Conference on Environmental Problem-Solving with Geographic Information Systems*, EPA/625/R-95/004, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio, September 1995, 31–41.

Schreppel, C.K. (2003). Distribution system security — setting the alarm for an early warning, *Opflow*, AWWA, Vol. 29, No. 6, June 2003, 1–6.

Schultz, G.A. (1988). Remote sensing in hydrology, *Journal of Hydrology*, Vol. 100, 239–265.

Seiker, F. and Verworn, H.R. (Eds.) (1996). IAHR/IAWQ *Proceedings of the 7th Annual Conference On Urban Storm Drainage*, Hannover, Germany. Vols. I, II, and III.

Sevier, C. and Basford, C. (1996). Hydraulic modeling helps district plan for future, react to challenges. *WaterWorld*, May 1996, 46–47.

Seybert, T.A. (1996). Effective Partitioning of Spatial Data for Use in a Distributed Runoff Model. Ph.D. dissertation, Department of Civil and Environmental Engineering, Pennsylvania State University, State College, Pennsylvania.

Shamsi, U.M. (2002). *GIS Tools for Water, Wastewater, And Stormwater Systems*, ASCE Press, Reston, VA. www.GISApplications.com.

Shamsi, U.M. (2002a). A new pipe replacement approach. *Journal of American Water Works Association*, AWWA, Vol. 94, No. 1, January 2002, 52–58.

Shamsi, U.M. (2001). GIS and Modeling Integration. CE News, Vol. 13, No. 6, July 2001, 46–49.

Shamsi, U.M. (2000). Internet GIS for water industry, Chapter 8 in *Practical Modeling for Urban Water Systems*, Monograph 11, W. James (Ed.), Computational Hydraulics International, Toronto, Ontario, 139–153.

Shamsi, U.M. (1999). GIS and Water Resources Modeling: State-of-the-Art. Chapter 5 in *New Applications in Modeling Urban Water Systems*, W. James (Ed.), Computational Hydraulics International, Guelph, Ontario, Canada, 93–108.

Shamsi, U.M. (1998). ArcView applications in SWMM modeling. Chapter 11 in *Advances in Modeling the Management of Stormwater Impacts*, W. James(Ed.), Vol. 6. Computational Hydraulics International, Guelph, Ontario, Canada. 219–233.

Shamsi, U.M. (1997). SWMM graphics. Chapter 7 in *Advances in Modeling the Management of Stormwater Impacts*. W. James (Ed.), CHI, Guelph, Ontario, 129–53.

Shamsi, U.M. (1996). Stormwater management implementation through modeling and GIS. *Journal of Water Resources Planning and Management*, ASCE, Vol. 122, No. 2, 114–127.

Shamsi, U.M. (1991). Three dimensional graphics in distribution system modeling. *Proceedings of the Specialty Conference on Computers in the Water Industry*, AWWA, Houston, TX, April 1991, 637–647.

Shamsi, U.M., Benner, S.P., and Fletcher, B.A. (1996). A computer mapping program for sewer systems. Chapter 7 in *Advances in Modeling the Management of Stormwater Impacts*, W. James (Ed.), Computation Hydraulics International, Guelph, Ontario, Canada, 97–114.

Shamsi, U.M., Benner S.P., and Fletcher, B.A. (1995). A computer mapping program for water distribution systems, *Distribution System Symposium*, American Water Works Association, Nashville, TN, September 10–13, 1995.

Shamsi, U.M. and Fletcher, B.A. (2000). AM/FM/GIS applications for stormwater systems, Chapter 7 in *Applied Modeling of Urban Water Systems*, Monograph 8, W. James (Ed.), Computational Hydraulics International, Toronto, Ontario, 123–139, 2000.

Shamsi, U.M. and Scally, C. (1998). *An Application of Continuous Simulation to Develop Rainfall versus CSO Correlations*, WEFTEC Asia, Conference and Exhibition on International Wastewater and Water Quality Technology, WEF, Singapore, March 7–11, 1998.

Shamsi, U.M., and Smith, P. (2004). ArcGIS and SWMM5: perfect for each other. *Proceedings of Stormwater and Urban Water Systems Modeling Conference*, Computational Hydraulics International, Toronto, Ontario, February 19–20, 2004.

Singh, V.P. (Ed.) (1995). *Computer Models of Watershed Hydrology.* Water Resources Publications, 1130.

Slawecki, T., Theismann, C., and Moskus, P. (2001). GIS-A-GI-GO. *Water Environment & Technology*, Vol. 13, No. 6, June 2001, Water Environment Federation, 33–38.

Somers, R. (1999). Project milestones: Measuring success, managing expectations. *Geo Info Systems*, October 1999, 20–26.

Somers, R. (1999a). Hire, rent, or train? Making effective investments in GIS expertise. *Geo Info Systems*, May 1999, 20–25.

Space Imaging (2001). Diverse data sources shed light on global change. *Imaging Notes*, Vol. 16, No. 2, March/April 2001, 24–25.

Sponemann, P., Beeneken, L., Fuchs, L., Prodanovic, D., and Schneider, S. (1996). Criteria for geographic information systems used in urban drainage. In F. Seiker and H.R. Verworn (Eds.) (1996) *IAHR/IAWQ Proceedings of the 7th Annual Conference on Urban Storm Drainage*, Hannover, Germany. Vol. III, 1689–1694.

Stalford, R.N. and Townsend, J.A. (1995). Water dstribution system modeling, GIS, and facility management systems, what do they have in common? A case study. *Proceedings of the 1995 International User Conference*, ESRI, San Diego, CA.

Stern, C.T. and Kendall, D.R. (2001). Mobile solutions deliver improved utility infrastructure and operations management. *Journal AWWA*, American Water Works Association, November, 2001, Vol. 93, No. 11, 69–73.

Stone, D. (1999). County reaps rewards from airborne laser survey, *CE News*, November 1999, 58–60.

Stupar, D., Heise, S. and Schielbold, J. (2002). Ocean county utilities authority takes a business-driven approach to GIS design and implementation. *Water Writes*, ESRI, Winter 2002, 6–7.

Suchanek, M. and Metelka, T. (2004). Sofia Drainage Area Plan. Case study submission in response to a call for case studies distributed by the author (U. M. Shamsi) for this book.

Swalm, C., Gunter, J.T., Miller, V., Hodges, D.G., Regens, J.L., Bollinger, J., and George, W. (2000). GIS blossoms on the mighty Mississippi. *Geo Info Systems*, April 2000, 1998, 30–34.

Taher, S.A. and Labadie, J.W. (1996). Optimal design of water-distribution networks with GIS, *Journal of Water Resources Planning and Management*, ASCE, Vol. 122, No. 4, July/August 1996, 301–311.

TenBroek, M.J. and Roesner, L.A. (1993). MTV — Analysis tool for review of computer models. *Proceedings of the Conference on Computers in the Water Industry*, Water Environment Federation, Santa Clara, CA, August 8–11, 1993, 173–181.

Teo, Frederick K.S. (2001). For Tacoma water GIS is a wireless wireless wireless world. *Geospatial Solutions*, 40–45, February 2001.

The San Diego Union Tribune (1998). GIS — New Technology's Scope is Almost Infinite. Sunday, July 26, 1998, A-12.

Thoen, B. (2001). Maximizing your search for geospatial data. *GEOWorld*, Vol. 14, No. 4, April 2001, 34–37.

Thomas, G. (1995). A Complete GIS-based Stormwater Modeling Solution, *Proceedings of the 1995 User Conference*, ESRI, May 22–26.

Thompson, S. (1991). Mapping the latest trends in GIS. *American City & County*, May 1991, 26–38.

Thrall, G.I. (1999). Low cost GIS, *Geo Info Systems*, April 1999, 38–40.

Thrall, G.I. (1999b). Want to think spatially? Raster GIS with MFworks, *Geo Info Systems*, September 1999, 46–50.

Thrall, G.I. (2003). Laptops for GIS warriors, *Geospatial Solutions*, Vol. 13, No. 10, October 2003, 49–52.

Thuman, A. and Mooney, K. (2000). TMDLs: How did we get here and where are we going? Environmental Perspectives, *The HydroQual Newsletter*, Fall 2000, Vol. 5, 1–4.

Tiewater, P. (2001). GASB statement 34: Not just for accountants and engineers. *Presentation in the 3rd Annual Sewer Rehabilitation Conference*, Three Rivers Wet Weather Demonstration Program, Warrendale, PA, September 25–26, 2001.

TOPAZ (2000). Topographic parameterization software, Grazing Labs Research Laboratory Website at http://grl.ars.usda.gov/topaz/TOPAZ1.HTM.

Trammel, D. (2004). Resource Planning and CIP Allocation. Case study submission in response to a call for case studies distributed by the author (U. M. Shamsi) for this book.

Turner, A.K. (2001). The future looks bright (in any spectrum) for GIS data. *GeoWorld*, Vol. 14, No. 5, May 2001, 30–31.

Turner, B.G., Arnett, C.J., and Boner, M.C. (2000). Wet weather monitoring and calibration of USEPA BASINS model provides basis for watershed management framework compliance with TMDL allocations. *Proceedings of the 1st World Water Congress of the International Water Association*, Book 5, Water Resources and Waste Management, Paris, France, July 3–7, 2000, 439–446.

Understanding GIS — The ARC/INFO Method, 1990. ESRI, Redlands, CA.

U.S. Department of Agriculture (1986). *Urban Hydrology for Small Watersheds*. Technical Release 55, Soil Conservation Service, U.S. Department of Agriculture, Washington, D.C., 2.5–2.8.

USGS (2000). USGS Digital Elevation Model Data, USGS Web site at http://edcwww.cr.usgs.gov/glis/hyper/guide/usgs_dem.

Venden, E. and Horbinski, T. (2003). Gurnee streamlines notification process. *Geospatial Solutions*, January 2003, Vol. 13, No. 1, 26–27.

Walski, T. (2001). Importance and accuracy of node elevation data. *Current Methods*, Vol. 1, No. 1, Haestad, Waterbury, CT.

Walski, T.M. and Male, J.W. (2000). Maintenance and rehabilitation replacement. In *Water Distribution Systems Handbook*, L.W. Mays Ed., McGraw-Hill, 17.1–17.28.

Walski, T., Toothill, B., Skoronski, D., Thomas, J., and Lowry. S. (2001). Using digital elevation models. *Current Methods*, Haestad, Vol. 1, No. 1, 91–99.

Wang, M., Hjelmfelt, A.T., and Garbrecht., J. (2000). DEM aggregation for watershed modeling. *Water Resources Bulletin*, American Water Resources Association, Vol. 36, No. 3, June 2000, 579–584.

Water Resources Bulletin, American Water Resources Association, Vol. 36, No. 3, June 2000, 579–584.

Waters, N. (2001). Network query language is the new, new thing in GIS. *GEOWorld*, April 2001, Vol. 14, No. 4, 30.

Wells, E. (1991). Needs analysis: Pittsburgh's first step to a successful GIS. *Geo Info Systems*, October 1991, 30–38.

Whittemore, R.C. and Beebe, J. (2000). EPA's BASINS model: Good science or serendipitous modeling? *Journal of the American Water Resources Association*, June 2000, Vol. 36, No. 3, 493–499.

Wild, M. and Holm, D. (1998). Consensus building at superfund sites using GIS. *Proceedings of the 1998 National Conference on Environmental Engineering*, T.E. Wilson (Ed.), ASCE, June 7–10, 1998, Chicago, IL, 326–331.

Wilson, J.D.P. (2001). The next frontier—GIS empowers a new generation of mobile solutions. *GEOWorld*, June 2001. Vol. 14, No. 6, 36–40.

Wilson, J.P., Mitasova, H., and Wrightang, D. (2000). Water resources applications of GIS. Articles currently under peer review by the URISA Journal, University Consortium for Geographic Information Science (UCGIS) Web site, www.ucgis.org/apps_white/water.html.

WinCan Web site (2001). WinCan America, www.wincanamerica.com.

WMS (2000). Watershed Modeling System Web site, http://emrl.byu.edu/wms.htm, Environmental Modeling Research Laboratory (EMRL), Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT.

Zeiler, M. (1999). *Modeling Our World*, ESRI Press, ESRI, Redlands, CA, 199 p.

Zimmer, R.J. (2001). GIS vision. *Professional Surveyor*, November 2001, Vol. 21, No. 10, 38–39.

Zimmer, R.J. (2001a). DEMs — A discussion of digital elevation data and how that data can be used in GIS. *Professional Surveyor*, Vol. 21, No. 3, March 2001, 22–26.

Zimmer, R.J. (2001b). GPS resource mapping for GIS. *Professional Surveyor*, Vol. 21, No. 4, April 2001, 44–47.

Zimmer, R.J. (2000). Parcel mapping part I. *Professional Surveyor*, Vol. 20, No. 9, October 2000, 52–55.

Index

Page numbers marked with the suffix n indicate that the index entry is located in the note on that page

3D Analyst, 23, 37, 103–105, 170, 207–208
3RWWDP, 184–186
4M applications, 6
8-direction pour point model, 86–87

A

Access, Microsoft
  ArcView and, 174–175
  AutoCAD and, 313–314
  Avenue and, 174–175
  CCTV and, 349
  demand and, 252–253
  format for, 196
  InfoWorks and, 287
  MapSheets and, 68
  NPDES permits and, 174–175
  PCSWMM GIS and, 281
  SRDMS and, 331
  SWMM and, 259
ACRIMSAT, 53
Adam County (Illinois), 2
Adjacent, 139, 146
Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), 54, 65
Advantica Stoner, 229, 232, 237
ADWR, 62
Aerial Images' TerraServer, 59, 73
Aerial photographs
  altitude of, 49, 50
  for base maps, 143
  definition of, 49
  DOQs and, 60
  format for, 139
  image displacement in, 59
  LULC maps, creation of, 63
  planimetric maps and, 143
AGNPS, 220–221
Agricultural areas, 63, 199, 200
Agricultural Nonpoint Source Pollution Model (AGNPS), 220–221
AGSWMM, 280–281
Air Weather Service (AWS), 181
Airborne Laser Terrain Mapping (ALTM), 85
Alaska, DEM data for, 83
Albany County (Wyoming), 14–15, 48–49
Albuquerque (New Mexico), 336–337
ALERT Flood Warning System, 178–179
Algae, 360
Allegheny County (Pennsylvania)
  3RWWDP, 184–186
  planimetric maps of, 143, 145
  precipitation atlas of, 185
  sewage system, 159, 170, 293, 331
ALTM, 85
AM, 298–301
AM/FM, *see* Automated mapping/facilities management (AM/FM)
AM/FM/GIS
  vs. AM, 298
  applications for, 11–12
  data mining software and, 122
  Denver case study, 229
  features of, 298, 300
  vs. FM, 298
  vs. GIS, 298
  relational databases and, 122, 301
  SAMS project and, 290
  types of software for, 301
  work orders and, 326
American Society for Photogrammetry and Remote Sensing (ASPRS), 49–50
American Society of Civil Engineers (ASCE), 2, 296–297
American Standard Code for International Interchange format, 112, 265

American Water Resources Association (AWRA), 5
American Water Works Association (AWWA), 5, 296
AMSA, 342, 343
Antwerp province (Belgium), 188, 356–358
Application programming, 20–24
Applications, definition of, 3, 6
ARC/HEC-2 program, 219–220
Arc Macro Language (AML), 21, 24, 93, 152
ArcCAD software, 154
ArcCatalog, 227
ArcExplorer software, 70, 124
ArcFM software, 151, 236, 301–303, 356–358
ArcGIS software
  AGSWMM and, 280–281
  Buncombe County case study, 347–352
  CCTV and, 321, 349
  Cityworks and, 304, 305, 326
  data conversion and, 150
  DEM and, 89, 90
  demand and, 352
  extensions and, 23
  external programs and, 23
  Hyperlink function in, 167
  InfoWater and, 234
  parcel selection in, 163
  Pidpa case study, 356–358
  pilot project, use in, 40
  pipe rating and, 349
  scripts in, 21
  SewerCAD and, 289
  Spatial Analyst and, 23, 90–92
  SRDMS and, 331
  storm runoff fees and, 174
  SWMM and, 280–281
  Tucson case study, 352–353
  Utility Network Analyst and, 327–328
  Water Utilities Data Model, 227
  WaterCAD and, 227
  WaterGEMS and, 236
  weather data and, 183
ArcGRID extension
  DEM and, 89, 90
  features of, 93
  images in, 51
  LULC mapping and, 64
  vs. Spatial Analyst, 90
  weather data and, 183
ArcIMS software
  H_2OVIEW and, 234
  for Internet GIS, 122, 124
  mobile GIS and, 133
  MrSID and, 70
  pilot project, use in, 40
  SCADA and, 188
ArcInfo software
  ArcFM and, 303
  ArcGrid extension for, 64, 93
  AVNET and, 238
  Data Explorer and, 212
  decision matrix for, 36–37
  DEM and, 89, 90
  FILL function in, 87
  Geo-STORM and, 219
  GISHydro modules and, 209
  HEC programs and
    HEC-1/-2 and, 210–211, 219–220
    HEC-GeoHMS and, 76–77
    HEC-GIS and, 208
  HYDRA and, 291
  for mapping, 152
    MMA and, 169
    pipelines, 69
    Ramsey, 157–158
    San Diego, 63
    of vegetation, 64
  MIKE NET and, 236
  for monitoring, 62, 178–179, 194–195
  MrSID and, 70
  RAS and, 208
  SAPAL and, 15
  scripts in, 21
  SRDMS and, 331
  SWMM and, 290–291
  SWMMDUET and, 283–284
  topology and, 154
  WMS and, 208
  XP-SWMM and, 289–290
ArcMap software
  isolation trace and, 328
  network analysis and, 227, 327–328
  storm runoff fees and, 174
  Tucson case study, 352–353
  water editing tool and, 152
ArcObjects software, 24, 37, 328
ArcPad software, 70, 131–132
ArcScripts, 21, 23, 24
ArcSDE software, 40, 174
ArcView software
  3D Analyst, *see* 3D Analyst
  absolute pathnames in, 320–321
  Access and, 174–175, 196
  AGSWMM and, 280
  Avenue and, 174–175, 270–271
  AVNET and, 237–241
  AVSWMM and, 270–278
  AVWater and, 237
  BASINS and, 213–216

CCTV and, 319–321
Cityworks and, 304, 326
curb length and, 201
decision matrix for, 36–37
DEM and, 89, 90
demand and, 252–253
demographic data in, 203
DLLs and, 24
DOQQs and, 61, 330
drawing tool in, 198
EPANET and, 237–238
features of, 237–238
GBA Master Series database and, 294–295
Geographic Transformer and, 71
GISHydro modules and, 209–210
GISRed and, 364
GPS data and, 115
HEC-GeoHMS and, 76–77, 207
Hotlink function, 167
HYDRA and, 291
hydrology extension in, 23, 90–92
Image Analysis and, 66, 69
InfoNet and, 287
InfoWorks and, 237, 287
isolation trace and, 328
MapObjects and, 294–295
for mapping, 152, 157–158, 167–168
master plan case study, 364
measurement tool in, 198
MediaMapper and, 322
MIKE BASIN and, 218
MIKE NET and, 236, 355–356
MIKE SWMM and, 286
for monitoring, 178–179, 187
MontanaView, 127–128
Mouse GIS and, 285–286
Mouse GM and, 286
MrSID and, 70
NAD27 vs. NAD83, 61
NPDES permits and, 174–175, 187, 189
NPSM and, 221
ODBC features of, 196–197
RAS and, 207–208
SAPAL and, 15
scripts in, 21
Spatial Analyst and, 90–92
SRDMS and, 331–333
SWMM and, 270–271, 277, 280
SWMMTools and, 278–280
WaterCAD and, 105, 235
WMS and, 208
Working for Water and, 15
Arizona
satellite imagery of, 62
Tucson case study, 352–353
Army Corps of Engineers, 208; *see also* Hydrologic Engineering Center (HEC)
ASCE, 2, 296–297
ASCII, data in, 112, 265
ASPEN software, 115
ASPRS, 49–50
Asset management systems, 295, 310–311, 342
Asset Surveyor, 115
Association of Metropolitan Sewerage Agency (AMSA), 342, 343
ASTER, 54, 65
Asymetrix Corporation's Digital Video Producer, 315
Atlanta (Georgia), 311
Attributes, definition of, 34
Audio recordings, 321, 322
Aurora (Colorado), 178–179
Australia, 15, 73
AutoCAD software
Access and, 313–314
GPS data and, 115
H_2ONET and, 233
Internet GIS and, 125, 127
Map GIS and, 313
MIKE NET and, 355
SewerCAD and, 289
SRDMS and, 331
StormCAD and, 289
WaterCAD and, 235
Autodesk
AutoCAD, *see* AutoCAD software
AutoLISP, 21
MapGuide, 120, 124, 127
MrSID and, 70
OnSite software, 132
AutoLISP software, 21
Automated mapping (AM), 298–301
Automated mapping/facilities management (AM/FM)
features of, 298, 300
vs. GIS, 146–147, 298
work orders and, 300, 301, 326
Automated mapping/facilities management/geographic information system, *see* AM/FM/GIS
Avenue scripting language
Access and, 174–175
applications development and, 24
ArcView and, 174–175, 270–271
CCTV and, 21, 320–321
definition of, 21
DLLs and, 24
GISHydro modules and, 209
interface development and, 270–271
NPDES permits and, 174–175

SWMM and, 270–271
water editing tool and, 152
AVNET program, 237–241
AVSand software, 261, 284
AVSWMM, 270–278
AVWater software, 232, 237
AWRA, 5
AWS, 181
AWWA, 5, 296
Azteca Systems
Cityworks, *see* Cityworks software
Web site for, 301

B

Back-end interface, *see* Postprocessor
Baldwin, Borough of, 330–331
Base maps, 142–145
BASINS program, 213–218, 222, 282–283
Bayou Chico (Florida), 290–291
Bays, modeling of, 219
Belgium, Antwerp province, 188, 356–358
Bentley Systems, 235; *see also* Haestad Methods
Bethel Park (Pennsylvania), 161, 170
Bethlehem (Pennsylvania), 188
Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) program, 213–218, 222, 282–283
Blacksburg (Virginia), 128
Bloomington (Indiana), 45
Blue Marble Geographics software, 67, 71–72
Bluetooth, 133
Bore holes, cost of, 13
BOSS International, 236
Boston (Massachusetts)
asset management system for, 311
catch basins in, 311
CCTV and, 314
drinking water in, 296
GIS implementation in, 13
sewer system in, 290
"Bottom-up" approach, definition of, 12
Boulder (Colorado), 217–218
Buffalo Bayou Watershed (Texas), 76–77
Buffers, 164–166, 255
Bulgaria, 346–347
Bull Run Watershed (Pennsylvania), 97–101
Buncombe County (North Carolina), 309, 347–352

C

CAFM, 298
CaiCE Software's Visual SWMM, 266–267
CALAMAR system, 185
California
Callegaus Municipal Water District, 133
Lake Elsinore, City of, 158–159
Los Angeles, 177
Los Angeles County, 138–139, 189
LULC data on, 64
Mariposa East, 75
San Diego, 13, 24, 62–63
San Francisco Bay area, 313–314
San Jose, 328
San Mateo County, 45
Stege Sanitary District, 313–314
Stockton, 291
Ventura, 62
California, Borough of (Pennsylvania), 125
California Gulch Superfund Site (Colorado), 62
Callegaus Municipal Water District (California), 133
Camp, Dresser, and McKee (CDM), 285
Canada, Toronto, 190
Capacity, management, operations, and maintenance (CMOM), 260, 312–314
Capital asset inventories, 295, 312
Capital improvement planning (CIP), 311
CarteGraph Systems, 301
CARTERRA, 49, 58
CASS WORKS
MapObjects and, 24
SAMS project and, 290
Web site for, 301
Casselberry (Florida), 188
CASSView software, 24
Catchments, *see also* Watersheds
in Boston, 311
in data dictionary, 157
definition of, 259
in DEMs, 86
mapping of, 142, 157
Mouse and, 285
storm runoff taxes and, 174
surveying of, 112
SWMMTools and, 280
types of, 259
CDs, 314–317
CEDARE, 15
CEDRA Corporation
AVSand, 261, 284
AVWater, 232, 237
Web site for, 232
Cell threshold, 86
Census Bureau
polygon layers from, 251n
TIGER data, 145, 203, 291
Center for Environment and Development for the Arab Region and Europe (CEDARE), 15

Center for Preparation and Implementation of International Projects (Russia), 15
Center for Research in Water Resources (CRWR), 91, 105, 208–210, 216
Change detection, 64
Channels, *see also* Streams
  ArcGRID extension and, 93
  culverts, 77
  DEMs and, 86
  flood profiles for, 211
  flow in, types of, 210
  flow in, velocity of, 211
  GPS mapping of, 108–109
  hydraulic characterization of, 179
  LIDAR and, 85
  roughness coefficient of, 205–206
  TOPAZ software and, 95
Charleroi, Borough of (Pennsylvania), 101–102
Charlotte (North Carolina), 14, 31, 230
Chatham County (Georgia), 31, 85
Chemical oxygen demand (COD), 221
Cherokee Metropolitan District (Colorado), 11
Chesapeake Bay, Landsat image of, 47
Chester Engineers
  AVNET, 237–241
  Internet GIS and, 125
  needs analysis by, 29
  pilot project by, 44
  SRDMS, 331
  water editing tool for maps, 151–152
  Wyemaker, 171–173
Chlorination, 355
CIFM, 298
Cincinnati (Ohio)
  cost-benefit analysis, 13
  Internet GIS for, 122
  MMA development, 169
  storm sewer system in, 11
  vulnerability assessments, 338
  water infrastructure project in, 31
CIP, 311
CITIPIX software, 59
Cityworks software
  ArcFM and, 301
  ArcGIS and, 304, 326
  ArcView and, 304, 326
  features of, 304–305
  isolation trace and, 328
  Oracle and, 304
  SQL and, 304
  Sybase and, 304
  Web site for, 301
  work orders and, 304, 326
Clark University's IDRISI software, 51, 89, 94–95
Clarke spheroid, 140
Clean Water Act
  3RWWDP and, 184–186
  BASINS model and, 218
  Internet GIS and, 127
  TMDLs and, 218
Cliffs, DEM data for, 77
Closed-circuit television (CCTV)
  Access and, 349
  ArcGIS and, 349
  Cityworks and, 304, 326
  equipment for, 314–319
  GBA Master Series database and, 294
  GIS links to, 21, 319–321
  GPS and, 322
  for inspections, 179, 260, 307, 314, 323–324
  mapping and, 321–322, 324–325
  physical characterization by, 179
  pipe rating and, 309, 347–352
  PipeTech and, 319
  SRDMS and, 331–333
  VMS and, 322
  WinCan and, 317–319
Clouds, 54
CMMS, 295, 298, 311
CMOM, 260, 312–314
COD, 221
Colorado
  Aurora, 178–179
  Boulder, 217–218
  California Gulch Superfund Site, 62
  Cherokee Metropolitan District, 11
  Colorado Springs, 68–69, 183, 184
  Denver, 174, 229, 312
  Gunnison River Basin, 212
  Leadville, 62
  Loveland, 69
Colorado Springs (Colorado), 68–69, 183, 184
Columbus (Georgia), 218
COM, 237
Combined sewer overflow (CSO)
  BASINS model and, 218
  control facilities, 263
  cost of, 296
  in data dictionary, 35
  diversion chamber, 293
  inflow, infiltration, and, 312
  LTCP, 260, 263
  mapping of, 159, 166, 260
  NMC, 260, 312
  NPDES permits and, 187, 260

Physical Characterization data, 179, 312
Plan of Actions, 263
SHC, 260
SWMM and, 263
System Inventory, 312
Compact disks (CDs), 314–317
Compaq's TerraServer, 59, 73
Component Object Model (COM), 237
Composite Reflectivity, 182
Computational Hydraulics
PCSWMM GIS, 261, 262, 281–282
Web site for, 261
Computer-aided design (CAD)
AM and, 298
FM and, 300
vs. GIS maps, 146–147
Computer-aided facility management (CAFM), 298
Computer-aided mapping (CAM), 298–301
Computer integrated facility management (CIFM), 298
Computerized maintenance management system (CMMS), 295, 298, 311
Conduits
AVSWMM and, 274–276
channels, *see* Channels
culverts, 77
gutters, 291
pipes, *see* Pipes
rivers, *see* Rivers
streams, *see* Streams
in XP-SWMM, 266
Containment, 139
Contour-based structures, 77
Coordinate system, definition of, 140
Copper, 190
Cost savings factor, 13
Coverage, 53, 148
CRWR, 91, 105, 208–210, 216
Crystal Decisions, 189, 190
Crystal Reports, 189
CSO, *see* Combined sewer overflow (CSO)
Culverts, 77
Customers
addresses of, 149, 190, 356
AVSWMM and, 271–272
consumption rate, 249–253, 352–353
critical service, 327
demographic data on, 202–203
service shutoff, 167–169
skeletonization and, 247
in SWMM, 267, 269
types of, 147, 249, 327
Czech Republic, 358–360

D

Danish Hydraulic Institute (DHI)
company, data on, 236, 284
MIKE BASIN, 218–219
MIKE NET, 232, 236–237, 355–356, 358–360
MIKE SHE, 219
MIKE SWMM, 261, 285, 286–287
Mouse, 261, 284–287, 346–347
Web site for, 232
Data
collection of, for maps, 148
conversion of, 33–34, 139, 141, 148–154
errors in, 141
formats for, 342–343
for monitoring, 179
processing of, for maps, 153–155
QA/QC of, for maps, 155
Data dictionary, 35, 112–113, 157
"Data-driven" approach, 12
Data Explorer software, 212, 289–290
Data Storage System (DSS), 76–77
Database, design of, 34–36
Database management system (DBMS), 35, 154–155
Datum, 140
DBMS, 35, 154–155
Decision support system (DSS), 261
D-8 model, 86–87
Delaware Department of Natural Resources, 283–284
Demand, 249–253, 352–353, 356
Demand Allocator module, 235
Demographic data, 202–203
DEMs, *see* Digital elevation models (DEMs)
Denver (Colorado), 174, 229, 312
Depressions, 87, 95
DFIS, 15
DHI Water and Environment, *see* Danish Hydraulic Institute (DHI)
Dialog Designer, 152, 174–175
Digital elevation models (DEMs)
3D visualization and, 79–80, 89, 103–105, 170, 253–254
accuracy of, 80, 254
aggregation of data in, 88
applications for, 79, 81
ArcGrid and, 89, 90, 93
BASINS program and, 216
boundary "cross-over" problem in, 98, 100–101
for Buffalo Bayou Watershed, 76–77
cost of, 330
definition of, 77
D-8 model, 86–87
depressions in, 87, 95

EROS and, 84
file format for, 82–83
flat areas in, 87–88
flooding and, 259
HEC programs and, 77, 208
Hydrologic Modeling Dialogue and, 91
IDRISI software and, 89, 94–95
IFSAR and, 85–86
LIDAR and, 62, 84–85
NED and, 83
procedures for, 254
resolution of, 80–81
satellite imagery and, 51
SDTS and, 83
sewers in, 101–103, 260
sinks in, 87, 95
SLOPE function in, 88
software for, 88–93
Spatial Analyst and, 89–92, 97–101
SRTM and, 84
streams in, 86–88, 95–101
structure of, 77
TOPAZ and, 89, 95
TRex Wizard and, 254–255
USGS and, 81–84, 89
water systems and, 103
watersheds and, 80–81, 95–101
Web sites for, 91, 105
Digital line graph (DLG), 76, 80
Digital orthophoto quadrangles (DOQs), 48, 58, 60–61, 141
Digital orthophoto quarter quadrangles (DOQQs), 48, 60–62, 330
Digital orthophotos (DOPs), 59–60, 143, 183, 254
Digital raster graphics (DRGs), 141–142, 144, 241, 330
Digital surface models (DSMs), 77
Digital terrain models (DTMs), 77, 170, 207, 208
Digital Video Producer, 315
Digitization, 149
Digitized Facilities Information System (DFIS), 15
Distributed models, 196
District of Columbia
DOPs of, 70
Pentagon, attack on, 337
VSAT and, 342
WSSC, 40
DLG, 76, 80
DLL, 23–24, 232
DOPs, 59–60, 143, 183, 254
DOQs, 48, 58, 60–61, 141
DOQQs, 48, 60–62, 330
Dover (United Kingdom), 14
Dowagiac River Watershed (Michigan), 220–221
DRGs, 141–142, 144, 241, 330
Drinking Water Research Division's EPANET, 231–232, 236–241, 339, 364
DSMs, 77
DSS (Data Storage System), 76–77
DSS (Decision support system), 261
DTMs, 77, 170, 207, 208
DTN Weather Services, 183
Dublin (Ohio), 120–121
DVDs, 314–316
Dye tests, 260
Dynamic link library (DLL), 23–24, 232

E

Earth Observing System (EOS)
ACRIMSAT, 53
Hyperion spacecraft, 65
image processing by, 84
Landsat, *see* Landsat
QuickSAT, 53
Terra, *see* Terra satellite
time-series data from, 72
Earth Resource Data Analysis System (ERDAS)
company, data on, 66–68, 115
IMAGINE, *see* IMAGINE software
MapSheets, 68
Stereo Analyst, 69
Web site for, 67
Earth Resource Mapping's ECW format, 70
Earth Resources Mapping's ER Mapper, 51, 67
Earth Resources Observation Systems (EROS), 84
EarthExplorer, 61, 73, 105
EASI/PACE software, 67, 71
Eastman Kodak's CITIPIX software, 59
Echo Tops, 182
ECW format, 70
Electric utilities, landbase accuracy for, 141
Elevation, 83, 103–105, 253–254, 259; *see also* Digital elevation models (DEMs)
Emergency Response Plans (ERPs), 337, 342
EMRL's WMS, 91, 208, 222
Engineering departments
applications for, 30
base map accuracy for, 142
landbase accuracy for, 141
needs matrix for, 34
Enhanced Thematic Mapper Plus (ETM+), 54–55
Enterprise relational database, *see* Relational databases

ENVI software, 51, 65, 67
Environmental Modeling Research Laboratory's WMS, 91, 208, 222
Environmental Protection Agency (EPA)
  BASINS program, 213–218, 222, 282–283
  CSO, *see* Combined sewer overflow (CSO)
  EPANET, 231–232, 236–241, 339, 364
  HSPF model, 214–216
  MS4 permits, 173
  national water quality database, 127
  NPDES permits, *see* National Pollution Discharge Elimination System (NPDES) permits
  river reach files from, 76
  security grants from, 337
  on sewage overflows, 159
  STORET, 127
  SWMM, *see* Storm Water Management Model (SWMM)
  on TMDLs, 214
  water quality monitoring by, 64
  Web site for, 159
Environmental Systems Research Institute (ESRI)
  AML, 21, 24, 93, 152
  application programming and, 21–24
  ArcCAD, 154
  ArcCatalog, 227
  ArcExplorer, 70, 124
  ArcFM, 151, 236, 301–303, 356–358
  ArcGIS, *see* ArcGIS software
  ArcGRID, *see* ArcGRID extension
  ArcIMS, *see* ArcIMS software
  ArcInfo, *see* ArcInfo software
  ArcMap, *see* ArcMap
  ArcObjects, 24, 37, 328
  ArcPad, 70, 131–132
  ArcScripts, 21, 23, 24
  ArcSDE, 40, 174
  ArcView, *see* ArcView software
  Avenue, *see* Avenue scripting language
  company, data on, 6
  Hydro extension, 89
  Image Analysis, 66, 67, 69
  MapObjects, 24, 70, 294–295, 311
  pilot project, software for, 38
  site plan file format, 42
  Spatial Analyst, *see* Spatial Analyst
  Water Distribution Data Model, 40
  water editing tool and, 151–152
  Web site for, 24
EOS, *see* Earth Observing System (EOS)
EPA, *see* Environmental Protection Agency (EPA)
EPANET software, 231–232, 236–241, 339, 364
ER Mapper software, 51, 67
ERDAS, *see* Earth Resource Data Analysis System (ERDAS)
EROS, 84
ERPs, 337, 342
ERS, 71
Escambia County (Florida), 290–291
ESRI, *see* Environmental Systems Research Institute (ESRI)
Estuaries, 219
ETM+, 54–55
"Euclidean Distance" approach, 102
Evapotranspiration, 66
Everglades (Florida), 62
Excel, 68, 185, 236
Extensible Markup Language (XML), 16
Extensions, 21–22, 66
External programs for GIS, 23–24

F

FAA, 181
Facilities management (FM), 29, 35, 298, 300, 301; *see also* Asset management systems
Fairfax County (Virginia), 312
Federal Aviation Administration (FAA), 181
Field GIS, *see* Mobile GIS
Fieldport, 133
Fire hydrants, *see* Hydrants
Fitchburg (Massachusetts), 116
Flat areas, in DEMs, 87–88
Fleet management, 111, 130
Flexidata pipe survey software, 316–317
Floodplains, HEC programs and, 219–220
Floods, 178–179, 181–182, 259
Florida
  Bayou Chico, 290–291
  Casselberry, 188
  Escambia County, 290–291
  Everglades, 62
  Miami-Dade, 186
  Pensacola Bay area, 290–291
Flowmeters, 179–180
FLY!, 71
Forests, 63, 199
Fort Wright (Kentucky), 289–290
FoxPro software, 289–290
FRAMME software, 301
France
  Massif Central, 92
  SPOT satellite, *see* SPOT satellite
Front-end interface, *see* Preprocessor

G

Gas, natural, 10, 141
GASB, 295, 312, 333
GBA Master Series, 294–295, 301
GE Smallworld software, 70
GEMS
  SewerGEMS, 289
  WaterGEMS, *see* WaterGEMS software
Generic Database, 71
Genesee County (Michigan), 229
GenScn tool, 216
Geo-STORM, 219
GeoCommunity's SDTS, 83
GeoCover LC, 63
Geoexplorer GPS, 114–115
GeoGateway, 71
Geographic Calculator, 71–72
Geographic Engineering Modeling Systems (GEMS)
  SewerGEMS, 289
  WaterGEMS, *see* WaterGEMS software
Geographic Information System for Montgomery and Prince George's Counties (GeoMaP), 40
Geographic information system (GIS)
  advantages of, 7–9
  vs. AM, 298
  vs. AM/FM, 146–147, 298
  applications for, 9–11
    evolution of, 16
    sewer systems, 259–260
    technical specifications and, 43–44
    water systems, 5–6, 227–228
  benchmarks for, 31
  benefit-cost ratio of, 8
  definition of, 3
  disadvantages of, 12
  vs. FM, 298
  future of, 18–19
  hardware for, 37–39
  history of, 4–6
  literature on, 5
  milestones for, 41
  needs analysis and, 28–46
  personnel for, 41–43
  pilot project for, 40–41
  related technologies, 49
  remote sensing and, 48–73
  resource inventory for, 32–33
  segmentation of, 235
  skeletonization and, 243–249
  vulnerability assessments, 338–339
Geographic Transformer, 67, 71
Geographically Referenced Information Program (GRIP), 120
Geography Markup Language (GML), 16, 190
GeoMAP, 40
Geomatica software, 51, 67, 70–71
GeoMedia, 124–125, 150, 287
GeoPlan software, 301
Georgia
  Atlanta, 311
  Chatham County, 31, 85
  Columbus, 218
  Middle Chattahoochee River Watershed, 218
Geospatial Information and Technology Association (GITA), 10, 141, 159
Geospatial Solutions magazine, 188–189, 190
GeoWasteWater software, 301
GeoWater software, 301
Germantown (Tennessee), 226–227, 228
GIF images, 195
GIS, *see* Geographic information system (GIS)
GISapplications.com Web site, 364
GISHydro modules, 208–210, 222
GISRed, 362–364
GITA, 10, 141, 159
Global Hawk aircraft, 72–73
Global Navigation Satellite System (GLONASS), 109
Global positioning system (GPS)
  accuracy of, 109–110, 113–114, 116, 254
  alternatives to, 111, 311
  applications for, 110
  CCTV and, 322
  data, format for, 112
  definition of, 109
  equipment, 113–117
  for fleet management, 111
  Image Analysis and, 69
  modes of, 110
  SA error in, 110
  for surveys, 111, 112–116, 133–134
  VMS and, 322
GLONASS, 109
Gloucester (Massachusetts), 296
GML, 16, 190
Goodwater Creek (Missouri), 88
Government Accounting Standards Board (GASB), 295, 312, 333
GPS, *see* Global Positioning System (GPS)
Graphical user interface (GUI), 20, 38–40, 264–266

Graphics Interchange Format (GIF) images, 195
Grasslands, 63, 199
Great Salt Lake (Utah), 87
Green-Ampt parameters, 269–270
Greenville (North Carolina), 303
Grids
  coordinate systems and, 140
  DEM data in, 77
  map types and, 139
GRIP, 120
GRS80 spheroid, 140
GUI, 20, 38–40, 264–266
Gunnison River Basin (Colorado), 212
Gurnee (Illinois), 162
Gutters, 291

H

H&H, *see* Hydrologic and Hydraulic (H&H) models
Haestad Methods
  company, data on, 235
  connection wizards, 211–212
  HEC-GIS system, 208
  SCADA Connect, 188
  SewerCAD, 188, 211, 261, 289
  SewerGEMS, 289
  Skelebrator module, 249, 353
  StormCAD, 211, 261, 289
  WaterCAD, *see* WaterCAD software
  WaterGEMS, *see* WaterGEMS software
  WaterSAFE, 340–342
  Web site for, 232
Hail, 181, 182
Hail Index Overlay, 182
Hampton (Virginia), 294–295
Hansen Information Technologies
  CCTV inspection database, 301–302
  IMS-ARC, 301
  IMS-AV, 301
  Infrastructure Management System, 301, 319
  Web site for, 301
Harrisburg (Pennsylvania), 241, 242
Hawaii, 83
Hazardous waste, 92
HEC, *see* Hydrologic Engineering Center (HEC)
HEC-GIS system, 208
Hell's Canyon (Idaho), 85
Hewlett Packard
  iPAQ Pocket PC, 133–134, 317
  Web site for, 134
HGL, 259, 266, 284
HMS, 76, 195, 207, 261
Homeland Security, Department of, 337
Hotlink function, 167
Houston (Texas), 76–77
HRAP, 183
HSG, 199–200
HSPF model, 214–216
H_2OMAP software
  features of, 234
  linkage method, 232, 261
  Protector module, 340
  Skeletonizer module, 249
  Web site for, 232
H_2ONET software, 232, 233
H_2OVIEW software, 234, 261
Huber, Wayne, 262
Huntington (West Virginia), 187
HYDRA software, 291
Hydrants
  in data dictionary, 157
  mapping of, 148–149, 157–158
  in MIKE NET, 237
  modeling of, 244
  premarking of, 149
  surveys and, 112
Hydraulic characterization data, 179
Hydraulic gradient lines (HGLs), 259, 266, 284
Hydrologic and Hydraulic (H&H) models
  contamination analysis in, 339
  GIS and, 14, 20–21, 195–197
  meteorological data for, 179
  for sewer systems, 259–260
  subbasin parameters for, 198
  surveys and, 112, 116
  SWMM, *see* Storm Water Management Model (SWMM)
  types of, 196
  updating of, 14
  vulnerability assessments, 339
  in WISE, 14
Hydrologic Engineering Center (HEC)
  ARC/HEC-2 program, 219–220
  background of, 207
  Buffalo Bayou Watershed and, 76–77
  HMS, 76, 195, 207, 261
  RAS, 207–208, 261, 289–290
  Web site for, 222
Hydrologic Modeling Dialogue, 91
Hydrologic Modeling System (HMS), 76, 195, 207, 261
Hydrologic Rainfall Analysis Project (HRAP), 183
Hydrologic soil group (HSG), 199–200
Hyperion spacecraft, 65
Hyperlink function, 167

I

IAHS, 5
IBM's Data Explorer software, 212, 289–290
ICAM, 311
Idaho, Hell's Canyon, 85
IDRISI software, 51, 89, 94–95
IFSAR, 85–86
IKONOS satellite
  Albany County images, 48
  coverage of, 53
  DEM data from, 86
  features of, 57
  format of images, 58
  image processing software for, 71
  imagery costs, 59
  Pentagon images, 337
  resolution of, 53, 56, 58
  Web site for, 73
  World Trade Center images, 337
Illinois
  Adam County, 2
  Gurnee, 162
  Quincy, 2
  Sycamore, 243–247
Image Analysis, 66, 67, 69
Image calibration, 69
ImageMap USA, 70
ImageWorks software, 71
IMAGINE software
  in ADWR water monitor system, 62
  components of, 68
  cost of, 67
  DEM and, 84, 88, 89
  DOQQs and, 62
  extensions and, 69
  features of, 68
  GPS data and, 115
  hyperspectral analysis by, 65
  images in, 51
  for pipeline mapping, 69
  SLOPE function in, 88
  Web site for, 67
Imaging spectroscopy, 62
Impervious areas
  AVSWMM and, 273
  classification and, 200
  impervious factor and, 173
  NPDES permits and, 173–174
  percent imperviousness and, 64, 65
  runoff curve number and, 199
  sewage flow and, 260
  in SWMM, 204, 267
  SWMMTools and, 280
Implementation plan, 41–43
IMS-ARC software, 301
IMS-AV software, 301
Indian Remote Sensing (IRS) satellite, 15
  features of, 55
  image processing software for, 71
  resolution of, 53, 56
Indiana, Bloomington, 45
Indonesia, 15
Inflow and infiltration (I/I)
  CCTV and, 314
  deterioration and, 290, 312
  documentation of, 260
  service laterals and, 170–171
  SWMM and, 259
InfoNet software, 232, 237, 261, 287
Information system, definition of, 3
Informix, 236
InfoWater software, 234
InfoWorks software, 232, 237, 261, 287
Infrastructure capital asset management (ICAM), 311
Infrastructure Management System, 301, 319
Innovative System Developers' Geo-STORM, 219
InService software, 132
Integraph, 331
Integration method
  description of, 212–221
  vs. interchange method, 197, 221
  vs. interface method, 197, 212, 221
  sewer modeling and, 261
  water modeling and, 230, 231, 232
IntelliWhere OnDemand, 132–133
Interchange method
  description of, 197–205
  features of, 221
  vs. integration method, 197, 221
  vs. interface method, 197, 205, 221
  sewer modeling and, 261
  water modeling and, 230, 232
Interface method
  description of, 205–212
  vs. integration method, 197, 212, 221
  vs. interchange method, 197, 205, 221
  sewer modeling and, 261
  water modeling and, 230, 232
Interferometric Synthetic Aperture Radar (IFSAR), 85–86
Intergraph
  FRAMME, 301
  GeoMedia, 124–125, 150, 287
  GPS data and, 115
  InService, 132
  IntelliWhere OnDemand, 132–133
  MGE software, 38, 115

MrSID and, 70
Notifast, 162–163
Web site for, 301
Intermap Technologies, 62, 86
International Business Machines, 212, 289–290
International Association of Hydrological Sciences (IAHS), 5
Internet, *see also* World Wide Web
GIS data and, 122–128
history of, 4
security, 123–124
wireless connections to, 133
Internet Map Server, 294–295
Iowa, Neal Smith National Wildlife Refuge, 108–109
iPAQ Pocket PC, 133–134, 317
Irrigation, 66, 218
IRS, *see* Indian Remote Sensing (IRS) satellite
Irving (Texas), 174
ISODATA classifier, 69

J

JERS, 71
Jesper County (Iowa), 108–109
Johnson County (Kansas), 119, 128

K

Kansas
Johnson County, 119, 128
NAD27 and, 61
Kentucky
Fort Wright, 289–290
University of, PIPE2000 Pro, 232
Kinematic wave models, 195
King of Prussia (Pennsylvania), 14
Kiski Valley (Pennsylvania), 144–145, 146
Kittanning, Borough of (Pennsylvania), 166
KnowledgeMiner, 122
Knoxville (Tennessee), 14
Kodak's CITIPIX software, 59
KYPIPE model, 230, 232, 237

L

Lake Elsinore, City of (California), 158–159
Lake Wylie area (North Carolina), 360
Land use/land cover (LULC) data
aerial vs. satellite imagery, 63
AVSWMM and, 272–273
BASINS program and, 216
from California, 64
changes in, 64
classification by, 63, 64, 200, 204–205
collection of, 63–65
demand and, 251–253
digitization of, 149
impervious areas and
AVSWMM and, 273
classification and, 200
impervious factor, 173
NPDES permits and, 173–174
percent imperviousness and, 64, 65
runoff curve number and, 199
sewage flow and, 260
in SWMM, 204, 267
SWMMTools and, 280
maps for, 139, 149
MISR and, 54
from Pennsylvania, 55–56
PSRM and, 204
runoff curve numbers and, 199–200
from San Diego, 62–63
SCS and, 200
for sewer systems, 260
SWMM and, 204
from urban areas, 58
Landsat
coverage of, 55
ETM+, 54–55
features of, 55
GeoCover LC and, 63
history of, 52
image processing software for, 71
ImageMap USA, 70
imagery costs, 59
MSS on, 52
resolution of, 52–55
TM on, 52, 54–56, 59, 63, 204
Web site for, 73
Laser imaging detection and ranging (LIDAR)
accuracy of, 85
ALTM, 85
analysis of data from, 19, 72
definition of, 62n
DEM and, 84–85
features of, 85
World Trade Center images, 337
Laterals, service
catchments and, 259
flow in, 195
inflow, infiltration, and, 170–171, 312
mapping of, 171–173
SRDMS and, 332
Layers
in AM, 298
buffers and, 164

coverage of, 53, 148
definition of, 148
features and, 156
MMA and, 169
themes in, 164
LBS, 18, 130
Lead, testing for, 190
Leadville (Colorado), 62
Lehigh County (Pennsylvania), 32–33
Leica Geosystems, 105, 115; *see also* Earth Resource Data Analysis System (ERDAS)
Lewisburg Borough (Pennsylvania), 97–101
Licking River, 290
LIDAR, *see* Laser imaging detection and ranging (LIDAR)
LizardTech's MrSID, 62, 66, 67, 69–70
Ljubljana (Slovenia), 286
LoadBuilder module, 250–252, 353
Location-based services (LBS), 18, 130
Lockheed Martin Information Services, 282–283
Long-term control plan (LTCP), 260, 263
Los Angeles (California), 177
Los Angeles County (California), 138–139, 189
Loveland (Colorado), 69
Lower Colorado River Authority (Texas), 253
LTCP, 260, 263
LULC, *see* Land use/land cover (LULC) data
Lumped-parameter models, 196

M

Maidment, David, 208
Mains, water
classification of, 158, 164
in data dictionary, 157
deterioration in, 190
isolation trace, 235, 327–328
mapping of, 157–158
modeling of, 229
monitoring of, Toronto, 190
repair/replacement of, 327
wye information for, 171
Maintenance
historical data on, 311
in implementation plan, 42–43
types of, 305
Maintenance management systems (MMS), 237; *see also* CMMS
Malcom Brook (New York), 194–195
Manholes
AVSWMM and, 273, 274–276
covers for, 103, 112, 116, 159
in data dictionary, 157
depth of, 112
inflow, infiltration, and, 312
inspection of, 179, 260, 323
maintenance of, 260
mapping of
Allegheny County, 159
in layers, 148
Los Angeles County, 138–139
for NPDES permits, 260
Ramsey, Borough of, 157–158
resolution and, 143
scale for, 142
themes in, 323
premarking of, 149
repair/replacement of, 331–333
SRDMS and, 331–333
Maps
3D visualization and, 169–170
accuracy of, 140–142
AM/FM vs. GIS, 146–147
attributes on, 146, 148–149
audio recordings and, 321
buffers in, 164–166
CAD vs. GIS, 146–147
CCTV and, 321–322
construction drawings and, 149
coordinates for, 140
creation of, 147–156
data dictionary for, 157
datums and, 140
digitization of, 149
feature detail on, 140–141
field notes and, 321
labels and legends on, 146, 148, 155
layers in, 148, 156
links to, 321
OGC Web Map Service, 190
overlay analyses of, 164
photographs and, 143, 321
projection systems for, 140
revisions to, 146
scale of, 140–142
scanning of, 150
vs. spreadsheet, 242
of stream flow, 206
themes in, 164
types of, 139, 142–145
video for, 321–322
Map Maintenance Application (MMA), 169
MapGuide, 120, 124, 127
MapInfo
GPS data and, 115
InfoNet and, 287
InfoWorks and, 237, 287
MapMarker J Server, 111
MapXtreme, 111, 124, 132

MediaMapper and, 322
MIKE NET and, 236
Routing J Server, 111
SWMM and, 259
MapObjects software, 24, 70, 294–295, 311
MapSheets software, 68
MapXtreme software, 111, 124, 132
Mariposa East (California), 75
Maryland, GeoMaP in, 40
Massachusetts
Boston
asset management system for, 311
catch basins in, 311
CCTV and, 314
drinking water in, 296
GIS implementation in, 13
sewer system in, 290
Fitchburg, 116
Gloucester, 296
metal levels in water, 190
MWRA, 102, 167, 290
Massachusetts Water Resources Authority (MWRA), 102, 167, 290
Massif Central (France), 92
Meades Ranch (Kansas), 61
Mecklenburg County (North Carolina), 70, 230, 360
Media Player, 21, 23
MediaMapper software, 322
Mesocyclone Overlay, 182
Meteorlogix, 183, 190
Mexico, 15
MGE software, 38, 115
Miami-Dade (Florida), 186
Michigan
Dowagiac River Watershed, 220–221
Genesee County, 229
Oakland County, 174
MicroDEM, 83, 89
Microsoft
Access, *see* Access, Microsoft
COM components, 237
Excel, 68, 185, 236
Internet Explorer, 265
Media Player, 21, 23
.NET initiative, 234, 236
Office, 68
Power Point, 68
SQL Server, 236
TerraServer, 59, 73
Visual Basic, 21, 132, 152, 328
Word, 68
Microstation software, 301
Middle Chattahoochee River Watershed (Georgia), 218
MIKE BASIN, 218–219
MIKE NET, 232, 236–237, 355–356, 358–360
MIKE SHE, 219
MIKE SWMM, 261, 285, 286–287
Miner and Miner
ArcFM, 151, 236, 301–303, 356–358
Web site for, 301
Minneapolis (Minnesota), 125
Minnesota, Minneapolis, 125
MISR, 54
Missouri, Goodwater Creek, 88
Mixels, definition of, 65
MMA, 169
MMS (Maintenance management systems), 237
MMS (Multimedia Mapping System), 322
Mobile GIS, 130–135
ModClark, 207
Modeling of Urban SEwers (Mouse) software, 261, 284–287, 346–347
Moderate Resolution Imaging Spectroradiometer (MODIS), 53, 54
MODIS, 53, 54
Mohawk Valley Water (New York), 40–41, 249
Monongahela (Pennsylvania), 35
Monongahela River (Pennsylvania), 99
Montana, water quality in, 127–128
Montgomery County (Maryland), 40
Mouse software, 261, 284–287, 346–347
MrSID, 62, 66, 67, 69–70
MS4 permits, 173
MSS, 52
Multiangle Imaging Spectroradiometer (MISR), 54
Multimedia Mapping System (MMS), 322
Multiresolution Seamless Image Database (MrSID), 62, 66, 67, 69–70
Multispectral images, 53, 65
Multispectral scanner (MSS), 52
Munhall, Borough of (Pennsylvania), 143, 145
Municipal Separate Storm Sewer System (MS4) permits, 173
MWH Soft
company, data on, 232–233
Demand Allocator module, 235
H_2OMAP, *see* H_2OMAP software
H_2ONET, 232, 233
H_2OVIEW, 234, 261
InfoWater, 234
Protector module, 340
Skeletonizer module, 235, 248, 249
Tracer module, 235
Web site for, 232
MWRA, 102, 167, 290

N

NAD27, 61, 81, 140
NAD83, 60, 61, 81, 140
NAPP, 60–61
NASA, 64, 65
National Aerial Photography Program (NAPP), 60–61
National Aeronautic and Space Administration (NASA), 64, 65
National Climatic Data Center (NCDC), 181, 190
National elevation dataset (NED), 83, 105
National Geodetic Vertical Datum of 1929 (NGVD), 81
National Geospatial Data Clearinghouse, 84, 105
National High Altitude Photography (NHAP), 60
National Map Accuracy Standards, 56, 59
National Mapping Program, 81
National Oceanic and Atmospheric Administration (NOAA), 64
National Pollution Discharge Elimination System (NPDES) permits
  ALERT Flood Warning System and, 179
  maps for, 260
  sewer overflow and, 187, 260
  storm runoff fees and, 173–175
  surveys and, 112
National Risk Management Research Laboratory's EPANET, 231–232, 236–241, 339, 364
National Spatial Data Infrastructure (NSDI), 61, 84
National Weather Service's NEXRAD, 181–184, 185
Natural Resources Conservation Service (NRCS), 199–200, 267–270
Naval Earth Map Observer (NEMO), 72
Navigation System by Timing and Ranging, *see* Global Positioning System (GPS)
Navman, 133–134
NAVSTAR, *see* Global Positioning System (GPS)
NCDC, 181, 190
Neal Smith National Wildlife Refuge (Iowa), 108–109
NED, 83, 105
Needs analysis, 28–42, 147
NEMO, 72
.NET, 234, 236
Network Analyst software, 37, 168, 201
Network Query Language (NQL), 16
Networks
  demand zones in, 247
  in FM, 300
  Mouse and, 285
  skeletonization and, 243–249
  topology and, 139
  vulnerability assessments, 339
New Jersey
  Ocean County, 28
  Ramsey, Borough of
    mapping of, 137, 156–158
    needs analysis, 29, 30, 38, 39
    pilot project for, 44–45
New Mexico
  Albuquerque, 336–337
  Truth or Consequences, 353–356
  Williamsburg, 355
New York City
  drinking water in, 296
  World Trade Center, attack on, 337
New York (state)
  Malcom Brook watershed, 194–195
  Mohawk Valley Water, 40–41, 249
  New York City, 296, 337
  Westchester County, 194–195
Newport News (Virginia), 252
NEXRAD, 181–184, 185
NEXRAIN Corporation, 183, 184, 190
Next Generation Weather Radar (NEXRAD) system, 181–184, 185
NGVD, 81
NHAP, 60
Nine Minimum Controls (NMC), 260, 312
Nitrogen, 221, 360
NMC, 260, 312
NOAA, 64
Node demand, 249–253, 352–353, 356
Node elevation, 83, 103–105, 253–254, 259; *see also* Digital elevation models (DEMs)
Nonpoint Source Model (NPSM), 214, 221
North American Datum of 1927 (NAD27), 61, 81, 140
North American Datum of 1983 (NAD83), 60, 61, 81, 140
North Carolina
  Buncombe County, 309, 347–352
  Charlotte, 14, 230
  Greenville, 303
  Lake Wylie area, 360
  Mecklenburg County, 70, 230, 360
  The Palisades, 360
  Pitt County, 303
Notifast, 162–163
NPDES permits, *see* National Pollution Discharge Elimination System (NPDES) permits
NPSM, 214, 221
NQL, 16
NRCS, 199–200, 267–270
NSDI, 61, 84

O

Oakland County (Michigan), 174
Object Linking and Embedding (OLE) technology, 68
Observation and Measurements Language, 190
OCEA awards, 2
Ocean County (New Jersey), 28
ODBC, 196–197, 236, 343
Office, Microsoft, 68
Office, Pathfinder, 109, 115
OGC, 190, 343
Ohio
  Cincinnati
    cost-benefit analysis, 13
    Internet GIS, 122
    MMA and, 169
    storm sewers in, 11
    vulnerability assessments, 338
    water infrastructure project in, 31
  Dublin, 120–121
  GRIP, 120
Ohio River, 290
Oklahoma, TRMM studies, 65
OLE, 68
One-Hour Precipitation, 182
OnSite software, 132
Open Database Connectivity (ODBC), 196–197, 236, 343
Open GIS Consortium (OGC), 190, 343
Optech, 85
Oracle
  Cityworks and, 304
  GeoGateway and, 71
  Internet GIS and, 122–125
  MGE software and, 38
  MMA and, 169
  RAPID, 59
  RDBMS data and, 45, 122–123
  SQL and, 304
Orange County, 38
OrbView satellites, 57, 65, 73
Oregon, Portland, 258–259
Oregon State University, 264
OrthoEngine Suite, 71
Outfall conditions, 112
Outstanding Civil Engineering Achievement (OCEA) awards, 2
Overflow structures, 112, 159, 260
Overshoot, definition of, 154

P

PAG, 353
The Palisades (North Carolina), 360
PalmPilot, 317
PaMAGIC, 56
PAMAP software, 71
Panchromatic images, 53
Paramax Systems Corporation, 183
PASDA, 61, 73, 84, 105
Pathfinder GPS, 108–109, 115
PCI Geomatics
  company, data on, 70–71
  EASI/PACE, 67, 71
  FLY!, 71
  GDB technology, 71
  GeoGateway, 71
  Geomatica, 51, 67, 70–71
  ImageWorks, 71
  OrthoEngine Suite, 71
  PAMAP, 71
  RADARSOFT, 71
  RAPID, 59
PCSWMM GIS, 261, 262, 281–282
PDAs, *see* Personal digital assistants (PDAs)
PDOP, 109
Pearpoint, 316–319, 333
Peninsular Technologies, 319, 333
Penn Hills (Pennsylvania), 122
Penn State Runoff Model (PSRM), 203, 204
Pennsylvania
  Allegheny County
    3RWWDP, 184–186
    CSO diversion chamber, 293
    planimetric map of, 143, 145
    precipitation atlas of, 185
    sewage system, 159, 170
  Baldwin, Borough of, 330–331
  Bethel Park, 161, 170
  Bethlehem, 188
  Bull Run Watershed, 97–101
  California, Borough of, 125
  Charleroi, Borough of, 101–102
  Harrisburg, 241, 242
  King of Prussia, 14
  Kiski Valley, 144–145, 146
  Kittanning, Borough of, 166
  Lehigh County, 32
  Lewisburg, Borough of, 97–101
  LULC maps of, 55–56
  Monongahela, 35
  Monongahela River, 99
  Munhall, Borough of, 143, 145
  PaMAGIC, 56
  PASDA, 61, 73, 84, 105
  Penn Hills, 122
  Philadelphia, 167–169
  Pittsburgh, 43–44, 170, 184–186
  Saw Mill Run Watershed, 170

Susquehanna River, 97–101
Terrabyte data for, 55–56
Union County, 97–101
Washington County, 328–330
White Haven, Borough of, 103
Pennsylvania Mapping and Geographic Information Consortium (PaMAGIC), 56
Pennsylvania Spatial Data Access (PASDA) system, 61, 73, 84, 105
Pensacola Bay area (Florida), 290–291
Percent imperviousness, 64, 65
Personal digital assistants (PDAs)
geographic data on, 16
iPAQ, 133–134, 317
operating systems for, 130–131
PalmPilot, 317
SRDMS and, 331
WinCan and, 317
Philadelphia (Pennsylvania), 167–169
Phosphorus, 221, 360
Photographs, *see also* Digital orthophotos (DOPs)
aerial, *see* Aerial photographs
ASPRS, 49–50
mapping and, 143, 321
NAPP, 60–61
NHAP, 60
Physical Characterization data, 179, 312
Pidpa, 188, 356–358
Pilot projects
definition of, 40–41
geographic limits, criteria for, 44
for maps, 147
Pima Association of Government (PAG), 353
PIPE2000 Pro software, 232, 237
Pipes
defects in, 309, 349–352
EPANET and, 231–232
head losses for, 242, 244
height of, 288
hydraulic characterization of, 179
inspection of, 179
KYPIPE and, 230
landbase accuracy for, 141
layout of, 244
location of, 356
mapping of, 148–149, 157–158
MIKE NET and, 355
modeling of, 229
pressure in, 242, 253, 255
rating of, 309–310, 347–352
repair/replacement of, 235, 296, 327, 331–333
roughness coefficient of, 205–206, 275
size of, 247–249
SRDMS and, 331–333
surcharge depth, 288
PipeTech, 319
Pipeworks software, *see* Cityworks software
Pitt County (North Carolina), 303
Pittsburgh (Pennsylvania), 43–44, 170, 184–186
Pixels
interpolation vs. spatial join, 254–255
in radar data, 183
in raster maps, 50, 65, 68, 72, 139
in Terrabyte, 56
Pizer's HYDRA, 291
Planimetric maps, 143
Planning departments
applications for, 30, 147
base map accuracy for, 142
landbase accuracy for, 141
pilot project for, 45
security assessments, 336–338
Plotters, 38, 142
Pollution
hydrology extension and, 92
Mouse and, 285
MS4 permits and, 173
NPDES permits and, 173, 189
NPSM and, 214
SWAT and, 360
TMDLs and, 214
tracking of, 11
WARMF and, 360
Portland (Oregon), 258–259
Positional dilution of precision (PDOP), 109
Postprocessor, 197, 205, 265
Power Point software, 68
Prague (Czech Republic), 358–360
Precipitation
cloud cover index for, 180
radar data on, 181–186
rainfall, *see* Rain
Rhea model of, 212
snow, 54
Preprocessor
definition of, 197
in interface method, 205
MrSID, *see* MrSID
for SWMM, 265
Prince George County (Maryland), 40
Professional Surveyor magazine, 117
Proximity, 139
Public Health Security and Bioterrorism Preparedness and Response Act, 337
Public safety, department of, 30, 147
Public works departments, 30, 45, 142, 147
Puerto Rico, NED resolution for, 83

Q

Qatar, 15
QUAL2E model, 214
Quality assurance/quality control (QA/QC) program, 155
QuickBird satellite, 53, 56, 57, 73
QuickSAT, 53
Quincy (Illinois), 2

R

Radar
  images from, 49, 71
  rainfall data from, 181–186
RADARSAT, 71
RADARSOFT, 71
Rain
  mapping of, 173–175, 184
  monitoring of, 178–179, 185–188
  remote sensing of, 66, 179–186
RainVieux, 186
Ramsey, Borough of (New Jersey)
  mapping of, 137, 156–158
  needs analysis study for, 29, 30, 38, 39
  pilot project for, 44–45
Rand Corporation, 123–124
RAPID, 59
Raster maps
  definition of, 139
  from satellites, 50, 65, 68, 72
Rational Method program, 208
RBE, 302
RDBMS, *see* Relational database management system (RDBMS)
Real-time Acquisition and Processing of Imagery Data (RAPID), 59
Real-Time Control (RTC) modules, 285, 287
Real time kinematic (RTK) GPS receivers, 109–110
Red Hen Systems, 322, 333
Reduce/skeletonize/trim (RST) process, 249
Reflectivity, base, 181
Regional Planning Technologies, 301
Relational database management system (RDBMS)
  ArcFM and, 303
  definition of, 35
  Internet GIS and, 123
  for maps, 155
  MMA and, 169
  Oracle and, 45, 122–123
Relational databases
  Access, *see* Access, Microsoft
  data storage in, 322
  GeoGateway and, 71
  Informix, 236
  MIKE NET and, 236
  Oracle, *see* Oracle
  SQL Server, 236
  Sybase, 236, 304
Remote sensing systems, 48–73, 179–186
Research Systems' ENVI software, 51, 65, 67
Resolution, spatial, 53
Rhea model, 212
River Analysis System (RAS), 207–208, 261, 289–290
River hydraulics and Geometry (RMA-2V) software, 289–290
Rivers, *see also* Streams
  floodplain and, 166
  Geo-STORM and, 219
  hydrology extension and, 92
  MIKE models for, 218–219
  swimming in, 296
RJN Group
  CASS WORKS, 24, 290, 301
  CASSView, 24
  Web site for, 24
RMA-2V software, 289–290
RMSE classification, 80
Roads, DEM data for, 77
Root mean square error (RMSE) classification, 80
Roughness coefficient, 205–206, 274, 275
Round Rock (Texas), 252–253
RST, 249
RTC modules, 285, 287
RTK receivers, 109–110
Rule base engine (RBE), 302
Runoff
  curve numbers, 199–200, 273–274
  DEM and, 80
  Geo-STORM and, 219
  hydrology extension and, 92
  LULC and, 64, 65
  mapping of, 173–175
  MIKE model for, 219
  Mouse and, 285
  WMS and, 208
Russia
  GIS implementation in, 15
  GLONASS, 109

S

Safe Drinking Water Act, 337
SAIC, Weather Monitoring System, 190
SAMS, 290
San Diego Association of Governments (SANDAG), 62–63

San Diego (California), 13, 24, 62–63
San Francisco Bay area (California), 313–314
San Jose (California), 328
San Mateo County (California), 45
SANDAG, 62–63
Sanitary sewer overflows (SSOs)
  CMOM and, 260, 312–314
  inflow, infiltration, and, 312
  mapping of, 159, 260
  pipe rating method and, 309
  rainfall and, 186
  structures for, 112
  SWMM and, 263
Satellite imagery, 48–73, 179–186
Saw Mill Run Watershed (Pennsylvania), 170
SCADA, 49, 187–188, 237
SCADA Connect, 188
Scanning, 150
Scientific Visualization System (SVS), 212
Scripts, 20–22, 198
SCS, 199, 200
SDE, 36–37, 303
SDTS, 83
Seattle (Washington), 169
Security
  ERP development and, 337
  Internet GIS and, 123–124
  modeling of, 229
  password access to software, 39
  planning and, 336–338
  software for, 340
  Tracer module and, 235
  vulnerability assessments, 338–339
Selective Availability (SA) error, 110
Sensor Web, 189–190
Service laterals, *see* Laterals, service
Severe Weather Probability Overlay, 182
Sewage analysis and management system (SAMS) project, 290
Sewer rehabilitation data management system (SRDMS), 331–333
Sewer systems, *see also* Stormwater systems; Wastewater systems
  applications for, 259–260
  AVSWMM and, 274
  catchments in, 259
  classification of, 349
  cost of, 296
  in data dictionary, 157
  database for, 34, 261
  DEM of, 101–103, 260
  demographic data and, 202–203
  "Euclidean Distance" approach to, 102
  flow in, 153–154, 171–173, 260, 263–264
  H&H modeling of, 259–291
  inflow, infiltration, and, 290, 312
  inspection of, 260, 322
  mapping of
    Allegheny County, 159, 170
    Bethel Park, 161, 170
    CCTV for, *see* Closed-circuit television (CCTV)
    in England, 148
    in layers, 148
    Los Angeles County, 138–139
    Ramsey, Borough of, 157–158
  monitoring of, 184–187
  sampling of, 260
  septic conversion to, 296
  SWMM and, 263–264
  vulnerability assessments, 338–339, 342
SewerCAD software, 188, 211, 261, 289
SewerGEMS software, 289
Sewersheds, 264; *see also* Catchments; Subbasins
SEWERview software, 301
Shapefiles
  3D Analyst and, 104–105
  connection wizard for, 211–212
  creation of, 68
  DEM data in, 84
  editing of, 68
  GPS data in, 112
  Hydro extension and, 90
  Image Analysis and, 69
  site plans in, 42
  Stereo Analyst and, 69
  TerraExplorer and, 126
  weather data and, 183
SHC, 260
Shuttle Radar Topography Mission (SRTM), 84
Sinks, 87, 90–92
Sioux Falls (South Dakota), 55
Sisterna de Agua Potabley Alcantarillado de León (Mexico), 15
Skelebrator module, 249, 353
Skeletonization process, 243, 244–249
Skeletonizer module, 235, 248, 249
Skyline Software System's Terra Explorer, 80, 126
Slope, 206, 260, 271
Slovenia, Ljubljana, 286
Small-scale maps, 144–145
Smoke tests, 179, 260, 322
Snow, 54
Sofia (Bulgaria), 346–347

Software, *see also* specific programs
application programming and, 20–24
hardware for, 37–39
Soil
AVSWMM and, 273
BASINS program and, 216
capillary suction of, 270
Green-Ampt parameters and, 270
hydraulic conductivity of, 270
moisture content of, 51, 65, 270
Mouse and, 285
multispectral imagery and, 58
NRCS data for, 267–270
pipe deterioration and, 190
runoff curve numbers and, 199–200
sampling of, 190
sewage flow and, 260
total suspended sediment criteria, 360
Soil and Water Assessment Tool (SWAT), 216, 360
Soil Conservation Service (SCS), 199, 200; *see also* NRCS
Source maps, 139
South Africa, 15
South Australian Water Corporation (Australia), 15
South Dakota, Sioux Falls, 55
Spacient Technologies' Fieldport, 133
Spain, 362–364
Spatial Analyst
ArcGIS and, 23, 90
vs. ArcGRID extension, 90
AVSWMM and, 271
decision matrix for, 37
DEM and, 89, 90–92, 97–101
features of, 90
HEC-GeoHMS and, 207
hydrology extension in, 90–92, 97–101
for monitoring rainfall, 178–179
NPDES permits and, 189
RAS and, 208
vs. TOPAZ, 95
Spatial Data Transfer Standard (SDTS), 83
Spatial Database Engine (SDE) software, 36–37, 303
Spatial Hydrology Analysis System, 209
SpatialAge Solutions' Knowledge Miner, 122
SPC projection, 140
Spectrum Width, Base, 181
SPOT satellite
features of, 55
image processing by, 71, 84
resolution of, 53, 56
watersheds and, 51
World Trade Center images, 337
SQL, 16, 174–175, 303, 304
SQL Server, 236
SRDMS, 331–333
SRTM, 84
Stakeholders, 30–32
STAR-3i system, 86
State Plane Coordinate (SPC) projection, 140
Stege Sanitary District (California), 313–314
Stereo Analyst, 69
Stockton (California), 291
Stoner Associates, *see* Advantica Stoner
Storm Structure, 182
Storm Total Precipitation, 182
Storm Tracking Information Overlay, 182
Storm Water Management Model (SWMM)
AGSWMM and, 280–281
ArcGIS software and, 280–281
ArcInfo and, 290–291
ArcView and, 270–271, 277, 280
Avenue and, 270–271
AVSand and, 284
AVSWMM and, 270–278
basement flooding and, 258–259
BASINS and, 282–283
curb length and, 200–201
demographic data and, 203
development of, 262
features of, 261–264
GIS layers in, 267, 268–269
Green-Ampt parameters and, 269–270
GUI for, 265–266
linkage method, 261
LULC data for, 204
MIKE SWMM and, 285
Mouse GM and, 286–287
NRCS data for, 267–270
overland flow width data for, 92
PCSWMM GIS and, 281–282
runoff quality in, 200–201
Spatial Analyst and, 271
SWMMDUET and, 283–284
SWMMTools for, 278–280
Visual SWMM and, 266–267
Web sites for, 261, 264
wet-weather models, 263–264
XP-SWMM and, 266–267
StormCAD software, 211, 261, 289
STORMview software, 301
Stormwater systems, *see also* Sewer systems
catchments in, 259
CCTV for, *see* Closed-circuit television (CCTV)

Cityworks and, 305
DEM and, 77
fee assessment for, 173–175
flow in, 92, 179, 260
Geo-STORM and, 219
GISHydro modules and, 210
H&H models for, 179
HEC-GeoHMS and, 77
inspection of, 305
maintenance of, 179, 260
mapping of, 141, 152, 156
monitoring of, 179, 189
MS4 permits and, 173
NPDES permits and, 173–175
remote sensing and, 59
Rule 34 requirements, 312
sampling of, 260
sewersheds and, 101
surveys and, 113
SWMM, *see* Storm Water Management Model (SWMM)
testing of, 305
threats to, 337
in XP-SWMM, 266
Streams, *see also* Rivers
in ArcGrid extension, 93
attributes of, 205
"burning in" of, 88
conditions of, 205
in DEMs, 86–88, 95–101
flood profiles for, 211
flow in
direction of, 153–154
documentation of, 260
Geo-STORM and, 219
GIF images and, 195
lateral, 170–171, 195
mapping of, 159, 166, 206
measurement of, 179–180
overland width data for, 92
types of, 210
USGS data on, 127–128
velocity of, 211
GISHydro modules and, 209–210
modeling of, 195, 205–206
monitoring of, 178–179
roughness coefficient of, 205–206
sampling of, 260
in Spatial Analyst, 90–92
TOXIROUTE model for, 214
types of, 205
USGS Internet site on, 127–128
Structured query language (SQL), 16, 174–175, 303, 304
Subbasins, 200, 203–205, 209–210, 264
Sueca (Spain), 362–364
Sun Microsystems' RAPID, 59
Supervisory control and data acquisition (SCADA) systems, 49, 187–188, 237
Surveys, 111–117, 253
Susquehanna River (Pennsylvania), 97–101
SVS, 212
SWAT, 216, 360
SWMM, *see* Storm Water Management Model (SWMM)
SWMMDUET, 283–284
SWMMTools, 278–280
SWS model, 229
Sybase, 236, 304
Sycamore (Illinois), 243–247
SynerGEE Water software, 232, 237
System Hydraulic Characterization (SHC), 260

T

Tablet PC, 131
Tacoma (Washington), 126–127
Taiwan, DEM resolution for, 81
Taxes, 30
TAZs, 353
Telecommunications, landbase accuracy for, 141
Telemetry, *see* Supervisory control and data acquisition (SCADA) systems
Television, *see* Closed-circuit television (CCTV)
Tennessee
Germantown, 226–227, 228
Knoxville, 14
Terra satellite, 53–54, 65, 73, 84
Terrabyte, 55–56
TerraExplorer software, 80, 126
TerraServer, 59, 73
TetraTech's BASINS program, 213–218, 222, 282–283
Texas
Buffalo Bayou Watershed, 76–77
CRWR, 91, 105, 208–210, 216
Houston, 76–77
Irving, 174
Lower Colorado River Authority, 253
Round Rock, 252–253
storm runoff fees in, 174
Theissen Polygons, 250
Thematic mapper (TM), 52, 54–56, 59, 63, 204
Thematic maps
AVNET and, 241
creation of, 155, 164
of flood risk, 259

of pipes, 242
of stream flow, 206
theme-on-theme selection in, 164
Themes
coverage and, 53, 148
curb length and, 201
definition of, 164
DOQQs and, 330
Thermal-infrared spectrum, 51
Three Rivers Wet Weather Demonstration Program (3RWWDP), 184–186
TIFF format, 58, 70, 183
TIGER data, 145, 203, 291
TIN, *see* Triangular irregular network (TIN)
TM, *see* Thematic mapper (TM)
TMDL, 213, 214, 217, 218
TOPAZ software, 89, 95
Topologically Integrated Geographic Encoding and Referencing (TIGER) System, 145, 203, 291
Topology
3D visualization and, 170
in AM, 300
CAD vs. GIS maps, 146
data conversion and, 153–154
definition of, 139, 327
digitizing errors, 154
in FM, 300
isolation trace and, 327
vulnerability assessments, 339
Tornadic Vortex Signature Overlay, 182
Toronto (Canada), 190
Total maximum daily loads (TMDLs), 213, 214, 217, 218
Total Station 4800 GPS, 116
TOXIROUTE model, 214
Tracer module, 235
Traffic analysis zones (TAZs), 353
TRex Wizard, 254–255
Triangular irregular network (TIN)
DEM data in, 77
HEC programs and, 208, 219
in IDRISI, 94, 95
storm runoff taxes and, 174
WMS and, 208
Trimble
ASPEN software, 115
Asset Surveyor, 115
Geoexplorer GPS, 114–115
Pathfinder GPS, 108–109, 115
Pathfinder Office, 109, 115
Total Station 4800 GPS, 116
Web site for, 117
TRMM, 65
Tropical Rainfall Mapping Mission (TRMM), 65
Truth or Consequences (New Mexico), 353–356
Tucson (Arizona), 352–353
Two Rivers Regional Planning Council, 2

U

Undershoot, definition of, 154
Union County (Pennsylvania), 97–101
UNITAR, 15
United Kingdom
Dover, 14
sewer system, maps of, 148
United Nations Institute for Training and Research (UNITAR), 15
United States Department of Agriculture's TOPAZ software, 89, 95
United States Geological Survey (USGS)
DEM data from, 81–84, 89
DOQs from, 48
DRGs from, 141–142
EarthExplorer, 61, 73, 105
EROS, 84
maps from, accuracy of, 141–142, 253, 254
NED, 83, 105
RAPID, 59
stream-flow data from, 127–128
TerraServer, 59, 73
Web sites for, 73, 105
United States Naval Academy's MicroDEM, 83, 89
United States of America, *see also* specific states
DEM data for, 83
NED resolution for, 83
regional-scale models of, 80
wastewater systems in, 295–296
water industry in, 2–3
water systems in, 295–296
Universal Transverse Mercator (UTM) projection, 60, 61, 81, 140
University of Kentucky's PIPE2000 Pro, 232, 237
Urban areas
InfoWorks for, 287
LULC data from, 58, 63
Mouse software for, 284–285
runoff curve numbers for, 199
SCS class, 200
USDA's TOPAZ software, 89, 95
User interface, 38; *see also* GUI
USGS, *see* United States Geological Survey (USGS)
Utah, Great Salt Lake, 87

Utility Network Analyst, 327–328
UTM projection, 60, 61, 81, 140

V

VAD Wind Profile, 182
Valves, water
  in data dictionary, 157
  isolation trace, 235, 327–328
  mapping of, 11, 148, 157–158
  modeling of, 244
  premarking of, 149
  pressure regulating, 255
  Tracer module and, 235
VCRs, 314–319
Vector maps, definition of, 139
Vectorization, 150
Vegetation
  evapotranspiration and, 66
  LULC data on, 63–64
  MISR and, 54
  multispectral imagery and, 58
  in vector layers, 143
Velocity, Base, 181
Ventura (California), 62
Vertically Integrated Liquid, 182
Video cassette recordings (VCRs), 314–319
Video Mapping System (VMS), 322
Vieux & Associates
  radar hyetographs from, 185–186
  RainVieux, 186
  Web site for, 190
Virginia
  Blacksburg, 128
  Fairfax County, 312
  Hampton, 294–295
  Newport News, 252
Visible spectrum, 51
Visual Basic, 21, 132, 152, 328
Visual SWMM, 266–267
VMS, 322
VSAT, 342, 343
Vulnerability assessments, 338–339
Vulnerability Self Assessment Tool (VSAT), 342, 343

W

WADSOP, 229
Wallingford Software
  company, data on, 287
  InfoNet, 232, 237, 261, 287
  InfoWorks, 232, 237, 261, 287
  Web site for, 232
WAP, 16
WARMF, 360
Washington County (Pennsylvania), 328–330
Washington (District of Columbia)
  DOPs of, 70
  Pentagon, attack on, 337
  VSAT and, 342
  WSSC, 40
Washington (state)
  Seattle, 169
  Tacoma, 126–127
Washington Suburban Sanitary Authority (WSSC), 40
WASP5 software, 289–290
Wastewater systems, *see also* Sewer systems
  CCTV for, *see* Closed-circuit television (CCTV)
  Cityworks and, 305
  demographic data and, 202–203
  flow in, 179, 260
  H&H models for, 179
  inspection of, 305
  maintenance of, 260
  mapping of, 169–173
  master plan for, 336–337
  monitoring of, 179
  Rule 34 requirements, 312
  sampling of, 260
  SWMM and, 263–264
  testing of, 305
  threats to, 337
  in USA, 295–296
  vulnerability assessments, 342
  in XP-SWMM, 266
Water Distribution Data Model, 40
Water Distribution System Optimization Program (WADSOP), 229
Water Environment Federation (WEF), 296
Water industry, definition of, 2n
Water Infrastructure Network (WIN), 296
Water pressure
  fire suppression and, 355
  node elevation and, 253
  pipe height and, 288
  thematic mapping of, 242
  zone trace, 255
Water quality
  ALERT Flood Warning System and, 178–179
  InfoWorks and, 287
  metal levels and, 190
  MIKE model for, 219
  monitoring of, 179, 190
  Mouse and, 285
  MS4 permits and, 173

NPDES permits and, 173
QUAL2E model and, 214
remote sensing of, 64
TMDLs and, 214
Water Quality Act (Montana), 127
Water Supply and Water Resources Division's EPANET, 231–232, 236–241, 339, 364
Water systems
applications for, 5–6, 227–228
CCTV for, *see* Closed-circuit television (CCTV)
Cityworks and, 305
cost of, 296
customers, *see* Customers
database for, 34
demand and, 249–253
design of, 229
drinking water in, 296
ERPs for, 337
flow in, 153–154, 171–173, 260
groundwater recharge, 66
inspection of, 305
landbase accuracy for, 10, 141
master plan for, 364
modeling of, 103–105, 230–255, 340–342
monitoring of, 179
remote sensing of, 51–52, 58–59
Rule 34 requirements, 312
surveys of, 111
testing of, 305
threats to, 337
in USA, 295–296
vulnerability assessments, 338–339, 342
WaterCAD software
ArcGIS and, 227
ArcView and, 105, 235
AutoCAD and, 235
connection wizard for, 211–212
demand and, 252–253, 353
features of, 235
linkage method, 232
node elevation in, 105
pilot project, use in, 40
SCADA and, 188
Skelebrator module, 249
WaterSAFE and, 340–342
Web site for, 232
WaterGEMS software
ArcGIS and, 236, 352
demand and, 250–252, 352–353
features of, 235–236
linkage method, 232
Loadbuilder module and, 250
SCADA and, 188
Skelebrator module, 249
TRex Wizard, 254–255
WaterSAFE and, 340
Web site for, 232
WaterSAFE software, 340–342
Watershed Analysis Risk Management Framework (WARMF), 360
Watershed Information System (WISE), 14
Watershed Modeling System (WMS), 91, 208, 222
Watersheds, *see also* Catchments
in ArcGrid extension, 93
BASINS program for, 213–218
data management files for, 216
DEM and, 80–81, 95–101
hydrologic elements in, 210
NRCS data on, 267–269
nutrients in, 63, 221, 360
in Spatial Analyst, 90–92
SWAT and, 216, 360
WARMF and, 360
WATERview software, 301
Waterways Experiment Station (WES), 208
Weather Surveillance Radar - 1988 Doppler (WSR-88D) sites, 181, 183
Web, *see* World Wide Web
WEF, 296
Weirs, 112, 179–180
Wells, 92, 166
WES, 208
West Java (Indonesia), 15
West Virginia, Huntington, 187
Westchester County (New York), 194–195
WGS84, 109
White Haven, Borough of (Pennsylvania), 103
Williamsburg (New Mexico), 355
WIN, 296
WinCan, 317–319
WinCan America, 317–319, 333
Wireless application protocol (WAP), 16
WISE, 14
WMS, 91, 208, 222
Woolpert, 210–211
Word, 68
Work orders
AM/FM and, 300, 301, 326
Cityworks and, 304, 326
CMMS and, 298
GBA software and, 294
generation of, 260, 325–326
Working for Water (South Africa), 15
World Geodetic System of 1984 (WGS84), 109
World Trade Center (New York City), 337

World Wide Web, *see also* Internet
  GIS data and, 122–128
  security, 123–124, 343
  Sensor Web, 189–190
  streaming video from, 321
WorldView satellite, 58
WSI Corporation, 183–184, 190
WSR-88D sites, 181, 183
WSSC, 40
Wyemaker, 171–173
Wyoming, Albany County, 14–15, 48–49

X

XP-GIS module, 267
XP-Software, 261, 266–267, 289–290
XP-SWMM, 261, 266–267, 289–290

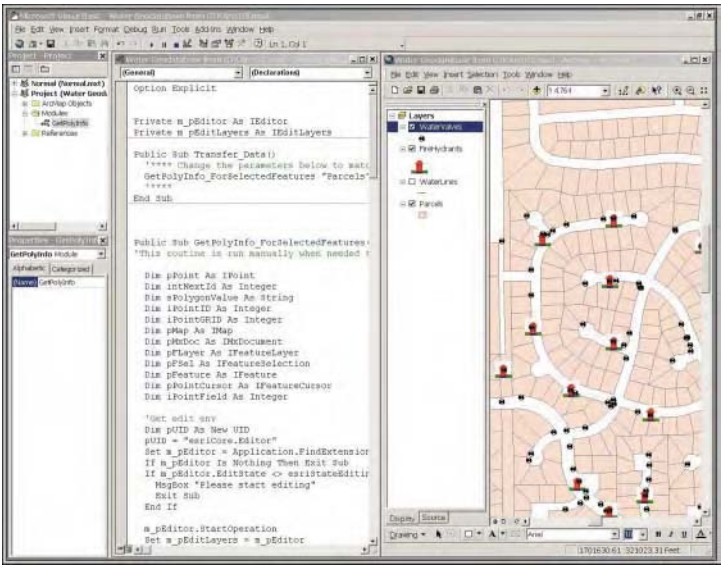


Figure 1.3 Integrated application development environment of ArcGIS (Version 8.3).

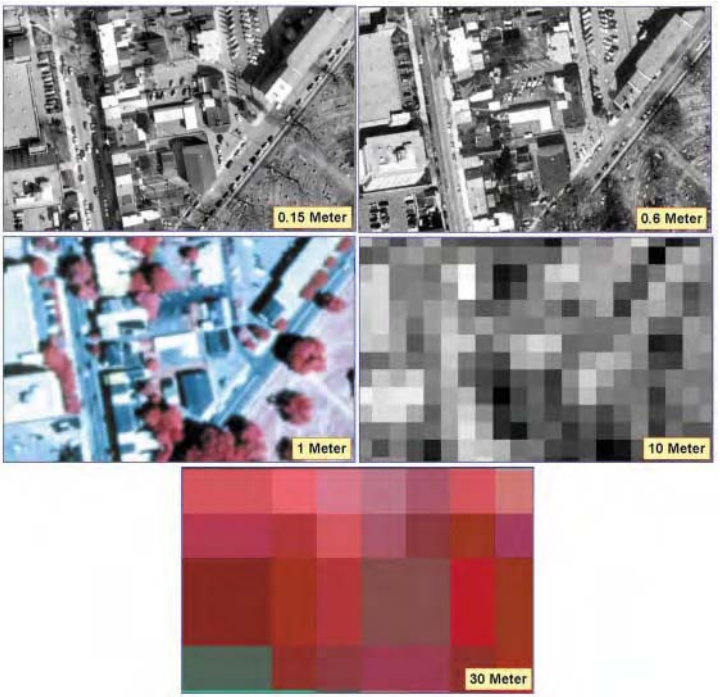


Figure 3.2 Image resolution comparison. Top left: 0.15-m B&W orthophoto (1993); top right: 0.6-m B&W orthophoto (1998); center left: 1-m color infrared orthophoto (1999); center right: 10-m simulated SPOT; bottom: 30-m Landsat TM (2000).

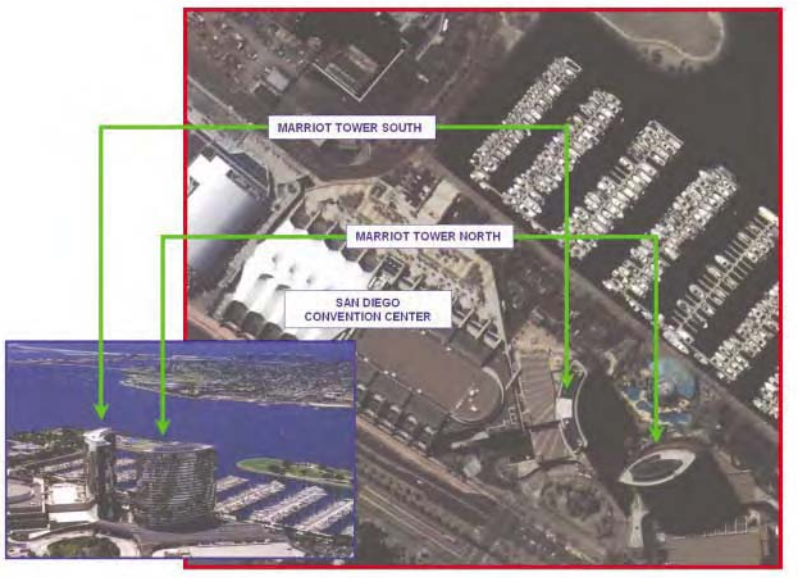


Figure 3.4 Sample QuickBird imagery for San Diego, CA. (Image courtesy of DigitalGlobe.)

Figure 4.12 Manual vs. DEM subbasins for cell threshold of 500 (best).

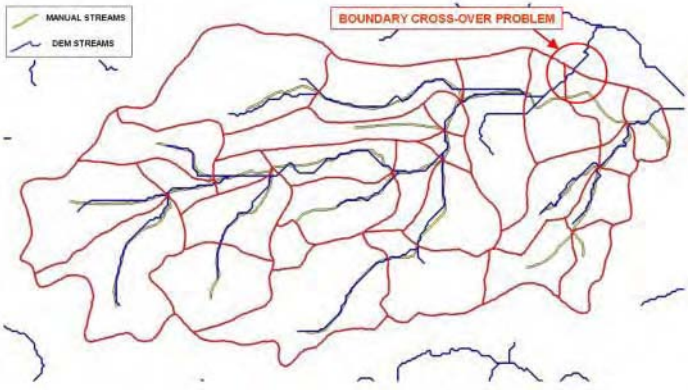


Figure 4.15 Manual vs. DEM streams for cell threshold of 500 (best).

Figure 5.3 Components of Trimble Pathfinder Pro XR Backpack.

Figure 6.1 Screenshot of Dublin (Ohio) Internet GIS Web Site.

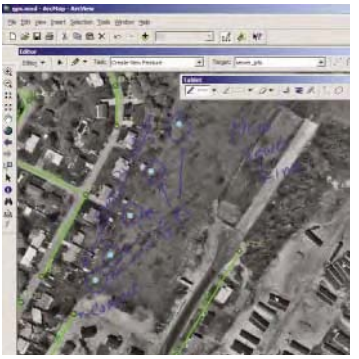


Figure 7.2 Tablet PC screenshot showing field notes and sketches.

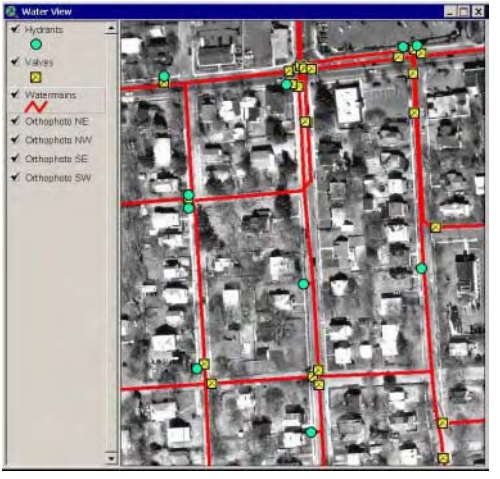


Figure 8.1 A water distribution system overlayed on a digital orthophoto base map.

Figure 8.2 A sewer system overlayed on a planimetric base map.

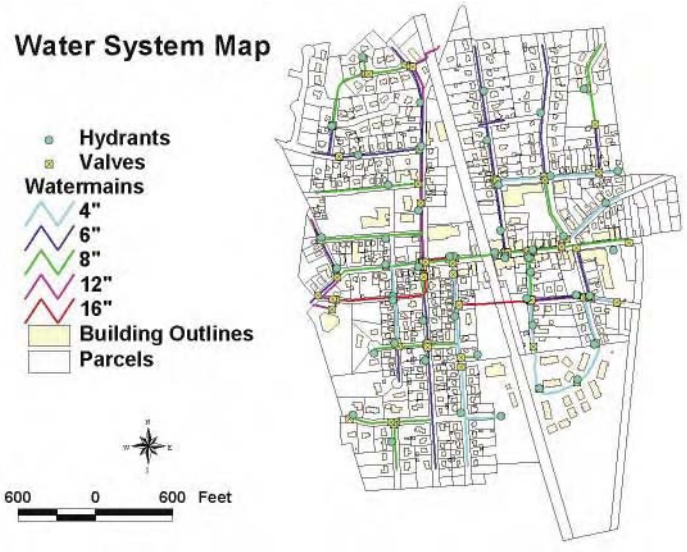


Figure 8.8 Water distribution system map showing pipes classified by diameter.

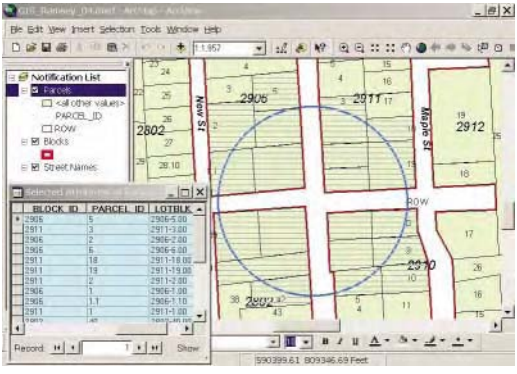


Figure 9.1 Selecting notification parcels in ArcGIS.

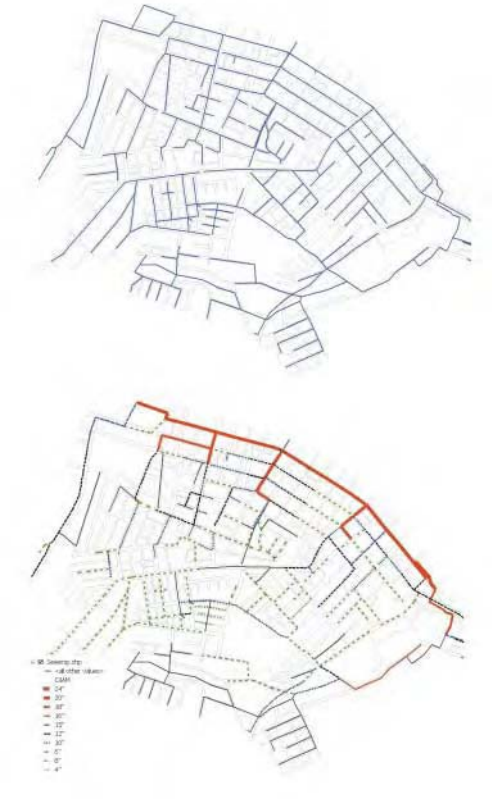


Figure 9.2 A sewer system map. Above: without attribute classification and symbology; below: with attribute classification and symbology.

Figure 9.5 3D fly-through animation of a sewer construction project.

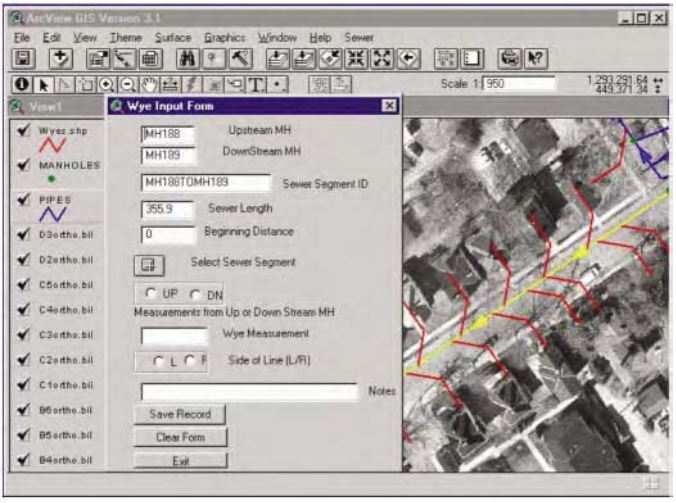


Figure 9.7 Service laterals view in ArcView.

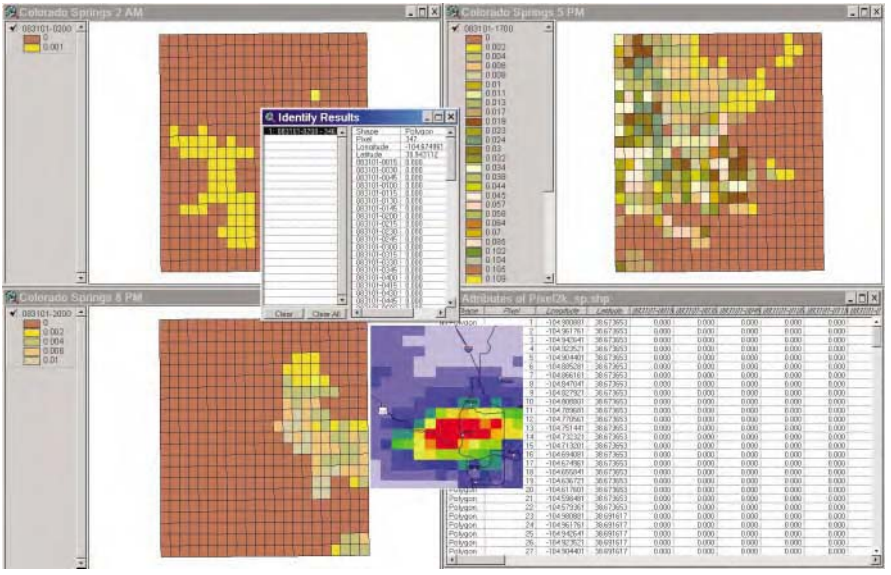


Figure 10.2 Shapefile of the NEXRAIN-2k radar pixels for Colorado Springs, CO, on August 31, 2001. Top left: rainfall at 2 A.M.; top right: rainfall at 5 P.M.; bottom left: rainfall at 8 P.M.; bottom right: attribute table; top center: pixel query window; bottom center: rainfall overlaid on street map.

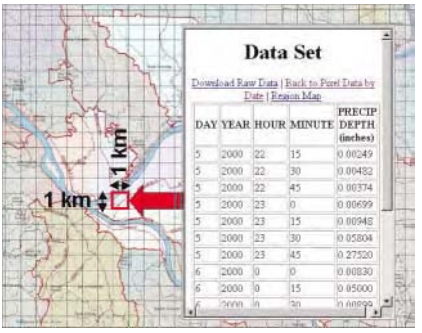


Figure 10.3 Radar-based virtual rain-gauge map and retrieved rainfall data.

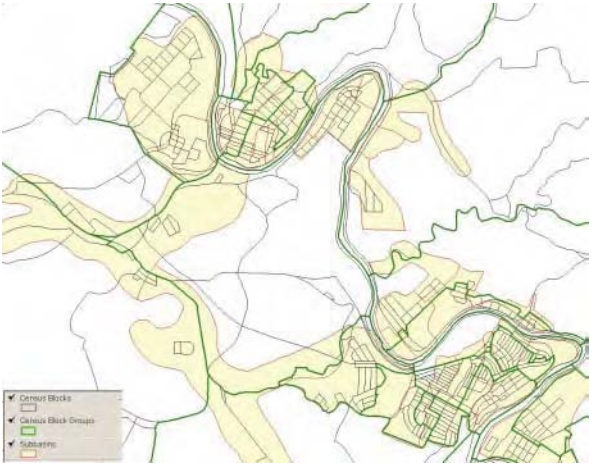


Figure 11.6 Sample census-block and block-group layers for estimating subbasin population.

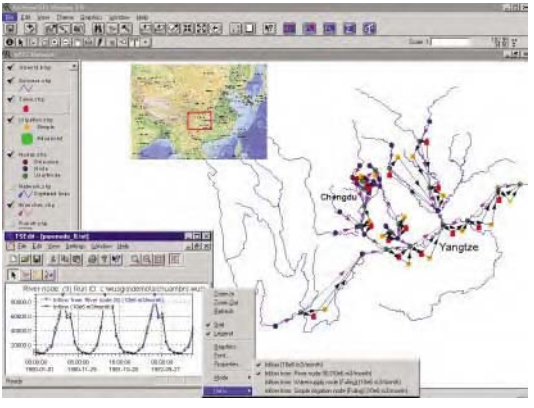


Figure 11.13 MIKE BASIN screenshot.

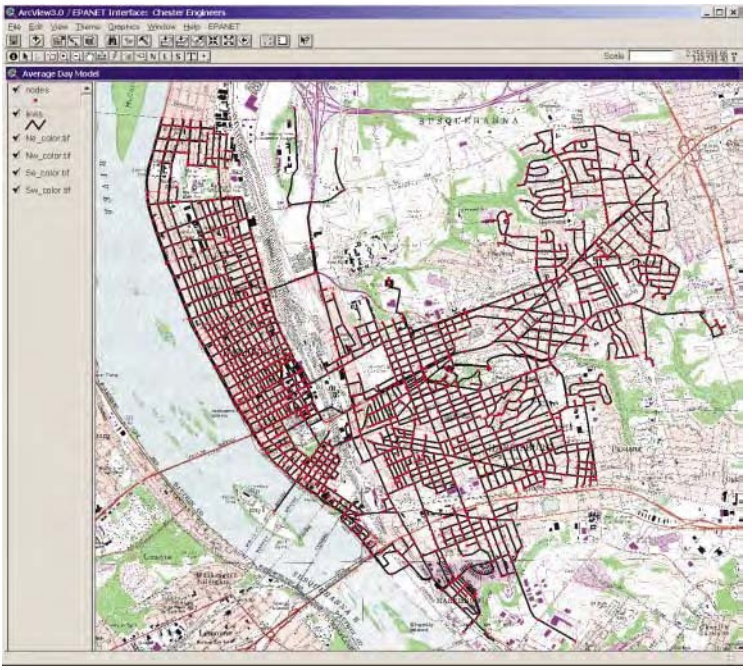


Figure 12.8 AVNET model for the City of Harrisburg, PA

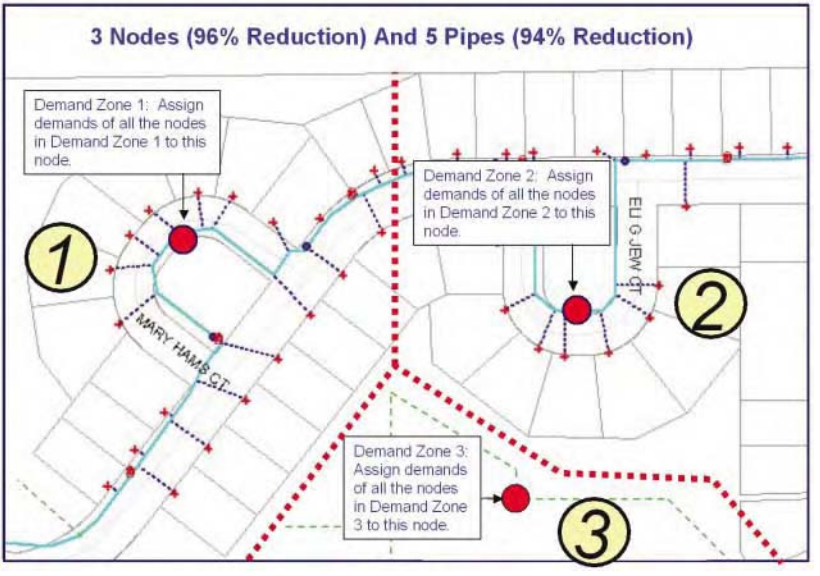


Figure 12.16 Skeletonized network model after aggregating demand nodes.

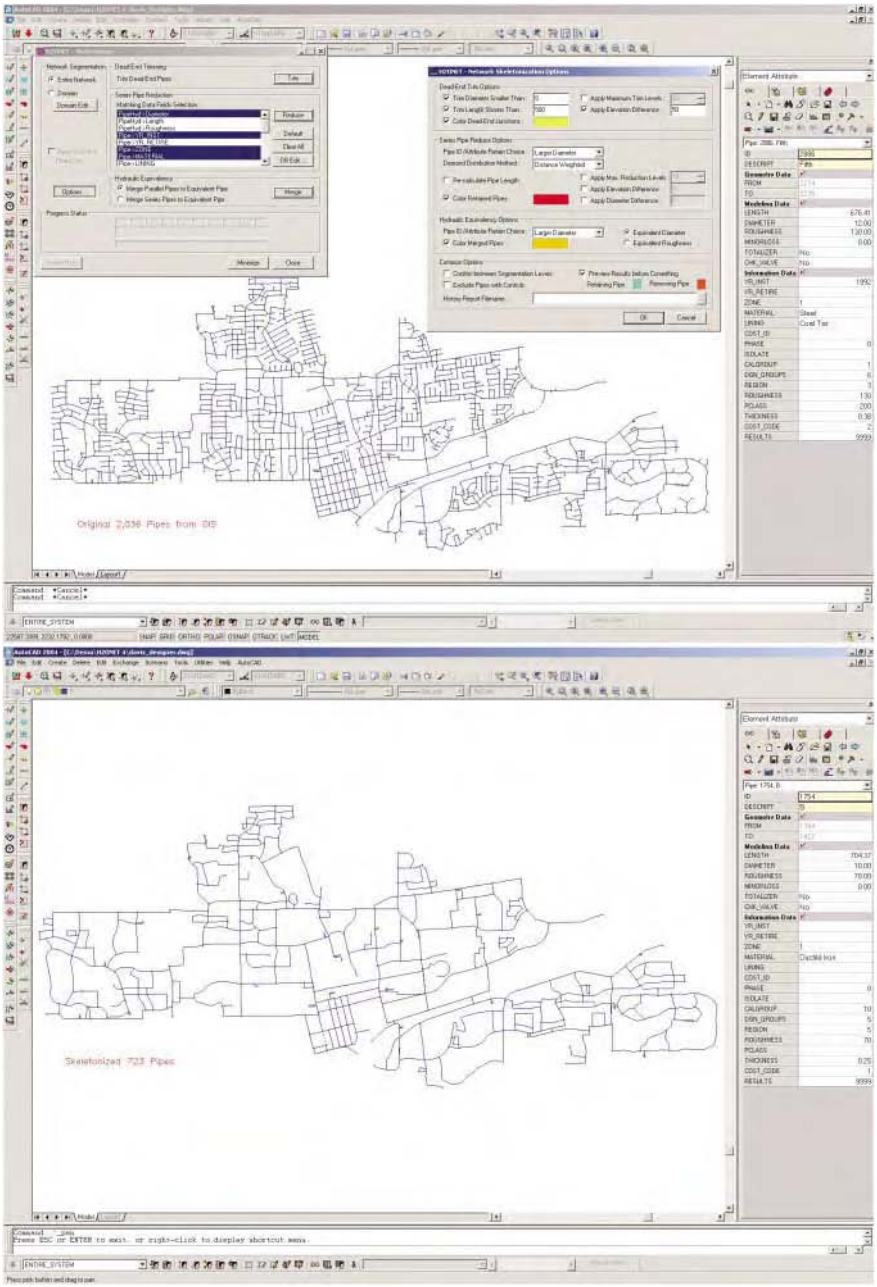


Figure 12.17 Skeletonizer's network simplification results. Top: original system with 2036 pipes. Bottom: skeletonized network with 723 pipes (65% reduction). (Maps courtesy of MWH Soft, Inc.)

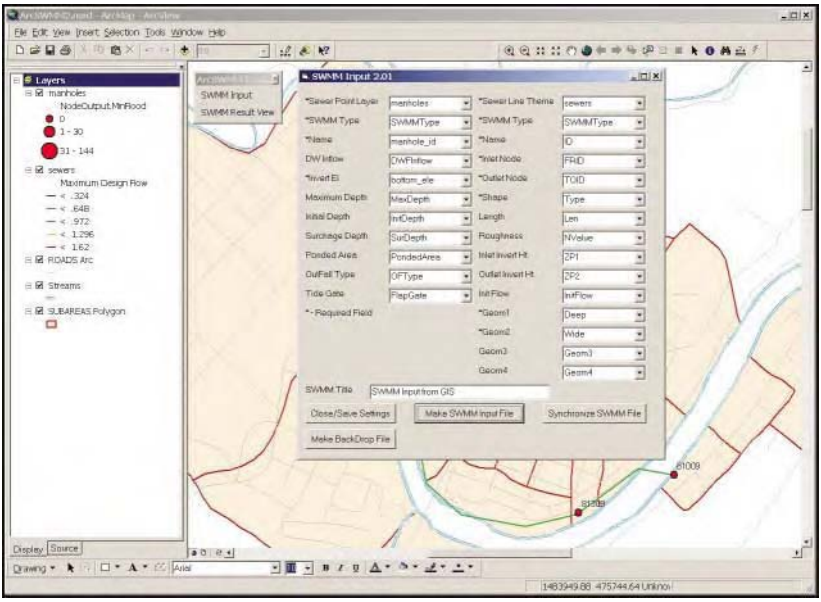


Figure 13.10 AGSWMM toolbar and input form.

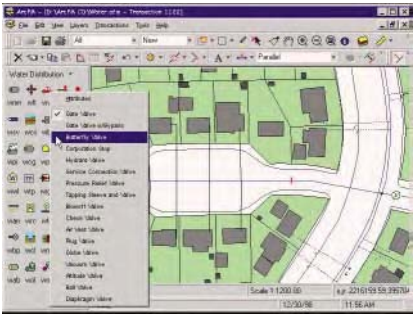


Figure 14.5 ArcFM screenshot showing water objects.

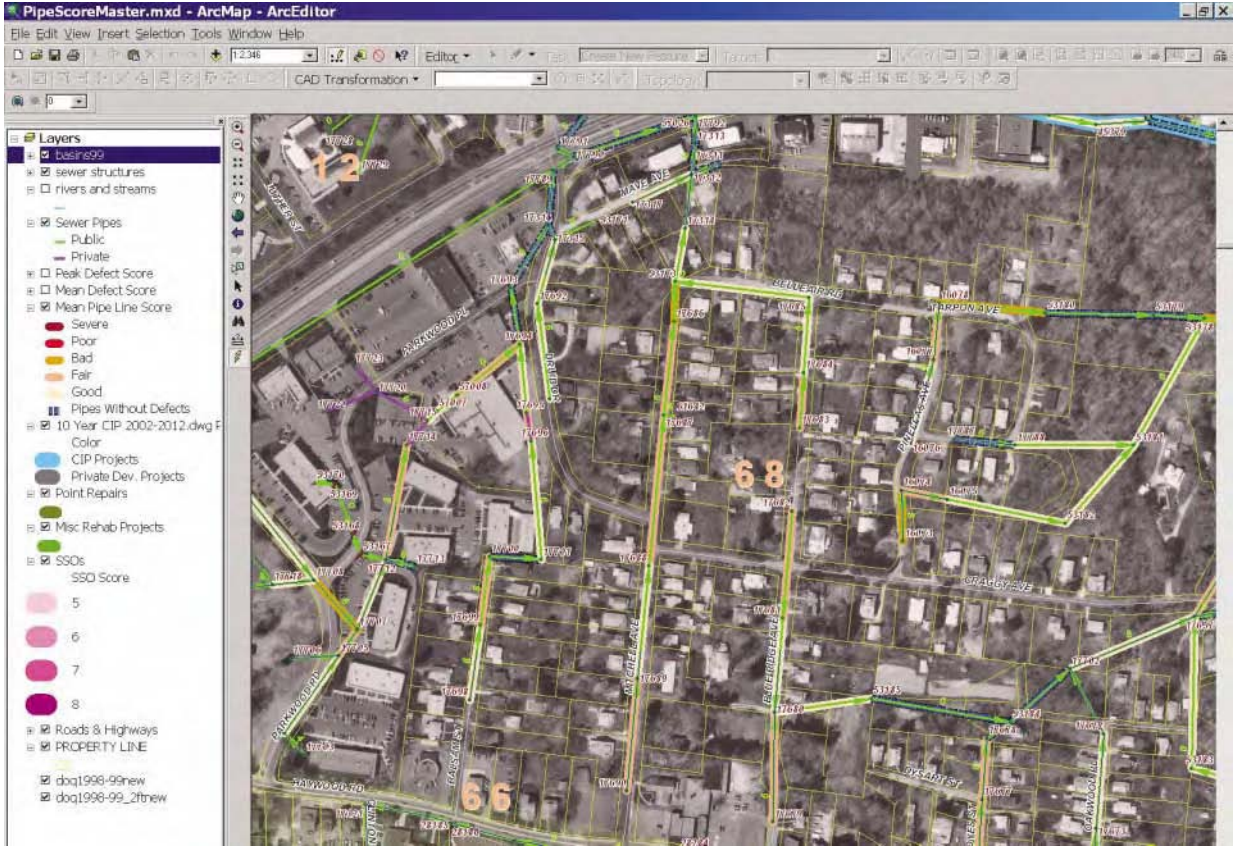


Figure 15.1 ArcGIS screenshot of mean pipeline ratings for the Metropolitan Sewerage District (MSD) of Buncombe County, NC.

Stephen R. Galati

Geographic Information Systems Demystified

Geographic Information Systems Demystified

For a listing of recent titles in the *Artech House Mobile Communications Series*, turn to the back of this book.

Geographic Information Systems Demystified

Stephen R. Galati

ARTECH
HOUSE
BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library.

ISBN-10: 1-58053-533-x
ISBN-13: 978-1-58053-533-5

Cover design by Robert Pike

© 2006 ARTECH HOUSE, INC.
685 Canton Street
Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

10 9 8 7 6 5 4 3 2 1

To my wife Janet and my children, Zachary, Nicholas, Sarah, and Jacob

Contents

Preface *xv*

Acknowledgments *xvii*

Introduction *xix*

Part I: What Is a Geographic Information System? **1**

1 How Does a GIS Work? **3**

The Fundamental Flow of Information 3

Geography and Geographic Data 5

Georeferencing 8

Queries: Locations and Attributes 8

Geoprocessing 9

Spatial Analysis 11

Geographic Information 12

2 Why Use a GIS? **15**

Discerning Your Need for GIS 16

Common GIS Products and Bi-Products 17
Database Management Systems 19
Quality Assurance, Six Sigma, and GIS 25
Project Visibility and Popularity 26
Benefits, Cost Savings, and Automation 27

3 The Structure of Geographic Data 29

Raster and Vector Data Structures 29
Vector Feature Geometry 31
Raster Image Structures 34
Topology 35
GIS Attribute Tables and Indices 38

4 Geospatial Metadata 41

What Is Metadata? 42
Geospatial Metadata and GIS 43
Metadata Schemata, Element Sets, and Syntax 45
Geospatial Metadata Standards 48
Content Standards for Digital Geospatial Metadata 50
Dublin Core Metadata Element Set 51
European CEN Metadata Standard 53
Australian Government Locator Service Metadata Element Set 54
The UK GEMINI Discovery Metadata Standard 54
Executive Orders 12906 and 13286 55
The National Spatial Data Infrastructure 56
Geospatial Metadata Clearinghouses and Publishing Resources 57

Part II: Geodesy, Earth Models, and Coordinate Systems 59

5 The Basics of Geodesy and Scale 61

What Is Geodesy? 61

Understanding Map Scale 65

Simple Map Scale Calculations 67

Scale and Accuracy 67

Geographic Data Generalization 69

6 The Ellipsoid 73

History of the Earth Model 74

Shape of the Earth 75

Mathematical Model of the Earth 77

Typical Earth Shape Calculation 81

WGS84 Constants 81

Reference Ellipsoids 82

7 The Horizontal Datum 85

What Is a Geodetic Datum? 86

Primary Types of Geodetic Datums 87

Standard Horizontal Datums 89

World Geodetic System 1984 89

International Terrestrial Reference Frame 90

Geodetic Reference System of 1980 91

North American Datums of 1927 and 1983 92

European Datum 1950 93

Horizontal Datum Transformation Models 94

The Molodensky Datum Transformation Method 95

The Bursa-Wolf Transformation Method 96

Horizontal Grid Transformation Methods 99

The North American Datum Conversion 99

The High Accuracy Reference Network 100
Canadian and Australian National Transformations 100
Reference 101

8 The Vertical Datum 103

Mean Sea Level and the Geoid 104
The Geoid-Ellipsoid Separation 106
Standard Vertical Datums 107
Earth Gravitational Model of 1996 107
GEOID03 and Predecessors 108
National Geodetic Vertical Datum of 1929 108
North American Vertical Datum of 1988 109
Vertical Datum Transformation Methods 110

9 The Map Projection 113

The Physicality of the Earth 113
Scale Reduction: From the Ellipsoid to the Globe 116
Transformation from the Globe 116
What Is a Map Projection? 117
The Grapefruit Peel Experiment 119
Choosing a Map Projection 120
Types of Map Projections 122
The Cylindric Type 122
The Conic Type 122
Planar Type 124
The Aspect of Projection 125
Classification of Map Projections 126
Equidistant Projections 128
Azimuthal Projections 128
Equal-Area Projections 129
Conformal Projections 130
Distortion 132

Distortion Type 132
Distortion Magnitude, Distribution, and Overall Effect 132

The Tissot Indicatrix 133

Methods for Distortion Management 136
Redistribution of the Scale Factor 136
Changing Projection Aspect 137
Other Successful Methods 137

10 The Coordinate System 139

What Is a Coordinate System? 139
The Geographic Coordinate System 140
The Cartesian Coordinate System 142
Universal Transverse Mercator 145
The Universal Polar Stereographic 147
The State Plane Coordinate System 148

Coordinate System Implementation in GIS 150

Part III: GIS Applications and Environments 153

11 Thematic Mapping 155

The Need for Thematic Mapping 156

Manipulating Thematic Layers 157

Types of Thematic Maps 159
Choropleth Maps 160
Graduated Symbol Maps 161
Dot Density Maps 162
Isopleth Maps 163

Mapping Data, Classification, Portrayal, and Distinction 165
Counts, Rates, and Densities 165
Data Classification 166
Projection and Scale 167
Legends, Color, and Symbols 167

Common Cartographic Elements 168

12 GIScience, Engineering, and Related Applications 173

Using GIS for Technical Applications 173

GIScience: Principles and Applications 174

Case Study: Using GIS to Study Relationships Between Bedrock Geology and Subtle Topography 174

Engineering Applications 175

Case Study: New Jersey Natural Gas Transmission Line Property Delineation 176

Marketing and Financial Applications 177

Case Study: Analyzing the Rapid Closure of a Cincinnati-Based Office Depot 178

13 GIS and the Environment 181

Using GIS for Environment Applications 181

Managing Natural Resources 182

Case Study: Lebanon Natural Resources Management with Participatory GIS 183

Case Study: Community-Based Natural Resource Management Program in Namibia 184

Environmental Investigation and Remediation Applications 185

Case Study: Illinois Underground Storage Tank Monitoring and Inspection Program 186

Case Study: The USEPA's Disaster Dioxin Air-Monitoring GIS System 186

Shoreline Management Systems 187

Case Study: The Coosa/Warrior Shoreline Management GIS—Alabama Power Co. 188

Other Environmental Applications 188

14 GIS and the Public 191

A Public Need for Information 191

Managing Public Health and Safety 192

Case Study: Detailing Areas at High Risk for Elderly Fall Injuries to Enhance Injury Prevention Programs in Alberta, Canada 193

Disaster: Planning, Response, and Recovery 194

Transportation and Information 195

Case Study: Portland's TriMet Transit Tracker Online 196

Homeland Security's Third Eye 196

Public Education Applications 197

Redefining Public Policy 199

Case Study: Public Policy Education in Gaston County, North Carolina 200

15 Getting Started with GIS—An Overview of Environments, Tools, and Data 201

Getting Yourself Up and Running with GIS 201

Leading For-Purchase GIS Environments 202

GRASS and Other Free GIS Environments 205

GIS Component Software 208

Geospatial Data Clearinghouses 210

Afterword 213

Appendix A: Reference Ellipsoid Parameters 219

Appendix B: Worldwide Geodetic Datums 223

Glossary 233

Bibliography 249

About the Author 259

Index 261

Preface

The field of geographic information systems (GISs) is emerging as one of today's most exciting and progressive technical areas. GIS technology has evolved during the past four decades and is only now starting to penetrate major industry and service hubs. With such a wide array of new users, there exists not only considerable interest in GIS but also a fair amount of uncertainty and misunderstanding surrounding the discipline.

Like other technical fields, there have been various methods employed to convey the fundamentals, implementation, and importance within the industry. My personal approach within this book is to encourage the highest quality understanding of GIS fundamentals before suggesting the reader study through application. In taking this route, this book attempts to disband user uncertainty and misgivings through guiding discussions about GIS rudiments and technical theory. With this in mind, I want to clarify the book's distinct structure, level of technical information, and somewhat hands-off approach.

The book is structured into three distinct parts:

- Part I: What Is a Geographic Information System?
- Part II: Geodesy, Earth Models, and Coordinate Systems
- Part III: GIS Applications and Environments

Although the specific content of each division is summarized in the Introduction, I want to offer some insight into each part's overall intent and degree of technical detail. On the surface, the technical level varies among the parts, but with very good reason. Let me briefly explain.

Part I presents the preliminary background of what a GIS is, why it exists, how it works, and what forms of data it uses. Within this first division, my primary focus is to get readers up to speed on GIS's composition and I take an explicitly less technical path to achieving this. A more technical approach to this background information would leave the reader swathed in tangential theory and would needlessly add further complexity to an already intricate subject. The topics in Part I are supportive in nature and are approached as such with distinct technical brevity.

Part II is the true "nuts and bolts" of GIS and, as such, is approached with a much greater technical treatment. My approach to Part II is to comprehensively explain the theory behind GIS's core components, such as the reference model of the Earth, the coordinate systems, and the map projection. As you will notice, Part II is dramatically more complex in nature than Part I and this is done for a specific reason: Most GIS uncertainty, misgivings, and misuse extend from a limited understanding of these core components. I believe a comprehensive understanding of these topics is essential for effective, proper, and creative GIS use.

Part III takes a step beyond the theoretical to provide an overview of geographic data presentation, real-world applications, and inexpensive (practice) systems and data. To this end, Part III is technically less complex than the previous part for the sole reason of conveying a clearer vision of how GISs are actually used, rather than overwhelming the reader with a thorough discussion on each element.

I want to highlight one other feature—a benefit—to my approach: It is theory based rather than hands-on based. I call it a benefit because true hands-on approaches lead the reader to focus primarily on the use of a GIS (typically on a specific system) rather than on understanding a GIS. Presented too early, this hands-on method, I feel, becomes an impediment to comprehension. A quality theoretical understanding is the key ingredient to effective and insightful real-world use. A hands-on approach restrictively focuses on a particular system, while a theory-based approach lets the user (reader) freely experiment and choose the system on which to learn. This is not to say a hands-on approach to GIS is a wrong learning method; I am simply stating that for the best use of a GIS and to gain the greatest amount of user perspective, the theory-based approach proves beneficial and leaves the reader at the doorstep of hands-on practice.

I hope you enjoy this book and gain a deep appreciation and excitement for geographic information systems.

Acknowledgments

Many people provided their support, guidance, and encouragement during the research and writing of this book. I wish to express my thanks to the following:

Dr. Tyler S. Pruett, for his significant input in the book's overall structure and his help with the writing and technical review of some important chapters, particularly Chapters 1, 3, 9, and 12. I am grateful for his early guidance and support of this book.

Dr. Paul Aplin at the University of Nottingham, United Kingdom, for his comprehensive technical review and keen eye for my many errors. His superior technical and structural suggestions have enabled me undoubtedly to write a much better book.

René Essiembre, for his assistance and support during the writing of some seriously technical chapters. I am grateful he let me "pick" his GIS-brain.

I have also benefited from the help and support of John R. Kubasek from John R. Kubasek & Associates, Inc.; John Pavek and Glenn Small, PE from T&M Associates, Inc.; and Van Bowersox from the National Atmospheric Deposition Program (NADP).

I wish to express my tremendous gratitude to Mark Walsh and Barbara Lovenvirth from Artech House. Mark's support of this project has remained steadfast from the very beginning and throughout the thick and thin of the book's development. Barbara has been equally committed to this project throughout its development and has shown me remarkable support and guidance. I thank them wholeheartedly for believing in this project.

My heartfelt thanks go to my beloved family—my wife Janet, my youngest son Jacob, and my adoring triplets, Zachary, Nicholas, and Sarah. My family has inspired me to continue writing "the book" and has shown me much needed enthusiasm when mine was waning. Their support and love is endless as is my love for them.

Finally, and most importantly, I want to thank God for giving me the fortitude to complete this book.

Introduction

The 2004 U.S. presidential elections marked a milestone in how information is used and portrayed. For the first time ever, analysts used geographic information systems (GISs) to generate election maps that showed which presidential candidate was leading in each county in every state throughout the country. This information was well received by political analysts and home viewers alike, though few understood the *thing* called GIS. To most, the highly visual election results were the products of some "black box" that offered amazing analysis tools. Understanding how the "black box" worked was not their concern.

The perception of GIS in the business world goes much the same way. Many engineers, environmentalists, and business professionals who use GIS do not necessarily know how it works. This is not unfounded since groundbreaking technology typically evokes initial skepticism and uneasiness. Do you remember the initial unease associated with the "smart machine" in the early 1980s? Back then, personal computers were new and quickly embedding into the worldwide society. This "smart machine" apprehension resulted in the typical person using computers only for word processing and merely tapping a small percentage of the computer's true capability.

GIS is not secluded from these issues of technology fear. Actually, modern GIS environments are almost 100% dependent on the "black box." Without software technology, GIS as we know it would cease. Consequently, the GIS community experiences the same uneasiness among its users. Many GIS novices, and even some intermediates, use but a fraction of the functional capability of GIS software systems, minimizing the system's ability to that of limited queries and static mapmaking. Others politely refrain from using GIS at all, believing it will be complicated, overly technical, or a waste of time altogether.

The rapid growth of GISs worldwide, particularly in the 1990s, has challenged users with the void of GIS standards and software market saturation. Company after company has injected GIS software products into the open market, transforming the flux of GIS capitalism into a sequence of software start-ups and reinventions. Sadly, this has left users, both novice and advanced, apprehensive about what is the right software environment for their time and money investments. Only recently has this lack of standards become an evolving topic in the GIS community, to which the ultimate resolution will be a homogeneous GIS user interface.

Pushing beyond the apprehension to learn GIS requires not only graphical consistency in the GIS platform and environment but also the recognition that GIS is an iterative discipline. *Learn by doing* is a caveat to any valid GIS approach, although to get the most out of a GIS, the user must not only use the system but also understand its potential. Even a rudimentary understanding, such as a basic grasp of fundamental concepts, can intrinsically help with the conveyance, analysis, and final usage of geographic information. With understanding, these modern-day pioneers have at their disposal a powerful analysis tool that will open new avenues of data analysis in their applications. The success of GIS for the 2004 elections only proved that a basic understanding could provide amazing tools for analysis and real-time information conveyance.

GIS has been around for nearly four decades but has gained a true user backing only within the past 10 years. With technical and industry professionals becoming aware of GIS's capability and benefits, the dedicated following of developers and users is now growing. The next 10 years will mark new milestones in geographic information, while everyday life will continue to embed with GIS applications.

A Definition of Geographic Information Systems

So what is a GIS? A *geographic information system,* as defined in the Environmental Systems Research Institute's (ESRI's) *Dictionary of GIS Terminology*, is a collection of computer hardware, software, and geographic data for capturing, storing, updating, manipulating, analyzing, and displaying all forms of geographically referenced information. This is a very general description for such a complex and wide ranging set of tools. GIS is, in essence, a central repository of and analytical tool for geographic data collected from various sources. The developer can overlay the information from these various sources by means of themes and layers, perform comprehensive analysis of the data, and portray it graphically for the user.

In the case of the 2004 elections, a layer discerning counties overlaid a base map of the United States. County demographics were developed into the GIS

prior to election day and integrated other geographic data sets such as candidate information. These data sets for county demographics are readily available.

Throughout election day, real-time data from each reporting precinct in every county fed into the GIS and was cross-referenced with the existing data. The result was three-dimensional mapping that illustrated not only candidate leaders and preelection poll accuracy in each county but also voter turnout by socioeconomic status and party affiliation.

GISs hold a wide range of functionality and analytical capability that with the right developer can provide a collection of useful data and maps. Although the method of presenting geographic information has transformed throughout the ages, the individual uses have only expanded. In fact, such yesteryear applications as farming, sailing, cartography, and warfare still use geographic information.

Poincaré's Maps and the Move Toward GIS

Jules-Henri Poincaré can be considered one of the first bridges between the old technology of cartography and the first traces of GISs. Poincaré was one of France's most innovative thinkers of the late nineteenth and early twentieth centuries. His work covered the landscapes of mathematics, time, physics, relativity, and geodesy. He used his geometric visualization methods, which portrayed non-Euclidean geometry through "visually" representative maps, to fill in the blanks on the world map.

Poincaré began his prestigious career examining coal mines throughout France. While working as a mining engineer, Poincaré was instrumental in determining the root cause for a deadly explosion of a mine pit in Magny. He conducted a meticulous and logical investigation of the events while maintaining consideration of the mine's structure, geology, worker locations, and reason. Poincaré drew highly detailed maps of the mine that indicated the flow direction of air through the mine shaft. This gathering of information on a single diagram assisted Poincaré in determining that an accidental puncture of a Davy "safety" lamp caused the ignition of the mine's resident methane gas. He also provided the exact location of the ignition and exact lamp punctured (lamp 476). After the Magny incident, Poincaré became reputed for his visualization techniques.

In 1889, Poincaré's notoriety spread with the presentation and acceptance of his visual mapping method. He plotted the habits of a small asteroid repetitively orbiting around Jupiter and the Sun (known as a three-body problem). He intentionally neglected the examination of the entire orbital path and, ingeniously, focused on the orbit's impact points within a single plane in space. The result was a stroboscopic diagram identifying a complex series of plotted impact points. This plotted diagram caused a stir in the scientific community because it

identified chaos, or instability, within the natural process of the universe. The mapping method, now known as Poincaré's Map, intertwined geography, location, visualization, data, and analysis.

In the following years, Poincaré gained access to France's bureaucracy, acting in various high-level, high-visibility positions. As the Bureau of Longitude's scientific secretary, Poincaré became heavily involved with the escalating and competitive telegraphic longitude work. France had not made any important impacts as had other worldwide entities, such as the United States and Britain. France and Poincaré wanted to establish telegraphic simultaneity, a primitive methodology compared to today's global positioning system (GPS) technology. At the same time, the International Geodetic Association desperately needed an accurate length measurement of a meridial arc to remeasure the shape of the Earth. France jumped at the chance.

In 1900, Poincaré headed a mission to help France's Bureau of Longitude accurately plot the region of Quito, Ecuador. Although his expedition teams in Ecuador met misfortune every step of the way in the difficult, obstacle-filled terrain, Poincaré was successful. After 7 years, Poincaré completed the mission and created a network of latitude and longitude measurements that later were connected by telegraphic cable and linked to the world telegraph network. Poincaré, once again, used his visual geometrical methods to help detail the measurement data and cable pathways upon a geographic map.

Through his many exploits, Poincaré used mathematics and ingenuity to expand the parameters of map visualization, physics, geodesy, and scientific concept. In doing so, he revolutionized methods of using geodetic information to widen the realm of analysis. Undoubtedly, Poincaré left his mark as an early pioneer to modern-day information systems, such as GIS and GPS, and his Poincaré Map was a predecessor to present-day chaos theory.

Old Technology, New Platform

As seen through the work of Poincaré, the concept of a GIS is rather old. Ancient cartographers used observation, field informants, celestial positions, and measuring tools (such as the sextant) as inputs to calculate distances and determine land formations. Hand-drawn maps then depicted this information. Maps indicating demographics, such as the layout of crops throughout the nation or population, were based upon field information and generalization and were not completely accurate. Usually informed generalization and estimation helped develop the maps.

As technology increased, so did the precision of maps. Graphic landforms more accurately portrayed reality, while maps better portrayed demographics. Color palettes, symbol sets, and attribute tables depicted data.

Today, geographic information is at its most precise. As the new platform for real-time data, GIS utilizes high-resolution satellite imagery, light intensity detection and ranging (LIDAR), GPS data, computer-aided design (CAD) files, and enterprise database management systems for complex analysis. Users can cross-reference these geographic data sources in a GIS and create impressive geographic displays catered to an area of interest and to individual or specific applications. Users now have the opportunity to customize their maps not only to show the information they want but also how they want it. For instance, an engineering firm widening a section of interstate roadway not only can use a GIS to mark a satellite map of the roadway section with bridge locations, exit and on ramps, and elevations but can also map soil boring points with hyperlinked data, traffic data, surrounding vegetation types, and roadway drainage.

GIS is appearing in almost every industry, forging its place as the new platform for geographic information. GIS is now used for education, land management, natural resource management, environmental, and aeronautical applications, just to name a few. GIS even crosses every industrial and humanitarian threshold with software and data development efforts. The only limits to GIS are the user's fundamental knowledge of the system and overcoming the stigma of creative GIS use.

The GIS Mystique

Whether it is the result of users' anxiety, disinterest, or uncreative whim, GIS bears a cross of a subtle mystery. This GIS mystique, as I call it, is the ultimate result of misconception and limited knowledge. Let me explain.

GIS holds a preconception that is grossly different from reality. For instance, GIS is sometimes solely viewed as mapmaking software. This is simply not true. GIS has found its way into everyday life, from the internal management of utility companies that deliver electricity and water to mapping distant planets, such as the European and U.S. expeditions on and around Mars and Venus. The reality of GIS, as this book details, is that it can do so much more than make a map.

Similarly, GIS is sometimes wrongly linked on a one-to-one basis with GPS technology, found in handheld receivers, E911 systems, and cellular phones. This user misconception "dumbs down" GIS technology to the point that it severely limits the ultimate capability and potential of a properly structured system. GPS data is but a small portion of the information a GIS manages and, in all honesty, is not needed for every user's application.

To overcome this communicable mystique, we must first understand that there is still not enough hands-on work with GIS by all levels of users. The users

are only interested in information at their fingertips from the "black box," while management rarely understands the levels of visualization in GIS or how it can be successfully implemented. Consequently, many GIS development efforts are underfunded and prematurely abandoned.

A properly structured GIS requires time, and many users are either not willing to invest this time and money or not afforded the opportunity to design a system that maximizes capability. As competition increases, more users hastily implement a GIS to stay in line with their market and, subsequently, promote the fact they use GIS technology. If truth were told, many industry people want the term "GIS" on their laurel list, though they have little intention of using the software. Many systems suffer from this lack of use and grow stagnant. Competition, along with users' indolence, dilutes GIS structure and scientific implementations.

About This Book

Geographic Information Systems Demystified is written to give readers a digestible approach to understanding what a GIS is and how it can be used, while revealing the falsehoods of the GIS mystique. This book is not all encompassing but rather focuses on GIS theory. It presents the important elements of GIS to get anyone started and ready to work with certainty, and provides candid insight into this often unfamiliar technology science.

While many GIS learning sources are software product biased, the book offers an unbiased approach to the fundamentals of learning GIS, helping readers become functional and knowledgeable users. This guidebook also presents an original, comprehensive, and innovative treatment of spatial coordinate systems, a topic rarely covered at all in most popular GIS learning titles. The book examines and explains critical GIS concepts in a consistent structure, demystifying an often misunderstood discipline.

The structured topics first lay down the basics of GIS and then build upon this knowledge foundation with more specific information. The book is laid out in three parts:

- Part I: What Is a Geographic Information System?
- Part II: Geodesy, Earth Models, and Coordinate Systems
- Part III: GIS Applications and Environments

Part I focuses upon what a GIS is, why is it needed, the structure of geographic data, and the importance of metadata. It offers an understanding of how geographic data transforms into geographic information through a succinct

presentation of geographic information basics and solidifies the knowledge foundation for the entailing topics found in Part II.

In Part II, the discussions tackle the more complex elements of GIS applications and, quite different from Part I, are much more technical and comprehensive. Core topics, such as map scale, Earth models, and map projections, are scrupulously defined and add to the host of GIS skills designed to give the user creative control over the analysis, portrayal, and use of data. The most common GIS tools are discussed in technical detail geared to get anyone quickly up and running with their own GIS applications.

Part III highlights some of the creative tools and techniques presently available within the discipline of GIS, as well as how GIS is used today in various industries. It details how geographic data can be uniquely used for analysis and presentation. Topics such as thematic mapping and real-world examples offer the user ways to creatively depict their geographic data and widen the spectrum of possible GIS applications. Also addressed are places to get free GIS environments, component software, and geographic data.

This book includes useful appendices and a glossary that may be used in tandem with the text. This information offers handy standards for ellipsoid parameters and datum transformation tables to use. This information was useful in the writing of this book and I felt it would be equally useful to the readers. The glossary defines words and phrases used throughout the book but should not be viewed as a comprehensive glossary of GIS terminology. If you wish for a more thorough glossary, check out some of the resources in the bibliography.

The Book's Approach to GIS Concepts

With GIS's realistic business and engineering models offering attractive cost savings and efficiency, geospatial systems can sometimes transcend practical concerns. The technology injects a touch of creativity into the veins of global infrastructure and serves decision makers on all levels of business and financial permanence. Important new GIS concepts are born every day, and industry trends fluctuate so frequently that a tempered approach to introducing GIS concepts hardly seems fitting. This book attempts to show GIS science and industry as they are, providing a literal roadmap through the industry's landscape.

We must remember that GIS is still brand new to the overwhelming majority of desktop computing users, as well as the technical and business fields. The importance of understanding fundamental GIS concepts is clearly underestimated in the industry today. This book addresses the most important concepts in earnest and details methods for using and developing geospatial systems. Every critical concept is discussed in a clear, consistent voice that is intended to raze any false preconceptions the reader may have.

Since GIS is an international infrastructure organization tool, the book takes a worldwide approach to the technology. GIS is booming throughout the world, especially in Europe, the Americas, India, Japan, and Australia. Topics are designed to further the overall understanding of GIS as it relates to worldwide applications.

The Goal to Demystify GISs

This book's main goal is to take the fear of the "smart machine" out of the user and expose the inherent mystery of GIS. Geared to the beginner and intermediate user, this book intends to forge a paced, seemingly trolling, pathway to productive understanding and unveil the wide range of user choice and creativity.

GIS is being used in almost every industry and affects some portion of every person's daily life, usually without awareness. Whether you realize it or not, your life is affected on some level by GIS technology. GIS is currently being used in almost every conceivable industry. Major GIS software companies boast more than a million daily application users. In fact, if you have ever used MapQuest, you have explored geographic data and performed analysis using GISs.

Geographic Information Systems Demystified intends to herald the broad availability of worldwide geographic information, diminish apprehension of the GIS unknowns, and widen the user's eyes to the great possibilities that can be achieved using GIS. In essence, this book aims to rip open the top of the "black box" and reveal the wonderful things that lie hidden inside.

Part I:
What Is a Geographic Information System?

From Geographic Data to Geographic Information

1

How Does a GIS Work?

In a geographic information system (GIS), geographic data are transformed into geographic information. This simple transformation, however, involves a complex series of functions and processes. In a nutshell, geographic data begins as raw positional feature data holding attributes. These data are then overlaid with complementary and/or contrasting data sets, which form coincident relationships. Data and relationships are analyzed, geoprocessed, and then presented as geographic information products. These geographic information products are often interactive software applications used to help people make decisions.

GISs are accessible to an array of users, from the expert GIS software developer to the GIS novice project manager, and, subsequently, offer visualization to users throughout the spectrum of skill levels. This diversity becomes a unique benefit of GIS and explains how quickly it can become visible within an organization, as well as offering project visibility to the public.

To fully understand the power of a GIS, we must first take a closer look at how raw geographic data becomes usable geographic information. The following sections depict the movement and manipulation of geographic data sets, as well as the areas where user ability is vital to usable output. The core geographic data transformation involves the following fundamental flow of information.

The Fundamental Flow of Information

Geographic data originate from actual locations and physical characteristics of features on or near the surface of the Earth (or other celestial bodies such as Mars or the Moon). These raw, positional data are the start points for every GIS by providing the basic geographic information needed for attribution, dataset

modeling, relationships, and analysis. Refer to Figure 1.1. Raw data can come from a range of sources, such as aerial photographs, previously digitized maps, and global positioning systems. These types of digital and nondigital geographic resources are readily available and, in many cases, are plentiful, well designed, and comprehensive. Additionally, many digital sources are free. Other more labor intensive sources are field data and measurements collected from site visits, and transformed maps, whereby old hardcopy maps are first scanned into a computer and then digitized.

In the most fundamental sense, raw data are either geographic data or are transformed into geographic data through a GIS, and are in turn used to produce geographic information products. Geographic information products are user-conceived information results created through a GIS and a user's ability to relate, manipulate, and present overlaid geographic data. These products are used to analyze data for a specific application.

The overall work of transforming data in a GIS can be summarized through its three distinct procedures:

- *A GIS leverages the flexibility of geographic data.* Raw data are static (nonchanging) and offer only a limited amount of flexibility on real-world applications. When raw data are transformed to geographic data through a GIS, the capability for enhanced data use and analysis (i.e., data flexibility) significantly increases. At a minimum, overlaying two geographic data sources provides sufficient new information, better analytical means, and additional flexibility to not only help someone visualize the real world but also help them make an informed decision.

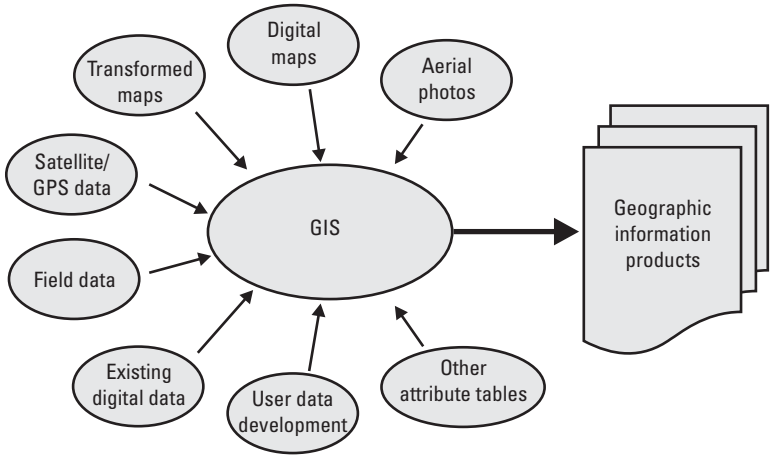


Figure 1.1 The path of geographic information.

- *A GIS performs functions and analysis within a single environment.* These data functions are the literal "doers" of a GIS solution and are known as *geoprocessing* and *spatial analysis.* These operations are available within a single GIS environment and include the generation of features, buffers, view sheds, and cross sections; the calculation of centroids, slopes, statistics, and suitability; and the manipulation of feature attributes, smoothing of lines, feature transformations, and clipping. Both geoprocessing and spatial analysis are discussed in more detail later in this chapter.
- *A GIS serves as a software application and creates useful information products.* GIS environments, foremost, serve as spatially enabled enterprise data management systems and data repositories. These systems are software applications that protect the value and usefulness of the information related to your project. The end result is an information product that enables the user to better manage his or her project.

Through this procedural flow of information, geographic data are transformed into geographic information. GIS environments centralize both data collection and information management to save time, minimize technical effort, and automate known repetitive administrative tasks.

The core data component of a GIS is often represented by a *geographic data model,* which is an industry or discipline-specific template for geographic data. A geographic data model offers the user flexibility in the design of the file management and database hierarchy. Geographic data models typically utilize a grid-based structure (known as raster) or a coordinate point structure (known as vector). Both models are explained in detail in the upcoming chapters.

Facilitating the model is one or more geodatabases. A *geodatabase* is a collection of geographic data sets, real-world object definitions, and relationships. Comparable to a Microsoft Access file, a geodatabase is a collection of geographic data sets and geometric features. A geodatabase furnishes the data organizational structure and workflow process model for the creation and maintenance of the core data product. In essence, the geodatabase is the heart of a GIS's management capability.

Let us now look at each element in the flow of information. We will start at the beginning with the various forms of raw geographic data and work our way to the GIS products.

Geography and Geographic Data

Geography is the study of the Earth's surface and climate, and is the founding science to a GIS. Geography furnishes information about the Earth and

distinguishes how features upon the Earth correlate with one another. For example, a basic geographic study involves how climate and landform interrelate with inhabitants, soil, and vegetation. Data collected from this study are geographically oriented and are therefore geographic data. Any study with a geographic component, regardless of form, produces geographic data.

By their very nature, geographic data comprise the physical locations of objects on or near the surface of the Earth. Data are intimately concerned with the properties of such objects and hold attributes that can be associated to other types of geographic data. For example, a user can have two types of geographic data about a 40-acre stretch of land in New Zealand: one detailing elevation above sea level, and one detailing the various types of soil composition throughout the parcel. Both forms of data can be combined for analysis in a GIS using the land parcel as the common link.

All physical relationships between layers of geographic data are interpreted as coincident relationships, meaning that features coincide in real world space. In the above example, the elevation data and soil composition data would go on separate "layers" in the GIS. Layers are discussed later in this book.

Many times, geographic data are modeled in what is known as vector space. A vector space is simply a platform for geographic vector data, which use *x-y* coordinates with lines and shapes to depict Earth features. Geographic vector data store nontopological coordinate geometry and attribute information for spatial features.

Most standard GIS vector file formats consist of a feature file, an index file, and a linked attribute table. A *feature file* contains geographic object feature information, such as representative point, line, and polygon information. An *index file* contains unique identifiers that comprise more detailed information and help speed spatial feature queries. A linked *attribute table* is a matrixed table that contains explicative attributes for a group of spatial features.

The index file links the feature file data to the attribute table. Consequently, attributes and features exist in a strict one-to-many relationship, whereby geographic features can have multiple attributes. The vector model is readily recognized by those who use computer-aided design (CAD) environments, using feature geometry in a GIS vector model to instantiate points, lines, and polygons as objects. Figure 1.2 details sample vector data.

In most cases geographic vector data are *discrete data* occurring in cases where there are well-defined boundaries for physical representations or limited data values. The above figure involves tax lots that have visibly evident and discrete boundaries. Vector data can be created or modified directly by digitizing features using GIS development environments.

There is another major type of geographic data known as raster data. Raster data are digital images represented by a grid of valued pixels, or *cells*. The image type and the number of colors represented determine the properties and

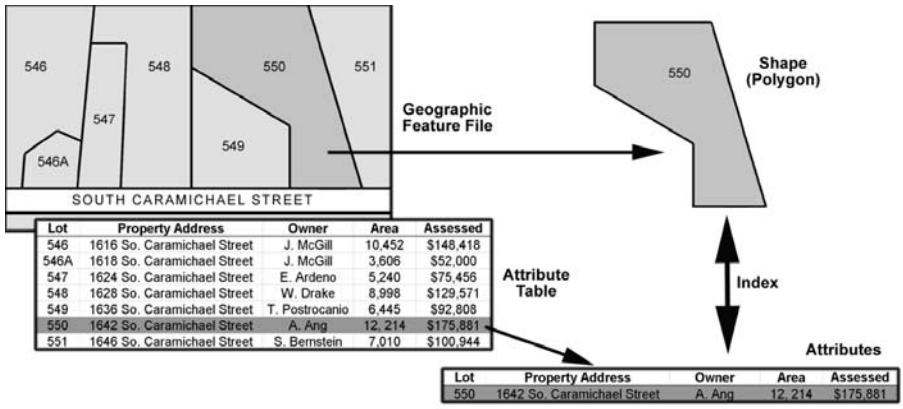


Figure 1.2 Vector data.

appearances of these pixels. Figure 1.3 details sample raster data that utilizes three different cell colors for three different property types.

In most cases, raster data are more suited for representing *continuous data* than vector data. Continuous data are a numeric form of data usually associated with the physical measurement of boundaries that are not well defined. Additionally, the surfaces represented are, in many cases, estimated through statistical

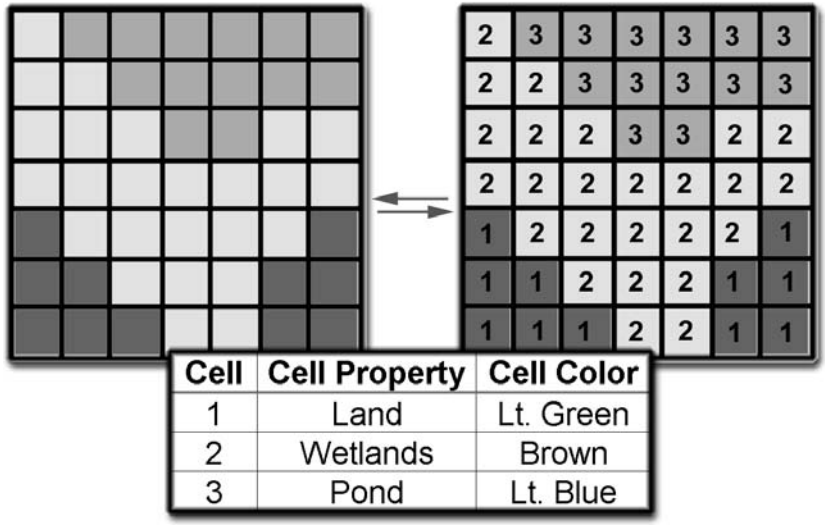


Figure 1.3 Raster data.

analysis. Raster imagery is perfectly suited for GIS sensor webs, a current industry trend that harnesses real-time observation and measurements over large, even global, regions.

Chapter 3 focuses specifically on the raster and vector data and offers a comprehensive discussion of each data model.

Georeferencing

Many data sources lack formal spatial referencing. Some CAD and GIS data sets are developed in a generic "design" space and have unique, often proprietary, types of referencing that simply need reinterpretation to be spatially integrated into a GIS environment. However, many of these sources are scanned raster data (digital imagery) that have only the coordinates of a raw pixel grid from an original scan. While these raster sources are often times unique and critical to a GIS project, images also need to be referenced from scratch, spatially transformed into a defined coordinate referencing system, then integrated and overlaid in a GIS environment. This process is known as *georeferencing*.

The ability to perform accurate and timely spatial referencing adds a measure of customization to any GIS operation or project. Raster imagery, such as hardcopy maps and aerial photography, is the most popular type of data to use with georeferencing, since it is the most commonly available type of data to use. Scanning imagery also alleviates the need to perform time-consuming and repetitive digitization efforts (i.e., transforming hardcopy to an electronic, digital file).

Georeferencing is the art of selecting common point locations in the real world using at least two data sources: an unreferenced source (such as a raster map) and a referenced source of the same area providing positional information. Basic georeferencing procedures involve point selection and transformation. For example, when a hardcopy map is scanned to an electronic file, it has no relationship to any real-world coordinate system. The georeferencing process establishes (or in some cases reestablishes) the relationship between image pixel locations and real-world locations. Georeferencing is accomplished by first selecting points on a source image (scanned raster map) with known coordinates for the real-world surface location (benchmarks, grid ticks, road intersections, and so on). These real-world coordinates are then linked to the corresponding pixel grid coordinates in the raster source image. After the image is georeferenced, each pixel has a real-world coordinate value assigned to it.

Queries: Locations and Attributes

Once data are adequately georeferenced and resident in a GIS, the user can then create a query expression to find the relevant data for a specific application. A

query enables the user to search geographic data to collect location, feature, and attribute information from a relational database management system or geodatabase. The most fundamental test of geographic coincidence between data sources is a query by location in tandem with a targeting expression, such as *Intersection, Within distance of, Contained by, Share features with, Touch the boundary of, Are crossed by the outline of,* and *Completely in.*

Queries can also be executed on feature attribute tables and performed on specific fields. Records are selected in this manner; each record (or "row") in the table corresponds and is linked to a distinct geographic feature data set. Features conforming to certain attribute criteria can be selected and extracted using this method.

Quite simply, queries are the user's refining tool for taking the massive available data and selecting only those pieces that pertain to the application at hand. The absolute function of queries is a GIS staple and, subsequently, will be highlighted both in example and in operation throughout this book.

Geoprocessing

Geoprocessing is the fundamental process of creating a derived set of geographic data from various existing data sets using operations such as feature overlay and data conversion. In a typical geoprocessing environment, the user applies GIS functions to a group of geographic (input) data to yield a precise output data set suitable for a particular application. Geoprocessing functions run the gamut from simple spatial clipping to more complex analytic operations. These software functions can stand alone or be chained to other processes. This ultimately opens the gates for virtually unlimited sets of geoprocessing models and potentially staggering sets of output data to solve specific problems.

Most professional GIS software environments include a mission-specific geoprocessing interface, or "workbench," of geospatial dialogs and tools. These software environments usually include extensible scripting tools and compilers to automate, customize, and document geoprocessing workflows. The most important contribution of geoprocessing to the GIS big picture is the automation of repetitive tasks. Geoprocessing is an elaborate turnkey for efficient and clean geographic output.

Geoprocessors come in different forms. Many geoprocessing functions are embedded in a GIS environment. A GIS environment is a package of integrated GIS components: a geographic map control, a map layout designer, a data tree catalog, and so forth, of which geoprocessing is a member. However, many powerful stand-alone software applications offer specific, related subsets of geoprocessing functionality. Some include file format translators or spatial referencing transformation tools. Many professional GIS efforts actually require

licensing these stand-alone accessories in order to reap the often-advantageous outputs of these stand-alone software applications.

Site selection is a prime example of a geoprocessing application. Another is the function of batch processing in a noninteractive manner. Geoprocessing environments can be considered "robots" that automate geographic data processes and provide storage of geographic data models. Geoprocessing is very reliable. In fact, within the GIS community geoprocessing environments are viewed as software-based insurance policies for the analyst trying to deliver geographic information products on time and within budget.

There are eight categories of predominant geoprocessing operations, or families of operations:

1. *Conversion.* Conversion is completely an issue of formatting. File format conversions (translations) and coordinate system referencing conversions are the most common geoprocessing conversion operations, and serve to characterize the conversion family.
2. *Overlay (union, intersecting).* Overlay involves superimposing two or more geographic data layers to discover relationships. In fact, overlay is intimately associated with the discipline of set topology, which defines the rules for valid spatial relationships between features in a geographic data layer.
3. *Intersect.* Geoprocessing computes a geometric intersection of the input features. The resultant features or portion of features common to all layers or assigned groups of same shape type (called a feature class) will be written to the output.
4. *Union.* Like intersect, union computes a geometric intersection of the input features. All features with the overlapping attributes from the input features will be written to the output feature class.
5. *Extraction (clip, query).* Like overlay, extraction is also intimately associated with the discipline of set topology. Queries help select the geographic data to be clipped or extracted, subject to a specific group of topology rules.
6. *Proximity (buffer).* Proximity is initiated through a query that selects geographic features based on their distance or proximity from other features. Geographic features include lines, points, and polygons.
7. *Management (copy, create).* GIS data management software is generally designed to facilitate the organization of a user's unique personal catalog (or collection) of geographic data. The intrinsic forms of all types of geographic data are accommodated by these applications.

8. *Transformation.* Typically in GIS, the term "transformation" means a spatial transformation, such as a datum transformation or reprojection. Geoprocessing, however, introduces transformations of different types, such as temporal or geometric transformations. This facet of flexibility finds favor among integrated user interfaces, complementing comprehensive spatial analysis within a geoprocessing software environment. Chapters 7 and 8 describe datum transformations in more detail.

In summary, the integration of geospatial data and geoprocessing interfaces into desktop computing makes possible the widespread use of interoperable geoprocessing software. Geoprocessing enables a clear pathway for geospatial data products in the information infrastructure. Geoprocessing opens the doorway to more defined data examination, primarily spatial analysis.

Spatial Analysis

By their very nature, geographic data are intimately related to locations and feature attributes. Spatial analysis harnesses this duplicity through the study of geographic feature locations and shapes. Spatial analysis offers the user a range of procedures, tools, and interfaces varying in application and complexity. For example, creating a simple map in a GIS environment is a basic form of spatial analysis. Spatial analysis is not necessarily complex, but it is a process of reducing complex relationships to something simpler, possibly bringing to attention things that otherwise would have remained hidden to the user.

Spatial analysis relies heavily on the first and most fundamental law of geography: Things closer together in space tend to be more alike than things that are far apart. This law is based upon Waldo Tobler's work, which essentially states that everything is related to everything else, but near things are more related than distant things. GIS does not wholly support this almost philosophical supposition, though it does acknowledge that near things are more related than distant things. In fact, geographic data contain spatial autocorrelation, which is the formal property that measures the degree to which near and distant things are related. GIS functions on spatial data to which it can be supposed that spatial autocorrelation is in essence the first law of spatial analysis.

Positive spatial autocorrelation occurs when features that are similar in location are also similar in attributes. Negative spatial autocorrelation occurs when features that are close together in space are dissimilar in attributes. Zero autocorrelation occurs when attributes are independent of location.

Spatial analysis, be it autocorrelation, overlay, or surface analysis, is an advanced and flexible form of geographic data analysis. A GIS offers the user a

platform for specific application analysis of location and feature attributes. The end result is highly usable and application-specific geographic information. Let us look at the final product: geographic information.

Geographic Information

As mentioned previously, the flow of information begins with a collection of geographic data and through various functions can smartly coalesce in the final geographic information product. This final information product is derived and interactive and offers the user a host of capability, organization, and material for analysis. Geographic information can facilitate the analysis of patterns within nature and human societies, can affect decisions at a low level of political influence, and can even return accurate and subtle answers to compelling questions. The geographic information output obtained from a GIS is virtually boundless, limited only by the adeptness of the user and availability of raw geographic data.

Actual geographic information products vary in form and appearance. They are characterized in three distinct groups: software environments, maps, and documents.

Software environments tend to be the most vital and characteristic of distributed geographic information systems. These environments are usually simple front-end interfaces presented in a dialog-based (or oftentimes wizard-based) package. They facilitate the automation of spatially enabled information management, usually alleviating the dependence upon repetitive tasks in the solution of a problem.

For instance, a shoreline-permitting GIS application can help municipal government administrative personnel make basic decisions concerning the approval of permits for new development on shore land. The administrative employee can enter the required basic information about a particular subject parcel (usually required information on a permit application), launch a dialog-based wizard to organize the input data, and then return an analysis of the site in relationship to natural, sensitive, or protected resources potentially impacted by the proposed construction.

A purer GIS product is much less restricted by process. Some outrageously inventive and complex GIS software environments offer geographic information products in a constant stream. This ultimately poses the convenience and challenge of interactive research for decision making powered by GIS functions. However, in the case of the environmental permitting application, complexity is not necessary and simple output from the system is sufficient. Environmental permitting at any level usually becomes such a contentious process that actual physical maps are required to provide graphic depiction and substantiation of research findings and assertions. The “map” is therefore the most obvious

example of a geographic information product, and satisfies the desire of the public at large to have everything presented in universally recognizable documents. *Maps* are the second facet of geographic information.

Maps represent the surface of the Earth—a surface with which all human beings are intimately familiar. Maps have a variegated array of uses, such as graphics in periodicals to supplement columns and articles, figures in permitting application documents, and large format prints mounted on the wall in the lobby of city hall showing planned municipal projects.

However maps, be they printed on paper or as an electronic image, are static information tools. They are noninteractive documents. GIS software environments are capable of organizing elaborate and diverse computations, occurring simultaneously with attribute or feature data. This is ideal for tabular figure information or in many instances complete report templates. This is the third facet of geographic information.

Tabular figures and other data *documents,* such as reports, charts, and graphs, are the third and final category of geographic information products. The document types help to disseminate computational, research, and spatial analysis results to anyone interested, be it the public or an engineering project manager.

These information products are usually developed and produced within a GIS environment through embedded controls native to the types of desired outputs. In some cases entire software programs for reporting, charting, Internet publishing, and database connectivity are "bundled" with a core GIS environment. This "bundled" convention encourages the user to stay right in a GIS software environment to complete tasks that are not commonly associated with GIS. Approaches like these help to expand the breadth of GIS as a modern technological discipline.

Even though geographic information is the end product of a GIS, it is not the ultimate finish line for geographic information flow. Geographic information moves beyond a GIS to the realm of limitless real-life applications. Traditional applications have been hydrologic plant relicensing efforts, infrastructure planning, shoreline zoning disputes, and wetland delineation. GISs have only just begun appearing in many nontraditional applications, such as defining statewide educational results, marketing efforts, public policy, and, yes, even presidential election result reporting. The only real stifling factor to the use of a GIS and the geographic information it offers is the user's ability to be creative, judicious, and purposeful.

2

Why Use a GIS?

In the previous chapter, we discussed to some degree how a GIS works, though we made no mention to why one would use such a system. For what reasons would someone want to progress geographic data through a transformation into geographic information? We begin this chapter with the intent to answer the question: Why use a GIS?

In 1854, Dr. John Snow did something revolutionary: He turned geographic data into geographic information and boldly stopped cholera. Let me explain. From 1831 through the mid-1850s, London, England, suffered from deadly cholera epidemics. Cholera is an acute infectious disease caused by a bacterium in the small intestines. Back in the early to mid-1800s, the disease traits of cholera were yet to be wholly identified, while outbreak sources were almost always unknown. An outbreak of the disease was a ruthless, communal disaster.

Then in 1854, a severe cholera outbreak caused horrific anguish to Londoners. In the area surrounding the intersection of Broad Street and Cambridge Street, more than 500 deaths were recorded in just 10 days. London's officials were baffled as to a solution and feared the absolute worst.

Dr. John Snow, a British physician, took matters into his own hands and mapped the cholera deaths on a hand-drawn street map of the London neighborhood. It soon became evident to Dr. Snow that a concentration of deaths centered around one particular water pump on Broad Street. He immediately presented his findings to the officials and had the pump handle removed. Immediately afterward, the spread of cholera ceased and the outbreak was finally contained.

Much like a modern-day GIS, Dr. Snow's geographic data, which included a street map and cholera death point locations, developed into a

unique and highly accurate geographic information map. Without the physician's information product, countless more Londoners would have died.

GIS enables developers to take a hard look at their project goals and offers capability for developers to maintain an underlying objective to improve the overall process. Like any pervasive process, a fundamental understanding as to why each goal must be achieved is imperative to the project's general success. This fundamental understanding affords the user new opportunities for use. In fact, many times after project goals are reached, additional objectives and project efficiencies arise out of the accessibility and plethora of GIS data. Dr. Snow reached his goals to preserve human life through technology, but his product and the process by which he reached his goal were reused for new outbreaks.

The following discussions highlight typical GIS products and outline several basic reasons why GIS is the forerunner in geographic intelligence. Even though the days of usable hand-drawn maps are long gone, creative information products, like Dr. Snow's creative methodology, set apart the typical GIS user from the atypical GIS thinker.

Discerning Your Need for GIS

Specific software is designed to satisfy a specific need. Tax preparation software (such as Quicken), for example, is designed to provide systematic tax form preparation, while Microsoft Word is designed to provide powerful word processing capabilities. Each software package has a definite, and separate, function, and ultimately achieves an explicit outcome. While the need for each software package is delineated and valid, the users, in the most basic sense, must first know what they want to accomplish before using the appropriate software package.

GIS is no different. A user must have an understanding of what overall outcome or expectation is desired before implementing and using geographic information systems. Discerning the particular need for a specific application early in the project is crucial to a GIS's overall success or failure. Many times a GIS fails because the project manager hastily implements a GIS for the sake of implementing it and pays little or no attention to what the system can actually do. Failure to initially define the expectations leads to a costly GIS implementation that can increase cost exponentially as the project phases progress. Consequently, a GIS's true capability and core benefits can be lost due to poor management planning.

Unlike GIS users, lawyers are trained to ask only those questions to which the answer is known. In the courtroom, this methodology proves quite efficient. However, in GIS, knowing the actual answer to a data request or the specifics of the geographic outcome is usually never possible. Therefore, GIS professionals and managers alike must generalize anticipated outcomes for the sole purpose of

management planning, such as anticipating a forest classification map of a 10-acre area of rainforest in Belize or a map detailing car theft in Madrid, Spain.

More times than not, the GIS's outcome provides greater detail and a wider range of possible analysis than first anticipated. A wholly environmental-related request can present additional detail for use with agriculture, conservation, and public safety analysis. Users need to define expectations early on, but be aware that the anticipated outcome may be a variant or enhancement of such expectations. As for lawyers and other professionals who need to know the outcome up front, take heed that GIS results can be enlightening and unexpected.

Common GIS Products and Bi-Products

GIS is one of those broad-ranging technologies that seem to have virtually no bounds other than the limit of the user's creativity. GIS, by abstract function, is a theoretical "black box" that takes data input and data requests and outputs the data results. If the user is vastly creative with both the input and requests, then the output will retain a certain degree of inventiveness. What gives GIS the aura of virtually no bounds is that the user can also implement a level of creativity with the data results, which, in the end, can produce an enormously innovative product.

Let us look briefly at the data process. As we know, standard geographic data resources are typically the only forms of basemap data available and used, such as digitized maps, census data, aerial photography, and so forth. These data are relatively static and the user is commonly bound by accessible data sources. Data requests, on the other hand, are formed by a particular application need and afford the user a suitable reason for specificity. This enables a level of user distinction with requests. An impetus for a data request may be the need for wetland delineation data in a square mile section of Brazilian shoreline or concentrations of emitted chlorofluorocarbons (CFCs) throughout Lithuania. Ultimately, geographic data requests allow creativity to enter the data process.

On the other end of the black box are the geographic data results that form the GIS product. We already discussed GIS information products in general and to some appreciable length. The GIS products are where the user's originality truly comes into play. Maps can be overlaid with multiple layers illustrating important and application-specific data.

For example, the New York City Department of Parks and Recreation used GIS to chart and manage the spread of the Asian longhorned beetle (ALB), a degenerative insect that bores holes in trees to lay eggs and whose larva bore additional holes to get out of the tree. ALBs cause the quick decimation of healthy tree populations. Through nontypical agency-created beetle data and Census Bureau data, the city was able to generate an ALB infestation geospatial

application that detailed infested tree density, quarantine zones, and infestation survey areas. These common, though project-specific, GIS products helped and continue to help the agency manage the ALB infestations. Overall, highly usable geographic data sets were created, which have ultimately helped contain the city's ALB infestations to just three incidents.

Another example of common GIS products used for unique scenarios involved a series of dreadful brushfires that ravished Sydney, Australia, in December 2001. Commissioned to discern the extent of the fires by the New South Wales state government, the Sydney Catchment Authority (SCA) utilized prefire and postfire imagery and developed the Satellite Pour L'Observation de la Terre (SPOT2) data set. SCA discriminated the normalized difference vegetation index (NDVI), a gauge of land cover suppleness, for both images and used GIS to calculate burn severity between the same areas in both images. Very little or no change in the NDVI value meant that the area was unaffected by the brushfires. A large change in the NDVI value indicated that the brushfires drastically burned the area. GIS also detailed a scale of varying degrees of forest burning, depicted on change analysis maps with a colored scale. Overall, the SCA's use of GIS not only helped detect the overall extent of the brushfires but also armed the SCA with a better grasp of the nature of similar brushfires.

Common GIS products such as the density maps and graphical visualizations are typical staples for user analysis but can remain distinct through the unique representation of data results. Depending upon the user's pioneering and innovative character, the common GIS information products are not the end of the data process. Many GIS information products possess what are called "bi-products," or developable information products formed as a result of a specific application data process.

Simply stated, geographic information bi-products are potentially exploitable proprietary sets or procedures that come out of a specific GIS process. The New York City Department of Parks and Recreation, for example, was left with a one-of-a-kind geographic data set of Asian longhorned beetle data (GIS bi-product) that may be usable by or salable to some other agency or private firm. Through the brushfire GIS application, the Sydney Catchment Authority developed two bi-products: a reusable procedure for the calculation of burn severity (which is now standard practice at the SCA for brushfires), and the forest burn data results (now being used for another GIS application—Sydney's spatial fire management system).

Bi-products spawn from creative approaches to specific applications. Some unique and profitable bi-products involve geographic data sets and demographics. The process of creating special data sets is tedious, time consuming, and often extremely difficult. In some cases, it can be impossible to get the raw data. Firms are many times willing to purchase prefabricated data sets from distributors in order to minimize project development expenditures and reduce

development time on the project schedule. This need has manifested in the creation and subsequent popularity of online geographic data marts/depots, where distributors compete for sales.

Specific demographics are usually harder to get and can be impossible at times without agency permissions and cooperation. Demographics add substance to a geographic data set and offer the user an enhanced capacity for data analysis. GIS developers know that qualified demographics, especially if you have proprietary access to such data, can be an enormous moneymaking bi-product from a GIS process. In fact, many firms now specialize in one-of-a-kind data sets and demographics, because of their high demand and financial value.

Database Management Systems

This section goes to great lengths to explain *how* database management systems (DBMSs) operate, but in the end the section clearly answers *why* DBMSs should be used and the functional role they play in a GIS. DBMS elements, such as queries and attributes, are covered in the previous chapter, since they are vital components to the overall flow of information but only a fractional part of a GIS's inner working.

A database management system is a program or collection of programs that enables the user to save, modify, classify, select, and extract information from a central database. DBMSs are used throughout the world and throughout almost every industry for the purposes of centralizing information and automating processes. For example, automatic teller machines (ATMs) function by means of a database management system. The user's card number and personal identification number (PIN) are entered into the ATM (database log-in security), information is displayed (database queries), user action is taken (information is sent to the central database, calculations are completed, and the database is appended with the new information), ATM dispenses money (database condition is met and approves machine's mechanical functions), and the user quits the ATM (database is closed, database log-in security activates again).

A typical DBMS is an organized, noninferential, and nonreferential data system that utilizes a file tree structure or file hierarchy. DBMSs are great for single purpose usage, where there are few users at one time and a modest amount of existing data at any one time. Without spatial coordinates, DBMSs can be described as information systems that are used for nongeographical databases (such as in the previous example of the ATM).

Ultimately, DBMSs are tremendous data management tools for small- to moderate-sized applications. These systems handle data rather easily and without much delay. However, problems arise in larger databases. Due to the

amount of data, large databases not only form object-oriented redundancies, they also slow down the query processes and become a substantial programming maintenance liability.

Back in the 1960s, these database problems were already evident. Surely you have heard the stories about the first mainframes and their physical bulk. Some of the largest corporations had buildings filled with room-sized computers that had limited processing capabilities and were getting "tied-up" with tedious database query functions. Databases were highly data dependent and, as a result, the functionality was impaired when large groups of records were accessed. The most prominent and baneful forms of data dependence included ordering dependence, indexing dependence, and access path dependence. The unanimous conclusion was the urgent need for less programming, simplified functions, and compressed data files.

In 1970, Dr. Edgar F. Codd offered a valid solution with his remarkable paper entitled, "A Relational Model of Data for Large Shared Data Banks." Dr. Codd's paper detailed a new method of simplifying database management using relational algebra for simple database functions. Codd's relational model broke new ground for eliminating redundancy and simplifying database functionality through user-defined relationships between data.

At first, Dr. Codd's paper went ignored until developers at International Business Machines (IBM) took notice. Based upon Dr. Codd's groundbreaking idea, IBM developed the first relational database called System/R. As the backbone to System/R, IBM developed and implemented a new form of relational programming language called Structured Query Language or SQL (pronounced "Sequel"). The space- and time-saving element of SQL was that it defined what data operations were used rather than how the operations were performed. As a result, the amount of programming code was drastically reduced through SQL, with many cases being reduced by more than half.

This leads us to the advent of the relational database management system (RDBMS), which is a program or collection of programs that enable the user to build data relationships to efficiently save, modify, classify, select, and extract information from a central relational database. RDBMSs are predominately SQL and Visual Basic-based with some of today's most widely used systems being Oracle, Microsoft SQL Server, Informix, DB2, and Sybase.

Relational databases are typically tabular data with a series of data columns and rows. The columns or *attributes* contain distinct field data, while the rows or *tuples* contain the individual records. The relational aspects of these databases are wholly contained within the attributes. In addition, each record has a unique attribute called a *primary key*, which distinguishes that record. These primary keys can be sequential record numbers or user-defined values that are unique and, thus, not duplicated anywhere within the same attribute throughout the records.

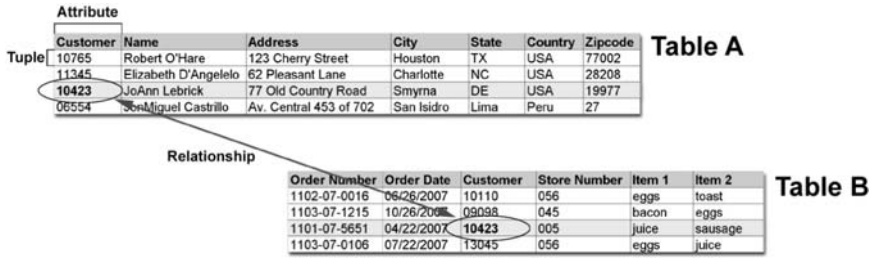
Consider Figure 2.1. Two distinct tables exist with various attributes and records. The relationship between both tables lies within the "Customer" attribute. This single relationship makes available to the user a wide range of query and data selection capability. Attributes "Customer" in Table A and "Order Number" in Table B are the primary keys.

RDBMSs can take on a variety of system architectural types, but the most fundamental and commonly used types include file-based systems, client/server systems, and multitier systems.

A *file-based system* is the simplest of relational database types and uses one file that contains the database tables, queries, security information, and forms/reports. Microsoft Access is a perfect example of a file-based system. The database file can reside on a single user computer or on a network. This type of architecture fails when more than a few dozen users access the database at one time.

A *client/server system* is designed to effectively handle many users accessing the central database at one time with speed and efficiency. The client/server system typically uses a network to connect the user workstations (called clients) to the central database residing on a file server. The system name evolves from this arrangement, hence "client/server." A client accesses the database using SQL and Visual Basic commands, and, in return, receives data from the database file server. The client, in turn, uses his workstation to process the data and perform the necessary calculations and report generation (i.e., information product). The amount of data transmitted from the file server is, in most cases, substantially less through the progressive use of SQL/Visual Basic query commands and user workstation processing.

The client/server system is a remarkable advance from the file-based database system but has one unmitigated fault: scalability. Data scalability involves the power to reduce or increase the amount of data being transmitted from the database to the user. Often, data scalability is warranted when information is presented on the Web or if data accessing rules must be employed.



Customer	Name	Address	City	State	Country	Zipcode
10765	Robert O'Hare	123 Cherry Street	Houston	TX	USA	77002
11345	Elizabeth D'Angelelo	62 Pleasant Lane	Charlotte	NC	USA	28208
10423	JoAnn Lebrick	77 Old Country Road	Smyrna	DE	USA	19977
06554	San Miguel Castrillo	Av. Central 453 of 702	San Isidro	Lima	Peru	27

Order Number	Order Date	Customer	Store Number	Item 1	Item 2
1102-07-0016	06/26/2007	10110	056	eggs	toast
1103-07-1215	10/26/2007	09098	045	bacon	eggs
1101-07-5651	04/22/2007	**10423**	005	juice	sausage
1103-07-0106	07/22/2007	13045	056	eggs	juice

Figure 2.1 A relationship between tabular data.

The solution to scalability is a *multitier system*, whereby there are one or more data task servers between the database server and the client. A multitier system (sometimes called a three-tier system) functions somewhat like the client/server system, except that there are middle servers, which contain software to perform certain tasks and process some of the transmitted data. These middle servers help reduce the amount of data being transmitted to the client.

The use of multitier systems results in easier application deployment, data scaling for the Web, enhanced security, and the ability to implement data or business rules. Since clients have reduced processing capabilities in a multitier system, they are called *thin clients*. Conversely, clients requiring their own individual rules, such as in the client/server system, are called *thick clients*.

Figure 2.2 depicts the three fundamental RDBMS architectural types and their differentiating physical arrangements.

So now that we have discussed the physical and functional infrastructure of relational database management systems, we can ask: *Where does GIS fit in?*

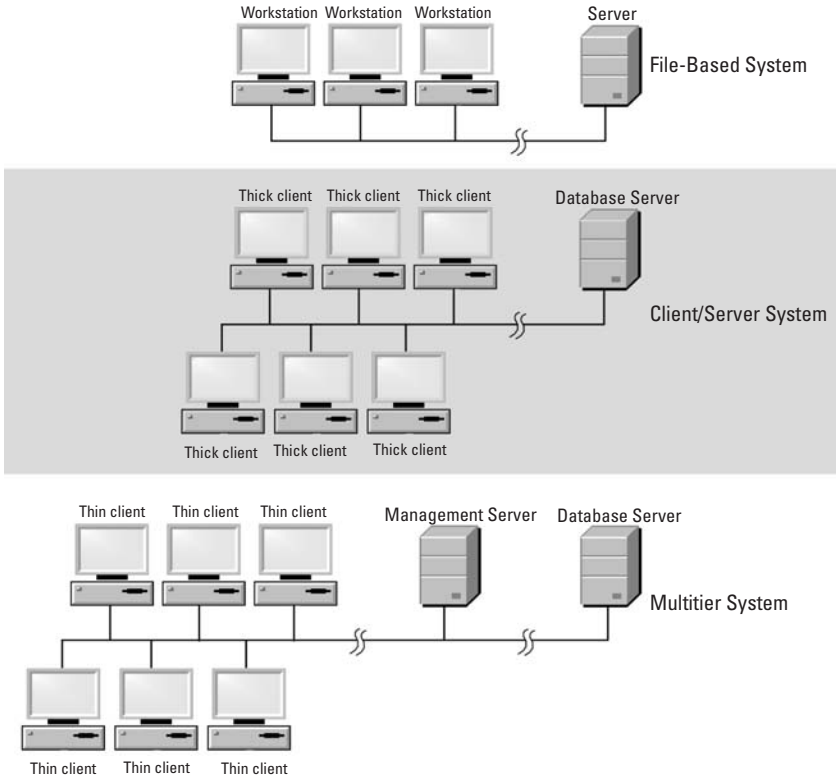


Figure 2.2 Three fundamental RDBMS types.

Geographic information systems follow the multitier system architectural type in that there are typically three tiers: a public presentation tier, a business and data management tier, and the GIS management tier.

The public presentation tier involves general viewing, Web-enabled data access, superficial analysis, and geographic visualization. This first tier is made up of general end users (or public thin clients). The business and data management tier involves users responsible for building the geographic database, site maintenance, critical analysis and visualization, and project management staff. This second tier acts as a filter to what information can and will be presented to the general public, as well as creates the project-specific geodatabase. The GIS development tier involves GIS professionals and systems responsible for data modeling and design, Web site development, low-level programming, and site administration. This third tier can also be called a data tier since it primarily functions to supply the entire database information.

Figure 2.3 depicts the three typical GIS tiers and their interaction with the RDBMS. Notice that the GIS development or data tier provides a relational and

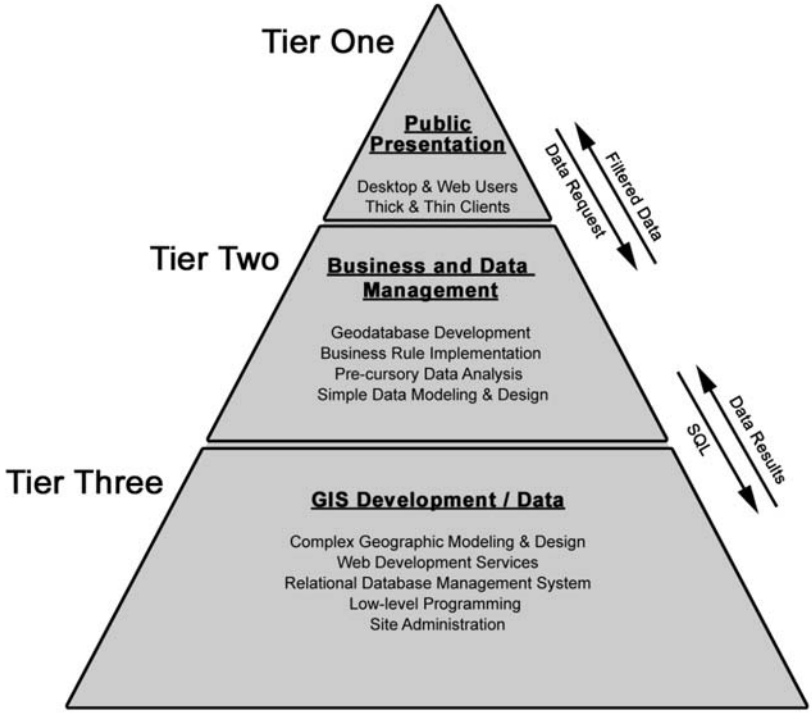


Figure 2.3 The three common GIS tiers.

comprehensive data resource. The business and data management tier takes the entire database and creates a project-specific geographic database, weeding out extraneous information and saving the pertinent. The thin clients then utilize both desktop and Web-enabled applications for viewing and analyzing this data. Overall, a massive relational database management system forms the backbone to GIS.

Let us look at a real-life example to help understand GIS's three-tier system. MapQuest.com is a free GIS-based map-generating product on the Internet, whereby a GIS backbone populates the Internet search capabilities. Although entirely different, GIS and the Internet integrate together to provide a Web-based front for geographic information. In brief, Web users have the ability to enter basic location information and view corresponding maps of that particular world location. Consider the user as the thin client in the public presentation tier (first tier). Users can view maps through minimal location input. Say, for instance, you are the thin client and you request a map of 415 Charles Street in Omaha, Nebraska. The software sends a query for data to the business and data management tier. This second tier receives the request from the user and processes the information. The second tier in turn sends a request to the RDBMS in the data (third) tier for Omaha, Nebraska, map data. The third tier contains every map for every city in the world, scalable to street level. Once it receives the request, all Omaha, Nebraska, map information is transmitted back. The second tier receives the RDBMS information and selects all the Charles Street information. This selected information is then searched by addresses until address 415 is found. The second tier transmits this select data to the first tier software, which, in turn, creates a graphic map. The thin client now has the correct map requested through the Web.

This leads us into the topic of how databases utilize the Web. In fact, there is a delicate bidirectional relationship between the RDBMS and the Web that patterns the very nature, flexibility, and importance of GIS Web services. All RDBMSs are designed to distribute information on the Web with a varying degree of user interaction. To achieve control, RDBMS utilize three types of data distributing Web pages: static, dynamic, and active. Because GIS is a relational geographic database management system, GIS products are portrayed and distributed in the same manner.

Static pages display database-driven information that cannot be changed, altered, or modified by the user. In fact, static pages display static Hypertext Markup Language (HTML) content. Dynamic pages, on the other hand, involve some user interaction, whereby the pages display the most recent or dynamic database information as defined by the user. Dynamic pages access a database through a query. MapQuest utilizes dynamic pages. The most common implementations of dynamic pages are with Java, a multiplatform, object-oriented programming language. Active pages involve full user interaction, whereby the user

can request information as well as make changes to and manage the database. Active pages can be considered a comprehensive database terminal with processing logic capabilities. In any case, HTML pages are the most common type of Web page utilized by GIS Internet map servers.

GIS Web services utilize all three Web page types when portraying geographic information. Noninteractive map products would typically be displayed using static pages, since user input is not required. GIS professionals use dynamic pages when the end user requires some functionality, such as turning on and off layers to a GIS product. The United States Environmental Protection Agency (http://www.epa.gov) utilizes dynamic pages so that users not only can search regions for reported environmental hazards but can customize the views through interactive layer selection. Lastly, GISs use active pages so that GIS developers have full control of the background data.

In conclusion, database management systems, both object oriented and relational, are important not only to how a GIS works but, more specifically, to why GIS is used rather than other methods. A geodatabase, for instance, is implemented directly on a RDBMS to take advantage of the system's data management capabilities and system administration tools. In turn, a GIS expands the functionality of a relational database so that it can proficiently handle geographic data, produce custom maps, and perform spatial analytic tasks. Together, GIS and a RDBMS offer the best in database management technology, geographic data distribution, information manipulation, and accelerated processing.

Quality Assurance, Six Sigma, and GIS

A fully implemented GIS adds value to a project while maximizing accuracy and data quality. In today's industries, quality is a big subject. You cannot enter any industry these days, let alone the engineering and environmental fields, whereby quality is not a major factor as to whether you will succeed or fail. In recent years, phrases such as "ISO-certified" (for International Organization for Standardization) and "ISO-9001 standards" are commonplace and, in many cases, required to provide services for certain agencies.

By its very structure, GIS elicits a high degree of data consistency and source data reliability. GIS information products, such as layered maps, contoured reliefs, and point data, offer various automated procedures and same-source data. GIS's process automation is just one placard for achievable higher quality output.

A rising new methodology for achieving quality that is quickly gaining corporate placement is called Six Sigma. Six Sigma stands for six standard deviations from the mean or average and involves a measure of quality that strives for near perfection. The lowercase Greek letter sigma is the mathematical representation

for standard deviation. In practice, Six Sigma is a disciplined, data-driven methodology for eliminating shortcomings and deficiencies in any business process. This perceptive approach can be implemented from the development stage to the service provision stage and involves staff trained in the Six Sigma principles. Large companies, such as GE and Motorola, already implement the Six Sigma approach and have experienced the benefits.

Unlike the ISO or ANSI (American National Standards Institute) standards, Six Sigma is not a regulation or a directive, per se. It is simply a unique way to achieve higher quality and save money. Incidentally, GIS can offer some of the same quality-enhancing results. GIS is complementary to the Six Sigma principles through GIS's data-driven structure, invariable consistency, and ability to improve processes. In addition, like Six Sigma, GIS's capability to reduce data variation and efficiently improve an organization's data quality is quickly becoming the new industry benchmark.

Developers, managers, and users are seeing first hand the far-reaching advantages to implementing an enterprise-wide GIS, from project efficiencies to fiscal brow beating. Its heightened quality enhancements complement existing corporate quality programs. Six Sigma and GIS can become the one-two punch for quality preeminence and knockout ISO/ANSI regulation compliance.

Project Visibility and Popularity

Through fluid Internet accessibility and high-octane map controls, GIS has become the preeminence in data sharing and distribution. Many times local governments and agencies find geographic information systems to be the one-size solution for their communication and public outreach programs. By default, GIS has become the technology of choice for detailing spatially based technical results to nontechnical viewers. The use of data representation and techniques makes analysis easy for those not directly involved with the project or experienced with the processes.

GIS is highly visible. Users can access data results easily through the Web and can manipulate (to some extent) how the data can be presented onscreen. For instance, after September 11, 2001, the U.S. Environmental Protection Agency (USEPA) let users view mapped environmental contamination and air quality data for Lower Manhattan. Users were able to turn on and off information layers over a street map baseline. The USEPA updated the air sample results daily and millions of users accessed the site. Through necessity and huge public interest, this basic GIS-based site experienced incredible visibility and popularity.

On a more fundamental and everyday level, project managers have embraced GIS use because they can efficiently manage the project's document and graphic-based results from a single data location. Through GIS, virtually

limitless project personnel and management alike can review the results in real time via the Web. The increase in project communication and quick project team debriefing has made GIS highly popular in project implementation and administration schemas.

As another example of GIS's popularity and visibility, a soil sample for a transportation rehabilitation project can be taken in the morning and its location, analysis type, and other relevant data can be easily entered into the GIS. Soil sample data locations and base information is instantly accessible by team members and others with GIS access. Soil sample information can be illustrated on a project site map with hyperlinks to soil boring logs and, eventually, the laboratory results. In the end, the benefits of GIS's one-stop shop of project data make GIS an extremely popular and highly exploited project tool.

Benefits, Cost Savings, and Automation

In many businesses where the bottom line drives decisions, GIS is often viewed as an expensive resource. Most decision-making managers have only the vaguest idea of what a fully enabled, successfully implemented GIS can do for their business proceedings. Although the concept of implementing a GIS is typically interesting, the front-heavy costs of implementing the system are not. These initial GIS costs are often enough to scare away most managers, while the proposed automated and cost-saving outcomes get classified as a project whim.

Today GIS is breaking new ground in project efficiencies, data communication, and value-added deliverables. Many private firms and public agencies are now requesting the implementation of a GIS on their projects because of the overwhelming long-term savings and advantages. Especially in local governments, GIS has made great strides to help consolidate the vast types of data and information into one categorical data repository. Local governments have begun learning the wide-ranging savings involved with distributed data throughout their projects. The dreaded fiscal black hole of government data communication is becoming a thing of the past.

One such local government is the Maryland–National Capital Park and Planning Commission (M-NCPPC). In a cost-benefit assessment report from 1999, M-NCPPC detailed the overall impact of the GIS implemented in Montgomery County, Maryland, highlighting the significant changes from the pre-GIS county environment and detailing the outstanding benefits of the GIS. Some of M-NCPPC's pre-GIS problems included inaccurate maps and data; incompatible maps and data sets; limited numbers of maps produced because of the required person-time; and end products not being standard. In total, M-NCPPC detailed nine crucial problem areas in the pre-GIS environment.

Since Montgomery County implemented GIS, a number of benefits were achieved. These benefits were grouped them into five areas: (1) overall improvements in existing operations from GIS, (2) additional capabilities not available in a non-GIS environment, (3) emergency response using GIS, (4) intangible improvements, and (5) product and bi-product sales revenues.

M-NCPPC's end result in its cost-benefit assessment report was that Montgomery County saved a typical annual amount of more than $500,000 from GIS functions alone. Additionally, GIS bi-product sales from fiscal year 1999 amounted to well over $44,000. M-NCPPC's overall conclusion was that the use of GIS had an overwhelming functional and cost benefit to Montgomery County.

Local governments have also experienced the benefit of increased government efficiency with the implementation of a GIS. For instance, the city of Ontario, California, used GIS to audit their billing files and reclaimed $190,000 per year in lost business license fees. In Bucharest, the Romanian National Railways Company implemented an ambitious GIS to automate their railway systems, which ultimately resulted in a geographic inventory of all railway assets, geographic maps for more than 22,000 km of track, and a centralized database to support all primary business unit applications. Additionally, a European Commission–backed project called Bacchus is using European Space Agency imagery and GIS to chart Europe's vineyards in unprecedented detail. The GIS information products provide vine growers with the tools to improve grape production, manage vines, and guarantee grape quality.

Just looking at the enormous benefits and heightened automation within the handful of examples presented, the intrinsic power and enormous impact that GIS has worldwide is easily understood. The key to maximizing all GIS benefits and cost savings revolves around knowing what the user wants from the geospatial application. The earlier the users can define their expectations, the cheaper the GIS implementation. Incidentally, user education is one sure-fire way to secure continued benefits from any geospatial application and helps the untrained user eye become more discriminating from the project start.

Possibly the greatest benefit of GIS is the ability to conduct tasks not previously possible. Whether it is Europe's vineyards or Maryland's local government, GIS has made an indelible mark on the procedures, efficiency, and future health of public and private entities worldwide. If you need additional proof as to the overall success of GIS and more reasons why GIS should be implemented, then peruse the voluminous case studies and success stories available on ESRI.com, Intergraph.com, and the hundreds of local government sites throughout the Web. You will soon recognize that the question *Why use GIS?* has but one accurate response: *Why not?*

3

The Structure of Geographic Data

Geographic data come from a variety of sources, such as digitized maps, aerial photography, GPS, and field data. Valid geographic data serve as a collective true marker as to what level of detail and accuracy a GIS can potentially attain. But, in reality, this is only a half-truth. The manipulation of this geographic data is as important and, often, more suggestive of the outcome's quality.

The appropriate structure of geographic data is considered a top reason why a GIS succeeds or fails. Inappropriate data handling or structure predictably leads to inappropriate geographic information products. To this extent, a user's grasp of the GIS application, the desired outcomes, and the geospatial system's limits are detrimental to realizing how best to treat the geographic data. Ultimately, the successful structuring of geographic data requires a combination of understanding and sound decision making.

Geographic information systems utilize two primary data models to manipulate and structure geographic data: the raster data model and the vector data model. We have already discussed both types briefly in Chapter 1 but now must build upon that initial introduction to gain a better understanding of both the elements and idiosyncratic nature of each type. By examining the nature of each data model you will have a much clearer picture of the important and institutional structure of geographic data, as well as a distinct understanding of the advantages and disadvantages of each.

Raster and Vector Data Structures

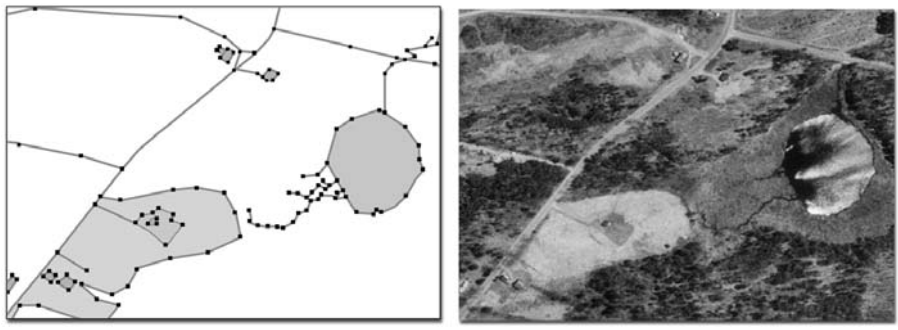
Like any complex data structure, raster and vector data have a myriad of different realizations that vary in complexity through use, appearance, format, and file size.

Although distinctly different, these affiliate data structures share two characteristics: (1) They visually represent real-world features, and (2) they are subject to orientation within the real world. By satisfying both of these characteristics, geographic data are born and made interoperable with other geographic data sources within GIS. Before proceeding further with this discussion, let 'us take a small step back to reacquaint ourselves with the raster and vector data structures.

Raster data structures characterize continuous data (such as imagery) and are exceptionally strong where boundaries and point information are not well defined. Raster data provide data as a pixel grid, whereby each pixel or cell is a feature capable of retaining properties and attributes. These pixels approximate pictures and images in an impressionistic way, with all of the smallish, monothematic cells contributing to a greater whole. Adding further identity, a raster image can vary in file format, color representation, resolution (size of pixels/number of pixels per set area), and potential properties.

Vector data are a bit different. Vector data structures characterize discrete data (such as roads, pipelines and topographic features) and are exceptionally strong where distinct boundaries and point information are well defined. This data structure is constructed on ordered two- and three-dimensional coordinates ([*x*,*y*] and [*x*,*y*,*z*], respectively). Features are represented as geometric shapes defined through single or grouped coordinates on a set grid.

For a clearer understanding of both models, refer to Figure 3.1, which illustrates the visible differences between raster and vector data, as well as how each data set represents Earth features. The original graphic is a raster, high resolution satellite image of a land parcel with infrastructure and a pond. Notice the distinct feature variance between its continuous data image and vector's point representation. It is clear that the raster image serves as a good overall representation, while the vector representation serves as a good individual feature representation.



Vector Representation

Raster Satellite Image

Figure 3.1 Raster versus vector data representation.

Both the raster and vector data structures have inherent advantages and disadvantages that allow GIS users a certain degree of choice. A full understanding of the native characteristics of each data model is a prerequisite for GIS success. Often certain requirements force the use of one data model rather than the other, such as the need for a better output resolution, easy image analysis, or enhanced spatial accuracy.

Raster data, for example, offer a truly simple data structure that involves a grid of row and column data. This simple grid structure allows for easy raster image analysis, as well as analysis among multiple images. Raster modeling is also much easier to implement due to the single-value cell structure and relatively simple software programming.

These advantageous raster capabilities are shadowed only by raster's native weaknesses. Disadvantages to raster data include general spatial inaccuracies and misrepresentations, low resolution, and massive data sets that require significant processing capability. The lack of accurate topology is also a major raster-based limitation.

Similarly, vector data offer their own variation of modeling strengths. Vector data, for instance, are spatially accurate and support a better, higher resolution than the raster data model. The ability to provide topology or feature relationships is a definite advantage, as well as the minimal data storage requirements.

Although seemingly ideal, vector data have their own share of weaknesses worthy of mention. Due to the complex data structure, vector data require a greater and more powerful processing capability. With this comes the need for better, faster workstations to minimize the data processing times that typical computers face. Inevitably, the costs to run a vector GIS and expeditiously process complex geographic data sets can become highly expensive.

Needless to say, both the raster and vector data models have their own brand of geographic data expertise and capability embedded within a GIS. Nowadays, GISs are offered as raster-based systems, vector-based systems, or raster-vector capable systems. When necessary, raster-to-vector/vector-to-raster conversions can be easily accomplished through GIS or third-party conversion software. Undoubtedly, the choice of data model lies with the user and the available software/hardware.

The next few sections examine raster and vector data capabilities that are vastly used in everyday GIS work.

Vector Feature Geometry

Often it is necessary to depict real objects as features on a map and designate object positions within a GIS. These features can range from the simplistic (i.e.,

linear transmission lines) to the complex (a multibranched, nonlinear river). To detail objects as features in a GIS, we must choose the best data model. In most cases, features are defined using the vector data model.

Given the nature of raster data, features are ambiguous in shape and general in position. For applications where specific form or position is not needed, raster may prove easiest. Some possible applications where raster data can be used to define features is on nonprecise maps, such as visitor maps of an amusement park, or macroscopic markers on an overview map, such as a colored box for approximate position.

Alternately, given the accurate, positional nature of vector data, features are best represented by coordinates and geometry. Real-world objects can be represented as individual or a group of geometric shapes called feature geometries. In any geospatial platform, there are three primary types of feature geometries: points, lines, and polygons. As a subset of these three primary types there exists a fourth geometric feature called a polyline. Figure 3.2 illustrates these major feature geometries in both the raster and vector data models. The differences between and usefulness of the models are obvious as they relate to depicting feature geometry.

A point is an individual position defined as a vector *x-y-z* coordinate or as one raster pixel. Lines are two connecting points with two distinct coordinates in vector or a linear block of pixels in the raster model. Polygons are a grouping of vector coordinates connected in a sequential fashion or a group of pixels

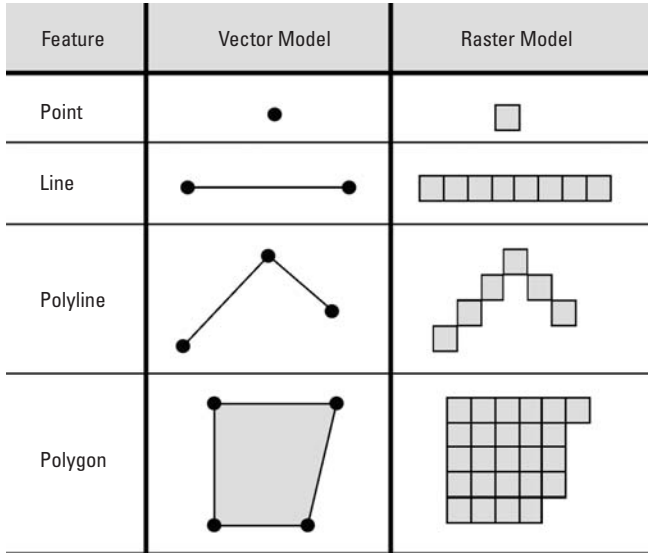


Figure 3.2 Feature geometries in raster and vector.

forming the object's general shape. Lastly, the subset polyline (known also as an arc) is a connected string of vector points or raster pixels. Often a polyline involves two lines sharing a same point or pixel (vertex).

Vector features are constructed on ordered pairs of vertices, whereby these ordered pairs reside in the design plane having known *x*, *y*, *z* locations. The positions of these ordered pairs are recorded in the vector file and often encoded as binary. These vertices are objects and can be grouped to form features that exist as independent entities. This is in contrast to the raster model, where the entire image is an object. In raster, features must first be sampled and then represented as image pixels, which, as we already know, visually approximate the features. Raster images have no independent vector ordered pairs, the actual building blocks of feature-driven spatial information.

Vector data for feature geometry comprise positional coordinates and, for some, other position-defining data, such as inside/outside and left/right. Figure 3.3 depicts the three primary types of geospatial feature geometries, as well as the polyline and the more complex polygon within a polygon (i.e., the "doughnut").

As illustrated in the figure, the point depicts a discreet position in space, defined by an individual geospatial coordinate. The line is defined by two geospatial coordinates. The open-ended third feature is a polyline, which comprises two vector lines sharing a common point (vertex). Interesting enough, polylines are defined not only by geospatial coordinates, but also by left and

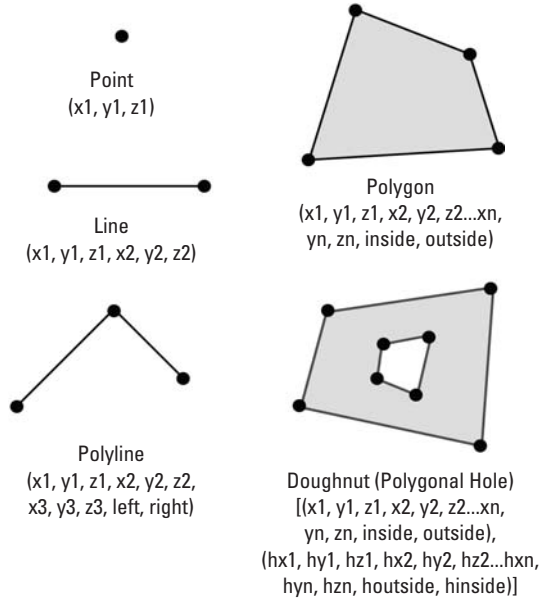


Figure 3.3 Major vector feature geometries.

right characteristics that indicate whether other features are directly left or right of the polyline.

Growing in complexity, the remaining two geospatial feature geometries in Figure 3.3 are closed features and depict specific areas. The polygon involves numerous vector points that are connected in sequence. Polygons are defined with inside and outside characteristics that delineate whether any overlapping geospatial feature geometries exist inside the polygon's boundaries.

The most complex of the geospatial feature geometries presented in Figure 3.3 is the doughnut or polygonal hole. A doughnut comprises a polygon within another polygon. Polygon 1 forms the boundary area and polygon 2 serves as the "cutout" within polygon 1. Think of a doughnut, whereby polygon 1 is the outermost doughnut boundary and polygon 2 is the cutout in the doughnut center (doughnut hole). Each polygon enables coordinate characteristics for inside and outside elements.

To uniformly structure, manage, and manipulate feature geometries like the ones just discussed, a GIS stores vector data in geospatial feature file formats. Common feature file formats, such as the *shapefile* developed by Environmental Systems Research Institute (ESRI) or the *TAB file* developed by MapInfo, are basic geometric containers compatible with an overwhelming majority of GISs. These geographic dataset types store nontopological vector (or coordinate) geometry in the form of real-world spatial features, as well as links to attribute information for these respective objects.

Raster Image Structures

The nature of images, such as aerial photographs and base maps, involves a continuous array of data. As discussed, the raster model provides the best solution for continuous data. Due to the lack of specified feature boundaries, point locations, and multiple objects, vector data are not intrinsically fit to handle the image information. A vector data representation of such an image produces a complex and often massive image structure.

Unlike a vector format, the raster palette is much more desirable for applications requiring the use of photographs or digital scans. Raster imagery instantiates outrageously complex geometries quite easily with the use of attributed image pixels. The essence and value of raster image structures are that raster represents real-world information with native visual properties—a feat vector imagery cannot reproduce.

Because raster images are physically continuous by nature, a reasonably accurate transformation process can be applied to fit raster data onto the real world. In brief, images are "rasterized" or digitally transformed to raster data through a matrix of pixels. Photographs are typically scanned with a set image

resolution defined by pixels per inch (ppi), more commonly known as dots per inch (dpi).

With the use of a GIS, the organic transformation into a raster image can enable users to characterize previously nonattributed existing geographic data material. In fact, solutions produced from this process result in alternative, often unexplored, geographic data sets.

There are times when an existing raster image is not at the desired resolution for the present application and an image adjustment is necessary. Manipulating resolution within raster images is a technique to adequately control and manage the amount of raster data to be processed and the overall image quality. As mentioned earlier, raster data resolution is the result of the amount of row and column pixels used to define an image. Raster image dataset size is directly proportional to the amount of raster pixels used to define an image. The rule of thumb is: The more pixels in the grid, the higher the image resolution, quality, and dataset size; the fewer pixels in the grid area, the lower the image resolution, quality, and dataset size.

Modifying the image's resolution presents varying results within the image. You can easily transform a high resolution raster into a low resolution raster without desecrating image quality. The image size often remains unchanged through an intelligent interpolation that transforms a grouping of pixels into one pixel. However, it is not always easy to reverse this transformation. Taking a low resolution into a high resolution form is not always feasible. To retain moderate image quality, a high resolution raster will take on a reduced, fractional size to that of the original low resolution raster. If image size remains unchanged, the new high resolution image will be relatively unusable and indistinguishable in content. Care should always be taken when modifying the resolution of raster images.

In actuality, geographic information obtained from these types of organic processes may be of critical relevance to future projects and unwittingly provide solutions to problems not yet known. Many times these unforeseen criticalities involve complications put forth by the presence of precious natural resources on a project. The manifestation (or realization) of these transformations is the fundamental building block for geospatial systems and makes available opportunity for other advanced GIS techniques.

Topology

It has already been established that in a vector-based GIS, primary geometries (i.e., points, lines, and polygons) represent real-world features. Topology is the set of rules through which a GIS represents features with the primary geometric shapes (i.e., point, line, and polygon). The vector data model utilizes topology

to organize spatial relationships between discrete features. In essence, the main functions of topology are to define: (1) feature-to-feature locality or, simply, where a feature is in relation to another feature, (2) what is shared between different features, and (3) how features are grouped or connected within a set.

In a GIS, topology establishes geometric harmony within a geographic data set. Illustrating this precept is ESRI's *shapefile*, in which the vector feature file employs a set of natural numbers plotted and structured in binary format. This set of natural numbers delineates the boundaries and coverage of a particular feature. The simple binary structure defines a topological space for the feature and establishes continuity with other features, forming topological relationships.

Topological relationships are defined in all types of feature files and are generally categorized into the three primary functions of topology (previously mentioned):

1. Feature-to-feature locality, called a *complement;*
2. What different features share, called an *intersection;*
3. How features are grouped, called a *union.*

In Figure 3.4, the simplistic Venn diagram depicting two overlapping, amorphous shapes highlights the base concepts of the three feature topology functions.

Within this figure, the lighter areas are considered *complementary* shapes (*A* and *B*), while the darkest area depicts the *intersection* of the two shapes (*C*).

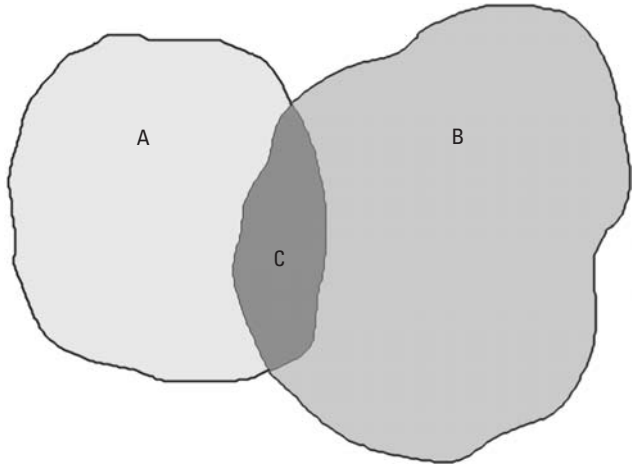


Figure 3.4 Venn diagram.

All three areas together (*A*, *B*, and *C*) form the *union* of the set. Applying these concepts to a vector GIS, shapes *A* and *B* can be considered two separate features that share topologic space (*C*).

Feature topology is also very sensitive to retaining the original shape of different features and the demarcation of each individual feature. In the context of traditional mathematics, topology is the examination of objects and groupings of objects that exhibit a predictable structure. Put simply, topology involves object aspects that remain unchanged even when the object itself is under some form of physical transformation (i.e., intersection). Similarly within the figure, shape *A* maintains its form even when sharing topological space with shape *B*.

All types of geographic data sets can utilize topology within geospatial software environments. These predictable structures (such as shapes *A* and *B*) constitute a set of features (*N*), which represents whole geospatial features. Topology implements a successor function: $s(x) = x + 1$, which succinctly indexes each separate feature for identification (FID). The successor function enables unique identification of each feature within the set. In this way topology serves as a geographic data quality overseer by ensuring the geometric integrity of geographic features between the real world and GIS, as well as retaining responsibility for producing truly clean and representative geographic data products.

Topology also helps to avoid repeating feature data, such as shared boundaries and shared nodes (points). The data model stores a single line to represent a boundary, as opposed to two lines with the same coordinates. This topological quality control helps maintain a smaller data set and vector feature file.

In Figure 3.5, polygon 1 (points *QRST*) and polygon 2 (points *RUVS*) are side by side, connected by a shared border (*RS*). Topology dictates that line *RS* for polygon 1 and line *RS* for polygon 2 are identical and the data model only accounts for *RS* once. Feature topology proves diligent in avoiding topological overlap.

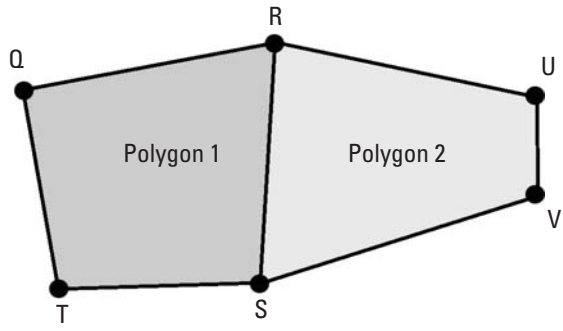


Figure 3.5 Topology of two polygons.

These geometric monitoring techniques allow GIS to control, query, and edit the topological coincidence between geospatial features (objects). Topology introduces the notion of absolute feature continuity, opening the door for numerous potential software compatibilities, including complex mathematics and engineering programs.

GIS Attribute Tables and Indices

In order for a GIS to successfully manipulate and portray geographic data, feature files must be indexed to attributes providing more information. As detailed in Chapter 1, most standard GIS data formats consist of a feature file, an index file, and a linked attribute table, whereby the feature file contains geographic object feature information, the attribute table (file) contains explicative attributes for spatial features, and the index file links attributes with features.

Briefly stated, index files act as attribute pointers. Indices contain unique identifiers that contain more detailed information about a specific feature. This embedded linkage helps speed up spatial feature queries within the GIS and serves as an index file's only true function.

Geospatial attribute tables, on the other hand, drive the spatially enabled database. There are various attribute field data types to handle the multitude of data, differentiated by a specific form of data and the degree of precision. To adequately handle the variants, attribute field data types signify different groupings of data bits, such as incorporating sign (i.e., "+" or "−"), binary, mantissa, and exponent bits, or signify data values, such as text, date, and object ID values. For clarity of this discussion, an *exponent* defines repeated multiplication (i.e., in $2^3 = 2 \times 2 \times 2$, the exponent is 3) and a *mantissa* is the value to the right of the decimal point in a common logarithm (i.e., in $\log 196 = 2.2923$, the mantissa is 0.2923).

The following attribute field data types are most common and are supported in many major GIS environments:

- *Short Integer.* A basic attribute data type that includes one signed bit and 15 binary bits.
- *Long Integer.* A more complex form of the basic attribute type that incorporates one signed bit and 31 binary bits. As you can imagine, the *Long Integer* offers greater precision than the *Short Integer.*
- *Float.* Contains one signed bit, seven exponent bits, and 24 mantissa bits.
- *Double.* A more complex form of the *Float* attribute type with one sign bit, seven exponent bits, and 56 mantissa bits. As with the *Long Integer,*

the *Double* attribute type holds greater precision than the *Float* attribute type.

- *Text.* Contains varying forms of data, such as numbers, letters, and symbols. The *Text* attribute type is a character string that can hold any amount of characters, but each character is stored using eight bits (called a byte). An interesting aspect to *Text* attribute data are that each text value in the same field must have the same number of characters. To achieve this, end blanks are used to fill in the empty slots.
- *Date.* Though not apparent from the attribute data type name, a *Date* type contains date and time data. The value is based upon a standard time format and is automatically transformed into the current day and time within the system's local time zone.
- *BLOB.* Short for *Binary Large Object*. A *BLOB* is a complex (and large) object stored in the database that may include an image, sound, video, or geometry. *BLOB*s allow users the ability to insert any type of multimedia data into the geodatabase.
- *GUID.* Acronym for *Globally Unique Identifier*. A *GUID* is a unique 128-bit (16 byte) number that is produced to identify a particular application, file, database entry, hardware, or user. Each generated *GUID* is "mathematically guaranteed" to be unique since the total number of unique keys is colossal and the probability of generating an identical *GUID* is virtually impossible.

As with any abstract concept, casual or novice users often misunderstand attributes in geospatial systems. With a better understanding of GIS's theory (this book) and real-world practice, these feature-attribute concepts will congeal.

4

Geospatial Metadata

Today's society is information rich. We can walk into any major bookstore and use a computer to locate individual or related topic books, obtain in-stock status, and get exact store location. Let us not forget we will probably be provided a host of other miscellaneous information about the book, such as its ISBN, the number of pages, and reviews. In a matter of a few minutes, we could identify whether or not our visit is fruitless. In that short time span, we not only have discovered whether our desired information is available but also have saved a great deal of time.

This act of information sharing and dissemination is, by all means, not by accident or some form of electronic wizardry. It is a well-organized data retrieval system that retains information in a real-time database. The search function utilizes the searcher's entered information and locates specific record fields or keywords with any associated matches.

Similar to the bookstore's search utility, data resources, such as data sets, images, and documents, are digitally marked with descriptive tags known as metadata and are organized in a specific, universal order. Searches can be made against the metadata to locate these resources quickly. Without metadata description or encoding, each resource must be searched in its entirety for a match. You can just imagine how long it would take to search millions of individual resources in their entirety.

This chapter attempts to take the wizardry and mystery out of metadata. It offers an overview of dataset metadata and demonstrates how a defined metadata structure can serve as a key time-saving protocol for GIS users. It details some of the efforts taken to organize and expand available resource data, while presenting the geospatial paths taken by worldwide governments. Due to

the nature of this book and the wide breadth of metadata, this chapter is limited to an overview. Specific field metadata and metadata authoring (see Table 4.1 later in this chapter) are subjects for further exploration.

What Is Metadata?

Metadata is defined as structured information that enables a resource to be easily identified, used, manipulated, and cataloged. Metadata is often identified as "data about data" and can be applied to various resources, such as images, online documents, maps, library records, data sets, or anything else inherently searchable. In simpler terms, metadata helps resources be suitably found, cataloged, and used.

The concept of metadata is nothing new. In fact, library card catalogs, which have been around since long before GIS, are a form of metadata. The older hardcopy information cards and newer online records hold a selection of basic information and identifiers for the material being sought. Typically, information supplied includes a title, author, date, ISBN or similar identifier, and a unique library location number. This library catalog card/record is identical to metadata since it is data (card information) about data (book or resource material). Figure 4.1 identifies some common manipulators of metadata.

Metadata sometimes is embedded within the resource and other times in a stand-alone document. For instance, Web pages have metadata embedded into their source code, often programmed in HTML, and image files have embedded

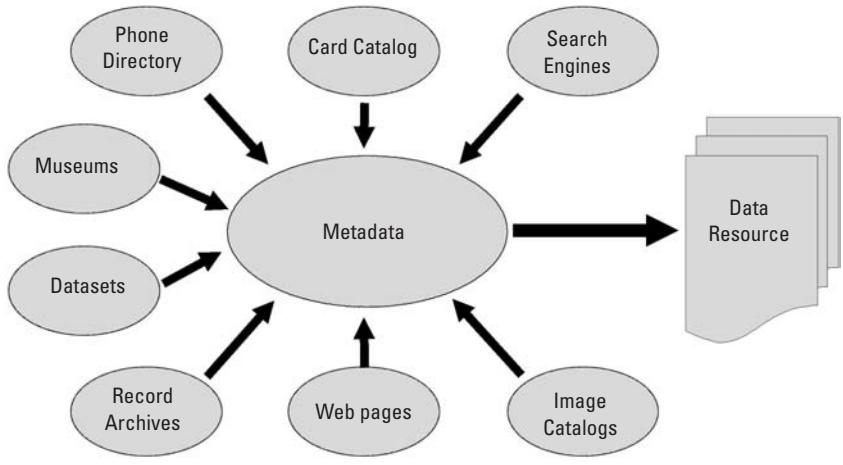


Figure 4.1 Various forms of metadata implementation and use.

metadata in the file headers. An archived map, on the other hand, may have a separate resource document containing the metadata for the map.

Embedded and separate metadata each have certain advantages. Embedding metadata within the resource ensures that the metadata will not be misplaced or lost and alleviates any problems with data-metadata linkage. Sometimes the capability to embed is not possible and, as a result, metadata has to be separate from the resource. This data separation offers easier resource location and retrieval, as well as offering a facile way to manage the metadata. Separate metadata are usually stored in a database and linked to the affiliate resource.

As you may have guessed, the topic of metadata is vast. In fact, it offers so many variations and covers such an array of fields that metadata can form (and has formed) the subject of its own book. Our interest in metadata goes only so far as the basics and those traits relevant to GIS.

Before jumping into the GIS-relevant side of metadata, it is important to understand the various objectives for the creation of metadata. The most primary objective for metadata is resource discovery. Metadata makes the resource "visible" to searchers, making it much easier for users to discover available resources. Other metadata goals include resource identification, enhanced structure, administrative controls, intellectual rights management, and preservation. These metadata objectives are equally important throughout the assorted fields of use.

Geospatial Metadata and GIS

We have already discussed how metadata are uniquely linked internally or externally to various types of data resources. In GIS, the data resource of paramount interest is the data set. Metadata that identifies and describes a data set, be it in GIS or another database management system, is classified as dataset metadata. Furthermore, since a GIS is a geospatial database utilizing a geospatial data set, its metadata are likewise geospatial metadata.

Users of geographic information systems rely heavily on the use of geospatial metadata for identifying and manipulating geographic data sets. Classified as dataset metadata, geospatial metadata are defined as structured information about a specific data set that enables the geospatial resource to be easily identified, used, and cataloged. Geospatial metadata includes at the very least the geographic data's owner, source, description, resolution, scale, and proper use.

Metadata helps GIS users understand the numerous parameters surrounding the data sets. Users can quickly discern the data set's level of precision and usefulness in tandem with other data sets and objects. Well-detailed geospatial metadata directs users on how best to use the resource and to what lengths the

dataset creator went to construct the resource. Through metadata review, an astute GIS user could identify which dataset resources are of high quality, which are beneficial for their application, and which are pure garbage.

Geospatial metadata is used for the wide array of areas that GIS handles but has proved to be particularly valuable in the technical and scientific arenas. New, refined definitions of geospatial metadata are concurrently being constructed to better define and catalog the available technical resources. New pathways for information exchange and sharing have been paved through past metadata initiatives and continue to expand through ongoing initiatives within many progressive disciplines, such as in the biological, natural resource, and social science fields.

Every initiative to define and beneficially structure discipline-specific metadata is based upon an already established metadata format called a *schema* (or scheme). Refer to Figure 4.2 for an overview of a schema. Traditionally, *schema* is used for diagrammatic description and *scheme* is used for textual description. For the sake of simplicity and the nature of GIS data sets and products, we will use the term *schema*. Moving forward, the next section offers insight into metadata schemata and their element sets, as well as offering the particulars of the ones more universally employed.

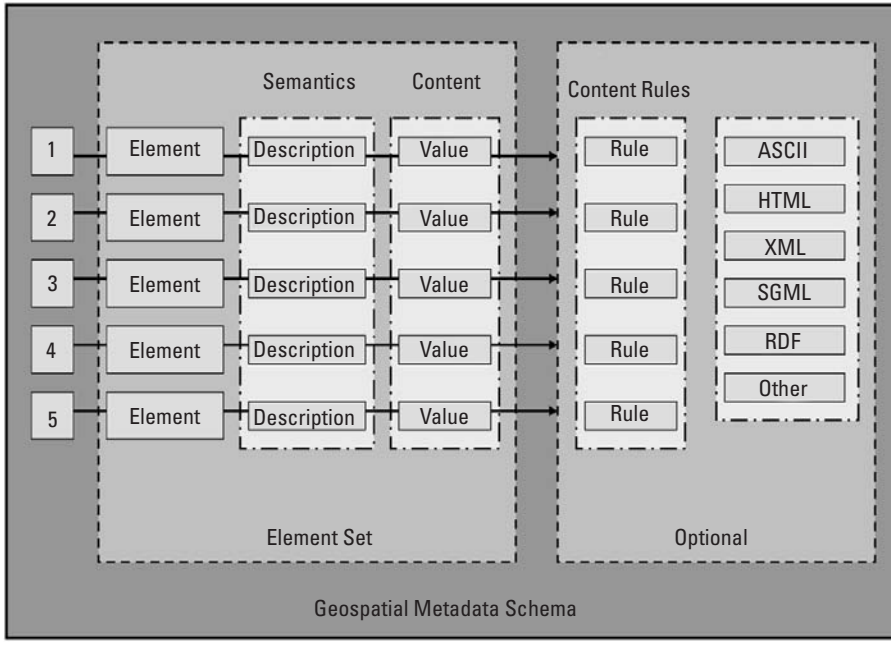


Figure 4.2 Geospatial metadata schema.

Metadata Schemata, Element Sets, and Syntax

A metadata schema is a description of the particular structure, arrangement, or content applied to a set of metadata elements. Alternately, the metadata element set is the collection of individual constituents that form the actual metadata. You can envision the metadata's schema as being the "database," the element set as being the database's primary table, and each element as being an individual data field within that primary table.

Metadata schema elements hold characteristics similar to any spoken or programmed language, whereby they must follow stringent rules for correct usage. The description of a schema element is considered the semantics of that schema and the values given to the schema elements are considered the content. This schema content may be further constrained through schema content rules. For instance, it may be required for a date-related schema element to be entered in an MM-DD-YYYY format. These elemental dictates are part of the overall structure of the schema and help maintain consistency throughout the developed resources.

Imposing further vigilance in metadata consistency, certain schemata must abide by well-defined content-encoding syntax rules. These rules offer a consistent metadata structure and further enhance the resource's discovery. As indicated, not all schemata are required to follow defined syntax rules and, as such, are called syntax independent.

To implement syntax, modern metadata schemata use various frameworks, such as ASCII, HTML, Standard Generalized Markup Language (SGML), Extensible Markup Language (XML), and Resource Description Framework (RDF). Figure 4.3 offers a few samples of metadata syntactic frameworks.

ASCII is the most basic of frameworks, simply because ASCII is nonsyntactic. ASCII is an original computer-based exchange format that is text only and is not capable of any embedded information. Therefore, however the metadata schema structures the data is how the data is displayed, without any content rules. ASCII proves to be minimally useful for resource discovery.

A better framework for geospatial metadata is one of the markup languages. HTML is a base framework that uses embedded element tags to describe the document. Although much more detailed than ASCII, HTML is rather simple and focuses solely on a document's style rather than its structure. XML, on the other hand, is an extension of HTML that uses tags to define and embed structure within the document. XML offers better versatility and enhances resource discovery and data interchange. In recent years, XML has become one of the syntactic frameworks of choice for metadata.

SGML departs from the role of essential resource description and structure to something more elaborate and comprehensive. SGML is an enhanced union of both HTML and XML and, as such, is often called an HTML-XML superset.

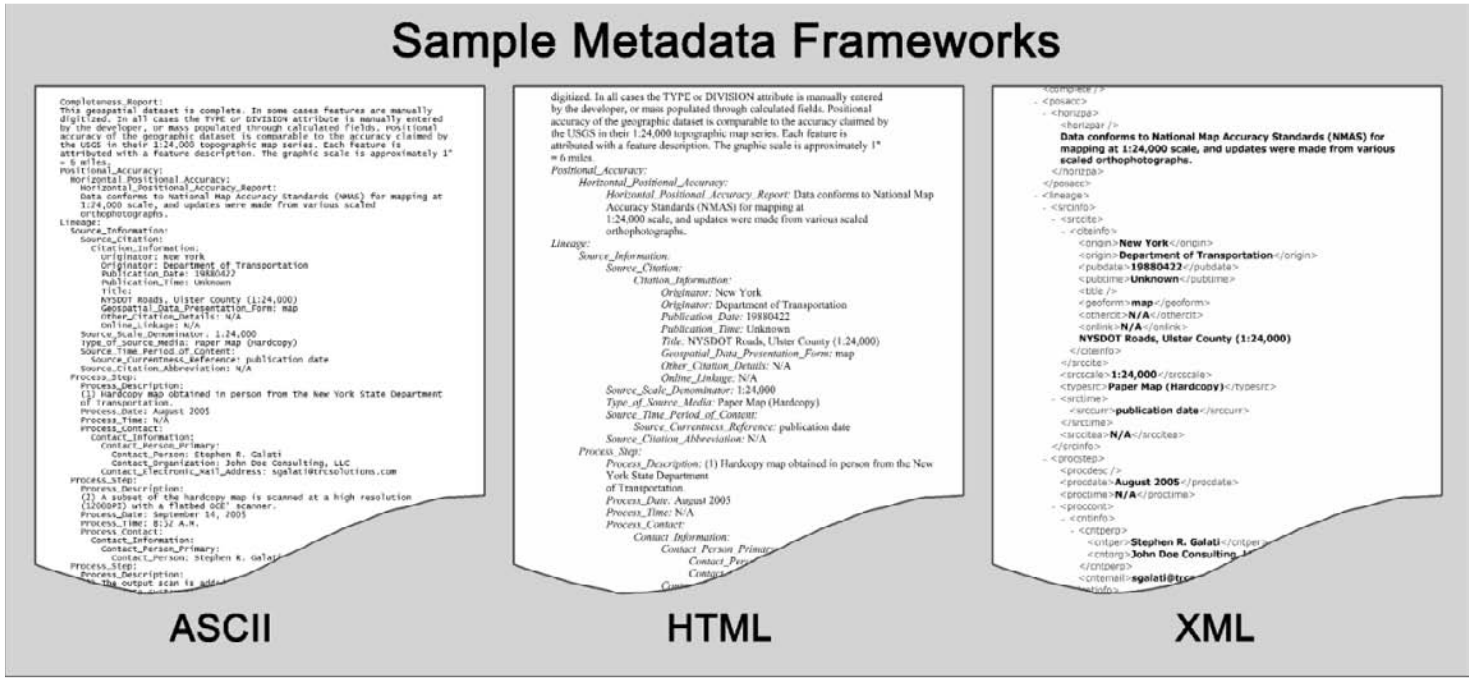


Figure 4.3 Sample metadata syntactic frameworks.

SGML looks much like an HTML document but is a structurally rich version of XML. With more structure descriptors, SGML offers copious description, both in style and structure, within the document. Additionally, the SGML framework offers the best availability for resource exposure and discovery among the discussed syntactic frameworks. However, with so much detail required, SGML is sometimes found too laborious. Figure 4.4 details the unique relationship between the markup language frameworks.

The final syntactic framework we will discuss is Resource Description Framework. RDF takes a different approach toward metadata architecture. As its name implies, RDF is a framework for describing a resource and sharing geospatial metadata. RDF is different from the markup language triad (HTML-XML-SGML) and focuses on the essential information that is needed to simplify metadata data searches, encourage interchangeability between syntaxes, and identify key resource properties.

In terms of depth of description, XML and RDF are the most similar but lack the comprehensive nature of SGML. You may be asking yourself, "Which one is better?" To be honest, it depends on the desired product, as well as the schema requirements, the desired amount of exposure, and the resource creator's effort to describe the resource.

For instance, some schemata (such as the Dublin Core discussed later) prefer RDF to one of the more traditional markup languages due to the ratio between metadata structure simplicity and essential content. Figure 4.5 depicts a sample RDF structure. Whatever syntactic framework and/or standard schema are used, the more description supplied will only benefit resource discovery, interoperability, and use.

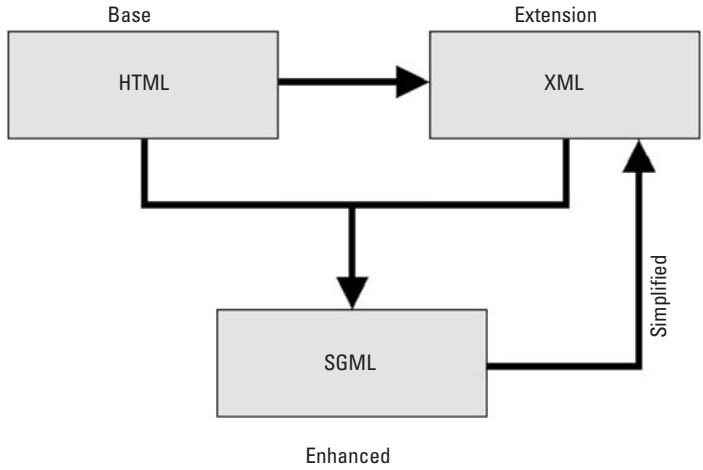


Figure 4.4 The relationship between HTML, XML, and SGML.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF SYSTEM "http://dublincore.org/documents/2002/07/31/dcmes-
xml/dcmes-xml-dtd.dtd">

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="\\metadata\standards\samples\rdfsample.html">

    <dc:title>
      GIS Demystified RDF Metadata Sample
    </dc:title>

    <dc:creator>
      Galati, Stephen R.
    </dc:creator>

    <dc:subject>
      Metadata.
    </dc:subject>

    <dc:description>
      Standards sample provides an example of Dublin Core standards
      as published in a Resource Description Framework form.
    </dc:description>

    <dc:publisher>
      Artech House
    </dc:publisher>

    <dc:type>
      Text
    </dc:type>

    <dc:format>
      text/html
    </dc:format>

    <dc:format>
      18425 bytes
    </dc:format>

  </rdf:Description>
</rdf:RDF>
```

Resource Description Framework (RDF)

Figure 4.5 A sample Resource Description Framework structure.

Geospatial Metadata Standards

Throughout the past years, metadata has become a hot topic for discussion. Many worldwide organizations feel that geospatial data should be standardized for swift interoperability and exchange, while others argue that many field- and location-specific structures are needed to keep metadata uniquely keyed toward the individual fields they represent. This also helps capture field-unique

elements that a general standard would omit. As a result, there are numerous initiatives currently defining and instituting field-specific tangents (called profiles) off already established schemata.

Established and officially documented schemata form the current standards for today's geospatial metadata. Standards, as a whole, are valuable within the GIS industry for a number of good reasons. First, standards provide universal terms within the various resources they describe. This universal terminology offers beneficial consistency within the thousands (even millions) of available data resources within which the terminology is enforced.

Second, universal terminology presents the capability of automatic searches for specific terms. Users can program workstations to search for any universal term, such as *author, publisher,* and *source,* and save enormous amounts of discovery time. Automated searches can be done for nonstandard metadata, but these will take much longer and will habitually provide less than satisfactory results.

Third, similar to the second reason, users can quickly search numerous data sets and geospatial objects for specific terms. Having various forms of resources tagged with universal terms once again helps link the resources and produce comprehensive search results. Thus, finding specific information is usually much easier and time efficient within standardized metadata than within a nonstandard format.

Fourth, standards enable simpler information interchange. Since terms are set for individual metadata elements, the crossover between two standards is uniform throughout all associated resources. For instance, Dublin Core metadata can easily be transformed into European CEN/TC 287 metadata, since both are established element sets.

Fifth and finally, certain standards are mandated by governments and are required by local agencies. The U.S. government requires all agencies to use the Federal Geographic Data Committee content standard, while the Australian Government requires the Australian Government Locator Service standard.

Geospatial metadata standards offer the GIS user an efficient way to organize and catalog available images and data sets. As you may already grasp, there are numerous standards available to the worldwide public, with some schemata better for geospatial metadata than others. Some available standards for geospatial metadata include:

- Content Standard for Digital Geospatial Metadata (CSDGM);
- Dublin Core Metadata Element Set;
- European CEN Metadata Standard (CEN/TC 287);
- Australian Government Locator Service (AGLS) Metadata Element Set;
- UK GEMINI Discovery Metadata Standard.

To get a better idea of metadata structure and how each schema differs, we take a brief look at these metadata keystones.

Content Standards for Digital Geospatial Metadata

Under an executive order from the U.S. government, the Federal Geographic Data Committee (FGDC), a 19-member interagency committee, was charged with the task of developing national geospatial metadata standards. As a direct result, the FGDC developed the Content Standards for Digital Geospatial Metadata, or CSDGM. Since its origin in the mid-1990s, the CSDGM has developed into the most widely used metadata standard to date. CSDGM has even become the basis for numerous standards worldwide.

The intent for CSDGM was to help the U.S. government minimize the costs of data acquisition efforts and to promote interagency data sharing. The CSDGM was developed for consistent geospatial dataset description and utilizes a group of seven core (primary) and three floating elements. The core elements make up the major element set, whereas the floating elements are nonessential and wholly optional. Figure 4.6 details the seven CSDGM metadata core elements.

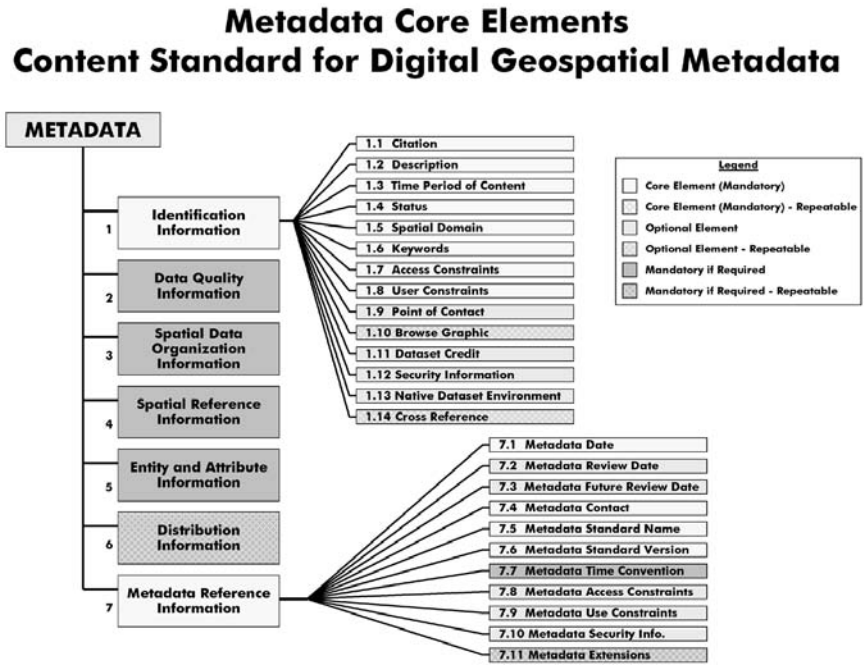


Figure 4.6 CSDGM metadata core elements.

The core elements are separated as *Mandatory, Optional,* and *Mandatory if Required.* The *Mandatory* elements are required in order to be minimally compliant with FGDC's standard. The *Optional* elements can be added or omitted from the metadata without affecting metadata compliance. The *Mandatory if Required* elements are metadata elements that may or may not be applicable, depending upon the resource being described. For instance, CSDGM metadata for a Web page may warrant only the minimal core elements, while a Geocoded map may also require one or more *Mandatory if Required* elements, such as data quality information. Although Figure 4.6 breaks down only the two *Mandatory* elements into their subcomponents, each element has a similar breadth of subsets.

The *Mandatory if Required* status of CSDGM elements adds a level of user flexibility and decision making. As with any form of metadata, the more description included in metadata, the greater the discovery potential. Hence, if the resource information is available to the metadata creator, it is in his or her best interest to include that additional description.

Certain elements and element subcomponents are repeatable, which means that they can appear in metadata more than once. Say, for instance, there was a set distribution for a particular metadata (e.g., 14 college professors). You have the capability within the CSDGM structure to include information about each of the 14 professors within the *Distribution Information* element. As depicted in Figure 4.6, this core element is repeatable and, thus, would be compliant with FGDC's standards.

As a major, widely available metadata element set, CSDGM sets the bar for geospatial metadata consistency and standards and continues to encourage international data exchange. As the leader in the geospatial metadata arena, its structure has been emulated for countless standardized element sets around the world. Adding credence to CSDGM is the fact that it still remains the national standard for the United States.

Dublin Core Metadata Element Set

The Dublin Core Metadata Element Set began as an initiative to advance the discovery of information resources. The initiative, called the Dublin Core Metadata Initiative, started in 1995 with a workshop in Dublin, Ohio, and involved various types of interested professionals, such as researchers, librarians, and content providers. The result of this initiative was the creation of a Dublin Core standards prototype, which involved a rather small group of information descriptors. The Dublin Core standards quickly gained worldwide attention and interest from a wide variety of metadata users.

Today, the Dublin Core Initiative's ongoing progress has produced the Dublin Core Metadata Element Set, a 15-element set that has attained

worldwide acceptance and use within a wide range of content areas. The Dublin Core sets the standard for multidomain resource description and data resource visibility.

Figure 4.7 represent the 15 primary elements that make up Dublin Core metadata. Unlike the CSDGM element set, all Dublin Core elements are optional and repeatable and may be used in any order. Therefore, the primary elements depicted in Figure 4.7 are in no particular sequence and can be changed, duplicated, or omitted from metadata. This innate flexibility has made the Dublin Core standards highly attractive to users worldwide, as well as an excellent schema for geospatial metadata.

Going way beyond the Dublin Core Initiative's original expectations, the Dublin Core Metadata Element Set is a means for individual preference and metadata richness. The element set offers the use of qualifiers, which are bits of additional information that are useful in further exposing the resource for discovery and promoting enhanced information interchange. Qualifiers include such additional information as various encoding syntaxes (i.e., HTML, XML, and RDF), greater content descriptors, and value type description. In fact, Dublin Core users can choose between Simple Dublin Core metadata and Qualified Dublin Core metadata.

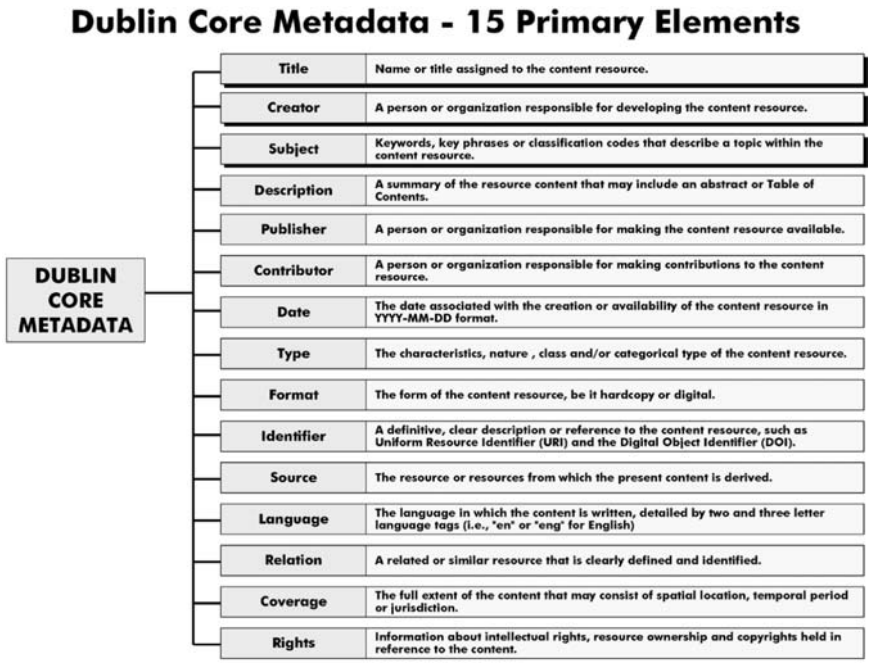


Figure 4.7 Dublin Core metadata—the 15 primary elements.

Simple Dublin Core metadata, also known as Unqualified Dublin Core metadata, involves the use of only the 15 primary elements with simple content and/or values. Qualified Dublin Core metadata goes beyond the simple content within the 15 primary elements and enables the metadata writer to use qualifiers to add a greater level of resource detail. Both forms of Dublin Core standards are efficient, but in some instances Qualified Dublin Core can reduce the interoperability of the metadata and add an unwanted layer of complexity.

As mentioned earlier, the Dublin Core Metadata Element Set also allows for various encoding syntax, which helps interchange Dublin Core descriptors into other metadata schemata. When Dublin Core standards first came out, they were predominately encoded in HTML. Nowadays, the prevailing encoding is XML and RDF, whereby XML is typically used for Simple Dublin Core metadata and RDF for the more complex products. Whatever format Dublin Core metadata takes, its beneficial flexibility and cost-effective simplicity has made it an internationally recognized metadata leader.

European CEN Metadata Standard

Departing from the individual "national standard" ideology, most European countries follow the European CEN Metadata Standard known as the CEN/TC 287 metadata standard for geographic information. Since the late 1990s, CEN/TC 287 has become the most widely adapted geospatial model in Europe that is fully maintained and standardized through a technical committee.

The CEN/TC 287 metadata standard is governed by the Comité Européen de Normalisation (European Committee for Standardization), or CEN for short. CEN was founded in 1961 and contributes to the standardization, exploration, and development of various research programs. It utilizes special technical committees (TCs) for each of its program areas, with TC 287 dedicated to geographic information.

Differing from the FGDC's and Dublin Core's hard-nosed structure, the CEN model offers flexibility with the number of different elements. For instance, the Dutch version of CEN/TC 287 (called NVN-ENV 12657) contains nearly 300 metadata elements. Interesting enough, if you were to compare, side by side, CEN/TC 287 to either the FGDC or Dublin Core standards, you could easily link one CEN/TC 287 element tag to many FGDC/Dublin Core element tags.

Now, GIS catalog software producers are taking notice of the European CEN Metadata Standard. ESRI, in particular, has developed a way to link CEN/TC 287 standards and apply them to FGDC's format. Their ArcCatalog software, for instance, takes CEN metadata and transforms it into the CSDGM schema in XML. As more and more software producers implement the CEN

model, these standards will gain even more popularity within Europe and, possibly, elsewhere.

In recent years, the initial publishing of ISO/TC 211, a widely accepted international geospatial metadata standard, brought the work of Technical Committee 287 to halt. As the ISO/TC 211 metadata standards were disseminated and implemented, CEN realized that TC 287 was again needed, but this time to help implement the new ISO standards within Europe through the CEN model. Therefore, the European CEN Metadata Standard is still one of the industry's major metadata leaders. For more information about the CEN standard, visit their Web site at http://www.cenorm.be.

Australian Government Locator Service Metadata Element Set

The Australian Government Locator Service (AGLS) Metadata Element Set first developed from a National Archives of Australia workshop in late 1997. Since then, the AGLS Metadata Element Set has become the leading geospatial metadata standard used throughout Australia. The AGLS element set is designed to enhance the exposure, discovery, access, and interchange of available government resource information and services.

The AGLS element set is a group of 19 descriptive elements that are intrinsically based upon the 15 element descriptors from the Dublin Core Metadata Element Set. The AGLS element set, however, is much more complex, since it allows richer element description through element qualifiers. These qualifiers enable the elements to describe a greater array of resource categories.

Because the AGLS element set is a derivative of the Dublin Core element set, both are entirely compatible with each other. This fluid data interchange adds to the greater endeavor of universal metadata standardization and enhanced resource access. The AGLS Metadata Element Set remains a modern and attractive standard that not only operates seamlessly with other standards but also adds a discriminating level of geospatial resource description.

The UK GEMINI Discovery Metadata Standard

The UK GEMINI Discovery Metadata Standard is a defined metadata element set used within the United Kingdom for describing geospatial data for faster, more accurate searches (metadata discovery). As such, this U.K. standard focuses entirely on discovery-level metadata, which is the specific metadata responsible for expediting resource information retrieval. The UK GEMINI standard is an element set derived from ISO 19115 (Geographic Information—Metadata) and the U.K. e-Government Metadata Standard (e-GMS). This metadata standard has gained industry respect and use, since it is nourished by the government, academia, and the private sector geospatial community.

The UK GEMINI Discovery Metadata Standard is a distant relative of the Dublin Core standard, since it has been developed to match explicitly with the element tags of e-GMS, a standard based upon the Dublin Core. The UK GEMINI standard uses 32 elements, ranging from *Title* and *Data Format* to *Use Constraints* and *Lineage*. Similar to the CSDGM, there are mandatory and optional elements, specifically 17 mandatory and 15 optional elements. Some of these elements have just a single occurrence, while others are repeatable.

Adding to the level of discovery the UK GEMINI standard offers is the depth of description and information attached to each element. Each element not only has the typical name, description, and value but also has extra description information including equivalent e-GMS and ISO 19115 element names and the element's data type. Given the added information, the UK GEMINI standard fulfils its main purpose of heightened geospatial metadata discovery.

Executive Orders 12906 and 13286

In the 1990s, the U.S. government became very interested in the retention, organization, and dissemination of geospatial data. By this time, GIS, GPS, and satellite imagery were being used in full force throughout the federal and local levels of government. As public interest in natural resource stewardship, environmental protection, and economic development soared, so did the plethora of data resources, duplication of data efforts, and inconsistencies in resource retention. It became clear to the government that standardization and assemblage of data resource dissemination was desperately needed.

Acting as a geospatial impetus, Federal Circular No. A-16, originally published in 1953 for survey and mapping issues and revised in 1967, was again revised in 1990 to include spatial data. This second revision marked the origin of the FGDC and took the first steps toward a comprehensive National Spatial Data Infrastructure (NSDI).

On April 13, 1994, President Bill Clinton signed Executive Order 12906, which was a giant step toward coordinating geospatial data acquisition and access. In it, the NSDI was formed and a mandate for the FGDC to create geospatial data standards was enacted. FGDC's Content Standard for Digital Geospatial Metadata was essentially born.

Even more specific to GIS, Executive Order 12906 detailed plans for a national digital geospatial data framework, whereby all federal agencies and organizations were required to document their geospatial data using the FGDC's standard. Additionally, all published metadata now had to be distributed and made available through a clearinghouse system. As you can imagine, this executive order proved to be the push needed to set the geospatial community in motion.

Contributing to the release and dissemination of geographic and geospatial information came a second significant executive order from the U.S. government. On March 5, 2003, President George W. Bush signed Executive Order 13286, which amended Executive Order 12906 to represent changes in government, as well as to represent changes in geospatial management, changes in technology, and clarifications to FGDC responsibilities. Without question, these federally imposed geospatial dictates helped to build the enormous data repository available today and to disseminate the importance and efficacy of geospatial metadata.

The National Spatial Data Infrastructure

With the signing of Executive Order 12906, the FGDC was charged with the task of developing the NSDI. The original concept was to have a centralized architecture for geospatial data policies, standards, and procedures. The NSDI was the conceived solution to disseminate geospatial data in a consistent fashion, while flattening the growing costs of duplicated governmental geospatial data collection.

As part of the vision, NSDI would provide an infrastructure through which data producers and users could share geospatial data. This producer-user relationship has already started to emerge from FGDC's efforts. Through the NSDI, the FGDC has begun to make key partnerships with public and private data producers to increase data availability to geospatial data users.

Incidentally, one valuable outcome of the NSDI is that it affords emergency services the capability of accessing geospatial data in an easy, expedited manner. In emergency situations, such as natural disasters, terrorism, and accidents, expedited and consistent data retrieval can truly hold life and death in balance. As the NSDI continues to develop and grow, the scope of available benefits and innate advantages will become ever clearer.

Although in recent years the FGDC has made great strides in implementing the NSDI, more work is needed. NSDI is still a work in progress. FGDC's effort is ongoing and is in cooperation with many organizations from state, local, and tribal government levels as well as the academic community and the private sector. The U.S. government places great trust in the fact that this infrastructure will provide substantial cost savings and facilitate enhanced decision-making. The truth of the matter is that the NSDI is still in its early stages and, in the coming years, may prove to be the resource that was originally envisioned. As with all things, only time will tell.

Geospatial Metadata Clearinghouses and Publishing Resources

As mentioned previously in this chapter, the U.S. government mandated that all published geospatial metadata be made available through a clearinghouse system. A metadata clearinghouse is an online or offline reservoir of published metadata that is being offered to the public. These clearinghouses in general offer freely distributed metadata, although some provide for-purchase variations of hard-to-obtain information.

Table 4.1 details just a small selection of international geospatial metadata clearinghouses. All of these sites have free metadata. There are, however, hundreds of available metadata clearinghouses available, mostly online.

Often these geospatial metadata clearinghouses also offer free and for-purchase geospatial data sets. We discuss these particular geospatial data clearinghouses in more detail in Chapter 15. For now, focus on the geospatial metadata these data bureaus provide and understand that the best way to create your own metadata is to examine what has been done by industry peers.

Take some time to explore the clearinghouses detailed in Table 4.1. You can gain a great deal of understanding by examining the language, syntax, and structure of a few good metadata samples. In fact, these samples will serve as distinct resources as to how compliant metadata are written.

If you want to try your hand at writing metadata, a good place to start is with the U.S. Geological Survey and, in particular, Peter Schweitzer's work with metadata. Mr. Schweitzer has written numerous documents on metadata basics and created software to help people write formal metadata (Unix-based XTME and Windows-based TKME editors), preparse it for proper structure [i.e., "chew and spit" (CNS) metadata preparser], and parse it for compliance with FGDC standards [metadata parser (MP) software].

In addition to the FGDC-compliant software, each metadata standard (e.g., Dublin Core) has its own editors and parsers for compliance with its standards. Check out the particular standard's Web site for more information on its available software and requirements. Once you become more comfortable with metadata and GIS, you will be able to create your own metadata and publish it to a clearinghouse.

Table 4.1
Geospatial Metadata Clearinghouses and Publishing Resources

Nairobi Metadata Explorer (GRID)	http://gridnairobi.unep.org/metadataexplorer/explorer.jsp
Colorado Plateau Environmental Metadata Clearinghouse	http://mprlsrvr1.bio.nau.edu/metadataexplorer
Eastern Sierra Geospatial Data Clearinghouse Metadata	http://www.wmrs.edu/resources/data%20access/clearinghouse/contents.htm
EPA Geospatial Data Clearinghouse	http://www.epa.gov/nsdi
FGDC National Data Geospatial Clearinghouse	http://clearinghouse1.fgdc.gov
Geospatial One-Stop Portal	http://www.geodata.gov
National Biological Information Infrastructure (NBII) Metadata Clearinghouse	http://www.nbii.gov/datainfo/metadata/clearinghouse
University of Sydney's ECAI Clearinghouse	http://ecaimaps.berkeley.edu/clearinghouse/
NSDI Geospatial One-Stop metadata publishing guidance	http://www.geo-one-stop.gov/metadata
National Oceanographic and Atmospheric Administration (NOAA) Coastal Data Locator	http://www.csc.noaa.gov/cgi-bin/id/cid2k/cid2k.cgi?page='cdl'
Tutorials on setting up a clearinghouse node	http://www.fgdc.gov/clearinghouse/tutorials/howto.html
U.S. Geological Survey (USGS) XTME metadata editor (Unix)	http://geology.usgs.gov/tools/metadata/tools/doc/xtme.html
USGS TKME metadata editor (Windows)	http://geology.usgs.gov/tools/metadata/tools/doc/tkme.html
USGS CNS metadata preparser	http://geology.usgs.gov/tools/metadata/tools/doc/cns.html
USGS MP metadata parser	http://geology.usgs.gov/tools/metadata/tools/doc/mp.html

Part II:
Geodesy, Earth Models, and Coordinate Systems

Understanding the Core GIS Environment

5

The Basics of Geodesy and Scale

The story of Christopher Columbus and his bold testament to the Spanish monarchy that the world was round and not flat is captivating and has remained an interesting tale within world history. Learning of Columbus as a child, I remember thinking: how could they believe the Earth was flat? I now look back and realize that science had proved the common understanding of the Earth's shape long before you and I were born. In fact, we were born into a society holding common and mature geographic knowledge, much different from Columbus, who was born into a society with varying images of the Earth, abundant misunderstandings, and science in its infancy.

So where does that leave us in terms of GIS? Well, for starters, through the discussions in Part I, we have begun clarifying the role of GIS by way of a primary understanding of GIS's groundwork. Let us now go further, tackling the core of GIS, the nature of geographic data, and the backbone to correct geographic representation and structure. Therefore, the science behind the Earth is our next destination, since it serves as a logical path to understanding GIS and its employment within our society. For this, we are directed toward the study of the Earth and the principles behind its shape, its size, and the laws that govern its portrayal. Let us begin Part II with geodesy.

What Is Geodesy?

Geodesy is the division of science associated with the measurement and portrayal of the Earth. Geodesy, known also as *geodetics,* is intimately concerned with the determination of the size and shape of the Earth, as well as its elements.

These Earth-based elements include its terrestrial gravity, magnetic field, tides, geologic and crustal movement, and polar motion.

As a study of the Earth and its rudiments, geodesy covers a large analytical landscape and combines the sciences, mathematics, physics, and observation. This analytical landscape and questions surrounding Earth's geodetic elements first initiated the science of geodesy. The study of geodesy began as mere curiosity and an undying human will to explain the Earth's unknowns through logic. There exist the first remnants of geodetic analysis starting as far back as the early Greeks. Homer, Pythagoras, Plato, and others all had ideas about the shape of the Earth. Homer, for instance, held the idea that Earth was a large flat disc, while Pythagoras, a mathematician, viewed Earth as an idyllic spherical figure. Plato logically guessed that the circumference of the Earth was 40,000 miles.

This early form of geodesy was based upon visual observations, logical assumptions, and crude calculations and transformed the notion of a flat Earth to the idea of a curved surface. In fact, Greek sailors took notice that as they came closer to land the visible high points seemed to ascend from the water. Others noticed that astronomical events such as a lunar eclipse occurred high in the night sky on one end of the Mediterranean while on the other end of the Mediterranean it occurred closer to the horizon. These observations added reason to the common consensus in a round or curved Earth.

Many, like Plato, made educated guesses as to the size of the Earth, but they were just guesses. Archimedes went one step further, providing an educated approximation that the circumference of the Earth was 30,000 miles. However resolute, no one could definitively prove the Earth was round or determine its true size. Along came Greek mathematician Eratosthenes who with curious mind and mathematical craft created a method to answer these geodetic unknowns.

Eratosthenes had taken notice during a summer solstice (longest day of the year) that a stick placed straight up in the sand cast no shadow and that the midday sun (directly overhead) shone to the very bottom of a well. Both observations occurred in his home town of Syene (now Aswan), Egypt. Eratosthenes also noticed that at the same time, the sun was not directly overhead in Alexandria, a town roughly 500 miles away. As a result, a shadow was cast against the vertical in Alexandria that was equal to 1/50th of a circle or 7 degrees 12 minutes (7° 12′). See Figure 5.1.

Eratosthenes took his summer solstice observations and applied some assumptions: (1) The distance between Syene and Alexandria was 500 miles; (2) since Syene's midday summer solstice sun was directly overhead and, on that day, the sun moves along the tropic of Cancer, it was concluded that Syene was on this tropic zone line; and (3) Syene and Alexandria lie on a direct north-south line.

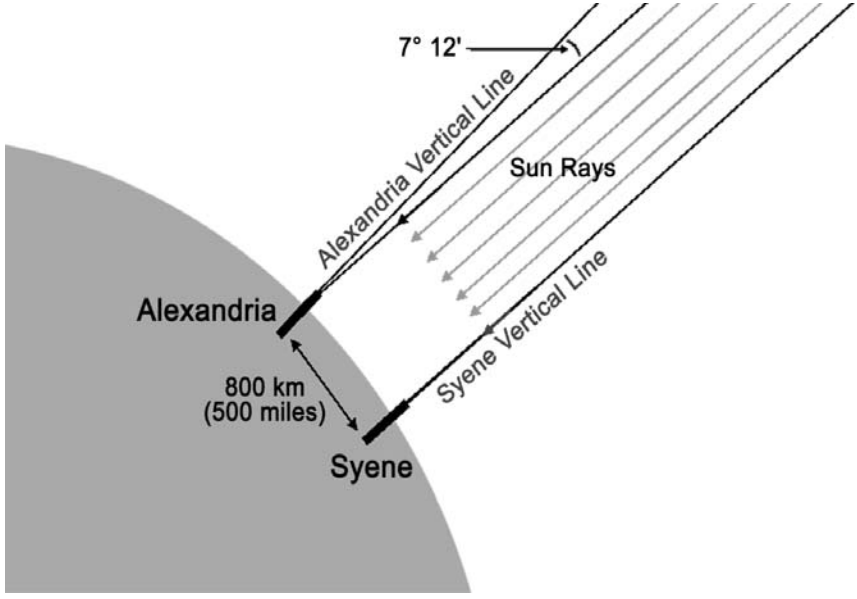


Figure 5.1 Eratosthenes's method of calculating the Earth's size.

Based upon these assumptions and the measured shadow, Eratosthenes concluded that the distance between Syene and Alexandria was 1/50th the circumference of the Earth. Mathematically it can be explained as:

$$1/50 \times \text{Earth's circumference} = 500 \text{ miles}$$

$$\text{Earth's circumference} = 50 \times 500 \text{ miles} = 25{,}000 \text{ miles}$$

Remarkably, Eratosthenes made a nearly accurate measurement of the Earth's size given that today's accepted World Geodetic System (WGS) calculation is 24,901 miles! However, Eratosthenes's calculations were somewhat incorrect, since every assumption was a bit erroneous. We now know that (1) the real distance between Syene and Alexandria is 453 miles; (2) Syene lies 37 miles north of the tropic of Cancer line; and (3) Syene lies 3 degrees 30 minutes (3° 30′) east of Alexandria's north-south line. Additionally, Eratosthenes's calculation of 7° 12′ from the vertical is actually 7° 5′. Despite these facts known in hindsight, Eratosthenes became the first to substantiate the calculation of Earth's size.

After Eratosthenes, others continued to revise the Earth's measurement with varying results. It was not until after the invention of the telescope,

development of the logarithm tables, and progressive understanding of the Earth's true shape that more exact Earth size measurements were calculated. With these new measurements and Earth size-shape concepts came the maturation of geodesy and the development of the current conventions.

Present-day geodesy is well beyond questioning Earth's overall shape and size; it now focuses upon Earth's actual shape with its unique physical terrain. Geodesy centers upon Earth's significantly vast landscape with its drastic elevations and varying land forms. This deviating terrain is a true surface and includes high mountain ranges, sweeping plains, and variable geologic surfaces. This true surface adds countless levels of scientific complexity to geodesy and, in conjunction with the desire to measure and represent the Earth, forces concern about both accurate Earth positioning and the determination of Earth's actual shape. Thus, reducing the Earth to a simple shape (i.e., a spheroid) is of menial scientific interest in current-day geodesy other than for simple calculations and geographic representation.

Present geodetic concerns subsist within a three-dimensional time-varying space. This means that over time the Earth's surface changes due to temporal surface variations, geologic events, the gravitational field, and other surface-altering episodes. Erosion, for instance, will permanently change a coastline over time, or a river basin swollen by excess storm water runoff will temporarily alter features on the Earth. Both, however, alter the true shape of the Earth at any given time. These sometimes unpredictable changes add layers of geodetic complexity to the accurate measurement and representation of the Earth.

Geodesy has also advanced and now, to some extent, relies upon two forms of terrestrial measurement: geomensuration and surveying. *Geomensuration* is the measurement of the Earth as a whole (i.e., on a global level), while *surveying* is the measurement of individual parts on the Earth's surface (the ever-changing physical terrain). Both techniques offer credible measurements wholly useable toward accurate geodetic positioning, surface shape, and size. GIS employs these forms of terrestrial measurement through various data sources, such as satellite, GPS, and field measurements. As such, geodesy serves as an important science and elemental backbone to GIS.

With this understanding, we are now ready to discuss the pertinent geodetic elements that define geodesy and provide crucial pieces of geographic information. Specifically, the reference Earth shape (the ellipsoid), geodetic positioning (geodetic datum and coordinates), the true Earth shape (the geoid and vertical datum), and the usable portrayal of the Earth (map projections). The next four chapters go into great detail about these geodetic elements; however, before we move into these areas, we must first discuss the principles of map scale. In short, map scale provides the exact geodetic relationship between the GIS information product and reality.

Understanding Map Scale

You are driving on a desolate back road in Montana and your car's gas light turns on. Nervously you reach down and grab a road map. You easily find your position and pan the paper map for the nearest town that would have a gas station. Three inches from your position is a reasonable sized town by the looks of it. From experience you know that when the gas light turns on, there is enough fuel in the tank for another 40 miles. Three inches isn't that far, right?

As the dread of waiting for a tow truck fills your mind, you notice in the corner of the map the words, "One inch represents 20 miles." You swiftly conclude that 3 inches must represent 60 miles and that you can reach the town before running out of gas. You continue toward town with newfound peace of mind.

We have all been in a similar situation, though maybe not as dramatic. The group of words that ultimately informed and relieved the Montana traveler is called a map scale. A map scale, in its most common treatment, relates to the size of map-based features in relation to their actual, real-world size. Specifically, map scale is the relationship between a distance on a map and its corresponding distance on the ground.

Although detailing the same information, not all map scales look alike. In fact, there are three common forms of map scale: verbal scale, graphic scale, and ratio scale. Refer to Figure 5.2 for a graphic illustration of each discussed map scale.

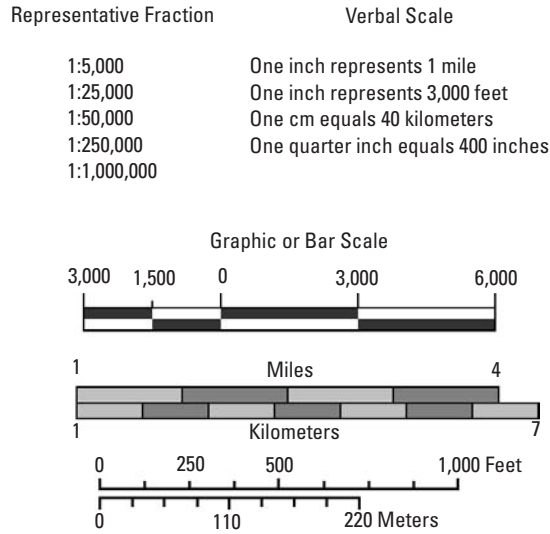


Figure 5.2 Key map scales.

In the example Montana situation, a *verbal* scale was used. Verbal scales express the map-to-ground relationship in words, such as "One inch represents 20 miles" or "One inch equals 20 miles."

A *graphic* scale or *bar* scale depicts a bar of a certain length that represents a distance length. Graphic scales enable the map user to visually recognize distances with relative ease. Because of this simple understanding, graphic scales are undeniably the most used form of map scale. One striking feature of the graphic scale over verbal and ratio scales is that graphic scales remain accurate if the map is photographically enlarged or reduced.

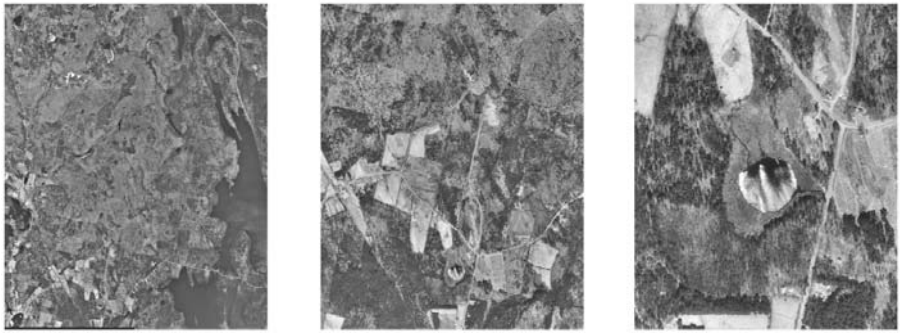
A *ratio* scale, more formally known as representative fraction and less formally as natural scale, reduces a map's distance to 1. One unit on the map corresponds to a distance in the same units on the ground. In a typical 30-minute series topographic map, the representative fraction is shown as either "1/125,000" or "1:125,000" whereby the 1 and 125,000 are the same unit of measure.

When dealing with GIS products, the term *scale* can also be used in a wholly different manner. The term holds significance to describe the working size of an application or project area, specifically *small scale*, *mid-scale,* and *large scale*. Parameters for each distinction are highly flexible, based upon individual judgment rather than a specific criterion.

Small-scale applications involve relatively large areas, such as a map of the world, the continent of Australia, or the city of Rome, Italy. A typical small-scale representative fraction can be assumed anything greater than 1:500,000. Large-scale applications involve small areas of focus, such as a community, a city block, or a university campus. A typical large-scale representative fraction can be assumed anything less than 1:50,000. Mid-scale applications, on the other hand, fall right in the middle of large- and small-scale applications, with a typical representative fraction ranging from 1:50,000 to 1:500,000.

As stated earlier, parameters for this scale nomenclature are determined by the individual user or the specific application. What is considered large scale for one application may be considered as mid or small scale on another. Figure 5.3 presents an example of this flexible scale nomenclature. Notice that the relationship between representative fraction and scale nomenclature is different from our previously stated "typical" ranges. This specific application's overall area covers roughly 20 square miles with the small-scale view covering the entire area and the large-scale view covering less than one square mile.

It is easy to get confused with this scale nomenclature, thinking a small-scale application should show only a small area. This, of course, is wrong. There is a rule of thumb to remember when using this scale terminology: Large scale has a large representative fraction [i.e., the denominator (number after the colon) is small] and small scale has a small representative fraction (i.e., the denominator is large).



1 : 126,000
Small Scale

1 : 61,400
Mid-Scale

1 : 7,880
Large Scale

Figure 5.3 Large-, mid-, and small-scale applications.

Simple Map Scale Calculations

Regardless of specific format, map scale can be calculated using the following equation:

$$\text{Map scale} = \text{Distance on map}/\text{Distance on the ground}$$

In the Montana example, every map inch (distance on the map) represented 20 actual miles (distance on the ground), thus, map scale = 1/1,267,200.

There are many instances when map scales need to be converted into another form, such as a verbal scale to a representative fraction or vice versa. These scale transformation calculations are simple and prove worthwhile for various cartographic and GIS applications. Let us take a look at these calculations in Table 5.1.

Scale and Accuracy

An important characteristic of geographic data is that it can be represented differently at different scales. Proper map scale within GIS and on hardcopy maps enables the user to display accurate geographic data that is easily discernable, representative of the particular application and wholly successful in its implementation. Geographic features that are prominent at 1:24,000 scale are many times indistinguishable at 1:1,000,000. As with any GIS function, a user's knowledge of the final use is pertinent to the quality production of geographic information products.

Table 5.1
Scale Transformation Calculations

Conversion from Verbal Scale to Representative Fraction
1 inch represents 4 miles
1 inch equals 4 miles
1 inch = 4 miles
1 inch = 4 miles × 5,280 feet/mile
1 inch = 21,120 feet × 12 inches/foot
1 inch = 253,440 inches
1:253,440
Conversion from Representative Fraction to Verbal Scale
1:100,000
1 inch = 100,000 inches
1 inch = 100,000 inches × (1 foot/12 inches)
1 inch = 8,333.33 feet × (1 mile/5,280 feet)
1 inch = 1.578 miles
1 inch represents 1.578 miles

For specific applications or to develop certain targeted information products, geographic data representation can be manipulated by increasing or decreasing map scale. For instance, state planners targeting an individual land parcel would employ parcel maps at 1:24,000 or larger scales. If they were targeting regional land parcels instead, area maps at 1:250,000 or smaller scales would be the most useful.

When scale is adjusted by enlarging or minimizing the area view, accuracy of the data becomes a concern. Typically, as scale decreases, the number of specific features and the detail of those features also decrease. As such, data accuracy is lessened to the extent that the overall representation is more general than descriptive. When scale of the original data is increased, accuracy is not improved, but features are enlarged.

For example, take another look at Figure 5.3. The frozen pond prominent in the large-scale view (1:7,880) is but an indistinct feature among the landscape when scale is decreased to a small-scale view (1:126,000). As apparent, the amount of feature (pond) detail in the large-scale view is much greater than in the other two views. Consequently, when the scale gets reduced (large to small), feature detail becomes too little to see and cannot be adequately represented.

Proper map scale is directly proportional to accuracy. Sometimes when map scale is decreased and maps become small scale, geographic detail is lost. This can cause serious user problems, since geographic unknowns arise and accuracy diminishes. For instance, imagine if a person new to Interstate 81 in Virginia used a map of North America instead of a state map to get from Pulaski to Harrisonburg. The user may or may not get to his destination, since the North American map is dramatically less accurate in terms of Interstate 81.

In short, scale and accuracy go hand in hand with both GIS-based products and hardcopy cartographic maps. Every data source has accuracy limitations, and users must make every attempt possible to define the legitimacy of data sources. If scale and accuracy are used appropriately, the geographic information product will be used as intended and offer a well-defined and highly efficient representation.

Geographic Data Generalization

Geographic technologies represent the real world mainly for convenience, similar to the role played by hardcopy paper maps used by our wandering ancestors. For example, a traveler cannot simply fold up a true representation of his 100-mile route and stick it in his coat pocket. For the sake of convenience and portability, the 100-mile route is reduced to a scale of 1:100,000 (i.e., 1 inch equals 100,000 inches), which is then drawn and portrayed on paper, folded, and placed in his pocket. This diminution of scale is adequately achieved through generalization.

Generalization at its core is a method to manage and reduce unwieldy or overly numerous input data sets. Through this process of reduction, the real-world data source is rendered in less detail and, as a result, is somewhat distorted from reality. Each level of generalization introduces a greater degree of distortion; however, immediately, generalization affects the overall appearance and, even more important, the precision of the output data. Smart geographic data users are aware of the precision of data they use because it has a direct correlation to and effect upon the ultimate usability of their geographic information products.

Even by today's technological standards, an extremely hefty 500-gigabyte high-resolution image is not a manageable quantity of input data. Like trying to eat a 6-foot sandwich all at once, it is simply not feasible. So what is considered a manageable quantity? Every user has an individual parameter for manageability, which takes into consideration the computer and hardware specifications, the patience for waiting out long processing tasks, and the ultimate goal of analysis. A median case could potentially involve generalizing the 500-gigabyte image to

a more tolerable 1.2-gigabyte image, resulting in a sleeker file more adaptable and friendly to geoprocessing.

All geographic data are in some way generalized. This is an inevitable, astringent process that alters the actual character of the data into a simplified and shrunken model of the real world. Ultimately, scale dictates the severity of generalization, varying in level from project to project. For instance, a map drawn to 1:100,000 scale (smaller scale) is more generalized than the same area drawn to 1:24,000 (larger scale), thus the smaller the scale, the greater the level of generalization. Chapter 11 take a much closer look at scale factors and the resultant degrees of distortion.

Figure 5.4 exemplifies geographic data generalization, wherein a physical structure is portrayed in relation with two different interpretations of the same roadway.

Figure 5.4 illustrates the quintessential "house on the wrong side of the road" problem. The 1:50,000 scale details a generalized road condition and offers severely less detail. The 1:10,000 scale details a close representation of the actual roadway, with the appearance of blatant roadway shape changes. A house (physical structure) is shown close to the roadway. In reality and as depicted on the larger-scaled (1:10,000) roadway drawing, the house appears left (or west) of the roadway. However, on the smaller-scaled (1:50,000) roadway depiction, the house is on the right (or east) of the roadway, forming a basic generalization of actual position and real-world physical shapes.

Decisions concerning the use of geographic data and their suitability for any given application are paramount to the validity of the geospatial system. Selecting an adequate scale for a project is a balance between positional accuracy and the amount of geographic data processing. For instance, take the more generalized road and apply it to a tourism booklet's map. Geographic data requirements are minimized. Additionally, the effect of the rounded road is perhaps negligible considering the native generality of the tourism application. However, take that same geographic data with the same "rounded" characteristics and apply it to an emergency vehicle deployment routing application. That mistake in judgment may have taken someone's life! The depicted geographic data at a 1:50,000 scale are generalized beyond minimum suitability for the emergency deployment application. The vehicle drivers are looking for a specific house on the left that is actually on the right! A 1:10,000 scale would be a more suitable level of generalization for ensuring the system's integrity for positional accuracy and left-side/right-side roadway events.

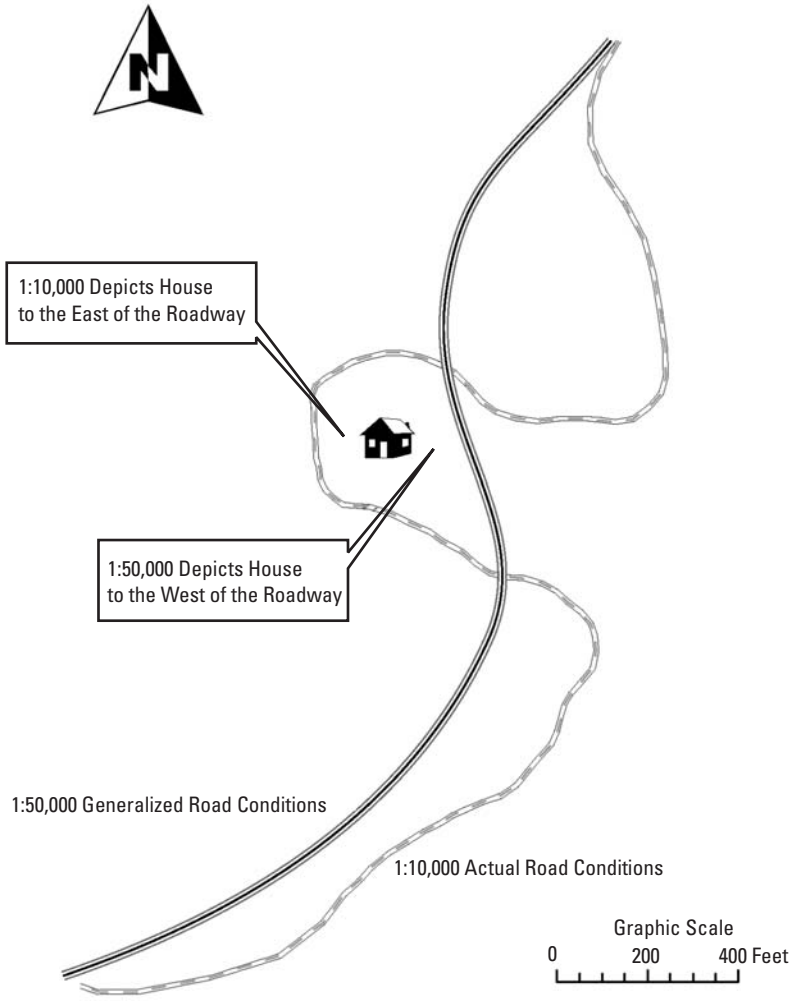


Figure 5.4 The "house on the wrong side of the road" problem.

6

The Ellipsoid

Throughout the early years of geodetic understanding, some of the most important components of geodesy were realized through innovative thinking, physical science, and debate. The spherical Earth-flat Earth debate previously discussed initiated the science of geodesy and spawned numerous attempts at defining the circumference of the Earth. By the seventeenth century, scientists, mathematicians, and others had begun to realize that the Earth wasn't quite perfectly rounded as first believed. New theories soon emerged out of the scientific communities, and the quest for the true Earth model began. This quest ended with the ellipsoid and subsequently recast physical science and geodesy forever.

Today's precision GIS is built upon the advances in modern geodesy and the finely tuned knowledge of the Earth's true shape. With satellite imagery and modern physical science, there is now no question as to what the world looks like and what shape best represents the Earth—a nearly round object that bulges in the middle. To this end, ellipsoids are exceedingly important to GIS because they pose a near resemblance to the Earth's shape without the inherent complexities. In short, ellipsoids simplify GIS.

To truly understand the core GIS environment, we must first understand how we got from Eratosthenes's curved Earth and the pure sphere to an ellipsoid, a much closer physical representation (model) of the real Earth. We will discuss the physical parameters of the ellipsoid and how this elliptical Earth model is manipulated and referenced within GIS.

History of the Earth Model

From the time Eratosthenes calculated the Earth's circumference in the third century B.C., the Earth model for many took on the shape of a sphere. Generally, a sphere is a symmetrical three-dimensional shape, in which the distance (radius) from the surface to a common center point is identical anywhere on the object (i.e., height equals width). It can be thought as a circle rotated around an axis.

The sphere shape for the Earth model was the initial conclusion based on a combination of minimal information and poor measurement tools. It was the logical curved successor to the "flat Earth.. To its credit, it remained a favored Earth model from the third century B.C. to the seventeenth century A.D., when science made it possible to measure distances and gravity with precision. Telescopes were invented and became indispensable tools for early geodesists.

Soon, imperfections in the Earth's curvature became widely apparent with the recognition and measurement of dissimilarities in different locations. This dissimilarity invoked new theories concerning the Earth model. As a direct result, two predominant and disputing theories emerged from the geodetic community, namely from the French and the British communities. Figure 6.1 details the competing Earth model theories.

The story of the British-French rivalry concerning the Earth model is legendary and quite interesting. Both wanted to be correct in order to attain a level of international excellence and scientific supremacy. Leading the French were the Cassinis, a renowned family of geodesists. After taking measurements in France from Paris to Perpignan, Jean-Dominique Cassini and his son, Jacques

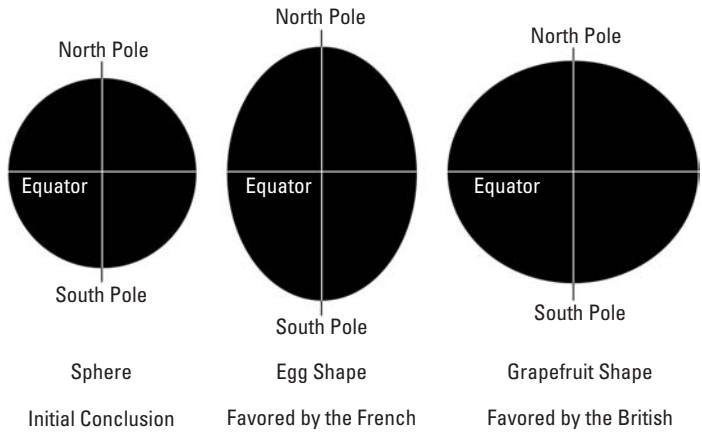


Figure 6.1 The predominant Earth models.

Cassini, were certain the Earth pointed toward the poles and flattened at the equator. As a result, the French model was a prolate ellipsoid and egg shaped.

The British took a different approach. They were armed with Isaac Newton's then recently released gravitational theory published in Newton's *The Principia*. The British were equally certain that the Earth was flattened but that it was occurring at the poles. The British model was an oblate ellipsoid—grapefruit shaped and bulging in the middle.

The competition became serious as international interest grew. In fact, the contrasting theories invoked an international debate with the Earth *elongators* favoring the French model and the Earth *flatteners* favoring the British model. In an effort to end the dispute and determine who was truly correct, the French Academy of Sciences devised a plan that involved measuring a 1° arc in the North and on the equator. If the arc in the North was shorter than the arc on the equator, then the French would be correct. If the North arc was longer, then the British would be correct.

The French Academy of Sciences sent out two teams to make 1° arc measurements in Peru and Lapland. Peru was a reachable point near the Equator and Lapland was reachable and near the North Pole. This meridian arc measurement expedition started in April 1735 as the first team was deployed to Peru. In April 1736, just two months shy of the first team reaching the site in Peru, the second team was deployed to Lapland.

By August 1737, the Lapland team was already finished and back in Paris. The Peru team, on the other hand, had a much harder time with its meridian arc measurement. Due to the extraordinarily rough terrain in the Andes mountains, the team did not complete its measurements until 1743 and did not return to Paris until February 1745. Incredibly, the French Academy of Sciences' resolution took nearly 10 years to complete. In the end, the British were proved correct and the Earth model took on its new form: a grapefruit-shaped oblate ellipsoid.

Shape of the Earth

The ellipsoid, also referred to as a *spheroid*, is a much better approximation for the shape of the Earth than the sphere; the poles are slightly flattened and the equator bulges. Unlike the sphere, the ellipsoid can handle dissimilar dimensions. GIS relies heavily upon geographic positioning and the Earth model; therefore, the closer a model comes to the actual surface of the Earth, the better for geographic positioning. A sphere proves adequate along the equator (equatorial plane) but fails at locations closer to the poles. For this reason a GIS based upon a sphere model would produce flawed information products for most locations throughout the world, especially within the arctic and antarctic circles.

This Earth flattening defines our contemporary understanding of the shape of the Earth. Pull out a table globe and examine its features. Although difficult to immediately recognize on a globe, notice how the nearly spherical shape that ancient theorists surmised as "perfect" slightly bulges at the equator. Now imagine placing an invisible measuring stick through the core from the north pole to the south pole. It would measure 12,713,505 meters. Imagine placing another measuring stick through the Earth's core from one point on the equator to another. That stick would measure 12,756,274 meters. This relatively tiny difference in measurements causes the Earth's subtle bulge and defines the nature of its flattening.

Figure 6.2 offers a bird's-eye look at ellipse flattening as a whole. As you can see, a circle has zero flattening because the horizontal and vertical measurements are equal. When the measurements differ, that is, when the horizontal measurement is larger than the vertical measurement, ellipse flattening increases. Alternately, a larger vertical measurement defines an ellipse elongation.

Figure 6.2 illustrates flattening with large, visually dramatic flattening scenarios. However, in the case of the Earth, the north-to-south measurement is only slightly shorter than the east-to-west measurement and has a small flattening of approximately 1/300.

What makes the ellipsoidal model useful in GIS and other real-life applications (i.e., GPS systems, military artillery, science, and so forth) is its ability to

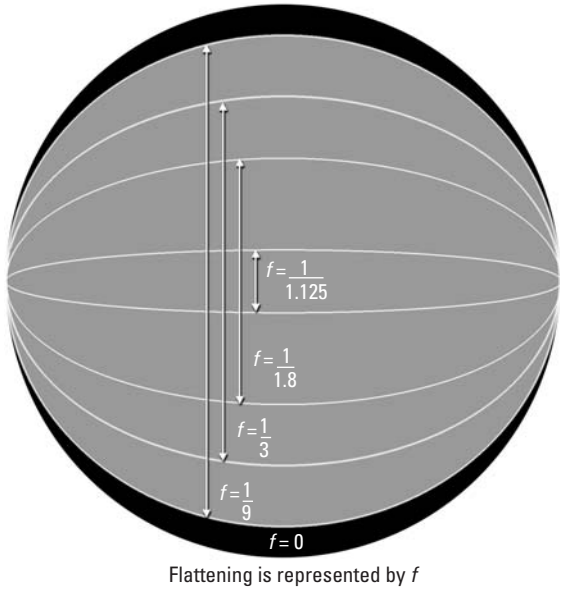


Figure 6.2 Various levels of flattening.

handle the Earth's flattening with ease and within the simple constructs of a relatively smooth shape. Now, the Earth is hardly smooth. We have mountains, plains, and valleys on land and deep, cavernous terrain within the oceans. We have cities that are below sea level (such as New Orleans, Louisiana) and others many miles above sea level (such as Katmandu in Nepal).

You might be thinking, how could the Earth's various terrains be served adequately by a smooth ellipsoid? Well, let us look at this question on a global level. The highest feature on Earth is Mount Everest at about 9 km above sea level. The lowest feature is the Pacific Ocean's Marianas Trench at roughly 11 km below sea level. The topographic height variation from highest to lowest is 20 km. Given that the shortest core-to-crust radius is 6,356.7 km, this topographic height variation would provide a minuscule surface change and is therefore negligible for Earth-modeling purposes.

We can liken these relatively subtle topographical height variations to the dimples on a golf ball; the overall shape is spherical, but the actual surface varies slightly in height. For ease, we use a smooth ellipsoid as the basis for calculations and geographic positioning. The difference between a smooth ellipsoid and one with topographic feature heights is negligible for many base-level calculations and for Earth referencing in GIS. We will look at these topographical height variations in greater detail in upcoming chapters.

Mathematical Model of the Earth

To adequately represent the shape of the Earth in scientific and real-life applications, a calculatable, formula-driven figure of the Earth is needed. Let us go to the basics of an ellipsoid to build a mathematical figure of the Earth. In its most basic sense, an Earth ellipsoid is a *flattened* sphere that bulges in the middle and has two imaginary lines traversing its core: one from north to south and one from east to west. Each imaginary line is called an *axis*.

The ellipsoid's flattening causes two unequal axes: a longer axis and a shorter axis. The north-to-south axis through the Earth's core is the shorter axis and, as such, is called the *minor axis* or *polar axis*. The east-to-west axis through the Earth's core is longer and is called the *major axis* or *equatorial axis*. Figure 6.3 illustrates a three-dimensional model and highlights the ellipsoid's flattening.

The Earth's ellipsoid is an ellipse rotated upon its minor axis, which is functionally called the *axis of rotation* or *axis of revolution*. The purpose of a mathematical model is to simplify calculations. Since an ellipse is a two-dimensional shape and an ellipsoid is a more complex, three-dimensional object, we utilize the figure of an ellipse to achieve this mathematical simplicity. Figure 6.4 details this two-dimensional mathematical model of the Earth.

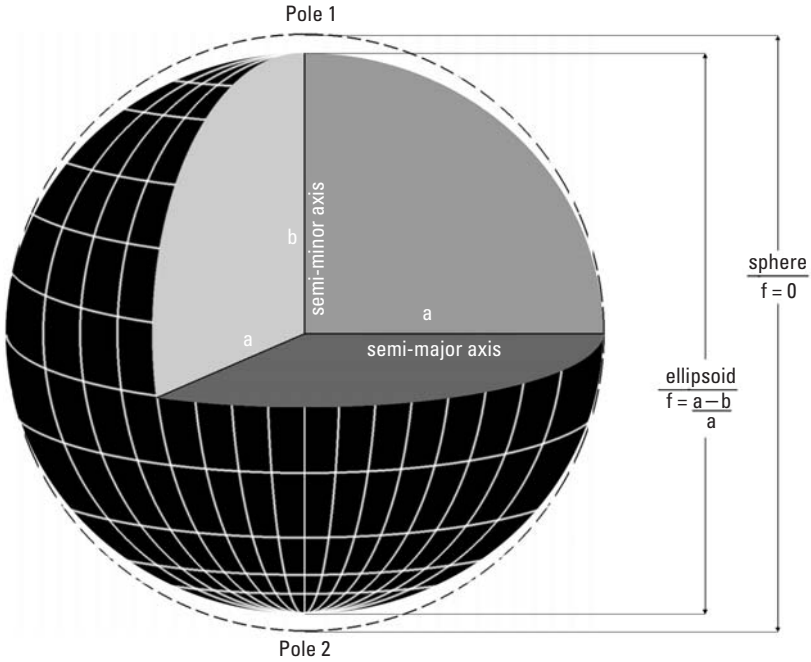


Figure 6.3 The three-dimensional oblate Earth.

Due to the Earth's symmetry and to minimize complexity, the mathematical model targets one quadrant in the ellipse and imparts six key parameters for calculation. The six parameters are the semimajor axis, the semiminor axis, flattening, inverse flattening, eccentricity, and second eccentricity. Let us go through each one while using Figures 6.3 and 6.4 for visualization.

The semimajor axis is an equatorial radius and is defined as one-half of the major axis. Alternately, the semiminor axis is a polar radius and is defined as one-half of the minor axis. Both hold major roles within the mathematical model and are represented by *a* and *b*, respectively. As we will see in the next two chapters, both the semimajor and semiminor axes are instrumental in defining accurate GIS point locations.

We have already discussed flattening but have not fully addressed its calculation. In brief, the flattening of the ellipse is directly related to the differences in both the semimajor and semiminor axes. It is represented by the formula

$$Flattening\ f = \frac{a-b}{a}$$

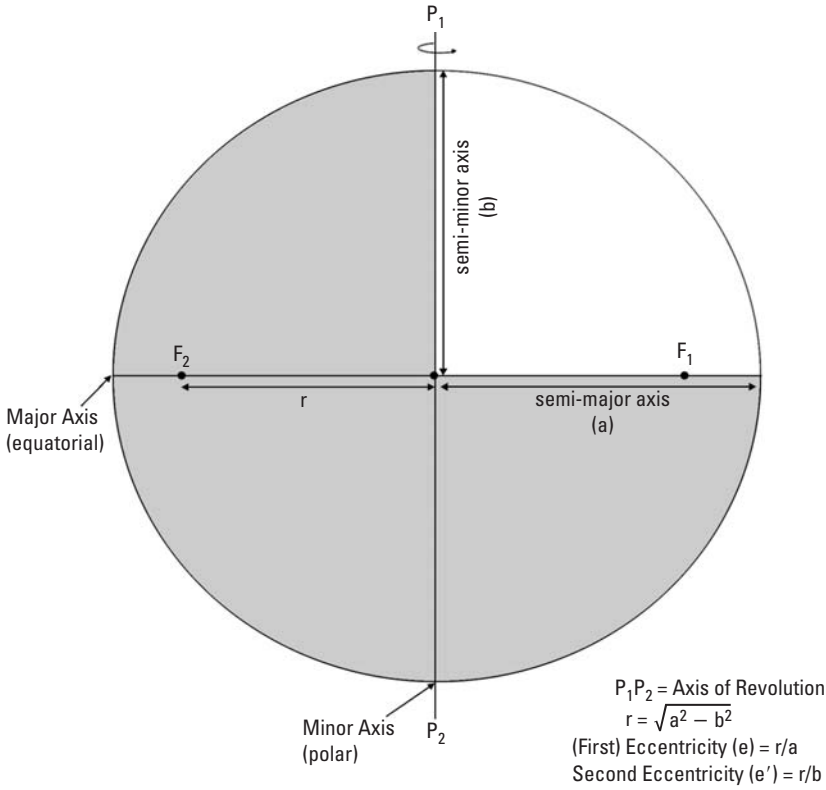


Figure 6.4 The two-dimensional, six-parameter mathematical model of the Earth.

Typical of this geodetic application, ellipse (and ellipsoid) flattening is a very small number holding many decimal places. In GIS, science, and mathematics, flattening results can be onerous to transcribe from tables and presents an opportunity for human error. To minimize error and help simplify conveyance, the Earth's flattening is sometimes displayed as a reciprocal called the *inverse flattening*. For instance, a flattening of 0.003389831 (or 1/295) can be portrayed as an inverse flattening of 295. As you can understand, the inverse flattening is much easier on the eyes and even easier to incorporate into dense calculations.

The formula for inverse flattening is straightforward, represented as:

$$\textit{Inverse flattening } f^{-1} = \frac{1}{f} \textit{ or } \frac{a}{a-b}$$

The final two parameters, namely eccentricity and second eccentricity, have more to do with the mathematical nature of the ellipse than its geographic use. Both eccentricities uniquely characterize the shape of the ellipse through a relational measurement for the degree of flattening.

Eccentricity e is the first eccentricity, whereby the descriptor *first* is typically implied. It measures the degree of flattening as the relationship between the focal radius, or radius from the center point to the foci (i.e., F_1, F_2), and the semimajor axis. It is calculated through the formula

$$(First)\ Eccentricity\ e = \frac{Focal\ radius\ r}{Semimajor\ axis\ a}$$

In any ellipse, the focal radius $r = \sqrt{a^2 - b^2}$; therefore,

$$(First)\ Eccentricity\ e = \frac{\sqrt{a^2 - b^2}}{a}$$

Second eccentricity e' measures the degree of flattening as a relationship between the focal radius and the semiminor axis. The formula for second eccentricity is

$$Second\ \ Eccentricity\ e' = \frac{Focal\ radius\ r}{Semiminor\ axis\ b} = \frac{\sqrt{a^2 - b^2}}{b}$$

Eccentricity (first and second) is constant throughout the entire ellipse and ranges from 0 to 1, or more simply from an ellipsoid with no flattening (circle) to one totally flat (line). Since a circle has no flattening, its eccentricity equals zero ($e = 0$). As e becomes larger and approaches 1, the ellipse becomes flatter until it is a straight line ($e = 1$).

Figure 6.5 illustrates this proportional relationship and endorses the eccentricity rule of thumb: The closer eccentricity is to 1, the flatter the ellipse.

Typical Earth Shape Calculation

To understand how the mathematical model is used for typical Earth shape calculations in GIS, let us calculate the Earth's six key parameters using the derived mathematical model as illustrated in Figure 6.4 and the standard measurements of the Earth's axes from the World Geodetic System 1984 (WGS84).

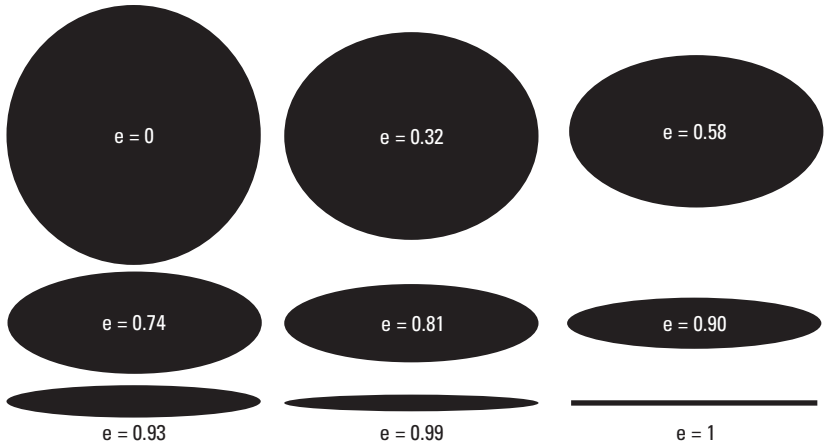


Figure 6.5 The constant eccentricity of an ellipse. (*Source:* D. Downing, *Dictionary of Mathematics Terms*, 2nd ed., Hauppauge, NY: Barron's Educational Series, 1995, after Figure 47.)

WGS84 Constants

$$\textit{Major axis} = 12{,}756{,}274 \text{ meters}$$

$$\textit{Minor axis} = 12{,}713{,}504.628 \text{ meters}$$

$$\begin{aligned}\textit{Semimajor axis } a &= \tfrac{1}{2}\,\textit{Major axis}\\ &= \tfrac{1}{2}(12{,}759{,}274)\\ &= 6{,}378{,}137 \text{ meters}\end{aligned}$$

$$\begin{aligned}\textit{Semiminor axis } b &= \tfrac{1}{2}\,\textit{Minor axis}\\ &= \tfrac{1}{2}(12{,}713{,}504.628)\\ &= 6{,}356{,}752.314 \text{ meters}\end{aligned}$$

$$\begin{aligned}\textit{Flattening } f &= \frac{a-b}{a}\\ &= \frac{6{,}378{,}137 - 6{,}356{,}752.314}{6{,}378{,}137}\\ &= \frac{21{,}384{,}686}{6{,}378{,}137}\\ &= 0.00335811\end{aligned}$$

$$\text{Inverse flattening } f^{-1} = \frac{1}{f}$$
$$= \frac{1}{0.003352811}$$
$$= 298.257$$

$$\text{Eccentricity } e = \frac{r}{a} = \frac{\sqrt{a^2 - b^2}}{a}$$
$$= \frac{\sqrt{6{,}378{,}137^2 - 6{,}356{,}752.314^2}}{6{,}378{,}137}$$
$$= \frac{\sqrt{272{,}331{,}609{,}224.645404}}{6{,}378{,}137}$$
$$= \frac{521{,}854.0114}{6{,}378{,}137}$$
$$= 0.081819191$$

$$\text{Second eccentricity } e' = \frac{r}{b} = \frac{\sqrt{a^2 - b^2}}{b}$$
$$= \frac{\sqrt{6{,}378{,}137^2 - 6{,}356{,}752.314^2}}{6{,}356{,}752.314}$$
$$= \frac{\sqrt{272{,}331{,}609{,}224.645404}}{6{,}356{,}752.314}$$
$$= \frac{521{,}854.0114}{6{,}356{,}752.314}$$
$$= 0.082094438$$

Reference Ellipsoids

Before leaving the subject of the ellipsoid, we must define its function within a GIS. Basically, every GIS and geographic information product is based upon a reference ellipsoid, which is defined as a standard ellipsoid with proven and measured parameters. The reference ellipsoid enables GIS to define point locations with regional accuracy.

Throughout the years, numerous reference ellipsoids have been developed by various geodesists with measurements taken at different source points on Earth. These reference ellipsoids are only slightly dissimilar. Because of this

differentiation in source data locations, certain reference ellipsoids work better than others for certain applications and for certain regions. For example, the reference ellipsoid Bessel 1841 is better suited for European GIS products than American GIS products and will produce more precise results due to its European-based measurement site.

Unlike Bessel 1841 and other major area-specific ellipsoids, WGS84 is much more universal and, as a result, is one of the most widely used reference ellipsoids available. It can be used for any location on Earth with a high degree of precision and flexibility and is backed by satellite measurements.

Appendix A, entitled "Reference Ellipsoid Parameters," lists many of the key reference ellipsoids available to GIS users. The appendix provides their associated equatorial radius, polar radius, and inverse flattening parameters, which are all necessary for proper designation in GIS. Since all geographic information products are built upon a reference ellipsoid, we will frequently refer to Appendix A throughout this book.

7

The Horizontal Datum

During the "Shock and Awe" military mission in Iraq, the U.S. armed forces were responsible for destroying enemy targets with missiles. The challenge was that these targets were in the heart of downtown Baghdad, near schools, mosques, and neighborhoods. The accuracy of missiles had to be extremely precise so that only the target was destroyed and the surrounding innocent people were not killed. Using standard ellipsoids to model the Earth helped bring the science of geodesy into a manageable and useful format for various applications. However, launching those missiles using positions on a standard ellipsoid would have been catastrophic, since positional accuracy would have been generalized, and thus the target's position would have been generalized. To maintain precision, U.S. missile systems implemented GPS-based targeting systems for the highest level of accuracy and to minimize collateral damage.

Sometimes standard ellipsoids provide enough accuracy and prove sufficient for small- and large-scale GIS applications. However, similar to the military missile systems, often a greater degree of precision and positional accuracy is needed. To help bridge the standard ellipsoid to a more exact Earth position, a precise reference point called a *geodetic datum* is used. Incidentally, specific geodetic datums are used to manipulate GPS properly and help guide missile targeting systems.

Let us examine these precision-enhancing ellipsoid references to obtain a clearer understanding of their importance and implementation within GIS. The complete discussion of datums is somewhat lengthy. In an effort to simplify and help organize the explanation of the topic, I have divided it into two logical parts: the horizontal control datums and the vertical control datums. This chapter focuses on datum theory, as well as horizontal control datums. The next

chapter completes the discussion with a detailed examination of vertical control datums.

What Is a Geodetic Datum?

By the purest sense of the word, a datum is any set of numeric or geometric parameters used to accurately measure or define another quantity. More specifically related to GIS, a geodetic datum is a "curved Earth" reference model that associates a geodetic reference ellipsoid (i.e., "parameters") to a coordinate system (i.e., "another quantity").

As previously discussed, a geodetic reference ellipsoid is attributed with certain characteristic parameters, such as the semimajor axis *a*, the semiminor axis *b,* and flattening *f,* while a coordinate system defines a geodetic space through orientation, position, and scale. Both are brought together with a geodetic datum, which, in itself, is purely a mathematical Earth model. Given this mathematical union it can be determined that any reference ellipsoid can support an infinite number of datums; however, each datum can only be defined by one ellipsoid.

Geodetic datums are generally classified as either a geocentric datum or a local geodetic datum. A *geocentric datum* is globally centered and adequately approximates the Earth's size and shape as a whole. The center of the reference ellipsoid coincides with the Earth's center of mass. See Figure 7.1.

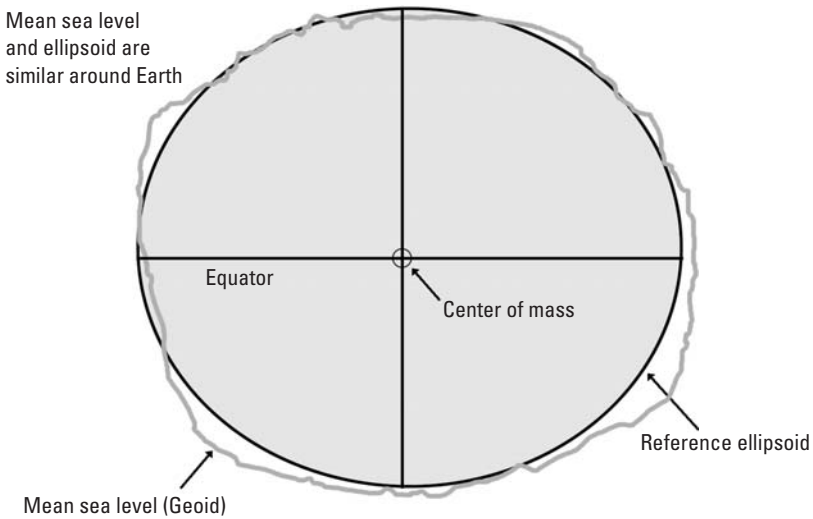


Figure 7.1 The geocentric datum. (*After:* [1], diagram 6.)

A geocentric datum is not designed to best characterize a particular area but instead is best used when applications are global in nature. GPS, for example, utilizes the common geocentric datum WGS84.

A *local geodetic datum* is used to characterize a particular region where the reference ellipsoid and the Earth's shape coincide. The center of the ellipsoid is often located away from the Earth's center of mass, thereby providing a poor representation of the Earth as a whole. Figure 7.2 illustrates how the ellipsoid's surface and Earth's surface coincide within a particular location and diverge elsewhere.

Both figures depict the Earth's mass as a geoid or mean sea level model. This is a typical convention with geodetic datums, especially when defining the height of the Earth from its core. Since we are detailing the relationship between the Earth's center of mass and the center of the reference ellipsoid, as well as the surfaces of both these shapes, the geoid proves to be a fair Earth shape parameter. We will examine the true function of the geoid in the next chapter.

Primary Types of Geodetic Datums

The classification of geodetic datums offers a degree of user control as it relates to which datum suits the application or location best. For instance, you would never choose NAD27, a North American datum, to map the city of Tokyo in Japan. The Tokyo datum has been devised specifically to handle that terrestrial location. This form of user control surrounds the act of selecting a

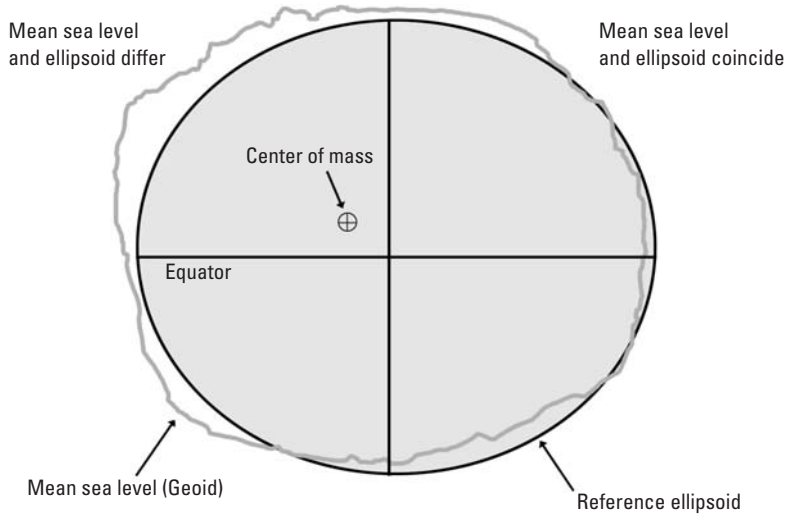


Figure 7.2 The local geodetic datum. (*After:* [1], diagram 5.)

location-specific datum based on individual characteristics, rather than manipulating a more general datum.

Geodetic datums empower the user with a referenced-based geographic coordinate system within which accurate point positions can be determined. In a two-dimensional space, point positions are defined by a coordinate pair (*x*, *y*) on a uniform plane. Think of a single point on an *x-y* grid drawn on a piece of paper. Every *x-y* positional coordinate on that paper is at a uniform height (i.e., uniform plane). In the real three-dimensional world, height (*z*) is as important a distinguisher as the *x* and *y* coordinates. The three-dimensional point positions involve a horizontal component (*x* and *y* coordinates) and a vertical component (*z* coordinate).

At any given point on the Earth's surface there exists a horizontal level that runs parallel with the surface (ground) and a vertical level that runs perpendicular to the surface (ground). Refer to Figure 7.3.

Imagine holding a plumb line or a weighted string so that it remains suspended in the air aiming toward the ground. The imaginary line running through the string and weight into the ground provides that point position's vertical level. Likewise, a glass or cylinder containing water and placed on the ground presents the horizontal level for that position. Since horizontal and vertical levels exist for every point position, positional coordinates can be adjusted as needed along these planes.

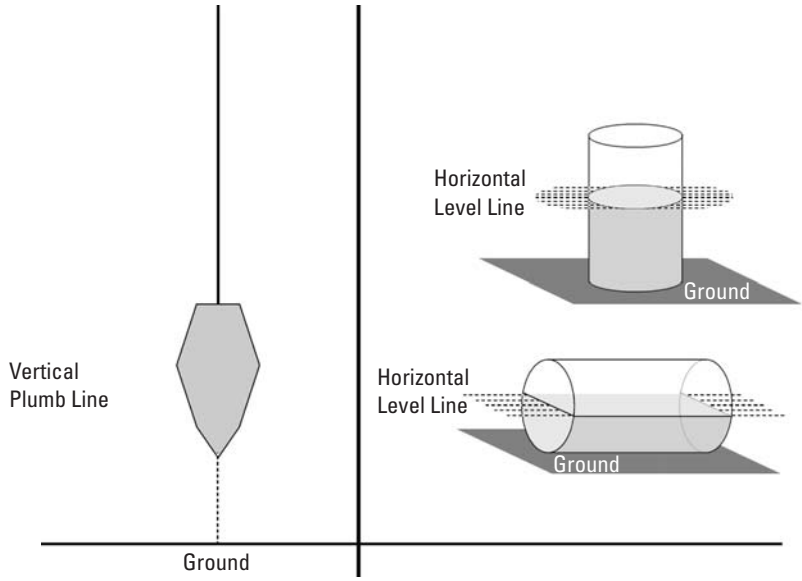


Figure 7.3 The vertical plumb line and the horizontal water line.

Geodetic datums, both local and geocentric, offer the opportunity to manage these horizontal and vertical levels for greater positional accuracy. This leads us into the two primary types of geodetic datums: horizontal control datums and vertical control datums.

A *horizontal control datum* (or *horizontal datum*) is used to manipulate and set a position in the *x* and *y* directions. Horizontal datums are often defined by a reference ellipsoid and a coordinate origin. Alternately, a *vertical control datum* (or *vertical datum*) is used to manipulate and set a positional height in the *z* direction. Vertical datums provide a level from which positional heights can be determined.

Both primary types of geodetic datums provide a favorable degree of positional control. As mentioned earlier, the remainder of this chapter will examine the standard types of horizontal datums and various methods to transform between two horizontal datums.

Standard Horizontal Datums

As we already understand, horizontal datums take on both geocentric and local geodetic forms. Each major province or country has its own standard datum or simply a datum of choice. Standard horizontal datums cover large geographic areas. They are used throughout the world for simplicity and local datum reference. We will briefly examine five commonly used horizontal datums and reference systems for the Earth model.

World Geodetic System 1984

The World Geodetic System 1984 (WGS84) is considered a global datum that defines an Earth-fixed global reference frame for the Earth. WGS84 is based upon the ellipsoidal parameters of the Geodetic Reference System of 1980. WGS84 is used for GPS satellites and is defined by a gravity model of the Earth.

WGS is a geocentric datum that made its first appearance in 1960. A product of the U.S. Defense Mapping Agency (DMA) for mapping and charting, the WGS has evolved from the primitive satellite measurements of 1960 (WGS60) to the sensitive Doppler satellite data and satellite data altimetry used in the latest and highly accurate WGS84. With its satellite-driven precision, WGS84 has proved to be a satisfactory reference datum for nearly every Earth location.

Given its Earth-fixed nature, WGS84 offers a standard reference ellipsoid with base parameters being:

Semimajor axis $a = 6{,}378{,}137$ meters
Semiminor axis $b = 6{,}356{,}752.314$ meters
Flattening $f = 0.003352811$
Inverse flattening $1/f = 298.2572236$

Due to its general accuracy worldwide, WGS84 is a highly used, worldwide datum (and reference ellipsoid). In fact, due to its global appeal, many applications require local datums to be converted to WGS84. The transformation parameters provided with the worldwide geodetic datums in Appendix B are conversion factors between the particular local datum and WGS84.

Incidentally, it is anticipated that WGS84 will remain valid until 2010, when it will be revised.

International Terrestrial Reference Frame

Another major global reference datum is the properly named International Terrestrial Reference Frame (ITRF). This reference system is used for a large selection of astronomical, geophysical, and geodetic analysis and information products worldwide, although it is typically more popular in Europe.

The goal of this geodetic datum is the realization of a terrestrial coordinate system through the creation and implementation of a terrestrial reference frame. Positions on the surface of the Earth are indeterminate observations or quantities without a terrestrial reference frame. Thus, the ITRF enables precise positional coordinates that are linked to the terrestrial landscape of Earth. This terrestrial landscape is referenced through an International Terrestrial Reference System (ITRS), which is a dynamic spatial referencing system that adheres to the Earth's rotation and geophysical changes.

The ITRF has proved to be a uniquely useful and precise tool for both the geodesist and GIS user alike. The ITRF maintains precision through use of modern geodesy, GPS, satellite laser ranging (SLR), lunar laser ranging (LLR), very long baseline interferometry (VLBI), and other analysis tools. Additionally, precision is maintained with the aid of measurement for more than 800 observation and control stations worldwide. These observation and control stations are typically located at small distances from major fault lines or areas of seismic activity. The resultant product, namely the ITRF, is a highly precise datum that simulates the present-day terrestrial Earth and captures the minute changes in the Earth that other geocentric datums cannot capture.

The ITRF's base parameters for its reference ellipsoid are identical to WGS84 down to the 10-cm level. This level of accuracy is precise enough to say that WGS84 and ITRF share the same ellipsoid parameters. In addition, ITRF

has no transformation factors to WGS84. As a result, ITRF coordinates can be expressed as WGS84 coordinates at a 10-cm level.

The first adequate version of ITRF came out in 1994. It was revised in 1996, again in 1997, and finally in 2000. At the time of this writing, the latest form is ITRF2000, although plans are set for an ITRF2005.

Geodetic Reference System of 1980

The Geodetic Reference System of 1980 (GRS80) is another widely used geodetic reference system that consists of a global reference ellipsoid and a gravity field model. The GRS80 model provides a geocentric reference ellipsoid used by the ITRF geodetic network. GRS80 supersedes GRS67. Both were developed and are maintained by the International Association of Geodesy (IAG), a division of the International Union of Geodesy and Geophysics (IUGG). The GRS80 model offers a simple yet accurate framework for the definition of a worldwide geodetic datum and the determination of the normal gravity field for the Earth.

The GRS80 ellipsoid encompasses the entire mass of the Earth, which goes beyond the terrestrial mass and includes the atmosphere. The inclusion of the Earth's atmosphere into the ellipsoid measurements was an innovation first introduced in the GRS80 model.

The GRS80 reference system is used by GPS through a standardized realization of the system: WGS84. Given this connection, GRS80 offers a standard reference ellipsoid with similar base parameters:

$$\textit{Semimajor axis } a = 6{,}378{,}137 \text{ meters}$$
$$\textit{Semiminor axis } b = 6{,}356{,}752.314 \text{ meters}$$
$$\textit{Flattening } f = 0.003352811$$
$$\textit{Inverse flattening } 1/f = 298.2572236$$

As mentioned earlier, GRS80 takes into account the Earth's atmosphere, which is conceptually a uniform, condensed layer of atmosphere on the surface of the ellipsoid. In the GRS gravity field model, the variation in density for this condensed atmospheric layer is ignored so that the geocentric gravitational field can be efficiently calculated. The GRS80 gravity field model is defined through the following physical constants:

$$\textit{Geocentric gravitational constant } GM = 3{,}986{,}005 \times 10^{8} \text{ m}^3/\text{s}^2$$
$$\textit{Dynamic form factor } J_2 = 108{,}263 \times 10^{-8}$$
$$\textit{Angular velocity of the Earth } \omega = 7{,}292{,}115 \times 10^{-11} \text{ rad/s}$$

The GRS80 geocentric model is a reference system through which other geocentric and local geodetic datums are based. Its worldwide accuracy has enabled GRS80 to remain a geodetic datum reference of choice for many regions. For instance, the government of Australia used the GRS80 geocentric ellipsoid to define GDA94, a popular Australian local datum.

North American Datums of 1927 and 1983

The North American Datums of 1927 and 1983 were established to provide accurate positional coordinates and increased horizontal control for any location within North America. By the end of 1926, the original horizontal reference system for the United States (North American Datum) was drastically outdated and imprecise, and the U.S. Coast and Geodetic Survey called for a revision to the standard datum. In 1927, the North American Datum of 1927 (NAD27) was developed to include the then newly measured meridian arcs of North America. The new datum was based on the Clarke ellipsoid of 1866. NAD27 proved to be a more accurate geodetic local datum that is still used today.

Similar to the original North American Datum, NAD27 is referenced through a base triangulation station in Meades Ranch, Kansas. This reference station is close to the geographic center of the United States, which helps NAD27 minimize differences between astronomic and geodetic latitudes and longitudes.

The development of GRS80 brought forth the many shortcomings of NAD27 and showed the need for a revised and more accurate North American Datum. The North American Datum of 1983 (NAD83) was subsequently developed through an international effort between the U.S. National Geodetic Survey, the Geodetic Survey of Canada, and the Danish Geodetic Institute.

The resulting NAD83 is a geocentric datum based upon the GRS80 Earth model that uses three-dimensional techniques to establish horizontal control. These (x, y, z) observations are facilitated through the implementation of high-precision measurement tools, namely GPS, Doppler satellite, and VLBI. In fact, precision of these measurement tools is within 2 meters.

As a geocentric datum, NAD83 uses the Earth's mean center of mass as a datum reference, whereas NAD27 uses a reference on the Earth's surface (the Meades Ranch station). Adding to NAD83's importance is its compatibility with the WGS84 (GPS) network. In fact, the differences between the two models are negligible. For the purposes of GISs, these datums are effectively equivalent.

The differences between coordinates in NAD27 and NAD83 range from 200 to 300 feet in the western United States to tens of feet in central and eastern U.S. regions. Coordinate conversion between NAD27 and NAD83 can be done using GIS software or geographic calculators or through the horizontal datum

transformation methods discussed later in this chapter. Eventually, there will be only one North American Datum, since NAD27 is gradually being superseded by NAD83.

The reference ellipsoid parameters for the North American Datum of 1927 are:

Semimajor axis $a = 6{,}378{,}206.4$ meters
Semiminor axis $b = 6{,}356{,}758.8$ meters
Flattening $f = 0.003390075$
Inverse flattening $1/f = 294.9786982$

The reference ellipsoid parameters for the North American Datum of 1983 are:

Semimajor axis $a = 6{,}378{,}137$ meters
Semiminor axis $b = 6{,}356{,}752.314$ meters
Flattening $f = 0.003352811$
Inverse flattening $1/f = 298.2572236$

European Datum 1950

During World War II, the accuracy and compatibility of latitude and longitude positions in Western Europe became an issue. Battles fought on the borders of Germany, France, Belgium, and the Netherlands proved that datum inconsistencies existed, while the mapping of these countries aired several incongruities. The common consensus was that a consistent mapping datum was needed in Western Europe. As such, the European Datum 1950 (ED50) was created after the war.

ED50 is a geodetic datum that proves to be highly accurate in Western European countries. ED50 has abolished the incongruities and incompatibility issues that were faced during the many battles of World War II, while providing a tool to connect the existing and individual country datums.

ED50 is based upon the International 1924 ellipsoid (known as the Hayford ellipsoid) and offers the following base parameters:

Semimajor axis $a = 6{,}378{,}388$ meters
Semiminor axis $b = 6{,}356{,}911.946$ meters
Flattening $f = 0.003367003$
Inverse flattening $1/f = 297$

Similar to other pre-1960 datums, ED50 is referenced through a base triangulation station. This European geodetic datum is referenced at Frauenkirche of Munich in Southern Germany, close to the center of Western Europe's geodetic networks during the Cold War. Before the implementation of GRS80 and WGS84, the ED1950 was widely used around the world. Today, it is still used by default in most of Western Europe except in countries with their own datums, such as Great Britain, Poland, Ireland, Hungary, Sweden, and Switzerland.

Horizontal Datum Transformation Models

With the multitude of horizontal geodetic datums available for any given region, the need for datum consistency is essential for correct and accurate results. Combining coordinates from different datums will produce incorrect results and can even put users in jeopardy. That being stated, often GIS users or others need to transform the coordinates from one geodetic datum to another for the sake of coordinate consistency and accuracy.

Imagine that you are trekking in the backlands of Malawi, Africa, with a region map based on the local datum (ARC1950) and realize you are moving beyond the boundaries of this map. You take out your GPS unit and quickly recognize it functions on a different datum (WGS84). To reach your destination and return safely, you will need to transform your positional coordinates from the local datum (ARC1950) into the GPS-friendly datum (namely WGS84). A datum transformation must be calculated not only to obtain your GPS coordinates but also to link the local map with the GPS unit.

This example is dramatic, but not unlikely. Geodetic datum transformations are common practice for GIS users who are working with existing coordinate data and different geodetic datums. In fact, datum transformation is the most utilized positional management function in core GISs. Geodetic network orientation is frequently mismatched between sets of geographic data and requires correction before properly coinciding in geographic space. A datum transformation provides such correction.

As you might have guessed, there are a handful of accurate transformation models from which to choose. Although many transformations can be calculated using software, the novice GIS user should understand the methods for calculating these transformations. Most applications of these methods seek to transform geographic data to the standard WGS84 from older or more obscure geodetic datums.

WGS84 serves as a common model for all geodetic networks and, therefore, acts as a common ground for many datum transformations. More specifically, many datums are first transformed into the common WGS84 and then transformed again into the desired datum. By expressing datum parameters as

"shifts" to WGS84, the need to define all geodetic networks in terms of all other geodetic networks is alleviated. The worldwide datums table in Appendix B lists an assortment of local geodetic datums from around the world with their *x*, *y*, and *z* datum shifts to WGS84.

Let us examine the two most common geodetic datum transformation formulas: the Molodensky datum transformation method and the Bursa-Wolf datum transformation method.

The Molodensky Datum Transformation Method

The Molodensky (or Molodenski) datum transformation method is by far the most efficient and simplest datum transformation method available. Although sufficiently accurate, its simplicity makes this transformation less precise than others, stemming from the fact that it does not factor in rotation or scaling between datums. The Molodensky transformation is commonly called the *three-parameter transformation method* for its minimal formula involving three datum shift parameters between the two ellipsoids. This transformation method is used when the *X*, *Y*, and *Z* axes of both datums are known to be parallel and at the same scale. Further simplifying the calculation, the Molodensky method employs geodetic coordinates for the Earth's curvilinear surface, which alleviates any rectangular coordinate system conversion calculations.

As illustrated in Figure 7.4, the three geodetic datum shift parameters are defined as

1. Delta X (ΔX), the difference in geodetic longitude in meters between the current datum (#1) and the desired datum (#2);
2. Delta Y (ΔY), the difference in geodetic latitude in meters between the current datum (#1) and the desired datum (#2);
3. Delta Z (ΔZ), the difference in geodetic height in meters between the current datum (#1) and the desired datum (#2).

The Molodensky transformation formulas utilize the three datum shifts to calculate the desired datum coordinates. The formulas are as follows:

$$X_1 = X_2 + \Delta X$$

$$Y_1 = Y_2 + \Delta Y$$

$$Z_1 = Z_2 + \Delta Z$$

S. M. Molodensky first presented his datum transformation method in his classic title, *Methods for the Study of the External Gravitational Field and Figure of*

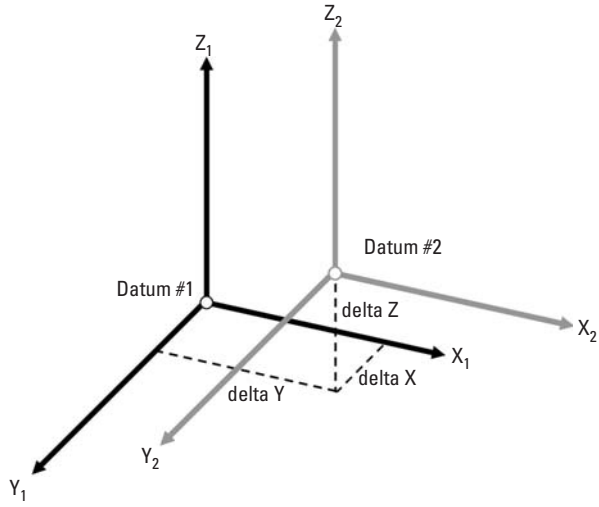


Figure 7.4 The Molodensky three-parameter datum transformation.

the Earth. His initial presentation involved a seven-parameter transformation method, which was further simplified to the current form: three parameters. In the end, the Molodensky datum transformation method retains adequate precision between the datums (within 5 meters) and proves to be a great geodetic tool for GIS.

The Bursa-Wolf Transformation Method

The Bursa-Wolf datum transformation method is a far more rigorous calculation than the three-parameter Molodensky formula. The Bursa-Wolf method is commonly called the *seven-parameter transformation method* for its lengthy formula involving three datum shift parameters, three datum rotations, and a scale correction factor. The Bursa-Wolf transformation is used when the *X*, *Y*, and *Z* axes of both datums are not parallel (axial rotation exists) and not at the same scale (scale change exists).

Similar to the Molodensky method, the Bursa-Wolf transformation method employs geodetic coordinates for the Earth's curvilinear surface, which, once again, alleviates any rectangular coordinate system conversion calculations. As illustrated in Figure 7.5, the seven datum parameters are defined as:

1. Delta X (ΔX), the difference in geodetic longitude in meters between the current datum (#1) and the desired datum (#2);

2. Delta Y (ΔY), the difference in geodetic latitude in meters between the current datum (#1) and the desired datum (#2);
3. Delta Z (ΔZ), the difference in geodetic height in meters between the current datum (#1) and the desired datum (#2);
4. X-axis rotation R_x, the rotation of the geodetic longitude in arc seconds from the current datum to the desired datum;
5. Y-axis rotation R_y, the rotation of the geodetic latitude in arc seconds from the current datum to the desired datum;
6. Z-axis rotation R_z, the rotation of the geodetic height in arc seconds from the current datum to the desired datum;
7. Scale change S_c, the scale difference between the current datum and the desired datum.

The Bursa-Wolf transformation formula utilizes all seven parameters to calculate the desired datum coordinates. The formula is as follows:

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + S_c \begin{bmatrix} 1 & -R_z & R_y \\ R_z & 1 & -R_x \\ -R_y & R_x & 1 \end{bmatrix} \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix}$$

The Bursa-Wolf transformation is a subset to the older Helmert seven-parameter transformation method. Both methods share the same formula, but they differ in their conventions for rotational polarity. The Helmert model (known also as *coordinate frame transformation* or *three-dimensional similarity transformation*) defines the rotation of the entire coordinate frame (i.e., all points on the axes). The Bursa-Wolf model (known also as a *position vector transformation*) defines the rotation of the point position with respect to a fixed coordinate frame. The conventions contrast one another and, thus, form two separate transformation methods. Compare Figure 7.5 (the Bursa-Wolf model) to Figure 7.6 (the coordinate frame transformation model).

The rotational polarity for the Bursa-Wolf model is positive clockwise around the datum 2 axes when viewed from the origin. Mathematically, the rotational polarity is reversed when viewing the origin from the axes. Whatever the convention used, the seven-parameter transformation model proves to be a much more accurate datum transformation method, with approximate accuracy held within 1 meter.

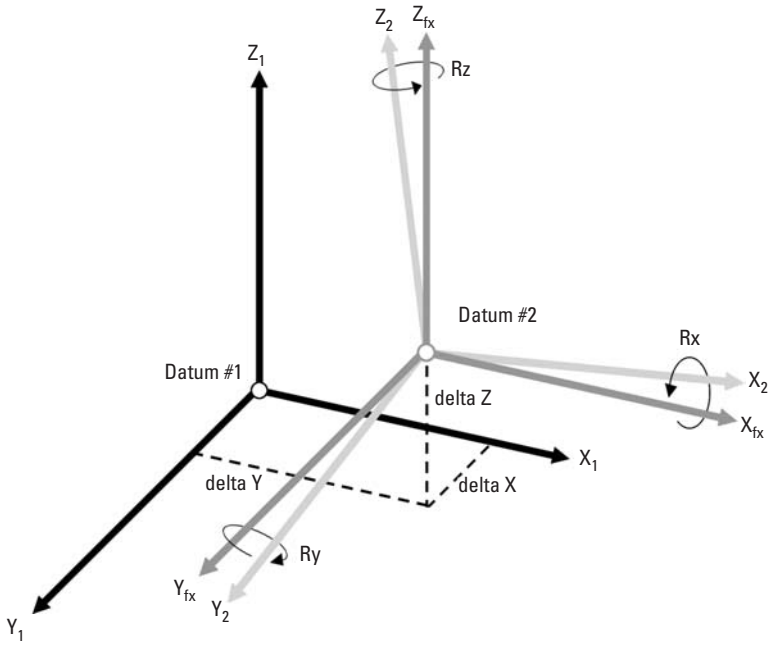


Figure 7.5 The Bursa-Wolf seven-parameter datum transformation.

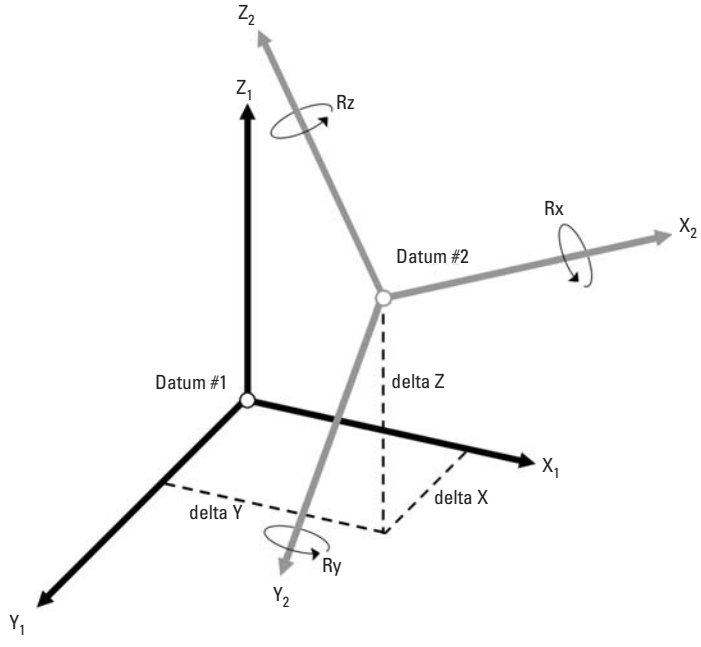


Figure 7.6 The coordinate frame transformation.

Horizontal Grid Transformation Methods

Another popular datum transformation type exists, called *horizontal grid transformation,* that enables discrete horizontal control within a datum. This transformation method is markedly different from both the Molodensky and Bursa-Wolf models. The latter models utilize a set of datum shift parameters that are attributed to a common source. Horizontal grid transformation methods use a set of field-based observations and statistical functions that describe a geodetic datum in a more comprehensive manner.

Adding to the complexity of the calculations but enhancing the accuracy of the results, horizontal grid transformations solve transformation through a mathematical function: bilinear gridded interpolation. This data interpolation technique facilitates the accurate transformation from older datum to newer, more accurate datum. Bilinear gridded interpolation is used by the U.S.-based North American Datum Conversion software and the Canadian National Transformation application. Additionally, this grid transformation technique has been adopted by the governments of Australia, New Zealand, and Great Britain for its high-accuracy horizontal control.

The standard horizontal grid transformation methods provide a greater level of horizontal datum control than that achieved using the Molodensky and Bursa-Wolf models. Adding to the better horizontal control is an enhancement in transformation accuracy, with grid transformations retaining an approximate accuracy within 0.1 to 0.3 meter. To reprise, the Bursa-Wolf accuracy was within approximately 1 meter and the Molodensky accuracy was within approximately 5 meters.

There are three very common horizontal grid transformation methods that warrant discussion: The North American Datum Conversion, the High Accuracy Reference Network, and the Canadian-Australian National Transformations. Let us examine these methods.

The North American Datum Conversion

The North American Datum Conversion (NADCON) is the U.S. federal standard for transforming latitude and longitude coordinates between NAD27 and NAD83. NADCON is a conversion utility that was first developed to facilitate transformations between the datums directly through known grid shifts, alleviating the need for datum parameters or scale changes.

A NADCON transformation is based upon a gridded set of control points that are geostatistically interpolated. NADCON grids are based on more than 150,000 of these horizontal control points. Transformations are based upon the grid shifts measured for the horizontal control point closest to the positional coordinate to be converted.

The NADCON program also provides datum correction for errors that exist in the NAD27 horizontal control datum. When NAD27 was first created, all measurements and calculations were done by hand and retained in hand-transcribed tables. These measurements were formulated in reference to a single ellipsoid surface datum point in Meades Ranch, Kansas. These outdated methods produced numerous inconsistencies and, in some areas of the country, geographic uncertainty due to imperfect measurements.

NADCON provides correction for some of these existing uncertainties and inconsistencies in NAD27 and, as such, provides a relatively efficient method to transform between both datums. With such a direct correlation between datums and focusing on whole grid shifts rather than individual axis or point shifts, the NADCON transformation method proves to be incredibly accurate, far more accurate than the previously discussed datum transformation methods.

The High Accuracy Reference Network

GPS and other present-day advances in geodetic measurement and coordinate position accuracy have allowed for the creation of a High Accuracy Reference Network (HARN). A HARN is a statewide or regional, high-accuracy geodetic network that improves the precision of older datum control networks. The ultimate goal for developing a HARN is to produce a denser geodetic network that will present validated results rather than human estimation.

More specific, a HARN is used to improve accuracy of NAD83 in specific state-based control networks. This increased accuracy has warranted the need for transformation grids to convert from the original NAD83 measurements (implemented in 1986) to current measured values. Prior to 1992, this datum conversion method was known as a High Precision Geodetic Network (HPGN) correction. The acronym HARN has now superseded HPGN.

In order to revise the original NAD83 measurements to the level of accuracy supported by GPS, which was introduced shortly after NAD83, HARNs were implemented throughout the United States. Starting in Wisconsin in 1989, HARNS were set up in every state to remedy and readjust the lower accuracy NAD83 coordinates. In the end, NAD83 was revised and the HARN measurements produced a coordinate set with greater local accuracies throughout the United States. Incidentally, HARN surveys are becoming more popular worldwide, with HARNs appearing in Europe and Central America.

Canadian and Australian National Transformations

Popular grid transformation methods have also surfaced outside the United States, with probably the most common and extensively implemented being the

Canadian Spatial Reference System's NTv2 software. NTv2, which stands for National Transformation Version 2, is a grid shift interpolation software that transforms coordinates (nationally) between NAD27 and NAD83.

NTv2 has been available since 1992, when it replaced the former version. The NTv2 software has proved to be immensely useful in Canada, as well as in other horizontal geodetic datum networks, and has the capability for accurate modeling based upon various national control survey and municipal networks. NTv2 has remained Canada's national grid transformation method.

In November 1997, Australia's Intergovernmental Committee on Surveying and Mapping (ICSM), faced with a recently revised national datum, turned to Canada's grid transformation method. The conversion utility was adopted for Australia's national conversion from the older, less accurate Australian Geodetic Datums of 1966 and 1984 (AGD66 and AGD84) to the GPS-referenced Geocentric Datum of Australia 1994 (GDA94).

Australia implements NTv2 to handle the differences between datums as a national grid shift, with various control stations surveyed throughout Australia. NTv2 is used to provide horizontal control when converting between the two datums and enables accurate positional coordinates through simple software-driven interpolation. Since AGD66 and AGD84 are equally used, Australia utilizes two grids: one to convert AGD66 to GDA94 and another to convert AGD84 to GDA94. In the end, implementation of Canada's NTv2 method will enable Australia to have greater local accuracy nationwide, purported to be within 10 cm.

Reference

[1] Jones, A., and G. Blick, *Where in the World Are We?*, ver. 1, Wellington New Zealand: Government of South Australia, 1999.

8

The Vertical Datum

Just as a horizontal control datum requires a reference grid on which all coordinates abide, a *vertical control datum* (or *vertical datum*) demands a contiguous, consistent reference from which topographical elevation and bathymetric depth are measured. Perhaps considered a complex geodetic construct, the vertical datum is used to manipulate and define a positional height as it relates to a determined reference. How can one accurately measure heights when the Earth's surface ranges in elevation from the peak of Mount Everest to the underwater depths of the Pacific Ocean's Marianas Trench? What could constitute a valid reference for precise vertical measurements?

These and other similar questions have plagued geodesists, mathematicians, and mariners alike. The geodesists had standard ellipsoids that proved to be a general, though innately imprecise, height reference. Mathematicians had their mathematical model of the Earth that was, for height, equally imprecise. Mariners had the tides as reference and frequently, due to the ever moving and inconsistent height of the seas, found themselves stuck in sandbars or crashed upon the rocks. The early need for a consistent and accurate height reference was desperate.

Today, vertical datums provide just such a reference from which positional heights on the Earth's topographical surface can be determined with reasonable accuracy. There are two primary reference surfaces used by leading vertical datums: sea levels (tidal) and standard ellipsoids (geodetic). References for sea level–based datums vary from high and low tide levels to an average sea level. Ellipsoid-based datums generally reference the same ellipsoid model used for calculating horizontal datums. By function, GIS utilizes these same reference surfaces when defining general and specific point heights.

Mean Sea Level and the Geoid

In many day-to-day uses, elevations are referred to as height above or below sea level. For example, in the case of the highest and lowest points on Earth, the peak of Mount Everest is 9 km above sea level and the base of the Marianas Trench is approximately 11 km below sea level. The Earth's topology varies 20 km, highest point to lowest point. In the grand scheme of ellipsoids, this variance is nearly negligible (remember the dimples on the golf ball from the ellipsoid chapter). However, when referenced to sea level, the elevation swing is perceivable and affecting. We can easily conclude that height measurements referenced to the ellipsoid are less accurate than when referenced to the sea level.

Additionally, since tides are constantly changing, rising and falling, the term "sea level" is considered the average of the tide levels. In terms of datums, we call this average sea level the *mean sea level* (MSL). Figure 8.1 illustrates this relationship.

As represented in the figure, the height measurement is the distance from the control station or geodetic point to the MSL. This is generally called the *mean sea level elevation.* Known formally as the *orthometric height,* the measurement is the height above the MSL and, depending upon position, can be represented as a positive (above MSL) or negative (below MSL) measurement.

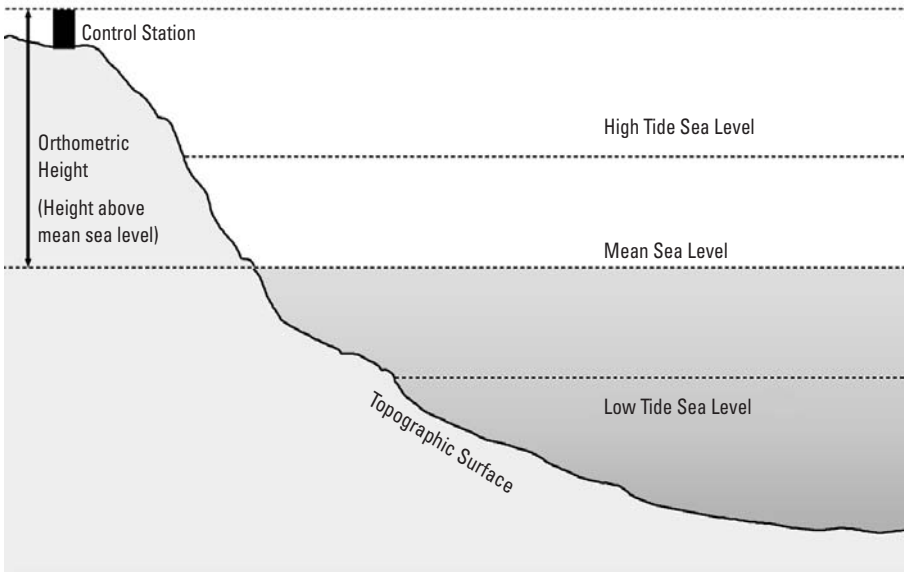


Figure 8.1 The relationship between sea level and height measurements.

As we know, tides are highly influenced by the Earth's gravity field. The Earth's gravity field is defined by two distinct components: the normal mathematical gravimetric earth model and the deviations and disturbances (known as anomalies) from this normal model. In the middle of the ocean, undisturbed by land, the MSL and the Earth's gravitational field coincide. On land, the terrestrial mass and density affect the gravity field and, thus, cause a very small variation between MSL and the Earth's gravity field. Nearly everywhere on Earth, this small variation can be ignored without diminishing the accuracy of height measurements.

Now, imagine if the MSL were to cut through the continents and create a level surface around the Earth. The result would be an exact model of the Earth defined by the MSL that would double as an adequate depiction of the Earth's gravity field. This model is known as the *geoid*.

The geoid is an equipotential (or level) surface of the Earth's gravity field, which coincides with the MSL. For all reasonable purposes, the geoid is defined by the MSL and is considered a good reference by which elevations or heights can be measured.

The geoid is not a smooth surface. The gravitational pull of the Earth is stronger in topology rich in iron and other dense materials, and, as such, causes the geoid to take on a less refined shape. Figure 8.2 depicts how the geoid looks in reference to the standard ellipsoid. Notice its irregular shape, resembling a potato.

The shape differences between the geoid and the standard ellipsoid illustrate just how dissimilar these vertical datum reference surfaces are. To understand the relationship between these reference surfaces, we must examine the separations between the accurate but complex geoid (MSL) and the general but

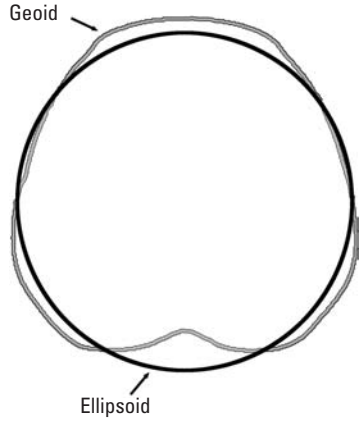


Figure 8.2 The geoid's surface.

simple ellipsoid. These surface separations are called *geoid-ellipsoid separations* or *undulations.*

The Geoid-Ellipsoid Separation

The geoid-ellipsoid separation is an important factor in defining the vertical position as it relates to the vertical datum. Since GPS and certain datums reference the ellipsoid, while other vertical datums reference the geoid, a firm understanding of their relationship is needed to link datums and determine precise vertical positioning between systems. Figure 8.3 details the specific relationship between the geoidal, ellipsoidal, and topographical surfaces

The components to vertical positioning and interrelation between datums can be determined mathematically using the following formula:

$$h = N + H$$

The geodetic height h (also called the *ellipsoid height*) is defined as the height above the ellipsoid to the topographic surface. The orthometric height H is defined as the height above the MSL. The geoid separation N is the distance between the ellipsoid and the geoid. Geoid separation (undulation) values are positive when the geoid is above the ellipsoid and negative when the geoid dips below the ellipsoid. Incidentally, throughout a large portion of the Earth, the geoid is above the ellipsoid (N is positive). In the United States, however, the geoid is below the ellipsoid and N is negative.

The compatibility between various vertical datums is paramount to controlling and manipulating the vertical component in a point position. Understanding geoid undulation enhances datum interoperability and vertical datum

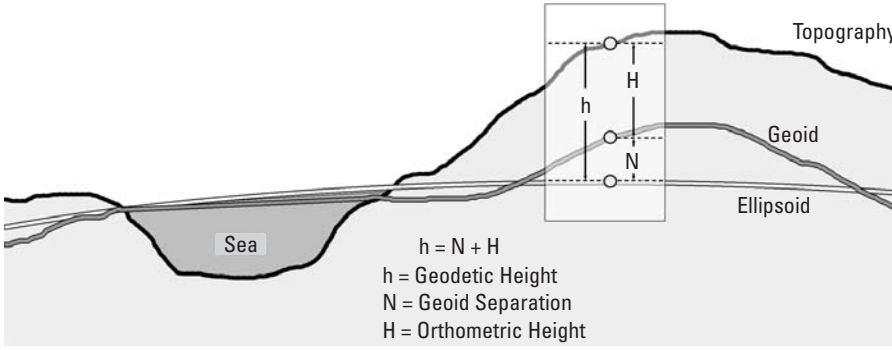


Figure 8.3 The geoid-ellipsoid-topography relationship.

transformations. Let us now examine some standard vertical datums and common transformation methods.

Standard Vertical Datums

The heights or depths of terrestrial points are determined through referencing the defined elevation surface of a vertical datum. Like the horizontal datum, the need for positional accuracy has prompted the development of standardized vertical datums. Once again, each major province or country has its own national standard datum of choice or design. They are used throughout the world for elevation-related reference and precision along the z axis. We will overview four commonly used vertical control datums for the Earth model, namely the Earth Gravitational Model of 1996, the GEOID03 reference family, the National Geodetic Vertical Datum of 1929, and the North American Vertical Datum of 1988.

Earth Gravitational Model of 1996

The Earth Gravitational Model of 1996 (EGM96) is the accurate collaborative product of the National Imagery and Mapping Agency (NIMA), the NASA Goddard Space Flight Center, and the Ohio State University. EGM96 is a geopotential model of the Earth, which means that orthometric height (height above MSL) is adjusted using the differential of gravity with latitude and elevation. Therefore, EGM96 offers heights that are influenced and adjusted by the Earth's gravity.

When working with models accounting the Earth's gravitation field, spherical harmonics are often used to approximate the shape and size of the geoid. A spherical harmonic is a single-valued, continuous, and complex calculation involving longitudinal and latitudinal angular coordinates. EGM96 utilizes a complete set (360 degrees) of spherical harmonic coefficients that offers accurate details of the entire geoid to regions as small as 55 km.

As a complex global reference surface, EGM96 details the variations between the MSL and the ellipsoid, discussed earlier as geoid-ellipsoid undulations (separations), and characterizes how uneven the Earth's gravitational potential is for the MSL.

Information for the development of EGM96 came from a collection of detailed surface gravity data, as well as direct satellite altimetry data, altimetry-derived gravity anomalies, and multisystem tracking of more than 20 satellites. This dynamic level of precision has earned EGM96 worldwide appeal and implementation. In fact, when EGM96 was first released, it was used as a geodetic reference to update WGS84.

As a well-defined global vertical datum, EGM96 is used as a modern reference for numerous applications worldwide, such as bathymetrical and geophysical studies.

GEOID03 and Predecessors

The U.S. Hybrid Geoid Model for 2003 (GEOID03) is a global vertical datum that calculates geodetic heights (height above the ellipsoid to the topographic surface) above the GRS80 ellipsoid. Developed by the National Geodetic Survey, GEOID03 is the latest revision to the U.S. GEOID model, superseding predecessors GEOID99, GEOID96, GEOID93, and GEOID90. With each revision came additional refinement and enhanced precision to the model. Although greatly used in the United States, GEOID03 is an excellent gravimetric geoid worldwide.

The GEOID03 model is considered a "hybrid" geoid, because it was developed using both a very precise gravimetric geoid (U.S. Gravimetric Geoid Model of 2003, or USGG2003) and updated NAD83 GPS ellipsoidal heights on leveled benchmarks (GPSBM2003). Incidentally, through this background infrastructure, GEOID03 is a derivative of EGM96.

GEOID03 remains the leading geoid model in the United States and the foundation for national models elsewhere in the world. The GEOID03 model is accurate to within 1.0 cm throughout the conterminous United States (CONUS). Since the geoid model is calculated using GRS80, it holds positional accuracy equivalent to GPS. In addition, conversion from NAD83 ellipsoidal heights to the North American standard vertical datum, NAVD88 (discussed later), are accurate to within a few centimeters.

The model's revision in 2003 is less obvious with respect to its predecessor GEOID99. Both are extremely similar, having backing from the national gravimetric geoid and NAD83 GPS benchmarks. The distinction is the countless, more explicit local definitions of orthometric heights throughout CONUS. Additionally, GEOID03 was modified in 2005 to reflect the orthometric height changes in the coastal Louisiana region resulting from a series of damaging hurricanes.

National Geodetic Vertical Datum of 1929

The National Geodetic Vertical Datum of 1929, known as NGVD29, is a geodetic reference for elevations based on a mean sea level datum. Formally known as the Sea Level Datum of 1929, NGVD29 has proved adequate for all areas in North America. Although it is an older vertical control datum that is slowly being superseded, it is still very much used throughout the United States and Canada. Many states and commonwealths still use maps based on NGVD29.

The main obstacle with using NGVD29 is that it is somewhat flawed. NGVD29 and its predecessor erroneously set a fixed elevation (zero elevation) by the MSL at 26 tide gauges in North America (21 in the United States and 5 in Canada). Since the Sea Level Datum of 1929 was first calculated more than eight decades ago, the MSL and the sea topological surface of the Earth have changed.

By May 1973, this gradual change in sea level was apparent and, technically, the implication of the old datum name was incorrect. The datum was no longer viewed as a reference to the MSL or geoid. This prompted the National Geologic Survey to change the name to the current National Geodetic Vertical Datum of 1929. Only the name changed; the outdated measurements remained unrevised. Needless to say, elevations measured using NGVD29 and Sea Level Datum of 1929 are different at different times.

During its lifetime, NGVD29 has grown flawed and has lost its usefulness in relation to positional accuracy. Geodesists can now measure these elevation differences more precisely with the latest advances in technology and a sophisticated geodetic network. Many U.S. state agencies and Canadian provinces have either begun or completed the chore of switching from NGVD29 to the more accurate North American Vertical Datum of 1988. In years to come, NGVD29 may be used less and eventually become obsolete.

North American Vertical Datum of 1988

The North American Vertical Datum of 1988 (NAVD88) is a vertical control datum first established in 1991. NAVD88 is considered a modern control datum that institutes what is called a minimum-constraint adjustment of leveling observations in Canada, the United States, and Mexico. NAVD88's minimum-constraint adjustment is a nominal modification to elevations based upon one tidal benchmark at Father Point/Rimouski in Quebec, Canada. The result is a very precise vertical control datum that is responsive with respect to geologic changes and is no longer tied to the time-altering MSL.

As mentioned previously, NAVD88 is superseding the older, less precise NGVD29. This process is not as easily implemented as one would expect given that each location in North America has a different conversion factor between the two datums. U.S. states and Canadian provinces have begun redrawing regional and height-sensitive maps (such as flood maps) with the new vertical control datum. Many regions provide area-specific conversions for use with the existing outdated maps. For instance, Wilkes County in North Carolina uses a datum offset of –0.48 feet and therefore calculates the NAVD88 elevation as the NGVD29 elevation – 0.48 feet.

Eventually, NAVD88 may be the only North American standard vertical datum used for vertical control and height accuracy. As we will discuss in the

next section, transformation software has been developed to help transform height data between NGVD29 and NAVD88.

Vertical Datum Transformation Methods

As with horizontal datums, there is a functional need to transform between different vertical control datums. As we know, heights are dependent upon the reference point (or zero elevation). There are many instances when height (elevation) transformation is warranted. You might want to reference a horizontal datum height to a gravimetric geoid or you might need to transform a height from an older vertical datum to a newer, GPS-precise datum.

Many horizontal datums provide a height quantity or distance along the z axis. To transform heights from one ellipsoid-based datum to a second ellipsoid-based datum, the Molodensky and Bursa-Wolf transformations continue to prove adequate. Both transformation methods were discussed in the previous chapter (horizontal datums), but allow for a position point's z-axis (height) transformation. The level of accuracy is based upon the ellipsoid origin and z-axis datum shift. This type of vertical position transformation does not account for the Earth's gravity field or the geoid.

A transformation from a horizontal datum height to a vertical datum height is sometimes required. These transformations involve referencing a gravimetric geoid, like GEOID03, between the horizontal and vertical control datum heights. For instance, transforming vertical positions from NAD83 to and from NAVD88 requires the typical height transformation with an implementation of GEOID96 or later. The transformation converts the position and the geoid model factors in the gravity field's potential.

The last form of vertical height conversion is a true vertical datum transformation—transformation from one vertical datum to another. The general impetus for implementing such a conversion is to transform older datum positions to newer, more precise positions.

The most common implementation of this transformation is between NGVD29 and NAVD88. To convert easily between the two vertical datums, the National Geodetic Survey devised VERTCON, a vertical model transformation method that computes the difference in orthometric height between NGVD29 and NAVD88. VERTCON was implemented to help state and county agencies revise their maps and charts from the older datum to the newer, GPS-based datum. Both datums utilize the same coordinate grid; therefore, VERTCON is considered a vertical grid transformation model.

The VERTCON model calculates the z value at any point in CONUS through the following calculation: z = NAVD88 − NGVD29. VERTCON highlights, through this simple mathematical equation, the varying inaccuracies

in the older NGVD29 datum. The VERTCON method utilizes a preprogrammed grid on which a transformation is calculated at every point on the ellipsoid. The resultant *z* values serve as a reference for every location in CONUS.

If you have an NGVD29 height to convert, the VERTCON model *z* value for that location should be added to the NGVD29 height (i.e., NAVD88 = NGVD29 + *z*). Alternately, if you have an NAVD88 height to convert, the VERTCON *z* value for that location should be subtracted from the NAVD88 height (i.e., NGVD29 = NAVD88 – *z*).

The VERTCON method is a common, much-used vertical datum transformation model that has warranted its own software. Whatever method you employ for height transformation, the end product will most likely be a more defined height position that not only will conform to modern accuracy standards but also will depict the precise orthometric heights in reference to a common Earth model.

9

The Map Projection

When was the last time you went into a gas station and purchased a projection of the Earth? Believe it or not, it was the last time you bought a road map. Atlases, world maps, and road maps are all projections of the Earth under well defined criteria.

The idea of depicting the Earth on a flat map is easy to imagine and rather simple to understand, but the actual process of a map projection is complex. Maps used in everyday life, as well as maps used in GISs, are created with a specific application in mind.

Mapmakers, or *cartographers*, use map projections to portray the real world with minimal aberrations. This chapter specifies these tools in detail and defines projection techniques and understandings that are not readily discernible from an everyday road map.

The Physicality of the Earth

Before discussing the *what* and *why* of a map projection, let us first recap a fundamental element of Earth science: the ellipsoid. The Earth's ellipsoid rotates upon a single, standard axis, making one full revolution every *sidereal day* (or 23 hours, 56 minutes, and 4.09 seconds). The Earth's overall shape is commonly defined and modeled as an *ellipsoid of revolution.*

More specifically, this axial rotation causes every location on Earth to experience sunlight and darkness, although within the artic and antarctic circles, the axial rotation creates increased lengths of sunlight and darkness. The Sun faces only a portion of the Earth at any given moment. This illuminated portion, over time, rotates beyond the range of sunlight and enters darkness. Almost

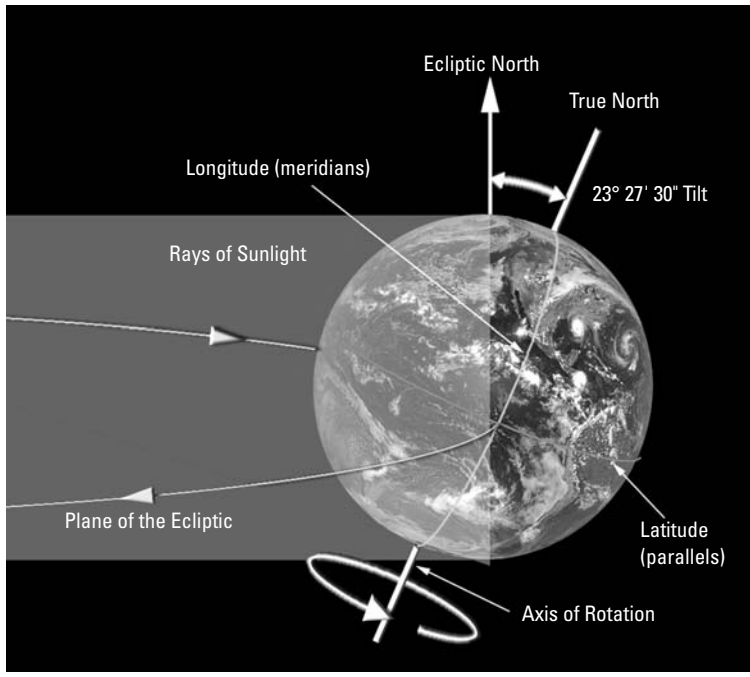


Figure 9.1 The physical Earth.

all locations on Earth experience sunlight and darkness in one day, but areas close to the poles (within the arctic and antarctic circles) experience longer daylight and nocturnal phases. This variance is a result of the Earth's tilt, which extends or diminishes a location's exposure to the Sun (see Figure 9.1).

The standard axis is perpendicular to the Earth's surface and intercepts the ellipsoid's surface through two distinct points known as the Earth's poles: the north pole in the arctic and the south pole in the antarctic. Figure 9.1 indicates this "axis of rotation."

The Earth revolves around the Sun with a single revolution taking 365.2564 days (known as a *sidereal year*) and defining a plane in space called the *plane of the ecliptic*.

The Earth's standard axis runs north and south and is tilted 23 degrees 27 minutes 30 seconds from the plane of the ecliptic. This tilt is the reason we experience the four seasons of the year. For example, in the summer, which starts with the summer solstice (June 21), the northern hemisphere is tilted toward the Sun, causing a rise in temperature. During the winter, which starts with the winter solstice (December 21), the northern hemisphere is tilted away from the Sun, causing a fall in temperature.

The Earth's ellipsoid is partitioned by a network of horizontal (east/west) and vertical (north/south) lines. The horizontal lines are called *parallels,* because they run parallel to one another, while the vertical lines are called *meridians*. The Earth has a prime meridian, which runs through Greenwich, England, and a central parallel called the equator. The equator separates the northern hemisphere from the southern hemisphere. In terms of the Earth's ellipsoid, each location along a parallel has the same *latitude*, or distance from the equator. Similarly, each location along a meridian has the same *longitude,* or angle between the target meridian and the prime meridian (0°). Although the meridians converge at the two poles, the parallels never converge.

The Earth is a complex three-dimensional object with physical dimensions, including height, width, depth, mass, and density. An important physical characteristic is its *curvilinear surface*, which is a surface consisting of or bounded by curved lines. The Earth's nonstraight, curvilinear surface is the sole reason why a latitude and longitude system is necessary. Figure 9.2 depicts the Earth model with its geographic coordinate system.

This curvilinearity also adds complexity to a map projection's transformation of the Earth model or *globe* to a flat surface. In fact, it is the primary reason why map projections are only cultured depictions of the globe and not exact representations. Due to the Earth's curvilinear surface, map projections always hold a degree of distortion.

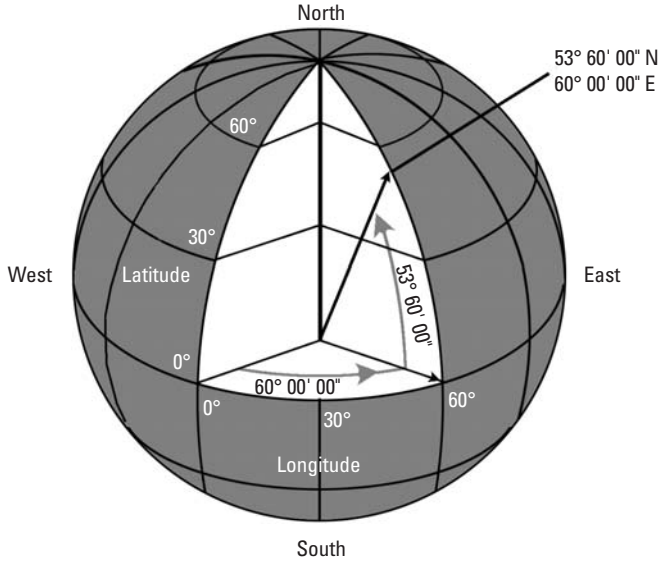


Figure 9.2 Earth model geographic coordinate system.

Scale Reduction: From the Ellipsoid to the Globe

Up until now we have discussed the Earth's ellipsoid in relation to its physicality and true size, but for the purposes of geospatial analysis the Earth's true dimensions are unwieldy. They are simply too large to realistically handle, manipulate, and compute. All GIS applications ultimately use an interpretation of the Earth at a reduced scale. For this reason, before a map projection can be created, the dimensions of the Earth must be reduced.

The process of proportionately reducing actual dimensions into much smaller, more convenient dimensions is called *scale reduction*.

Scale reduction depicts the projecting ellipsoid as a drastically scaled-down model, much like a tabletop globe of the Earth. The Earth's globe offers the user an easily accessible object, capable of depicting a range of geographic data at flexible vantages. Relative to GIS applications, the Earth's globe is typically used in the creation of map projections, since the ellipsoid proves unmanageable.

Let us look at the use of globes in map projections in a more fundamental way: Think of the Earth's globe as being made of transparent glass with a light emanating from its center in a darkened room. The images appearing on the walls in this room are projections of the globe, similar to the shadows around a lampshade as its light bulb burns bright. This Earth model is easily manipulated and projected and properly characterizes the importance of and need for scale reduction.

A scaled-down Earth is absolutely necessary for any GIS application. For instance, a road map printed on a 50-foot mainsail will do you very little good in the car at the stoplight. A letter-size map of the United States will prove futile in assisting you in finding the downtown of a major city from the suburbs. In summary, scale reduction must be used in the creation of map projections to transform the bulky ellipsoid of revolution into the more manageable and easily projected globe.

Transformation from the Globe

The basic purpose of a map projection is to transform the Earth's globe into a useable format for a specific application. A projection systematically transforms the geographic latitudes and longitudes on the surface of the ellipsoid into locations on a plane. Locations in three-dimensional space are arranged to correspond to a two-dimensional representation. Think of the Earth's globe in relation to a flat map of the world.

Projections that perform these systematic transformations from the globe are classified as *geometric*. Geometric projections are subclassified and defined according to the type of object the projection uses.

The geometric map projection methodically transforms all geographic coordinates on a globe into corresponding plane coordinates of a map space. Figure 9.3 illustrates this transformation. A basic projection formula can be expressed as

$$X = Function\,1\,(Latitude, Longitude)$$

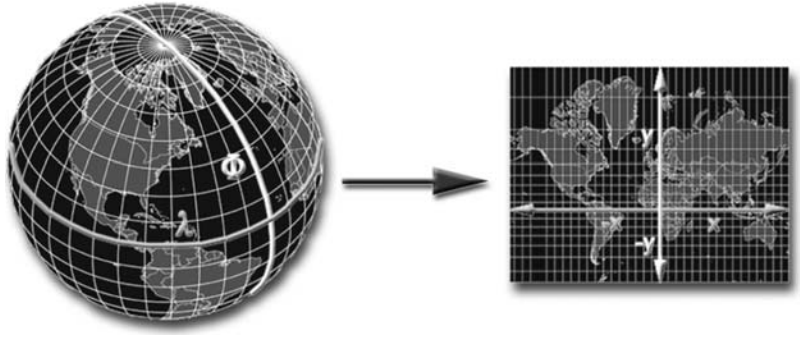
$$Y = Function\,2\,(Latitude, Longitude)$$

X and Y positions on the projected surface are functions of latitude and longitude, commonly represented by the Greek letters lambda and phi (λ and Φ), respectively. These functions must be unique to ensure that a particular point will appear at only one position on the map. They must be finite, or within a limit, so that a particular point will not appear at infinity. They must also be continuous, ensuring that although stretching or shrinking may occur, no gaps are introduced.

The geometric projection uses this formula and these projection methods to initiate a transformation, transferring latitudes and longitudes to locations on a surrounding surface.

What Is a Map Projection?

A *map projection* portrays a three-dimensional object, such as the Earth's globe, in a two-dimensional format. The map projection is quite simply the most intriguing component of coordinate system referencing because it offers a high



Geographic Coordinate System Plane Coordinate System

Figure 9.3 Moving from the Earth's geographic coordinate system to the projection's plane coordinate system.

level of flexibility. Projections are wholly graphical, as evident in the highly used Goode Homolosine map projection of the world (Figure 9.4).

The map projection employs projection formulas that perform the critical task of transferring a three-dimensional spheroid onto a two-dimensional plane surface. This is a complex process because the Earth, like other celestial bodies, is a complex object. While the projection is the easiest coordinate system component to visualize, the projection application is the most difficult. As you will quickly discover, an understanding of the projection application is vital to the success and usability of the transformation.

Map projections, by default, are not true or accurate portrayals of the globe. A two-dimensional plane cannot accurately represent large portions of the rounded, curvilinear surface of the Earth. Figures 9.5 and 9.6 illustrate the Mercator model and projection. Notice the shapes of North America and how different they are from in each other.

To show regions of the Earth on any appreciable area with accuracy, geographic data must be drawn to compromise the distortions of shapes, distances, and directions introduced by the spheroid. The various methods of preparing a two-dimensional plane from the surface of the Earth are critical for the accessibility and presentation of GISs.

The map projection is an important element in GIS. As discussed in the previous two chapters, we can comfortably surmise that datum superimposed upon the surface of the Earth's globe, or *ellipsoid of revolution*, establishes vertical and horizontal control for the specific area. The projection transforms this specific area from the curvilinear surface of the ellipse to a flat plane upon which an image is projected. This projected area is then implemented within a GIS.

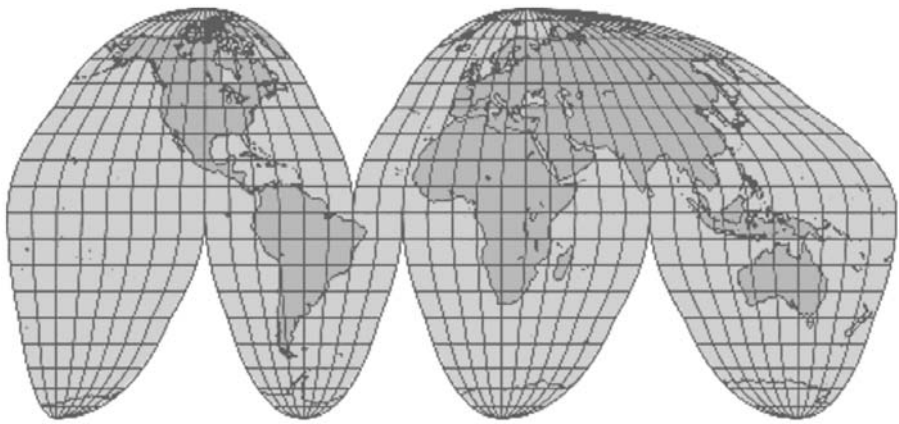


Figure 9.4 A Goode Homolosine 10° equal-area map projection of the Earth.

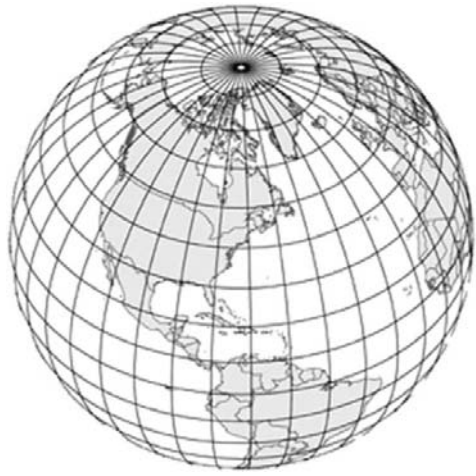


Figure 9.5 Geographic Mercator model.

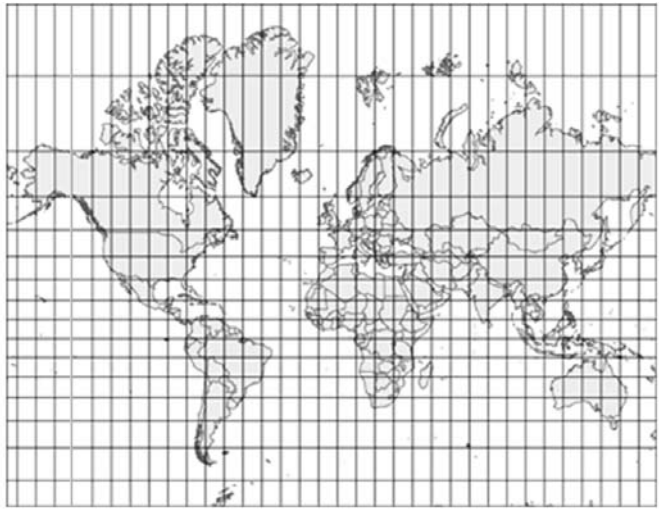


Figure 9.6 Planar Mercator projection.

The Grapefruit Peel Experiment

Let's look at the map projection a different way. Take an ordinary grapefruit and observe its spherical form. This tasty spheroid is a three-dimensional object that is similar, in form, to the Earth's globe. We will make a physical projection of one half of the fruit to help visualize a GIS map projection.

Using a flexible measuring tape and marker, draw a two-inch vertical line on the peel close to its "pole." With a knife, cut the peel perpendicular to the drawn line in one straight cut all around the grapefruit. The drawn line should be away from the peel cut. You can easily imagine each half of the cut grapefruit as a hemisphere. Now, carefully lift off the peel hemisphere (the one with the drawn line) and place it on a flat table so that the widest area is face down on the table. Notice that this object is still three dimensional and has retained the approximate curvature of the grapefruit.

Using the knife, cut the grapefruit's skin from the pole to the edge, keeping away from the drawn line. This step helps control where the skin will rip. With your palm, push down on the peel until it flattens. Creases and tears will form in various locations to compensate for the flattening of the curvature and to conform to the table surface. Once completely flat and neglecting the peel's thickness, the object has theoretically become a two-dimensional representation of the grapefruit hemisphere. Now let's examine this physical projection.

The flattened grapefruit skin does not look like a spheroid anymore and, with the peel's control cut and tears, has become distorted. Take that measuring tape and remeasure that two-inch drawn line. You will notice it is slightly less than two inches now, or that it is torn. The flattening process distorted the line's length, which is a measurement of distance, one of the four major map distortions discussed latter in this chapter.

You have now made a physical (and conic type) projection of the grapefruit. A GIS map projection functions in much the same way as the grapefruit peel experiment, though not quite as crude. When projecting a three-dimensional object onto a two-dimensional plane, various distortions and aberrations occur to functionally portray the object. Maps of the Earth's globe do not precisely depict the real shape of each continent but project the adjusted flat version of the curved landforms. Unlike the grapefruit, tears, aberrations, and distortions can be easily manipulated through projection equations.

Choosing a Map Projection

One of the more difficult parts of the map projection process is the selection of the best projection type for the application. Not only must you fully understand the object being transformed onto the flat surface, but you must also understand the desired properties you want to exhibit.

When choosing a projection, the purpose of the application must first be recognized. Then, once the purpose is understood, a strategic plan can be developed to determine those features that need to be preserved and those features that can be compromised by relative distortion. Even variations of the same projection can be selected for more accurate depictions in specific applications, such as the two Bonne projections in Figure 9.7.

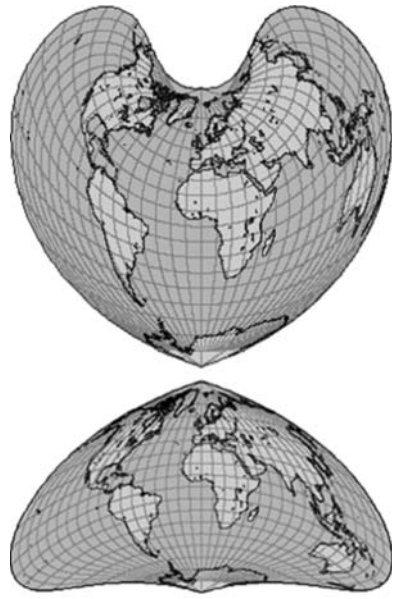


Figure 9.7 Two variations of the Bonne projection, a large-scale topographic map projection: central parallel 45° north (top) and 15° south (bottom).

The projection process is, by far, not a natural rearrangement of aspects from the three-dimensional surface. It is a mathematical condition superimposed on a natural surface to accommodate a man-made interpretation. As a result, the projected plane always has some degree of inherent distortion.

There are four important properties of the Earth model that can be tactically preserved during a projection process: *area*, *shape*, *distance,* and *angles*. These four elements are the essential pawns when projecting a surface. You must come to terms with the hard truth involved in portraying the spherical shape of the Earth on a nonspherical surface: Preserving accurate representations of all four elements simultaneously is impossible.

Important, even vital, characteristics for any projection must be identified and specifically defined early in the process. The proper type and classification of projection capable of preserving these characteristics should be tested, selected, and then implemented. If time is taken to accurately plan a projection, control of the projection will be retained and the various ill side effects from the process will be minimized.

Choosing a map projection involves more than planning; it involves decisions and an understanding of how best to represent the spheroid. Aside from manipulating the four projection elements, there are other choices and degrees

of flexibility available when creating the best map projection for a particular application. The ability to choose the projection's *type*, *aspect,* and *classification* is the ultimate tool for achieving the desired projection properties and for controlling distortion. The following discussions categorically detail each decision element.

Types of Map Projections

As tools in the projection toolbox, there are three distinct types of projections that can be utilized on the projected plane: *cylindric, conic,* and *planar*. Commonly known as *developable projection surfaces*, these types are selected on the basis of characteristic curves. The projected plane inherits the display characteristics associated with each projection type and forms what is called a *graticule*. A graticule is basically a grid of meridians and parallels.

Each projection type implements a particular graticule as a base projection surface. During a projection, a three-dimensional object is transformed onto the projection type graticule and takes advantage of the built-in characteristics. Let us discuss each major projection type further.

The Cylindric Type

The *cylindric* projection type takes on the appearance of a rectangular graph with an *x* (parallel) and a *y* (meridian) axis (see Figure 9.8). The globe's longitudal meridians and latitudes are represented by equidistant, parallel straight lines that intersect one another at right angles. The cylindric type projection is a strict grid representation of the curvilinear surface that is true at the equator and increases in distortion toward the poles.

The Mercator projection is a good example of the cylindric type. The graticule consists of equally spaced meridians but unequally spaced parallels. As the parallels get closer to the poles, the spacing between each becomes wider. With this increase in width, there is an increase in distortion. Therefore, the distortion increases as the projection moves toward the poles. This projection type also safeguards the integrity of features at the central parallel (such as the equator in a world projection) and the regions in proximity.

The Conic Type

The *conic* projection type is fan shaped, characterized by an upside down cone over the sphere. The meridians are represented by a system of equally inclined concurrent straight lines, and the parallels are represented by two concentric circular arcs. The angle between any two meridians on the projection is less than their true difference in longitude on the sphere. Refer to Figure 9.9.

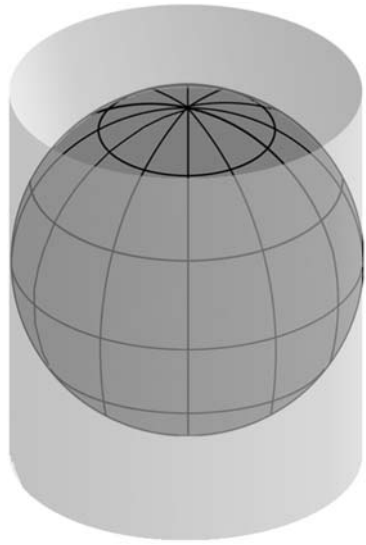


Figure 9.8 The cylindric projection.

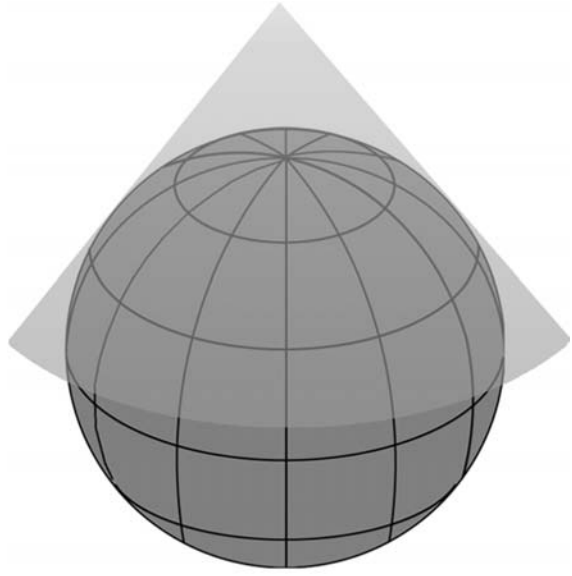


Figure 9.9 The conic projection.

Conic classifications are at an exact scale along a particular parallel, or standard parallel, between the equator and a pole. Distortion increases away

from the standard parallel. According to famed map projectionist John P. Snyder, a typical conic projection will require the definition of two true-scale standard parallels demarking an area of interest.

GIS conic projections will match between geographic data sets only if the standard parallels are the same for every geographic data layer. Current popular conic type projections are the Lambert conformal conic, also known as the American polyconic, used in state plane systems, and Albers equal area, used in small-scale mapping.

Planar Type

The *planar* projection type is an azimuthal orientation of a projected surface, which means that it pertains to the angle (usually in degrees) of an object around the horizon that is measured from north to east. The planar type protects the integrity of azimuths, bearings, and directions from a central point to other more remote points in the plane. Therefore, planar projections are true only at their center point.

The planar graticule represents meridians as straight lines that incline toward each other at their true longitudal difference. Parallels are represented by a system of concentric circles with their common center at the point of converging meridians or pole. The distortions in the planar type tend to be most prevalent along the edges of the projected plane surface. Figure 9.10 illustrates a simple planar projection scheme.

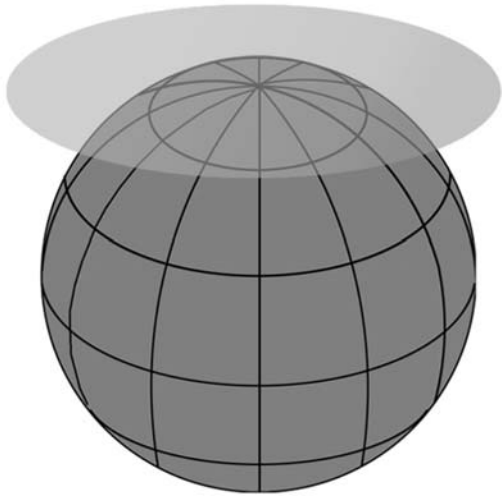


Figure 9.10 The planar projection.

The Aspect of Projection

Projection *aspect* is the relative orientation of the projected plane and the ellipsoid of revolution with respect to the location of an observer. Aspect is a natural projection property. Common with every projected surface is a standard aspect—a point or line of bearing and magnitude on which any given projection is centered.

Take another grapefruit and draw a large dot to represent a pole on opposite ends. First hold it out straight, observing its side with the poles to the north and south. Now twist it around and observe different views of the spheroid. You have the ability to observe the grapefruit from any side, from either pole, at any angle, and somewhere in between. This is exactly what a projection aspect is like; you can view the map projection from various viewpoints by manipulating the angle.

Bringing this projection tool into perspective, there are four predominant types of projection aspects: *normal*, *transverse*, *oblique,* and *polar*.

The *normal aspect* (Figure 9.11) is based upon a standard line along which distortion is minimized. The farther you move away from the standard line, the more the distortion. If this standard line is coincident with the equator, or central parallel, its aspect is considered to be *equatorial*. A normal aspect is considered *azimuthal* if the standard line connects to one of the poles while the axis of

Figure 9.11 Normal aspect.

the projection (conic or cylindric) is coincident to the axis of the ellipsoid of revolution.

An aspect is said to be *transverse* (Figure 9.12) when a standard line (normal) aspect is rotated 90 degrees. This transverse aspect projection minimizes distortion along meridians and can have azimuthal or equatorial alignment.

In the *oblique* aspect (Figure 9.13), the axis of the Earth and the axis of the projection are oriented in an arbitrary manner. The oblique aspect is uniquely suited to be used for geographic areas that are centered along lines that are neither meridians nor parallels but are assumed to be "great circles" passing through the region. Distortion is minimized along the central or prime meridian in a projection. An oblique aspect is often difficult to represent as the graticule for the projected plane. Of the nine potential directions, oblique aspect types tend to be directed northwest, northeast, southwest, and southeast.

The *polar* aspect (Figure 9.14) centers on a single point, either the north pole or the south pole. Meridians are subsequently depicted as straight lines radiating from this pole outward toward the equator. Angles between these meridians are always true in relation to the longitudal distance on the Earth's globe. Latitudes (parallels) are equally spaced concentric circles radiating from this same source pole. Polar aspect projections retain symmetry about any meridian and distort between them.

Classification of Map Projections

The classification of map projections, or *projection classifications*, is used to further organize projection types and serve as another tool to gain better results for

Figure 9.12 Transverse aspect.

Figure 9.13 Oblique aspect.

Figure 9.14 Polar aspect.

an application. While the *types* determine the specific method of physical projection, the *classification* defines the properties of the projected surface. Projection classifications ensure that not only that the desired traits required of the application are satisfied but also that the desired parameters are preserved.

There are four primary projection classifications:

1. Equidistant;
2. Azimuthal;

3. Equal area;
4. Conformal.

Projected surfaces can exhibit the properties of several classifications simultaneously, though any surface will potentially distort shapes, areas, angles, or distances at some level. Since all distortions can be measured or estimated, the selection of an adequate projection requires specific knowledge of the application. As a result, the three types of projections (cylindric, conic, and planar) can be combined with one or more of the projection classifications to control the appearance and distortion for any particular application.

For example, a planar equal-area projection preserves both angles and areas. The region of interest is accurately represented in size while angles to other locations are also preserved. The *distances* between the area of interest and other locations are absorbed by the distortion.

To wholly understand how these classifications aid in accurately projecting three-dimensional objects, each is discussed in detail.

Equidistant Projections

A projection is considered *equidistant* (Figure 9.15) when scale is true along at least one line (a focal line) or from one or two points (focal points) to all other points on the projected surface. The focal points are usually at the projection center or some other approximately central location on the surface.

Although no projection can provide a developable surface with a perfectly uniform scale, equidistant projections retain a true scale along these explicit focal features (focal lines and points). Consequently, if these focal features are properly developed, the entire projection surface will benefit from their true scale. For example, a sinusoidal projection can be classified as equidistant because all parallels and the central meridian are shown at true scale.

As another example, an equidistant projection centered on Bangkok, Thailand, will provide true distances from Bangkok to any other city in the world, such as Dallas, Texas, or Dublin, Ireland. However, that same projection will distort the relationship of distance between Dallas and Dublin.

Azimuthal Projections

In an *azimuthal* projection, all points in relation to a central point (such as a pole) are not deformed during the projection process from globe to plane. The central point occurs at the intersection of the tangential plane to the ellipsoid of revolution, or, simply, where the projection plane and the projected object meet. Azimuthal projections preserve the angular relationship of all features in a plane to the central point. This central point has zero distortion because it is the

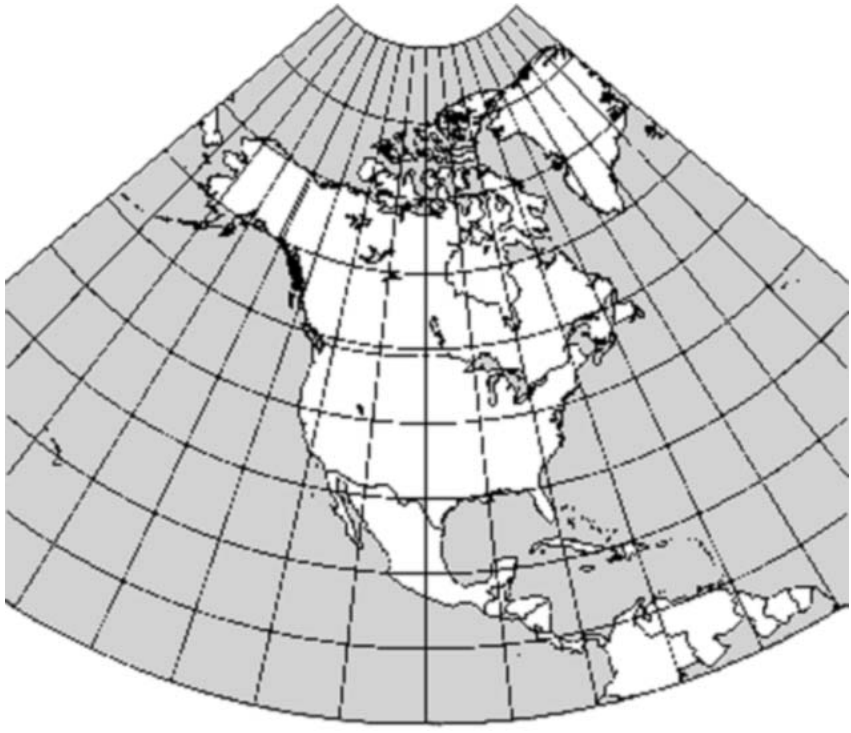


Figure 9.15 An equidistant conic projection of North America.

only point that is being truly represented in the projection. Consequently, distortion gradually increases as distance from this central point increases.

Azimuthal projection classifications are somewhat limited in that they correctly depict directions or angles to all other points in the projection with respect to one or rarely two central points. Figure 9.16 has only one central point, the North Pole.

Equal-Area Projections

The classification is considered *equal area* when the relative sizes of all features on a globe are maintained during the projection process. An equal-area projection vigilantly retains area properties of the spheroid on the plane surface through the use of compensatory scale factors. This means that if area is to be preserved but scale cannot be, then any given feature on a globe (such as a state) requires a scale factor greater than 1.0 in one direction and less than 1.0 in the other direction. Both scale factors compensate for one another and, in doing so, retain the area characteristic.

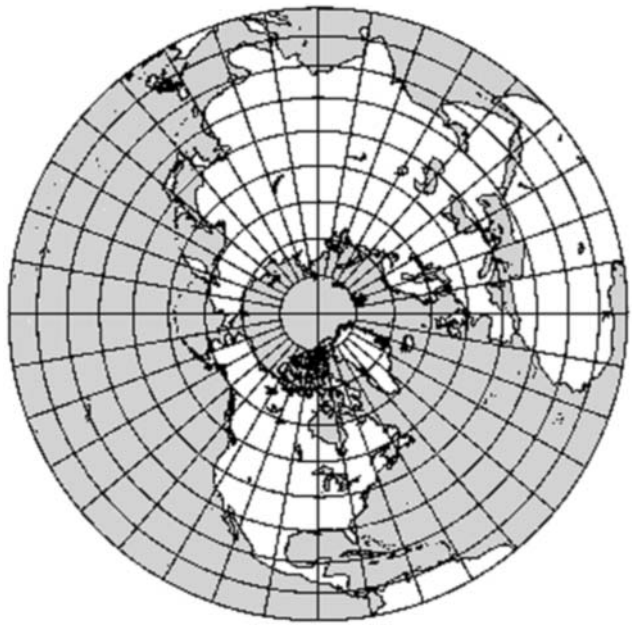


Figure 9.16 Lambert azimuthal equal-area projection of the Earth.

Using compensatory scale factors in equal-area projections, a circle on a globe will project as an ellipse on a plane surface. For example if a coin is placed on any area of a projected surface, it will cover as much of the surface of the sphere as if it were placed elsewhere on the projected surface.

When an equal-area projection is used for small-scale applications that show larger regions, such as in Figure 9.17, the distortion of both angles and shapes increases as the distance of an area from the projection origin increases. Although multiple projection classifications can be concurrently employed, a projection surface cannot exhibit equal-area and conformal classification properties simultaneously.

Conformal Projections

Unlike equal-area projections, *conformal projections* have equal scale factors in all directions at any one point on the projection surface. Instead of preserving area or size, conformal projections preserve shape. For this reason they are also referred to as *orthomorphic* projections.

A conformal surface is also defined as a plane on which all angles at infinitely small locations are correctly depicted. This projection classification increasingly distorts areas away from the point or lines of true scale (a scale

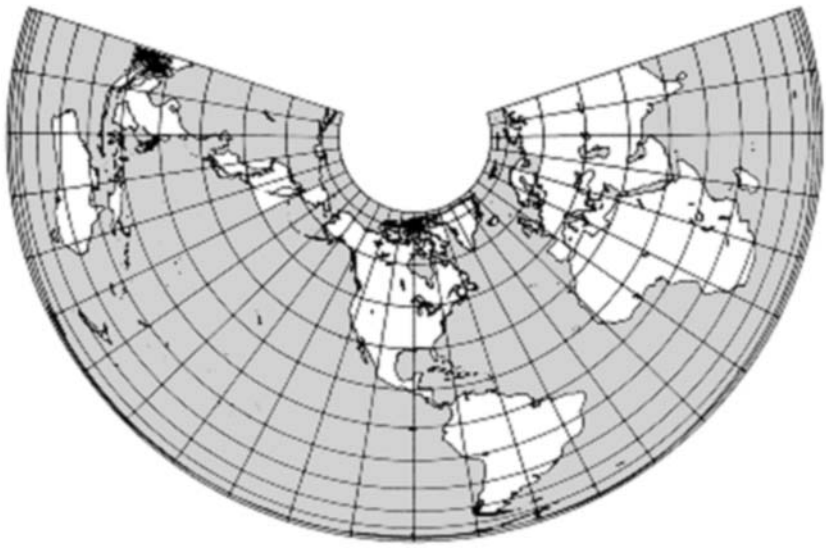


Figure 9.17 Albers equal-area conic projection.

factor of 1.0) on the plane. A good example is the Lambert conformal projection (Figure 9.18). Conformal projections, like the Lambert projection, are ideal for medium- to large-scale projection applications.

Conformal projections are also ideal for constructing plane coordinate systems for use in regional grids. These projections are conducive to managing groups of adjacent systems. Since there is a single scale factor controlling the

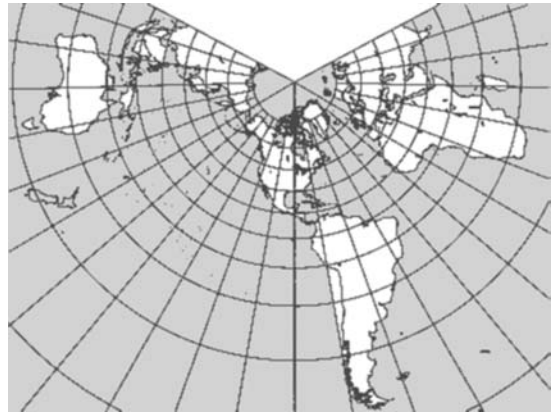


Figure 9.18 Lambert conformal conic projection.

projection, angles retain their integrity in the projected surface. This scale factor, generally not true scale, applies to all distance measurements, unless the distance is of great length or precision. For example, a circle on the globe will render as another, smaller circle in the conformal projected surface.

Distortion

Now that we have discussed the basic tools to transform a three-dimensional object and manipulate the projection, we must look at distortion. We know that the presence of feature distortion is unavoidable when a three-dimensional surface is reduced and projected to a two-dimensional plane. In order for a feature or set of features to be projected without appreciable distortion, it must be *explicitly accommodated* by the entire projection structure (i.e., classification, type, aspect, and so on). Features that are not explicitly accommodated may or may not incur the distortion of geometric properties, regardless of the intention.

The goal of any projection is to minimize the effects of unwanted distortion; therefore, a clear segregation of explicitly and nonexplicitly accommodated features must be made. Geometric integrity is compromised or sacrificed in one feature for the benefit of another. The distribution of these features varies by situation and the specific application.

Distortion Type

Projection distortion is a window into the projection, where its true strengths and weaknesses are revealed. Though most projection distortion is angular or *areal*, a projection can potentially distort five spatial properties of a developable surface: *area*, *angle*, *shape*, *distance,* and *direction*.

For each of these spatial properties there are two primary considerations: How much distortion is present and what is the effect of distortion in the projected surface?

Distortion Magnitude, Distribution, and Overall Effect

Distortion magnitude, or the amount and density of distortion in a projection, answers the question: How much distortion is present? A general rule of thumb is that the larger the area, or *region of interest*, the more pronounced the distortion becomes. This is particularly noticeable in small-scale applications, where distortion magnitude is at its peak. A fine example of a small-scale application is the world map. In small-scale applications, the projection area is large and the scale factor is small, whereas in a large-scale application the projection area is small and scale factor is large.

Small-scale distortion, especially in the world map application, is an issue that has captivated scientists for nearly 3000 years. In fact, Greek astronomer Claudius Ptolemy (circa 150 A.D.) worked with geospatial projections on a small scale, while projections, in general, are known to have been in use some three centuries earlier.

The *overall distribution* and effect of substantial geospatial distortion is often apparent to the eye. By just looking at the projection, you can see where the distortion occurs and how this distortion affects the surface features. For example, a Mercator projection of the world shows the area of Greenland larger than the area of South America, when in reality the area of Greenland is a fraction of the area of South America. These landforms are profoundly affected by the increasing distortion of the Mercator projection toward the poles. The Mercator projection retains the integrity of angles but not areas, resulting in a distributed surface that visibly distorts the appearance of substantial northerly and southerly features.

As discussed previously in this chapter, projection classifications can be used to minimize certain forms of distortion. Classifications enable the user to manage the overall distortion distribution and effect for a specific application. For instance, if a Mercator is projected on an equal-area surface, the area of landforms is retained while the other feature characteristics, such as angles, shapes, and distances, are compromised. Manipulating distortion for a particular application is your best means of properly creating and using a projection.

The Tissot Indicatrix

In 1881, cartographer Nicolas Auguste Tissot published his groundbreaking approach to distortion analysis, now known as the *Tissot indicatrix*, in *Memoire sur la Representation des Surfaces et les Projections des Cartes Geographiques.* The Tissot indicatrix is a helpful tool for visually modeling distortion and for graphically analyzing the distortion properties in a projected surface. Tissot's approach to distortion analysis was clearly the most innovative in his day and still remains an effective method of interpreting a projection. He first began describing this approach as early as 1859, and it has now joined other nineteenth-century mathematical and geographic innovations like the least-squares method or conformal mapping.

The success of the Tissot indicatrix resides in that it is a simple method for visualizing distortion. It has two primary elements: the projected graticule and geometric deformation indicators. Simply stated, the indicatrix depicts the projected graticule as a distortion graph and shows distortion as circles or ellipses at the specific intersections in the plane. A circle represents conformality (no distortion in shapes), while the relative flattening of an ellipse represents the extent

and type of local distortion at the intersection where the object is plotted. These circular or elliptical objects, known as *geometric deformation indicators*, are fixed at a finite scale with explicit locations in the projected surface.

Figure 9.19 depicts a basic Tissot indicatrix with deformation indicators. Figure 9.20 illustrates a standard transverse Mercator projection with its related Tissot indicatrix. Notice how the distortion indicators become larger as they move away from the two central points.

Geometric deformation indicators are circles or ellipses that represent the distortion scale factor at the intersection of any perpendicular graticule line set on the ellipsoid of revolution. The determination of such indicators is the primary driving concept of Tissot's indicatrix and projected surface deformation theory.

Tissot also treats the projected space as a plane with the same size as the region of interest on the reference ellipsoid. For instance, an infinitely small circle on the surface of the Earth model (reference ellipsoid) is identical to an infinitely small circle on the surface of the projected plane.

For the purposes of enhanced visualization and practical appearance, the indicators are significantly enlarged in the indicatrix graph. Although not drawn to scale, the indicators are scaled proportionately to other indicators in the same solution. Accordingly, these geometric deformation indicators depict the characteristics of the projected plane at or nearly at the intersection where they are plotted.

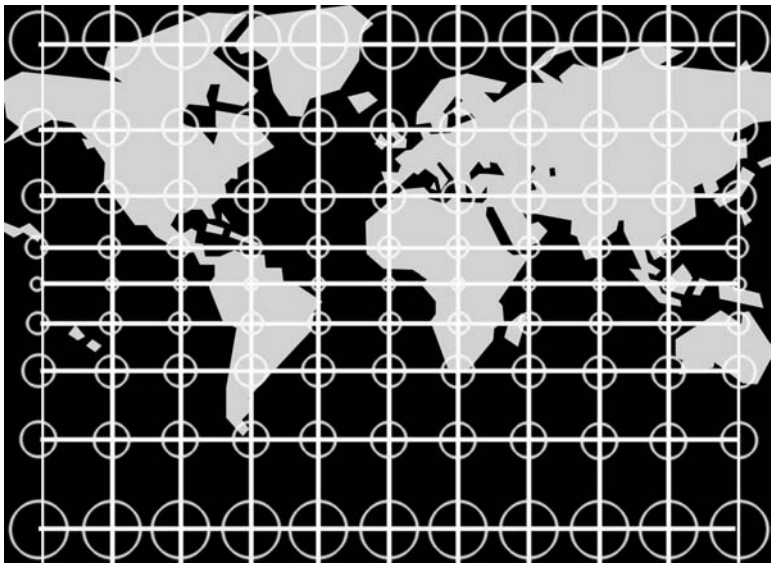
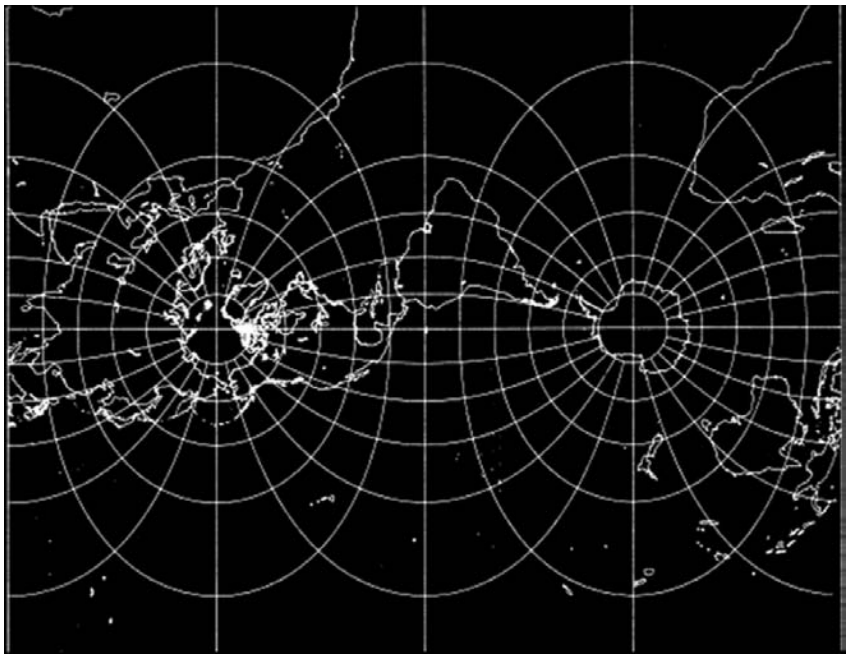
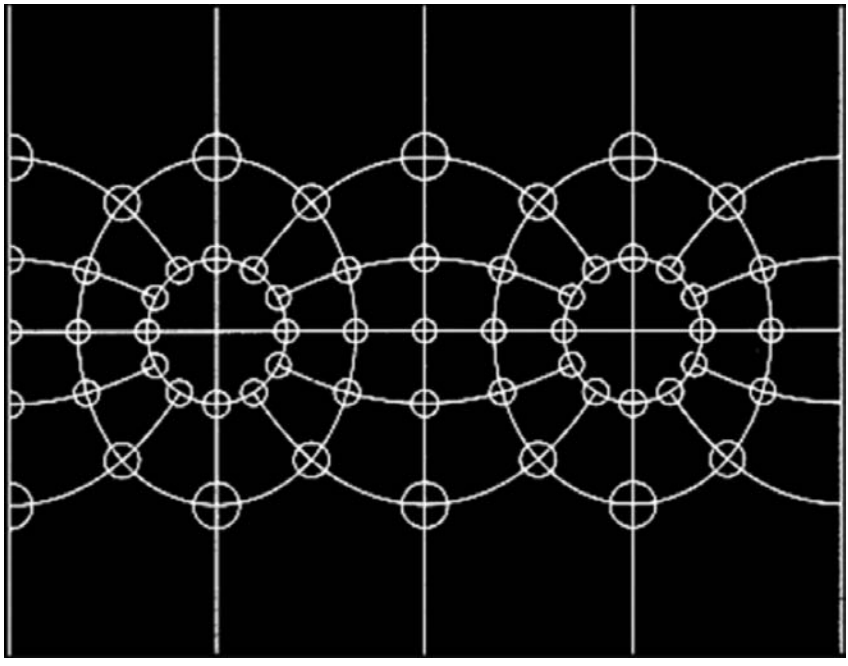


Figure 9.19 A simple Tissot indicatrix of the world.



(a)



(b)

Figure 9.20 (a) Transverse Mercator projection and (b) its Tissot indicatrix.

Methods for Distortion Management

We have discussed map projection distortion in great detail within this chapter but have mildly addressed methods for distortion management. The following methods are tried and true ways to manipulate and manage distortion within any geospatial projection.

Redistribution of the Scale Factor

A proven method of manipulating the effects of and controlling the location of distortion is to modify or redistribute the scale factor. As discussed in detail previously in the chapter, all dimensions of the Earth are changed proportionately when the Earth is reduced from actual dimensions to a globe model. As the surface of this generating globe is further transformed into a projected surface, some dimensions remain unchanged. For example, in a cylindrical projection, as the cylinder is unrolled and developed, the equator maintains its original length and true scale. This *standard line* or line of exact scale is represented by a true scale factor value of 1.0.

The scale factor on any projected surface can vary at any location, since no projected surface has a uniform scale. Scale within a projected surface may also vary at different *directions* at the same location. For instance, in a particular conformal projection the scale is 1.0 (true) along each standard parallel, while other points between the standard parallels are at a scale *less than* 1.0. Accordingly, points outside the standard parallels are at a scale *greater than* 1.0. The resulting conformal projection creates a continuous error surface, in which any line (other than a standard parallel) within the plane has a unique scale factor and any line of substantial distance crosses many different scale factors.

Proportional to the original object, scale factor has innate characteristics that are linear in nature. For example if a line is doubled in length by the projection in the output surface, the scale factor of the projection is 2.0.

A basic formula to use when calculating the scale factor for distortion manipulation is

$$Scale\ factor = \frac{Distance\ on\ the\ projected\ surface}{Distance\ on\ the\ spheroid}$$

Distortion can be manipulated by redistributing the scale factors of the standard lines. The shorter the distance between the standard lines, the lower the distortion, and conversely, the longer the distance, the greater the distortion. A good implementation of this method is to position your particular area of interest in between two standard parallel lines. The projected area benefits from the true scale factor of the standard lines, while distortion is significantly reduced within the projected area.

Changing Projection Aspect

A change in the projection aspect is the most obvious and fundamental method to minimize unwanted distortion. Simply put, it is a change in the location of the projection's center point. When a projection aspect is changed, the distortion pattern in the projected map space is modified whereby areas of least distortion are relocated. An adept user can determine which projection aspect offers the most advantageous distortion relocation for the specific application.

Other Successful Methods

Aside from the two most prominent projection distortion management techniques, there are alternative methods that have proved adequate for particular applications. Here are some fine methods to employ:

- *Visual-logical analysis.* A visual examination of graticule patterns is useful for identifying local distortions (anisotropy).
- *Familiar shapes.* Recognizable features, such as coastline shapes or inland characteristics, provide important clues to distortion patterns and imbue an overall impression of major distortion points.
- *Color differentiation.* Mapping three color scales at locations within the developable surface provides insight into the projection distortion patterns. Color scales are imposed upon the *x, y* range and angular convergence parameters. Prominent differentiation between color scales offers a clear layout of definable distortion points. This color differentiation method lends itself to the structure of raster data.
- *Isarithms (isolines).* The isarithm method provides a good vantage of the overall distortion patterns in a projection. Isolines are basic contour lines that identify the differing plateaus of distortion within the projection. The result can be likened to a topographical map with designations not of elevation but of distortion magnitude. This is particularly helpful in the analysis of equal-angular or equal-area distortion. However, like the indicatrix, isarithms are static objects that are unable to accommodate the diversity of continuous application.

As we have discussed throughout this chapter, a well-planned and adeptly created map projection is critical for the basis upon which to build a GIS. In fact, the success or failure of the application resides initially with the projection. A good understanding of the final application and a basic command of projection elements will ensure a solid GIS foundation.

10

The Coordinate System

Up to this point, we have only touched the surface of coordinate systems, both in theory and function. In the past chapters, especially the chapter on map projections, coordinate systems have taken a nominal role in support of the main topic. We have already heard the likes of geographic and plane coordinate systems but have discussed very little about what these systems do. Concluding this book's theory-intensive Part II to understand the core GIS environment, we will explore coordinate systems and the role they play in a GIS.

What Is a Coordinate System?

A *coordinate* is a number set that denotes a specific location within a reference system. Typical coordinates are the *x*-*y* set ([*x*, *y*]), which is used in a two-dimensional system, and the *x*-*y*-*z* set ([*x*, *y*, *z*]), which is used in a three-dimensional system. With that stated, a coordinate system is the reference system upon which coordinates are defined. A coordinate system is often structured in a two-dimensional or three-dimensional plane that consists of a set of reference points and rules that define the spatial position of points.

In reality, coordinate systems are extensions of mathematics, particularly algebra and trigonometry, and come in various formats. Planar systems have *x*- and *y*-axes, while three-dimensional systems have an additional *z*-axis for height. Terrestrial coordinate systems, such as the ones on geographic maps, utilize a system of latitudinal and longitudinal coordinates. Polar reference systems utilize positional coordinates consisting of the distance from the origin *r* and its angle of inclination Θ. Ultimately, each coordinate system format is used on a specific type of plane.

Coordinate reference systems can also be created for specific regions or particular applications. For instance, the New York City is considered a structured, planned city, whereby streets and avenues intersect, forming a defined grid. Essentially, this is a citywide reference system upon which positions can be easily defined and understood. You can say your position is at 44th Street and 6th Avenue and immediately your position is understood within the boundaries of the city.

This same premise goes for road maps, which commonly have a zoned grid defined by horizontal numbering and vertical lettering. These zones do not define exact positions, but a grid rectangle (area) in which the target resides and the locator can search. For example, you can find a particular town by locating and searching its particular zone (i.e., A6). If there are multiple page maps, the zone must be detailed with the page number (i.e., page 6, A6) since most multipage maps use the same grid identifiers on each page. This procedure of "zeroing in" on a target's position dramatically reduces search times and is comparable to stating the city street you are on (zone) and not the actual address (position).

Mathematics and the physical Earth require the use of various standard coordinate systems. In relation to GIS and geodesy, there are a handful of important reference systems worthy of discussion and necessary for the novice GIS user to understand. These coordinate systems include the Earth-based geographic coordinate system, the vector-based Cartesian coordinate systems (both two-dimensional and three-dimensional versions), the zone-based Universal Transverse Mercator and Universal Polar Stereographic systems, and the U.S.-based State Plane Coordinate System. Let us take a brief look at each.

The Geographic Coordinate System

A geographic coordinate system is a three-dimensional positional reference that utilizes latitude, longitude, and ellipsoidal height. We have briefly explored the geographic coordinate system in our map projection discussion. This positional reference system is among the most used today for global locations and is typically associated with geodetic datums.

To recap, the Earth's three-dimensional ellipsoid is mapped through a series of horizontal and vertical reference lines, which form a set of standard reference lines. As we know, the horizontal lines are called *longitude* lines. Longitude lines are parallel with one another and are therefore often called *parallels*. The vertical lines are called *latitude* lines. Since latitude lines meet at the poles, they are often called *meridians*. Figure 10.1 illustrates Earth's latitude and longitude lines.

Earth's parallels intersect every meridian at right angles. The central parallel, located exactly between the poles, is called the equator. This central,

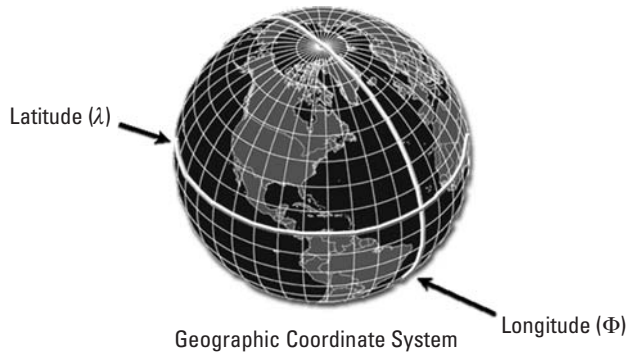


Figure 10.1 Earth's latitude and longitude.

standard parallel holds a longitudinal value of 0°. The remaining parallels are assigned angular values relative to their direction and distance from the equator. The lines of longitude progressing toward the north pole are assigned parallels ranging from 0° to 90° north. Likewise, the lines of longitude progressing from the equator toward the south pole are assigned parallels ranging from 0° to 90° south.

Earth's meridians run from pole to pole, and unlike Earth's parallels, meridians convene at each pole. The center meridian is called the prime meridian, which is the line of latitude that crosses through Greenwich in the United Kingdom. The prime meridian is the primary latitude reference line and is assigned an angular value of 0°. The remaining meridians are assigned angular values relative to their direction and distance from the prime meridian. Meridians progressing east of this center meridian are assigned angular values ranging from 1° to 180° east. Similarly, the lines of latitude progressing west of the center meridian are assigned angular values ranging from 1° to 180° west.

The geographic coordinate system uses geodetic latitude and longitude to define position on an ellipsoidal surface and, as such, to form the geographic coordinates. Geodetic latitude and longitude are identical to the Earth-based system. In addition, the geographic coordinate system provides a quantity for geodetic height. Geodetic height at a point is the distance from the reference ellipsoid to the point in a direction normal to the ellipsoid. Figure 10.2 illustrates the geodetic latitude, longitude, and height.

The geographic coordinate system provides a solid reference through which any location on Earth can be uniquely distinguished and identified. As we have seen with map projections, GIS generally transforms three-dimensional geographic coordinates to other coordinate systems in order to clearly portray point positions on a two-dimensional information product. This leads us to our second coordinate reference system, namely, the Cartesian coordinate system.

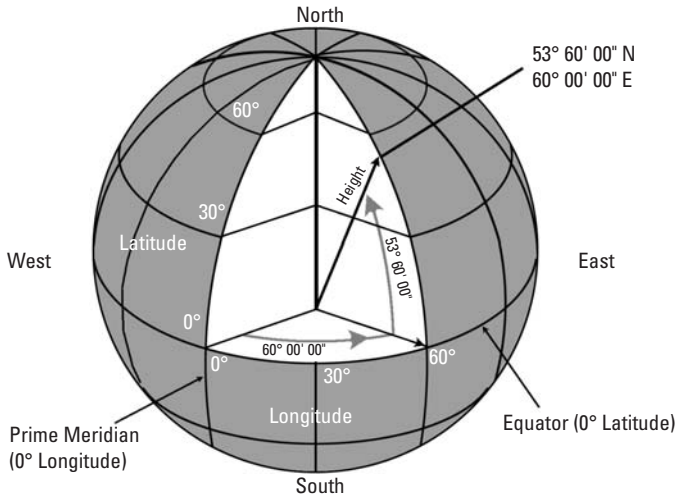


Figure 10.2 Geodetic latitude, longitude, and height.

The Cartesian Coordinate System

Perhaps the most used coordinate reference system in mathematics, science, and GIS is the Cartesian (or rectangular) coordinate system. The Cartesian system is a reference structure in which point positions are measured along intersecting planes in two and three dimensions. The common coordinate or intersection point for all planes is called the origin. All measurement distance increments are consistent on every intersecting plane and throughout the system. These planes are detailed along the *x*, *y*, and/or *z* axes.

The Cartesian coordinate system was named for famous French mathematician, scientist, and philosopher René Descartes (1596–1650). Descartes, known also as Cartesius, is often dubbed the "Father of Modern Mathematics" for his mathematical genius and his creation of analytic geometry. In his *Discourse on the Method to Rightly Conduct the Reason and Search for the Truth in Sciences* (usually shortened to *Discourse on Method)*, Descartes detailed his method of translating geometric problems into algebra using a coordinate system. This mathematical innovation paved the way for later mathematical advances, particularly calculus.

The Cartesian two-dimensional coordinate system involves two axes: the horizontal *x* axis and the vertical *y* axis. The origin is detailed as coordinate (0, 0). This planar, two-dimensional model is divided into four quadrants by the perpendicular, intersecting axes. These are labeled quadrants I through IV, starting in the upper right-hand quadrant and progressing counterclockwise. These quadrants take on the following structures:

- Quadrant I: *x* axis is positive, *y* axis is positive (+*x*, +*y*);
- Quadrant II: *x* axis is negative, *y* axis is positive (−x, +*y*);
- Quadrant III: *x* axis is negative, *y* axis is negative (−*x*, −*y*);
- Quadrant IV: *x* axis is positive, *y* axis is positive (+*x*, −y).

Figure 10.3 offers a good line drawing of the standard two-dimensional Cartesian coordinate system.

As you might have guessed, the Cartesian coordinate system is perfect for plotting and measuring vector-based feature geometry. A GIS references the positions of feature-defining points, lines, and polygons using the Cartesian system of coordinates. In addition, you may remember that a map projection brings three-dimensional geographic coordinates (latitude and longitude) into two-dimensional Cartesian (or planar) coordinates.

Often points are located in a three-dimensional space and a height measurement is necessary. Height is measured along a third axis—the *z* axis.

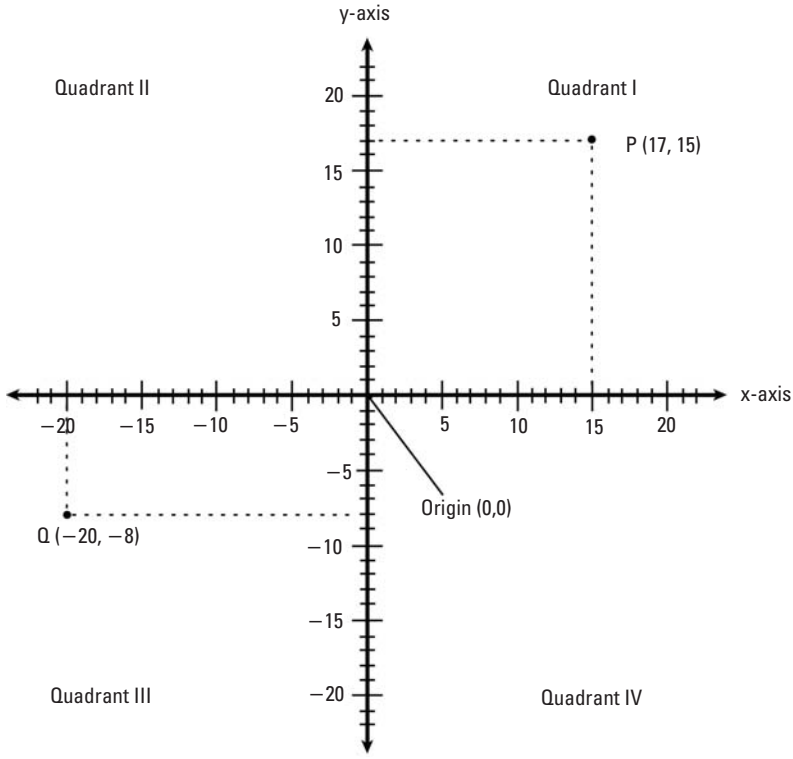


Figure 10.3 Two-dimensional Cartesian coordinate system.

Three-dimensional objects, like the Earth's ellipsoid, utilize a height measurement for positional accuracy. Consequently, geodetic datums use this system. Figure 10.4 details the three-dimensional Cartesian coordinate system.

Frequently, the Earth is set to a three-dimensional Cartesian system called the *Earth-centered, Earth-fixed* (ECEF) Cartesian coordinate system. ECEF is used to define three-dimensional positions with respect to the earth's mass center of gravity and the center of the reference ellipsoid. The Earth's center serves as the origin (0, 0, 0). You may want to refer back to Figure 10.2, which details an ECEF system as it relates to the standard lines: the equator and prime meridian.

This form of Cartesian coordinate system is especially useful for geodetic structures. It is called Earth fixed because each axis is fixed in respect to the orientation of the Earth. For example, if the Earth rotates, the axes rotate sympathetically and proportionately.

Before leaving the Cartesian system discussion, we need to explore an important terminological matter. Habitually, cartography and surveying practices use Cartesian coordinates to define positions on maps. As a matter of function and ease, the *x*-axis is considered an *easting* and the *y*-axis is considered a

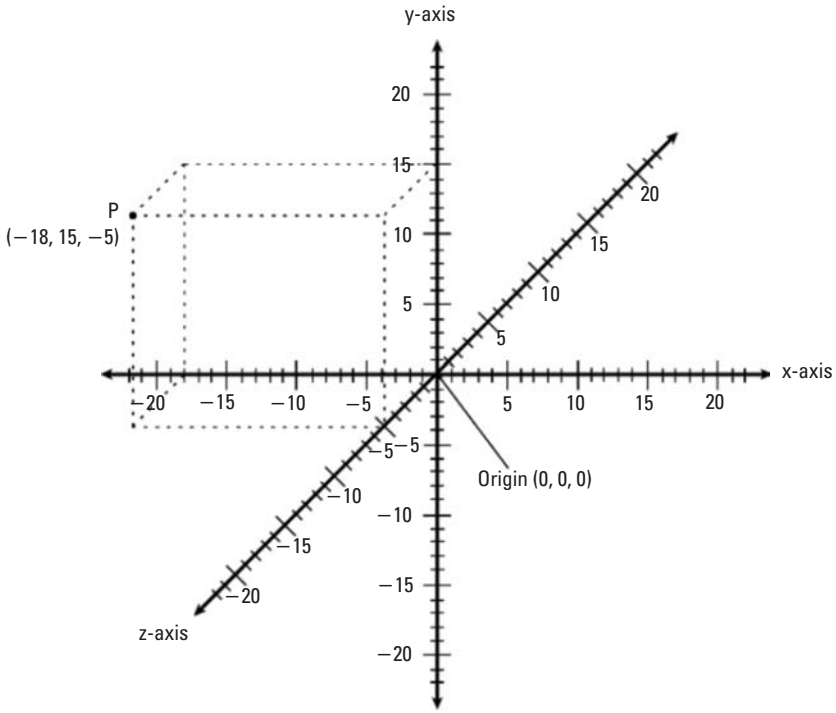


Figure 10.4 Three-dimensional Cartesian coordinate system.

northing. The northing-easting terminology is used to help efficiently relay position, while avoiding signage and coordinate errors.

An easting is the distance east from a north-south reference line, or meridian. Often the prime meridian is used as this reference line. Similarly, a northing is the distance north from an east-west reference line, or parallel. The equator is often used as this reference line.

Furthermore, to simplify positions in the northing-easting reference method, negative numbers are eliminated. To do this, *false easting* and *false northing* constants are used. A false easting is an adjustment constant added to *x*-axis coordinates to eliminate negative numbers. Likewise, a false northing is an adjustment constant added to *y*-axis coordinates to eliminate negative numbers. On maps using the northing-easting terminology, false easting and false northing values are supplied.

Universal Transverse Mercator

The Universal Transverse Mercator (UTM) is a global map projection that transforms the three-dimensional world into a two-dimensional system. The UTM employs an international plane coordinate system that extends around the world from 84° north above the equator to 80° south below. It is extended an extra 4° in the North to cover the northernmost land on Earth. This bounded coverage is because the UTM projection distorts near the north and south poles. The Earth's polar regions are handled by the Universal Polar Stereographic projection, which is discussed later in this chapter.

The UTM coordinate system is set upon a zoned grid, which divides the Earth into 60 equal zones that are all 6° wide in longitude (east-west). The UTM zones are numbered 1 through 60, starting at the international date line (zone 1 at 180° west longitude), progressing east past the prime meridian (zone 30), and back to the international date line (zone 60 at 180° east longitude). Figure 10.5 illustrates the complete UTM zones.

As depicted in the figure, each zone extends in both the northern and southern hemispheres and employs the northing-easting methodology for position identification. Since the equator separates the hemispheres, it serves as the zone's standard parallel. Each UTM zone refers to its own central meridian located at each zone's middle (3°).

Given that the northing-easting method does not allow for negatives, the equator separates each zone into two positive sections: a north section and a south section. Each section then employs its own rectangular grid system using the equator and the zone's central meridian as the two standard references for each system. In total, the UTM comprises 120 separate coordinate systems.

Coordinate positions referencing the UTM system must be indicated with the zone number as well as its northing and easting values in meters. The easting

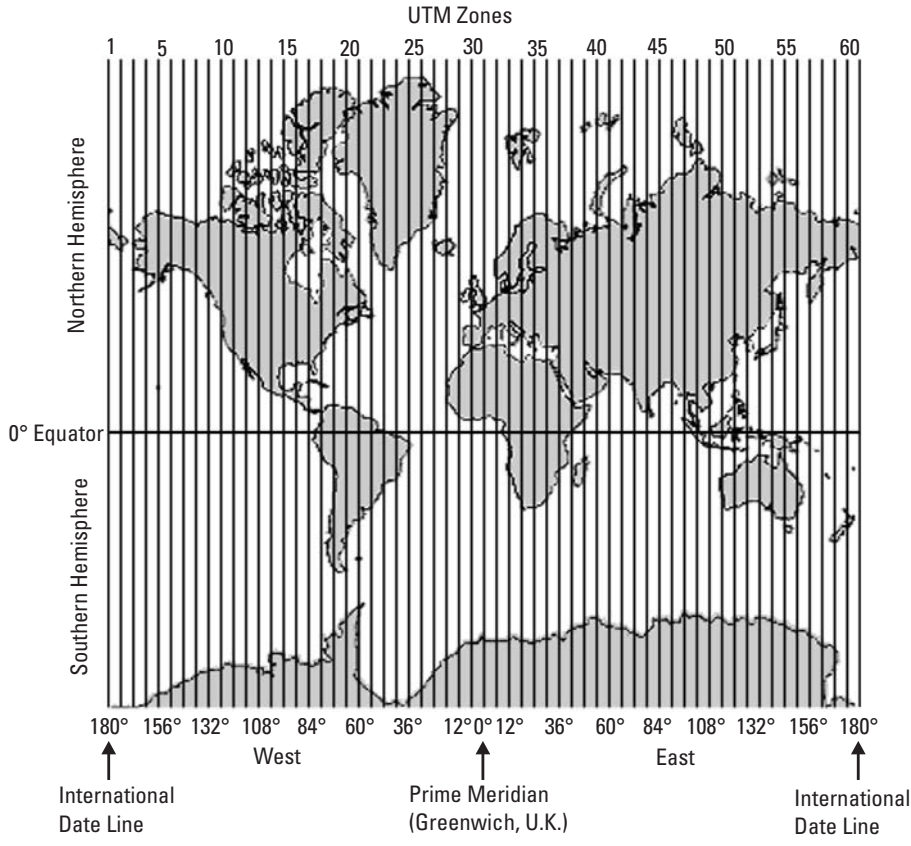


Figure 10.5 The Universal Transverse Mercator zones.

value is always a six-digit value in meters, while the northing value is nearly always a seven-digit number in meters. For example, a typical UTM coordinate for zone 13 would look like *13 645,000mE 2,870,000mN*. Furthermore, to ensure positive coordinates, each UTM zone utilizes a false northing value of 0m in the northern hemisphere and 10,000,000m in the southern hemisphere, and a false easting value of 500,000m.

The UTM system is used worldwide for a multitude of applications, although it may not always be an identical system. For example, in Europe, the Kruger-Gauss projection is used, and the system is consequently called the Kruger-Gauss Transverse Mercator. In the United States, the military uses a modified UTM system called the Military Grid Reference System (MGRS), which further divides the UTM zones into 8° latitudinal divisions (the last division with 12°) designated with a letter. The MGRS starts at C at the 80° south latitude and progresses to X ending at the 84° north latitude. The X latitudinal

division is the only division 12° high. Additionally, to avoid user error, the letters I and O are not used because of their similarity with numbers.

In summary, the UTM is a widely used planar coordinate system quite capable of representing the Earth's geographic coordinates on a flat, northerly-easterly plane. The UTM is used worldwide, and in many countries it serves as a basis for their national grid, particularly the United States and the United Kingdom. Incidentally, the approximate scale distortion of the UTM system is only 1 part in 2,000, proving that the UTM is a highly accurate system. Still, as mentioned earlier, the UTM projection cannot be used for the Earth's polar regions due to amplified distortion. Let us now examine UTM's corresponding system that focuses on theses polar regions—the Universal Polar Stereographic.

The Universal Polar Stereographic

Serving as the polar version of the UTM, the Universal Polar Stereographic (UPS) system uses a universal polar stereographic projection to transform the three-dimensional polar regions into a two-dimensional system. The UPS takes over where the UTM projection distorts, which is defined as above 84° north latitude and below 80° south latitude, but also includes an additional 30' of latitude, extending into the UTM grid for a small degree of overlap between both systems. Figure 10.6 portrays the approximate combined geographic coverage of Earth through the UTM and the UPS systems.

As depicted, the UPS system has two zones, namely, the north and south polar zones. Like the UTM, each zone has its own grid structure and coordinate system. Northings and eastings are used to characterize coordinate positions for each zone and are subsequently calculated using a polar aspect stereographic projection. Figure 10.7 presents a polar aspect stereographic projection.

Similar to Figure 10.7, the UPS system employs a conformal map projection that is free of angular distortion. The projection retains a constant scale along the parallels and is true scale at 87° 7′ north and south. As with the UTM, the U.S. military utilizes a gridded version of the UPS system in which each grid block is identified with a distinguishing two-letter combination (e.g., ZA, where the easting coordinate is defined by Z and the northing coordinate is defined by A).

Overall, the UTM and UPS systems are complementary to one another and combined provide a whole flat view of Earth. The zoned, multigrid approach offers a wonderful positional mapping capability to cartography and surveying, while the minimized distortion provides a near true-to-life representation. In short, the UTM-UPS combined system is a good way to characterize the Earth as a whole and to lay a foundation for both coordinate representation and measurement.

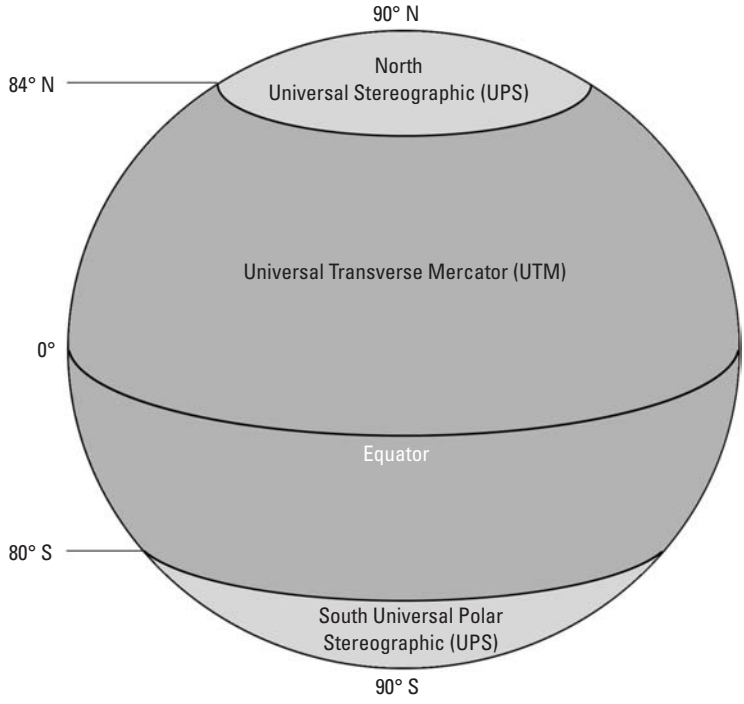


Figure 10.6 The UTM-UPS global coverage.

The State Plane Coordinate System

The State Plane Coordinate System (SPCS) is a zoned grid system that divides all 50 United States into more than 120 numbered zones. The SPCS was first introduced in the 1930s by the U.S. Coast and Geodetic Survey to serve as a common reference for cartographers, surveyors, and engineers. The SPCS is used only in the United States and allows coordinates from an entire state or portions of a state to be transformed into planar (rectangular) coordinates on a single grid. The SPCS was developed to have even greater precision than the UTM, with its maximum scale distortion at just 1 part in 10,000.

Unlike the UTM system that covered both sections of the northern and southern hemispheres within one zone, SPCS zones are bounded by state borders. Smaller states are divided into one or two zones, while larger states are divided into as many as six zones. Each zone has its own origin, central meridian, and standard parallel which often follow state county boundaries. Additionally, for easy identification, each zone has its own assigned code.

Figure 10.8 illustrates the SPCS used in Arizona. It is important to recognize that the state is divided into three zones, which are attributed for their

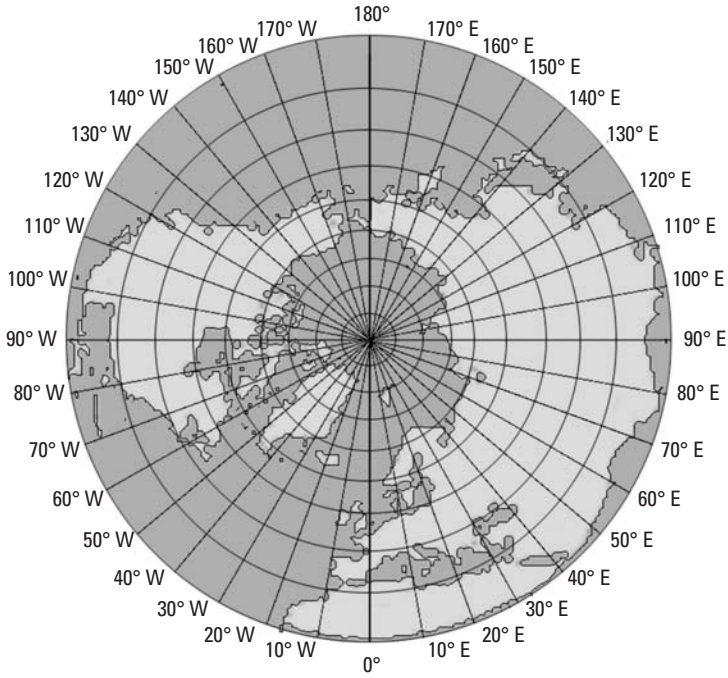


Figure 10.7 A north polar aspect stereographic projection of the Earth.

spatial location within the state: Arizona West, Arizona Central, and Arizona East. Each zone is assigned a unique four-digit code encompassing a common state prefix 02 and an incremental two-digit zone suffix. As illustrated, Arizona's zones follow the state's county lines.

The original SPCS was developed using the standard national datum NAD27 (North American Datum of 1927). Appropriately called the U.S. State Plane Coordinate System of 1927 (SPCS27), this early version references the Clarke 1866 ellipsoid and creates a set of local grids for each state. Like NAD27, the SPCS27 system is based on the foot measurement—a U.S. standard that has since changed.

After NAD27 was superseded by the more accurate NAD83, the SPCS was revised. The new system, similarly called the U.S. State Plane Coordinate System of 1983 (SPCS83), is based upon the latest national standard horizontal datum (NAD83). SPCS83 moved away from the antiquated foot measurement to the U.S. standard, the meter. Due to its NAD83 foundation, SPCS83 offers greater geodetic control and cross-links to the GPS-based GRS80 datum.

The SPCS is based upon two map projections: the transverse Mercator and the Lambert conformal conic. For states that are elongated north to south,

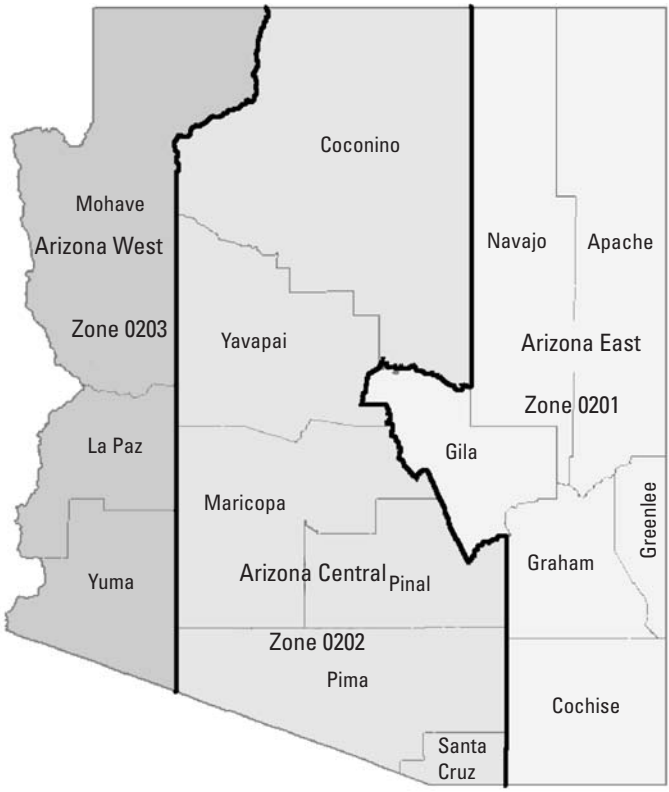


Figure 10.8 The State Plane Coordinate System in Arizona.

such as California and Florida, the SPCS utilizes the Lambert conformal conic for minimal projection distortion. For states that are elongated east to west, such as Pennsylvania and New York, the SPCS utilizes the transverse Mercator.

The SPCS also utilizes the northings-eastings method of coordinate identification. Like other planar coordinate systems, false northings and false eastings are used to keep coordinates positive for every zone. Since every zone is slightly different, each has its own values for false northing and false easting. Overall, the SPCS system is an exceptionally precise and easy system to employ throughout the entire United States.

Coordinate System Implementation in GIS

Coordinate systems play an important role within GISs. They are used throughout every form of geodetics from geographic information referencing and vector

feature geometry to geodetic map projections and datums. The coordinate system forms the definitive core of GIS and serves as the primary tool for modern geodetics.

It is important to understand that the implementation of a coordinate system in a GIS enables geographic data to be defined by vector data. The positional accuracy of vector data on a coordinate system allows for real-world features to be precisely represented in the GIS, as well as on the ever-important geographic information products. Vector data also enables feature topology and its resulting, advantageous efficiency. In short, GIS accuracy and vector data availability are both direct products of utilizing a coordinate system.

Another beneficial effect of coordinate system implementation is that geodata from various (different) sources can be used with one another through layering techniques and coordinate system transformations. By knowing the source data's reference system, feature coordinates can be converted to a single, uniform GIS coordinate system. With all the data sources converted, the information can be manipulated on the same grid, and associations between sources can be made. It can be noted that without the implementation of a coordinate system, a GIS would remain just a database.

Part III:
GIS Applications and Environments

Geographic Data for Analysis and Presentation

11

Thematic Mapping

In his published anecdotes, Samuel Johnson, noted eighteenth-century writer and philosopher, wrote, "The use of traveling is to regulate imagination by reality, and instead of thinking how things may be, to see them as they are." We are those travelers in that, throughout this book's final part, we will be traveling from GIS's theory into the territories of true GIS products and real-life applications. As part of this demystification of GIS, we will depart from discussing how information products should be (imagination) to how they really are (reality).

We already know that in order to produce useable geographic information products, the user must manipulate geodetics through the selection of a standard ellipsoid and datum for their particular focal area. We also know that, once selected, a beneficial map projection must be chosen and applied upon a defined coordinate system. After these elements are set, valid geographic information products can be created and analysis can be conducted with a degree of quantified certainty.

One of the most common and valuable products a user can produce with GIS is a map. Some map products depict pure cartographic information. These products include road maps, subway maps, and the like. Other map products offer a mechanism for comprehensive geographic data analysis and, as such, are typically more valuable. These specialized maps utilize user selection and the benefits of a GIS to offer a unique, yet comprehensive, description of the focal geographic terrain. This specialized map product is called a *thematic* map.

The Need for Thematic Mapping

We have all seen and used a thematic map of some sort. Think of any major city subway or bus map with the stations or stops depicted, or of a TV weather map with temperatures and high- and low-pressure areas represented. Well, the stations, stops, temperatures, and highs and lows are all attributes of geospatial data. These dataset attributes are called themes, and subsequently the subway, bus, and weather maps are all variations of thematic maps.

A thematic map is defined as a specialized map that depicts the spatial distribution of one or more explicit themes. These map themes portray dataset attributes and can detail anything quantifiable, such as population density, income distribution, crime, natural resources, water distribution, pipelines, and disease occurrence. Since GISs are used to manipulate, analyze, and convey geospatial information, thematic maps offer the GIS user a better way to understand the geographic area in focus, as well as an excellent means to inform others.

Now, just imagine how difficult it would be for the user if the bus stops were identified in a data table rather than graphically on the map or if the temperatures for 20 neighboring locations were spoken by the broadcaster and not represented on the map. It is quite possible that some commuters would miss their stops, while the comprehension of temperatures for 20 locations would become an exhausting test of mnemonics and listening. Without thematic, visual indicators, the delivery of this data would be less than effective.

Thematic maps offer a means to define data attributes in a way that is not fully available from a data table. In fact, thematic mapping offers the GIS user a better instrument to convey geographic information and geospatial data that is relative to their information objectives. This, in part, provides a greater means to analyze data sets and find trends and spatial patterns that may remain hidden in data tables. Figure 11.1 offers some insight into how a data table from the U.S. Census Bureau can be interpreted differently when portrayed as a thematic map.

As demonstrated in Figure 11.1, the data table containing household and county information for the state of Florida "hides" the housing trends that the map freely depicts. One definitive trend, in particular, is the state's household per county ratio, whereby the highest ratio is in southeastern Florida and the lowest ratio in northern Florida. The thematic map helps visualize these trends and spatial patterns, while the data table, although informative, seemingly obscures Florida's popular household areas within its data.

Thematic maps offer a bird's-eye view of a particular attribute within the context of a geographic area. As such, they can provide either qualitative information or quantitative information. Examples of qualitative information are vegetation types, soil conditions, and land use categories. Quantitative information includes such attributes as number of Hispanic homeowners, carbon

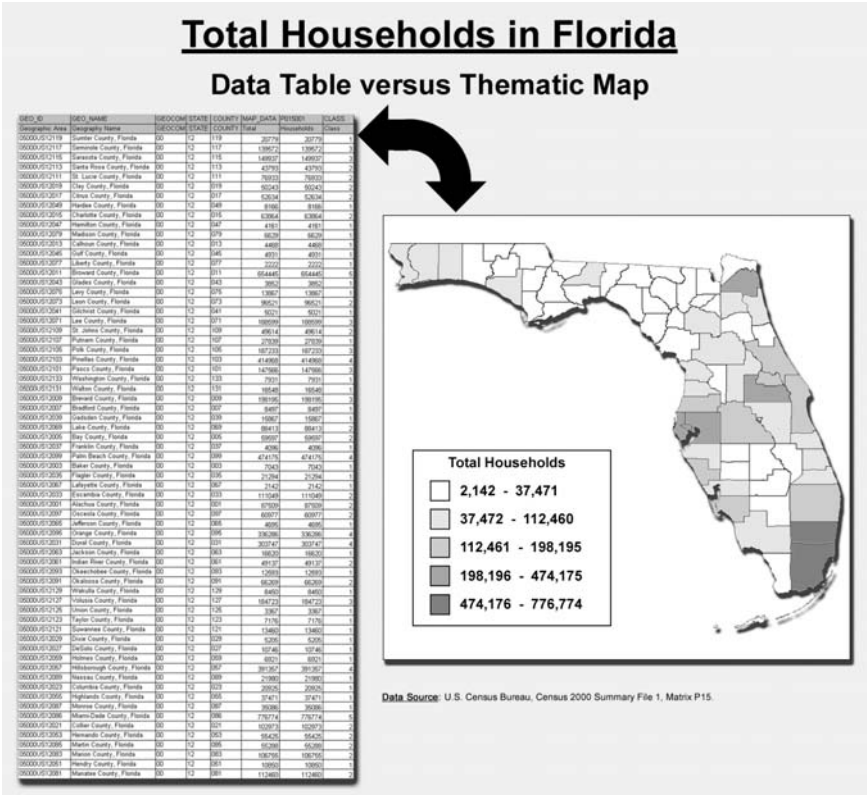


Figure 11.1 A data table versus a thematic map.

monoxide emission measurements, and soil depths. The thematic map can depict both qualitative and quantitative data only if two or more themes are used in tandem.

In the end, the absolute need for well-designed thematic maps becomes apparent whenever medium to complex attribute analysis is undertaken. Since users have inherent limitations to their mental acuity, visual models are necessary for enhancing data comprehension and vitalizing the data's analytic potential.

Manipulating Thematic Layers

A GIS offers users an easy method to manipulate specific data attributes through the application of thematic layers. A layer is a set of theme-based data that is described and stored in a map's data library. Layers do not contain geographic

data, per se, but are georeferenced to geographic data. Additionally, layers complement geographic data through a basemap or satellite image, or such information as roadways, pipelines, soil sample locations, soil types, vegetation types, and water distribution. Layers can be turned on or off in a GIS and can be ordered relative to viewing priority. For instance, a higher priority theme would be placed atop a less important theme, such as a railroad track layer atop a background map of the area.

Figure 11.2 offers a good, visual explanation of thematic layers. The composed thematic map is a combination of five thematic layers: cities, railroads, interstates, major roads, and states. These layers are logically ordered for best overall viewing, with the attribute layers overlying the background layer (states). If the states layer had the highest priority for viewing, then it would block the other themes from showing through. Therefore, part of mapmaking with a GIS is both realizing the importance of logical layer order and keeping data conveyance at a premium.

For GIS users, thematic layers prove to be incredibly useful when developing information products. What makes thematic layers so advantageous is the fact that they provide data attributes in the form of features and surfaces that are uniquely linked to the geography. These features and surfaces are layered and georeferenced to help advance the overall understanding of the terrain and offer insight into what objects affect the landscape, and to what degree. Incidentally, this forms the foundation of thematic mapping's enhanced geospatial analysis.

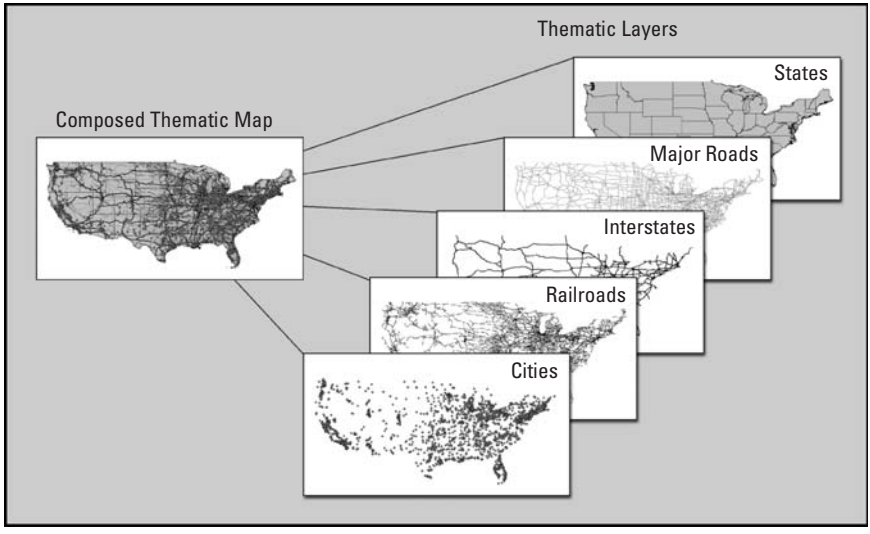


Figure 11.2 How thematic layers constitute a thematic map.

Thematic layer *features* are either natural or manufactured objects that are present or exposed to the terrain, either above ground or below ground. Thematic features include natural objects, such as wetlands, mudflats, and hydrology, or manufactured objects, such as pipelines, roadways, and electrical distribution lines. Each map feature has a defined geospatial position, shape (i.e., point, polygon, or line), and identifier (or symbol) that represents one or more of its characteristics.

Thematic layer *surfaces*, on the other hand, are slightly different and offer a different type of information. These layer surfaces offer a perspective to the terrain in terms of specific characteristics. Examples of thematic surfaces include elevations, soil density, temperature, pressure, precipitation, and flooding. Unlike features, surfaces do not always have defined shapes.

Proper thematic mapping involves a certain level of user discipline relative to thematic layers, as well as a consistent layering scheme. For effective GIS use and map control, layers should be consistently named, with easily identifiable labels. These layer names should suggest the layer's function and indicate some type of hierarchy. Suppose you are entitling a thematic layer of regional streams. A title such as "rty65sc" would make a poor thematic layer name, since it is not easily ascribed to streams or anything else. A better choice would be "Streams001" since it is obvious and has a method to define a stream hierarchy.

Now imagine having 20 different thematic layers for a particular region, including seven stream-related layers, six soil-related layers, and seven layers defining various utilities (gas, electric, and so forth). Which label nomenclature would you rather use? What if you had to produce 22 multithemed maps of that region for a client? Without a disciplined approach to layer taxonomy, this thematic mapping undertaking can rapidly become an onerous task.

As with everything in GIS, there is no single correct way to manipulate or display thematic information. GIS is all about personal choice, creativity, and user flexibility. As such, GIS offers various thematic mapping types from which to choose. We will take a closer look at each thematic map style to understand how and when each is appropriate.

Types of Thematic Maps

GIS offers a fair degree of flexibility and choice when it pertains to what geospatial information is used, how it is used, and the manner in which it is presented. Similar control is offered when depicting attribute data in a thematic map. Consequently, there are numerous ways thematic information can be portrayed over a geographic region. Depending upon the particular data portrayal or analytic objective, the GIS user has a certain amount of freedom over the mapping design.

Given the widespread implementation of thematic mapping, four prominent map types have surfaced. Each offers a particular analytic benefit and, as mentioned previously, can depict qualitative or quantitative data. These four primary thematic map types are choropleth maps, graduated symbol maps, dot density maps, and isopleth maps. Let's examine each one.

Choropleth Maps

Perhaps the most prevalent type of thematic map, the *choropleth map* offers a definite method of portraying ratios, densities, or proportions aggregated within a specific region. These regions are defined by bounded areas, such as census tracts, counties, states, or countries. Choropleth maps, also known as *graduated color maps,* depict thematic information in these bounded areas through color shading and color/value attribution.

Figure 11.3 provides a good example of a choropleth map and illustrates how state median age data can be portrayed in grayscale. The map quickly offers spatial patterns associated with age and individual states that would have been nearly impossible to discern from the data alone. Notice how the eastern states have a higher median age, while the central states hover closer to the middle of the age spectrum. It is easy to visually identify that Utah has the lowest age group (youngest median population).

A choropleth map assigns a value or range of values (called *data classes*) to a specific color or shading. In the figure, the data classes were assigned to five different shades of gray. Hence, the figure has five data classes. For best results,

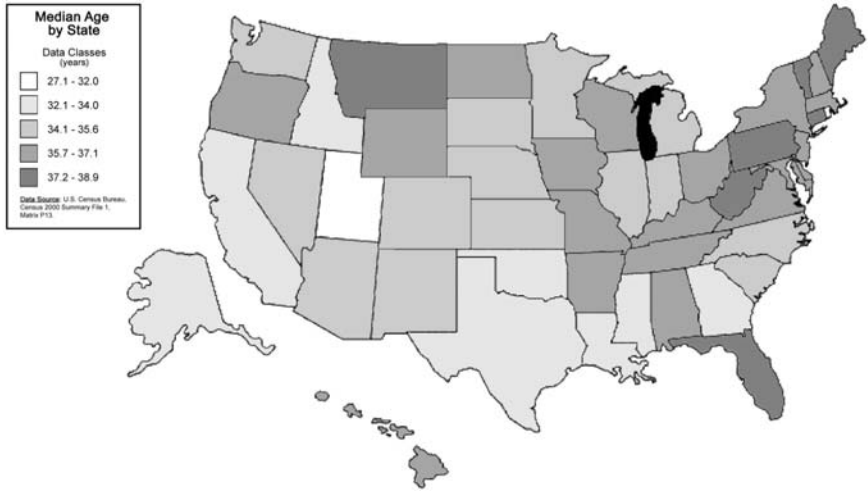


Figure 11.3 A choropleth map of the U.S. census data on median age by state.

choropleth maps should utilize no more than eight data classes since most individuals have a hard time distinguishing between many subtle variations in colors.

Choropleth maps work best with discrete data over defined, bounded areas. Problems arise when choropleth maps are detailed with too many unique values attributed by unique colors. In such cases, the map turns into a geographic rainbow and loses its analytical worth. Poor map designs lead to spatial patterns and communal trends that get lost in the sea of differentiated color.

Choropleth colors for data classes follow particular conventions, whereby the rule of thumb is: The lighter the color, the smaller the value; the darker the color, the greater the value. In the median age map (Figure 11.3), the color white indicates the youngest age range (smallest value), while the darkest gray indicates the oldest age range (greatest value). These maps are portrayed in grayscale or color, typically using a predefined color scheme called a *color ramp.* The grayscale colors in the figure form a predefined color ramp. Users also have the option to bypass the color ramp scheme to define their own color-data class attribution. Whatever color scheme you use, the overall objective is to provide choropleth maps that are simple to understand and offer the greatest potential for spatial trend analysis.

Graduated Symbol Maps

The second type of map is the *graduated symbol map,* which is similar to choropleth maps in function but uses symbols (typically a circle or dot) instead of a color fill. Graduated symbol maps are highly beneficial when indicating quantity or magnitude of a certain phenomenon in relation to geographic positioning. These thematic maps indicate positional data and attribute values through a common symbol of varying size. As the choropleth had a rule for its colors, graduated symbol maps have a similar rule of thumb: The smaller the symbol, the smaller the value; the larger the symbol, the larger the value. Figure 11.4 offers a sample graduated symbol map that details total population by U.S. state.

In Figure 11.4, state populations are shown in relative approximation using various sizes of dots. Each size is assigned a specific value (e.g., 1 million) and is placed in the appropriate states. You can easily recognize a pattern for the country's population through this particular map type, with the east coast having more states with a greater populations than the west coast.

Aside from the "standard" graduated symbol map, a few variations exist for enhanced data portrayal. One variation is called the proportional circle map, wherein the map symbols (graduated dots) are proportionally sized according to their represented value. For example, a proportional circle map of the national population (i.e., Figure 11.4) would perhaps have a miniature dot to represent 1

Figure 11.4 A graduated symbol map of U.S. state populations.

million people, a dot that is 4.5% larger to represent 4.5 million, and another dot that is 10% larger than the miniature dot to represent 10 million people. Value and symbol size are duly proportional.

Another variation is the pie chart map, wherein the symbol used to represent data is a proportional pie slice. This type of map depicts two pieces of discrete data for each bounded region (e.g., state). Last, a third variation similar to the pie chart map is the bar/column chart map. This map portrays bar graphs as a symbol for each region or state and, like its pie chart counterpart, offers a way to depict two forms of discrete data. Ultimately, the most popular of all the variations of graduated symbol maps is the "standard" version.

Dot Density Maps

Often, there is a need to show theme concentrations over a defined region. Choropleth and graduated dot maps are ineffective in the conveyance of thematic densities when they apply to specific geographic locations (as opposed to aggregate areas). For these instances, a *dot density map* is the choice. As the name implies, dot density maps depict spatial densities of particular thematic elements or events.

These specialized thematic maps are excellent for discerning area-theme relationships and can provide insight into the concentrated patterns of a bounded area (e.g., a state). Dots on the map are always uniform in size. Unlike other thematic map types, each dot (indicator) on the map represents a defined number of thematic elements for a specific geographic position.

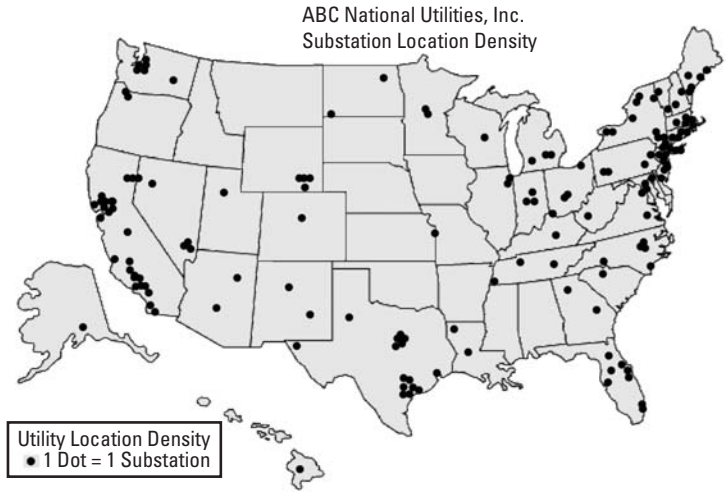


Figure 11.5 Dot density map.

The dot density map portrayed in Figure 11.5 offers an example of a typical theme density map and a perspective on how the thematic densities assist dataset analysis. Just glancing at the map provides the user with important substation concentrations throughout the United States. Ideally, it could help ABC National Utilities understand where it has an abundance of utility facilities and where it has few or none. Each dot represents one substation and provides an approximate position within each state.

Dot density maps are sometimes problematic relative to understanding exact numbers or precise coverage. In areas of abundance, such as the New York City–New Jersey–Connecticut area on Figure 11.5, the dots occupy a good portion of the map space. Since New York City is completely covered with black dots, it can be mistaken as meaning hundreds of substation locations rather than the actual, much smaller amount. In addition, dot density maps could give the wrong impression of an overabundance in a region when the amounts are adequate for the given population and area.

Aside from these few inequalities, dot density maps offer an excellent way to depict attribute densities throughout a geographic region. They are simple to understand and effectively represent concentration patterns and spatial trends across the entire mapped terrain.

Isopleth Maps

The last type of thematic map we will discuss is called the *isopleth map.* Known also as an *isoline map,* isopleth maps visually depict continuous or nearly

continuous data in contiguous patterns of color or shading. Like choropleth maps, values are broken down into data classes and are represented by a color ramp (or user-defined ramps). Isopleth maps connect same data classes together with lines that form the perimeter to same data class areas over the geographic region. The resultant map of same value areas presents a sophisticated view of the spatial patterns in the region, as well as a means to define regional trends.

Due to the manner in which continuous data are represented, isopleth maps are frequently used by scientific, environmental, and meteorological professionals. Typical continuous data mapped with isopleths includes temperature, emissions, precipitation, rainfall, elevation, and contamination. Isopleth mapping can depict a variable data feature within a thematic map or as part of a topographic map (isoline). As an example, Figure 11.6 represents a real-world isoline map created by the U.S. National Atmospheric Deposition Program (NADP) to depict the country's total precipitation in 2004.

In Figure 11.6, the continuous data has no inner boundaries other than the perimeters defined by the explicit data classes. The precipitation trends for the country are well documented in that two "hot" precipitation zones are identified: the southeastern states (i.e., the Louisiana–Alabama region) and the immediate northwest (i.e., the Washington–Oregon region). Because there are 11 colors in the color ramp, particular data points are indicated within the map to facilitate the overall data representation.

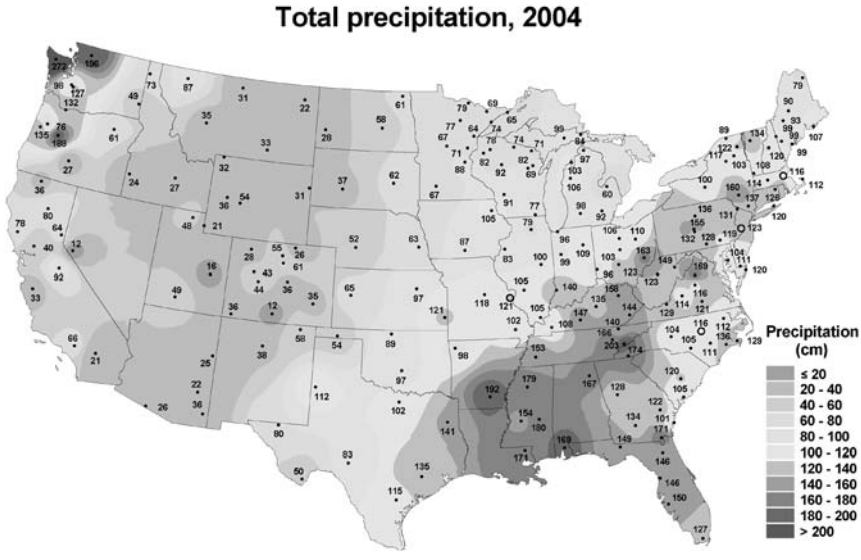


Figure 11.6 An isopleth map of the total U.S. precipitation in 2004. (*Source:* National Atmospheric Deposition Program (NRSP-3), 2005. Reprinted with permission.)

The NADP's isopleth map is used by various professionals to understand not only national precipitation trends but also where precipitation could pose problems. Suppose you were a manager for a major national lumber company and were responsible for picking the location of a new facility. One specification for the facility is that it would need five acres of open land for drying and storing wood. NADP's isopleth map would certainly help minimize the potential effect of rainfall and moisture on the open-air wood storage, and, ultimately, help distinguish good facility locations from poor ones.

Aside from the professional realm, isopleth maps are commonplace in television and published weather reports. Temperature belts and pressure bands are typically illustrated with isopleth maps and high-contrast color ramps (i.e., red means hot and blue means cold). Just imagine how much harder it would be to understand these weather reports if broadcasters used choropleth maps or, even worse, graduated symbol maps. The successful use of isopleths, as with the other map types, is defined by the data being displayed, as well as the anticipated use of the map.

Mapping Data, Classification, Portrayal, and Distinction

Thematic maps, by design, fortify attribute data dissemination and promote analysis not readily obtainable through data tables. These theme-based representations offer highly useful information only if the information being presented is organized logically. Effective thematic mapping involves understanding the data to be presented, a sensible breakdown of that data, selection of a proper map appearance, and well-defined map features. Let us briefly discuss these important thematic mapping elements.

Counts, Rates, and Densities

As discussed, each type of thematic map has its own characteristics as they apply to the data representation. Thematic data can be quantified and represented as numeric counts, rates, and densities. *Counts* are real, whole numbers that represent distinct data entries. Examples of count data include population, number of facilities, and total soil samples. Individual counts are typically represented by a symbol for each item counted. If the counts are too large or exact locations are indeterminable, then one symbol can represent a group of count data (e.g., "one dot represents 100 people").

Rates are a proportion between two counts and are devised by dividing one count by another. Rates are mapped to illustrate how an attribute has changed over time for a specific region or to show the relationship between two

attributes. Rate data includes percent of the total and percent change. Choropleth maps are primarily used for rate data mapping.

Densities are similar to rates in that they show a relationship between two counts; however, densities are intimately related to the geographic region. One count in a density is an area of the geographic region where the second or other count was conducted. Densities include persons per square mile and concentrations per square foot. As with rates, a choropleth map is the superior mapping method for density data.

Data Classification

Data classification can determine whether a thematic map is effective. Data classes were briefly mentioned during the map type discussions, though we must now examine the different methods of determining proper data classes. There are four primary methods for data classification: equal interval, quantile, standard deviation, and natural breaks.

The *equal interval* method is self-implied: It breaks the data into classes of equally distributed intervals. For instance, town population data can be broken into equal intervals of 1,000 people, such as class 1 equals 0 to 1,000, class 2 equals 1,001 to 2,000, and so on. Equal intervals offer an unbiased yet simple way to break represented data.

The *quantile* method divides data into classes having equal amounts of features. For example, if a thematic map had 100 attribute features, you could use five data classes with each representing 20 features. Data are ordered by rank in the classes and, as such, the quantile method proves excellent for maps with ranked values. However, one potential disadvantage to the quantile classification is that it distorts the natural distribution of attribute data and may skew the overall attribute data analysis.

Unlike other data classification methods, there is a defined nomenclature associated with quantile maps. For instance, maps with three classes are called *tertile,* maps with four classes are called *quartile,* maps with five classes are called *quintile,* and so on. Additionally, due to the same number of features per class, quantile classifications often produce visually pleasant maps.

Standard deviation classifications create classes that represent standard deviations from the average (mean) attribute value. Typically, the standard deviation method requires an even number of data classes so that there are an equal number of data classes above and below the mean attribute value. This data classification offers simple comparisons between above-mean and below-mean data. One disadvantage to this classification is that a level of statistical understanding is needed to fully interpret the resultant map.

Finally, the *natural breaks* method recognizes that the breaks in distribution may be directly related to the characteristics of the theme being mapped.

GIS mapping software is able to search for these natural breaks in the charting or histograms of the data using a statistical equation called Jenks method. The natural breaks classification maintains uniformity throughout the entire thematic map and avoids any obviously saturated data classes. For example, a choropleth map with 49 out of the 50 United States in one color offers no real analytic advantage. This method of data classification would find minute differentiations in the data and adjust the data classes accordingly, thus producing a more useful analytical map.

Projection and Scale

The topics of projection and scale have been discussed throughout this book because they are, among other things, functional map essentials. Thematic mapping requires the designer to judiciously select elements and features that promote understanding and efficacy. To do this, map designers utilize both map projections and scale to present the data in the best light, so to speak. As we know, map projections are selected to best highlight the region being mapped through minimizing unwanted distortion. Each projection type has variations that may or may not affect the mapped region. Careful consideration must be applied when selecting the best projection, based on an understanding of the desired properties you want to exhibit.

Working in tandem with a projection is map scale. Depending upon the level of attribute data refinement, map scale may be manipulated to best depict the available data. For instance, a total population count of each U.S. state would be best scaled on a nationwide level, rather than state or county level. However, if population data were available for Michigan's counties, a state-level map of Michigan may be warranted if applicable to the user's mapping objectives.

Both scale and projection are wholly user defined and, if used right, can be significant in transforming a data set into a highly useable geographic information product.

Legends, Color, and Symbols

A legend is a map reference that lists and distinguishes various map colors, lines, shapes, patterns, symbols, and annotation. Often, the legend highlights other map-related information such as the scale, source, and projection. Symbols defined in a map legend are assigned to particular data attributes or themes within the map. Map colors, on the other hand, distinguish between similar legend elements and symbols (e.g., colors can distinguish different soil characteristics). Legends, colors, and symbols, when used together, give the user total

disclosure of the relevant information needed to discern spatial trends within the data set.

An interesting and important aspect of these mapping elements is that they compose the geographic product's metadata. Back in Chapter 4, metadata was discussed at length. The map legend elements, such as map scale, title, and classification, make up the metadata used to catalog the geographic product within a GIS or other database management system. The more information provided in the map's legend, the greater the document's access and searchable data.

Figure 11.7 offers a sample thematic map legend from a wetlands inventory. Notice that there are very distinct elements within the wetlands inventory legend that define distinct forms of data within the map. We have the map title and/or identifier that label the map for the user and distinguishes the map's purpose. Under "Area Transportation" there are different symbols used for various transportation-type elements. For instance, a thick multishaded line is used for an interstate highway, solid colored lines are used for roadway features, and roadway sign symbols are used for roadway identification. You should also recognize how different colors are used to differentiate between the various features (i.e., state road, streams, and so on). Under "Wetlands" in the legend, color fills distinguish attributes among the various wetland types.

Figure 11.7 also highlights other important map elements such as the county identifier, map scale, map number in the series, and date of creation. These tidbits of information add to the user's scope of understanding and enable universal use.

In short, the sole purpose for legends, color, and symbols is to enable anyone with any background to pick up a map and understand, at least on a minimal level, what information the map is conveying.

Common Cartographic Elements

Before closing the topic of thematic mapping, we must examine the cartographic elements that frequent and add dimension to maps. Thematic maps have two basic separations: the data frame and the map information frame. The data frame is the actual thematic map with themes overlying a basemap. The map information frame is the area of the information product that contains cartographic elements. Figure 11.8 offers an advanced view of the most common map elements

The elements depicted in Figure 11.8 are typical features that are found alongside or below a thematic map (data frame). Since proper map identification is needed for user comprehension, thematic maps utilize a map title block and a more detailed map identifier. The map's title block typically contains the company responsible for the map, as well as the type of thematic map (e.g., a

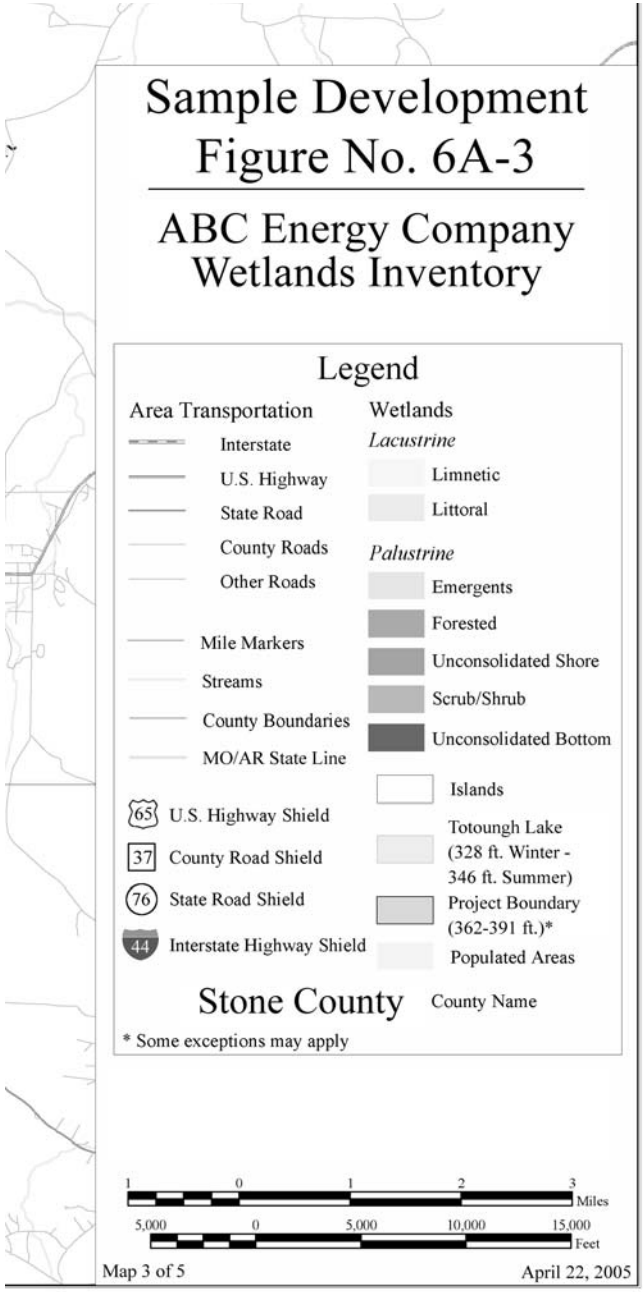


Figure 11.7 A sample thematic legend.

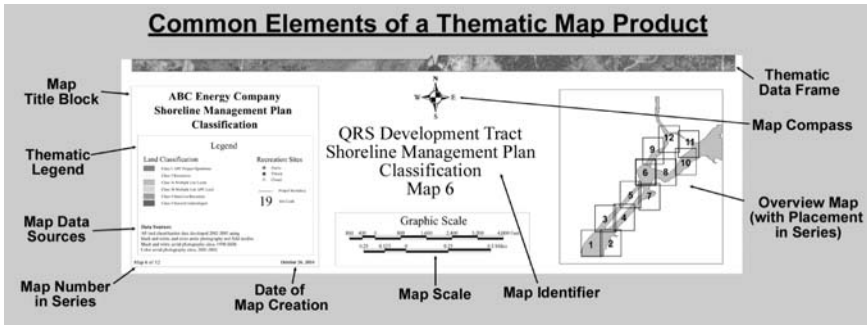


Figure 11.8 Common elements of a thematic map product.

shoreline management plan classification). The map identifier highlights the land involved, the nature of the theme involved, and the map's position (number) within the grouping of area drawings.

As discussed in the previous section, a common element found on a thematic map is a legend. The map legend offers the user a pure "nuts and bolts" understanding of the map's thematic phenomenon. The figure's legend contains an assortment of data class information, with color fills used for land classification and dot/line symbols for recreational sites.

The map's legend may also contain more background information, such as data sources, a textual description of the map's number in the entire series (if applicable), and the official date of creation. This background information will serve as a key component in useful information products.

Any cartographic product utilizes both a map scale and a map compass. The map scale, as previously discussed, helps present the attribute data in as much detail as possible. In fact, scales can be manipulated to portray a larger region of focus, as well as aid in the recognition of actual feature size. The map compass has a simple function: It signifies the true direction of north. Often, the map compass is just an arrow pointing north, while other times (as in Figure 11.8), the map compass is more elaborate.

Another common cartographic element employed is the overview map, which graphically illustrates the individual map's placement among the entire series of maps. For instance, the ABC Energy Company's map was map 6 in a 12-map series. The overview map shows the arrangement of the maps, with map 6 highlighted. As you can imagine, this map element supports user understanding of the region depicted in the data frame and offers insight into how it relates to both the landscape and the other maps.

Overview maps enable individual maps to retain a high level of detail and portray the focus region in an easily viewable format. Just think how ineffective

and small featured a thematic map would be if it had to represent all 12 maps in one data frame.

Overall, no matter which cartographic layout and elements are used to describe the attribute data set, the mapping must be consistent, well detailed, and easy to read. These GIS end products are ultimately used by interested people engaged in a solution. It is always important to remember that when the thematic maps are of good quality, the GIS and support efforts are held in high regard. An ineffective or poorly executed thematic map will, subsequently, mar geographic information efforts and vandalize the data credibility of the GIS.

12

GIScience, Engineering, and Related Applications

Up until now we have discussed geographic information systems on a basic level, describing the *what* of GIS with some of the fundamental elements that make the spatial technology tick. Now we must tackle the reasons why a GIS would ever be needed on a project and the advanced, innovative ways GIS is being used today.

As stated earlier, GIS is used in many different industries for a wide range of reasons. This chapter, as well as the next two chapters, attempts to clarify some of the *whys* of GIS with a compendium of different and viable real-life applications.

Using GIS for Technical Applications

In *The Quest for Certainty*, John Dewey, a prominent American philosopher and educator, wrote, "Every great advance in science has issued from a new audacity of imagination." GIS is clearly such an "audacious" technology. Within its short existence, GIS has proved itself an excellent analytical tool and a leading pathway to innovation. GIS's architecture, unlike any enterprise database system, is well designed for the widely diverse implementation field.

Even more dramatic is GIS's usefulness in highly technical and specialized arenas, such as the sciences, engineering, and technology. With these and other technical applications, GIS has helped companies stay one step ahead of competition, while emblazoning their technological contributions and advances. Many of the world's largest and most successful technological companies and agencies use GIS as a means to centralize their information, automate their analytical processes, and analyze the current and future markets.

In tandem with the scientific and engineering advances, GIS helps companies enhance their business operations from financial endeavors to marketing and potential use analysis. GIS's inherited automation offers the opportunity for higher efficiency over time while keeping procedures consistent and information available. The following discussions cover general science and engineering applications, while offering some perspective on the business end of the technology.

GIScience: Principles and Applications

Geographic information science (known as GIScience) is the study of geographic concepts, such as places, things, and phenomena on the Earth's surface, using GISs. GIScience is concerned not only with where objects are on the Earth's surface but also what these objects are and their relationship to other entities. Through GIScience, GIS becomes both an information technology and an Earth science. With its innovative software technologies, GIS showcases information about the Earth's surface (and other celestial bodies) with geographic data and, in turn, offers the visually gratifying form of geographic features. These features supply the specific knowledge about where things are in the world (their locations) as well as what they are (their attributes).

GIScience also provides nearly effortless analytic methods for describing spatial autocorrelation, or the spatial relationships between geographic features. Geographic data are diverse in character. Some data sets are very detailed and contain volumes of attribute information, while others are restricted and modest in size. All of these elements define the broad interpretation of GIScience.

Though GIS is most commonly classified as a computer or physical science, GIScience opens the gates for the infusion of sociology, the empirical social sciences, and history. GIScience is now used for geoarcheology, psychological analysis, social services, and historical analysis. Additionally, this trend of interdisciplinary applicability is most closely echoed by the field of geostatistics, a discipline offering a synthesis between two disciplines: traditional statistics and geospatial interfaces.

Make no mistake, the definition of GIScience is very broad and the proverbial ground it covers will grow exponentially as object-oriented geospatial functions and techniques become ever more important.

Case Study: Using GIS to Study Relationships Between Bedrock Geology and Subtle Topography

GIS plays a vital role in the geosciences and is increasingly becoming a source for revolutionary scientific analysis. One such study involved the definitive quantification of relationships between bedrock geology and topography, where the

bedrock contained sandstone and shale, and the topography involved a subtle but well-defined surface. The study area was in Oklahoma in the United States and contained geology fractured by area drainage. The researchers used GIS, importing a digitized basemap of Oklahoma's bedrock geology and a topographical map. Unfortunately, topographical maps prove difficult for detailed evaluation of landforms over large areas and small reliefs, or elevation differentials. The researchers decided to use a digital elevation model (DEM) for a more precise analysis of the study area's subtle topography.

Ultimately, by using GIS the researchers were able to distinguish that the study area's topographic variations rely heavily upon the ratio of sand to shale in the bedrock in areas where sand overwhelmingly exceeds shale. Sandstone was found to be thicker in higher relief areas and thinner in lower relief areas. Additionally, the researchers found that there was no real difference in geology bulk density within these variant regions. Without GIS, these relationship findings would not have been made or determined easily. GIS enabled the researchers to easily quantify the subtle differences in local topography within the different bedrock geologic units—an onerous feat using conventional field techniques. The study has afforded future research and quantifiable topographic studies a viable basis for better land use and an enhanced understanding of subtle topographies.

Engineering Applications

As mentioned previously, the field of GIScience is broad in its interpretation and crosses borders with other applicable disciplines, such as the environment, conservation, mathematics, and engineering. More specifically, the physical sciences have long shared common ground with the engineering discipline in such areas as natural resource design, feasibility studies, and environmental impacts from engineering applications.

Similar to GIScience, the engineering industry has maintained its own history of geographic techniques. From the early days to the present, drawn maps and plans have played a huge role throughout the wide stretch of engineering applications. Amongst the civil, structural, environmental, transportation, industrial, aerospace, mechanical, and electrical industries, for instance, engineers regularly use site maps and plans to depict the proposed scope of work, particular designs, and the proper implementation of such designs in a defined geographic space. Old hand-drawn prints and layout designs showcased early geographic mapping techniques perfectly and provided engineers with a planning foundation. These painstaking hand drawings became extinct with computers and computer-aided design programs. Today, engineers are supplementing CAD with the dynamic drawing accessories offered by GIS.

Today's geospatial standards in the engineering industry are actually inherited from the days of plan views in black pen on a Mylar transparency, displayed through an overhead projector in a dimmed conference room. Black pen markings on the site plan can now not only be accurately scaled and drawn using CAD but can also be geospatially defined in real time and geocoded within GIS. The engineering industry has readily embraced this conjoining of CAD with GIS and has realized GIS's positive impact on company bottom lines, project budgets, quality assurance, safety, and the project schedule.

But GIS goes well beyond engineering site plans and geographic products; engineers use GIS for various engineering applications, such as property delineations, conduit/transmission runs, pipeline siting, marine science, surveying, engineering scenario planning, terrain analysis, and traffic accident research. GISs, for example, help utility companies define their electrical substation placement among distribution networks over entire coverage areas. As another example, oil companies use GIS to pinpoint offshore drilling locations, hyperlinking drilling reports and detailing product quantities for these particular geographic locations.

Some of the more effective engineering applications are oil and water pipelines. The pipeline industry's use of GIS is highly progressive and embraces a wide range of disciplines and engineering methods. Typical pipeline projects involve, to some degree, surveying, GPS and remote sensing, pipeline siting, vegetation mapping, wetlands delineation, natural river/waterway channel analysis, parcel delineation, slope analysis, and defining existing pipe locations. In fact, the pipeline industry has moved in the direction of *trenchless mapping,* or full data collection and scenario planning before construction starts. GIS proves exceptionally advantageous in gathering all data, defining construction scenarios, conducting analysis, and detailing the entire project in one effort without breaking ground. GIS-based trenchless mapping helps the pipeline constructors understand the situations they will face while minimizing expensive project changes, enhancing their public outreach programs, and keeping the risk for litigation at bay.

We could review each engineering discipline and highlight literally hundreds of valid, real-life applications where GIS is used and has made a positive difference. There are many available publications and periodicals that solely detail variant and innovative applications. For this discussion we will take a look at one specific application, New Jersey Natural Gas' Transmission Line Parcel Delineation, whereby the critical GIS functions used are representative for many engineering and utility projects.

Case Study: New Jersey Natural Gas Transmission Line Property Delineation

To prepare for law-mandated public and emergency notifications, New Jersey Natural Gas (NJNG), a utility provider in central New Jersey, required an

accurate record of existing properties within 700 feet of its underground gas transmission lines. These central gas lines service a massive area that covers Ocean, Middlesex, Passaic, and Morris counties. NJNG quickly realized the absolute complexity of such a large-scale property (parcel) delineation effort. Accordingly, the New Jersey utility turned to a local geospatial and engineering firm, T&M Associates, Inc., to manipulate the existing data and develop the necessary GIS backbone for the project.

The geospatial system was built upon ESRI's ArcGIS 9 environment. Existing digital parcel information was initially obtained from the state and other sources and imported into the system. Original tax Mylar sheets were scanned into a raster image and consistently referenced with existing tax file naming conventions. This scanned raster image was then rectified with New Jersey's high-resolution orthoimagery (from 2002) and digitized. The digitalization process defined lot lines, boundary lines, and roadways and streams. The outer edges of primary building rooflines (or *footprints*) within the project area were also delineated. The lots created from individual tax sheets were saved as a merged geodatabase file, with adjacent sheets edge matched as necessary. A best-fit approach was used to create the overall parcel layer.

Finally, the parcel layer and subsequent feature attribute table was linked with New Jersey's tax list database, called MOD IV. Unique identifiers and quality assurance protocols were used to facilitate the joining of information, preventing data duplication and parcel ownership errors. In the end, the parcels 700 feet to each side of the transmission lines were successfully delineated and a base GIS was developed for future NJNG efforts.

Marketing and Financial Applications

Aside from the purely technical aspects of the engineering and scientific communities resides the ever-important business bottom line. No matter how advanced a company is or how many innovations a firm has brought to the market, bottom line profitability is the true make-or-break for any business venture. To stay successful, business managers must understand the needs of their market while staying abreast of new potentials. These endeavors take resources: people, time, and money.

The keystones to business bottom line profitability and industry success prominently include business marketing and financial stability. To help managers with their market and financial analysis, business managers are reaching out to new technology and industry trends (i.e., Six Sigma quality discussed previously). In recent years, GIS has played a huge role in conducting valid business sustainability analysis while offering managers an amazing tool for understanding business shortcomings.

GIS offers business marketers the valuable capability of real-time market analysis and customer profiling. Customers can be researched through demographics and purchase histories so that marketing can be redirected (if needed) and the precise customer reached. Some firms advertise GIS-based marketing services that are geared for medium to large businesses of any venue. Internal marketing departments have teamed with in-house or outside GIS consultants to target their market campaigns. In the end, this business marketing strategy has a complementary effect on the fiscal end of the business.

Business financial departments are using GIS to model sales patterns and market histories while aiding in the discovery of profit model inefficiencies. GIS can help accountants prepare or conduct company audits, as well as handle the overwhelming requirements of state- or country-imposed fiduciary regulations. Sometimes, GIS is used for combined marketing and financial analysis when product areas or markets have drastically decreased in sales or new services have proved unprofitable. Commonly, marketing and finance go hand in hand to keep businesses profitable and in touch with the markets. The following case study melds both business ends together through the investigation and interpretation of a rapid chain store closure.

Case Study: Analyzing the Rapid Closure of a Cincinnati-Based Office Depot

In mid-2001, Office Depot, a major chain store for office supplies and equipment, opened a new branch in an eastern suburb of Cincinnati, Illinois. After just six months of operation, the new store closed. Was it a failure in marketing? Was competition surrounding the store too harsh? How could a formidably successful chain have such a fiscal miscalculation? These questions prompted a retail marketing study using GIS analysis tools to offer insight as to the reasons for the store closure.

The branch in question (Beechmont Avenue) was a sister store to five other Office Depots and had competition from five Office Max stores and nine Staples stores in the greater Cincinnati area. The researchers decided to focus on geographic location, since location impacts sales potential and the closure of the Beechmont Avenue branch was related to its low sales potential. The researchers developed a sales potential model within GIS based upon a retail industry standard (D. L. Huff's sales potential model) to evaluate the store's overall performance.

GIS analysis constituted the known spatial distribution and spending habits of potential customers in coincidence with the store's overall highway accessibility, square footage size, approximate travel time, and distance. These characteristics were diagnosed against preopening and postopening data from the Beechmont store, the other greater Cincinnati–based Office Depots, and the area competitors. Through GIS analytical tools and the Huff-based sales model, the

researchers were able to ascertain that one reason the store was underperforming was poor customer accessibility. Closing the branch may have been more cost effective than keeping the failing store open, although as noted in the study, if this type of GIS-based analysis had been conducted before opening the branch, management would have saved a great deal of time and money.

13

GIS and the Environment

Since the world's conscience about the Earth's resources began to peak in the mid-1970s, there has been increasing public concern and interest in environmental issues. Lead-based paint, asbestos-containing materials, PCBs, mercury, underground storage tanks, and chlorofluorocarbons (CFCs) have been just a few of the many environmental issues holding public scrutiny throughout the years. However, environmental concerns migrate well beyond hazardous materials and manufactured contaminants. In fact, a cavalcade of public and governmental issues have addressed environmental conservation and those arenas governed by Mother Nature. Animal habitat preservation, deforestation, endangered species, geology, floodplain analysis, water resources, and mining are just a handful of the myriad areas where the public, environmental agencies, and nature convene.

With this ever-growing and substantial public inquiry, state and governmental agencies need accurate information at their immediate disposal that can be quickly discriminated, analyzed, and conveyed to the mass public. To this extent, GIS has become the tool for both the field scientist and the desktop environmentalist.

Using GIS for Environment Applications

Environmental issues became vividly clear to the public in August 1978 when the small, populated community near Niagara Falls, New York, filled headlines. The community was called Love Canal. Hidden underneath this 15-acre neighborhood was a former chemical landfill. After reaching maximum capacity in 1953, the landfill was closed and covered under layers of dirt. Buried within the

soil were metal drums that, throughout the years, began to rot, causing hazardous chemicals and wastes to leak into the soil. By the mid-1978, the leaking drums became horribly evident when hazardous wastes began percolating to the surface of Love Canal's backyards and cellars. It was later reported (in November) that 200 tons of dioxin, a lethal chemical, and more than 200 chemical compounds were buried in the canal.

In 1978, the Love Canal neighborhood included approximately 800 single-family homes and 240 low-income apartments. Three schools also graced the Love Canal parcel, of which the 99th Street Elementary School was built almost directly above the former landfill's epicenter. A posttoxic release study revealed a high rate of birth defects and miscarriages among Love Canal families throughout the years and offered documented claims of birth deformities and pregnancy-related problems within the community. Ultimately, the Love Canal incident remains one of America's worst environmental disasters and one of America's most remembered tragedies.

Beginning in the 1990s, GIS has grown as a chief facilitator for major environmental efforts. This technical evolution was inevitable since most environmental efforts are deeply entrenched within the fabric of the land and hold geographic importance. GIS is used for hazardous material issues and natural occurrences alike and has now become the popular instrument for land disputes and owner delineation lines. Researchers are now applying the tool to past contamination for further analysis and understanding. In fact, the U.S. Environmental Protection Agency (USEPA) has reprised analysis on the Love Canal incident using GIS so that environmentalists, agencies, scientists, the public, and the victims have a heightened clarity as to what really happened and a sharpened view of every angle of which people were affected.

The following sections detail common GIS applications as they relate to environmental issues and conservation, though in theory GIS can be used for analysis in many environmental situations. Although cost and feasibility are typically the prohibitive factors for GIS in many environmental decisions, when GIS is applied, the results are sound, intrinsically useable, and many times astounding.

Managing Natural Resources

Natural resources are said to be the Earth's wondrous gifts to humanity. The soil upon which we walk, the vegetation we utilize, the animals with whom we coexist, and the water we drink are all innate elements to life and the recognized gifts of our planet. Natural resources can be considered the prize possession of all humans, although, throughout the ages, humans have neglected to care for these gifts.

There is not a day that goes by without our hearing or seeing examples of human neglect, such as the greenhouse effect, air pollution, toxic emissions, hazardous waste, oil spills, corroding shorelines, and endangered species. During the past two decades, a new surge of environmental conservation of the Earth's resources has surfaced and caught the eye of the public. The management of these natural resources has become the breeding ground for conservationists, public outcry, activism, and lawsuits. This surge has evoked many ideological and creative uses for GIS.

Natural resources encompass rivers and streams, lakes and inlets, landforms, geology, forests and woodlands, mines, wildlife habitats, and ecosystems. The management of these natural earthly gifts involves a deep understanding of each system and a proactive plan for these resources to remain unharmed. Natural resource management covers a large platform for potential analysis, from reclaiming brownfields and facilitating wildlife habitat protection to defining oil spills and delineating land classification. GIS is used by conservationists of all types to define potential infringements upon the resources from nature, as well as any residential, commercial, or industrial imposers. It helps form a stable cadre of geographic information and analysis from which to solidify environmental campaigns or to cast the proverbial stone. Conservationists look toward GIS for the position they need in the fight against combatant corporations, developers, and residents.

But GIS doesn't satisfy just the conservationists; it holds valid analytical and environmental defense support for those on the other end of the conservation fight (i.e., the corporations, developers, and residents). For instance, GIS may be used by a developer to show the self-benefiting, attractive results of little or no intrusion against the natural aesthetics of a lake. Dam owners may use GIS to show how a dam removal would positively affect the area, while not disclosing the owner's high-yielding advantages. Whatever side you are on, GIS proves to be a magnificent propaganda tool.

Many times GIS is used toward natural resources when there is no head-to-head fighting at all, but just as a way to quell rising public concern or help with preemptive actions as to appropriate for the specific natural resources. Energy companies tend to use GIS to explore the impacts of operational by-products, such as the impact of rising waters on wildlife, and uses analysis from the GIS products to determine if actions are warranted.

Case Study: Lebanon Natural Resources Management with Participatory GIS

Entrenched within the Mediterranean, Lebanon boasts a diverse socioeconomic society throughout its relatively small population (4.1 million people). Remarkably, Lebanon is a war-torn land. By 1991, 17 years of war ended, leaving Lebanon overwhelmingly disabled. Its once profitable agricultural industry was

deteriorating, and the country looked toward its natural resources for help. However, after years of brutal war, Lebanon's resources were unmanaged and in disrepair.

The Environmental and Sustainable Development Unit (ESDU) at the American University in Beirut took notice and began a series of research programs that meshed technology with community support. The ESDU developed the "participatory GIS" or PGIS, whereby the community provided participatory research, land data, and information for a GIS. The program approach proved highly successful, evidenced by the fact that the ESDU utilized PGIS from 1995 to 2003.

A participatory GIS is a fresh approach to getting relatively accurate measurements and geodata for areas that are hard to measure. The nature of participatory GIS offers many incentives to native volunteers to measure and register land-based and geological features. Such incentives include comprehensive area resource identification and profitable agricultural analysis. In exchange for these incentives is the ability for a highly customized and unique data set. Participatory GISs are gaining popularity in underprivileged communities worldwide and within the GIS community.

Lebanon used PGIS for various natural resource management applications, such as to delineate indigenous agro-ecological zones, plan orchard development, delineate poorly or inappropriately used land, plan rainwater harvesting reservoirs, and develop georeferenced databases for the study area.

Ultimately, PGIS served Lebanon well, providing accurate georeferenced land information that has proved invaluable to researchers, developers, communities, and land stakeholders alike. The participatory GIS information products have given Lebanon a new lease on their natural resources and a feasible approach for socioeconomic growth.

Case Study: Community-Based Natural Resource Management Program in Namibia

In Namibia, Africa, GISs are playing an important role in supporting community decisions as they relate to the management and use of Namibia's natural resources. Namibia's communities are using GIS information products, particularly printed geographic maps, to manage the natural resources on an individual community level. Namibia's Community-Based Natural Resource Management (CBNRM) program involves participants in every community to encourage the proliferation of sustainable development. Namibia's community-based program is very similar to Lebanon's participatory GIS program, but with a heavier focus on natural resources.

The CBNRM's use of geographic data and information is distinct, primarily land-use planning, resource conservation, game and wildlife monitoring, tourist and commercial development, and communication. One of the chief

activities threaded throughout the various GIS uses is boundary delineation, especially for conservancy and land development.

Namibia's community GIS program relies heavily on an intercommunity sharing of geographic data, information products, and tools. The program runs on a needs-based approach and involves dedicated community field users who trust the accuracy of the GIS information, have fostered a set process to delineate community needs, and have built a reservoir of information to ensure the program's longevity.

Environmental Investigation and Remediation Applications

A continuously growing area for intensive GIS use is in the hazardous waste arena. With the uprising in brownfield redevelopment, environmental impact statements, building toxics, and hazardous material remediation, GIS has become the premier project database, saving companies and agencies substantial amounts of money over long-term projects.

Interestingly enough, the USEPA is the largest user of geospatial information as it relates to the environment. The USEPA Web site, http://www.epa.gov, boasts a fully hyperlinked and geospatially coded (or *geocoded*) mapping tool, whereby any user can search a U.S. location and retrieve graphically enhanced maps depicting all recorded environmental concerns. The Web site provides a wonderful example of GIS use in an environmental application.

Hazardous materials come in many forms and can infiltrate the human ecosystem in an innumerable number of ways. Just a sampling of the possible environmental episodes include leaking underground storage tanks; toxic releases; the presence of volatile organic compounds (VOCs) and semi-volatile organic compounds (SVOCs) in soils; and micro-fiber silica and polycyclic aromatic hydrocarbons (PAHs) in the air. As with the USEPA site, GIS can and is being used to geographically reference release locations for various agencies and private firms.

GIS has made an indelible mark on the way sites are remediated and monitored. In particular, heavily contaminated sites such as the USEPA superfund sites and chemical dump sites use GIS to track soil boring locations, with hyperlinked sampling results. For environmental site assessments, GIS is used to organize Sanborn maps (historical site characteristic maps) that detail site history, to depict locations of potential contamination, to hyperlink sample analysis, and to represent underground structures, such as storage tanks, and known soil contamination.

On asbestos and lead-based paint contaminated sites, GIS is sometimes used for large contaminated buildings and facilities, whereby asbestos "hot spots" and known lead-based paint locations are meticulously defined in the

system. Sample analysis, air-monitoring results, abatement information, and containment locations are delineated, and sample chain-of-custody information is accurately tracked.

Case Study: Illinois Underground Storage Tank Monitoring and Inspection Program

Within the United States, Illinois has more than 28,000 facilities with underground storage tanks (USTs). These facilities range from industrial to commercial uses, though many are simply gas stations. USTs hold petroleum or other forms of fuel in metal-encased tanks. These systems are well built, but after years of neglect and metal shell deterioration, fuel can leak from the storage tanks into the soil. This causes severe human health issues. Every state in the United States requires owners to report USTs.

Illinois needed a comprehensive yet efficient way to manage the locations and information associated with its USTs statewide. Many former gas station sites still have USTs below ground with newer building structures above. These USTs are of greatest concern since they may not have been drained and closed properly, or residual fuel may have leaked into the soil during tank closure. Fuel-laden soil creates VOCs and SVOCs that are known carcinogens. To help accurately monitor and inspect USTs throughout the state, Illinois's Office of the State Fire Marshal (OSFM) turned to GIS.

The use of GIS virtually reinvented the way OSFM handled USTs, from the everyday monitoring and inspection to the data retrieval and data quality. Before GIS, the field inspectors carried boxes of files with them and had to complete nearly 20 different forms to collect the necessary tank inspection information. They then had to mail these forms to OSFM headquarters for processing and entry into the database. Now that GIS has been implemented, the field inspectors need only their laptop computer and GPS receiver. Tank information is entered directly into the computer and the once-tedious forms are quickly completed. Sites get geocoded with the collected information, forms, and GPS location data. Field inspectors can create maps to locate tanks and facilities, as well as perform spatial queries on facilities. In the end, GIS has made Illinois's UST program highly efficient, flexible, and accurate.

Case Study: The USEPA's Disaster Dioxin Air-Monitoring GIS System

After the September 11 terrorist attacks on the World Trade Center in New York City, the world became horribly familiar with air carcinogens and how latent environmental hazards can become time bombs to human health. With the Trade Center disaster, large, concentrated amounts of dust were propelled throughout the Lower Manhattan area, reaching upper Manhattan and parts of Brooklyn, Queens, and Staten Island. Within the fine dust were varying levels of

carcinogens such as asbestos and silica fibers. In Lower Manhattan there were many other elements of potential concern, such as fiberglass, mercury, PAHs, PCBs, lead dust, dioxins, and trace metals. The collapse of the building structures caused the material to granulate and become extremely fine dust, an ingestible state that proves most dangerous.

The USEPA used GIS to assist in the analysis of the laboratory results and the field samples. The USEPA set up fixed air-monitoring stations throughout New York City, with a concentration of stations in Lower Manhattan. These stations produced results that were fed into the GIS, from which environmental exposure trends were determined. The GIS information products were used to disseminate accurate, up-to-the-moment results to the public through USEPA's Web site and media presentations.

The GIS products served the numerous environmental agencies and scientists with an accurate geodatabase of sampling results that were at their proverbial fingertips, while offering the nonenvironmentalists a visual depiction of the concentrations and areas where environmental exposure was greatest. The USEPA response was a marvel in the way geographic information and environmental analysis can collaborate to serve a wide range of applications for a much greater array of users.

Shoreline Management Systems

GIS use with shoreline management has provided owners and conservationists a stable repository of geographic landform data and a tool to provide environmental impact analysis. A shoreline management system is a proactive structure by which a shoreline is planned, managed, and/or delineated, with potential human and natural encroachers easily determined. Coastal planning, shoreline planning, and wetlands management are popular topics for a wide range of individuals, including developers, environmentalists, regulators, and lawyers. All three shoreline systems are linked in scope and audience, as well as through their proclivity for dispute and participant division. GIS helps all participants understand the overall scope of the effort in dispute and offers GIS information products (such as maps and land classification charts) to help end these disputes through information.

Environmental regulators use GIS to legally validate their shoreline and coastal management planning efforts, while conservationists use GIS to aid their efforts to preserve natural areas. Developers typically use GIS to further communicate their project's scope to residents and community groups while facilitating the evaluation of the potential area impacts.

GIS information products result from careful shoreline management planning efforts that can lend tremendous credence to any environmental

management undertaking and potentially influence political opinions or legal decisions. You might be asking yourself, "How does GIS really do this?" GIS increases the visibility of the effort and showcases facts to substantiate claims, ideas, or complaints. The ability of GIS to utilize plentiful, accurate, and freely available geographic data can be the "equalizer" in big business or bureaucratic environmental litigation. Any users with modest GIS knowledge and desktop computer proficiency can build a system to "make a case" for their shoreline, coastal, or wetlands management cause.

In the United States alone, shoreline, wetland, and coastal management planning is truly big business. Management planning is backed by the government through enactments to state and federal law, including the USEPA's Coastal Protection and Management Act of 1995 and Coastal Protection and Management Regulation 2003, the U.S. Geologic Survey's Shoreline Management Act of 1971, and the USEPA's Wetlands and Watershed Act of 1997.

Case Study: The Coosa/Warrior Shoreline Management GIS—Alabama Power Co.

As a direct result of the U.S. Federal Power Act of 1935, the Federal Energy Regulatory Commission (FERC) was granted the right to regulate nonfederal hydroelectric projects. Owners of hydroelectric facilities must periodically relicense their facilities with the government in order to reap governmental aid and receive revenue from their hydroelectric enterprise. These relicensing efforts are time consuming and costly and must meet FERC and local agency requirements. Requirements are strict, cover various areas of importance, and usually involve the management and planning of impacted shorelines and owner-occupied coasts.

As part of Alabama Power Company's relicensing efforts, a geospatial database was developed to help facilitate the relicensing of seven major hydroelectric facilities on the Coosa and Warrior watersheds in Alabama and Georgia. The project involved spatial database design, spatial analysis, image processing, environmental shoreline management, and land classification.

A complete Shoreline Management Plan (SMP) system was developed for Alabama Power's work at the sites, classifying more than 3,000 miles of shoreline and project areas. With the application of GIS, the SMP mapping effort consisted of more than 150 map layouts that were developed for public presentation and project team analysis.

Other Environmental Applications

The list of geospatial environmental applications is far more extensive than the popular areas discussed in this chapter. You can say that the list is virtually ever

growing. Through user creativity and out-of-the-box thinking, GIS has reached the small corners of the environmental industry and instilled renewal into the traditional project management controls. Areas such as pollution prevention, groundwater modeling, environmental compliance, and water resources (among others) have gained new life through GIS and as a result have promulgated higher efficiency rates for technical analysis and the dissemination of information.

By its very nature, GIS is fully migratory within the industries and, as such, has shifted into firm practice worldwide. Whether because of its power of efficiency or its sheer management capabilities, GIS has proved itself an incomparable project tool. Realizing this, many large consulting firms and organizations now have in-house GIS capability and thus offer geospatially based solutions as part of their core environmental offerings.

14

GIS and the Public

Since the advent of the Internet, more and more people, whether they know it or not, use GISs on a frequent basis. Although the Internet and GIS are separate systems, some Internet sites offer GIS-processed information, such as turn-by-turn driving directions or up-to-the-moment weather information. The public has the ability to search neighborhoods for health and safety hazards, search communities for public notifications and alerts, or even find out the air quality index for that specific day in any defined location. The general public relies on these sites for free, quick information, though the users rarely understand that the information they are manipulating is supplied from a GIS or that they comprise the majority of Internet-supported GIS users.

The general public is also the largest group uniquely affected by GIS technology. Information systems have infiltrated everyday life, from tangible products, such as news and home electricity, to abstract products, such as emergency 911 system routing and port vessel tracking. The reality is that GIS and the public are intertwined, and as the technology grows in use, the number of ways one person is affected by GIS becomes incalculable.

A Public Need for Information

Without question, it is vital that the general public have a source or group of sources that provide accurate information. A means to easily distribute geographic information to the public can sometimes mean life and death. For instance, during the September 11 attacks in New York City, the NYC Police Department used GIS to maintain public safety, control Lower Manhattan access, and disseminate community information. Through various state and

federal Web sites, the public had the comprehensive ability to access area maps, timetables for redirected traffic patterns, rescheduled public transportation, information about community meetings, and results of air quality samples. This intrinsic use of GIS helped maintain control of a very unstable situation while keeping the public well informed. During the emergency, the public demanded accurate information quickly and continuously, and a GIS proved to be the most comprehensive and easily implemented of solutions.

This need goes much further than just community safety, disaster, and emergency situations; the need resonates throughout public schools systems, public transportation systems, and within the policies and regulations that govern facilitation. To this end, GIS has infiltrated the populace as a subtle but powerful information medium. It can be suggested that the immediacy of GIS information products, in tandem with an Internet delivery, makes daily newspapers and the evening news seem like "pony express" reporting.

Nevertheless, immediacy is not always the main propellant for GIS use. Many times the public user requires the power of the information database rather than data immediacy, since it offers immense public documentation catered to a specific application. For instance, school principals can monitor crime statistics in a school's district, and social service providers can tap into local demographics and social structures. These types of public applications necessitate accurate, time-nurtured data rather than up-to-the-moment data.

Similar to the other disciplines of GIS (i.e., education, business management, environmental, and so forth), public praxis is evolving and is limited only by user creativity. The following sections highlight typical applications for the public's use of GIS and information management. As you will quickly realize, the general public comprises a very diverse suite of users that viewed macroscopically is wide ranging and viewed microscopically is highly user particular.

Managing Public Health and Safety

Public health and safety have always been subjects of debate, interest, and downright concern. From the early days when there was minimal medical knowledge and no technology, people with a communicable disease such as tuberculosis could infect the neighborhoods or towns where they lived. News of a new case would send fear racing throughout the communities. Neighbors would inform others of new cases. People fearful of infection would stay indoors or somehow try to hide from the disease. Invariably, others would be infected. The end limits to any disease management efforts would be to either quarantine and treat the infected or wait for their death.

Nowadays, medical knowledge is painstakingly sophisticated, and the once dreaded communicable diseases are fervently managed by professionals. The

public fear of such community-infectious diseases has tapered to mere interest. However, the concern about deadly new outbreaks has not diminished. As in the days of old, public information remains the first and best barrier to a total outbreak, and new technology provides a second barrier for professional analysis and understanding. As you can imagine, GIS plays an integral part within both barriers and the meticulous spatial tracking and management of disease.

Disease management is not the only public health and safety issue where GIS plays a management role. Scientists and medical professionals use GIS to spatially analyze areas for air contaminants and toxins, higher rates of miscarriage and birth defects, spikes in asthma and other respiratory stressors, proficiency in child inoculations, increased worker safety concerns, and occurrence of repetitive motion injuries, just to name a few. GIS aids in public awareness programs, offering accurate information online while providing health and safety analysts a compendium of source data.

The process of managing public health and safety is imperative to lessen human loss or injury when outbreaks and emergencies occur or as an ongoing medical prevention program. GIS is used to inform the public of "hot spots," or places to avoid, while indicating information for preventive measures and locations for proper treatment. Many governmental and state public health agencies use GIS technology to identify or defuse a potential health safety threat, conduct communal human health risk assessments, make better decisions in terms of public safety, or just stay ahead of area health trends. The superfluity of demographics, client data, and resource distribution is among the greatest reasons why a GIS is used.

Case Study: Detailing Areas at High Risk for Elderly Fall Injuries to Enhance Injury Prevention Programs in Alberta, Canada

Researchers from the University of Alberta's Division of Emergency Medicine, the Health Surveillance Branch of Alberta Health and Wellness, and the British Columbia Rural and Remote Health Research Institute conducted an investigation detailing the geography of fall injuries in the elderly within Canada's Capital Health Region in Alberta. The researchers used GIS to analyze the spatial character of fall injuries within the study area in an effort to enhance fall prevention methodologies. The data set for this particular study comprised Alberta residents greater than 66 years of age who visited an emergency department for a fall. Postal codes from emergency vehicle and department reports comprised valid location data.

Using GIS, spatial analysis was completed to identify general fall injury patterns and high-incidence areas in both Alberta proper (such as rural and medium-population areas) and within a domicile (such as on a stair or off a ladder). The researchers relied heavily on geographic information products, such as

empirical Bayes estimate maps of incidence data. Empirical Bayes estimates invoke British mathematician Thomas Bayes's theory of statistical sampling, which states that a probability distribution can be assigned to an unknown parameter in a statistical problem. In this application, the researchers assigned an underlying probability distribution to the fall incidence ratio. In the end, the researchers proved their argument: that fall injuries in the elderly occur in distinguishable areas, making region-specific fall prevention programs possible and definitively beneficial.

Disaster: Planning, Response, and Recovery

We have already discussed in some detail how a GIS can be used for information distribution in emergencies or for analysis in the wake of a disaster. However, there are additional functions relevant to natural and human-made disasters that we have not covered, namely disaster planning, response, and recovery.

As a result of the recent natural events and terrorist activities, the need for a heightened focus on disaster planning is sobering. Property owners, bankers, insurance professionals, and state and federal agencies have taken preliminary steps to increasing their disaster preparedness, while being proactive toward disaster response procedures and the minimization of lost business days during recovery. As you might guess, GIS has been playing a greater role in these areas.

Disaster planning involves understanding historical events, analyzing each historical event's "before" and "after," and learning to better deal with a similar situation. Field- and satellite-based GIS data in time-lapse maps are used to study events ranging from the levels of floodwaters and the nature of particular tornado damage to past errors in fire safety and building collapse characteristics. Disaster planning highlights offsets to catastrophic ruin and ways to back up current systems. Overall, disaster planning covers a wide ground, but it can be reigned in through the analytical and database power of GISs.

Disaster response goes hand in hand with planning, but with a deeper focus on past response–related errors. The main purpose of a response is to lessen the amount of loss, be it human loss or material loss. Every past error indicates how a response should not be conducted, while highlighting those areas where additional planning is needed. A well-planned GIS can help route emergency personnel to the correct area or alert them of the correct protocol, transplanting accurate data for potentially fatal guesswork. A GIS can calculate the best route, querying the most recent street-position data, construction detour data, and traffic information. Using a GIS for disaster or emergency responses can be the time-saver needed to save critical material or to save a person's life.

This leads us into disaster recovery, which is the readjusting of a disaster event to a level of normality or stability. For example, banks implement offsite

data recovery stations, where vital financial information is transmitted and can be instantly recovered in the event that the origin bank is incapacitated. Disaster recovery goes much further, such as in the reoccupancy of buildings or the desperate search for human life in a building's rubble. GISs are used to inform (once again) rescuers and emergency personnel of vital location data, such as the number and identity of building occupants from an electronic security badge database or maps of predisaster floor layouts. The main purpose for disaster recovery is to minimize loss while stabilizing the situation in a timely manner.

Transportation and Information

Public transportation systems have always been a natural candidate for embracing GIS and the geographic information products the technology offers. Subway maps, bus routes, highways and public access roads, and ferry paths have historically involved graphical street maps overlaid with route paths, route numbers, and legend symbols. Maps offer the general public an easy-to-understand bird's-eye view of the entire transportation system and engage the simple human task of recognition. There are no in-depth text instructions to comprehend or composed tables to siphon through in order to find the desired information. Just graphics, labels, and overlays: the basis for GIS.

With the use of geographic information products, public transportation and the distribution of transportation-related data are much more efficiently facilitated. Transportation GISs not only can map routes for the public user but can also detail construction locations, alternate routes and reroutes, city gridlock areas, and sites of heavy congestion. These systems help keep the public commuter well informed while offering a means of informed transportation choice. This side to GIS use in transportation is extremely visible to the public user.

There is another side to public transportation that is less visible to the end user and much more instrumental in application. GISs are used to maintain subway tracks and roadway, maintain transportation fleets and equipment, identify system faults, route emergency vehicles, pinpoint accidents, and make immediate changes to information products. These applications, of course, are just a sampling of the copious functions that public transportation managers, engineers, and operators require.

Going one step further, transportation planners and engineers use GIS to better analyze particular traffic situations in any study location. These professionals use spatially defined or geocoded data to help discern the causes and effects of the particular situation, while the information gained from this type of geospatial analysis aids in the development of a practicable, cost-effective solution. A GIS enables transportation professionals to build upon existing transportation network data, update information with GPS, and easily integrate spatially referenced attribute data.

You need not look too far from your own town or nearby city to fully observe the power of GIS in relation to public transportation and information dissemination. Visible to the public or not, many of today's smaller public transportation systems (and almost every large public system) are now spatially enabled in some fashion. Check out a transportation Web site for a local agency for indications of geospatial data. If you want to grasp how completely public transportation and GIS can meld, take a look at Portland, Oregon's TriMet interactive system map at http://www.trimet.org.

Case Study: Portland's TriMet Transit Tracker Online

Portland's TriMet system offers its public ridership real-time bus information and train (TriMet's MAX trains) arrival data through an ESRI ArcInfo-backed system called Transit Tracker. The real-time information is accurate wherever the user may be. The system works off GPS mobile tracking, wireless technology, and a GIS. Transmitting sensors on the public buses send location information to the Transit Tracker system database, while sensors built within the train tracks transmit train speed and sensor location. The system displays actual bus locations and calculates an estimated arrival time for MAX trains based upon transmitted information. Public users can access this data wherever they are through handheld devices, the Internet, or telephone.

TriMet's interactive system map uses a high-resolution aerial photograph in conjunction with bus and train data layers. Maps highlight bus stops, MAX stations, points of interest, and park-and-ride locations. As with any GIS map product, map views can be zoomed in and out, information can be retrieved about stops, stations, and routes, and specific locations can be searched and viewed. An interesting feature to TriMet's system is their MAX train "up-to-the-minute arrival time" window that displays train arrival countdowns. Portland's use of an online GIS for public transportation is highly progressive and will set the path for future public transportation information systems.

Homeland Security's Third Eye

Although we have discussed GIS's role with disaster planning, not a day goes by when the phrase "homeland security" is not heard, along with the notion that something more is needed for terrorist events than pure disaster planning. In these undetermined times, the subject of national security is of great importance to both the government and the public. Governments want to maintain a status quo on threat control and homeland sustainability, while the public wants full information disclosure and a stable sense of life safety. Security analysts and managers now use GIS in conjunction with other in-place protocols for threat

and vulnerability assessments; security control procedures; sea craft and aircraft tracking systems; surveillance, detection, and intelligence gathering; data sharing; and overall preparedness.

The interesting aspect about using GIS with domestic security is that typical security projects meld the public applications just discussed: public safety, disaster planning, and transportation systems. Its use with public safety is seemingly obvious. GIS offers a means to identify safety concerns and areas for potential vulnerability, whereby officials can close off these vulnerable areas to the general public. GIS products enable agencies to identify and subsequently control public access areas to help maintain public safety and enhance early threat detection.

GIS offers powerful capability in understanding how a safety plan will work in the event of a disaster or terrorist situation. The technology helps security analysts detail the best areas for a rapid mass public exodus from any area, as well as defining the easiest paths for emergency personnel. GIS proves to be most useful with interagency information sharing, in addition to serving as a central repository for information-gathering exercises. This single system, single data methodology may be the deciding factor in preventing a terrorist attack or in maintaining control during a sudden, disastrous event.

The transportation portion of domestic security revolves around mass public threats and evacuation protocols. Because transportation systems handle a large, diverse ridership, with people looking over unknown faces every day, the importance of early detection is critical to human safety. GIS and GPS are used to monitor train and bus locations and can help officials identify any abnormalities with service. These abnormalities can act as the early detector in public safety concerns or hostage and terrorist situations (on September 11, planes had abnormal flight patterns). Additionally, GIS can collect ridership data for any given moment, which can aid in vulnerability analysis and rescue/recovery initiatives.

Traditionally, in mystical circles a person's "third eye" was considered the eye that saw the unseen and invoked a person's extrasensory perception. GIS is quickly becoming the third eye for worldwide homeland security protocols, since it helps analysts "see" what is not readily obvious and helps keep agencies one step ahead of terrorists. In reality, the technology's automation is its greatest asset to these security professionals, while the data's accuracy and consistency enables more concentrated and successful threat mitigation measures.

Public Education Applications

GIS technology has become the newest tool for public education analysis and proliferation. GIS is now used in tandem with teaching curricula to enhance a

student's ability to comprehend lessons and perform hands-on problem solving for complex problems. Science, physics, mathematics, Earth studies, and chemistry are among the major subjects most positively affected by GIS, though the scope of in-class analytics is wide. The information products form intriguing, exciting manipulatives that tend to engage students and help progress their understanding of the topic.

Teachers are not the only educators using the information technology; school officials, principals, district leaders, and superintendents have begun using GIS to better understand their district's or school's demographics. Educators and school officials can evaluate their school system occupants through a mix of demographical data that ranges from racial background and economic status to familial situations and standardized test scores. GIS offers a new level of detailed analysis that helps educators and officials pinpoint where deficiencies lie and successes hide. When used correctly, a GIS can be a potent facilitator for a better school system.

The power to conduct accurate educational analysis through information sharing between districts/schools is far more effective than with past methodologies. A GIS, by default, becomes a central clearinghouse for school data, which aids in quantitative and qualitative educational understanding. School officials can precisely rank their schools with other same-level schools, district leaders can compare their successes with those of other districts, and superintendents can monitor the overall effectiveness of their public education domain.

Using GIS in tandem with education helps students and educators become community aware and more firmly involved with the events and activities around the school. This interchange of information breaks the spatial barriers of traditional learning and gives everyone involved a refined vision of data exchange while making each user aware of their community's spatiality.

Aside from the educational advantages of GIS, school facility managers can implement a Web-based GIS for campus/school facility information. Students and educational personnel alike can easily access real-time campus maps, building information, campus activities, and community notifications. Facility managers can use such a system for building inventories and utilities, as well as to dispense facility closure and emergency evacuation information.

In the same vein as facility information, the campus community can use the technology to tap into available neighborhood information for identification of potential threats, to learn about local evacuation procedures, and to realize areas of caution. Some district police departments conduct spatial crime analysis through geocoding community felonies (i.e., turning crime addresses into grid coordinates) to keep neighbors, neighborhood watch groups, and officials informed of local incidents. The availability of neighborhood data depends upon the capabilities and funds of the particular community. Nevertheless,

through GIS, public education participants have a viable means of being zealous in their community while encouraging active learning and participation.

Redefining Public Policy

The field of public policy remains one of the most radical arenas for GIS implementation. That stated, one reason for its radical prowess is public policy's diverse and ever-changing atmosphere. Public policy covers every agreeable public domain, spreading from education and the environment to social services and international applications. If that weren't diverse enough, public policy initiatives involve the lawmakers in government, the decision makers in industry, and the eventual public end users. In short, public policy GIS applications cover every niche we have discussed so far in this book.

Since public policy's applicatory enormity can be overwhelming, let me define its common scope for spatiality. Typically, GIS is used to help secure governmental financing, define public issues to the legislators and voters, determine the effectiveness of current policy, help streamline new and existing governing protocols, and identify voids in foreign policy.

Governments finance various efforts, such as agency requests (e.g., war analysis and border patrol), industry innovations (e.g., power generation and aircraft development), and public needs (e.g., public assistance programs and education). Consequently, a deeper understanding of such financial distributions and contributions is required, and the overall effects of these fiscal changes must be adequately assessed.

To simplify fiscal assessment, GIS maps have become the popular tools of public policy applications. Maps can provide straightforward graphical depictions of current policy and highlight the advantageous elements of proposed policy changes. Color scale maps tend to be the presentation of choice for policy makers, since the graphical displays are easy to understand and can be manipulated for any specific purpose. For example, a request for additional fiscal allotment to the U.S.'s Women, Infants, and Children (WIC) program can be proved to be warranted with the help of a needs evaluation map using a confirmed data set and a differentiated color scale. The shades of color can signify increases or ranges in a target need and give a visual, easily distinguishable peek at how an allotment of funds can fill the understood needs.

Going beyond fiscal analysis and the definition of public issues, GIS has been a proponent of public policy education. Many times agencies need to educate the public on their current policies to enhance public awareness of active programs, prevent misunderstandings, and diminish unnecessary outcries for policy reform. GIS can play and has played a huge role in public policy education through detailed maps, explanatory literature, and online information

depots. With GIS-enhanced policy education programs in place, the general public is better informed, and ultimately, the policy making process is progressively focused and successful.

Case Study: Public Policy Education in Gaston County, North Carolina

In 1988, the Quality of Natural Resources Commission (QNRC), a committee of appointed citizens in Gaston County, North Carolina, was chartered to identify the county's air, surface water, and groundwater environmental concerns. A substantial portion of QNRC's work was to recommend policy changes or new policy to protect Gaston County's natural resources. Early in the project, the committee was faced with incomplete and fractured data, as well as data too technical for effective policy making. The QNRC's work progress quickly became stifled, at which point a multidisciplinary group from North Carolina State University (NCSU) stepped in to help. The group undertook a two-phase approach to the study, wherein the first phase highlighted a countywide environmental assessment and the second phase involved refining that assessment and developing a public policy education program with educational materials.

Throughout both phases of the work, the NCSU-QNRC team used geographic information systems. In phase one, GIS was used to help the research team analyze the nature and impact of Gaston County's environmental problems. In phase two, GIS was used in a more educational function, with information products being developed and used for policy education tasks. GIS maps highlighting hazardous waste sites, community well locations, water supply intakes, and surface water quality monitoring sites were among the various types of public materials created. These maps were implanted within educational literature and distributed to the public.

The NCSU group developed a public policy education program that piggybacked the publicly distributed educational materials and helped in policy evolution. Geographic information products aided the NCSU group and QNRC to realize the extent of environmental problems within Gaston County, locate the faults within existing public policy, and portray their findings to the public in a wholly comprehensible manner. In the end, NCSU and QNRC met their charter and created an environment whereby Gaston County's natural resources are protected.

15

Getting Started with GIS—An Overview of Environments, Tools, and Data

As we now understand, GIS, by its very nature, caters to visual understanding. A well-planned information product captures the important elements of the location like a snapshot captures moments and delivers it in a digestible format. Individually, text and graphical representation offer essential information for any need, but together, a whole story is told.

The same premise goes for the discovery, demystification, and total comprehension of GIS itself. A textual description of GIS and its many elements such as this book can only deliver a theoretical understanding of the science and systems. The learning of GIS demands interaction and hands-on use for a true and total grasp.

We are now at that crossroad where the essential theory has been discussed and the hands-on portion must commence. This chapter is designed to get you up and running with GIS and to pinpoint key data and software destinations on the Web for you to explore. As you work more with a true GIS environment, the theory will coalesce and an advanced understanding will form.

Getting Yourself Up and Running with GIS

Up until this point you have been armed with GIS theory and are now likely eager to start using a geospatial environment. Like an athlete mastering his sport, the GIS novice must practice in a GIS environment for the theory to fully make sense. But how can you move past the theory if you own no such GIS software and have no data with which to practice?

You can purchase the required software; however, the overall expense may be too great for someone just learning the craft. To gain experience quickly, a more practical solution (at least in these beginning stages) is to try out the free GIS software made available through various outlets in the geospatial and GIS communities. Some of the software offerings have been around for many years, while others are relatively fresh in the field. The array of available programs includes complete GIS environments, component software, basemaps, and geospatial data.

Before moving too far, we will take a brief look at the leading geospatial systems on the market that are considered the industry "standards." Once we examine what is available for purchase, we can set our sights on the easy-to-acquire free programs. Bear in mind that these systems are somewhat scaled-down environments. However, some free GIS environments, such as GRASS (discussed later), are comprehensive and are used as the sole systems in many agencies around the world. With only modest use, you will gain an appreciation of the power and functionality of this available software.

Leading For-Purchase GIS Environments

GIS is a hot commodity in the technical world. Many engineering, environmental, and technology firms use GIS regularly and subscribe a sometimes sizable budget each year for upkeep on the latest and greatest GIS environment. Luckily, today's market is highly competitive and, as such, offers an excellent array of GIS packages from which to choose. As with many software niches, there are many software packages offered, but only a handful are true benchmarks.

The GIS industry has many comparable packages available to the public. Table 15.1 lists diverse GIS environments available for purchase. However, within this listing there are seven GIS environments that lead the pack in terms of package features, flexibility, overall use, and cumulative sales throughout the industries. These systems are ESRI's ArcGIS, Intergraph Corporation's G/Technology, General Electric's Smallworld, Clark Labs' IDRISI Kilimanjaro, Autodesk's GIS Design Server, DeLorme's XMap, and MapInfo.

Although these leading software packages are considered popular and are being discussed within this context, I encourage you to examine the whole range of available GIS environments to identify the right software for your need. All of the software packages on the market are quite similar, but each offers particular variations to define functional autonomy. Equally, every GIS user has a different prerequisite and need, and to select software based on popularity may, in the end, shortchange efforts and minimize project goals. Keeping that in mind, let us take a brief look at these leading geospatial environments.

Table 15.1
For-Purchase GIS Environments

ArcGIS	http://www.esri.com/software/arcgis/index.html
Maptitude	http://www.caliper.com
GIS Design Server and Autodesk Map	http://www.autodesk.com
G/Technology and GeoMedia	http://www.intergraph.com
IDRISI Kilimanjaro	http://www.clarklabs.org
Manifold Release GIS	http://www.manifold.net
MapGraphix GIS	http://www.comgrafix.com/mapgrafix_gis.html
MapInfo	http://www.mapinfo.com
MetaMap GIS	http://www.metamap.com
MyWorld GIS	http://www.myworldgis.org/myworld/
Smallworld	http://www.gesmallworld.com
TatukGIS	http://www.tatukgis.com/Home/home.aspx
TNTmips	http://www.microimages.com/product/tntmips.htm
XMap	http://www.delorme.com/professional/xmap

ESRI's ArcGIS is by far the most popular geospatial environment available. The Environmental Systems Research Institute is an offshoot of the U.S. Geological Survey (USGS) and has grown in popularity since its inception in 1969. Due in part to its ties with USGS, ArcGIS has become the favored system for many U.S. and international agencies. The ArcGIS environment encompasses an integrated set of geospatial products that, grouped together, form a powerful enterprise GIS environment.

ESRI's ArcGIS suite is designed for the private and public sectors alike, making available software for desktops, laptops, tablets, servers, and mobile devices. The GIS software enables users to analyze, manage, and map geographic information and can be deployed throughout an organization by Internet and intranet connections. Users can install plug-ins, called *extensions*, to add functionality to their GIS viewing package. ESRI also offers ArcExplorer, which is a free GIS data viewer that allows basic mapping and spatial querying. Overall, given their solid industry presence, ESRI's GIS suite has remained a strong leader in the GIS software market.

Another strong leader in overall use and popularity is Intergraph's G/Technology and GeoMedia program suites. Intergraph's packages offer powerful analysis, management, and design tools geared toward users in virtually every business sector. Intergraph's G/Technology suite of programs offers specialized industry-specific data models for the various utility, pipeline, water, and communications industries. Intergraph boasts that G/Technology addresses the

best practices for cost containment, process improvement, systems integration, customer base retention, and new service opportunities. Additionally, the software was designed as an Open GIS and can subsequently work with many different GIS formats.

Intergraph's GeoMedia software suite provides user flexibility through strong, quality-focused programs designed for efficient map design, presentation, and sharing. GeoMedia allows users the flexibility of desktop, laptop, and enterprise-wide compatibility. Additionally, Intergraph offers GeoMedia Viewer, a free GIS viewer geared toward desktop geospatial viewing and the distribution of data. The combination of G/Technology and the GeoMedia software suites proves a formidable GIS environment.

General Electric's Smallworld suite offers a fresh "feel" to GIS through its advanced spatial technology and seamless existing system integration. The Smallworld suite encompasses a selection of powerful geospatial programs, including Smallworld's Core Spatial Technology, Spatial Intelligence, Enterprise Integration Tools, and Design Manager. Smallworld's program suite is designed for desktop, laptop, and Internet interoperability.

GE's software architecture is somewhat different from ESRI's all-purpose design and Intergraph's general and specialized software programs. Unlike the broad focused programs, Smallworld's design centers upon engineering, scientific, and business-oriented applications. Like its competition, GE's flexible GIS platform is an exceedingly capable geospatial environment.

In an industry with so many competing environments, a good many users prefer user simplicity and image workability over an abundance of functions and project-specific decision-making. For this, Clark Labs offers IDRISI Kilimanjaro, a widely used raster-based GIS and image processing environment. IDRISI was developed as part of an academic research program and has been designed (and redesigned) to stay user friendly and highly accessible, while serving as a benchmark in geospatial standards.

IDRISI Kilimanjaro offers the same advanced GIS and image functions as the other leaders, such as GIS modeling, database querying, spatial data development, and geostatistics; however, the software uses advanced object-oriented development tools that are especially suited for focused research and environmental modeling. Aside from its raster-based functionality and ease of use, its low cost has maintained IDRISI as a viable competitor among the GIS leaders.

No discussion about GIS environments should exclude Autodesk, a strong software company in the engineering and technical fields for many years. Autodesk is best known for its impressive AutoCAD software and has now grown its technical market to include GISs and mapping, since CAD and GIS are complementary. Autodesk Map is the company's GIS environment that is built around AutoCAD and, as such, invokes user comfort through its widely familiar AutoCAD "feel."

Autodesk Map offers the user flexibility in designing, maintaining, and producing maps and geographic data. The software's geospatial capability is solid and acutely comparable to the competition. The program suite supports desktop, laptop, tablet, mobile, and enterprise systems and has a huge following. Through their enterprise GIS system, called GIS Design Server, Autodesk offers easy data integration with reliability and security. Although the user can work on various GIS file types through the software, the most seamless of file exchanges are with files that were originally designed in an AutoCAD environment. That being said, with functionality excellence, user trust, and an already familiar program exchange, Autodesk Map proves itself a GIS forerunner.

Taking the route of minimized expense, maximum function is mapping giant DeLorme's XMap software. XMap is a relatively low-cost, full-featured GIS environment that offers GIS and GPS mapping capabilities. XMap uses a modular design for expandability, interoperability, and enhanced functionality. XMap also supports Open GIS and most GIS data formats. DeLorme's software offers a less expensive alternative for quality, but standard, GIS capability.

This leads us to MapInfo's suite of GIS software, led by their MapInfo Professional program. MapInfo Professional is a popular GIS environment that is comparable to the other leading systems in terms of geospatial capability, analytical tools, and add-ons.

MapInfo's development side is where this software package shines, through MapX, an Active X component that enables alternative software embedding. What this means is that developers can embed mapping applications within other applications, such as Microsoft Word and Excel, Lotus 123, and others. This specific capability added to the suite of other GIS analytic capabilities has made MapInfo highly attractive to business developers and users. Incidentally, MapInfo's product uniquely exemplifies how geospatial environments differ, each offering advantageous characteristics amenable to particular goals.

Now that we have examined the most popular for-purchase environments, let me just reiterate that for-purchase systems require available funding and, as with any purchase, should not be chosen by text alone. You would never dare purchase a new car without taking it out for a test drive. Identify the software packages you are interested in and examine functionality, cost, and file interoperability. Then take your GIS test drive by downloading demos of the software and discovering which software meets your needs and which fall flat.

GRASS and Other Free GIS Environments

To many GIS novices, the thought of spending hundreds, even thousands, of dollars on GIS software is alarming. What may be even more disturbing is the fact that comprehensive software is ultimately needed to provide the user with

the full functions, science, and geospatial mechanisms of GIS. As mentioned previously, the proper learning of GIS requires theory intermixed with hands-on use. Without access to GIS software, the latter part of the learning equation is undermined.

Past users have been in a similar situation of learning on a budget or staving off expensive software purchases until project requirements and full system benefits became apparent. The need for inexpensive or totally free GIS software surfaced in the early 1990s, and over the many years a catalog of quality, free GIS environments developed. In many cases, the free GIS software, called *freeware*, is a scaled-down version of the more popular fee-based software. These scaled-down versions are sometimes as basic as GIS viewers and ESRI-based shapefile editors, while other versions are fully functional base GIS systems. Both forms of freeware offer the novice GIS user a surefire means to gain the hands-on skills needed for proper understanding without a hefty monetary investment.

Answering this call for a free, fully functional GIS environment is GRASS, one of the oldest and most recognized GIS software applications available. GRASS, which stands for Geographic Resources Analysis Support System, was originally developed by the U.S. Army Construction Engineering Research Laboratories (USACERL), a branch of the U.S. Army Corps of Engineers, in 1982 as a military tool for land management and environmental planning. The USACERL maintained and revised the software up until 1995, at which time the software was deemed well developed and suitable for a wide range of applications. Today, GRASS is maintained by Baylor University in Waco, Texas, and has various mirror sites throughout North America and Europe.

By design, GRASS is a raster and topological vector-based GIS that supports image processing, geographic data management, spatial modeling, map and graphics production, visualization, and geospatial analysis. The software, with its straightforward functions and solid architecture, has gained wide appeal throughout the industry, whereby users have chosen GRASS over one of the more popular for-purchase GIS environments. The software contains more than 350 programs that offer add-on functions, such as raster/vector manipulation and enhanced mapping capabilities.

The GRASS GIS environment is used for academic and commercial applications worldwide by public and private organizations alike. In fact, the software is used by the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Agriculture (USDA), National Park Service (NPS), U.S. Census Bureau, and USGS. With wide industry use and public domain status, GRASS proves to be an excellent starting point for beginning GIS users.

As previously indicated, all types of skill sets have turned to GRASS and other free packages to produce viable GIS products. The demographics for

system development and use are interesting. Free GIS software packages have been developed by university researchers, programmers, philanthropists, geospatial organizations, governmental agencies, and private developers since the 1980s. As with any mass assortment of developed software, a large number of the available titles were created to serve specific roles, either in application (e.g., mining and business statistics) or in function (e.g., file format conversions and image analysis). Beyond these specific titles are the base geospatial environments (such as GRASS) and standardized software that enable users on all platforms and of all skill sets to develop, manage, analyze, and map geographic data.

To guide you in this maze of freeware, Table 15.2 offers an abridged listing of popular free GIS environments, viewers, and shapefile editors available. More comprehensive listings of geospatial freeware exist on the Web through FreeGIS.org (http://www.freegis.org), GIS Lounge (http://www.gislounge.com), and the Open Geospatial Consortium (OpenGIS; http://www.opengis.org). You will be amazed at the amount of free GIS software available. In fact, be prepared when you visit FreeGIS.org; the site is astonishingly broad and has a cavernous array of good GIS software links.

The quality of freeware ranges from the very good to the moderately bad, which really means from high operability to difficult to use. Since GIS outputs

Table 15.2
Free GIS Environments, Viewers, and Shapefile Editors

ArcExplorer	http://www.esri.com/software/arcexplorer/index.html
AGISMap	http://www.agismap.com
Chameleon	http://chameleon.maptools.org
deegree	http://deegree.sourceforge.net
DIVA-GIS	http://www.diva-gis.org
Forestry GIS (fGIS)	http://www.digitalgrove.net/fgis.htm
GeoMedia Viewer	http://www.intergraph.com/gviewer
Geo/SQL	http://www.geosql.com
GRASS	http://www.baylor.edu/~grass
JGRASS	http://www.hydrologis.com/html/jgrass/jgrass_en.html
MapWindow GIS	http://www.mapwindow.com
idctlparmonoGIS	http://www.monogis.org
SAGA GIS	http://www.saga-gis.uni-goettingen.de/html/index.php
SPRING	http://www.dpi.inpe.br/spring/english/
TatukGIS Viewer	http://www.tatukgis.com/products/viewer/viewer.aspx
uDig	http://udig.refractions.net/confluence/display/UDIG/Home

are directly related to the geodata entered and not the system per se, quality really can be judged only on ingrained capability and functionality.

GIS Component Software

Once you begin using GIS software, whether free or purchased, you will soon realize that the software has limits. Most times the software will be satisfactory within these limits; other times you may want more. Perhaps the limit breaker is a data set that is in a different format that will take twice as long to integrate into your system or maybe a specific map analysis function that currently doesn't exist in the software. When more is desired, you have two options: Work around the limitations of your GIS environment or turn to GIS component software.

GIS component software is a building block that, when added to GIS software, forms an enhanced, personalized environment for the user. A mix of component software will make any GIS environment distinct for that user—making the environment the user's own. GIS component software allows the user to go far beyond the base system boundaries, limited only by a user's ability.

Component software is analogous to an office work desk. The actual furniture is like your GIS environment software—solid structure, yet standard for a multitude of users. The user's belongings—family pictures, reference books, or project paperwork—relative to the desk, just like component software to the GIS, make it the user's own. When used correctly, these belongings (components) add to the overall experience, use, and personalization of the desk (GIS environment) for a distinct user.

GIS component software comes in many variations, including specific function programs or user development software. Specific function components perform a dedicated task that adds to the GIS environment's tools. Such components include data format converters, flow data analyzers, and image processing software. User development software, on the other hand, is a development toolkit that enables the user to program components to perform specific functions. Say, for instance, you need to embed maps into a particular non-GIS program and there is no predeveloped component for your raster-only GIS. Developable component software would be the likely choice, especially if this operation is to be done numerous times. Simply put, component software helps save user time and automate redundant functions.

If component software interests you, there are many excellent products on the market. Perhaps the most popular component software available is Geotools (http://www.geotools.org), an open source GIS development toolkit that is freely distributed. The software is Java based and offers users the ability to develop OpenGIS-compliant products. OpenGIS is dedicated to developing

and standardizing geospatial and geoprocessing specifications. What makes Geotools attractive to users is its modular design, which allows easy installation and removal of components. The software works well with the major fee-based and free GIS environments mentioned previously.

There are various component software and development toolkits available. Table 15.3 lists a variety of useful component software links for you to explore. Sites involve fee-based and free software. Care should be taken when selecting components to download, since not all components work with every GIS environment.

Once you get hands-on familiarity with a GIS environment and are comfortable with the base functions, you should try a component. Test out a few different components to decide what these building blocks can do for your efficiency and total GIS experience. In the end, only you can decide what (or if) components are right for you. Sometimes it depends just on the objective of a project. But as with your office desk, only you can decide what you need within arm's reach.

Table 15.3
GIS Component Software

ALOV Map/TMJava	http://www.alov.org/index.html
Blue Marble Geographics	http://www.bluemarblegeo.com
FlowMap	http://flowmap.geog.uu.nl
FME objects	http://www.safe.com/products/developer-tools/fme-objects/index.php
Generic Mapping Tools (GMT)	http://gmt.soest.hawaii.edu
GeoTools	http://www.geotools.org
JShape software	http://www.jshape.com/frame.jsp
MapE Library	http://www.mapesoft.com/MapELibrary.htm
Map Fusion	http://www.geosolutions.com/globalgeo/default.aspx
MapObjects	http://www.esri.com/getting_started/developers/mapobjects.html
MapX	http://extranet.mapinfo.com/products/Overview.cfm?productid=1041&productcategoryid=1
Open Map	http://openmap.bbn.com
SpatialFX	http://www.objectfx.com/products/spatialfx.asp
SylvanMaps/CF	http://www.sylvanmaps.com/sylvanmapsCF.htm
TatukGIS Developer Kernel	http://www.tatukgis.com/products/Dk/kernel.aspx

Geospatial Data Clearinghouses

It can be said that a GIS system is only as good as the user's ability and the data set used. No discussion on how to get started with GIS should dismiss the acquisition of geospatial data. There are fundamentally two ways to obtain geospatial data: creating it yourself or acquiring previously developed data sets. Creating a quality data set is neither simple nor cheap, as one would expect. Quality dataset development requires, at a minimum, field visits and/or field personnel, satellite image analysis, aerial image analysis, GPS coordinate attainment, geospatial data entry, preexisting dataset investigations, and any existing basemaps. The effort expended on creating custom data sets can be arduous and expensive; therefore, often firms or individuals creating the data set sell their product on the market or create a geodata "bi-product" to sell. You can refer to our discussion on data bi-products in Chapter 2.

Acquiring a previously developed data set is frequently more practical and, as such, the path most GIS users take. You can acquire much needed data instantly over the Internet either at a price that is far lower than creating the data set yourself or through freely distributed data sets. For specific data sets, perhaps the 2005 bushfires in Yellowstone National Park, previous data sets may not be available and the burden of creating it is on the user.

Previously created base and bi-product data sets are distributed primarily through geospatial data clearinghouses, data warehouses, and data depots. The terms *clearinghouse, warehouse,* and *depot* are nearly synonymous with each other. For our purposes, we will view them similarly, as collections and collaborative sources of geospatial data sets. These clearinghouses sell for-purchase data sets as well as distribute free data sets under a public domain license. Easier-to-obtain datasets, such as satellite-based data, are distributed free and less available data, such as detailed bi-products about specific areas, are often fee based.

Geospatial data clearinghouses are widely used throughout the numerous industries using GIS. As a learning GIS user, you should visit these clearinghouses quite frequently and obtain free data sets and metadata with which to hone your GIS skills. Advanced users frequent these clearinghouses regularly for their project maps and geographic information products, as well as for background metadata. Whatever the purpose, these data clearinghouses are among the best places to obtain viable geospatial data and metadata.

Table 15.4 details only a small handful of international data clearinghouses. All of these sites have free data sets, and some, such as the Data Depot, are totally dedicated to free data and metadata. Bear in mind that these are just a selection of the many hundreds of available clearinghouses. For a more comprehensive listing of data sites, refer to the University of Edinburgh-Association of Geographic Information's GIS Resource list (http://www.geo.ed.ac.uk/home/giswww.html) for links to several hundred U.S. and international GIS data sites.

Proper use of free (or any) geospatial data sets is essential for attaining effective and accurate GIS products. Users must carefully consider and understand the guidelines about how the data were obtained and how they should be used. This information lies within the data set's geospatial metadata, which we discussed back in Chapter 4. No matter whom the data comes from or what typical parameters are normally used, each data set is individual and should be treated as such for the sake of your product's quality and applicability. Assume nothing and keep the data quality bar raised.

At last, you are ready to turn theory into practice within the world of geospatial information and GIS. With the mystery and uncertainty about GIS removed by understanding, you are well on your way to creating geospatial answers to today's worldwide geographic questions. Armed with these discussions and the starting knowledge of the available GIS environments, you can keenly get yourself started with GIS.

Table 15.4
Geospatial Data Clearinghouses

The Australian Consortium for the Asian Spatial Information and Analysis Network	http://www.asian.gu.edu.au/index.html
Blue Marble Geographics World Map Dataset	http://www.bluemarblegeo.com/products/worldmapdata.php
The Data Store	http://www.data-store.co.uk
EROS Data Center	http://edc.usgs.gov
Federal Geographic Data Committee Geospatial Clearinghouse	http://www.fgdc.gov/clearinghouse/clearinghouse.html
GEO-DATA Explorer (GEODE)	http://geode.usgs.gov
The Geography Network	http://www.geographynetwork.com
Geo-Gratis	http://geogratis.cgdi.gc.ca/clf/en
Geoscience Australia National Geoscience Datasets	http://www.ga.gov.au/map/broker/wms_info.php
GIS Data Depot	http://data.geocomm.com
MapMart	http://www.mapmart.com
ResMap: Earth Image Source	http://www.resmap.com
United Nations Environment Network	http://www.unep.net
USDA Forest Service geospatial data clearinghouse	http://svinetfc4.fs.fed.us
USGS geographic data download	http://edc.usgs.gov/geodata
University of Edinburgh Data Library	http://datalib.ed.ac.uk/holdings.html

Afterword

The Future of the Growing GIS Technical Revolution

It is hard to believe that GISs have been around for more than four decades. Within this time frame, the world has experienced rather turbulent and exciting times, wherein society has dealt with human inequality issues, a superpower Cold War and escalating world unrest side by side (almost hand in hand) with the proliferation of the Internet, powerful microprocessors, and terrestrial and extraterrestrial exploration. Today, the thought of GIS as a "fresh" technology still exists, but in reality it has been reinvented and revised so many times that it now appears as something much different from its original form.

The first 40 years of GIS have proved that the core theory has remained unchanged, but the method and functionality of manipulating this theory and producing informative products have drastically changed. When GIS first came into being, it was developed and used primarily for land use issues. Its simple database structure offered a repository of land-based information and analysis. Through the years came the availability of satellite imagery, relational databases, desktop computers, powerful software, and world data infrastructures, just to name a few. Add these innovations with users who are better educated and have a multidisciplinary interest and the sum is an established, reinfused technology.

You may be asking yourself, "Where does GIS go from here?" or "How can this already established technology survive in the years to come?" Every user most likely has a personal opinion as to the future of GIS. In fact, there has been a good amount of print and online space dedicated to the dissemination of opinion and prediction.

A common theme throughout many opinion papers is the belief that history is the best indicator for technological evolution, since it reconstitutes in

continuous yet variant forms. I hold this same belief. Consequently, if history proves a valid gauge for GIS, then the technology will ultimately survive through further revision and enhanced redirection. Although this sounds as generic a prediction as a newspaper horoscope, the mere fact that GIS still exists today and is a developed technology adds value to this claim.

Let us move beyond today's version of GIS and postulate about the direction it is going. For this, I have developed what I call a GIS compass. This GIS compass (and this entire discussion for that matter) represents my own personal musings as to the direction of GIS. The following compass directions are selections of real issues that, to me, may influence the course of GIS's evolution.

The Developing GIS Compass

A compass is a worthy instrument, used to navigate terrains and help the user get from a point of origin to one of destination. It also helps the user discern what direction he or she is going. Understanding the value of a compass and the need to gauge the GIS industry's direction, I created the GIS compass. My objective was to help convey a selection of emerging terrains for GIS and identify the chief directions that I believe the industry is currently moving. Figure A.1 depicts the developing GIS compass.

As evident through today's GIS use, the industry is quickly moving in four distinct and primary directions. I define them as:

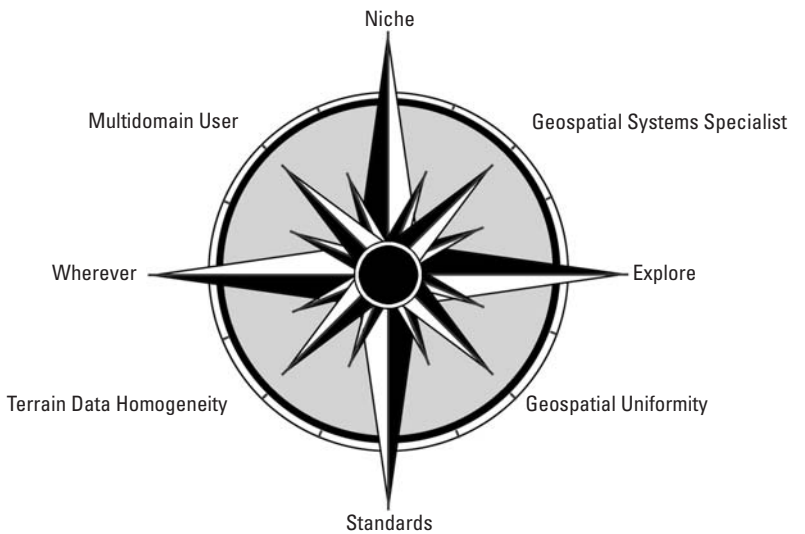


Figure A.1 The GIS compass.

- *N for niche.* Moving from the present-day GIS professional to an information technology niche field;
- *E for explore.* Moving from a geodatabase to enhanced geospatiality by exploring the increasing geospatial landscape;
- *W for wherever.* Progressing from the Earth to other topographies (both macroscopic and microscopic), such as interstellar terrains and inner body mapping;
- *S for standards.* Developing universally respected global standards.

These primary directions are speculative in nature and were determined through my interactions in the industry. User types are diversifying, applications are broadening, and the knowledge that any mapable surface can be geospatially coded is spreading. Among these primary directions are meldings in between that, I believe, will ultimately form the specifications to industry survival. Let us now take a brief look at these key pathways.

Niche

The GIS professional was always just that, a professional dedicated to GIS. These dedicated persons form the core staff responsible for (and knowledgeable about) spatially enabled geodatabases. Typically, they are the project members who provide the geospatial grunt work, such as uploading basemaps and data, validating data, geocoding resources, troubleshooting geospatial inaccuracies, and producing specific information products. They are the dedicated professionals for GIS.

With greater interest and better software abounding, it appears that today's GIS professionals are sometimes sidetracked by GIS do-it-yourselfers or project managers who skimp on their budgets. It is not out of the ordinary for GIS professionals to be tasked with other non-GIS project responsibilities. Most commonly, the GIS professional is slated as the information technology (IT) specialist, wherein the workload consists of both geographically based and general information–based system tasks. It has been my experience that the GIS professional's role also involves the actual IT hardware configurations.

This leads us to our first GIS direction: niche. I believe the future GIS professionals will not be GIS professionals per se but IT specialists with a GIS or geospatial niche. The need for a dedicated geospatial person, at least in the engineering and technological fields, is already being thwarted by minimized budgets and a push for multitasking. This is not the case for every company. Many large companies have teams of GIS professionals providing services that have sustained over time, and for federal and local agencies, the need for dedicated GIS professionals will undoubtedly remain.

Like everything else, only time will determine to what extent and in what sectors the industry will go in this direction. One thing is clear, though: that the upcoming GIS professionals will need to diversify their capabilities in order to remain valid in the marketplace.

Explore

As GIS moves toward and into every feasible technical and nontechnical crevice, the understanding (even enlightenment) of geospatiality continues to occur. When GIS began, it was a geographic database; today it is a geospatial data infrastructure. The idea that environments can be geospatially linked to other information is exciting to most, especially when it offers forms of analysis that never existed before. We have school systems exploring the geospatial terrain in their districts, such as crime statistics, neighborhood demographics, district and nondistrict school information, and more. Today's industries are just beginning to understand the absolute value of geospatial data and how they can use that information to make better decisions and grow their businesses.

Agencies have traditionally been the geospatial leaders and are continuing to broaden their range of analysis. GIS has progressed from land use applications to such exciting applications as geospatially referenced rainforest inventories. Homeland security has emerged in the geospatial realm, using available and project-acquired data to ensure public safety, while remaining a few steps ahead of the malcontents.

Exploring the geospatial landscapes involves increased knowledge of how to use and manipulate data. Geospatial training is already a large market in today's marketplace and, I trust, will likely grow into a huge on-demand market. More firms and agencies will likely continue to invest money for geospatial awareness training, since the potential cost savings and advantages to proprietary information are great.

Wherever

Similar to GIS's "explore" direction is a trend that is voraciously gaining speed: moving from Earth's immediate surface to diverse terrains and topographies. GIS has traditionally been applied to Earthbound geography, but in past years GIS has emerged as the system of choice for various different topographies. For instance, NASA has used GIS to precisely map and conduct unique analysis of the Moon's surface characteristics and, with the aid of its two robotic Martian rovers, has geospatially mapped the surface of Mars.

This "wherever" direction is not necessarily macroscopic or solely applicable to extraterrestrial landscapes; it also applies to landscapes too small to see. The medical industry is in its infancy in using GIS technology. I'm not talking

about disease tracking and prevention analysis, but rather the inner landscapes of the body, such as geocoding a map of the human brain or the human body's constituents.

The scientific community is progressing with GIS, as well. The potential for a union of underwater microscopy and GIS has only begun to be explored. I foresee GIS being used to monitor and analyze the various life forms and living structures on Earth's declining coral reefs. The absolute potential for once unimaginable analysis is made available through geospatial technology.

But more must be done to move farther in this primary direction. As stated previously, training will play a giant role in progressing industry understanding. The medical and scientific fields are just two of the myriad disciplines GIS can benefit. The tricky part, however, is getting GIS integrated within these sectors. If that weren't complicated enough, professionals must work "outside the intellectual box" and be creative with their applications in order to expand the overall system benefits. Now, I agree that GIS may not prove to be the be all and end all tool, but I do think it will play a growing role in many fields new to GIS. As with any technology, the overall efficiency and usefulness of the technology is intimately dependent upon the user's capability.

Standards

An area where there is presently the greatest amount of activity and conflict is with GIS standards. Today's GIS industry presents numerous and varying standards that are supported by some organizations and disputed by others. The procedures, protocol, format, and implementation of various GIS elements, such as metadata, are continuously experiencing change and redirection. In fact, there are some geospatial standards prominent in the United States that differ from the prominent standards in Europe and Australia.

As a promising pathway for geospatial consistency, some globally focused organizations are gaining the support and validation needed to create new standards and fluid interoperability. In recent years, the Open Geospatial Consortium has made great strides in creating the OpenGIS, a global system dedicated to universal GIS compatibility. I recognize that defining global standards will afford all GISs and geospatial data sets a method to seamlessly interoperate with each another.

Similarly, there are numerous metadata formats that are progressing toward global standardization, such as the FGDC, Dublin Core Initiative, and ISO efforts. The Open Geospatial Consortium has even posted guidelines for metadata, which has proved to be one stepping stone for international metadata standardization. With each metadata schema, there are very different minimum requirements, in addition to differing metadata semantics and syntax. I foresee this direction progressing further.

Even though great strides have been made through the ISO metadata efforts, I sense that global standards may never occur, simply because it would require a good many organizations worldwide to buy into the structure, discipline, and protocol of a specific standard. If these universal standards are achievable, it will most likely be countless years before global standards truly manifest and get approval by the whole GIS community. That being stated, if this direction's destination is reached, it has the potential of redefining the way GIS is done in the years to come.

All Directions in Between

The GIS compass is unique in that between my primary anticipated directions (refer to Figure A.1) there are meldings—geospatial systems specialist, geospatial uniformity, terrain data homogeneity, and multidomain user—that offer likely products. The geospatial systems specialist, for instance, is a melding of two directions—a niche IT person specializing in all things geospatial. Likewise, geospatial uniformity, to me, involves a melding of geospatial exploration (all things geospatial) and standardization.

Terrain data homogeneity melds the idea of any terrain (on or off the Earth) with standards and homogeneous use of the data. The multidomain user direction melds the notion of using GIS wherever necessary with specialized (niche) yet multidisciplinary users.

As you may have realized already, it is my opinion that the progression of the four primary directions can be met through creative combinations of each. These combinations are manifesting today worldwide and, as a guess, will increase in both occurrence and availability. It will not be too long before doctors and building managers alike are using GIS to conduct business tasks, or geospatial products from different continents are cataloged the same exact way throughout the world. Apparently, there will also be a time in the not too distant future when GIS systems will read all formats without fail. Right now, these aspirations are in process and remain to fully manifest. I feel, given time, they will. You can say they are whims of fancy and hold no certainty as yet; however, one thing is definite: that GIS remains a true technical revolution.

Appendix A: Reference Ellipsoid Parameters

Considered a worldwide ellipsoid, WGS84 is one of the most widely used reference ellipsoids available. However, there are several major ellipsoids that are focused to specific areas and more locally used. Table A.1 highlights key reference ellipsoids and provides the associated semimajor axis (equatorial radius), semiminor axis (polar radius), and inverse flattening parameters. Please note that semiminor axis constants were calculated using the following formula:

$$Semiminor\ axis = Semimajor\ axis \times \left(1 - \frac{1}{Inverse\ flattening}\right)$$

Table A.1
Reference Ellipsoid Parameters

Ellipsoid Name	Semimajor Axis (meters)	Semiminor Axis (meters)	Inverse Flattening (f^{-1})
Airy 1830	6,377,563.396	6,356,256.909	299.3249646
APL 4.5 (1968)	6,378,144	6,356,757.339	298.23
Australian National	6,378,160	6,356,774.719	298.25
Average Terrestrial System 1977	6,378,135	6,356,750.305	298.257
Bessel 1841 (Ethiopia, Indonesia, Japan, and Korea)	6,377,397.155	6,356,078.963	299.1528128
Bessel 1841 (Namibia)	6,377,483.865	6,356,165.383	299.1528128

Table A.1 (continued)

Clarke 1858	6,378,235.6	6,356,560.14	294.2606768
Clarke 1858–Modified	6,378,293.645	6,356,617.938	294.26
Clarke 1866	6,378,206.4	6,356,583.8	294.9786982
Clarke 1880	6,378,249.145	6,356,514.87	293.465
Danish–Andrae (1876)	6,377,104.43	6,355,847.415	300
Delambre 1810	6,376,985.228	6,356,323.664	308.64
Everest	6,377,276.345	6,356,075.413	300.8017
Everest (India 1830)	6,377,276.345	6,356,075.413	300.8017
Everest (India 1956)	6,377,301.243	6,356,100.228	300.8017
Everest (Pakistan)	6,377,309.613	6,356,108.571	300.8017
Everest (Sabah and Sarawak)	6,377,298.556	6,356,097.55	300.8017
Everest (W. Malaysia 1969)	6,377,295.664	6,356,094.668	300.8017
Everest (W. Malaysia and Singapore 1948)	6,377,304.063	6,356,103.039	300.8017
Fischer 1960 (Mercury)	6,378,166	6,356,784.284	298.3
Fischer 1968	6,378,150	6,356,768.337	298.3
GEM 10C	6,378,137	6,356,752.314	298.2572236
Germaine (Djibouti)	6,378,284	6,356,589.156	294
GRS67	6,378,160	6,356,774.516	298.2471674
GRS75	6,378,140	6,356,755.288	298.257
GRS80	6,378,137	6,356,752.314	298.2572236
Helmert 1906	6,378,200	6,356,818.17	298.3
Hough 1960	6,378,270	6,356,794.343	297
Indonesian 1974	6,378,160	6,356,774.504	298.247
International 1924	6,378,388	6,356,911.946	297
Krassovsky 1940	6,378,245	6,356,863.019	298.3
Krayenhoff 1827	6,376,950.4	6,356,356.341	309.65
Modified Airy 1849	6,377,340.189	6,356,034.448	299.3249646
Modified Everest	6,377,304.063	6,356,103.039	300.8017
Modified Fischer 1960	6,378,155	6,356,773.32	298.3
NWL 10D	6,378,135	6,356,750.52	298.26
NWL 9D	6,378,145	6,356,759.769	298.25
OSU 86F	6,378,136.2	6,356,751.517	298.25722
OSU 91A	6,378,136.3	6,356,751.616	298.25722
Plessis Modified 1817	6,376,523	6,355,862.933	308.64
Plessis Reconstituted (Modified DeLambre 1810)	6,376,523.994	6,355,862.907	308.624807

Table A.1 (continued)

SGS 85	6,378,136	6,356,751.302	298.257
SGS 90	6,378,136	6,356,751.362	298.2578393
South American 1969	6,378,160	6,356,774.719	298.25
Struve 1860	6,378,298.3	6,356,657.143	294.73
Svanberg	6,376,797	6,355,837.971	304.2506
Walbeck 1819–AMS (1963)	6,376,896	6,355,834.847	302.78
Walbeck 1819–Planheft (1942)	6,376,895	6,355,834	302.7821565
War Office 1926 (McCaw 1924)	6,378,300	6,356,751.689	296
WGS60	6,378,165	6,356,783.287	298.3
WGS66	6,378,145	6,356,759.769	298.25
WGS72	6,378,135	6,356,750.52	298.26
WGS84	6,378,137	6,356,752.314	298.2572236
Xian 1980	6,378,140	6,356,755.288	298.257

Appendix B: Worldwide Geodetic Datums

Using geodetic datums correctly is an important part of working with GIS and producing accurate information products. Although similar, geodetic datums are distinct, having unique geodetic characteristics. Therefore, conversion between datums is needed. To convert from one datum to another, specific conversion parameters and the three-dimensional coordinate shifts between both datums must be known. In more complex conversions, where coordinate axes are not parallel and/or scaled, coordinate axes rotational parameters and a scale difference factor must be known as well.

Since the World Geodetic System 1984 uses an Earth-centered, Earth-fixed coordinate system, it is a highly used, worldwide datum (and reference ellipsoid). In fact, GPS is based upon WGS84. Due to its worldwide appeal, many applications require local datums to be converted to WGS84. As with any simple geodetic datum transformation, basic datum transformation parameters must be known.

Table B.1 contains a brief listing of geodetic datum transformation parameters relative to WGS84. It highlights some of the more conventional datums used throughout the world, but by no means is it a comprehensive listing of available datums. Further information and more complete local listings are freely distributed through the datum sources cited.

The following base parameters were used to create Table B.1:

$$WGS84 \ semimajor \ axis = 6{,}378{,}137$$

$$WGS84 \ flattening = 0.003352811$$

Table B.1
Worldwide Geodetic Datums

Location	Datum Name	Ellipsoid	Semimajor Axis (meters)	Flattening (f)	Datum Shift X	Datum Shift Y	Datum Shift Z
Africa							
Algeria	North Sahara 1959	Clarke 1880	6,378,249.145	0.003407561	–186	–93	310
Algeria	Voirol 1960	Clarke 1880	6,378,249.145	0.003407561	–123	–206	219
Botswana	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–138	–105	–289
Burkina Faso	Adindan	Clarke 1880	6,378,249.145	0.003407561	–118	–14	218
Burkina Faso and Niger (mean)	Point 58	Clarke 1880	6,378,249.145	0.003407561	–106	–129	165
Burundi	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–153	–5	–292
Cameroon	Adindan	Clarke 1880	6,378,249.145	0.003407561	–134	–2	210
Cameroon	Minna	Clarke 1880	6,378,249.145	0.003407561	–81	–84	115
Congo	Pointe Noire 1948	Clarke 1880	6,378,249.145	0.003407561	–148	51	–291
Djibouti	Ayabelle Lighthouse	Clarke 1880	6,378,249.145	0.003407561	–79	–129	145
Egypt	Old Egyptian 1907	Helmert 1906	6,378,200	0.00335233	–130	110	–13
Egypt	European 1950	International 1924	6,378,388	0.003367003	–130	–117	–151
Ethiopia	Adindan	Clarke 1880	6,378,249.145	0.003407561	–165	–11	206
Ethiopia (Eritrea)	Massawa	Bessel 1841 (Ethiopia, Indonesia, Japan, and Korea)	6,377,397.155	0.003342773	639	405	60
Gabon	M'Poraloko	Clarke 1880	6,378,249.145	0.003407561	–74	–130	42
Ghana	Legion	Clarke 1880	6,378,249.145	0.003407561	–130	29	364
Guinea	Dabola	Clarke 1880	6,378,249.145	0.003407561	–83	37	124
Guinea–Bissau	Bissau	International 1924	6,378,388	0.003367003	–173	253	27

Table B.1 (Continued)

Kenya	ARC 1960	Clarke 1880	6,378,249.145	0.003407561	–157	–2	–299
Lesotho	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–125	–108	–295
Liberia	Liberia 1964	Clarke 1880	6,378,249.145	0.003407561	–90	40	88
Madagascar	Tananarive Observatory 1925	International 1924	6,378,388	0.003367003	–189	–242	–91
Malawi	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–161	–73	–317
Mali	Adindan	Clarke 1880	6,378,249.145	0.003407561	–123	–20	220
Morocco	Merchich	Clarke 1880	6,378,249.145	0.003407561	31	146	47
Namibia	Schwarzeck	Bessel 1841 (Namibia)	6,377,483.865	0.003342773	616	97	–251
Nigeria	Minna	Clarke 1880	6,378,249.145	0.003407561	–92	–93	122
Senegal	Adindan	Clarke 1880	6,378,249.145	0.003407561	–128	–18	224
Sierra Leone	Sierra Leone 1960	Clarke 1880	6,378,249.145	0.003407561	–88	4	101
Somalia	Afgooye	Krassovsky 1940	6,378,245	0.00335233	–43	–163	45
South Africa	Cape	Clarke 1880	6,378,249.145	0.003407561	–136	–108	–292
Sudan	Adindan	Clarke 1880	6,378,249.145	0.003407561	–161	–14	205
Swaziland	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–134	–105	–295
Tanzania	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–175	–23	–303
Tunisia	Carthage	Clarke 1880	6,378,249.145	0.003407561	–263	6	431
Tunisia	European 1950	International 1924	6,378,388	0.003367003	–112	–77	–145
Zaire	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–169	–19	–278
Zambia	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–147	–74	–283
Zimbabwe	ARC 1950	Clarke 1880	6,378,249.145	0.003407561	–142	–96	–293
Antarctica							
Camp McMurdo Area	Camp Area Astro	International 1924	6,378,388	0.003367003	–104	–129	239
Deception Island	Deception Island	Clarke 1880	6,378,249.145	0.003407561	260	12	147

Table B.1 (Continued)

Asia							
Afghanistan	Herat North	International 1924	6,378,388	0.003367003	–133	–222	114
Bahrain Island	Ain el Abd 1970	International 1924	6,378,388	0.003367003	–150	–250	–1
Bangladesh	Indian	Everest (India 1830)	6,377,276.345	0.003324449	282	726	254
Brunei and East Malaysia	Timbalai 1948	Everest (Sabah and Sarawak)	6,377,298.556	0.003324449	–679	669	–48
Con Son Island (Vietnam)	Indian 1960	Everest (India 1830)	6,377,276.345	0.003324449	182	915	344
Hong Kong	Hong Kong 1963	International 1924	6,378,388	0.003367003	–156	–271	–189
India and Nepal	Indian	Everest (India 1956)	6,377,301.243	0.003324449	295	736	257
Indonesia	Indonesian 1974	Indonesian 1974	6,378,160	0.003352926	–24	–15	5
Indonesia (Sumatra)	Djakarta	Bessel 1841 (Ethiopia, Indonesia, Japan, and Korea)	6,377,397.155	0.003342773	–377	681	–50
Iran	European 1950	International 1924	6,378,388	0.003367003	–117	–132	–164
Iwo Jima	Astro Beacon "E" 1945	International 1924	6,378,388	0.003367003	145	75	–272
Japan, South Korea, and Okinawa (mean)	Tokyo	Bessel 1841 (Ethiopia, Indonesia, Japan, and Korea)	6,377,397.155	0.003342773	–148	507	685
Okinawa	Tokyo	Bessel 1841 (Ethiopia, Indonesia, Japan, and Korea)	6,377,397.155	0.003342773	–158	507	676
Oman	Oman	Clarke 1880	6,378,249.145	0.003407561	–346	–1	224
Philippines (except Mindanao Island)	Luzon	Clarke 1866	6,378,206.4	0.003390075	–133	–77	–51
Philippines (Mindanao Island)	Luzon	Clarke 1866	6,378,206.4	0.003390075	–133	–79	–72
Qatar	Qatar National	International 1924	6,378,388	0.003367003	–128	–283	22
Republic of Maldives	Gan 1970	International 1924	6,378,388	0.003367003	–133	–321	50
Russia	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	28	–130	–95
Saudi Arabia	Nahrwan	Clarke 1880	6,378,249.145	0.003407561	–243	–192	477
Saudi Arabia	Ain el Abd 1970	International 1924	6,378,388	0.003367003	–143	–236	7
Singapore	South Asia	Modified Fischer 1960	6,378,155	0.00335233	7	–10	–26

Table B.1 (Continued)

South Korea	Tokyo	Bessel 1841 (Ethiopia, Indonesia, Japan, and Korea)	6,377,397.155	0.003342773	–147	506	687
Sri Lanka	Kandawala	Everest (India 1830)	6,377,276.345	0.003324449	–97	787	86
Taiwan	Hu-Tzu-Shan	International 1924	6,378,388	0.003367003	–637	–549	–203
Thailand	Indian 1954	Everest (India 1830)	6,377,276.345	0.003324449	217	823	299
United Arab Emirates	Nahrwan	Clarke 1880	6,378,249.145	0.003407561	–249	–156	381
Vietnam (near 16° north)	Indian 1960	Everest (India 1830)	6,377,276.345	0.003324449	198	881	317
West Malaysia and Singapore	Kertau 1948	Everest (W. Malaysia and Singapore 1948)	6,377,304.063	0.003324449	–11	851	5
Atlantic Ocean							
Greenland	NAD27	Clarke 1866	6,378,206.4	0.003390075	11	114	195
Greenland (South)	Qornoq	International 1924	6,378,388	0.003367003	164	138	–189
Iceland	Hjorsey 1955	International 1924	6,378,388	0.003367003	–73	46	–86
Puerto Rico and Virgin Islands	Puerto Rico	Clarke 1866	6,378,206.4	0.003390075	11	72	–101
Trinidad and Tobago	Naparima, BWI	International 1924	6,378,388	0.003367003	–10	375	165
Australia							
Australia and Tasmania	Australian Geodetic 1966	Australian National	6,378,160	0.003352892	–133	–48	148
Australia and Tasmania	Australian Geodetic 1984	Australian National	6,378,160	0.003352892	–134	–48	149
Canada							
Alberta and British Columbia	NAD27	Clarke 1866	6,378,206.4	0.003390075	–7	162	188
Canada	NAD83	GRS80	6,378,137	0.003352811	0	0	0
Manitoba and Ontario	NAD27	Clarke 1866	6,378,206.4	0.003390075	–9	157	184
New Brunswick, Newfoundland, Nova Scotia, and Quebec	NAD27	Clarke 1866	6,378,206.4	0.003390075	–22	160	190

Table B.1 (Continued)

Northwest Territories and Saskatchewan	NAD27	Clarke 1866	6,378,206.4	0.003390075	4	159	188
Yukon	NAD27	Clarke 1866	6,378,206.4	0.003390075	–7	139	181
Caribbean and Central America							
Antigua, Barbados, Barbuda, Caicos Islands, Cuba, Dominican Republic, Grand Cayman, Jamaica and Turks Islands	NAD27	Clarke 1866	6,378,206.4	0.003390075	–3	142	183
Belize, Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua (mean)	NAD27	Clarke 1866	6,378,206.4	0.003390075	0	125	194
Canal Zone	NAD27	Clarke 1866	6,378,206.4	0.003390075	0	125	201
Cuba	NAD27	Clarke 1866	6,378,206.4	0.003390075	–9	152	178
Mexico	NAD27	Clarke 1866	6,378,206.4	0.003390075	–12	130	190
Mexico and Central America	NAD83	GRS80	6,378,137	0.003352811	0	0	0
Europe							
Albania	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	24	–130	–92
Austria, Denmark, France, West Germany, Netherlands, and Switzerland (mean)	European 1950	International 1924	6,378,388	0.003367003	–87	–96	–120
Austria, Finland, Netherlands, Norway, Spain, Sweden, and Switzerland (mean)	European 1979	International 1924	6,378,388	0.003367003	–86	–98	–119
Cyprus	European 1950	International 1924	6,378,388	0.003367003	–104	–101	–140
Czechoslavakia	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	26	–121	–78
rin0Czechoslavakia (prior to January 1, 1993)	S-JTSK	Bessel 1841 (Ethiopia, Indonesia, Japan and Korea)	6,377,397.155	0.003342773	589	76	480
England	Ordinance Survey of Great Britian 1936	Airy 1830	6,377,563.396	0.003340851	371	–112	434
England, Ireland, Scotland, Shetland Islands, and Channel Islands	European 1950	International 1924	6,378,388	0.003367003	–86	–96	–120

Table B.1 (Continued)

England, Wales, and Island of Man	Ordinance Survey of Great Britian 1936	Airy 1830	6,377,563.396	0.003340851	371	–111	434
Estonia	Coordinate System 1937 of Estonia	Bessel 1841 (Namibia)	6,377,483.865	0.003342773	374	150	588
Finland and Norway	European 1950	International 1924	6,378,388	0.003367003	–87	–95	–120
Greece	European 1950	International 1924	6,378,388	0.003367003	–84	–95	–130
Hungary	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	28	–121	–77
Ireland	Ireland 1965	Modified Airy 1849	6,377,340.189	0.003340851	506	–122	611
Italy	Rome 1940	International 1924	6,378,388	0.003367003	–225	–65	9
Italy (Sardinia)	European 1950	International 1924	6,378,388	0.003367003	–97	–103	–120
Italy (Sicily)	European 1950	International 1924	6,378,388	0.003367003	–97	–88	–135
Kazakhstan	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	15	–130	–84
Latvia	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	24	–124	–82
Malta	European 1950	International 1924	6,378,388	0.003367003	–107	–88	–149
Poland	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	23	–124	–82
Portugal and Spain	European 1950	International 1924	6,378,388	0.003367003	–84	–107	–120
Romania	S42 (Pulkova 1942)	Krassovsky 1940	6,378,245	0.00335233	28	–121	–77
Scotland and Shetland Islands	Ordinance Survey of Great Britian 1936	Airy 1830	6,377,563.396	0.003340851	384	–111	425
Wales	Ordinance Survey of Great Britian 1936	Airy 1830	6,377,563.396	0.003340851	370	–108	434
Yugoslavia (before 1990), Slovenia, Crotia, Bosnia, Herzegovina, and Serbia	Hermannskogel	Bessel 1841 (Namibia)	6,377,483.865	0.003342773	653	–212	449
Pacific Ocean							
American Samoa Islands	American Samoa 1962	Clarke 1866	6,378,206.4	0.003390075	–115	118	426
Easter Island	Easter Island 1967	International 1924	6,378,388	0.003367003	211	147	111
Guam	Guam 1963	Clarke 1866	6,378,206.4	0.003390075	–100	–248	259
Johnston Island	Johnston Island 1961	International 1924	6,378,388	0.003367003	189	–79	5
Marshall Islands	Wake-Eniwetok 1960	Hough 1960	6,378,270	0.003367003	102	52	–38
New Zealand	Geodetic Datum 1949	International 1924	6,378,388	0.003367003	84	–22	209

Table B.1 (Continued)

Viti Levu Island (Fiji Islands)	Viti Levu 1916	Clarke 1880	6,378,249.145	0.003407561	51	391	–36
South America							
Argentina	Campo Inchauspe	International 1924	6,378,388	0.003367003	–148	136	90
Argentina	South American 1969	South American 1969	6,378,160	0.003352892	–62	–1	–37
Baltra and Galapagos Islands	South American 1969	South American 1969	6,378,160	0.003352892	–47	26	–42
Bolivia	Provisional South American 1956	International 1924	6,378,388	0.003367003	–270	188	–388
Bolivia	South American 1969	South American 1969	6,378,160	0.003352892	–61	2	–48
Brazil	Corrego Alegre	International 1924	6,378,388	0.003367003	–206	172	–6
Brazil	South American 1969	South American 1969	6,378,160	0.003352892	–60	–2	–41
Chile	South American 1969	South American 1969	6,378,160	0.003352892	–75	–1	–44
Chile (Northern)	Provisional South American 1956	International 1924	6,378,388	0.003367003	–270	183	–390
Chile (Southern)	Provisional South American 1956	International 1924	6,378,388	0.003367003	–305	243	–442
Colombia	Bogota Observatory	International 1924	6,378,388	0.003367003	307	304	–318
Colombia	Provisional South American 1956	International 1924	6,378,388	0.003367003	–282	169	–371
Colombia	South American 1969	South American 1969	6,378,160	0.003352892	–44	6	–36
Ecuador	Provisional South American 1956	International 1924	6,378,388	0.003367003	–278	171	–367
Ecuador (excluding Galapagos Islands)	South American 1969	South American 1969	6,378,160	0.003352892	–48	3	–44
Guyana	Provisional South American 1956	International 1924	6,378,388	0.003367003	–298	159	–369
intblGuyana	South American 1969	South American 1969	6,378,160	0.003352892	–53	3	–47
Paraguay	Chua Astro	International 1924	6,378,388	0.003367003	–134	229	–29
Paraguay	South American 1969	South American 1969	6,378,160	0.003352892	–61	2	–33
Peru	Provisional South American 1956	International 1924	6,378,388	0.003367003	–279	175	–379
Peru	South American 1969	South American 1969	6,378,160	0.003352892	–58	0	–44
Suriname	Zanderij	International 1924	6,378,388	0.003367003	–265	120	–358
Uruguay	Yacare	International 1924	6,378,388	0.003367003	–155	171	37
Venezuela	Provisional South American 1956	International 1924	6,378,388	0.003367003	–295	173	–371

Table B.1 (Continued)

Venezuela	South American 1969	South American 1969	6,378,160	0.003352892	–45	8	–33
United States							
Alaska (except Aleutian Islands)	NAD27	Clarke 1866	6,378,206.4	0.003390075	–5	135	172
Alaska (except Aleutian Islands)	NAD83	GRS 80	6,378,137	0.003352811	0	0	0
Aleutian Islands	NAD83	GRS 80	6,378,137	0.003352811	–2	0	4
Bahamas (except San Salvador Island)	NAD27	Clarke 1866	6,378,206.4	0.003390075	–4	154	178
Continental U.S. (east of Mississippi River, including Louisiana, Minnesota, and Missouri)	NAD27	Clarke 1866	6,378,206.4	0.003390075	–9	161	179
Continental U.S. (west of Mississippi River, excluding Louisiana, Minnesota, and Missouri)	NAD27	Clarke 1866	6,378,206.4	0.003390075	–8	159	175
Florida and Bahamas (mean)	Cape Canaveral	Clarke 1866	6,378,206.4	0.003390075	–2	151	181
Hawaii	Old Hawaiian	Clarke 1866	6,378,206.4	0.003390075	89	–279	–183
Hawaii	NAD83	GRS 80	6,378,137	0.003352811	1	1	–1
Kauai, Hawaii	Old Hawaiian	Clarke 1866	6,378,206.4	0.003390075	45	–290	–172
Kauai, Hawaii	Old Hawaiian	International 1924	6,378,388	0.003367003	185	–233	–337
Maui, Hawaii	Old Hawaiian	Clarke 1866	6,378,206.4	0.003390075	65	–290	–190
Maui, Hawaii	Old Hawaiian	International 1924	6,378,388	0.003367003	205	–233	–355
Oahu, Hawaii	Old Hawaiian	Clarke 1866	6,378,206.4	0.003390075	58	–283	–182
Oahu, Hawaii	Old Hawaiian	International 1924	6,378,388	0.003367003	198	–226	–347

Glossary

absolute accuracy Map accuracy based on geographic coordinates and specific to an object's geographic location in relation to its true location on the surface of the Earth.

active pages Web pages that offer full user interaction, whereby the user can request information, as well as make changes to and manage the database.

aspect The relative orientation of the projected plane and the ellipsoid of revolution with respect to the location of an observer.

attributes Columns in a relational database management system (RDBMS) that contain distinct field data.

axis A line of symmetry within a shape.

axis of revolution *See* axis of rotation.

axis of rotation An axis upon which a two-dimensional shape is rotated to form a three-dimensional shape. For instance, an ellipse is rotated upon its major axis to form an ellipsoid.

azimuth The horizontal component of a line's direction, measured in degrees clockwise from the baseline; the baseline is typically the vertical axis (*y*-axis), and both the line and the baseline share a common center point.

azimuthal projection — *See* planar projection

bar scale — *See* graphic scale

bi-products — Developable information products formed as a result of a specific application data process.

Cartesian coordinate system — A two-dimensional, planar coordinate system, where horizontal and vertical distances are measured as *x* and *y* coordinates.

cartography — The organization, design, collection, portrayal, and reproduction of geographically based information in various formats, such as on graphic maps or within digital formats.

Cassini, Jean-Dominique (1625–1712) — An eighteenth-century patriarch of a renowned French geodesic family who erroneously concluded that the Earth pointed toward the poles and flattened at the equator. Cassini's Earth model was a prolate ellipsoid and resembled an egg shape. It was the favored model among the French geodetic community until the French Academy of Sciences proved him incorrect in 1745. Cassini was born Giovanni Domenico Cassini in Genoa, Italy, and after moving to France changed his name to Jean-Dominique.

centroid — The center of an area, region, or polygon represented as discrete *x-y* coordinates.

circle — A symmetrical two-dimensional shape in which any point on the outer shape is a set distance (radius) from a center point.

client-server system — A relational database type that is designed to effectively handle with speed and efficiency the many users accessing the central database at one time. This system uses a network to connect the user workstations (called clients) to the central database residing on a file server.

Codd, Edgar F. (1923–2003) — The author of "A Relational Model of Data for Large Shared Data Banks," who detailed a new method of simplifying database management using relational algebra for straightforward database functions. He is considered the forefather of the relational database management system. *See* Relational database management system (RDBMS)

compensatory scale factors Variant scale factors that compensate for one another to preserve a projection's area characteristic. These scale factors are typically greater than 1.0 in one direction and less than 1.0 in another direction. In equal-area projections, using compensatory scale factors will project a circle on a globe as an ellipse on a plane surface.

computer-aided design (CAD) Computer-based design software used to produce various technical and design drawings, specifically in the engineering, architectural, and environmental fields.

conformal projection A projection classification that maintains the shape and angles of a small area while minimizing distortion. As regions become larger, projections and area shapes become more distorted.

conic projection A type of map projection in which the globe is represented at an exact scale along a particular parallel (standard parallel) between the equator and a pole. Distortion increases away from that standard parallel. It can be envisioned that this map projection is made by projecting the features of the Earth onto a conic wrap (like a birthday hat) then cutting it to make a flat map.

continuous data A numeric form of data usually associated with the physical measurement of boundaries that are not well defined.

coordinate A number set that represents a specific location within a reference system. An *x-y* set is used in a planar coordinate system and an *x-y-z* set is used in a three-dimensional coordinate system.

curvilinear surface A surface consisting of or bounded by curved lines.

cylindric projection A type of map projection where the globe is represented by equidistant, parallel straight lines that intersect one another at right angles. The projection is true at the equator and increases in distortion toward the poles. It can be envisioned that this map projection is made by projecting the features of the Earth onto a cylindrical wrap then cutting it to make a flat map.

database management system (DBMS)	A program or collection of programs that enables the user to save, modify, classify, select, and extract information from a central database. DBMSs are used throughout the world and throughout almost every industry for the purposes of centralizing information and automating processes.
datum	Any set of numeric or geometric parameters used to accurately measure or define another quantity.
developable projection surfaces	A shape that can be flattened as a surface without any residual distortion. Predominant shapes include cones, cylinders, and planes.
digital elevation model (DEM)	Digital terrestrial or terrain elevation data that is typically defined by *z*-values and referenced to a common datum.
digital terrain model (DTM)	A digital, three-dimensional model of the Earth's surface. DTMs are typically used to define terrain and landform reliefs.
discrete data	Data about geographic features with well-defined boundaries that are represented by points, lines, or polygons.
distortion	The warping of shape, direction, scale, or area on a map projection or image in relation to real-world measurements.
distortion magnitude	The amount and density of distortion in a projection.
distortion pattern	A discernible pattern in a projection's distortion, such as the misrepresentation of landforms near the poles or distortion of areas.
dynamic pages	Web pages that offer some user interaction, whereby the pages display the most recent or dynamic database information as defined by the user.
Earth model	A physical shape that wholly represents the Earth and is used for mathematical and scientific purposes.
Earth-Observing Satellites (SPOT)	*See* Satellite Pour L'Observation de la Terre.
Eccentricity	An ellipse-specific constant that measures the degree of flattening as the relationship between the focal radius and the semimajor axis.

ellipse A two-dimensional shape that is flattened on one axis. The shape resembles a flattened circle and consists of a major and a minor axis.

ellipsoid A three-dimensional shape that is flattened on one axis. The shape resembles a flattened sphere and is formed from an ellipse rotated around an axis (axis of rotation). Like the ellipse, it consists of a major and minor axis.

ellipsoid of revolution An ellipsoid that revolves around a major axis. *See* axis of rotation

enterprise system A system that supports a wide variety of users having different functional requirements over an entire organization.

equal-area projection A projection classification determined when the relative sizes of all features on a globe are maintained during the projection process. An equal-area projection vigilantly retains area properties of the ellipsoid on the plane surface through the use of compensatory scale factors. This projection classification may distort shape, angle, and/or scale.

equator An imaginary line that covers the widest circumference on Earth. It is indicated as 0° latitude.

equatorial Relating to or dependent upon the equator.

equidistant projection A projection classification determined when scale is true along at least one line (a focal line) or from one or two points (focal points) to all other points on the projected surface.

Eratosthenes (276–194 B.C.) An innovative third-century B.C. Greek mathematician who made a nearly accurate calculation of the Earth's circumference using base scientific understanding, crude measuring instruments, and meticulous observation. His observations occurred in Egypt, within his home town of Syene (now Aswan) and at Alexandria, a town roughly 500 miles away.

feature file A file that contains geographic object feature information, such as representative point, line, and polygon information.

file-based system The simplest of relational database types that uses a single file for the database tables, queries, security information, and forms/reports.

finite Within a limit or set parameter.

first eccentricity *See* eccentricity.

flattening A measurement of axial compression in an ellipse that is directly related to the differences in both the semimajor and semiminor axes.

focal radius The distance from the center point to the foci (i.e., F_1, F_2) in an ellipse.

geocode Geospatially coded information products.

geodesy The division of science associated with the measurement and portrayal of the Earth. Geodesy, known also as geodetics, is intimately concerned with the determination of the size and shape of the Earth, as well as its elements. These Earth-based elements include its terrestrial gravity, magnetic field, tides, geologic and crustal movement, and polar motion.

geographic coordinate system A geographic reference system that uses latitudal and longitudal measurements to define point locations on a sphere or ellipsoid.

geographic data Geographic-related information about a map's features, including point locations, shapes, and descriptions.

geographic information *See* geographic data.

geographic information system (GIS) A collection of computer hardware, software, and geographic data for capturing, storing, updating, manipulating, analyzing, and displaying all forms of geographically referenced information. GIS is, in essence, a central repository of geographic data collected from various sources, including satellites, the global positioning system (GPS), and topographical maps.

geography The study of the Earth's surface that furnishes information about the Earth and distinguishes how features upon the Earth correlate with one another. It is the founding science to GISs.

geoid An exacting model of the Earth defined by the mean sea level. The model continues through the continents and is not defined by true topological heights or standard shape references.

geomensuration The measurement of the Earth as a whole, such as on a global level, that offers credible measurements useable toward precise geodetic positioning, surface shape, and size.

geometric A term that represents the spatial relationship between features and objects.

geometric deformation indicators The circular or elliptical indicators on an indicatrix. A circle represents conformality (no distortion in shapes), while an ellipse represents local distortion at the intersection where the object is plotted. These circular or elliptical objects are fixed at a finite scale with explicit locations in the projected surface.

geoprocessing The fundamental process of creating a derived set of geographic data from various existing data sets using operations such as feature overlay and data conversion.

georeferencing The process of referencing images, spatially transforming them into a defined coordinate referencing system, and integrating and overlaying them in a GIS environment.

geospatial analysis The process of analyzing and interpreting geospatial models.

global positioning system (GPS) Handheld and mobile devices that communicate with a system of satellites orbiting the Earth to record *x*, *y*, and *z* coordinates and calculate precise positional data. GPS devices are used for a multitude of applications, including surveying, mapping, hiking, and navigation.

globe A scaled, physical representation of the Earth.

graphic scale A map scale that depicts a bar of a certain length to represent a distance length. Graphic scales are commonly called bar scales and enable the map user to visually recognize distances with relative ease.

graticule A basic grid of meridians and parallels.

hemisphere One-half of a symmetrical shape, such as an ellipsoid or sphere. Earth is divided by its equator, forming a northern and a southern hemisphere.

horizontal control datum Any datum that serves as a geodetic reference for a precise northerly or southerly location (latitude) and easterly or westerly location (longitude).

index file A file that contains unique identifiers that comprise more detailed product information and help speed spatial feature queries.

indicatrix A visual distortion analysis tool that depicts a projected graticule as a distortion graph and shows distortion as circles or ellipses at the specific intersections in the plane.

inverse flattening The reciprocal of flattening used to minimize human error and help simplify conveyance.

irregular triangular surface model *See* triangulated irregular network.

isarithms *See* isoline.

isoline A line on any surface that connects points of equal value.

large-scale A project or map signifier that refers to applications involving small areas of focus, such as a community, a city block, or a university campus. A typical large-scale representative fraction can be assumed to be anything less than 1:50,000.

latitude The distance along a meridian that is either north or south of the equator. Latitudes are measured in degrees and are commonly called parallels.

layer A set of theme-based data that is described and stored in a map's library. Layers may contain a basemap or satellite image or can include such information as roadways, pipelines, soil sample locations, soil types, vegetation types, and water distribution.

legend A map reference that lists and distinguishes various map colors, lines, shapes, patterns, symbols, and annotation. Many times, the legend highlights map scale, source, projection, and other information.

light intensity detection and ranging (LIDAR) Remote sensing tool that uses a laser to precisely measure distances to various surfaces and terrain.

linked attribute table A matrixed table that contains explicative attributes for a group of spatial features.

longitude The angular distance between a target meridian and the prime meridian (0°). Longitude lines converge at the north and south poles, with measurements typically expressed in degrees, minutes, and seconds.

major axis A line of symmetry that passes through the widest portion of a shape. In an ellipse, the axis passes through both foci.

map projection A process that transforms a three-dimensional object, such as the Earth's globe, into a two-dimensional format. Some projections preserve shape; others preserve the accuracy of distance, area, or direction.

map scale The relationship between a distance on a map and its corresponding distance on the ground. It relates the size of map-based features with their actual, real-world size.

map space A platform for geographic data in depicting Earth features on a distinct coordinate system.

meridian A vertical line extending from the north pole to the south pole. Locations along the same meridian are said to have the same longitude.

metadata Information about a specific data set, including the geographic data's source, resolution, scale, and proper use.

mid-scale A project or map signifier that refers to applications that are not overly large or small, but rather fall right in the middle. A typical representative fraction ranges from 1:50,000 to 1:500,000.

minor axis A line of symmetry that passes through the shortest portion of a shape. In an ellipse, the minor axis passes through the center point and is perpendicular to the major axis.

multitier system A relational database type that functions similarly to the client-server system, except that there are middle servers, which contain software to perform certain tasks and process transmitted data. *See* three-tier system.

natural resources Natural elements of the Earth, such as rivers, streams, wetlands, and landforms.

natural scale *See* ratio scale.

Newton, Sir Isaac (1642–1727) Considered the father of modern mathematics, his gravitational theory published in *Philosophiae Naturalis Principia Mathematica* became the backbone for Britain's Earth model. The British model was an oblate ellipsoid—grapefruit-shaped and bulging in the middle. In 1745, the French Academy of Sciences proved the British model was correct and, subsequently, validated Newton's theory.

normal aspect A projection aspect based upon a standard line along which distortion is minimized. The farther you move away from the standard line, the greater the distortion.

oblate ellipsoid A grapefruit-shaped ellipsoid that bulges in the middle and is flat on top and bottom.

oblique aspect A projection aspect used when the axis of the Earth and the axis of the projection are oriented in an arbitrary manner. This aspect type is uniquely suited for geographic areas that are centered along lines that are neither meridians nor parallels. Consequently, this aspect type tends to be directed northwest, northeast, southwest, or southeast.

orthomorphic projection *See* conformal projection.

orthophotography The photographic process by which aerial photographs are rectified to produce an accurate image of the terrain. The process removes the Earth's tilt and any land relief displacements that occurred at the time the photo was taken.

parallel Two or more lines wholly separated by a constant distance. Geographically speaking, latitude lines, such as the equator, are in parallel with one another.

photogrammetry The process of identifying objects and measuring their true physical dimensions from photographs.

planar projection A type of map projection that is an azimuthal orientation of a projected surface. This projection type is true only at the center point and protects the integrity of azimuths, bearings, and directions from a central point to other more remote points in the plane. It can be envisioned that this

	map projection is made by projecting the features of the Earth onto a flat, touching plane, like a piece of flat paper touching a globe.
planar space	A map space utilizing a horizontal (*x*) and vertical (*y*) axis.
planar surface	A two-dimensional surface defined by an *x-y* coordinate grid.
plane coordinates	Positional measurements from the origin (0,0) in a planar coordinate system. Coordinates are typically represented by an *x-y* pair.
plane of the ecliptic	The imaginary line that defines the elliptical path the Earth takes as it orbits the Sun.
polar aspect	A projection aspect that centers on a single point, either the north pole or the south pole. This aspect retains symmetry about any meridian and distorts between them.
poles	The points where the axis of rotation intersects the surface of a three-dimensional shape.
primary key	A unique attribute in a relational database management system (RDBMS) that distinguishes each record. They can be sequential record numbers or user-defined values that are unique and, thus, not duplicated anywhere within the same attribute throughout the records.
prime meridian	A reference longitudal line that flows through Greenwich, England. The prime meridian is typically called the Greenwich meridian and is located at 0° east and west.
projection aspect	The relative orientation of the projected plane and the ellipsoid of revolution with respect to the location of an observer.
projection type	A term that relates to the distinct types of map projections available for use on the projected plane, such as the cylindric, conic, and planar types.
prolate ellipsoid	An egg-shaped or elongated ellipsoid that is flattened in the middle.
quadrangle	A map that defines a focal area's topography within latitude and longitude lines. It usually refers to the U.S. Geological Survey's 7.5-minute series or 15-minute series.

raster data Spatial data represented as a grid of valued pixels or cells. Each cell contains an attribute value and positional coordinates. These positional coordinates are wholly associated with the grid and its defined order.

ratio scale A map scale that reduces a map's distance to 1. One unit on the map corresponds to a distance in the same units on the ground. In a typical 30-minute series topographic map, the representative fraction is shown as either 1/125,000 or 1:125,000, whereby the 1 and 125,000 are the same unit of measure. A ratio scale is more formally known as a representative fraction and less formally as a natural scale.

reference ellipsoid A standardized ellipsoid with officially proved and measured parameters. The reference ellipsoid enables GIS to define point locations with regional accuracy.

region of interest The area where focus has been placed.

relational database management system (RDBMS) A program or collection of programs that enables the user to build data relationships to efficiently save, modify, classify, select, and extract information from a central relational database.

relative accuracy Map accuracy specific to an object's geographic location in relation to the true location of other objects on Earth.

relief map A map detailing the elevations and depressions of the Earth's surface with contour lines, shading, and digital terrain modeling. Relief maps are and appear to be three dimensional in nature.

remote sensing Any form of distance-based or remote viewing, observation, and measurement of the Earth. Typical forms include aerial photographs, satellite images, global positioning systems, LIDAR, and sonar.

representative fraction *See* ratio scale

resolution A measurement of image sharpness, usually indicated by pixels per line or dots per inch.

Satellite Pour L'Observation de la Terre (SPOT) A series of Earth-observing satellites that offer the ability to observe the whole Earth (minus a strip over the equator) in a single day. These satellites are part of a remote sensing program first designed by France in partnership with

	Belgium and Sweden. Current satellites include the SPOT1 and SPOT2 units. Use of the SPOT3 satellite ceased in 1996.
scale reduction	The process of proportionally reducing actual dimensions into much smaller, more convenient dimensions.
second eccentricity	An ellipse-specific constant that measures the degree of flattening as a relationship between the focal radius and the semiminor axis.
semimajor axis	Equal to one-half of the major axis.
semiminor axis	Equal to one-half of the minor axis.
sidereal day	A true day measurement, with a length of 23 hours, 56 minutes, and 4.09 seconds.
sidereal year	A true year measurement, with a length of 365.2564 days.
Six Sigma	Stands for six standard deviations from the mean or average and involves a measure of quality that strives for near perfection. This quality protocol is a disciplined, data-driven methodology for eliminating shortcomings and deficiencies in any business process.
small-scale	A project or map signifier that refers to applications involving relatively large areas, such as a map of the world, the continent of Australia, or the city of Rome, Italy. A typical small-scale representative fraction can be assumed to be anything greater than 1:500,000.
Snow, John (1813–1858)	A British physician who mapped all cholera deaths on a hand-drawn street map of a London neighborhood and helped end the epidemic of 1854.
spatial	A term used to describe anything located within a space that has physical, measurable dimensions.
spatial analysis	The study of geographic feature locations, attributes, and spatial dimensions and the relationships between each.
sphere	A symmetrical three-dimensional shape, in which the distance (radius) from the surface to a common center point is identical anywhere on the object. It can be thought as a circle rotated around an axis.
spheroid	*See* ellipsoid

standard axis *See* axis

standard line A line of exact scale that is represented by a true scale factor value of 1.

State Plane Coordinate System A two-dimensional, planar coordinate system established by the U.S. government to define positions on, above, and below the Earth's surface by *x-y* coordinates.

static pages Web pages that display database-driven information that cannot be changed, altered, or modified by the user.

Structured Query Language (SQL) A relational programming language used in complex database management systems. It is pronounced "Sequel."

Summer Solstice The longest day of the year, which occurs when the Sun is farthest north. In the northern hemisphere, the day is around June 21. In the southern hemisphere, the day is around December 22.

surveying The measurement of individual parts on the Earth's surface that offers credible measurements useable toward precise geodetic positioning of surface shape and size of the ever-changing physical terrain.

tagged-image file format (TIFF) A standard raster data file format that supports black-and-white, grayscale, limited color, and true color images.

thematic mapping The process of creating specialized maps using select information (i.e., a basemap) relating to one or more explicit themes. Themes include such information as roadways, pipelines, soil sample locations, soil types, vegetation types, and water distribution.

thick clients Users with a majority of processing capabilities in a relational database management system (RDBMS), such as is in the client-server and file-based relational database systems.

thin clients Users with reduced processing capabilities in a relational database management system (RDBMS), such as in the multitier relational database type.

three-dimensional Term pertaining to a spatial object or phenomenon dealing with three physical dimensions, typically *x*, *y*, and *z*.

three-dimensional mapping Spatial mapping that is detailed against three spatial coordinate planes, usually *x*, *y*, and *z*. Sometimes the *z* plane is substituted for an *m*, *n*, or k plane.

three-dimensional space A map space utilizing a horizontal (*x*), vertical (*y*), and perpendicular (z) axis.

three-tier system A relational database setup involving three distinct tiers: the client or user tier, the application tier, and the development tier. *See* multitier system.

Tissot indicatrix A helpful tool for visually modeling distortion and for graphically analyzing the distortion properties in a projected surface.

Tissot, Nicolas Auguste A nineteenth-century cartographer best known for his groundbreaking approach to distortion analysis, known now as the Tissot indicatrix, in his *Memoire sur la Representation des Surfaces et les Projections des Cartes Geographiques.* His indicatrix became a helpful tool for visually modeling distortion and for graphically analyzing the distortion properties in a projected surface.

Tobler's First Law of Geography A fundamental law of geography that states that things close together in space tend to be more alike than things that are far apart.

topographic map *See* quadrangle

topography A definition of the terrain's shape or configuration, typically represented in map form by layered contour lines.

topology The spatial relationships between contiguous or adjacent geographic features.

transformation The conversion of map or image coordinates from one coordinate system to another. Standard transformation techniques include rotation, skewing, shifting, and scaling.

transverse aspect The projection aspect when a standard line (normal) aspect is rotated 90°. This transverse aspect projection minimizes distortion along meridians and can have azimuthal or equatorial alignment.

triangulated irregular network (TIN) A network or series of continuous, nonoverlapping triangles used to define elevation data over a focus area. These

	triangles are used for surface representation and display, sometimes called an irregular triangular surface model.
true scale	A scale that is devoid of distortion, represented by a scale factor of 1.
tuples	Rows in a relational database management system (RDBMS) that contain the individual records.
two-dimensional	Term pertaining to a spatial object or phenomenon dealing with only two physical dimensions, typically *x* and *y*.
two-dimensional space	*See* planar space
UTM coordinate system	A two-dimensional, positional coordinate system based upon the Universal Transverse Mercator (UTM) map projection. The UTM system partitions the earth's surface into 60 zones wherein each zone (cell) is roughly 6° east to west and 8° north to south.
vector data	Spatial data expressed as an ordered list of coordinate points (*x-y* coordinates) to represent geographic features with associated attributes. Geographic features are represented by points, lines, and polygons.
vector space	A platform for geographic vector data, which use *x-y* coordinates with lines and shapes to depict Earth features. It stores nontopological coordinate geometry and attribute information for spatial features.
verbal scale	A map scale that expresses the map-to-ground relationship in words, such as "One inch represents 75 miles" or "One inch equals 75 miles."
vertical control datum	Any datum that serves as a geodetic reference for elevations or height above and below the reference shape, be it a geoid or standard ellipsoid.
Winter Solstice	The shortest day of the year, which occurs when the Sun is farthest south. In the northern hemisphere, the day is around December 22. In the southern hemisphere, the day is around June 21.
zenithal projection	*See* planar projection

Bibliography

Advanced Information Systems, et al., *Oracle Unleashed*, 2nd ed., Indianapolis, IN: SAMS Publishing, 1997.

American National Standards Institute, *The Dublin Core Metadata Element Set,* ANSI/NISO Z39.85-2001, Bethesda, MD: NISO Press, 2001.

Angel, A. R., and S. R. Porter, *A Survey of Mathematics with Applications*, 6th ed., Reading, MA: Addison Wesley Longman, 2001.

Anton, H., *Calculus with Analytic Geometry*, 3rd ed., New York: John Wiley & Sons, 1988.

Audet, R., and G. Ludwig, *GIS in Schools*, Redlands, CA: ESRI Press, 2000.

Australian Drilling Industry Training Committee Ltd., *Drilling: The Manual of Methods, Applications, and Management*, 4th ed., A. W. Starkweather, Jr., (ed.), Boca Raton, FL: CRC Press—Lewis Publishers, 1997.

Australian Government Geoscience Australia, "Parameters for Commonly Used Ellipsoids," http://www.ga.gov.au/geodesy/datums/ellipsds.jsp (accessed June 27, 2005).

Baumann, G., and J. Weinstein, *Poised for Growth: Taking Your Business to the Next Level*, New York: McGraw-Hill, 1997.

Baumann, J., "GIS Helps Transportation Industry Keep on Truckin' in Czech Republic," *DirectionsMag.com*, http://www.directionsmag.com/article.php?article_id=532 (accessed May 2, 2005).

Belt, K., and S. T. Paxton, "GIS as an Aid to Visualizing and Mapping Geology and Rock Properties in Regions of Subtle Topography," *GSA Bull.*, Vol. 117, Nos. 1 and 2, 2005.

Berry, J. K., "Use 'Shadow Maps' to Understand Overlay Errors," *GeoWorld*, Vol. 17, No. 9, September 2004, pp. 18–19.

Board of Earth Sciences and Resources, *Down to Earth: Geographical Information for Sustainable Development in Africa*, Washington, D.C.: The National Academies Press, 2002.

Borkowski, K. M, "Accurate Algorithms to Transform Geocentric to Geodetic Coordinates," *Bull. Geod.,* Vol. 63, 1989, pp. 50–56.

Bothamley, J., *Dictionary of Theories*, London, U.K.: Gale Research International, 1993.

Boucher, C., et al., *The ITRF2000,* IERS Technical Note No. 31, Marne-la-Valle, France: International Earth Rotation and Reference Systems Service, 2004.

Brainerd, J., and A. Pang, *Floating Ring: A New Tool for Visualizing Distortion in Map Projections*, Santa Cruz, CA: Computer Science Department University of California, ftp://ftp.cse.ucsc.edu/pub/reinas/papers/cgi98.map.ps.Z.

Brainerd, J., and A. Pang, *Interactive Map Projections and Distortion*, Santa Cruz, CA: Computer Science Department University of California, ftp://ftp.cse.ucsc.edu/pub/reinas/papers/cgi98.map.ps.Z.

Bray, T., "RDF and Metadata," *XML.com*, August 1998, http://www.xml.com/pub/a/98/06/rdf.html (accessed November 10, 2005).

Brennan, P., "Dynamic Segmentation in ArcGIS," *ArcUser*, October–December 2002, pp. 52–54.

Briggs, R., "The Future of GIS: If I Really Knew, Would I Still Be Here?," University of Dallas, GISC 6381 GIS Fundamentals, 2004, http://www.utdallas.edu/~briggs/poec5319/future.ppt (accessed October 25, 2005).

Burke, R., *Getting to Know ArcObjects: Programming ArcGIS with VBA*, Redlands, CA: ESRI Press, 2003.

Bursa, M., "Dimension of the Earth's Mean Ellipsoid Specified by Brovar's Modified Conditions," *Acta Geod. Geoph. Hung.*, Vol. 39, No. 1, 2004, pp. 15–25.

Bursa, M., et al., "Dimension of the Earth's General Ellipsoid," *Earth, Moon, Planets*, Vol. 91, 2002, pp. 31–41.

Bush, G. W., "Executive Order 13286: Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure (Amended)," *Federal Register*, Vol. 68, No. 43, March 5, 2003, pp. 10619–10633.

Butler, A., "Metadata Freedom: Exchange Data Without National Standards," *GeoWorld*, Vol. 18, No. 8, August 2005, pp. 36–39.

Butler, A., et al., "The National Geospatial Coordinating Council: A Dramatic New Approach to Build the NSDI," *GeoWorld*, Vol. 18, No. 10, October 2005, pp. 38–41.

Butte County Association of Governments, "GIS Enhances Transportation Planning," *BCAG Bulletin*, Fall 1997, p. 1.

Clarke, K. C., *Getting Started with Geographic Information Systems*, Upper Saddle River, NJ: Prentice Hall, 1997.

Clarke, K. C., S. L. McLafferty, and B. J. Tempalski, "On Epidemiology and Geographic Information Systems: A Review and Discussion of Future Directions," *Emerging Infectious Dis.*, Vol. 2, No. 2, April–June 1996, pp. 85–92.

Clinton, W. J., "Executive Order 12906: Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," *Federal Register*, Vol. 59, No. 71, April 13, 1994, pp. 17671–17674.

Clinton, W. J., "Executive Order 12951: Release of Imagery Acquired by Space-Based National Intelligence Reconnaissance Systems," *Federal Register*, Vol. 60, No. 39, February 28, 1995, pp. 10789–10790.

Cloud, J., "American Cartographic Transformations During the Cold War," *Cartography and Geographic Information Science*, Vol. 29, No. 3, 2002, pp. 261–282.

Codd, E. F., "A Relational Model of Data for Large Shared Data Banks," *Commun. ACM*, Vol. 13, No. 6, June 1970, pp. 377–387.

Coddington, W., *Environmental Marketing: Positive Strategies for Reaching the Green Consumer*, New York: McGraw-Hill, 1993.

Cova, T. J., and R. L. Church, "Modeling Community Evacuation Vulnerability Using GIS," *Int. J. Geographical Information Sci.*, Vol. 11, No. 8, 1997, pp. 763–784.

Crossley, R., "Future of GIS," *GeoCommunity*, http://spatialnews.geocomm.com/features/futuregis.html (accessed November 5, 2005).

Dana, P. H., "Geodetic Datum List," *The Geographer's Craft Project*, Department of Geography, University of Texas–Austin, July 12, 1997, http://www.colorado.edu/geography/gcraft/notes/datum/dlist.html (accessed June 23, 2005).

Dana, P. H., "Reference Ellipsoids and Geodetic Datum Transformation Parameters (Local to WGS-84) from NIMA 8350.2 4 July 1977 and MADTRAN 1 October 1996," http://www.colorado.edu/geography/gcraft/notes/datum/edlist.html (accessed June 23, 2005).

Daniels, T., and K. Chhabra, "A Mapping Marriage: Cartography and Mapping Forge a Beneficial Union," *GeoWorld*, Vol. 18, No. 10, October 2005, pp. 34–37.

Davis, B. E., *GIS: A Visual Approach*, Santa Fe, NM: OnWord Press, 1996.

Davis, B. E., *GIS for Everyone,* 3rd ed., Redlands, CA: ESRI Press, 2003.

Defense Mapping Agency, "Datums, Ellipsoids, Grids, and Grid Reference Systems," *Defense Mapping Agency Technical Manual 8358.1*, Washington, D.C.: Defense Mapping Agency, 1990.

Defense Mapping Agency, "Error Theory as Applied to Mapping, Charting, and Geodesy," *Defense Mapping Agency Technical Manual 8400.1*, Washington, D.C.: Defense Mapping Agency, 1991.

Defense Mapping Agency, Geodesy for the Layman, Defense Mapping Agency Technical Report 80-D03, Washington, D.C.: Defense Mapping Agency, 1984.

Defense Mapping Agency, "The Universal Grids: Universal Transverse Mercator (UTM) and Universal Polar Stereographic (UPS)," *Defense Mapping Agency Technical Manual 8358.2*, Washington, D.C.: Defense Mapping Agency, 1989.

Defense Mapping School, Department of Defense, *Basic Geodesy,* Student Pamphlet 4MF1-F-010-010, Washington, D.C.: National Oceanic and Atmospheric Administration, 1977.

Dewhurst, W. T., *NADCON: The Application of Minimum Curvature-Derived Surfaces in the Transformation of Positional Data from the North American Datum of 1927 to the North American Datum of 1983,* NOAA Technical Report NOS NG-50, Rockville, MD: National Oceanic and Atmospheric Administration, 1990.

DiBiase, D., "Lesson 2: Scales and Projections," Geography 5121 Course, Pennsylvania State University, 2004.

DiBiase, D., "Lesson 3: Census Data and Thematic Maps," Geography 5121 Course, Pennsylvania State University, 2004.

Digital Geographic Information Working Group, *The Digital Geographic Information Exchange Standard (DIGEST): Part 3 Codes and Parameters*, 2.1 ed., September 2000.

Downing, D., *Dictionary of Mathematics Terms*, 2nd ed., Hauppauge, NY: Barron's Educational Series, 1995.

Dracup, J. F., *Preprint—Understanding the State Plane Coordinate Systems,* Rockville, MD: U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, 1994.

Dracup, J. F., *Horizontal Control,* NOAA Technical Report NOS 88 NGS 19, Rockville, MD: National Oceanic and Atmospheric Administration, 1980.

Environmental Systems Research Institute, "About Topology," ArcGIS Desktop Help, *ArcMap 9.0 Build 560,* Redlands, CA: ESRI Press, 2004.

Environmental Systems Research Institute, *ArcIMS 4 Architecture and Functionality*, ESRI White Paper, Redlands, CA: ESRI Press, May 2003.

Environmental Systems Research Institute, *ArcMap 9.0 Build 560,* Redlands, CA: ESRI Press, 2004.

Environmental Systems Research Institute, *Dictionary of GIS Terminology*, H. Kennedy, (ed.), Redlands, CA: ESRI Press, 2001.

Environmental Systems Research Institute, "Environmental Site Assessment," *Environmental Observer*, Winter 2005, pp. 4–5.

Environmental Systems Research Institute, *ESRI Map Book Volume 16–Geography–Creating Communities*, Redlands, CA: ESRI Press, 2001.

Environmental Systems Research Institute, *ESRI Map Book Volume 17–Geography and GIS–Sustaining Our World*, Redlands, CA: ESRI Press, 2002.

Environmental Systems Research Institute, *ESRI Map Book Volume 18–Geography and GIS–Serving Our World*, Redlands, CA: ESRI Press, 2003.

Environmental Systems Research Institute, *ESRI Shapefile Technical Description*, ESRI White Paper, Redlands, CA: ESRI Press, July 1998.

Environmental Systems Research Institute, *Metadata and GIS*, ESRI White Paper, Redlands, CA: ESRI Press, October 2002.

Environmental Systems Research Institute, "To Monitor and Inspect Underground Storage Tanks, the Illinois Office of the State Fire Marshal Uses GIS," *ArcNews*, Vol. 25, No. 3, Fall 2003, pp. 15.

Environmental Systems Research Institute, "Topology Rules," ArcGIS Desktop Help, *ArcMap 9.0 Build 560,* Redlands, CA: ESRI Press, 2004.

European Space Agency, "Benefits for Europe," http://www.esa.int/esaCP/SEMO75XO4HD_Benefits_0.html (accessed February 7, 2005).

Executive Office of the President, Office of Management and Budget, *Circular A-16 Revised*, Washington, D.C.: Executive Office of the President, 2002.

Federal Energy Regulatory Commission, *Guidance for Shoreline Management Planning at Hydropower Facilities*, Washington, D.C.: FERC, 2001.

Federal Geographic Data Committee, *Content Standard for Digital Geospatial Metadata Workbook,* Ver. 2.0, Washington, D.C.: FGDC, 2000.

Federal Geographic Data Committee, "FGDC Don't Duck Metadata: Metadata Quick Guide," Ver. March 2005, http://www.fgdc.gov/metadata/education/MetadataQuickGuide.pdf (accessed September 15, 2005).

Federal Geographic Data Committee, *Geospatial Positioning Accuracy Standards—Part 1: Reporting Methodology,* FGDC-STD-007.1-1998, Reston, VA: FGDC, 1998.

Federal Geographic Data Committee, Metadata Ad Hoc Working Group, *Content Standard for Digital Geospatial Metadata,* FGDC-STD-001-1998, Washington, D.C.: FGDC, 1998.

Galison, P., *Einstein's Clocks, Poincaré's Maps: Empires of Time*, New York: W. W. Norton & Company, 2003.

GIS Portal.com, "GIS Data and Software Library," http://www.gisportal.com/gis3f.htm (accessed October 5, 2005).

GIS.com, "What Is GIS?" http://www.gis.com/whatisgis/whatisgis.html (accessed November 15, 2004).

GISLounge.com, "Data and Resources Links," http://gislounge.com/ll/data.shtml (accessed October 1, 2005).

GlobalSecurity.org, "Appendix E: Ellipsoids and Datums," http://www.globalsecurity.org/military/library/policy/army/fm/6-2/Appe.htm#tabe_23 (accessed June 27, 2005).

Goodchild, M. F., *The Fundamental Laws of GIScience*, University of California Santa Barbara, http://www.csiss.org/aboutus/presentations/files/goodchild_ucgis_jun03.pdf (accessed April 5, 2005).

Goodchild, M. F., and D. M. Mark, "The Fractal Nature of Geographic Phenomena," *Ann. Assoc. Am. Geographers*, Vol. 77, No. 2, 1987, pp. 265–278.

Greene, R. W., *Confronting Catastrophe: A GIS Handbook*, Redlands, CA: ESRI Press, 2002.

Greene, R. W., *GIS in Public Policy: Using Geographic Information for More Effective Government*, Redlands, CA: ESRI Press, 2000.

Grise, S., and B. Brodaric, *ArcGIS Geology Data Model*, Redlands, CA: ESRI Press, 2004.

Haagmans, R., et al., "A Procedure for Combining Gravimetric Geoid Models and Independent Geoid Data, with an Example in the North Sea Region," *DEOS Prog. Lett.*, Vol. 98, No. 1, pp. 89–99.

Hammond, A. D., "An Application of GIS in Underground Mining," *Earth Observation Magazine,* Vol. 10, No. 9, September 2001, pp. 13–16.

Hart, D., and H. Phillips, "Metadata Primer: A 'How To' Guide on Metadata Implementation," *National States Geographic Information Council,* http://www.lic.wisc.edu/metadata/metaprim.htm (accessed August 12, 2005).

Heiskanen, W. A., and F. A. Vening Meinesz, *The Earth and Its Gravity Field,* New York: McGraw Hill, 1958.

Hightower, A. W., and R. E. Klein, "Building a Geographic Information System (GIS) Public Health Infrastructure for Research and Control of Tropical Diseases," *Emerging Infectious Dis.*, Vol. 1, No. 4, October–December 1995, pp. 156–157.

Hirsch, S., "19thCent_R.RTF (Nineteenth Century Map Projections in RTF)," *Map Projections and Coordinate Systems,* Department of Geography and Earth Systems Science 661, George Mason University, Fall Semester 1999, http://classweb.gmu.edu/shirsch/Map Projections/ (accessed September 27, 2004).

Huff, D. L., "Parameter Estimation in the Huff Model," *ArcUser,* October–December 2003, pp. 34–36.

Hunter Department of Geology & Geography, "Map Projection Distortion," http://www.geography.hunter.cuny.edu/mp/distort.html (accessed August 27, 2004).

Inter-Governmental Committee on Surveying and Mapping, *Standards and Practices for Control Surveys (SP1)*, ICSM Publication No. 1, Ver. 1.5, Australia: Geocentric Datum of Australia, 2002.

Intergraph Mapping and Geospatial Solutions, "Case Study," http://imgs.intergraph.com/ (accessed February 7, 2005).

International Organization for Standardization, Technical Committee 46/Subcommittee 4, *Information and Documentation—The Dublin Core Metadata Element Set,* ISO 15836:2003(E), Geneva, Switzerland: International Organization for Standardization, 2003.

Ireson, W. G., et al., *Handbook of Reliability Engineering and Management*, 2nd ed., New York: McGraw-Hill, 1996.

Jafrullah, M., et al., "An Integrated Approach for Banking GIS," GISdevelopment.net, http://www.gisdevelopment.net/application/business/mi03140pf.htm (accessed June 3, 2005).

Jo, M.-H., et al., "The Management System Development of Campus Facility Information Using Web-Based GIS," *ACRS 2000—Proc. 21st Asian Conf. Remote Sensing*, Taipei, Taiwan, December 4–8, 2000.

Johnston, K., *Using ArcGIS Spatial Analyst*, Redlands, CA: ESRI Press, 2002.

Jones, A., and G. Blick, *Where in the World Are We?*, Ver. 1.0, Wellington, New Zealand: Government of South Australia, 1999.

Kubal, M. T., et al., *Building Profits in the Construction Industry*, New York: McGraw-Hill, 2000.

Kulik, L., "Reasoning About Cardinal Directions Using Grids as Qualitative Geographic Coordinates," *COSIT'99, Lecture Notes in Computer Science,* D. M. M. Freksa, (ed.), Vol. 1661, 1999, pp. 205–220.

Lang, L., *Managing Natural Resources with GIS*, Redlands, CA: ESRI Press, 2003.

Laton, W. R., "Guide to Map Scale," California State University at Fullerton, Department of Geological Sciences, http://geology.fullerton.edu/Faculty/wlaton/Map-scale.htm (accessed June 30, 2005).

Letham, L., *GPS Made Easy*, 3rd ed., Seattle, WA: The Mountaineers, 2001.

Liebschultz, H. D., *The Oracle Cookbook: For Design, Administration, and Implementation*, New York: M&T Books, 1996.

Longley, P. A., et al., *Geographic Information Systems and Science*, New York: John Wiley & Sons, 2001.

MacEachren, A. M. *How Maps Work: Representation, Visualization, and Design*, 3rd ed., New York: Guilford Press, 1995.

Manning, J., and J. Steed, "Positional Accuracy," *Int. Symp. Spatial Data Infrastructure*, University of Melbourne, Australia, November 19–20, 2001.

Maryland–National Capital Park and Planning Commission, *Geographic Information System: Cost/Benefit Assessment Report M-NCPPC*, submitted to the Montgomery County Council's Management and Fiscal Policy Committee, April 12, 1999.

McCoy, J., *Geoprocessing in ArcGIS 9*, ESRI Software documentation, http://support.esri.com (accessed December 7, 2004).

McManus, B., "Thematic Mapping," GIS Resource Document 05-70, Geographic Information Analysis Core of the Population Research Institute, Pennsylvania State University, April 2005, http://www.pop.psu.edu/gia-core/pdfs/gis_rd_05-70.pdf (accessed November 20, 2005).

Mendelson, B., *Introduction to Topology*, 3rd ed., New York: Dover Publications, 1990.

Merrigan, M. J., et al., "A Refinement to the World Geodetic System 1984 Reference Frame," *Proc. ION-GPS-2002*, Institute of Navigation, Portland, OR, September 24–27, 2002, pp. 1519–1529.

Mitchell, R. L., "Web Services Put GIS on the Map," *Computerworld*, December 15, 2003.

Miugglier, C. J., "The French Republic," *Photogrammetric Eng. Remote Sensing*, January 2001, pp 33–35.

Molodensky, S. M., *Methods for the Study of the External Gravitational Field and Figure of the Earth*, New York: Ministry of Geology and Control of Mineral Resources of the USSR, Israel Program for Scientific Translations, 1962.

Molodensky, S. M., "On the Models of the Lower Mantle Viscosity Consistent with the Modern Data of Core-Mantle Boundary Flattening," *Stud. Geophys. Geod.*, Vol. 46, 2002, pp. 411–433.

Monmonier, M., *Air Apparent: How Meteorologists Learned to Map, Predict, and Dramatize Weather,* Chicago, IL: University of Chicago Press, 2000.

Monmonier, M., *How to Lie with Maps*, 2nd ed., Chicago, IL: University of Chicago Press, 1996.

Moore, L., *Transverse Mercator Projections and U.S. Geological Survey Digital Products*, Washington, DC: Mid-Continent Mapping Center, U.S. Geological Survey, 1997.

Morales, M., "GIS Software Components: Example Using OpenMap and MapObjects," *GISLounge,* http://gislounge.com/ucon/ucopenmap.shtml (accessed September 29, 2005).

Mueller, I. I., and J. D. Rockie, *Gravimetric and Celestial Geodesy*, New York: Frederick Ungar Publishing Co., 1966.

National Archives of Australia, *AGLS Metadata Element Set, Part 1: Reference Description,* Ver. 1.3, Commonwealth of Australia: National Archives of Australia, 2002.

National Geospatial Intelligence Agency, "Africa Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryAfricaTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, "Asia Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryAsiaTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, "Atlantic Ocean Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryAtlanticTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, "Australia Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryAustraliaTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, *Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems*, 3rd ed., NIMA Stock No. TR8350.2, Bethesda, MD: NGIA, January 3, 2000.

National Geospatial Intelligence Agency, "Europe Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryEuropeTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, "Indian Ocean Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryIndianTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, "North America Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryNorthAmericaTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, "Pacific Ocean Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountryPacificTable.html (accessed June 25, 2005).

National Geospatial Intelligence Agency, "South American Datums Table," http://earth-info.nga.mil/GandG/coordsys/onlinedatum/CountrySouthAmericanTable.html (accessed June 25, 2005).

National Information Standards Organization, *Understanding Metadata,* Bethesda, MD: NISO Press, 2004.

Natural Resources Canada, Geodetic Survey Division, "Geodesy," http://www.geod.nrcan.gc.ca/index_e/geodesy_e/whatis_e.html (accessed May 18, 2005).

Noobeed.com, "List of Predefined Datums," http://www.noobeed.com/nb_ex_geod_datum_list.htm (accessed June 25, 2005).

Oberlin College, "Thematic Mapping: Some Basics," http://www.oberlin.edu/OCTET/HowTo/GIS/Tutorial/3_ThematicMapping/Thematic%20Maps.htm (accessed November 17, 2005).

Open GIS Consortium, "OpenGIS Implementation Specification: Coordinate Transformation Services," *Document OGC 01-009*, Rev. 1.00, Wayland, MA: Open GIS Consortium, 2001.

Open GIS Consortium, "OpenGIS Reference Model," Document OGC 03-040, Kurt Buehler, (ed.), Ver. 0.1.2, Wayland, MA: Open GIS Consortium, 2003.

Open GIS Consortium, *OpenGIS® Simple Features Specification for SQL*, Rev. 1.1, Project Document 99-049, Wayland, MA: Open GIS Consortium, May 5, 1999.

Open GIS Consortium, "Topic 11: OpenGIS™ Metadata (ISO/TC 211 DIS 19115)," *The OpenGISTM Abstract Specification,* Ver. 5, Wayland, MA: Open GIS Consortium, 2001.

Papathoma, M., et al., "Assessing Tsunami Vulnerability: An Example from Herakleio, Crete," *Nat. Hazards Earth Syst. Sci.*, Vol. 3, 2003, pp. 377–389.

Price, M., "Incident Command and GIS," *ArcUser*, Vol. 7, No. 4, October–December 2004, pp. 34–37.

Raina, S., and V. Amanna, "GIS for Rural Marketing," *Proc. 6th Ann. Int. Map India Conf. 2003*, New Delhi, India, January 28–31, 2003.

Ratcliffe, J. H., "Geocoding Crime and a First Estimate of a Minimum Acceptable Hit Rate," *Int. J. Geographical Information Sci.*, Vol. 18, No. 1, January–February 2004, pp. 61–72.

Richards, T. B., et al., "Geographic Information Systems and Public Health: Mapping the Future," *Public Health Reports*, Vol. 114, 1999, pp. 359–373.

Riordan, R. M., *Designing Relational Database Systems*, Redmond, WA: Microsoft Press, 1999.

Ritchie, D., *C Reference Manual*, Murray Hill, NJ: Bell Telephone Laboratories, http://cm.bell-labs.com/cm/cs/who/dmr/cman.pdf (accessed February 4, 2005).

Robinson, A. H., et al., *Elements of Cartography*, 6th ed., New York: Wiley & Sons, 1995.

Ryan, B. J., et al., "A Clear Vision of the NSDI," *Geospatial Solutions,* April 1, 2004.

Ryburn, R. J., "The Evolution of Geoscientific Metadata," *AGSO Research Newsletter,* Vol. 31, November 1999, pp. 12–13.

Schweitzer, P. N., "Metadata in Plain Language," *US Geologic Survey, Formal Metadata: Techniques and Software,* http://geology.usgs.gov/tools/metadata/tools/doc/ctc/ (accessed August 12, 2005).

Schweitzer, P. N., *National Spatial Data Clearinghouse: The Real Story,* U.S. Geological Survey Open-File Report 03-471, Digital Mapping Techniques 2003 Workshop Proceedings, 2003.

Silva, E., *Cost–Benefit Analysis for Geographic Information System: Implementation Justification*, submitted to Bruce Oswald, Chair of the NYS GIS Coordinating Body, March 4, 1998.

Slocum, T. A., et al., *Thematic Cartography and Geographic Visualization*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2003.

Smith, J. R., *The Meridian Arc Measurement in Peru 1735–1745*, International Institution for the History of Surveying and Measurement, http://www.ddl.org/figtree/hsm/index.htm (accessed August 14, 2005).

Snay, R., and T. Soler, "Modern Terrestrial Reference Systems Part 2: The Evolution of NAD83," *Professional Surveyor*, February 2000.

Snay, R., and T. Soler, "Modern Terrestrial Reference Systems Part 3: WGS84 to ITRS," *Professional Surveyor*, March 2000.

Snowdon, N., *Oracle Programming with Visual Basic*, San Francisco, CA: Sybex, 1999.

Snyder, J. P., *Map Projections: A Reference Manual,* London, U.K.: Taylor and Francis, 1995.

Snyder, J. P., *Map Projections: A Working Manual*, Washington, D.C.: U.S. Government Printing Office, 1987.

Snyder, J. P., and P. M. Voxland, "An Album of Map Projections," U.S. Geological Survey Professional Paper 1453, Washington, D.C.: U.S. Government Printing Office, 1989.

Sonoma State University, Geography and Global Studies, "Map Scale," http://www.sonoma.edu/geography/resources/mapscale.htm (accessed July 2, 2005).

Strahler, A. N., *Introduction to Physical Geography*, 3rd ed., New York: John Wiley & Sons, 1973.

Switala, K., "Crime Stoppers: Web-Based Mapping Makes Safer Neighborhoods," *Geoworld*, Vol. 18, No. 4, April 2005, pp. 36–38.

Tang, W., and J. Selwood, *Connecting Our World: GIS Web Services*, Redlands, CA: ESRI Press, 2003.

Tenzer, R., "Discussion of Mean Gravity Along the Plumbline," *Stud. Geophys. Geod.*, Vol. 48, 2004, pp. 309–330.

Tissot, A., *Mémoire sur la Représentation des Surfaces et les Projections des Cartes Géographiques,* Paris, France: Gauthier-Villars, 1881.

Tomlinson, R. F., *Thinking About GIS: Geographic Information System Planning for Managers*, Redlands, CA: ESRI Press, 2003.

Tracy, J., *Plane Surveying*, New York: John Wiley & Sons, 1906.

University Archives, University Libraries, and State University of New York at Buffalo, "Introduction to the Love Canal," *Love Canal Collection*, http://ublib.buffalo.edu/libraries/projects/lovecanal/introduction.html (accessed April 2, 2005).

U.S. Centers for Disease Control and Prevention, "John Snow," http://www.cdc.gov/ncidod/dbmd/snowinfo.htm (accessed February 7, 2005).

U.S. Environmental Protection Agency, *Indoor Air Pollution: An Introduction for Health Professionals*, Washington, D.C.: EPA, 2002.

U.S. Environmental Protection Agency, *Integrated Environmental Management Systems: Implementation Guide*, EPA 744-R-00-011, Washington, D.C.: EPA, October 2000.

Weiss, A., *Million Dollar Consulting: The Professional's Guide to Growing a Practice*, 2nd ed., New York: McGraw-Hill, 1992.

Wilkins, V., and L. Lundquist, "Innovations at Work: Rail," *MassTransit*, Vol. 31, No. 2, April 2005, pp. 40–44.

Xu, Y., and L. Liu, "GIS Based Analysis of Store Closure: A Case Study of an Office Depot Store in Cincinnati," *Geoinformatics 2004: Proc. 12th Int. Conf. Geoinformatics—Geospatial Information Research: Bridging the Pacific and Atlantic*, University of Gavle, Sweden, June 7–9, 2004.

Yiannakoulias, N., et al., "Zones of Prevention: The Geography of Fall Injuries in the Elderly," *Social Sci. Med.*, Vol. 57, 2003, pp. 2065–2073.

Yurkina, M. I., "To Pizzetti's Theory of the Level Ellipsoid," *Stud. Geophys. Geod.*, Vol. 46, 2002, pp. 435–454.

Zeiler, M., *Modeling Our World: The ESRI Guide to Geodatabase Design*, Redlands, CA: ESRI Press, 1999.

Zhao, L., "Integrating Rank Correlation Techniques with GIS for Marketing Analysis," *GeoComputation 2000, Proc. 5th Int. Conf. GeoComputation*, University of Greenwich's School of Earth and Environmental Sciences, Chatham Maritime, Kent, U.K., August 23–25, 2000.

Zilkoski, D. B., et al., *Guidelines for Establishing GPS-Derived Ellipsoid Heights (Standards: 2 cm and 5 cm),* NOAA Technical Memorandum NOS NG-58, Ver. 4.3, Silver Spring, MD: National Oceanic and Atmospheric Administration, 1987.

Zurayk, R., "Participatory GIS-Based Natural Resource Management: Experiences from a Country of the South," *Arid Lands Newsletter*, No. 53, May–June 2003.

About the Author

Stephen R. Galati is the Northeast proposal manager for TRC Environmental Corporation. With more than 13 years of experience in technical writing, engineering, and proposal management, Mr. Galati's work centers upon environmental, marketing, and GIS applications throughout the United States. He has written numerous articles for industry periodicals.

Mr. Galati holds a B.S. in electrical engineering from Pratt Institute and an M.A. in English from the College of Staten Island–City University of New York. He is a member of the American Society of Photogrammetry and Remote Sensing, the American Geophysical Union, and the New York City Chapter of Geographic Information Systems and Mapping Operations (GISMO), as well as a GeoData Alliance Affiliate.

Index

Accuracy
  absolute, 233
  relative, 244
  scale and, 67–69
Active pages, 233
Alberta elderly fall injuries case study, 193–94
ArcGIS, 202, 203
ASCII, 45, 46
Asian longhorned beetle (ALB), 17–18
Aspect, projection, 125–26
  changing, 137
  defined, 125, 233
  normal, 125–26, 242
  oblique, 126, 127, 242
  polar, 126, 127, 243
  transverse, 126, 247
Attributes
  data types, 38–39
  defined, 233
  tables, 38–39
Australian Geodetic Datums of 1966 and 1984 (AGD66 and AGD84), 101
Australian Government Locator Service (AGLS) Metadata Element Set, 54
Autocorrelation, 11
Autodesk Map, 203, 204–5
Axis
  defined, 233
  major, 241
  minor, 241
  of revolution, 233
  of rotation, 77, 233
  semimajor, 78, 245
  semiminor, 245
  standard, 114
Azimuth, 233
Azimuthal projections, 128–29
  defined, 128, 234
  illustrated, 130
  See also Map projection(s)

Bacchus, 28
Bar scale, 66, 234
Bi-products, 18–19, 234
BLOB data type, 39
Bonne projection, 121
Bursa-Wolf data transformation method, 95, 96–98
  defined, 96
  formula, 97
  illustrated, 98
  parameters, 96–97
  rotational polarity, 97
  as seven-parameter transformation method, 96
  See also Horizontal datums

Canadian Spatial Reference System NTv2 software, 101
Cartesian coordinate system, 142–45, 234
  defined, 142

Cartesian coordinate system (continued)
false easting/northing, 145
three-dimensional, 144
two-dimensional, 143
See also Coordinate systems
Cartographers, 113
Cartographic elements, 168–71
data frame, 168
illustrated, 170
legend, 170
map compass, 170
map information frame, 168
map scale, 170
overview map, 170–71
title block, 168–70
Cartography, 234
Cassini, Jean-Dominique, 234
Centroid, 234
Chew and spit (CNS) metadata, 57
Choropleth maps, 160–61
data classes, 160, 161
defined, 160
illustrated, 160
NADP, 165
See also Thematic mapping
Circle, 234
Clearinghouses
geospatial data, 210–11
geospatial metadata, 57–58
Client/server systems, 21, 234
Codd, E. F., 20, 234
Color differentiation, 137
Computer-aided design (CAD), xxiii, 235
Conformal projections, 130–31, 235
Conic projection, 122–24, 235
Content Standards for Digital Geospatial Metadata (CSDGM), 50–51
Continuous data, 7, 30, 235
Conversion, 10
Coordinate frame transformation, 97, 98
Coordinate systems, 139–51
Cartesian, 142–45
defined, 139–40
geographic, 140–41, 238
implementation, 150–51
State Plane Coordinate System (SPCS), 148–50, 246
terrestrial, 139
Universal Polar Stereographic (UPS), 147–48
Universal Transverse Mercator (UTM), 145–47, 248
Coosa/Warrior shoreline management case study, 188
Counts, 165
Curvilinear surface, 115, 235
Cylindric projection, 122, 123, 235

Data
consistency, 25
continuous, 7, 30, 235
discrete, 6, 30, 236
generalization, 69–71
raster, 7, 29–31, 244
structure, 29–39
vector, 7, 29–31, 248
Database management systems (DBMSs), 19–25, 236
as data management tool, 19–20
defined, 19
relational (RDBMSs), 20–25
Data frame, 168
Data management software, 10
Date data type, 39
Datums. *See* Geodetic datums
Demographics, 19
Densities, 166
Density maps, 18
Descartes, René, 142
Developable projection surfaces, 236
Dictionary of GIS Terminology, xx
Digital elevation model (DEM), 236
Digital terrain model (DTM), 236
Disaster planning, response, recovery, 194–95
Discrete data, 6, 30, 236
Disease management, 193
Distortion, 132–33
defined, 236
magnitude, 132, 236
management, 136–37
overall distribution, 133
pattern, 236
small-scale, 133
type, 132
See also Map projection(s)
Dot density maps, 162–63
defined, 162

illustrated, 163
problems, 163
See also Thematic mapping
Double data type, 38–39
Dublin Core Metadata Element Set, 51–53
15 primary elements, 51–52
defined, 51
encoding syntax, 53
See also Geospatial metadata
Dynamic pages, 236

Earth
curvilinear surface, 115, 235
flattening, 76–77
geographic coordinate system, 115
latitude, 141, 240
longitude, 141, 241
mathematical model, 77–81
meridians, 115
parallels, 115, 140, 242
physical, illustrated, 114
physicality of, 113–16
scaled-down, 116
shape, 75–77
standard axis, 114
symmetry, 78
three-dimensional oblate, 78
typical shape calculation, 81–82
Earth Gravitational Model of 1996 (EGM96), 107–8
Earth models
defined, 236
history, 74–75
illustrated, 74
predominant, 74
Earth-Observing Satellites (SPOT), 236, 244–45
Eccentricity, 80–81
constant, 80
defined, 236
in Earth shape calculation, 82
first, 80, 81
second, 80, 81, 82, 245
Education applications, 197–98
Ellipse, 237
Ellipsoid, 73–83
defined, 75, 237
Earth, 77
flattening, 76–77
geoid separation, 106–7
to globe, 116
GRS80, 91
height, 106
oblate, 242
precision-enhancing references, 85
prolate, 243
reference, 82–83, 244
reference parameters, 219–21
of revolution, 113, 118, 237
WGS84, 89–90
Engineering applications, 175–77
case study, 176–77
examples, 176
GIS use, 175–76
Enterprise system, 237
Environmental investigation/remediation applications, 185–87
GIS uses, 185–86
Illinois UST monitoring/inspection case study, 186
USEPA's disaster dioxin air-monitoring case study, 186–87
Environment applications
additional, 188–89
GIS for, 181–89
investigation and remediation, 185–87
natural resource management, 182–85
shoreline management systems, 187–88
Equal-area projections, 129–30, 237
Equal interval method, 166
Equator, 237
Equidistant projections, 128, 129
defined, 128, 237
illustrated, 129
See also Map projection(s)
Eratosthenes, 62, 63, 237
European CEN Metadata Standard, 53–54
European Datum 1950 (ED50), 93–94
Executive Orders 12906/13286, 55–56
Explore, GIS compass, 216
Extensible Markup Language (XML), 45, 46, 47
for metadata, 45
relationship, 47
Extraction, 10

False easting, 145
False northing, 145
Feature files, 6, 237
Feature for identification (FID), 37

Federal Geographic Data Committee (FGDC), 50, 55
File-based systems, 21, 238
Flattening
  defined, 238
  Earth, 76–77
  in Earth shape calculations, 81
  inverse, 79, 82, 240
  levels, 76
Float data type, 38
Focal radius, 238
For-purchase GIS environments, 202–5
FreeGIS.org, 207

Gaston public policy education case study, 200
Generalization, 69–71
  defined, 69
  "house on the wrong side of the road" problem, 70, 71
Geocode, 238
Geodatabases, 5
Geodesy
  advancement, 64
  defined, 61, 238
  early form, 62
  present-day, 64
Geodetic datums, 86–87
  defined, 85, 86
  globally centered, 86
  illustrated, 86
  local, 87
  primary types, 87–89
  worldwide, 224–31
  See also Horizontal datums; Vertical datums
Geodetic Reference System of 1980 (GRS80), 91–92
  defined, 91
  ellipsoid, 91
  geocentric model, 92
  See also Horizontal datums
Geographic coordinate system, 140–41
  defined, 238
  latitude lines, 140
  longitude lines, 140
  See also Coordinate systems
Geographic data
  defined, 238
  generalization, 69–71
  model, 5
  structure, 29–39
  See also Data
Geographic information, 12–13
  defined, 238
  as end product, 13
  fundamental flow, 3–5
  geography and, 5–8
  path, 4
  uses, 12–13
Geographic information systems. *See* GISs
Geography
  defined, 5–6, 238
  geographic data and, 5–8
GEOID03, 108
Geoids
  defined, 105, 239
  ellipsoid separation, 106–7
  surface, 105
GeoMedia, 203, 204
Geomensuration, 64, 239
Geometric, 239
Geoprocessing, 9–11
  categories, 10–11
  defined, 9, 239
  site selection, 10
Georeferencing, 8, 9, 239
Geospatial analysis, 239
Geospatial data clearninghouses, 210–11
  data use, 211
  defined, 210
  list of, 211
  use, 210
Geospatial metadata, 41–58
  Australian Government Locator Service (AGLS) Metadata Element Set, 54
  clearinghouses, 57–58
  CSDGM, 50–51
  defined, 43
  Dublin Core Metadata Element Set, 51–53
  elements, 43
  European CEN Metadata Standard, 53–54
  GIS and, 43–44
  publishing resources, 57–58
  schema, 44, 45
  standards, 48–55
  standard types, 49
  UK GEMINI Discovery Metadata Standard, 54–55
  uses, 44

GIScience, 174–75
  case study, 174–75
  defined, 174
  uses, 174
GIS compass, 214–18
  all directions in between, 218
  defined, 214
  definitions, 214–15
  explore, 216
  illustrated, 214
  niche, 215–16
  standards, 217–18
  wherever, 216–17
GIS component software, 208–9
  as building block, 208
  types, 209
  variations, 208
GIS Design Server, 202, 205
GIS Lounge, 207
GISs
  attribute tables, 38–39
  coordinate system implementation, 150–51
  data consistency, 25
  data transformation, 4
  defined, xix, xx–xxi, 238
  demystification goal, xxvi
  discerning need for, 16–17
  engineering applications, 175–77
  for environment applications, 181–89
  functioning, 3–13
  functions and analysis, 5
  growth, xx
  indices, 38–39
  leading for-purchase environments, 202–5
  marketing and financial applications, 177–79
  move towards, xxi–xxii
  mystique, xxiii–xxiv
  products/bi-products, 17–19
  public and, 191–200
  reasons for using, 15–28
  as software applications, 5
  for technical applications, 173–74
  technical revolution growth, 213–14
  tiers, 23
  up and running with, 201–2
  visibility, 26–27
GIS Web services, 25
Global positioning system (GPS) technology, xxii, 239
Globe
  defined, 239
  ellipsoid to, 116
  transformation from, 116–17
Graduated color maps. *See* Choropleth maps
Graduated symbol maps, 161–62
  defined, 161
  illustrated, 162
  pie chart, 162
  proportional circle, 161–62
  See also Thematic mapping
Grapefruit peel experiment, 119–20
Graphical visualizations, 18
Graphic scale, 66, 239
GRASS, 202, 205–7
  defined, 206
  GIS environment, 206
Graticule, 239
G/Technology, 202, 203–4
GUID data type, 39

Health and safety management, 192–94
  case study, 193–94
  disease management, 193
  process, 193
  See also Public
Hemispheres, 240
High Accuracy Reference Network (HARN), 100
Homeland security, 196–97
Horizontal datums, 85–101
  Bursa-Wolf datum transformation method, 95, 96–98
  defined, 89, 240
  European Datum 1950 (ED50), 93–94
  GRS80, 91–92
  horizontal grid transformation methods, 99–101
  International Terrestrial Reference Frame (ITRF), 90–91
  Molodensky datum transformation method, 95–96
  North American Datum of 1927 (NAD27), 92–93
  North American Datum of 1983 (NAD83), 92–93
  standard, 89–94
  transformation models, 94–98

Horizontal datums (continued)
WGS84, 89–90, 223–31
See also Geodetic datums
Horizontal grid transformation
Australian Geodetic Datums of 1966 and 1984 (AGD66 and AGD84), 101
Canadian Spatial Reference System NTv2 software, 101
defined, 99
High Accuracy Reference Network (HARN), 100
methods, 99–101
North American Datum Conversion (NADCON), 99–100
"House on the wrong side of the road" problem, 70, 71
HTML, 45, 46, 47
relationship, 47
sample metadata framework, 46
Hypertext Markup Language (HTML), 24

IDRISI Kilimanjaro, 202, 204
Illinois UST monitoring/inspection case study, 186
Index files, 240
Indicatrix, 240
Indices, 38–39
International Terrestrial Reference System (ITRS), 90–91
Intersect, 10
Inverse flattening, 79, 82
defined, 240
formula, 79
parameters, 83
See also Flattening
Isarithms, 137, 240
Isopleth maps, 163–65
defined, 163–64
illustrated, 164
NADP, 165
uses, 164, 165
See also Thematic mapping

Lebanon natural resource management case study, 183–84
Legends, 167–68, 169, 240
Light intensity detection and ranging (LIDAR), xxiii, 241
Linked attribute table, 241
Local geodetic datums, 87
Long Integer, 38
Lunar laser ranging (LLR), 90

Management, 10
Map compass, 170
MapInfo, 202, 205
Map information frame, 168
Map projection(s), 113–37
application purpose and, 120
aspect, 125–26
aspect, changing, 137
azimuthal, 128–29, 130
classification, 126–32
classification types, 127–28
color differentiation, 137
conformal, 130–31, 235
conic, 122–24, 235
cylindric, 122, 123, 235
default, 118
defined, 117, 241
distortion, 132–33
distortion management, 136–37
equal-area, 129–30, 237
equidistant, 128, 129, 237
familiar shapes, 137
geometric, 116–17
isarithms, 137
planar, 124, 242–43
projection formula, 118
properties, 121
redistribution of scale factor, 136
selection, 120–22
types, 122–24
visual-logical analysis, 137
Map scales, 65–67, 170
accuracy and, 67–69
applications, 67
bar scale, 66
defined, 241
graphic scale, 66
key, 65
proper, 69
ratio scale, 66
simple calculations, 67
transformation criteria, 68
understanding, 65–66
Map space, 241
Marketing and financial applications, 177–79
case study, 178–79
GIS use, 177–78

Maryland-National Capital Park and Planning Commission (M-NCPPC), 27–28
Mathematical model, Earth, 77–81
  axis of rotation, 77
  illustrated, 78
  two-dimensional, six-parameter, 79
Mean sea level (MSL), 104
  elevation, 104
  negative, 104
  positive, 104
Mercator model, 119, 120
Meridians, 115, 140, 141
  defined, 241
  prime, 243
Metadata
  chew and spit (CNS), 57
  defined, 42–43, 241
  embedded, 43
  geospatial, 41–58
  syntactic frameworks, 46
  Unqualified Dublin Core, 53
  uses, 42
Military Grid Reference System (MGRS), 146
Minor axis, 241
MOD IV, 177
Molodensky datum transformation method, 95–96
  defined, 95
  formulas, 95
  three-parameter transformation, 96
  See also Horizontal datums
Multitier systems, 22

Namibia natural resource management case study, 184–85
National Geodetic Vertical Datum of 1929 (NGVD29), 108–9
  defined, 108
  obstacle, 109
  See also Vertical datums
National Spatial Data Infrastructure (NSDI), 55
  defined, 56
  implementation, 56
Natural breaks method, 166–67
Natural resource mapping, 182–85
  GIS use, 182–83
  Lebanon case study, 183–84
  Namibia case study, 184–85
Negative spatial autocorrelation, 11
Newton, Sir Isaac, 75, 242
Niche, GIS compass, 215–16
Normal aspect, 125–26, 242
North American Datum Conversion (NADCON), 99–100
North American Datum of 1927 (NAD27), 92–93
North American Datum of 1983 (NAD83), 92–93
North American Vertical Datum of 1988 (NAVD88), 109–10

Oblate ellipsoid, 242
Oblique aspect, 126, 127, 242
Open Geospatial Consortium, 207
Organization, this book, xxiv–xxv
Orthometric height, 104
Orthomorphic projections, 130, 242
Orthophotography, 242
Overall distribution, 133
Overlay, 10
Overview maps, 170–71

Parallels, 115, 140, 242
Photogrammetry, 242
Pie chart maps, 162
Planar projection, 124, 242–43
Planar space, 243
Planar surface, 243
Plane coordinates, 243
Plane of the ecliptic, 114, 243
Poincaré, Jules-Henri, xxi–xxii
Polar aspect, 126, 127, 243
Poles, 243
Portland TriMet case study, 196
Position vector transformation, 97
Positive spatial autocorrelation, 11
Primary key, 20, 243
Prolate ellipsoid, 243
Proportional circle maps, 161–62
Proximity, 10
Public
  disaster planning, response, recovery, 194–95
  education applications, 197–98
  GIS and, 191–200
  health and safety, 192–94
  homeland security, 196–97

Public (continued)
information need, 191–92
policy, redefining, 199–200
transportation and information, 195–96
Publishing resources, geospatial metadata, 57–58

Quadrangle, 243
Quality of Natural Resources Commission (QNRC), 200
Quantile method, 166
Queries, 8–9

Raster data, 7
advantages, 31
continuous, 30
defined, 244
feature geometries, 32
structure, 29–31
vector data versus, 30
See also Data
Raster images
physically continuous nature, 34
structures, 34–35
Rates, 165–66
Ratio scale, 66, 244
Reference ellipsoid parameters, 219–21
Reference ellipsoids, 82–83, 244
Region of interest, 244
Relational DBMSs (RDBMSs), 20–25
attributes, 20
client/server, 21
defined, 20, 244
file-based, 21
multitier, 22
primary key, 20, 243
tabular data relationship, 21
types, 22
Web relationship, 24
Relief maps, 244

Satellite laser ranging (SLR), 90
Scale factor
compensatory, 235
redistribution, 136
Scale reduction, 116, 245
Schemas
defined, 44, 45
elements, 45
See also Metadata
Semimajor axis, 78, 245
Semiminor axis, 245
Semi-volatile organic compounds (SVOCs), 185
Seven-parameter transformation method. *See* Bursa-Wolf data transformation method
Shoreline Management Plan (SMP), 188
Shoreline management systems, 187–88
case study, 188
GIS use, 187–88
Short Integer, 38
Sidereal day, 113, 245
Sidereal year, 114, 245
Site selection, 10
Six Sigma, 25–26, 245
Smallworld, 202, 204
Snow, John, 15–16, 245
Software environments, 12
Spatial analysis, 11–12
characteristics, 11
defined, 245
Spheres, 245
Spheroid. *See* Ellipsoid
Standard deviation classification, 166
Standard Generalized Markup Language (SGML), 45, 47
comprehensive nature, 47
relationship, 47
Standard line, 246
Standards, GIS compass, 217–18
State Plane Coordinate System (SPCS), 148–50
defined, 148, 246
illustrated, 150
map projections, 149–50
northings/eastings method, 150
original, 149
See also Coordinate systems
Summer Solstice, 246
Surveying, 64, 246
Sydney Catchment Authority (SCA), 18
Symbols, 167–68

Tagged-image file format (TIFF), 246
Tertile maps, 166
Text data type, 39
Thematic layers
defined, 157
illustrated, 158

manipulating, 157–59
surfaces, 159
user discipline, 159
uses, 158–59
Thematic mapping, 155–71
cartographic elements, 168–71
choropleth, 160–61
color, 167–68
counts, 165
data attribute definition, 156
data classification, 166–67
data table versus, 157
defined, 155, 246
densities, 166
dot density, 162–63
equal interval method, 166
graduated symbol, 161–62
isopleth, 163–65
layers, manipulating, 157–59
legends, 167–68, 169
natural breaks method, 166–67
need, 156–57
projection, 167
quantile method, 166
rates, 165–66
scale, 167
standard deviation classification, 166
symbols, 167–68
types, 159–65
Thick clients, 22, 246
Thin clients, 246
Three-dimensional Cartesian coordinate system, 144
Three-dimensional mapping, 247
Three-dimensional similarity transformation, 97
Three-dimensional space, 247
Three-tier system, 247
Tissot, Nicolas Auguste, 247
Tissot indicatrix, 133–35
defined, 133, 247
illustrated, 134
projected space, 134
success, 133
transverse Mercator projection and, 135
Title block, 168–70
Tobler's First Law of Geography, 247
Topography, 247
Topology, 35–38
defined, 247
elements, 38
feature, 37
relationships, 36
of two polygons, 37
Transformation, 11, 247
Transportation and information, 195–96
GIS use, 195–96
Portland case study, 196
See also Public
Transverse aspect, 126, 247
Triangulated irregular network (TIN), 247–48
True scale, 248
Tuples, 248
Two-dimensional Cartesian coordinate system, 143
Two-dimensional space, 248

UK GEMINI Discovery Metadata Standard, 54–55
Underground storage tanks (USTs), 186
Union, 10
Universal Polar Stereographic (UPS), 147–48
Universal Transverse Mercator (UTM), 145–47
defined, 145, 248
summary, 147
UPS global coverage, 147
use, 146–47
zoned grid, 145
zones illustration, 146
See also Coordinate systems
USEPA's disaster dioxin air-monitoring case study, 186–87

Vector data, 7
advantages, 31
defined, 248
discrete, 30
feature geometries, 32
raster data versus, 30
structure, 29–31
See also Data
Vector features
construction, 33
geometry, 31–34
Vector space, 248
Venn diagrams, 36
Verbal scale, 248

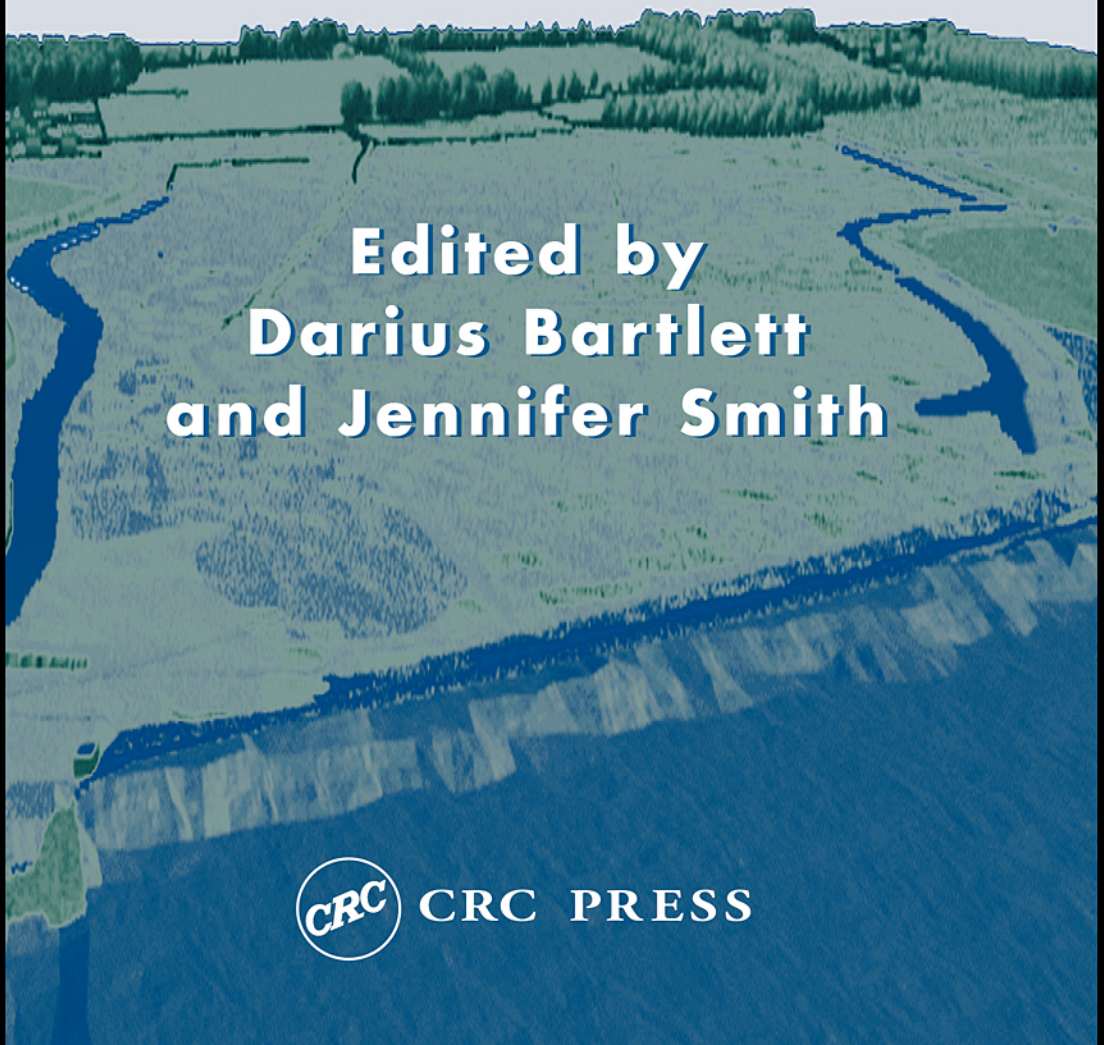
VERTCON, 110–11
  calculation, 110–11
  defined, 110
  software, 111
Vertical datums, 103–11
  defined, 89, 103, 248
  Earth Gravitational Model of 1996 (EGM96), 107–8
  GEOID03, 108
  National Geodetic Vertical Datum of 1929 (NGVD29), 108–9
  North American Vertical Datum of 1988 (NAVD88), 109–10
  reference, 103
  standard, 107–10
  transformation methods, 110–11
  VERTCON, 110–11
  See also Geodetic datums
Very long baseline interferometry (VLBI), 90
Visual-logical analysis, 137
Volatile organic compounds (VOCs), 185

WGS60, 89
WGS84, 89–90, 223–31
  base parameters, 223
  as common model, 94
  constants, 81–82
  defined, 89
  Earth-centered, Earth-fixed coordinate system, 223
  standard reference ellipsoid, 89–90
  worldwide geodetic datums, 224–31
  See also Horizontal datums
Wherever, GIS compass, 216–17
Winter Solstice, 248
Women, Infants and Children (WIC) program, 199

XMap, 202, 205

Zenithal projection. *See* Planar projection

GIS *for* Coastal Zone Management



Edited by
Darius Bartlett
and Jennifer Smith

CRC PRESS

GIS
for
Coastal Zone Management

GIS
for
Coastal Zone Management

Edited by
Darius Bartlett and
Jennifer Smith

CRC PRESS

Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

CoastGIS '01 Conference (2001 : Halifax, N.S.)
GIS for coastal zone management / edited by Darius J. Bartlett and Jennifer L. Smith.
p. cm.
Includes updated and edited presentations made to the CoastGIS '01 Conference in Halifax, Canada 18-20 June 2001.
Includes bibliographical references and index.
ISBN 0-415-31972-2
1. Coast changes—Congresses. 2. Environmental mapping—Congresses. 3. Coastal zone management—Congresses. 4. Geographic information systems—Congresses. I. Bartlett, Darius J., 1955- II. Smith, Jennifer L. III. Title.

GB450.2.C625 2001
333.91'7'0285-dc22 2004050302

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

All rights reserved. Authorization to photocopy items for internal or personal use, or the personal or internal use of specific clients, may be granted by CRC Press LLC, provided that $1.50 per page photocopied is paid directly to Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 USA. The fee code for users of the Transactional Reporting Service is ISBN 0-415-31972-2/04/$0.00+$1.50. The fee is subject to change without notice. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press LLC for such copying.

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

Visit the CRC Press Web site at www.crcpress.com

© 2005 by CRC Press LLC

No claim to original U.S. Government works
International Standard Book Number 0-415-31972-2
Library of Congress Card Number 2004050302
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Contents

Foreword *ix*
Ron Furness and Andy Sherin

Preface *xiii*
Darius Bartlett and Jennifer Smith

Contributors *xv*

Chapter One *1*
Coastal Spatial Data Infrastructure
Roger A. Longhorn

Chapter Two *17*
Bridging the Land-Sea Divide Through Digital Technologies
Simon Gomm

Chapter Three *27*
A Comparative Study of Shoreline Mapping Techniques
Ron Li, Kaichang Di and Ruijin Ma

Chapter Four *35*
Monitoring Coastal Environments Using Remote Sensing and GIS
Paul S.Y. Pan

Chapter Five *51*
Spatial Uncertainty in Marine and Coastal GIS
Eleanor Bruce

Chapter Six *63*
New Directions for Coastal and Marine Monitoring: Web Mapping and Mobile Application Technologies
Sam Ng'ang'a Macharia

Chapter Seven *77*
Exploring the Optimum Spatial Resolution for Satellite Imagery: A Coastal Area Case Study
Chul-sue Hwang and Cha Yong Ku

Chapter Eight 95
Visualisation for Coastal Zone Management
Simon R. Jude, Andrew P. Jones and Julian E. Andrews

Chapter Nine 109
Application of a Decision Support System in the Development of a Hydrodynamic Model for a Coastal Area
Roberto Mayerle and Fernando Toro

Chapter Ten 125
Decision-Making in the Coastal Zone Using Hydrodynamic Modelling with a GIS Interface
Jacques Populus, Lionel Loubersac, Jean-François Le Roux, Frank Dumas, Valerie Cummins, and Gerry Sutton

Chapter Eleven 141
Towards an Institutional GIS for the Iroise Sea (France)
Françoise Gourmelon and Iwan Le Berre

Chapter Twelve 153
Culture Intermixing, the Diffusion of GIS and its Application to Coastal Management in Developing Countries
Darius Bartlett and R. Sudarshana

Chapter Thirteen 165
The Use of GIS to Enhance Communications of Cultural and Natural Resources and Contamination
John A. Lindsay, Thomas J. Simon, Aquilina D. Lestenkof and Phillip A. Zavadil

Chapter Fourteen 181
GIS Applications in Coastal Management: A View from the Developing World
Peter C. Nwilo

Chapter Fifteen 195
High-Resolution Elevation and Image Data Within the Bay of Fundy Coastal Zone, Nova Scotia, Canada
Tim Webster, Montfield Christian, Charles Sangster and Dennis Kingston

Chapter Sixteen 219
Mapping and Analysing Historical Shoreline Changes Using GIS
Courtney A. Schupp, E. Robert Thieler and James F. O'Connell

Chapter Seventeen *229*
GIS for Assessing Land-Based Activities that Pollute Coastal Environments
J.I. Euán-Avila, M.A. Liceaga-Correa and H. Rodríguez-Sánchez

Chapter Eighteen *239*
Applying the Geospatial Technologies to Estuary Environments
David R. Green and Stephen D. King

Chapter Nineteen *257*
A Territorial Information System (TIS) for the Management of the Seine Estuary – Environmental and Management Applications
Jean-Côme Bourcier

Chapter Twenty *269*
Developing an Environmental Oil Spill Sensitivity Atlas for the West Greenland Coastal Zone
Anders Mosbech, David Boertmann, Louise Grøndahl, Frants von Platen, Søren S. Nielsen, Niels Nielsen, Morten Rasch and Hans Kapel

Chapter Twenty-one *281*
Environment Canada's Atlantic Sensitivity Mapping Program
André Laflamme, Stéphane R. Leblanc and Roger J. Percy

Epilogue: Meeting the Needs of Integrated Coastal Zone Management *295*
Jennifer L. Smith and Darius J. Bartlett

Index *301*

Foreword

The material presented in this volume comprises updated and edited presentations first made to the CoastGIS'01 Conference conducted in Halifax, Canada between the 18th and 20th June 2001 together with chapters commissioned by the Editors.

The CoastGIS series of conferences have been the outcome of a fruitful collaboration between the International Geographical Union's Commission on Coastal Systems and the International Cartographic Association's Commission on Marine Cartography. Generally entitled "International Symposium on GIS and Computer Mapping for Coastal Zone Management" we have seen five successful CoastGIS conferences held over this decade-long collaboration.

These conferences were held in Ireland (Cork, 1995), Scotland (Aberdeen, 1997), France (Brest, 1999) Canada (Halifax, 2001) and Italy (Genoa, 2003). A closely allied CoastalGIS conference was conducted in Wollongong, Australia in July 2003. Future conferences are planned for Scotland, Australia and Barbados.

At the first meeting in Cork, we had the honour of being addressed in a keynote presentation by Lord Chorley, who referred in his address to the House of Commons Environment Select Committee's 1992 Report on coastal zone protection and planning. Reflecting on the findings of that Report, Lord Chorley was then struck by three main points.

"First, it is only in recent years that the coastal zone has been recognised as one important topic in its own right. Second, the huge range of relevant aspects or considerations. (Thirdly): the huge number of agencies involved, often with overlapping and perhaps incompatible responsibilities, jurisdictions and objectives."

These themes have recurred throughout the conferences that followed.

Halifax 2001

CoastGIS'01 was convened in Halifax, Canada, at Saint Mary's University between 18th – 20th June 2001. The conference attracted over 150 delegates from the Americas, Europe, Africa, Asia and Australia who presented 36 oral presentations and live demonstrations in a single stream format and 50 posters. The theme selected by the 2001 Science Committee was "Managing the Interfaces," a theme with a multitude of possible interpretations. Overall, the shift in emphasis towards *integration* of systems for coastal management and the

growing interest in coastal spatial data infrastructures were especially in evidence at this meeting. So too was the international dimension of coastal GIS.

Notably, for the first time, financial assistance from the Canadian International Development Agency and the Geomatics Association of Nova Scotia permitted CoastGIS 2001 to fund ten delegates from the developing world to participate in the conference. Two chapters in the book result from this initiative (those by Nwilo and Euán-Avila *et al*.).

CoastGIS 2001 also instituted demonstrations of live GIS systems. Three chapters in this volume arose from this innovation (Laflamme *et al.,* Mosbech *et al.* and Bourcier). There were field trips to Nova Scotia Community College's Centre for Geographical Sciences in the Annapolis Valley and to the Bay of Fundy, where we considered the coastal issues facing a region that experiences the highest tides in the world.

This setting drove home the dynamic nature of the coastal zone interface of land, sea and air. Within this framework, several of the conference presentations that evolved into chapters in this volume deal with the dynamics of the coastal zone, while others address approaches to bridging the land-sea divide.

Many presentations at the Halifax gathering focused on the need for an effective interface amongst the range of participants and stakeholders involved in coastal management. The chapters in this volume that describe applications and case studies and those that include traditional ecological knowledge demonstrate the impact of effective communication between these parties.

The use of increasingly advanced technologies in the coastal zone (notably remote sensing, web mapping and mobile application technologies, visualization techniques, and LiDAR) to support research and management was a highlight of the conference and is detailed in several chapters.

The reader will no doubt be well aware of the amazing developments of GIS capability over the last decade in particular. Nonetheless, the development of standards, formats and data models together with the sheer genius of GIS technological developments and ultimate cost effectiveness are perhaps still hampered by the paucity of available data sets. At CoastGIS 2001, the development of spatial data infrastructures was highlighted. Susan Lambert, then the Executive Director of the Kentucky Office of Geographic Information and now with the United States Geological Survey, presented a keynote lecture on the development of the GeoData Alliance, a nonprofit organization open to all individuals and institutions committed to using geographic information to improve the health of our communities, our economies and the Earth. A presentation was also made on Canada's Marine Geospatial Data Infrastructure. The Editors of this volume invited Roger Longhorn to summarize the progress in the development of coastal spatial data infrastructures for this volume.

Genoa 2003 and beyond

One advantage the authors have in writing this Foreword is being able to do it immediately after the successful conclusion of this successful conference, which ran from October 16 to 18, 2003. Exploration of the SDI theme continued at the 2003 CoastGIS conference held in Genoa, Italy. At this gathering emphasis was placed upon many of the non-science and non-technology issues that continue to adversely impact the success and long-term sustainability of many coastal zone projects and wider coastal zone monitoring initiatives at national and regional levels due to barriers to the access to data and information.

In summarizing the outcomes of the Genoa conference on behalf of the Scientific Committee, Roger Longhorn noted that virtually all presentations had covered coastal zone research, monitoring or management work in a *single* nation, often in a single sub-national region. Very few therefore faced the added difficulties that can arise when trying to locate, access, understand and agree on the usage and dissemination terms for data from owners outside not only one's own discipline, but outside one's national legislative infrastructure for information use. The ocean, as Longhorn pointed out, has a "nasty" way of connecting one piece of coastline to another, and neither the ocean nor the physical coastline show any respect for national boundaries and differing jurisdictions.

Some of the key points highlighted by different members of the Scientific Committee at a meeting held on the last day of the conference included the following:

1. We need political champions to help guard our interests in seeing that coastal information needs are not forgotten as larger national and regional (trans-national) spatial data infrastructure (SDI) frameworks are created.

2. GIS in the coastal zone is certainly about supporting "science work", but there are also non-scientific and non-technical issues to be considered, hence the need for a policy level of collaboration.

3. Data usability is a key concern and continues to require both research and information management focus for continued development of ways to harmonise data for wider use.

4. Information infrastructure developments are needed that permit easier discovery of existing data and use of data once located, in a variety of forms, from multiple data owners or custodians.

5. We need to find ways to engage stakeholders (data creators, custodians and users) even more widely in the data management and access issues.

With the launch of the EU Water Framework Directive in 2000, to be fully implemented by December 2003 in all EU Member States, we have seen the first institutionalized, regional (trans-national) legal requirement that GIS be used in

monitoring the implementation of a major EU policy, and one of extremely high importance on a global level - i.e., maintaining good quality water resources in river basins, groundwater, coastal zones and the off-shore transitional waters leading to the coastal zone. In all likelihood this is only the first such legal requirement that we will see coming from major international institutions for use of spatial information and GIS tools for planning and monitoring purposes.

Therefore, coastal GIS practitioners need to address their next efforts towards effective usability of coastal knowledge (not just coastal data) as a major contribution to regional planning and monitoring, even at transnational level. In regard to this perceived need, two issues arise:

1. The landscape/seascape paradigm offers a comprehensive perspective of both the physical and human/cultural aspects and their interaction, defining the present state and heritage. At the European level, the European Landscape Convention (2000) may be assumed as a reference for Coastal GIS attuned for Administration in the governance, planning and design phases. Definition of relationships with non-European landscape policies should be sought.

2. The operation of data, jointly with the implementation of data infrastructures, may be regarded as a chief subject for GIS optimisation. It is hoped to create a link with the running global and pan-European initiatives and/or policies by offering a contribution for Data Infrastructure Profiles suitable for coastal GIS and or promoting these achievements towards non-GIS and non-ICAM specialists, addressing the concerned stakeholders in public administration and industry.

The legal requirement to use GIS for monitoring the Water Framework Directive, and the implied directive to use GIS to monitor the EU ICZM Recommendation both focus on primarily physical data, i.e., coastal or benthic flora and fauna, geomorphology, etc. Yet for wider planning and monitoring purposes, many administrative and non-physical data sources will be needed. These must somehow be accommodated by the evolving coastal SDI.

As we write, it is fairly clear that CoastGIS as a gathering will be around for the foreseeable future. Perhaps the main intangible, but nonetheless very real, benefit from the series of gatherings has been the camaraderie and consequent networking of many of the main contributors. However, as researchers and practitioners we all need more tangible records of these significant events. We trust that the presentation of part of the ongoing record in this volume will contribute to the development and improvement of coastal zone management around the globe.

Ron Furness *– Chair, International Cartographic Association Commission on Marine Cartography*
Andy Sherin *– Chair, CoastGIS 2001 Science Committee, Co-chair of the CoastGIS 2001 Organizing Committee and Coastal Information Specialist, Geological Survey of Canada*

Sydney and Ottawa, 13th November 2003

Preface

Darius Bartlett and Jennifer Smith

This book has arisen out of a decade-long collaborative initiative between the Commission on Marine Cartography of the International Cartographic Association and the Commission on Coastal Systems of the International Geographical Union, and manifested in the series of conferences known as the CoastGIS Symposia. The first CoastGIS meeting was held in Cork, Ireland, in February 1995. Since then, successive events have taken place in Aberdeen, Scotland (1997), Brest, France (1999) and Halifax, Nova Scotia (2001). The majority of chapters presented in the pages that follow had their origins in papers presented at the Halifax meeting, supplemented by a selection of additional contributions commissioned by the editors specifically for this volume.

Previous volumes have focused on GIS research in the marine and coastal realms (Wright and Bartlett, 2000) and on the application of GIS to oceanography and fisheries (Valavanis, 2002). The current volume is, to the best of our knowledge, the first to focus specifically on the role of GIS in integrated coastal zone management. We hope it will provide guidance, inspiration, encouragement and, where merited, a degree of caution, for all those tasked with the stewardship of the world's coasts, as well as for those whose interests are more academic and research-oriented.

The wide diversity of perspectives that can and must be brought to bear on the challenge of coastal zone management is reflected in the range and organisation of chapters in this book. Thus the opening chapters focus on technical issues, ranging from the incorporation of GIS within wider information infrastructures to techniques of visualisation, the importance of error and uncertainty in coastal databases, and the interfacing of GIS with simulation and process models. This is followed by a number of chapters that step back from technology, and which seek to put coastal zone GIS into a more human context, particularly through examination of cultural issues and exploration of techniques for incorporating traditional ecological knowledge within GIS-enabled coastal management regimes; and, finally, attention focuses on the use of GIS to historic shoreline change analysis, the application of geomatics to estuary management,

and to better understanding and management of environmentally sensitive shorelines.

We are particularly delighted that contributions to this volume have come from each of the inhabited continents of the world, namely from Africa, Asia, Australia, Europe and North and South America. The diversity of perspectives on coastal management arising from the cultural and professional backgrounds of the authors, and also from the range of geographic locations used in the case studies and applications reported on, underscores the truly international dimension of coastal management today.

As always, compilation of an edited collection of papers depends on the support, encouragement and assistance of a vast number of people who have worked "behind the scenes." It is, of course, a pleasure to thank the authors who have contributed chapters to the book, and who have borne with cheerful patience the many demands – some reasonable, some perhaps less so – of the editors. We also acknowledge with gratitude the support of the International Geographical Union and the International Cartographic Association.

On an individual level, to merely "thank" Ron Furness and Andy Sherin seems woefully inadequate: it is no exaggeration to say that, without the sustained friendship and cheerful encouragement of both these gentlemen, this volume simply would never have seen the light of day. No less valued was the encouragement of our friends and colleagues on the International CoastGIS Scientific Committee, past and present.

Closer to home, Darius Bartlett wishes to thank friends and colleagues within the Geography Department and the Coastal and Marine Resources Centre at University College Cork; his postgraduate students for their lively discussions and thought-provoking questions; and, above all Mary-Anne, Becky and dog Jessa for putting up with my irregular hours, my absences from home and my all-too-frequent neglect of domestic duties and responsibilities. For her part, Jennifer Smith would like to thank Andy Sherin and the Canadian CoastGIS committee who facilitated her involvement in this project.

Finally, both authors acknowledge with gratitude the assistance, support and guidance of Tony Moore at Taylor & Francis in London and Randi Cohen and Jay Margolis at CRC Press in Florida, who helped steer production of this volume from conception through all stages of publication to its final appearance on booksellers' shelves.

REFERENCES

Valavanis, V. D., 2002, *Geographic Information Systems in Oceanography and Fisheries*, (London: Taylor and Francis).

Wright, D.J. and Bartlett, D.J., 2000, *Marine and Coastal Geographical Information Systems.* (London: Taylor and Francis).

Contributors

THE EDITORS

Darius Bartlett
Department of Geography, University College Cork, Cork, Ireland. Phone: +353 21 4902835;
Fax: +353 21 4902190; e-mail: djb@ucc.ie

Darius Bartlett first encountered GIS as a postgraduate student at Edinburgh University in about 1982, and has been researching and writing on conceptual, institutional and related issues arising out of coastal zone applications of GIS since the mid-1980s. More recently, he has started investigating the incorporation of marine and coastal areas into SDI initiatives, issues surrounding the diffusion to and use of GIS in the Developing World, and use of GIS by community groups, voluntary organisations and Non-Governmental Organisations (NGOs). An avid traveller, he has so far visited over 65 countries around the world, and looks forward to visiting the remainder in due course. He is a Member of the International Geographical Union's Commission on Coastal Systems; and was one of the founder organisers of the CoastGIS series of biannual conferences.

Jennifer L. Smith
World Wildlife Fund Canada, Atlantic Office, Suite 1202, Halifax, Nova Scotia, Canada, B3J 1P3. Phone: (902) 482-1105; Fax: (902) 482-1107; e-mail: jsmith@wwfcanada.org

Jennifer Smith manages the application of GIS in conservation planning for World Wildlife Fund Canada’s Marine Program. She holds an Honours degree in Geography from McGill University, Montreal. Her interests in work and studies have focused on environmental monitoring, change in ecological systems, seagrass ecosystems, developing areas and GIS-based decision support for protected areas network design.

THE AUTHORS

Roger A. Longhorn
ral@alum.mit.edu

Roger Longhorn is an independent ICT policy consultant who holds B.Sc. and M.Sc. degrees in Ocean Engineering and Shipping Management from M.I.T, Cambridge, MA, USA. After more than a decade of implementing marine information systems for global maritime clients, in 1989 Roger became an external ICT expert to the European Commission, where since 1992 he focused on GIS technology and markets in the evolving Information Society. His special area of interest is GIS applied to the coastal zone. Currently a Ph.D. candidate in Information Policy at City University, London, his research focuses on regional spatial data infrastructures.

Simon Gomm
Ordnance Survey, Romsey Road, Southampton. SO16 4GU. UK. Phone: (+44)023 80305149

Simon has worked for Ordnance Survey in a variety of roles including geodetic surveying and data quality assurance and is now a Senior Research Leader where he is responsible for coordinating research on topics related to the capture, maintenance and use of spatial data.

Rongxing (Ron) Li
Department of Civil & Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210-1275, Tel. (614) 292-6946, Fax. (614) 292-2957; e-mail: li.282@osu.edu, http://shoreline.eng.ohio-state.edu

Dr. Ron Li is a professor at the Department of Civil & Environmental Engineering and Geodetic Science of The Ohio State University. His research interests include digital mapping, coastal and marine GIS, spatial data structure, Mars Rover localization and landing site mapping.

Kaichang Di
Department of Civil & Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210-1275, Tel. (614) 292-4303, Fax. (614) 292-2957; e-mail: di.2@osu.edu

Dr. Kaichang Di is a research associate at the Department of Civil & Environmental Engineering and Geodetic Science of The Ohio State University. His current research interests are coastal mapping using high-resolution satellite imagery, Mars Rover localization and landing site mapping.

Ruijin Ma
Department of Civil & Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210-1275, Tel. (614) 292-4950, Fax. (614) 292-2957; e-mail: ma.106@osu.edu

Mr. Ruijin Ma is a Ph.D. candidate at the Department of Civil & Environmental Engineering and Geodetic Science of The Ohio State University. His current research interests are coastal mapping and GIS, 3D model reconstruction from LiDAR and photographs, and remote sensing applications.

Paul Pan
62 Llanishen Street, Cardiff, CF 14 3QD, United Kingdom; e-mail: pan@cardiff.ac.uk

Paul Pan was the Principal Investigator for a number of innovative coastal monitoring projects in South Wales. He lectured and researched on GIS with the University of Wales, Swansea and Cardiff University for over 10 years. Paul now works as an independent consultant

Eleanor Bruce
School of Geosciences, University of Sydney, Sydney, NSW, 2006, Australia. Phone: +61 2 9351 6443; Fax: +61 2 9351 3644; e-mail: ebruce@geosci.usyd.edu.au

Eleanor Bruce is a senior lecturer in the School of Geosciences at the University of Sydney. She teaches GIS, coastal management and advanced spatial data analysis. Her research interests include the use of GIS for marine park zoning, nearshore habitat mapping and coastal system modelling. Currently, she is Assistant Director of the Spatial Science Innovation Unit at the University of Sydney.

Sam Macharia Ng'ang'a
Department of Geodesy and Geomatics Engineering, University of New Brunswick, P.O. Box 4400, Fredericton, New Brunswick, Canada, E3B 5A3. Phone: (506) 447-3259 or (506) 455-7073; e-mail: sam.nganga@unb.ca

Sam Macharia Ng'ang'a obtained a Bachelor's degree in surveying from the University of Nairobi, Kenya and a Master's degree (Land Information Systems) from the Department of Geodesy and Geomatics Engineering, University of New Brunswick (Canada). He is a part time lecturer at UNB and is currently completing his Doctorate degree on marine protected area information systems. He holds memberships in (among other institutions) the Canadian Institute of Geomatics (CIG) and the Canadian Hydrographic Association (CHA).

Chul-sue Hwang
Department of Geography, Kyung Hee University, Seoul 131-701, Republic of Korea. Phone: +82-2-961-9313; e-mail: hcs@khu.ac.kr

Chul-sue Hwang is assistant professor of geography at Kyung Hee University, Korea, and a member of the editorial board of the *Journal of the Geographical Information System Association of Korea*. His recent research focuses on uncertainty of remote sensing data, spatial data mining, and exploratory spatial data analysis.

Cha Yong Ku
Department of Geography, Sangmyung University, Jongro-gu, Seoul 110-743, Republic of Korea. Phone: +82-2-2287-5043; e-mail: koostar@smu.ac.kr

Cha Yong Ku is assistant professor of geography at Sangmyung University. His research interests include the integration of GIS and remote sensing, classification accuracy assessment and scale effects in remote sensing, and land use/land cover information extraction and modelling for coastal wetlands. He received his Ph.D. in Geography from Seoul National University.

Simon Jude
School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, United Kingdom, NR4 7TJ. Phone: +00 44 1603 591360; e-mail: s.jude@uea.ac.uk.

Dr. Simon Jude is a research associate in the School of Environmental Sciences at the University of East Anglia UK. His research involves developing the use of GIS, virtual reality and visualisation techniques for coastal decision-making.

Andrew Jones
School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, United Kingdom, NR4 7TJ. Phone: +00 44 1603 593127; e-mail: a.p.jones@uea.ac.uk

Dr. Andrew P. Jones is a lecturer in the School of Environmental Sciences at the University of East Anglia UK.

Julian Andrews
School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, United Kingdom, NR3 7TJ. Phone: +00 44 1603 592536.

Dr. Julian E. Andrews is a sedimentologist at the School of Environmental Sciences (University of East Anglia, Norwich UK) with special interest in modern and Holocene coastal sediments.

Roberto Mayerle
Coastal Research Laboratory (Corelab), Christian Albrechts University, Otto Hahn Platz 3, 24118 Kiel, Germany; e-mail: rmayerle@corelab.uni-kiel.de

Professor Mayerle is a specialist in numerical modelling of waves, currents, sediment transport and morphological changes in rivers, coastal areas and near hydraulic structures, with extensive experience in physical modelling of hydroelectric power schemes. He graduated as a civil engineer at the Federal University of Paraná in Brazil in 1979 where he worked for an energy concern involved in the construction and operation of several major hydropower schemes such as Itaipu. In 1988 he got his Ph.D. degree at the University of Newcastle-upon-Tyne in England. From 1989 till 1992 he was in-charge of a team working on the development of numerical models at the National Centre for Computational Hydroscience and Engineering in Oxford, USA. From 1992 till 1996 he worked at the Institute of Fluid Mechanics of the University of Hanover in Germany, being in charge of several research projects dealing with the investigation of the impact of climate changes on the morphological development both on the German North and Baltic Seas. Since 1996 he heads the Coastal Research Laboratory (Corelab) of the University of Kiel. Corelab is a research and teaching unit established to foster research in coastal environments. The Laboratory is engaged in applied research using a combination of in situ measurements and investigations as well as databases and numerical models embedded into decision support systems to help in the management of coastal areas.

Fernando Toro
Wilrijkstraat 37, 2140 Antwerpen, Belgium; e-mail: toro.gunst@skynet.be

Fernando Toro was born in Medellin, Colombia. He received his degree in civil engineering in 1993. He worked in Colombia in the construction of the Metro in the city of Medellin and in an engineering consultant company for three years. He pursued a Master of Science in computational hydrosciences, in the National Center for Computational Hydroscience and Engineering, at The University of Mississippi, USA, from 1996 to 1998. In 1998, he moved to Germany and did his Doctorate studies in the Coastal Research Laboratory, at the Christian Albrechts University in Kiel, until Summer 2003. He is presently working in an engineering consultant company in Antwerp, Belgium. His interests are numerical models and GIS applied to engineering problems.

Jacques Populus
Service des Applications Opérationnelles DEL/AO, IFREMER, BP 70, 29280 Plouzané, France
Phone: 0298224310; Fax : 0298224555; e-mail: jpopulus@ifremer.fr

Jacques Populus is a civil engineer who originally specialised in applications of high resolution remote sensing to coastal studies. His current activities concern the handling of geo-information for coastal applications, with a view to making it available to practitioners, in both developed and developing countries. This concerns GI as output of data analysis, remote sensing and hydrodynamic

modelling. Current applications deal with water and its use for aquaculture sustainability. More recently, he has focused on acquisition and processing of physical data (topography and bathymetry) in the coastal zone by way of innovative techniques such as the LiDAR.

Lionel Loubersac
Head of the Coastal Environment and Living Resources Laboratory of Ifremer in Sète (France). He is in charge of the organisation, coordination and planning of scientific and technical programs dealing with a) the monitoring of coastal environment and living resources quality along the Mediterranean shores, b) the development, interfacing and transfer of tools in the field of coastal oceanography; i.e. environmental monitoring networks and databases, hydrodynamic modelling, geographical data bases and their integration within standardized coastal GIS, new technologies of information and communication for transferring to the public scientific results on coastal environment quality. He is involved in training courses, scientific committees of symposiums as well as numerous projects at national, European and international levels in the field of the applications of Remote Sensing and Information Systems to coastal management. He has been selected as European evaluator and expert in the framework of the DG INFSO Programme Information Society Technology, Key Action "Systems and Services for the Citizen; Applications related to environmental protection."

Jean-François Le Roux
Jean-François is a computer engineer with a Master's Degree in computer science. He has been a software engineer at Simulog, a high-technology service company specializing in technical software engineering, between 1997 and 2000, detached at Thomson CSF Optronics, involved in several projects of technical and operational simulation (real-time 3D simulation). Now he is a system and software engineer at Ifremer for the operational applications service of the Environment and Coastal Planning Division (DEL/AO). He works on interfacing GIS, web and hydrodynamic models.

Franck Dumas
A coastal ocean circulation modeler, Franck graduated from Ecole Nationale Supérieure des Techniques Avancées. For the past seven years, he has been using the MARS-2D-3D modelling system intensively in various frameworks: in the context of pure physical oceanography and circulation along the French coast, but also in pluridisciplinary contexts for the understanding of complex coastal ecosystems such as that of the Baie du Mont Saint-Michel. He is now in charge of developing Ifremer's coastal ocean modelling system, and integrating all coastal ecosystem components developed over the past years within the Institute.

Valerie Cummins
Coastal and Marine Resources Centre, Environmental Research Institute, University College Cork, Naval Base, Haulbowline Island, Cork Harbour, Ireland. Phone: +353 (21) 4703100; Fax: +353 (21) 4703132; e-mail: v.cummins@ucc.ie

Valerie Cummins (B.Sc., M.Sc.) received her B.Sc. in marine geography from University of Wales, College of Cardiff, and her M.Sc. through the Department of Zoology and Aninmal Ecology in University College Cork. After several years working in the field of spatial data analysis/GIS in the UK at the British Oceanographic Data Centre, and subsequently at Landmark Information Group, Valerie returned to Ireland in 1999 to join the Coastal & Marine Resources Centre. She is currently manager of the CMRC with responsibility for the co-ordination of nineteen national and European funded research programs, with a current staff of 17.

Gerry Sutton
Coastal and Marine Resources Centre, Environmental Research Institute, University College Cork, Naval Base, Haulbowline Island, Cork Harbour, Ireland. Phone: +353 (21) 4703100; Fax: +353 (21) 4703132; e-mail: gerry.sutton@ucc.ie

Gerry Sutton graduated from University of Wales, Bangor, with Joint Honours degree in marine biology and zoology in 1984, following which he worked for two years as a fisheries officer with the Department of Fisheries in Sabah, East Malaysia. Here he focussed on research, development and subsequent local adoption of macro-algal cultivation techniques (*Eucheuma* spp.). Following his return to Ireland Gerry joined Hydrographic Surveys Ltd where, as senior hydrographic surveyor between 1991 and 1998, he was responsible for planning, conducting, processing and charting high precision coastal surveys associated with national and international civil engineering, dredging, and oceanographic projects. Since joining the CMRC in 1998 Gerry has been actively conducting research within the CMRC team, contributing to a number of EU and nationally funded projects, and consultative reports. Gerry recently received his M.Sc through the Department of Geography at University College Cork, in the field of marine resources and GIS. His primary research interests are currently in the fields of seabed mapping (specialising in multibeam sonar acquisition and processing); geophysics; oceanography; marine resources and technology; and marine geographic information systems. Gerry has been a professional member of the Irish Institution of Surveyors (IIS) since 1997.

Françoise Gourmelon
Laboratoire Géomer, LETG UMR6554 CNRS, European Institute for Marine Studies (IUEM), Technopôle Brest-Iroise, 29280 Plouzané, France. Phone: + 00 33 2 98 49 86 83; Fax: + 00 33 2 98 49 86 86; e-mail: francoise.gourmelon@univ-brest.fr

Françoise Gourmelon is a senior research worker at the CNRS (French National Center for Scientific Research). She is a member of the Géomer laboratory (LETG UMR 6554 CNRS), part of the European Institute of Marine Studies. She has been using GIS in ICZM studies since 1989, when she met François Cuq (Director of the CNRS Géosystèmes laboratory). She has been working with coastal and marine applications such as landuse and landcover changes in protected islets, habitat of marine fauna, in temperate and tropical coastal zones. Her current focuses are on the coupling of models with GIS for better knowledge of social/natural dynamics; on the use of remote sensing data for the long term monitoring of protected areas; and on the design of geomatic infrastructures dedicated to long term environmental observation.

Iwan Le Berre
Laboratoire Géomer, LETG UMR6554 CNRS, European Institute for Marine Studies (IUEM), Technopôle Brest-Iroise, 29280 Plouzané, France. Phone: 00 33 2 98 49 86 80 ; Fax: 00 33 2 98 49 86 86; e-mail: iwan.leberre@univ-brest.fr

Iwan Le Berre acquired his first experience with GIS when he joined the Géosystèmes laboratory as a Master of Science student in 1990 (and has stayed ever since!). His research interests deal with the implementation of GIS for synthetic mapping of coastal and marine environment. In 1992, he got a grant from the MaB committee of UNESCO for the achievement of a synthetic map of the Iroise Sea Biosphere Reserve. He defended his Ph.D., focused on the implementation of a coastal and marine GIS for the Iroise Sea, in 1999 and has been awarded by the Regional Council of Brittany in 2001. After several studies with different organisations in France (National Park Service, National coastal zone heritage agency, Ministry of Equipment) he got his position at the Western Brittany University in 2001 and he now teaches cartography, GIS and remote sensing for graduate, master and Ph. D. students.

R. Sudarshana
RI227, National Remote Sensing Agency, Dept. of Space, Govt. of India, Balanagar, Hyderabad 500 037, India; e-mail: sudarshana_r@nrsa.gov.in

R. Sudarshana conducted his doctoral research on seabed ecology in the Arabian sea and taught in Karnatak University, India for a while. He then joined the Department of Space, Government of India, wherein he developed an academic niche in marine science, developing and propagating remote sensing application skills. After several years of work in the field of marine applications of remote sensing, he now heads the programme planning affairs of the national remote sensing activities. He has been a consultant to

UNESCO in Africa, Arabia and Iran, besides being a visiting professor in institutions in Europe and Japan.

John Lindsay
John A. Lindsay, NOAA Pribilof Project Office, 7600- Sand Point Way NE, Building 3, Seattle, WA 98115. Phone: (206) 526-4560; e-mail: john.lindsay@noaa.gov

John A. Lindsay is the director of the Pribilof Islands Environmental Restoration Project. He began work with NOAA more than fifteen years ago following a long career as a marine invertebrate taxonomist and ecologist. At NOAA, Mr. Lindsay implemented the agency's natural resource trustee responsibility nationwide at hazardous waste sites, and chemical and oil spills. He also represented the U.S. Department of Commerce on the Joint U.S./Canada Atlantic Regional Environmental Emergencies Team.

Thomas J. Simon
NOAA Environmental Compliance, Health & Safety, and Security Office, 7600-Sand Point Way NE, Building 1, Seattle, WA 98115. Phone: (206) 526-6295; e-mail: Tom.Simon@NOAA.gov

Tom Simon is a geographer and GIS specialist who helped develop the NOAA Pribilof Project Office GIS project. His interests include the use of GIS to integrate varying technologies and develop interactive presentation tools.

Aquilina D. Lestenkof
Aleut Community of St. Paul Island, Tribal Government, P.O. Box 86, St. Paul Island, AK 99660. Phone: 907-546-2641; Fax: 907-546-2655; E-mail: aquilina@tdxak.com

Aquilina D. Lestenkof of St. Paul Island continually seeks ways to balance and blend the cultural knowledge of her people – Unangan (Aleut) – with present day life. Aquilina currently co-directs the Ecosystem Conservation Office of the Aleut Community of St. Paul Island's Tribal Government.

Phillip A. Zavadil
Aleut Community of St. Paul Island, Tribal Government, P.O. Box 86, St. Paul Island, AK 99660,
Phone: 907-546-2641; Fax: 907-546-2655; e-mail: pazavadil@tdxak.com

Phillip A. Zavadil of St. Paul Island co-directs the Ecosystem Conservation Office of the Aleut Community of St. Paul Island's Tribal Government. Mr. Zavadil enthusiastically seeks means to understand and improve the quality of life on St. Paul Island.

Peter Nwilo
Department of Surveying and Geoinformatics, University of Lagos, Lagos, Nigeria.

Dr. Nwilo is an associate professor of surveying and geoinformatics at the University of Lagos, Lagos, Nigeria, and is currently the sub-dean of the Postgraduate School of the University. Dr. Nwilo has a B.Sc. and M.Sc. in Surveying from the University of Lagos and a Ph.D. in environmental resources from the University of Salford, United Kingdom. His special research interests cover sea level variations and the impacts along the coast of Nigeria; GIS and its applications in coastal management; subsidence phenomenon along the coast of Nigeria; management of the navigable rivers of Nigeria; and management of Nigerian coastal areas. He has published extensively on these topics.

Tim Webster
Email: Timothy.Webster@nscc.ca

Tim Webster is a research scientist with the Applied Geomatics Research Group (AGRG) at the Centre of Geographic Sciences (COGS), a component of the Nova Scotia Community College. He currently is also pursuing his Ph.D. in earth sciences from Dalhousie University where he is applying DEM technologies including LIDAR to mapping geological landforms.

Charles Sangster
Email: sangstcw@gov.ns.ca

Charles Sangster is a graduate from the Applied Geomatics Research Program at COGS, and works in the field of conservation GIS. Currently, he is the GIS specialist for the Protected Areas Branch of the Nova Scotia Department of Environment and Labour.

Montfield Christian
Email: c_montfield@hotmail.com

Montfield Christian is a graduate from the Applied Geomatics Research Program at COGS, and works in Toronto, Ontario as an independent GIS analyst. His company is Green Eminent Consulting and they specialise in customised GIS projects.

Dennis Kingston
Email: Dennis.Kingston@nscc.ca

Dennis Kingston is the department head of the Information Technology Department, Annapolis Valley Campus, NSCC. Formerly, he taught GPS technology at COGS in the Survey Program.

Courtney Schupp
Department of Physical Sciences, Virginia Institute of Marine Science, Rt. 1208 Greate Road, Gloucester Point, VA 23062

Courtney Schupp focuses her research on shoreline behavior and its relationship to nearshore processes and underlying geology. After earning her B.S. in geology from Duke University, she worked as a GIS analyst at the U.S. Geological Survey and as a consultant to the Maryland Geological Survey. She is currently working towards an M.S. in marine science at the Virginia Institute of Marine Science at the College of William and Mary.

Rob Thieler
U.S. Geological Survey, Coastal and Marine Geology Program, 384 Woods Hole Road, Woods Hole, MA, 02543-1598, USA. Phone: (508) 457-2350; Fax: (508) 457-2310; e-mail: rthieler@usgs.gov; http://marine.usgs.gov

Rob Thieler is a research geologist with the U.S. Geological Survey in Woods Hole, MA. His research includes continental margin sedimentation, late Quaternary sea-level change, coastal hazards, and development of software to quantify rates of shoreline change.

James F. O'Connell
Woods Hole Oceanographic Institution Sea Grant Program, 193 Oyster Pond Road, MS#2, Woods Hole, MA, 02543-1525, USA. Phone: (508) 289-2993; e-mail: joconnell@whoi.edu

Jim is presently the coastal processes specialist with the Woods Hole Oceanographic Institution Sea Grant Program and Cape Cod Cooperative Extension (past 5 years). Prior to joining WHOI, Jim was the coastal geologist and coastal hazards specialist with the Massachusetts Office of Coastal Zone Management for 13 years, followed by a year with the Cape Cod Commission as a marine resources specialist. Jim's speciality is shoreline change analysis, analyzing human effects on coastal processes and coastal landforms, coastal floodplains, erosion control alternatives, and coastal hazard mitigation issues.

Jorge I. Euán-Avila,
Department of Marine Resources, CINVESTAV, Mérida, México; e-mail: euan@mda.cinvestav.mx.

Jorge I. Euán-Avila, Ph. D., is an engineer and researcher in the Department of Marine Resources, CINVESTAV, Mérida, México.

María de los Angeles Liceaga-Correa
Department of Marine Resources, CINVESTAV, Mérida, México; e-mail: liceaga@mda.cinvestav.mx.

María de los Angeles Liceaga-Correa,Ph.D, is a mathematician and researcher in the Department of Marine Resources, CINVESTAV, Mérida, México.

Héctor Rodríguez-Sánchez
Department of Marine Resources, CINVESTAV, Mérida, México; e-mail: hrs@mda.cinvestav.mx.

Héctor Rodríguez-Sánchez, M.Sc., is a biologist and research assistant in the Department of Marine Resources, CINVESTAV, Mérida, México.

David R. Green
Centre for Marine and Coastal Zone Management (CMCZM) / Aberdeen Institute of Coastal Sciences and Management (AICSM), Department of Geography and Environment, School of Geosciences, College of Physical Sciences, University of Aberdeen, Elphinstone Road, Aberdeen, AB24 3UF, Scotland, UK. / Tel. +44 (0)1224 272324 /Fax. +44 (0)1224 272331; e-mail: d.r.green@abdn.ac.uk Internet. www.abdn.ac.uk.cmczm or www.abdn.ac.uk/aicsm

David R. Green was educated in geography at the Universities of Edinburgh, Pennsylvania and Toronto. He is currently director of the Centre for Marine and Coastal Zone Management (CMCZM), and assistant director of the Aberdeen Institute for Coastal Sciences and Management (AICSM) at the University of Aberdeen in Scotland. He lectures in the environmental applications of the geospatial technologies (including remote sensing and GIS) with special research interests in online Internet-based GIS, mobile mapping, and user interfaces for use in integrated coastal zone management. David is currently president of the EUCC-The Coastal Union, and GIS Editor of the *Journal of Coastal Conservation* (JCC). He is also chief assessor and examiner for the ASET GIS Programme, chairman of the AGI Marine and Coastal Zone Management GIS Special Interest Group, and deputy chair of the ICA Commission on Marine Cartography (International Cartographic Association). David has presented papers and workshops at a number of national and international conferences, and has published a number of books on the use of GIS in landscape ecology, school education, marine and coastal applications, and use of remote sensing for coastal applications. He is currently involved in organising the Littoral 2004 and CoastGIS 2005 conferences.

Stephen D. King
Centre for Marine and Coastal Zone Management, Department of Geography and Environment, University of Aberdeen, Elphinstone Road, Aberdeen, AB24 3UF, UK.; Tel. +44 (0)1224 272324, Fax +44 (0)1224 272331; e-mail: s.d.king@abdn.ac.uk

Stephen is currently a research assistant in the Department of Geography and Environment at the University of Aberdeen. He has a B.A. in geography

(University of Liverpool) and an M.Sc. in environmental remote sensing (University of Aberdeen). His main interests are the application of remote sensing and GIS to coastal zone management and, in particular, the development of Internet-based coastal information systems and the use of mobile communications technology and GIS for field data collection.

Jean-Côme Bourcier
8 Rue Georges Laroque, Appartement 1021, 76300 Sotteville-Les-Rouen FRANCE. Phone: +33 02 63 87 30 or +33 06 88 43 54 95; e-mail: jean-come.bourcier@wanadoo.fr or geoinformatique@wanadoo.fr

J.-C. Bourcier obtained his Ph.D. at University of Le Havre, France, and works as a GIS specialist, environmental engineer and territorial planning engineer. His research concerns the use of GIS as a tool designed to create a better understanding of coastal processes, functioning and change and, more specifically, better management and development of coastal spaces such as estuaries. He is currently the president of GeoInformatique, an association which helps people to better use geographical information.

Anders Mosbech
National Environmental Research Institute, Department of Arctic Environment, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark. Phone: +45 4630 1934; Fax: +45 4630 1914; e-mail: amo@dmu.dk

Anders Mosbech, Ph.D., is a senior scientist who performs ecological research in Greenland on marine birds, geese, muskoxen, marine mammals and vegetation. He is an advisor to the regulatory agencies in the Greenlandic and Danish governments concerning environmental impacts of oil exploration and development in Greenland.

David Boertmann
National Environmental Research Institute, Department of Arctic Environment, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark. Phone: +45 4630 1934; Fax: +45 4630 1914; e-mail: dmb@dmu.dk

David Boertman is a senior research biologist with long experience in arctic ornithology and Arctic natural history, and has worked with environmental issues related to oil exploration in the arctic. He is currently associated with the National Environmental Research Institute, Denmark.

Louise Grøndahl
National Environmental Research Institute, Department of Arctic Environment, Frederiksborgvej 399, P.O. Box 358, DK-4000 Roskilde, Denmark. Phone: +45 4630 1934; Fax: +45 4630 1914; e-mail: lgr@dmu.dk

Louise Grøndahl holds an M.Sc. in physical geography. She has worked as research assistant at NERI with GIS and is now doing a Ph.D. on carbon flux in Northeast Greenland at the National Environmental Research Institute, Denmark.

Frants von Platen
Geological Survey of Denmark and Greenland (GEUS). E-mail: fph@geus.dk

Frants von Platen is a geographer and GIS-specialist who has been employed since 1983 at GEUS, where he works with geological databases and GIS development.

Niels Nielsen
Institute of Geography, University of Copenhagen, Denmark. Phone: +45 35322508; Fax: +45 35322501; e-mail: nn@geogr.ku.dk; www.geogr.ku.dk

Niels Nielsen is an associate professor in physical geography (coastal geomorphology) at the Institute of Geography of the University of Copenhagen.

Søren Stach Nielsen
Greenland Institute of Natural Resources, P.O. Box 570, DK-3900 Nuuk, Grønland. Phone: +299 328095 /ext. 243; Fax: +299 325957; email: SorenSN@natur.gl

Søren Stach Nielsen holds a Master of Technological and Sociological Planning degree from Roskilde University and is at present a project-researcher at the Greenland Institute of Natural Resources, investigating socioeconomic relations between hunting and hunting-gear and hunting and household economy, and incorporation of local knowledge into scientific research.

Morten Rasch
Danish Polar Center, Strandgade 100H, DK-1401 Copenhagen K, Denmark. Direct phone: +45 32880110 Mobile phone: +45 23490645; Secretary: +45 32880100; Fax: +45 32880101; e-mail: mr@dpc.dk; homepage, Zackenberg: www.zackenberg.dk, homepage, Danish Polar Center: www.dpc.dk

Morten Rasch, Ph.D, is a coastal geomorphologist educated at the Institute of Geography of the University of Copenhagen. He is a specialist in arctic coastal geomorphology and Holocene relative sea-level changes in Greenland. He has been working as station manager of Zackenberg Station, North East Greenland and is also head of logistics at Danish Polar Center.

Hans Kapel
SILA, The Greenland Research Center at The National Museum of Denmark, Frederiksholms Kanal 12, DK 1220 Copenhagen K. Phone: +45 33473256; e-mail. hans.kapel@museum.dk

Hans Kapel is a curator and archaeologist who has conducted archaeological research in Greenland since 1969. His field of responsibility includes The Central Register of Protected Monuments in Greenland.

André Laflamme
Environment Canada - Environmental Emergencies Section, Dartmouth, Nova Scotia, CANADA. Phone: 902-426-5324; Fax: 902-426-9709; e-mail: Andre.laflamme@ec.gc.ca

André Laflamme graduated from University of Sherbrooke, Quebec, in physical geography in 1994. He has worked since with Environment Canada Atlantic Region with the environmental emergencies section as an environmental emergencies officer. André is also the regional sensitivity mapping coordinator for Environment Canada. He was a co-author on scientific papers produced during several conferences such as the International Oil Spill Conference (IOSC), and the Arctic and Marine Oil Spill conference (AMOP). André was also part of a team tasked to develop a national oil spill contingency plan for the Republic of Chad (Africa) where he spent some time.

CHAPTER ONE

Coastal Spatial Data Infrastructure

Roger A. Longhorn

1.1 INTRODUCTION

The term Spatial Data Infrastructure (SDI) is now in common use in countries around the world, although definitions for the term differ quite considerably. The stated objectives of SDI initiatives vary as much as do the definitions, legal mandates, types of organisation responsible for specifying and implementing SDI and actual progress achieved in creating national and regional SDIs.

One complication in specifying any SDI is the nature of spatial information, i.e. information with an important location attribute, often said to represent 80% of all information held, especially at government level. The visionaries and designers of SDI must accommodate the widely varying information needs of highly diverse disciplines and sectors of society, business and government. Health epidemiologists are seldom interested in the same spatial data as geological surveyors, air traffic controllers or coastal zone managers. Yet an important overlap in jurisdiction and information needs may arise, e.g. when a potential health epidemic is generated by toxic chemical concentrations in marine fauna later consumed by area residents. Then knowledge of the coastal zone flora and fauna, hydrography, tidal states, nearby land use practices of industry and agriculture and transport routes, fishing practices and zones all become intertwined. The complex relationships between different types of spatial information are one reason that countries take different routes to specify their SDI, ranging from visions to strategies to goals to detailed content (data and standards) and implementation plans (rules and regulations).

We all recognize that the coastal zone is a difficult geographical area to manage due to temporal issues (tides and seasons) and the overlapping of physical geography and hydrography (offshore, near shore, shoreline, inshore), of jurisdictions, legal mandates and remits of government agencies and the often competing needs of stakeholders. Typically, many different local, national and regional government agencies are responsible for different aspects of the same physical areas and uses of the coastal zone, e.g. fisheries, environment, agriculture, transport (inland and marine), urban planning and cadastre, national mapping agency and the hydrographic service.

Conflict resolution in this complex piece of real estate called the “coast” is sometimes used as justification for special attention for the coast within NSDI initiatives. As Bartlett (2000) expresses it: “Given the diversity of interest groups, stakeholders, managerial authorities and administrative structures that converge at

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

the shore, conflicts are almost inevitable between and among coastal users, managers, developers and the wider public, as well as between human society and the natural environment."

Because of such complex physical and institutional relationships, it is not possible to develop a coastal SDI (CSDI) in isolation from the broader National SDI (NSDI) for a nation or Regional SDI (RSDI). CSDI will necessarily be a subset of a more comprehensive NSDI because the coastal zone covers multiple physical and institutional spaces included in the generic NSDI. For that reason, it is important that people and agencies with specific knowledge and experience of the coastal zone and marine offshore areas and information requirements be an integral part of the NSDI and RSDI planning process.

1.2 CSDI WITHIN NATIONAL, REGIONAL AND GLOBAL SDI

Few nations have specified separate SDI components for individual sectors such as the coastal zone. In the USA, a Coastal NSDI vision exists based on four goals that relate to the USA NSDI (NOAA, 2001). Practical implementation work is undertaken by the National Oceanic and Atmospheric Administration (NOAA) and the Federal Geographic Data Committee (FGDC).

Canada has proposed a Marine Geospatial Data Infrastructure initiative (MGDI) within the national Canadian Geospatial Data Infrastructure (CGDI) (Chopin and Costain, 2001). MGDI is seen as an extension to the CGDI in response to "the need for a comprehensive, integrated and common infrastructure of marine data and information … accessible to all stakeholders."

1.2.1 Regional SDI Initiatives

At regional level, the INSPIRE project (Infrastructure for Spatial Information in Europe) recognizes hydrographic data as one of its "selected topographic themes" for data content for a Regional SDI for Europe (INSPIRE, 2002). More detailed data requirements are presented in the European Commission's Water Framework Directive (European Commission, 2000), which underpins the need for INPSIRE. These two European regional projects are discussed more fully in a later section as they have legal mandates from a recognized regional government body—the European Commission—and have developed detailed descriptions of spatial data requirements, including marine and coastal zone data.

Work on specification for an Asia-Pacific SDI (APSDI) began in 1994 with formation of the Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP) under the auspices of the UN Regional Cartographic Conferences (UNRCC). The APSDI model stresses institutional frameworks, technical standards, fundamental datasets and an access network (PCGIAP, 1998).

A new initiative started in 2000 for a Regional SDI for the Americas (PC-IDEA), but it is too new to have achieved anything more than creating a basic project organisational framework at this stage.

1.2.2 Global SDI Initiatives

Global level SDI initiatives include the Global Map Project (ISCGM, 2001), which began in 1996, the Global Spatial Data Infrastructure (GSDI) Association, which began as a series of international conferences in 1996, and the Digital Earth initiative sponsored by the US government (Evans, 2001). These initiatives intend to build on existing national and/or regional SDI work and existing data resources (primarily remote sensing imagery). They typically mention shoreline data, boundary data (including marine, ocean, inland waters) and references to cadastre that may or may not include marine cadastral data. GSDI and the Global Map project are discussed more fully later.

Of more immediate interest—and importance—to coastal and marine managers and researchers is the work of the Intergovernmental Oceanographic Commission (IOC) International Oceanographic Data and Information Exchange (IODE) working group mentioned later and described in some detail by Longhorn (2002).

1.3 CSDI IN THE US

The US NSDI initiative began formally in April 1994 with a presidential Executive Order (Clinton, 1994) in response to a national performance review of federal government. One of the main elements of the US NSDI is the creation, harmonisation and promulgation of technical standards by which all federal spatial information can be recorded, metadata being of highest importance so that existing spatial data resources can be identified, accessed and used. This major task fell to the Federal Geographic Data Committee (FGDC), which was created in 1990. Responsibilities with regard to NSDI implementation for all federal agencies in the US was set out in the revised OMB Circular A-16 in 2002 (OMB, 2002).

In the US, the Coastal SDI vision is based on four goals relating to the NSDI (NOAA, 2001). The US CSDI initiative is led by the Coastal Services Centre of NOAA (National Oceanic and Atmospheric Administration). The main goals of the US CSDI are:

- that the coastal management community should understand and embrace the vision, concepts and benefits of the NSDI,
- spatial coastal and marine framework data should be readily available to the coastal management community,
- technologies to facilitate discovery, collection, description, access and preservation of spatial data should be widely available to the coastal management community, and
- NOAA should help develop and implement spatial data applications to meet the needs of the coastal and marine communities.

Note the emphasis on support for coastal resources management, a role assumed by NOAA under the 1972 Coastal Zone Management Act, as opposed to

NOAA's historical focus and initial primary mission of ensuring safety of navigation.

The main elements of CSDI that NOAA promote are bathymetry, shoreline identification and marine cadastre, although other types of data are under review for attention as a formal part of CSDI, e.g. coastal imagery, marine navigation, tidal benchmarks and benthic habitats (Lockwood and Fowler, 2001). Development of technical standards to support the CSDI is undertaken by the FGDC's Marine and Coastal Spatial Data Subcommittee, under chairmanship of NOAA/CSC (FGDC, 2002). The committee's primary mission is to develop and promote the Marine and Coastal NSDI so that "current and accurate geospatial coastal and ocean data will be readily available to contribute locally, nationally, and globally to economic growth, environmental quality and stability, and social progress" (NOAA, 2003). In 2000, this committee assumed the remit of the FGDC Bathymetric and Nautical Charting Data Subcommittee, formed in 1993.

Bathymetric data is treated as a sub-layer of the Elevation layer data in the NSDI Framework. Marine cadastre is being examined within both the FGDC Marine Boundary Working Group (MBWG) which includes members of the FGDC Cadastral Subcommittee. The MBWG was formed in 2001 to address issues relating to the legal and technical aspects of marine boundaries, with the goal to alleviate cross-agency problems concerning marine boundaries, plus provide outreach, standards development, partnerships and other data development critical to the NSDI. A major product of the MBWG work to date is the FGDC's Shoreline Metadata Profile (FGDC, 1998).

Two further standards nearing completion for the CSDI are the National Hydrography Data Content Standard for Inland and Coastal Waterways (FGDC, 2000a) and Accuracy Standards for Nautical Charting Hydrographic Surveys (FGDC, 2000b).

The US CSDI initiative should help identify the basic reference data required to achieve goals for national programs such as those stated in the Clean Water Action Plan - Coastal Research and Monitoring Strategy, which represents "the first effort to integrate coastal monitoring and research activities on a national scale to provide thorough, cross-cutting assessments of the health of the nation's coastal resources" (CRMSW, 2000). These goals include:

- Improving monitoring programs for integration at national level
- Integrating interagency research efforts
- Conducting national and regional coastal assessments, and
- Improving data management.

Interestingly, a report from the US Commission on Ocean Policy found a tight connection between inland systems for development and agriculture to areas traditionally designated as coastal. "The coastal zone is not a narrow band. It's the whole country" (US CoOP, 2002). Regarding data collection and sharing, the US Ocean Policy report further found that "there is no marine equivalent to the networks of meteorological observation stations distributed on land on all continents. Ocean observation efforts are limited temporally and spatially." This

leads to the conclusion that "there is a need for a better and more comprehensive way to link the work of different disciplines in a manner that offers a more integrated understanding of the marine environment and the processes that control it… there is a need for standardized practices and procedures" (US CoOP, 2002). These findings reinforce the premise that CSDI cannot and should not be developed in isolation from the broader NSDI of a nation or region. Actions like the Clean Water Action Plan mentioned earlier provide important input into the political process for securing added support for creating a CSDI.

1.4 CSDI IN CANADA

In Canada, the NSDI is called the Canadian Geospatial Data Infrastructure (CGDI) with the snappier, market-oriented title GeoConnections. From its beginning, the CGDI recognized that "governments have a responsibility to make geospatial information available … for developing a knowledge economy in response to the needs of citizens, industry and communities in support of economic, social and environmental well-being" (Labonte, Corey and Evangelatos, 1998). Marine navigation and charting for pollution control, coastal zone management and environmental monitoring were considered important applications to be fostered under CGDI. The CGDI has five technical and policy thrusts: access, data framework, standards, partnerships and creating a supportive policy environment. The CGDI vision is "to enable timely access to geo-info data holdings and services in support of policy, decision-making and economic development through a co-operative interconnected infrastructure of government, private sector and academia participants. From the outset, CGDI planners realized that "institutional issues will likely eclipse technology as an impediment to CGDI development and implementation" (Labonte *et al.*, 1998).

Within the framework of the CGDI, the Marine Geospatial Data Infrastructure (MGDI) is being developed in Canada, "to enable simple, third party access to data and information that will facilitate more effective decision-making" (Gillespie *et al*., 2000) for anyone involved in coastal zone management. MGDI is described as comprising data and information products, enabling technologies as well as network linkages, standards and institutional policies. The concept for an MGDI-like information network was first proposed in 1988 as the "Inland waters, Coastal and Ocean Information Network (ICOIN)" (after Butler *et al.*, 1998 in Gillespie *et al.*, 2000).

In support of the five main elements of GeoConnections (the CGDI), a Marine Advisory Committee was created in 1999 at the time of publication of a draft concept report for the MGDI (CCMC, 1999). The committee's remit is to ensure the full functionality of the CGDI in providing service to all marine stakeholders. To aid in this, a Marine Advisory Network has been set up to act as the physical focal point for stakeholder outreach and consultation (GeoConnections, 2003).

The MGDI recognizes the need for common standards permitting data to be used seamlessly across disciplines and systems. Because of the principally land-based focus of most NSDI developments, MGDI also recognizes that standards that apply perfectly well to land-based applications and data are sometimes

incompatible with the marine world. It is also understood that simply making spatial data available across the Internet does not always provide solutions to specific problems. Data is not information and information is not knowledge. MGDI also recognizes that data pricing and related policy issues dealing with intellectual copyright will be crucial to the success of both CGDI and MDGI and that, while these potential barriers may not fall easily, they are not insurmountable. The needs of users and potential stakeholders were identified more completely in an extensive requirements analysis study in 2001 (DFO, 2001).

The MGDI architecture includes:

- A common spatial data model
- An integrated process and data modelling environment
- A common spatial language and data exchange format
- Methods for managing, querying and delivering data with integrity, and
- Open source productivity tools ensuring access for all.

Progress in implementing the MGDI as a coherent system has been slower than anticipated, due mainly to two things that affect nearly all such initiatives known to this author, i.e. lack of resources (especially any new or extra resources) and institutional barriers inherent in government whether local or national. Note that even the Presidential Executive Order creating the USA NSDI did not offer any new money, rather all work to achieve the NSDI had to be conducted from within existing budgets.

1.5 CSDI ELEMENTS IN REGIONAL SDI

Regional SDI (RSDI) initiatives have started in Europe (INSPIRE), Asia-Pacific (PCGIAP) and the Americas (PC-IDEA). These projects all intend to build on National SDIs located within their geographic regions. Unfortunately, since few NSDIs have reached an advanced degree of completion—and many have not yet even been specified or received legal mandates—it is still early to see how coastal and marine SDI requirements will be accommodated within the RSDI envelopes.

The Permanent Committee for SDI for the Americas (CP-IDEA/PC-IDEA) was only formed early in 2000 under the auspices of the UN Regional Cartographic Conference. This new initiative has yet to produce any substantial documents relating to how SDI will be coordinated across the nations of Central and South America. Initially it is focusing on institutional strengthening, metadata, access policies, outreach and similar issues to those being pursued by PCGIAP and within the global GSDI Association.

1.5.1 INSPIRE—Infrastructure for Spatial Information in Europe

INSPIRE recognizes hydrographic data as one of its selected topographic themes for data content for the proposed Regional SDI for Europe (INSPIRE, 2002). Hydrography is defined as "surface water features such as lakes and ponds, streams and rivers, canals, oceans and shorelines." During the many years that European SDI consultations progressed from the GI2000 initiative (1995–1999) through ETeMII (European Territorial Management Information Infrastructure 1999–2001) to INSPIRE (2001–present), reference to specific themes for spatial data content in the proposed SDI specification was initially resisted. After much debate, the hydrographic component (including coastal zones) was included, along with only two other "topographic themes" - transport and height. Hydrography is defined in INSPIRE as "surface water features such as lakes and ponds, streams and rivers, canals, oceans and shorelines." The "height" topographic theme includes "contour data showing heights by isolines, and including with the same data set spot heights, high and low water lines, breaklines and bathymetry" (INSPIRE, 2002). At the end of the consultation period, INSPIRE was extended to cover a wide range of coastal and marine data components, including "bathymetry, coastline, hydrography, surface water bodies, water catchments, oceans and seas, oceanographic spatial features, sea regions, aquaculture facilities, polluted areas and more (INSPIRE, 2003).

1.5.2 European Water Framework Directive

The EU Water Framework Directive (European Commission, 2000) represents the culmination of five years of consultation and negotiations for implementing a harmonized and integrated water policy for all European Union Member States (now 15, soon to be 25). The WFD places quite detailed—and some say onerous—monitoring and reporting requirements for the status of surface water in four regimes: rivers, lakes, coastal waters and transitional waters (estuaries and similar bodies of water which are partly saline but strongly influenced by freshwater flows).

The data relating to coastal and transitional waters required in order for EU Member States to report to the European Commission on water condition include: detailed location (boundary) information, various biological data for aquatic flora and benthic fauna, hydromorphological data including depth variations, tidal regimes, transparency, thermal conditions, oxygenation conditions, salinity, nutrient conditions and pollution.

In order to do this reporting, a very detailed GIS specification has been produced (European Communities, 2002). For coastal waters, the data to be captured include shape, name, various identifying codes, type of water body, status of the water body (artificial, heavily modified), salinity typology, depth typology, tidal typology, and more. The specification document is nearly 170 pages long—a far cry from the "hydrography" or "marine boundary" terms found in isolation, without further explanation, in other regional and all global SDI specifications. In

fact, one worries if the European government authorities and their national oceanographic and coastal research institutions will be able to provide data at the level of detail required within the timeframe set by the Directive, which comes into force in December 2003.

The WFD Directive has been used by various Directorates General at the European Commission to justify the INSPIRE European Regional SDI initiative, on the basis that much harmonised, integrated and interoperable basic reference data will be required if WFD reporting requirements are to be met at least cost. It remains to be seen if the level of detail already specified in the INSPIRE working documents will in fact be sufficient to help fulfil that requirement.

1.5.3 Asia-Pacific Regional SDI (APSDI)

In the Asia-Pacific SDI (APSDI), what are termed "fundamental datasets" are only presented in the most general terms and with no detail. The discussion on what data should be included in the APSDI began in 1998 and has yet to reach a conclusion. Due to the large number of participants—55 nations across Asia and the Pacific are members—and very limited development resources, we should not expect the situation to change any time soon in regard to formal data specification. At the most recent meeting of the Working Group on Fundamental Data held in April 2002, the working group recommendations included (PCGIAP, 2002):

- The contents of the Asia-Pacific Regional Fundamental Dataset shall be defined, and further, the technical specifications of the Asia-Pacific Regional Fundamental Dataset shall be developed, with the Global Mapping Specifications and Administrative Boundary Pilot Project Specifications, as reference.

- For the first version of the Asia-Pacific Regional Fundamental Dataset, the vector data shall be no less accurate than 1:1 million in map scale, and the raster data shall be no less than 1KM in ground resolution.

At a 2001 meeting of the PCGIAP Working Group on Cadastre, an item was recommended for the work plan covering "identification of broad types of marine cadastres in use throughout the world as they become established" (PCGIAP, 2001). Unfortunately, the meeting also noted "concern about the apparent direction of the Cadastre WG in majoring in activities on marine cadastre, boundaries being a very sensitive issue in PCGIAP."

One surmises that little practical progress has been achieved in more detailed data specification in nearly six years of work and that the small scale of the data will be of minimal use to local coastal managers or researchers. Also, political issues may limit the degree of actual detail that will be agreed within the very broadly based group.

1.6 CSDI AT GLOBAL LEVEL

At the global level, the not-for-profit Global Spatial Data Infrastructure (GSDI) Association primarily offers a forum for discussion and exchange of experiences, especially in relation to creating SDI at national level. The GSDI's 'SDI Cookbook' (GSDI, 2001) does not specify what types of data nations should have in their SDI, but rather it concentrates on general, though still important, matters such as metadata creation, access policy and other administrative and institutional issues. In a general listing of the types of data that should be considered for including in a National SDI, the cookbook does mention cadastre and hydrography. No further details are given and there is no working group specifically targeting marine or coastal data requirements. A more thorough review of how coastal zone information is related to the GSDI can be found in Longhorn (2003b).

The Global Map specification (ISCGM, 2000) makes reference only to "coastline/shoreline" and "ocean/sea" boundary data, "ferry routes" under a "transportation" heading and edges of "islands" and "inland water" under "drainage (hydrography)." The remainder of the specification is devoted to identifying existing public domain data sources (all at scales of 1:1 million or smaller), data models for vector and raster data, precision and accuracy and metadata issues. The primary data sources listed are VMAP Level 0 from the US National Imagery and Mapping Agency (NIMA), the Global 30 Arc Second Elevation Data Set (GTOPO30) and Global Land Cover Characteristics Database, both available from the US Geological Survey and other sources. These resources have been available in the public domain for some years now, and include only small scale data, some of which are rather old compared to the time frames of interest for much coastal management and research work. For these reasons, there is little new in the Global Map project that would excite coastal managers and researchers.

What approximates a useful global marine SDI does exist for oceanographic data due to long established information management and data exchange programmes of the Intergovernmental Oceanographic Commission (IOC) (Longhorn, 2002). At its most recent meeting in June 2002, the IODE working group considered including coastal data in its remit, only to have the possibility removed from its future agenda by refusal or reluctance of the governments of some important developing country members to permit ready access to coastal zone data for national security and economic reasons.

1.7 PRINCIPAL COMPONENTS FOR CSDI

Having reviewed the few formal coastal SDI initiatives that exist, all at national level, it appears that a CSDI mainly comprises: data sources, standards, enabling technologies and institutional policies. Work in relation to the first two components needs to be carried out with a specific marine or coastal focus, which is sometimes missing from generic SDI initiatives. The latter two aspects apply to

the wider requirements of any information infrastructure, not just that of coastal SDI or even for spatial data alone.

1.7.1 Basic Reference Data for the CSDI

The presence or absence of basic reference spatial data (also sometimes called 'framework' data) in National, Regional and Global SDI implementations that relate to the needs of a viable coastal SDI are shown in table 1.1. As one might suspect, the most common are bathymetry, shoreline and boundary data. For most Regional and Global SDI initiatives, there is not sufficient detail in specification of data elements to determine whether or not the needs of coastal and marine resource managers and researchers will be met (with the exception of INSPIRE/WFD). Since the basic data will be collected at national level, this might not appear to be a serious problem at the moment. Yet when data exchange is required for research purposes, resolving boundary disputes or to satisfy a nation's responsibilities regarding various international data exchange conventions, then the absence of regional and global agreement on SDI contents and access issues will become noticed.

Table 1.1 Coastal SDI components in national, regional and global SDIs

SDI Component	USA	Canada	INSPIRE-WFD	APSDI	GlobalMap
bathymetry	yes	yes	yes	maybe	maybe
shoreline	yes	yes	yes	yes	yes
marine cadastre	yes	yes	no	yes	maybe
coastal imagery	maybe	maybe	yes	no	no
marine navigation	maybe	yes	maybe	no	maybe
tidal benchmarks	maybe	maybe	yes	no	no
benthic habitats	maybe	maybe	yes-WFD	no	no

A 'yes' indicates that the component in the left-hand column is formally listed as an important data component in the definition of spatial data infrastructure at national, regional or global level. The number of 'no' and 'maybe' entries is disquieting. Fortunately, the 'nos' appear mainly in regional or global initiatives while 'maybes' indicate that detailed user requirements or specifications have been identified and published, along with existing data sources that might provide this data. However, no firm decisions have been made as to how or if this data will be included within the higher-level NSDI or not.

1.7.2 CSDI Is More Than Data

Other components of an SDI mentioned in most SDI descriptions cover: metadata creation and standards, technical guidance (including standards) on spatial precision, accuracy and data formats (both raster and vector), data access policies (some of which are more liberal than others) and intellectual property guidance.

Standards issues in the spatial data world are now much better addressed than a mere five years ago due to the extensive work of the International Standards Organization's Technical Committee 211 (TC/211) on Geographic Information/Geomatics which is creating "a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth" comprising some 40 new GI-related standards (ISO, 2002). In parallel with ISO, the global reach and uptake of the GIS interoperability work of the Open GIS Consortium, Inc., and OGC Europe, Ltd., is providing a clear way forward in regard to integrating GI applications and data sources, especially using the Web as the service delivery machinery (OGC, 2002).

Other important standards developments relating to coastal and marine data include the S-57 (Special Publication No. 57) cartographic standard developed for and maintained by the International Hydrographic Organisation (IHO) and International Hydrographic Bureau (IHB) in Monaco (IHO, 1996). This standard is used for collection and exchange of hydrographic data among national Hydrographic offices globally. It is also very important for marine navigation as applied to the new Electronic Chart Display and Information Systems (ECDIS) now being introduced throughout the maritime industry. S-57 comprises a hydrographic data model, an object catalogue and an electronic nautical chart (ENC) product specification that are standard for ECDIS data.

Various shoreline and boundary data metadata standards have been developed at national level, as mentioned previously for the USA within the FGDC shoreline metadata working group. Looking to the future, two projects are underway that aim to specify a globally agreed standard for marineXML, a marine-specific implementation of the eXtensible Markup Language (XML) now used widely on the Internet for conveying semantic content of information as opposed to only the display specifications provided by HTML (Hypertext Markup Language). MarineXML is described as "an interoperability framework for global ocean observation systems" (ICES, 2003) which will encompass coastal zone elements as well. The development work on marineXML is undertaken by the European Union, via a part-funded project in the EU's Framework RTD programme (IOC, 2003a) in conjunction with the Intergovernmental Oceanographic Commission's Committee on International Oceanographic Data and Information Exchange (IODE) based at UNESCO headquarters in Paris (IOC, 2003b). National initiatives on creating marineXML specifications are also underway, for example in Australia and the USA (Sligoeris, 2002; Davis *et al.*, 2002).

More serious barriers remain in regard to harmonised data access policies and exploitation rights for spatial information, particularly that collected by public sector agencies. The European Union is trying to address this problem at a regional level via a new Directive setting out an agreed EU-wide framework for access to

and exploitation of public sector information (European Commission, 2002). Similar initiatives are under discussion, in consultation or being implemented in both developed and developing nations across the globe. The impact that these initiatives will have on the coastal zone (and larger research) community cannot be underestimated. Even in countries with strong "freedom of information" cultures, such as the USA, some public sector marine information is not disclosed due to fear of liability actions against the data providers (Lockwood and Fowler, 2001). With intellectual property (IP) legislation in a state of flux across the globe while attempts are made to accommodate the prior IP regime and existing international IP conventions to the digital world, resolving many of the non-technical data-related problems is far easier than the institutional ones.

1.8 CONCLUSIONS

The CZM community should be aware that their input, based on knowledge and experience of the coastal zone, is imperative during specification of SDI initiatives, whether at national, regional or global level. What we can conclude from the above examination of various types of spatial data infrastructures that are under development—national, regional and global—is that, at regional and global level, the emphasis is virtually all on process, standards, metadata and institutional issues, and rarely on data specifications for the content of an SDI. Thus, for the most part, these initiatives will almost certainly never have a separate strong coastal or marine focus. The only exception at Regional SDI level is that of INSPIRE, justified partly by the legal data reporting requirements of the European Water Framework Directive, the latter of which focuses very heavily and in great detail on coastal and near-shore waters. As more such regional (non-European) environmentally focused initiatives evolve, we can expect a similar emphasis on the need to arise for wider access to and use of coastal and marine spatial data in regard to environmental monitoring.

At the global level, some of the most important work in regard to major elements of what comprise an SDI has been implemented within major international environmental monitoring and research programmes, such as those of IOC, the World Meteorological Organization (WMO), International Geosphere-Biosphere Program (IGBP) and others (Longhorn, 2003a).

At national level, where one tends to find more detail in specification of the SDI and a more advanced state of actual implementation of elements of SDI, there is evidence that only the more obvious data needs are being considered, e.g. shoreline, hydrography/bathymetry and (sometimes) cadastre/boundaries. Yet other important elements such as habitat and navigation do not fit well within the primarily land-focused sections of NSDI specifications. This may be due partly to the absence of coastal zone/marine professionals on the SDI development teams or for other reasons, e.g. a national decision that the SDI should focus on only the most basic 'reference data' that is considered commonly needed to underpin broadly based economic activities and governance.

Finally, one should not forget that spatial data is only one facet of an SDI implementation. Important institutional, jurisdictional, data policy and standards/interoperability issues also Figure high on the agenda. These appear to

be the principal focal points for most regional and global SDI initiatives, rather than detailed basic reference data specifications.

1.9 REFERENCES

Bartlett, D.J., 2000, Working on the Frontiers of Science: Applying GIS to the Coastal Zone, In *Marine and Coastal Geographical Information Systems* edited by Wright, D. and Bartlett, D. (London: Taylor & Francis).

Butler, M.J.A., LeBlanc, C. and Stanley, J.M., 1998, *Inland Waters, Coastal and Ocean Information Network (ICOIN)*, Report submitted to the Canadian Hydrographic Service, Department of Fisheries and Oceans (DFO) by MRMS, Inc. pp. 85.

CCMC, 1999, *Draft Concept Outline Marine Geospatial Data Infrastructure (MGDI)*, Canadian Centre for Marine Communications, http://cgdi.gc.ca/english/geospatial/MGDI/pdf/mgdi.pdf (May 22, 2003).

Chopin, T. and Costain, K., 2001, *Beyond 2000: An Agenda for Integrated Coastal Management Development*, Post-Conference Report of the Coastal Zone Canada 2000 International Conference, Saint John, New Brunswick, September 17-22 2000. (Dartmouth, Nova Scotia: Coastal Zone Canada Association).

Clinton, W., 1994, *Executive Order 12906, April 13, 1994*, Federal Register, vol. 59, no. 720, pp. 1771–1774. (Washington, D.C.: Government Printing Office).

CRMSW, 2000, *Clean Water Action Plan: Coastal Research and Monitoring Strategy, September 2000*, Coastal Research and Monitoring Strategy Workgroup (Washington, D.C.: US Environmental Protection Agency). http://www.cleanwater.gov/coastalresearch/H2Ofin.pdf (May 19, 2003).

Davis, D. *et al.*, 2002, *Using XML Technology for Data and System Metadata for the MBARI Ocean Observing System (MOOS)*, (Moss Landing, CA, USA: Monterey Bay Aquarium Research Institute). http://www.mbari.org/ (10 October 2003).

DFO, 2001, *Marine User Requirements for Geospatial Data: Summary 2001*, Geospatial Projects Integration Office, (Ottawa, Canada: Department of Fisheries & Oceans), http://www.geoconnections.org/english/MGDI/Key_Docs/Marine_User_Requirements.pdf (May 20, 2003).

European Commission, 2000, *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*, (Luxembourg: Office of Official Publications of the European Communities).

European Commission, 2002, *Proposal for a European Parliament and Council Directive on the Re-use and Commercial Exploitation of Public Sector Document. June 5, 2002, COM(2002) 207*, (Brussels: European Commission), 20 p.

European Communities, 2002, *Guidance Document on Implementing the GIS Elements of the WFD, EUR 20544 EN*, edited by Jürgen Vogt, (Ispra, Italy: DG Joint Research Centre). http://agrienv.jrc.it/publications/pdfs/GIS-GD.pdf (May 19, 2003).

Evans, J. (editor), 2001, The New Digital Earth Reference Model, NASA Digital Earth Office, http://www.digitalearth.gov/derm/v05/index.html (May 22, 2003).

FGDC, 1998, *Shoreline Metadata Profile of the Content Standards for Digital Geospatial Metadata, FGDC-STD-001.2-2001*, Marine and Coastal Spatial Data Subcommittee, (Reston, VA: Federal Geographic Data Committee), http://www.csc.noaa.gov/metadata/sprofile.pdf (May 20, 2003).

FGDC, 2000a, *National Hydrography Data Content Standard for Inland and Coastal Waterways - Public Review Draft - January 2000*, Marine and Coastal Spatial Data Subcommittee, (Reston, VA: Federal Geographic Data Committee), http://www.fgdc.gov/standards/documents/standards/hydro/HydroStd_pr_draft.pdf (May 19, 2003).

FGDC, 2000b, *Geospatial Positioning Accuracy Standards Part 5: Standards for Nautical Charting Hydrographic Surveys - Public Review Draft - November 2000*, (Reston, VA: Federal Geographic Data Committee), http://www.csc.noaa.gov/cgibin/goodbye.cgi?url=http://www.fgdc.gov/standard/status/sub1_4.html (May 22, 2003).

FGDC, 2002, *Charter for Subcommittee on Marine and Coastal Spatial Data*, NOAA Coastal Services Center, http://www.csc.noaa.gov/fgdc_bsc/overview/charter.htm (May 23, 2003).

GeoConnections, 2003, *Marine Advisory Network Node Terms of Reference*, (Ottawa, Canada: GeoConnections Secretariat), http://www.geoconnections.org/english/MGDI/index.html (May 22, 2003).

Gillespie, R., Butler, M., Anderson, N., Kucera, H. and LeBlanc, C., 2000, MGDI: An Information Infrastructure to Support Integrated Coastal Management in Canada, *GeoCoast*, **1**, pp. 15-24, http://www.theukcoastalzone.com/geocoast/Volume1/Gillespie.pdf (May 22, 2003).

GSDI, 2001, *Developing Spatial Data Infrastructures: The SDI Cookbook version 1.1*, edited by D. Nebert, Gloal Spatial Data Infrastructure Technical Working Group Chair, http://www.gsdi.org/pubs/cookbook/cookbook0515.pdf (May 22,2003)

IHO, 1996, *IHO Transfer Standard for Digital Hydrographic Data Edition 3.0 - March 1996. Special Publication No 57*, (Monaco: International Hydrographic Bureau).

INSPIRE, 2002, *Reference Data and Metadata Position Paper*, RDM Working Group edited by Rase, D., Björnsson, A., Probert, M. and Haupt, M-F.

INSPIRE, 2003, *Contribution to the extended impact assessment of INSPIRE*, INSPIRE FDS Working Group and Dr. M. Craglia, Univ. of Sheffield, (Environment Agency for England and Wales, UK).

IOC, 2003a, *The MarineXML Project portal - inaugural meeting summary report*. http://ioc.unesco.org/marinexml/contents.php?id=9 (13 October 2003).

IOC, 2003b, *Marine XML Web portal*. http://ioc.unesco.org/marinexml/ (May 13, 2003).

ISCGM, 2000, *Global Map, version 1.1 Specifications, 16 March 2000, Capetown*, http://www.iscgm.org/html4/index_c5_s1.html#doc13_3787 (May 20, 2003).

ISCGM, 2001, *Resolutions of the 8th ISCGM Meeting, Cartagena, Columbia, 25 May 2001*, http://www.iscgm.org/more-iscgm/reso-8th-meeting.pdf (May 18, 2003).

ISO, 2002, *ISO/TC 211 Geographical information/Geomatics Scope*, International Standards Organisation, Geneva, http://www.isotc211.org/scope.

htm#scope (May 23, 2003).

Labonte, J., Corey, M., and Evangelatos, T., 1998, Canadian Geospatial Data Infrastructure (CGDI)–Geospatial Information for the Knowledge Economy, *Geomatica*, **52**, pp. 214–222. http://www.geoconnections.org/english/publications/General_information/publications_geomatica_cgdi_e..pdf (May 24, 2003).

Lockwood, M. and Fowler, C., 2000, Significance of Coastal and Marine Data within the Context of the United States National Spatial Data Infrastructure. In *Marine and Coastal Geographical Information Systems* edited by Wright, D.J. and Bartlett, D.J. (London: Taylor & Francis).

Longhorn, R, 2002, Global Spatial Data Sharing Frameworks: The Case of the Intergovernmental Oceanographic Commission (IOC). In *Proceedings of GSDI-6, Budapest, Hungary, 16–19 September 2002*. http://www.gsdi.org/ (April 10, 2003).

Longhorn, R., 2003a, A comparison of spatial information access policies of transnational environmental modelling and global climate change programs. In *Geoinformation for European-wide Integration* edited by Benes, T., (Rotterdam: Millpress), pp. 305–313.

Longhorn, R. 2003b, European CZM and the Global Spatial Data Infrastructure Initiative (GSDI). In *Coastal and Marine Geo-Information Systems* edited by Green, D.J. and King, S.D. (Dordrecht, NL: Kluwer Academic Publishers), pp. 543–554.

NOAA, 2001, *Coastal NSDI*, NOAA Coastal Services Center, http://www.csc.noaa.gov/themes/nsdi/ (May 23, 2003).

NOAA, 2003, *FGDC Marine and Coastal Spatial Data Subcommittee 2003 Work Plan*, http://www.csc.noaa.gov/fgdc_bsc/accomp/2003plan.htm (May 22, 2003).

OGC, 2002, *Open GIS Consortium Vision, Mission and Values*, http://www.opengis.org/info/vm.htm (May 22, 2003)

OMB, 2002, *Circular No. A–16 Revised, August 19, 2002*. (Washington, DC: Office of Management and Budget), http://www.whitehouse.gov/omb/circulars/a016/a016_rev.html (May 27, 2003).

PCGIAP, 1998, *PCGIAP Publication Number 1*, http://www.gsi.go.jp/PCGIAP/tech_paprs/apsdi_pub.htm (May 24, 2003).

PCGIAP, 2001, *Working Group 3 report of Group Meeting at Tsukuba, 25th April 2001*, PCGIAP Working Group 3: Cadastre, http://www.gsi.go.jp/PCGIAP/tsukuba/WG3_back.htm (May 22, 2003).

PCGIAP, 2002, *Report of working group 2 meeting, 18 April 2002*, PCGIAP Working Group 2: Fundamental Data, http://www.gsi.go.jp/PCGIAP/brunei/wg2_report.htm (May 22, 2003).

Sliogeris, P. 2002, *An XML Based Marine Data Management Framework*. (Potts Point, NSW, Australia: Australian Oceanographic Data Centre), http://www.aodc.gov.au/ (20 May 2003).

U.S. Commission on Ocean Policy, 2002, *Developing a National Ocean Policy: Mid-term report, September 2002*, http://www.oceancommission.gov.

CHAPTER TWO

Bridging the Land-Sea Divide Through Digital Technologies

Simon Gomm

2.1 INTRODUCTION

There are many different types of users of coastal zone information, from the casual user who may only want to browse, to the sophisticated user who makes frequent use of mapping and demands continuous improvement. These user communities are diverse in the topics they address, covering such areas as Local and Central Government, environmental and economic analysis, and also increasingly leisure use.

A common mapping framework that bridges the land-sea divide allows users to build applications and decision-making tools necessary to promote the shared use of such data throughout all levels of Government, the private and non-profit sectors and academia. A consistent framework also serves to stimulate growth, potentially resulting in significant savings in data collection, enhanced use of data and assist better decision-making.

As well as a physical division, the land-sea divide has also, for many spatial data suppliers, acted as a limit to their area of responsibility, or formed a data product boundary. As a result users wanting to model the diverse aspect of the coastal zone across this divide have had to identify, obtain and combine separate datasets to provide the data coverage they require. The combination process must resolve integration problems resulting from the differing projections, scale of capture and other specification issues of the source datasets. This process can be time consuming, result in inconsistent data and can cause a hindrance to the management of a particularly sensitive environmental zone.

This chapter will look at the technical issues involved with the integration of data across the land-sea divide and identify means for resolving these. Examples within this chapter have been drawn from the work done by Ordnance Survey of Great Britain, United Kingdom Hydrographic Office and the British Geological Survey on integrated coastal zone mapping project (ICZMap) (Gomm, 2001).

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

2.2 DATA SPECIFICATIONS

At the commencement of a project the user will need to have assessed the project's spatial and non-spatial requirements, and these will provide the key criteria for defining and selecting suitable data. Typically such criteria will include: physical extent of the project area, data content, attribution, positional accuracy, spatial resolution, currency, projection, datum and transfer format. These criteria are all discussed in more detail below.

2.2.1 Spatial Extent

The extent of the area for which data are required needs to be defined clearly in the coordinate system to be used in the project. At a minimum the extent should be defined by a bounding rectangle using the x and y coordinates of 2 diagonal corners. Ideally the extent should be delineated by a bounding polygon at an appropriate spatial resolution. A bounding polygon will allow for better selection of relevant information and itself provide a tool for analysis within the project. In defining the extent a sufficient margin should be included to allow for inclusion of features which may have an influence on the application.

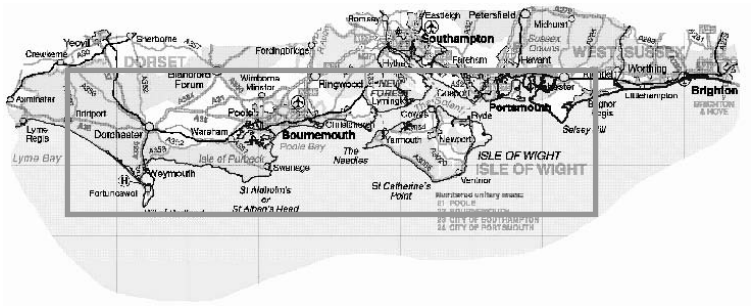


Figure 2.1 Bounding Rectangle. © Crown Copyright 2004. All rights reserved.

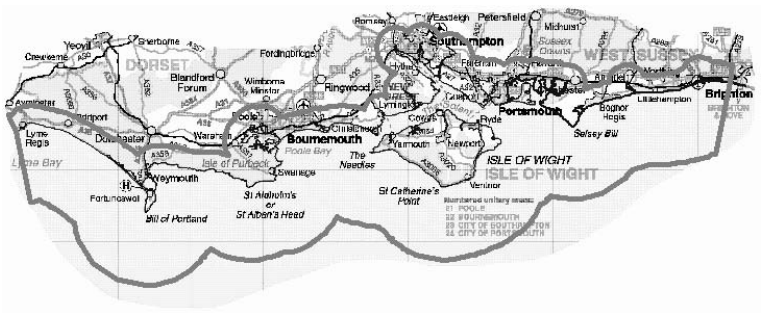


Figure 2.2 Bounding Polygon. © Crown Copyright 2004. All rights reserved.

In Figure 2.1 the extent of the ICZMap coastal zone project is defined as a simple bounding rectangle and in Figure 2.2 as a polygon defined by buffering the coastline 20km offshore and 5km inland. In practice, the latter was more appropriate for the application and datasets of the ICZMap project, with its emphasis on the processes and dynamics of the land-sea interface.

Whilst it is normal to consider defining extent only in two dimensions, it is worth noting that there will be applications where height/depth extents, along with the temporal aspect may need to be specified.

2.2.2 Data Content

The use to which data are to be put will define the features that are required. At the simplest level this may merely be raster imagery for use as backdrop mapping within an application. At a more complex level, where the data are required to form part of the analysis, the user will need to consider what features and object classes are required. In defining these requirements there will be some features that fall uniquely on the land or in the sea, but there will also be others that by their nature bridge this land-sea divide.

2.2.3 Attribution

Attribution of features within a dataset can be at a number of different levels. As a minimum this could take the form of feature coding, allowing selection of required features for analysis and symbolisation. Beyond this, additional attribution concerning the real-world properties of the feature and how it was captured will increase the versatility of the data. These attributes will vary for different features within a dataset depending on the real-world object they represent, and the uses for which the dataset was intended.

2.2.4 Positional Accuracy and Spatial Resolution

An appreciation of the positional accuracy requirements of a dataset is important to ensure that the data are used in an appropriate way with other datasets, and to put results of any analysis in to the correct context. The positional accuracy of spatial data can be expressed both in terms of its absolute accuracy and its relative accuracy.

Absolute accuracy is a measure to which a coordinated position in the dataset corresponds to the true position of the real world feature it represents. Relative accuracy expresses the positional accuracy between points in a dataset, and is a comparison of the distance between features in a dataset with the real world distance. Datasets with a high relative accuracy but low positional accuracy may indicate a systematic shift in the data with respect to the coordinate system.

The spatial resolution represents the coordinate precision to which data are stored in the dataset, and affects the maximum achievable accuracy for a dataset,

e.g. if the spatial resolution is only 10m then this could affect the absolute accuracy of a point by up to 7m.

2.2.5 Currency

Currency is sometimes overlooked as an aspect of a dataset's specification. Attribution should ideally allow for the recording of temporal information at feature level. This can include creation date, capture date and modification date (with nature of modification). Occasionally, temporal information will be limited to the date of the dataset's last update. The temporal information is important both from an analysis point of view, but also for the initial selection of data for the project. For example, the ability to select the position of the top and bottom of coastal slopes for a given year can allow predictive analysis of coastal erosion processes to be studied.

2.2.6 Projections and Datum

Coordinates within a dataset will be relative to a given projection and datum. Map projections attempt to represent the curved surface of the Earth on a flat plane. All projections are approximations and some will better represent large areas of the world, albeit with large distortions, while others are more suitable for small specific areas with much smaller distortions.

A geodetic datum or spheroid is a mathematical approximation of the surface of the Earth, which itself is an imperfect sphere. Numerous geodetic datums have been calculated over the years, some suitable for global applications and others calculated to minimise errors on a country-by-country basis.

Height datum represents a base level from which elevations are measured and each information source may have its own. Datasets on land will often share a common height datum (e.g. use of Ordnance Survey Newlyn Datum in Great Britain); however different datums will usually be used for marine datasets. Frequently, with marine datasets having their origin from navigation charts, depth values will be typically based on local lowest astronomic tide values.

For a project it is important to determine what projection and datum are most suitable for the application, and to ensure that you are aware of what the projections and datum of the source datasets are.

2.2.7 Data Transfer Format

Datasets are transferred between media using a chosen transfer format. Whilst there is no single common transfer format there are a number of commercial ones, such as Autodesk-AutoCAD DXF, ESRI Shapefile and MapInfo TAB and MIF/MID formats, which are becoming more widely adopted as *de facto* standards. Such formats can be limited as they are primarily intended for transfer between users of the same software. When read by other software, information may be lost due to data model differences.

Ideally a neutral, non-software-specific, transfer format is needed. Many attempts have been made to try to implement these, one of the most recent examples being GML (Geography Markup Language) promoted by the OpenGIS consortium as a development of the more widely used XML Web document markup language. However, in practice, the software being used for a project may act as a limiting factor in what formats it can and cannot read.

2.2.8 Metadata

Much of the information describing a dataset should be included in its metadata if present. As well as accompanying a dataset, metadata are now frequently being held on readily accessible databases, allowing users to identify datasets suitable to their requirements. The international standard for digital geospatial metadata (ISO 19115) is now being adopted by many national bodies, such as US Federal Geographic Data Committee (FGDC) for its Content Standard for Digital Geospatial Metadata (CSDGM) (http://www.fgdc.gov/metadata/metadata.html/), and the UK's Association for Geographic Information (AGI) with its GIGateway project (http://www.gigateway.org.uk/default.asp/).

2.3 DATA SELECTION

The properties of datasets that have been outlined above cover some of the points that need to be considered in establishing a project and specifying data requirements. Metadata services provide a tool for locating data, but they will not tell you if they are best suited for your purposes. It is unlikely that any dataset will fully match your requirements, but what should be considered is the ability to integrate and modify them to satisfy your needs. Typically, more flexible datasets will have a richness of feature classes and attribution, and good positional accuracy, but these will come at an increase in cost and data volumes.

2.4 DATA INTEGRATION

It is unlikely that a single dataset will satisfy the full needs of a project. This is particularly true when modelling the coastal zone, where there is likely to be one source for the land, another for the sea and potentially other subsidiary datasets straddling both. In these cases there will inevitably be some data integration issues. Typical problems and potential solutions are discussed here.

2.4.1 Differences in scale

Datasets which have been captured from graphic products can be said to have a nominal scale at which they were intended to be used. In these datasets, differences in scale will show themselves in the positional accuracy and spatial resolution, as well as in the features and attributes that are present e.g. field

boundaries may be shown at a scale of 1:25,000 but not at 1:50,000. At smaller scales data will also normally have been generalised, simplifying the geometry of the features. These factors are not a problem in themselves, as long as the data meet application requirements, but problems do nonetheless occur when trying to join together data captured at different scales.

The following example illustrates the relevance of this to the coast, using datasets from the Ordnance Survey of Great Britain and from the UK Hydrographic Office respectively.

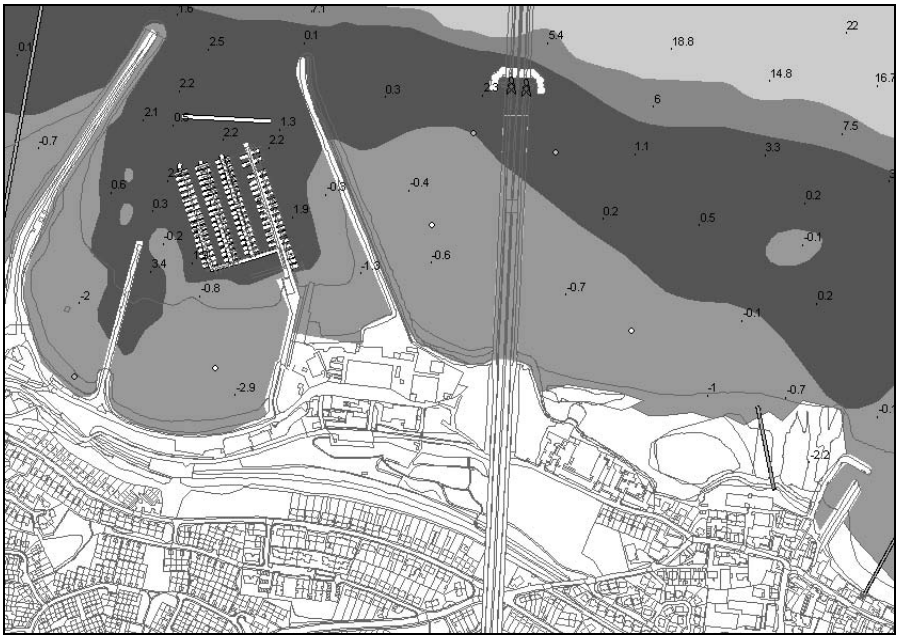


Figure 2.3 Overlay of hydrographic and terrestrial datasets. © Crown Copyright 2004. All rights reserved.

In Figure 2.3 the landward data have been captured at a scale of 1:2,500 and on the seaward side at a scale of 1:25,000 (this being the highest resolution data available for the area). The result of this is a disparity in the features common to both zones, and a greater density of detail on the land compared with the sea. In such situations a choice can sometimes be made as to which feature is most useful to the application, and the software system functionality used to remove or suppress the other. In a similar way features that are superfluous to the application can also be removed or suppressed by the software, provided that features can be distinguished from each other by their attributes.

If the geometry of the large-scale data is too complex, then filtering this using line-simplification or other generalisation algorithms can be considered. Within a project for a user this is probably only appropriate to features where line vertices can be removed whilst keeping the basic shape of the feature.

2.4.2 Projection

Where the datasets to be used are in different projections a choice will have to be made as to which to use, based on the application and the type of output required. Transformation facilities, with parameters to convert between most common map projections, are supplied with mainstream GIS packages. For projections not included these can also be transformed, provided the parameters defining the projection are known.

How the transformation is applied depends on the functionality of the software being used. Frequently this will allow transformation of data coordinates to be done either as a permanent process, storing new coordinates for the features in the new projection, or as a real-time transformation when the dataset is required for display or analysis.

2.4.3 Currency

Where different datasets abut each other or overlap, the same real-world features may be represented in both datasets. In such cases a decision will normally need to be made as to which features to retain. Such selection may be done on the currency of the data i.e. the date of capture of the feature. However, this may not give the best quality in terms of fidelity or resolution. In these situations the final decision as to whether the feature gets included will be based on the specific needs of the application.

2.4.4 Accuracy

As with currency, an appreciation of the relative positional accuracies of datasets that are to be integrated will ensure that data are combined in an appropriate manner. Where features representing the same real-world objects exist in two or more datasets, the positional accuracy will offer one means of selecting which to keep. In some instances it will not be appropriate just to take the data with the highest positional accuracy. For some applications the more accurate data may take up more space, be slower to manipulate, and not enhance the analysis, so ultimately it is up to the user to decide what best fits their needs.

2.4.5 Data Overlap

Many of the issues discussed above cover situations where data in different datasets overlap spatially and represent the same real world objects. It is assumed that in such situations a decision will need to be made as to which features to keep, and that this selection will be driven by the requirements of the application. In the case of a marine and a terrestrial dataset bounded by a coastline, due to the different representations of the coastline in the two datasets, features may overlap (as shown in Figure 2.4).

These overlaps can be treated in a number of ways:

1) Ignore them, accepting that both overlapping features are valid in the context of the source datasets and their specifications. This may not be a real solution when the application requires a single seamless layer of information with no duplication of common features.

2) Spatially intersect overlapping features and retain all the features that result.

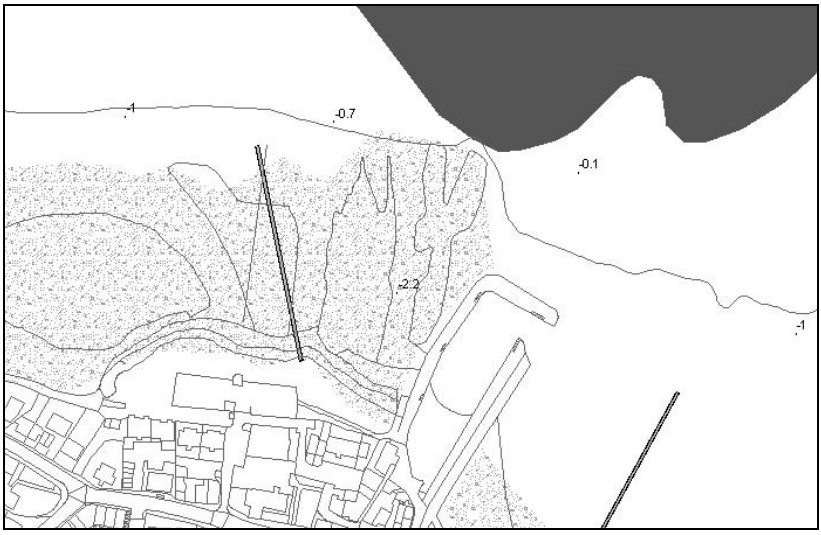


Figure 2.4 Data overlap example. © Crown Copyright 2004. All rights reserved.

Features will be split by others that overlap them and the resulting features will need to retain the attributes of both source datasets to be flexible. This is the most flexible solution but has a potential storage overhead of having to store the two sets of attributes.

3) Spatially intersect overlapping features and retain only one set of features. This will ensure no overlap, but some features will be truncated or split and as a result may present problems when used for analysis.

4) Edit datasets together using a single boundary between them. This can be labour-intensive, but will give the best solution. However, if datasets are updated independently then this work will need to be repeated to maintain currency of the combined dataset.

2.4.6 Height Datum Correction

2D plan data need not necessarily be physically combined into a common dataset to be of use. This is not true in the case of 3-dimensional data. Terrestrial and marine elevation data can be displayed graphically in 2D, but where any use of terrain modelling is required, such as for an analysis of the inter-tidal zone, a combined height model needs to be created. As mentioned previously, the terrestrial and marine elevation data will frequently be relative to different datums. As a result, prior to combining them into a single height model, one or more of the datasets need to be corrected to the other's datum as well as any re-projection of the data onto the other's coordinate system (Milbert and Hess, 2001).

In the simpler cases, correction of a datum can be achieved by a single difference applied to all elevation values in a dataset, but more frequently the difference will vary across an area. In the case of UK Hydrographic Office (UKHO) data, bathymetric values are based on lowest astronomic tides for navigation purposes, as computed for specific locations on each chart. The relationships between these different datums are shown in Figure 2.5.

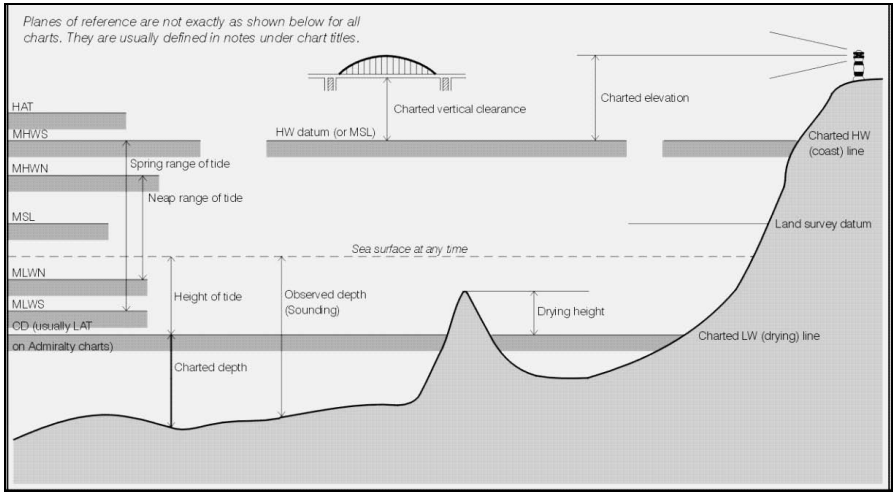


Figure 2.5 Datum Relationships. © Crown Copyright 2004. All rights reserved.

There is a complex and varying relationship between Ordnance Survey of Great Britain's single, standard Newlyn height datum (which applies to all terrestrial mapping by the OS) and the values used in Hydrographic charting, with the latter differing widely along the coast, due to the varying tidal conditions that occur. To convert the UKHO bathymetry values to Newlyn datum, a surface of differences has to be created so that correction value can be calculated for all points (Capstick & Whitfield, 2003). Once a surface of differences has been created this can be used to interpolate values for all bathymetry points. The corrected points can then be used with the terrestrial elevation data to generate an integrated land-sea model for display and analysis.

Figure 2.6 shows a representation of integrated elevation data for the Isle of Wight, UK, with the bathymetry information coming from UKHO sounding values and the terrestrial elevation from Ordnance Survey Land-Form PROFILE contours and spot heights.

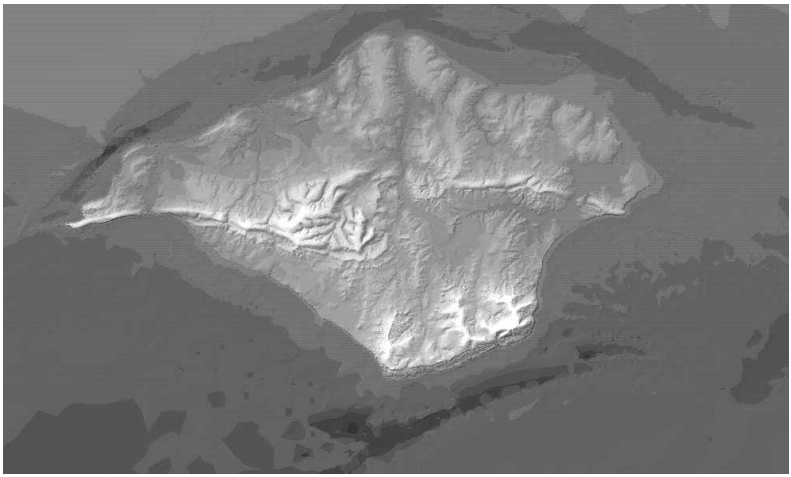


Figure 2.6 Integrated Height Model. © Crown Copyright 2004. All rights reserved.

2.5 CONCLUSIONS

Whilst there may be limited availability of single datasets covering the land-sea divide, digital technologies offer the capabilities to combine and resolve available adjacent datasets to satisfy diverse applications for the coastal zone. In satisfying an application, the user must have a clear specification for their data requirements so as to be able to assess the suitability of candidate datasets, and the issues that are likely to be encountered in their integration. Although not exhaustive, this chapter has attempted to highlight the main areas to be addressed in the selection and integration of spatial datasets for the coastal zone.

2.6 REFERENCES

Capstick, D. and Whitfield, M., 2003, Height integration at the coastal zone. In *Proceedings of the GIS Research UK, 11th Annual Conference 2003*.

Gomm, S., 2001, Integrated mapping of the marine and coastal zone. In *Proceedings of the GIS Research UK, 9th Annual Conference 2001*.

Milbert, D.G. and Hess, K.W., 2001, Combination of topography and bathymetry through application of calibrated vertical datum transformations in the Tamapa Bay region. In *Proceedings of the 2nd Biennial Coastal GeoTools Conference 2001*.

CHAPTER THREE

A Comparative Study of Shoreline Mapping Techniques

Ron Li, Kaichang Di, and Ruijin Ma

3.1 INTRODUCTION

The shoreline is a unique and complex feature and has been recognized as such by the International Geographic Data Committee (IGDC), which designates it as one of 27 key distinct features on the earth's surface. A shoreline is defined as the line of contact between land and a body of water. Although it is easy to define, the shoreline is difficult to capture because of the natural variability of water levels. An accepted method of capturing this line of interface is to determine the *tide-coordinated shoreline*, which is the shoreline extracted from a specific tide water level. NOAA uses Mean Lower Low Water (MLLW) and Mean High Water (MHW) in this way to map shorelines that can be georeferenced. Both the MLLW and MHW are calculated from averages over a period of 18.6 lunar years.

Like the shoreline, shoreline mapping techniques are changing and being improved upon. At present, photogrammetric techniques are employed to map the tide-coordinated shoreline from aerial photographs that are taken when the water level reaches the desired level. Aerial photographs taken at these water levels are more expensive to obtain than remote sensing (RS) imagery. With the development of remote sensing technology, satellites can capture high-resolution imagery with the capability of producing stereo imagery: one such source is the IKONOS satellite images. The question is: is it possible to use RS imagery to map the shoreline so that we can reduce the costs and improve mapping efficiency and accuracy?

During our study we extracted shorelines from aerial photos, simulated and actual IKONOS imagery, and the intersection between a Coastal Terrain Model (CTM) and the water level. We estimated the accuracies of these shorelines and analyzed the potential of these techniques for practical shoreline mapping by comparing the extracted dataset with the shorelines from the United States Geological Survey (USGS) topological maps and National Oceanic and Atmospheric Administration (NOAA) nautical charts.

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

3.2 SHORELINE EXTRACTION

To evaluate the new shoreline mapping techniques it is necessary to compare the results with a current shoreline product. In our study, the shoreline extracted from the tide-coordinated aerial photos was used as the reference shoreline and the other shorelines were compared with it to assess their relative accuracies. Shorelines from USGS Digital Line Graphs (DLG) and a NOAA Nautical Chart were digitized to facilitate the evaluation of the new results. The extraction methods used are discussed in the following sections.

3.2.1 Shoreline from aerial photos

A shoreline was digitized from tide-coordinated aerial orthophotos. This product was chosen as the baseline in our analysis due to its high accuracy. The aerial photos were commissioned by NOAA in 1997 when the water surface reached the specified level (MLLW). The photo scale is about 1:20,000 and the pixel size is about 0.6 meters on the ground. This shoreline is tide-coordinated to the mean lower low water level. A Global Positioning System (GPS) survey was carried out in 2000 to obtain control points for the bundle adjustment. Eight control points were used in the adjustment. The Root Mean Square (RMS) numbers indicate a good precision in this adjustment: 0.106, 0.107 and 0.073 meters in the X, Y and Z directions, respectively. Based on the bundle adjustment results, Digital Terrain Models (DTMs) and orthophotos were generated in sequence. Then, the orthophoto mosaic was created with a resolution of 1 meter. The shoreline was digitized manually from this orthophoto mosaic. The task was accomplished using ERDAS software.

3.2.2 Shorelines from simulated and actual IKONOS images

IKONOS imagery has two resolutions, 1-meter panchromatic and 4-meter multispectral. We first obtained 4-meter multispectral geo-referenced imagery for our project. The stereo 1-meter images were still pending when we produced the shorelines in this paper. To study the potential of 1-meter stereo imagery in shoreline mapping, we simulated the imaging procedure and obtained the simulated IKONOS 1-meter stereo imagery using rational functions (RFs). The shoreline extracted from aerial orthophotos was used for back-projecting to the simulated IKONOS raw backward- and forward-looking images. The resulting shorelines on IKONOS raw images show that RFs work very well. This indicates that we can use RFs to extract high accuracy 3-D shorelines from real IKONOS 1-meter imagery provided that we can find well-matched conjugate points on raw IKONOS images along the shorelines. To improve the accuracy of 4-meter IKONOS real data, 8 control points were used to perform planar polynomial geo-correction. The RMS in the X and Y directions are 1.048 meters and 1.1028 meters respectively. A shoreline was extracted from the refined IKONOS image.

3.2.3 Shoreline from the intersection between a CTM and the water level

This idea comes from the definition of a shoreline as the contact or intersection line between the water level and the land. We tried to obtain the land model and water surface model (WSM) and then compute the intersection line of these two models. This method is intuitive to people's perception of shoreline (See the profile in Figure 3.1 and the colour insert following page 164). The CTM depicts the elevation and bathymetry in the coastal area. The water surface model is obtained from the Great Lakes Forecasting System at The Ohio State University (OSU). To extract the shoreline, these two models were overlaid and the WSM was subtracted from the CTM to get an output layer. Theoretically, the shoreline should be the line with the zero value in the output layer. In practice, the zero pixels do not form a line due to random errors in these two data models introduced from data acquisition and sampling. We chose a small value range for the study and obtained a shoreline strip. We then took one side edge of the strip as the shoreline for our study. We employed this method because the strip is very thin and it is very small compared to the error of the CTM.

3.2.4 Shorelines from other sources

Besides the above shorelines, four other shorelines were used for the analysis in our study. Among them, two digital shorelines were provided by the Ohio Department of Natural Resources (ODNR), the third one was extracted from a USGS DLG and the fourth one was digitized from a NOAA Nautical Chart.

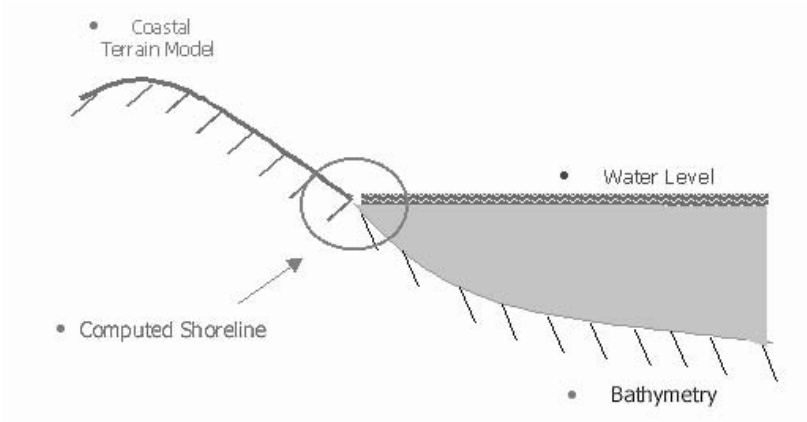


Figure 3.1 Generation of Digital Shorelines

3.3 ESTIMATION OF THE ACCURACY OF THE SHORELINES

The accuracy of these shorelines was analyzed in our study by considering the data sources and the extraction methods. More extensive work will be done in the future to further investigate this important issue.

From the bundle adjustment and conjugate point matching, the DTM standard deviation derived from the aerial photographs is estimated to be 2.1m in X and Y coordinates. Considering the error introduced by resampling from the aerial photos of a 0.6m resolution, the 1-meter orthophoto has an estimated standard deviation of about 2.1m. The identification error of the shoreline from the orthophoto may introduce 1.5 pixels, that is 1.5m, considering those situations where the water and land separation is very hard to define. Therefore, the shoreline from the aerial orthophoto has an estimated standard deviation of about 2.6m if we apply a simple model of error propagation.

The accuracy of the shoreline derived from 1-meter simulated IKONOS imagery should be about 2-4m (Zhou and Li, 2000), considering the fact that the accuracy of 3D ground points reaches 2-3m with GCPs and the accuracy of identifying and locating conjugate shoreline points is about 1.5 pixels (1-2m). The 4-meter IKONOS images have an accuracy of 23.6m (http://www.computamaps.com/ikonos_tech.html). After polynomial georectification using 8 GPS control points, the difference between the GPS and computed coordinates at the control points is about 6m. An optimistic estimation of the shoreline accuracy derived from the 4-meter IKONOS images in this specific case is about 8.5m.

The accuracy of the shoreline derived from a CTM and the water level is affected by the accuracy of the CTM and the accuracy of the water level. The CTM comes from the DTM generated in turn from the aerial photos and the bathymetry. The DTM has an accuracy of 2.1m and the bathymetry has an estimated accuracy of 40m. When we merged these two data sets, we select the DTM grid points in the overlapping area. If there is gap between the two data sets, we interpolate the elevations using weights of 2/3 from DTM and 1/3 from bathymetry where the standard deviation of the CTM is about 13.4m. Overall the accuracy of the CTM is from 2.1m to 13.4m for the area where the water surface model intersects with the CTM. The water level data is accurate to several centimeters. Taking 5 degrees to be the coastal slope in worse cases and 5cm the water level error, we can estimate the horizontal accuracy of the intersected shoreline caused by the water level error to be about 0.6m. Therefore, the final accuracy of the digital shoreline is about 13.4m with the CTM contributing the greatest source of error.

The NOAA T-sheets have large and medium scales: from 1:5,000 to 1:40,000. Taking 0.5 mm on the map as error source, the accuracy of the digitized shoreline from the T-sheet is about 2.5m to 20m. The shoreline digitized from the USGS DLG (1:24,000) should have an accuracy of within 12m. The ODNR map scale is 1:12,000. Again, using 0.5mm as the shoreline digitizing error, the estimated error of the shoreline is about 6m.

Table 3.1 Estimated accuracy of the shorelines derived from various sources

Shoreline	Estimated standard deviation
T-sheet	2.5m to 20m depending on scale
USGS DLG	12m (1:24,000)
ODNR map	6m (1:12,000)
Orthophoto	2.6 meters
CTM and water level	2m-13m dep. on CTM quality
IKONOS 1-meter simulated image	2-4 meters
IKONOS 4-meter image	8.5 meters

3.4 DIFFERENCE AND SHORELINE CHANGE ANALYSIS

The shorelines were acquired at different times and differences can be seen in the results (Plate 3.1). There are two possible interpretations of the shoreline differences. One is that the shoreline indeed changed in the real world. The other possibility is that the differences were introduced as shoreline mapping errors.

From our analysis of the accuracies above, the shoreline from the aerial orthophotos has the highest accuracy, and so we used it as the baseline for the difference analysis. The analysis was performed in raster format because of the efficiency of this method. The first task was to convert the vector shorelines into raster shorelines. The resulting pixel size was 2.5 meters. The procedure is outlined in Figure 3.2.

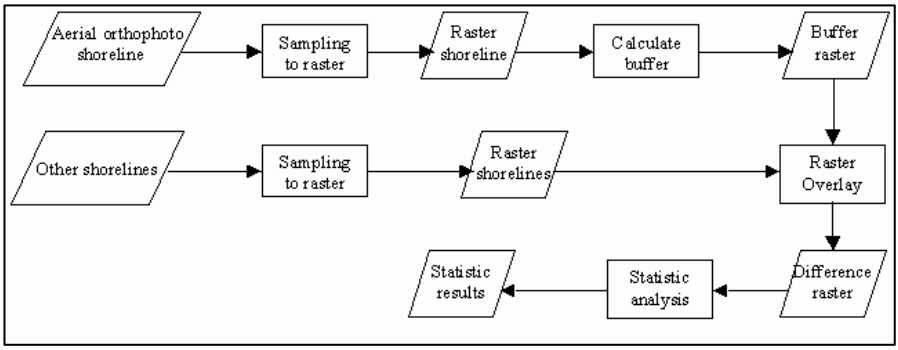


Figure 3.2 Difference Analysis Procedure

This analysis involved comparing every point on the test shorelines with the reference shoreline, which was derived from the aerial orthophoto, in a perpendicular direction toward the reference shoreline. In the buffer zone of the reference shoreline we created pixels whose value represented the distance of each

pixel to the reference shoreline. By overlaying other raster shorelines with the buffer image we were able to discern and quantify differences between the shorelines and the reference shoreline. For example, for the shoreline from the simulated IKONOS images, each point on the shoreline has a value of the shortest distance to the reference shoreline. Finally, we calculated some statistic indicators including RMS of the differences for the shorelines derived from the 4-m IKONOS images and the digital models (table 3.2).

Table 3.2 Statistics of the differences between derived shorelines and the reference shoreline

	Points used	**Avg. Dif. (pixel)**	**Avg. Dif. (m)**	**Max. Dif. (pixel)**	**Max. Dif. (m)**	**RMS (pixel)**	**RMS (m)**
Shoreline from 4-m IKONOS images	6652	4.02	10.04	16	40.0	4.91	12.27
Digital shoreline	4590	2.34	5.84	9	22.5	2.74	6.84

The accuracies of the shorelines derived from the 4-meter IKONOS images and the aerial orthophoto are 8.5m and 2.6m respectively; as a result, the estimated accuracy of the difference between them should be 8.9m, which is smaller than the RMS value 12.27m in table 3.2. This may indicate the actual shoreline change during the period from July 1997 when the aerial photos for the reference shoreline were taken to October 2000 when the 4-meter IKONOS images were taken. In addition, the IKONOS shoreline is an instantaneous shoreline, instead of a tide-coordinated shoreline. The water level in October was 0.2-0.3 meters lower than in July creating artificial shoreline accretion. We are developing methods that correct instantaneous shorelines to tide-coordinated shorelines and then compare them. For the digital shoreline from the CTM and WSM intersection, the RMS value 6.842m in table 3.2 is smaller than the worse case shoreline error of 13m in table 3.1. The difference can be introduced by two different factors. One is the seasonal water level difference, which can introduce a large difference in a flat area like Sheldon Marsh. Another factor is the difference between the accuracies of DTM and bathymetry in CTM. If the DTM covers the coastal surface up to the MLLW or better (this is the case in our research), the water surface model intersects with the DTM portion of the CTM and results in smaller errors; otherwise, the shoreline would be intersected by the water surface model, and the bathymetric part or the interpolated part of the CTM produces large errors.

Observing the differences between the reference shoreline and earlier shorelines from the maps, they may represent the shoreline changes (erosion) that occurred in long time periods. The earliest shoreline in our study is the shoreline from 1973 ODNR map, 24 years before the aerial photos for the reference shoreline were taken in 1997.

3.5 CONCLUSIONS

From the analysis above, we can draw several conclusions. First, we determined that the method for generation of digital shorelines from CTM and WSM can be used in mapping instantaneous shorelines. The generated shorelines can meet the accuracy requirement for certain user communities. Additionally, this work shows that high-resolution satellite imagery, such as IKONOS imagery, has the potential to become a tool in shoreline mapping and coastal change detection, and that it may reduce mapping costs. Finally, our study confirms that to detect coastal changes, tide-coordinated shorelines must be derived from the instantaneous shorelines so that the amount of shoreline change can be estimated objectively.

3.6 ACKNOWLEDGEMENTS

The research was supported by a Sea Grant – NOAA Partnership program with the collaboration of CSC, NGS and OCS of NOAA. Special thanks go to Dr. Keith Bedford and Dr. Philips Chu for providing the water level data. We appreciate the assistance and co-operation of Dr. David Schwab of the GLERL of NOAA during Ruijing Ma's PEGS Professional Development Project in Ann Arbor, MI. We also appreciate the continuous support and help from Dr. Scudder Mackey of the Lake Erie Geology Group of the ODNR in Sandusky, OH.

3.7 REFERENCES

Cheng, P. and Toutin, T., 2000, Orthorectification of IKONOS Data Using Rational Function. In *Proceeding of ASPRS Annual Convention*, Washington D.C.

Dowman, I., and Dolloff, J.T., 2000, An Evaluation of Rational Functions for Photogrammetric Restitution. *International Archives of Photogrammetry and Remote Sensing*, **33**(B3), pp. 254–266.

Horstmann, O. and Molkenthin, F., 1996, *Advanced Grid Modeling for Coastal and Nearshore Regions*. The International Conference for Hydroinformatics, Zurich.

Li, R., 1998, Potential of High-Resolution Satellite Imagery for National Mapping Products. *Photogrammetric Engineering and Remote Sensing*, **64**(2), pp. 1165–1169.

Li, R. and Zhou, G., 1998, Coastline Mapping and Change Detection Using One-Meter Resolution Satellite Imagery. Research proposal, Dept. of Civil Engineering and Geodetic Science, the Ohio State University.

Li, R., Zhou, G., Gonzalez, A., Liu, J. K., Ma, F. and Felus, Y., 1998, Coastline Mapping and Change Detection Using One-Meter Resolution Satellite Imagery. Project Report submitted to Sea Grant/NOAA, pp. 88.

Liu, Ernie, 1998, *Developing Geographic Information System Applications in Analysis of Responses to Lake Erie Shoreline Change*, MS. thesis, The Ohio State University.

NOAA, 1997, Shoreline Mapping. http://anchor.ncd.noaa.gov/psn/shoreline.html.

Peuquet, D.J. and Duan, N., 1995, An event-based spatio-temporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems*, **9**(1), pp. 7–24.

Raper, J. and Livingstone, D., 1995, Development of a Geomorphological Spatial Model Using Object-oriented Design. *International Journal of Geographical Information Systems*, **9** (4), pp. 359–384.

Schwab, D. J. and Sellers, D. L., 1980, NOAA Data Report ERL GLERL-16, ftp://ftp.glerl.noaa.gov/publications/tech_reports/glerl-016/dr-016.html.

Tao, C.V. and Hu, Y., 2000, Investigation of the Rational Function Model. In *Proceeding of ASPRS Annual Convention*, Washington, D.C.

Wang, Z., 1990, *Principles of Photogrammetry (with Remote Sensing)*. (Wuhan, China: Wuhan Technical University of Surveying and Mapping and Publishing House of Surveying and Mapping).

Worboys, M. F., 1992, A Model for Spatio-temporal Information. In *Proceedings of the 5th International Symposium on Spatial Data Handling*, pp. 602–611.

Worboys, M.F., 1994, Unifying the Spatial and Temporal Components of Spatial Information. Advances in GIS Research In *Proceedings of the 6th International Symposium on Spatial Data Handling*, Edinburgh, pp. 505–17.

Yang, X., 2000, Accuracy of Rational Function Approximation in Photogrammetry. In *Proceeding of ASPRS Annual Convention*, Washington, D.C.

Zhou, G., and Li, R., 2000, Accuracy Evaluation of Ground Points from IKONOS High-Resolution Satellite Imagery. *Photogrammetric Engineering and Remote Sensing*, **66** (9), pp. 1103–1112.

CHAPTER FOUR

Monitoring Coastal Environments Using Remote Sensing and GIS

Paul S.Y. Pan

4.1 INTRODUCTION

This paper concerns the monitoring of the marine and coastal environment in South Wales using state-of-the-art survey techniques and a geographic information system (GIS). One of the most important natural resources in South Wales is its marine aggregate. This resource is vital to the regional economy in that it provides the building industry with that most essential of raw materials, sand and gravel. However there are growing concerns as to the possible effects of the commercial extraction of aggregate on the coastal and marine environment, and a number of environmental monitoring procedures are in place to detect changes. These range from traditional beach profile surveys to state-of-the-art airborne remote sensing techniques. The National Assembly for Wales has pioneered the use of airborne LiDAR (Light Detection and Ranging) for the acquisition of highly detailed topographical data on beaches, and CASI (Compact Airborne Spectrographic Imager), for the determination of the state of the vegetation along part of the coastline. LiDAR is capable of accurately detecting changes in beach levels. The procedure began in 1998 and will continue until at least 2003, giving an unprecedented insight into coastal changes over time.

Another remote sensing technique has also been deployed. Close-range photogrammetry has been used to determine the degree of retreat of unstable sea-cliffs. All the data collected is used to populate a GIS. Data acquired in this way is compared with that from various monitoring procedures carried out previously (aerial photography and beach profiles). A number of advanced techniques have been developed in parallel to the GIS for the interpretation, analysis and visualization of the data. A number of invaluable lessons have been learned.

Apart from the site-specific monitoring procedures, other strategic data sets such as the macro-fauna community distribution, modelled parameters, etc., have also been acquired from a number of sources. Amongst these parameters, the most important of all is the sediment environment. It defines the fuzzy geographical boundaries in which distinctive hydrodynamic regimes operate. A summary on the resources and constraints is generated for each of the sediment environments. These resources and constraints summaries together with the GIS form the basis of a decision-support system for assisting the formation of policy for the management

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

of the marine resource. The findings will shape future decisions about the sustainable use of the marine resource in South Wales.

4.2 BACKGROUND

The marine and coastal environment is important to Wales. About 75% of the length of the country is coastal, allowing the waters of the Irish Sea and the Bristol Channel to lap its shores (Figure 4.1 and the colour insert following page 164). South Wales lies adjacent to the Bristol Channel, a body of water largely unaltered in its current dimensions since its beginnings as a marine transgression in the early Holocene. (The Holocene, or post-glacial epoch, covers the period for the end of the Pleistocene about 10,000 years ago to the present day). Here the coastal areas are not only characterized by a number of urban and industrial centres, such as Newport, the capital city Cardiff, Bridgend, Port Talbot and Swansea, but also by many cherished areas of special landscape and nature conservation interest, such as: the Gwent Levels; the Kenfig National Nature Reserve and candidate Special Area of Conservation (cSAC) designated under the European Habitats Directive; the Gower Area of Outstanding Natural Beauty (AONB) – the first to be designated in the United Kingdom; and the Glamorgan Heritage Coast (Figure 4.2 and colour insert). Together, and for different reasons, these coastal areas attract thousands of visitors every year, providing employment opportunities for some of the local population. One of the unseen resources of the Bristol Channel, however, contributes to the local economy in a different way – by yielding high quality building sand for local industry, an essential prerequisite for many forms of economic activity.

Whilst this economic activity brings undoubted prosperity to South Wales and neighbouring regions in mid Wales and South-West England, many people perceive changes to their familiar coastlines, and in particular, to the sandy beaches. There are growing concerns as to the possible effects of the extraction of marine sand from the Bristol Channel, and not everyone thinks the removal of this resource is acceptable or sustainable.

Generally, in Welsh waters, the extraction of sand from the marine environment by dredging is licensed by the Crown Estate. (The Crown Estate Commission is the representative of the Crown, which, in the UK, constitutes the owner of the seabed out to a 12-mile territorial limit). However, the decision on whether a production license should be granted essentially rests with the Environment Minister at the National Assembly for Wales – a devolved and autonomous arm of central Government that came into being in 1999. With its inception has come a desire to increase outside involvement in policy-making and administration, and to increase transparency and accountability in deciding major issues. The challenge for the Assembly's civil servants is to continue to provide objective advice to Ministers in an ever-evolving social, economic, environmental, cultural and political context. This advice must be based on facts and scientific evidence, together with the specialist and professional judgement of the officers involved. All of the dredging licenses granted in recent years have stringent environmental monitoring conditions attached. It is the scientific data from these monitoring procedures that form the basis of sound advice.

This paper discusses the introduction of a Geographic Information System (GIS) into part of the National Assembly for Wales for the analyses of the monitoring data. It highlights the technique's influence on the monitoring procedures and the way it has helped reshape the environmental monitoring requirements in relation to dredging licenses. By way of illustration, a number of the state-of-the-art procedures currently deployed are examined.

Figure 4.1 Wales and its surrounding areas.

4.3 DATA ANALYSES USING GEOGRAPHIC INFORMATION SYSTEMS

The benefits of modern computerized GIS have been well documented by others, for example, Clark *et al.*, 1991 and Maguire, 1991. GIS was first introduced to the former Welsh Office (now the National Assembly for Wales) in early 1997 for analysis of the monitoring data acquired in respect of the dredging license at Nash Bank. A prototype was developed using ESRI's (Environmental Systems Research Institute Inc.) ArcView GIS. The benefits of the technique over the traditional paper-based reporting were immediately apparent. It:

- provided a stable platform for the integration of disparate data from different sources;
- allowed a large quantity of data to be stored and processed;
- provided a seamless geographical database overcoming the restrictions of traditional map/chart boundaries;

- provided facilities for sophisticated analysis and cross-examination of data; and
- provided advanced facilities for the display and visualization of data to a wider audience.

After a period of evaluation, the technology was adopted for operational use. It would play a key role in re-shaping the environmental monitoring procedures. GIS highlighted the weaknesses of some of the established procedures in terms of both the quality and coverage of the data. The capability of GIS in handling spatial data, in particular, has also presented new opportunities for the introduction and subsequent adaptation of more efficient and cost-effective procedures. The following sections discuss some of these new and innovative environmental monitoring techniques. They are: the assessment of cliff instability using close-range photogrammetry; repeated topographical surveys using Light Detection and Ranging (LiDAR); and habitat mapping of sensitive sites of nature conservation interest using Compact Airborne Spectrographic Imager (CASI).

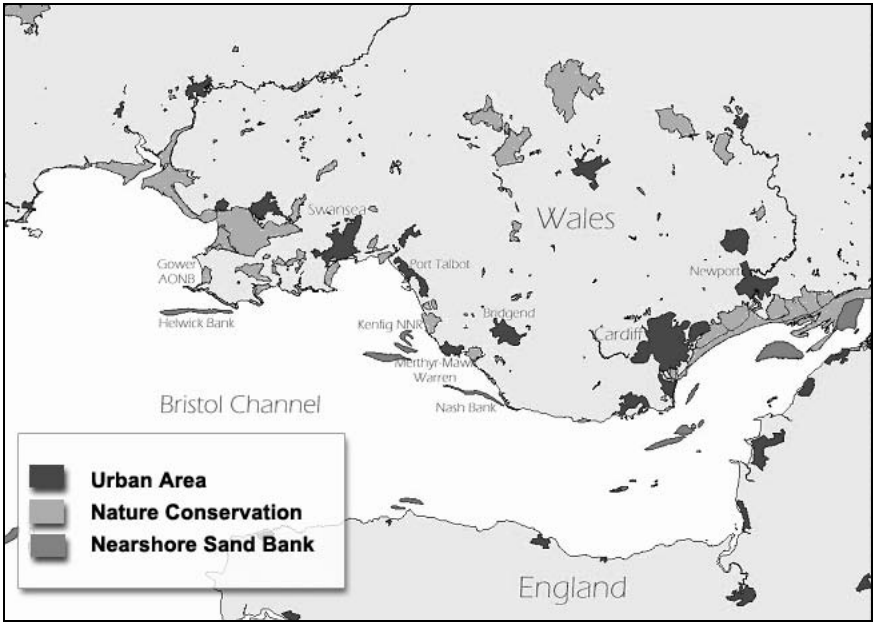


Figure 4.2 South Wales and the Bristol Channel

4.3.1 Assessment of Cliff Instability Using Close-Range Photogrammetry

The Nash Bank lies in the Bristol Channel very close to part of the Glamorgan Heritage Coast. The sandbank contains about 200 million tonnes of material and is a relic feature of the last (or "Devensian") glaciation. The bank is formed ostensibly from pre-existing glacial and glacio-fluvial deposits subsequently

moulded by advancing seas. At low water its eastern end is often exposed above the surface of the sea, and it resides, at its nearest point, only 300 metres from the shore. Its physical orientation and proximity to this area of sensitive coastline means that it acts as a "barrier" to incoming waves, and is therefore an important structure in terms of coast protection. Physical processes such as the movement of sediment in, on and around the feature, as well as human activities, are altering the shape of this sandbank. One way of measuring the "vulnerability" of the nearby coastline to such changes is by careful scientific examination of instability in the highly unstable Blue Lias cliffs that predominate. Extremely accurate measurements of the geometry of a representative 800-metre section of this geologically special part of the Glamorgan Heritage Coast, and including part of the Southerndown Coast Site of Special Scientific Interest (SSSI), have been undertaken since 1997 using close-range photogrammetry. Its application in this context is unique in the United Kingdom.

Annual surveys have been undertaken in August of each year between 1997 and 2000. A further one took place in February 2001.

4.3.1.1 Close-Range Photogrammetry: The Technique

Close-range photogrammetry operates on the same principle as aerial photography. It produces highly detailed geometric data of three-dimensional structures. Data capture involves the use of a specialist metric camera oriented horizontally on a tripod, and usually mounted on top of a theodolite. The lenses are calibrated and their distortion characteristics considered for subsequent data processing. The technique is also known as terrestrial photogrammetry. Both aerial and close-range photogrammetry have been described in detail in Wolf (1985).

Close-range photogrammetry involves photographing the features or structures being surveyed using the metric camera in a known orientation from two positions. The camera positions and their orientation can be established by "traditional" surveying methods, such as using visual intersection combined with theodolite measured distances from known control points. Alternatively, a number of control points can be set up in the area being surveyed and subsequently included in the photogrammetric survey. These control points are used to compute the position and orientation of the two camera positions.

Data is derived from the photographic images by simulating the relative orientation of the two camera positions, and processing involves the generation of a three-dimensional stereo model representing the geometry of the structure. It is relatively labour-intensive because the orientation of the two camera positions is not always parallel, and this means additional computational requirements. However, the generation of the 3D geometric model has been helped by advances in automatic image matching techniques in the past decade.

Close-range photogrammetry has been used widely for surveying architectural structures worldwide. Its main advantage is that it can survey the physical dimension of any structure that does not lend itself readily to traditional surveying techniques. It has also been used extensively by traffic accident investigators for collecting data from the scenes of accidents where a very limited period is available for data capture. It is therefore a technique which is characterized by a short set up time, data capture by relatively straightforward

photographic means and an ability to survey structures and features which are difficult to examine by any other means.

The intention was twofold: to (a), "test" the technique in the "real-world" of unpredictable environmental change, and (b), establish accurate rates of retreat which could be looked at against historical records. The technique is well suited to the task for several reasons. First, it allows data capture by photographic means without the need to get too close to the area being surveyed – an important consideration given the inherent instability of the Blue Lias cliffs! Also, because the Bristol Channel has the second highest tidal range in the world, only limited periods of time are available for data capture. The technique allows surveyors to record data very quickly around mean low water. Finally, the photographic images captured can be manipulated and rectified to provide a geometrically-correct picture called an orthoimage. An orthoimage is a distortion-free map-like image of an original photographic record. It is produced by rectifying the original photographic image using geometric data derived subsequently. It provides a definitive visual record of the conditions of the cliff at the time of the survey. Figure 4.3 (see colour insert following page 164) is illustrative.

Figure 4.3 The rectified image of the study area for the year 2000.

In addition, the distortion-free orthoimages of the study area can be displayed in the GIS environment where they can be used to help make accurate quantitative measurements, while the map-like orthoimages can be used in conjunction with the 3D geometry derived earlier in the process to produce a realistic representation of the cliffs using advanced computer visualisation techniques such as Virtual Reality.

The combined use of orthoimages and 3D geometric models in a GIS environment has provided a robust platform for analysis of changes to the cliffs. Quantitative contours are interpolated at 0.5 metre and 1-metre intervals using information derived from the geometric data sets. These contours, representing the degrees of change, are displayed on the rectified orthoimage. This innovative visualisation technique allows quantitative measurements of yearly change to be related directly to the cliff face and its geology. This may be vital in guiding interpretative analysis of why change occurs and in attempting to understand whether or not rates of recession are increasing.

An example will help illustrate the techniques. A large rock fall occurred in the winter of 2000-01. Figure 4.4 illustrates the extent and magnitude of the event (in the middle right of the picture) from an oblique angle.

A geometric representation of this event can be displayed and visualised in a couple of interesting ways. Firstly, the interpolated surface can be colour-coded in relation to distance between the viewer and the cliffs. Figure 4.5 shows the data for a 200-metre wide area centred on the fall. Darker shades represent features further away from the observer.

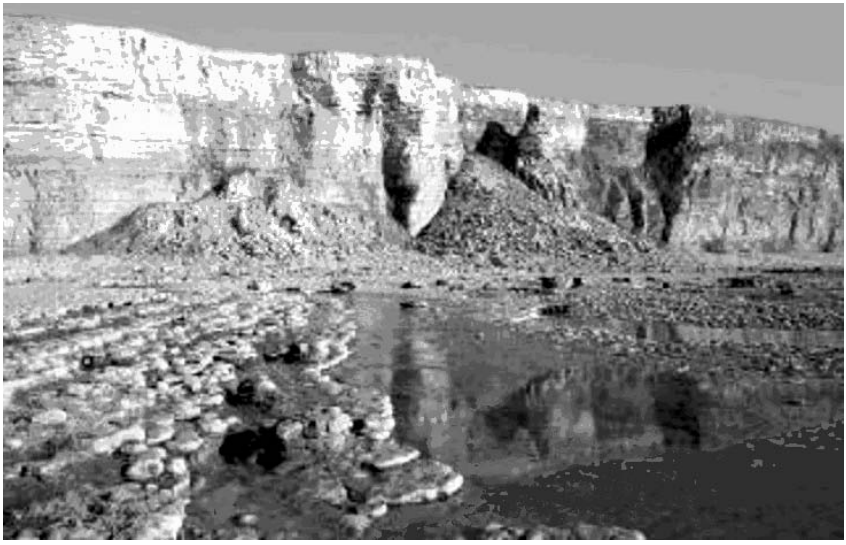


Figure 4.4 The Large rock-fall of winter 2000-01. The cliffs are comprised of Blue Lias rocks and are about 65 metres high. In the foreground is part of a wave-cut platform in the same formation that has been developing for about 7,000 years.

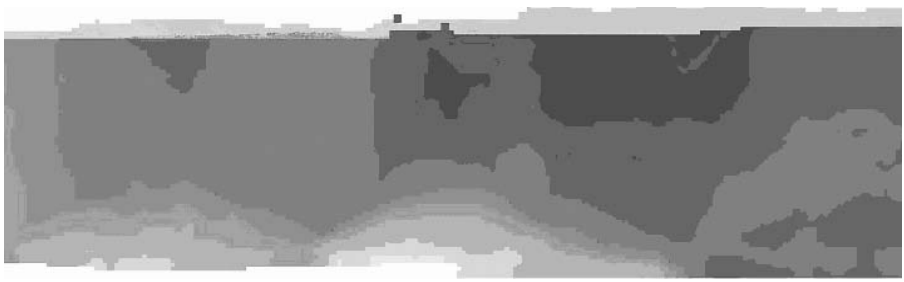


Figure 4.5 Data for a 200-metre stretch of coastline north-west of Cwm Bach

Secondly, the changes brought about by the fall between surveys can be colour-coded to illustrate the changes to the cliffs, and interpolated surfaces may produced using the GIS. Changes at the sampling positions are shaded in reds and blues, depending on the nature of the change recorded. Data collected in February 2001 was compared with that for the summer of 2000. Colour Plate 4.1 (following page 164) shows the results.

The fairly large area of the cliffs represented by deeper reds "failed," causing materials to be deposited as a "classic" debris fan at the base (in deeper blues). The thin area in blue towards the right of the failure represents an "advance." This is, in reality, a rock column which has become detached from the main body of the cliff, and which has been left as a freestanding pillar with a considerable void behind (see Figure 4.4). Future movements will be interesting to monitor!

4.3.2 Topographical Survey Using Light Detection and Ranging

Light Detection and Ranging (LiDAR) is a modern airborne remote sensing technique for surveying topography. It operates on the same basic principle as traditional Radio Detection and Ranging (RADAR), but uses laser as the detection medium. It accurately measures the distance between the instrument and its surrounding environment, and consists of three components: a laser-scanning device, an on-board Differential Global Positioning System (DGPS), and an Inertial Reference System (IRS).

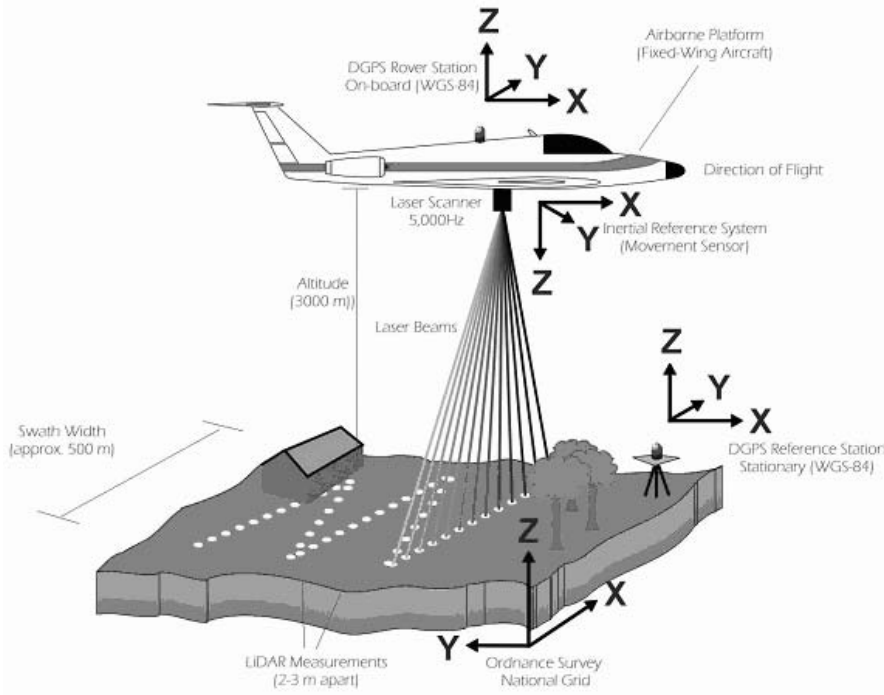


Figure 4.6 A systematic diagram of the three components of airborne LiDAR.

The laser-scanning device is used for the measurement of the distance between the instrument and the topography being surveyed. It emits laser pulses in perpendicular directions to the flight path. By measuring the time required for the laser pulse to return, the distance between the instrument and the topography can be accurately determined. At the same time, an on-board DGPS communicates

with navigation satellites and a stationary ground reference station. It gives the positions of the airborne platform, usually a fixed-wing aircraft, in relation to the earth's surface. The last component of the system is the IRS. This is essentially a motion sensor, and detects minute movements of the aircraft in terms of its yaw, pitch and roll (i.e. movements about its three axes). It corrects the orientation of the laser pulses that can become distorted because of aircraft movement. When the three components work in unison, the surveyed topography is recorded very accurately. Routine field validations of the technique by the Environment Agency have indicated an accuracy of better than 15 cm generally. Ground resolution of data is directly dependent on flying height and speed of the surveying aircraft, as well as the scanning rate of the laser pulses. For example, when flying at 70 knots at an altitude of 3,000 feet LiDAR, with a 5,000 Hz scanning rate, can acquire topographical data with a ground resolution of about 2.5 to 3.5 metres. At the time of writing, the Environment Agency has plans to operate a newer version of LiDAR with a scanning rate of 33,000 Hz. The new instrument should produce topographical data with a 0.5 metre ground resolution! Figure 4.6 shows the three main components of airborne LiDAR.

In South Wales, the coastline is diverse, characterized as it is by bays, inlets, estuaries, beaches, rocky-shores, high cliffs, sand dunes, near-shore sand banks and mud flats. Traditional monitoring on some of the beaches using linear profiles has provided some limited data. In addition, accurate identification of ground features from aerial photographs has been difficult and is relatively expensive. LiDAR, however, has provided a high-value and cost-effective solution. Its capability in capturing very detailed topographical data, without the need to physically visit most of the locations being surveyed, makes it well suited for environmental monitoring purposes, especially for inter-tidal zones where it can survey a large area within a short space of time. One of the few restrictions for the deployment of LiDAR is the tidal window.

LiDAR was first available in the UK in early 1998. The National Assembly for Wales has long realized the potential of the technique and has pioneered its use in Wales. As early as March 1998, in relation to dredging at Helwick Bank, the National Assembly for Wales decided to include the technique as one of the environmental monitoring requirements to measure change along the coastline. As such, the procedure would be deployed annually to monitor levels of key beaches in South Gower. This will continue until at least June 2003. This requirement is believed to be the first of its kind in the UK, and highlights the fact that the analyses of the highly accurate topographical data over time is innovative and pioneering.

To complement these activities, in early 1999, a collaborative project called the "Glamorgan Coastal Monitoring Initiative" was launched by the Welsh Office. This was designed to further embrace the potential of this airborne remote sensing technique, and to marry the data to high-resolution habitat information derived from the Compact Airborne Spectrographic Imager (CASI). This is discussed in the following section.

LiDAR produces extensive data sets. In Kenfig National Nature Reserve (NNR), for example, over 30 million data points have been collected under the Glamorgan Coastal Monitoring Initiative. A number of custom-built computer programs have been written to reduce the data files to more manageable sizes. The data sets have been transferred to an ArcView GIS for interpolation and further

analysis. Some important derivatives, such as slope and aspect, have also been computed. Plates 4.2 to 4.4 show examples of the interpolated elevation surface, and subsequent derivations of slope and aspect from LiDAR data of the same area in the Kenfig NNR.

4.3.3 Habitat Mapping Using Compact Airborne Spectrographic Imager (CASI)

Compact Airborne Spectrographic Imager (CASI) is a state-of-the-art remote sensing instrument for measuring radiation reflectances in the electro-magnetic spectrum. The classification of CASI data can produce a "map" by positively identifying the radiation reflectances of different "objects." In the case of habitat mapping, CASI can "see" a clear boundary between different types of vegetation as long as they emit different levels of radiation – the so-called spectral signature. Habitat mapping using CASI involves three components: an airborne survey using CASI; a simultaneous ground-truthing exercise; and image classification to produce the habitat map (Plate 4.5).

Table 4.1 The classified values, the classifications and their corresponding display colours in plate 4.6, the classified habitat map of Kenfig NNR.

Category	Classification	Colour
1	water	dirty green
2	bare sand	yellow
3	scrub	dark green
4	*Phragmites*	red
5	*Calluna*	light green
6	*Pteridium*	pale blue
7	successionally-young grassland	blue
8	orchid-rich slack	purple
9	successionally-young slack	light brown
10	tall rank grassland	light blue
11	embryo dune slack	pale blue

The data acquisition stage is characterized by an airborne survey using CASI. The ideal timing for the airborne survey is around midday. This is the time when the sun's angle is at its highest, casting minimum shadow and giving the most accurate representation of radiation reflectance. While the CASI survey is being carried out, a team of specialists undertakes a ground-truthing exercise to determine the ecology of selected locations within the study area. The ecological parameters are measured using traditional methods such as 1m x 1m quadrats with the aid of modern DGPS for accurate determination of position. The exercise provides important information to allow the airborne CASI data to be cross-referenced with known vegetation types. The last stage is the classification of CASI data in conjunction with the ground-truthing information to generate a map of the vegetation types. This classified map is subsequently interpreted to form the basis for a habitat map.

CASI has been used as an integral part of the "Glamorgan Coastal Monitoring Initiative." The main objective has been to provide a benchmark survey for Kenfig NNR which together with Merthyr Mawr Warren makes up the European Union's candidate Special Area of Conservation (cSAC). Both sites are of national and international importance. For example, Kenfig NNR holds over 40% of UK's population of the rare fen orchid (*Liparis loeselii*), whilst Merthyr-Mawr Warren is the second highest mobile sand dune system in Europe.

Kenfig has many other features of conservation interest. The successionally-young stages of dune development, in particular, are species-rich, with more than 40 plant species usually found in a 1m x 1m quadrat. These habitats have been prioritised for management and monitoring purposes under the current European Habitat Directive. However, there are growing concerns that the over-stabilization of the mobile sand dunes at Kenfig NNR is threatening the biodiversity of the cSAC.

For discussion purpose, Kenfig NNR has been divided into two broad areas: a southern and northern section (Figure 4.7). The southern section is mostly dominated by mature dune vegetation, whereas successionally-young seral stages are more prominent in the north of the site. The boundary is situated where one starts to grade into the other.

Figure 4.7 Kenfig NNR divided. The section in darker grey represents the ecologically older and more mature southern region of the Reserve. The northern area is depicted in pale grey.

The area Figure for the successionally-young grassland has been calculated by carrying out a pixel count for the habitat type based on a pixel size of 2 x 2m (Table 4.2). The area measurement for the whole site has been calculated in the same way.

Table 4.2 Area measurements for successionally-young grassland at Kenfig NNR, taken from the classified CASI images.

Section	Area of Section ($m^{2)}$	Area of successionally-young grassland (m^2)	%
Southern	1,771,276	201,248	11.3
Northern	2,706,684	782,120	28.8
Whole site	4,477,960	983,368	21.9

Preliminary results show that successionally-young grassland not only covers a much greater area in the Northern Section (which might have been expected because it is considerably larger than the Southern Section), but also occupies a much larger percentage of the area than in the South. This evidence confirms observations that the dunes in the southern part of Kenfig are generally in an advanced state of stabilisation, while this is less so in the northern part of the site where successionally-young grassland is relatively well distributed.

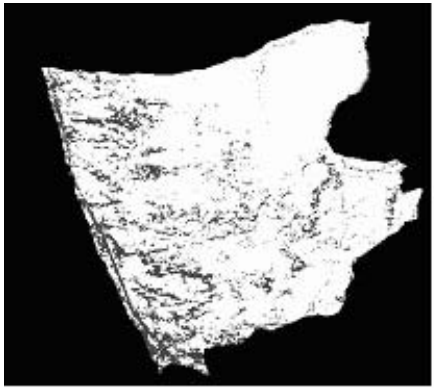


Figure 4.8a Areas classified as successionally-young grassland in the southern section

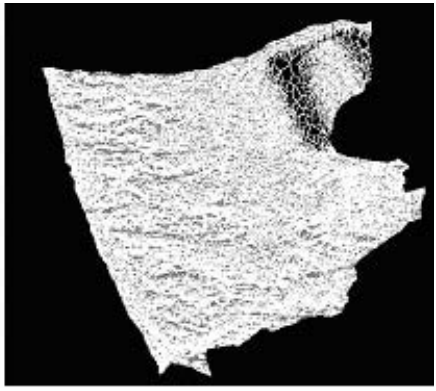


Figure 4.8b Areas with a south-facing aspect

Figure 4.8c Areas classified as successionally-young grassland in the south-western part of the southern section

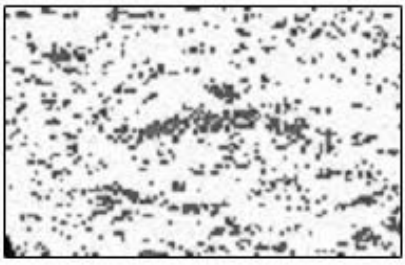


Figure 4.8d Areas with a south-facing aspect

Evidence from the joint analysis of topography and habitat classification in the southern section provides interesting comparison. A high proportion of the areas classified as successionally-young grassland occur close to the top of south-facing slopes. Figures 4.8a and b show the areas classified as successionally-young grassland in the Southern Section, juxtaposed with those showing a south-facing slope. There is also some correlation between the successionally-young grassland and aspect, and this is particularly evident in the south-western part of the section. Figures 4.8c and d are illustrative.

4.4 THE NEXT FRONTIER – MARINE AGGREGATE RESOURCE MANAGEMENT AND PLANNING SYSTEM (MARMPS)

Data collected from these innovative monitoring procedures helps to shed light on some of the plethora of natural processes causing change to the coastal environment. It also helps to expose the implications of some human activities. The data sets are used to populate a GIS. This provides a technical tool to underpin an embryonic decision-support system which will be used for the management and planning of the marine aggregate resource in South Wales to ensure it is only extracted in a sustainable way.

This system, entitled the "Marine Aggregate Resource Management and Planning System (MARMPS)," will operate using data and information on three layers: the site-specific monitoring data sets acquired in relation to the dredging licences, including those relating to bathymetry, 3D cliff analysis and topographical beach data; the regional data from the Glamorgan Coastal Monitoring Initiative, including data from both the LiDAR topographical survey and CASI habitat mapping survey; and the broad-scale and comprehensive strategic data sets derived from the "Bristol Channel Marine Aggregates: Resources and Constraints" project. This research, completed in August 2000, examined the available marine aggregate resources, and the technical, economic, environmental and cultural constraints operating to deter extraction. Two of the many significant outcomes of this work were, firstly, the delivery of an abstractive GIS database with about 100 data themes for the Bristol Channel and, secondly, the derivation of sediment environments representing sub-areas of the physical system that are considered to exhibit similarities across various components of the sediment regime. Together, these will help form the basis for future strategic policy and decision-making.

MARMPS will adopt a multi-criteria approach to aid decision-making at a more local level. In the Bristol Channel, a complicated number of conflicting objectives co-exist, for example, the economic demand for marine aggregate and possible environment impact. MARMPS will define the relationship between these by quantifying them in commensurate terms. In doing so, the conflicting objectives will be summarized and "optimised." The decision-makers will also take part by defining, in their own way, the relative importance of these conflicting objectives, and a weighting mechanism will be applied to take account of these views. Clearly, the weighting may change over time. MARMPS, when completed, will be the first system of its kind in the United Kingdom. It will be operational, designed to tackle real world issues. It will assist decision-makers in reconciling the conflicting interests of different users in the coastal zone.

It is not within the scope of this paper to discuss MARMPS in extended detail. However, the availability of good quality data sets is essential to the success of the development of the system. The National Assembly for Wales will continue to strive to acquire data sets that are robust and meaningful. Continual monitoring of the coastal and marine environment is key to an understanding of the mechanisms of coastal change.

But MARMPS is not just about data collection. There are, for example, a number of identifiable interfaces: the relationship between the coastal and the marine environment; the interaction between various bodies and organizations who have a common interest in the coastal zone; and the need to find lines of communication between science and the general public. The most important interface is, however, the one that brings about an understanding of the strengths and limitations of the natural environment and the formulation of a complementary sustainable policy to ensure wise use of its resources. The outcomes from this interface will, of course, have socio-economic implications.

The development of MARMPS is the latest in an ever-increasing list of new ways of collecting, analysing and using data and information by the newly devolved National Assembly for Wales — the introduction of GIS; routine deployment of LiDAR to monitor coastal change; habitat mapping using CASI; and 3D geometric monitoring of cliff instability. MARMPS will no doubt assist decision-makers at the National Assembly for Wales in making better informed and evidence-based decisions on the future sustainable use of our coastal heritage for many generations to come.

4.5 EPILOGUE

A dissemination seminar for the “Glamorgan Coastal Monitoring Initiative” took place in Porthcawl, South Wales, United Kingdom, on 11 July 2001. (Further details are available at: http://www.swansea.ac.uk/geog/gcmi.) The main objective was to report the results of the Initiative to participating partners, as well as to communicate knowledge about these advanced scientific techniques to the general public.

4.6 ACKNOWLEDGEMENTS

The authors would like to thank the National Assembly for Wales and the University of Wales for their support in the production of this joint work. They would also like to thank Clive Hurford of the Countryside Council for Wales and Graham Thackrah of the University of Wales, Swansea (now at University College London) for their assistance in the habitat mapping of Kenfig cSAC.

The authors are responsible for the contents of this paper, and none are attributable to either the National Assembly for Wales or the University of Wales.

4.7 REFERENCES

Clark, D.M., Hastings, D.A., and Kineman, J.J., 1991, Global databases and their implications for GIS. In *Geographical Information Systems: principles and application, vol. 2* edited by Maguire, D.J., Goodchild, M.F., Rhild, D.W. (London: Longman) pp. 217-231.

Countryside Council for Wales, 2000, *Conservation Report on Kenfig NNR.*

Countryside Council for Wales, 2001, *Habitat Monitoring for Conservation Management and Reporting Volume 1: Case Studies.*

Dargie, T.C.D., 1995, *Sand Dune Vegetation Survey of Great Britain: Part 3 – Wales. JNCC*. Peterborough.

Hurford, C., 1997, *Year 1 Report on the Fen Orchid* Liparis loeselii *Species Recovery Programme at Kenfig NNR, Glamorgan.* Countryside Council for Wales. Unpublished contract report.

Jones, P.S., 1996, *Kenfig National Nature Reserve: A Profile of a British West Coast Dune System, in Studies in European Coastal Management* (Tresaith, Cardigan: Samara Publishing Ltd.) pp. 292.

Maguire, D.J., 1991, An overview and definition of GIS. In *Geographical Information Systems: principles and applications, vol. 1* edited by Maguire D.J., Goodchild M.F., Rhild, D. W. (London: Longman) pp. 9-20.

Pan, P.S.Y. and Morgan, C.G., 1998, Monitoring of the Welsh Coastal Environment, in *Wavelength*, Department of the Environment, Transport and the Regions, Issue 2.

Pan, P.S.Y. and Morgan, C.G., 1999, Glamorgan Coastal Monitoring Initiative 1999, in *The Proceedings of the 25th Annual Technical Conference of the Remote Sensing Society*, Cardiff, September 1999, ISBN 0 946226 27 X.

Pan, P.S.Y., Morgan, C.G., Hurford, C. and Thackrah, G., 2001, Remote Sensing, GIS and the Coastal Environment in South Wales, in *The Proceedings of the GIS Research UK 9th Annual Conference GISRUK 2001*, pp 14-15.

Posford Duvivier & ABP Research & Consultancy, 2000, *Bristol Channel Marine Aggregates: Resources and Constraints*, National Assembly for Wales.

Sanjeevi Shanmugam and M.J. Barnsley, 1996. Monitoring vegetation succession in coastal dunes using remote sensing: a case study from the Kenfig NNR, South Wales, in *The Proceedings of the 22nd Annual Technical Conference of the Remote Sensing Society*, Nottingham, pp 236-243.

CHAPTER FIVE

Spatial Uncertainty in Marine and Coastal GIS

Eleanor Bruce

5.1 INTRODUCTION

The dynamic nature of coastal landscapes and the inherent complexity of the biophysical processes operating in these environments challenge the application of GIS methods. It is well recognised that spatial data models representing static objects are rife with uncertainty (Fisher, 1999; Foody, 2003). However, the mobility of many coastal and marine phenomena and the nebulous nature of boundaries in these environments provide an additional dimension to the problems associated with spatial data uncertainty. In abstracting the infinite complexity of reality into a finite computer based storage structure, multiple levels of uncertainty are introduced. The more encompassing or inclusive a data set, often the more complex the process of abstraction. Users of coastal and marine GIS are faced with both uncertainty in the information derived from spatial data, and uncertainty that inherently exists in the models. The ubiquitous nature of uncertainty in spatial analysis highlights the need to examine the implications for coastal and marine decision-making. This chapter examines the sources of uncertainty, methods for assessing reliability, model uncertainty and the cognitive and practical implications associated with the communication and incorporation of uncertainty in coastal and marine GIS.

5.2 SOURCES OF UNCERTAINTY IN COASTAL AND MARINE GIS

Uncertainties arise in the conceptualisation, measurement and analysis of geographic phenomena (Longley *et al.*, 2001). Distinctions have been made between different types of spatial uncertainty, to assist in identifying the problems and consequences that may emerge. There are a range of terms used to describe uncertainty including accuracy, error, reliability, precision and indeterminacy. Fisher (1999) identified three major types of uncertainty: these include error, vagueness and ambiguity.

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

Uncertainty is associated with error when both the class of object and the individual represented in the GIS data model are clearly defined (Fisher, 1999). This form of uncertainty is probabilistic in nature. However, some uncertainties are inherent in the geographic phenomena, such as the transition zone between two coastal vegetation communities, and cannot be attributed to randomness (Cheng, 1999; Fisher, 1999; Longley *et al.*, 2001).

Vagueness relates to the difficulty in assigning classes or crisp boundaries to poorly defined objects. For example, the question of whether an area can be described as a saltmarsh requires a threshold value of a measurable parameter, such as the relative or absolute occurrence of saltmarsh species, or recourse to expert opinion (Fisher, 1999; Longley *et al.*, 2001).

Ambiguity occurs when there are differing perceptions of a phenomenon and how it should be classified (Fisher, 1999; Foody, 2003). For example, different standards may be used to define shoreline position, such as Mean High Water (MHW) or Mean Low Water (MLW). These differences in definition will lead to ambiguities in shoreline change analysis (Anders and Byrnes, 1991). Sources of uncertainty frequently encountered in coastal and marine GIS include indeterminate environmental boundaries, the fragmentation of jurisdictional responsibilities over data capture and the scientific uncertainty associated with models.

Despite the prominence and in some cases the requirement for sharply defined boundaries in GIS databases, most objects in geographic space have unclear, fuzzy or non-existent boundaries (Couclelis, 2003; Frank, 1996). This is particularly evident in coastal and marine environments, in which the transition or transient state of many phenomena exacerbates the boundary delineation problem. For example, in defining the extent of a coastal wetland, which is subject to seasonal inundation, there will be temporal variation in boundary location. A range of local interacting environmental factors, including surface hydrology, soil property and groundwater recharge, will influence rate of inundation and consequently the position of the wetland boundary. Potentially compounding this problem, there may be uncertainty in the appropriateness of the definition adopted as the basis for determining the inclusion of an area in the description of wetland. If a wetland has been artificially modified to restrict tidal influences can it still be classified as a coastal wetland? This is described as uncertainty due to indefinite correspondence between concepts and the represented world, where the information may be precise but the classification unclear (Freksa and Barkowsky, 1996). Maps delineating coastal habitats such as wetlands or mangrove communities often form the spatial framework for establishing reserve areas, applying environmental protection legislation or restricting development. Users may assume data certainty based on the neatness of the precise line used to represent an area of habitat on a map while the zone of uncertainty surrounding each boundary remains ignored in the decision process. Management protocols and legislative frameworks often require crisp boundaries for coastal and marine phenomena that are continuous in nature.

Although the legacy of disparate government and private-sector responsibility in coastal and marine management is being addressed in many countries, through the adoption of integrated resource management (Harvey and Caton, 2003), such fragmentation is often still reflected in the capture and custodianship arrangements of spatial data. Bathymetric data is primarily

compiled by hydrographic agencies for marine navigation, transport planning and maritime safety (Ward *et al.*, 2000). Terrestrial elevation data is collated by state or federal land mapping agencies operating under different spatial data capture programs. However, the mapping and spatial modeling of coastal phenomena that traverse the water line, such as historical shoreline change, nearshore dynamics or intertidal sediment accretion, requires these datasets to be merged. The merging of these data involves careful consideration of the reference ellipsoid and water level datum (Li, 2000). Paradoxically, data discrepancy, and consequently uncertainty, is often greatest in the zone of highest research significance, the inter-tidal region. GIS based models designed to examine the coastal impacts of sea-level rise scenarios are particularly sensitive to inconsistencies between merged elevation surfaces (Bruce *et al.*, 2003). Recent developments in coastal applications of Airborne Laser Hydrography are providing detailed 'seamless' Coastal Terrain Models (CTMs) (Sinclair *et al.*, 2003). However, if these data are to become available to broader GIS research applications, there will be a need to reconsider the appropriateness of the shoreline as the delineation for the division of spatial data capture responsibilities.

A second consequence of multiple jurisdictional responsibilities is the incompatibility of classification systems used to describe coastal and marine biophysical environments that extend administrative boundaries. This may generate greater uncertainty at state or national boundaries. For example, in Australia the system of Marine Protected Areas (MPAs) is required to meet principles of comprehensiveness, adequacy and representativeness through strategic inclusion of diverse examples of marine life. In achieving this, the mapping of nearshore waters has involved the use of benthic habitats as a surrogate for biodiversity by most jurisdictions (state and territory governments). However, jurisdictional differences in the spatial scale of responsibility, the nature of marine environments (water clarity and depth), and conceptual divergence in definitions of mapping elements have resulted in considerable variation in mapping standards in Australian nearshore waters (ANZECC TFMPA, 2000). An attempt has been made to organise different jurisdictional mapping approaches and construct a single hierarchical ecosystem-based classification (ANZECC TFMPA, 2000). These issues are particularly evident in coastal and marine environments in which biophysical processes transcend jurisdictional boundaries.

It is clear that uncertainty in the conception and measurement of geographic phenomena will lead to uncertainty in analysis results (Longley *et al.*, 2001). However, in many coastal and marine GIS applications uncertainty in the model must also be considered. The exact values of model parameters are rarely known; therefore it is often necessary to estimate values, but there is always an associated error in that estimation (Heuvelink *et al.*, 1989). Increasing concern by researchers and environmental managers for the potential impacts of projected climate change to coastal environments has highlighted the importance of comprehensive model-based approaches for predicting and assessing longer-term change (Capobianco *et al.*, 1999). However, in developing appropriate spatial models there is a need to account for the scientific uncertainty that surrounds predicted sea-level conditions on which these models are based (Cowell and Zeng, 2003). The sensitivity of model results to variation in sea-level predictions can be demonstrated through alternative impact scenarios or coastal risk maps derived by modifying sea-level

curves. However, coastal managers are then faced with the challenge of transferring these alternative predictions into response policies.

5.3 METHODS FOR EXAMINING LEVELS OF DATA UNCERTAINTY

The reliability of spatial data sets can be assessed using techniques that extend beyond simple field sampling and verification to provide useful indicators of data uncertainty. Assessment methods provide indicators of data uncertainty that can be used by decision makers to determine a data set's fitness of use for the intended application. These techniques deal with measurement and processing uncertainty in the capture of spatial data. There is a broad range of techniques for assessing the reliability of spatial data that will reflect the nature of the data (categorical or continuous) and the form of uncertainty being examined (position and/or attribute). Two of these include the error matrix and the epsilon boundary model.

An accepted method for reporting accuracies in categorical data derived from remotely sensed classifications is the error matrix (Congalton and Green, 1993; Congalton and Mead, 1983; Veregin, 1999). Error matrixes cross tabulate the classified or mapped data with the referenced or field sampled data (Janssen and van der Wel, 1994). The technique has been applied to provide accuracy measures in the mapping of shallow temperate water habitats including seagrass communities and coral reefs (Bruce *et al.*, 1997; Mumby and Edwards, 2002). This approach allows the appropriateness of the assigned map classification to be assessed. For example, when mapping nearshore coastal habitats using remotely sensed data, light attenuation may distort reflectance signals resulting in confusion in the interpretation of classes such as dense seagrass and reef or low biomass seagrass species and sandy substrate. Direct comparison of field and mapped data in the error matrix provides reliability measures that allow end users to assess the level of misclassification associated with each class providing insight on the appropriateness of the data for its intended use. However, this approach assumes a formal conception or definition of the phenomena being mapped and that there is consistency in the descriptions assigned by the map producers and the field samplers. For example, the rule sets used in the interpretation of satellite imagery to define areas of dense seagrass based on reflectance signatures may not translate in the field for divers measuring the biomass density of sampling quadrats.

A useful technique for the assessment and representation of uncertainty associated with the cartographic representation of mapped boundaries is the epsilon error model (Blakemore, 1984; Chrisman, 1982; Chrisman, 1989). The epsilon error is based on the principle that the true position of a boundary line will occur at a displaced distance from the represented or mapped line (Chrisman, 1989). The epsilon area can be described geometrically as the zone extending either side of the mapped line delineating a probability density function of the boundary's real world location (Chrisman, 1989). True boundary position can be measured in the field through transect surveys to record changes in cover type. Using the line intercept method transect results are then overlaid with the mapped data to determine boundary offset (Skidmore and Turner, 1992). In the marine environment this can be conducted using an underwater video camera towed behind a vessel or detailed dive transects. The width of epsilon bands can be set uniformly according to the standard deviation of the uncertainty of the mapped

boundary; in this way the epsilon zone represents a form of mean error (Chrisman, 1989). However, these boundary; error concepts originated in forestry science (de Vries, 1986) and are not easily transferred to the marine and coastal environment. The distribution of error and the corresponding shape of the error band may not be uniform in width (Caspary and Scheuring, 1993; Veregin, 1999). In the nearshore environment factors such as water depth and turbidity may influence the distribution of error around the mapped line. Research conducted in temperate shallow-water environments in southeastern Australia has demonstrated that the depth at which seagrass boundaries can be detected is not an absolute measure but rather a relative measure influenced by the same conditions that control seagrass growth (Mount, 2003).

Quantifying uncertainties in measurement and classification will guides users and decisions makers in assessing the limitations of the data and their decisions. Estimates of measurement error also provide valuable input in examining the implications of uncertainty in analysis results. If spatial data are to be incorporated into coastal and marine models it is important to examine how measurement and processing errors influence model output.

5.4 APPROACHES FOR EXAMINING AND REPRESENTING MODEL UNCERTAINTY

Uncertainty in coastal and marine data and models is inevitable and useful descriptions of the properties of error are provided in the literature (von Meyer *et al.*, 2000). An important consideration for coastal and marine GIS users is how this uncertainty can be managed. Innovative approaches designed to model and represent underlying data uncertainty have emerged from recent spatial uncertainty literature (Morris, 2003). This section briefly describes three of these approaches and presents examples in which they have been applied to modelling coastal and marine environments. These include error propagation, sensitivity analysis and fuzzy set theory. Error propagation and sensitivity analysis measures the impacts of uncertainty in the input data on the analysis or model results. Fuzzy set theory allows an object's inclusion in a class to be assigned along a continuum, offering a definable approach for dealing with inexact concepts (Cowell and Zeng, 2003; Zadeh, 1965).

Analysis of error propagation allows the way in which uncertainties accumulate and affect the end result of the model to be explored (Burrough and McDonnell, 1998). Errors may continue when the output of one spatial operation (such as map algebra or polygon overlay) is used as the input for a subsequent operation (Heuvelink, 1999). The exactness of quantitative results derived from multiple input data layers in spatial models often makes it difficult for the end user to recognise the effects of data and model error. Although uncertainty in both the input data and the model will influence the value of results, it is also critical to examine the way the data and the model interact (Burrough and McDonnell, 1998). For example, uncertainties in elevation will propagate when bathymetric surfaces compiled at different dates are used to compute estimates of sedimentation rates (Van der Wal and Pye, 2003). If sedimentation is measured as the change in elevation the standard error is the root of the squared error deviations of the two bathymetric surfaces (Van der Wal and Pye, 2003). This has significant

implications for the quantification of morphological change if standard errors exceed finer scale change estimates. In tracking error behaviour throughout the model comparisons can be made between the contribution of data input and model error (Heuvelink, 1999). This allows developers and those implementing the model to assess the appropriateness of the model in relation to the available spatial data. However, an error propagation analysis can only yield plausible results if the errors associated with each model input have realistic values (Heuvelink, 1999).

Sensitivity analysis is the controlled variation of parameter values in isolation and in combination, and the observed response of model output (Conroy *et al.*, 1995; Lodwick *et al.*, 1990). By imposing variations on the input data sets, measurement of effects on the model results can provide an indication of reliability. An example of this in spatial analysis is the alteration of input data resolution. Identification of the variables with the greatest influence on model results is not only critical to model calibration but also provides direction for future research (Hamby, 1994). The effects of an input parameter on model outcome can be divided into two types. Parameters whose uncertainty contributes to the results are referred to as 'important' parameters, and those that have a significant influence on the results are 'sensitive' parameters (Hamby, 1994). Sensitivity analysis has been used to examine levels of uncertainty in marine habitat modeling (Bruce, 1997). Bayesian probability theory was applied to predict dugong distributions in a semi-enclosed coastal embayment in Western Australia (Bruce, 1997). Sensitivity analysis was applied to examine the effect of species mobility on the model results as dugongs may range over an extensive area within a preferred habitat potentially contributing to uncertainty in sighting location. It was demonstrated that the model was sensitive to positional error in species sighting data at positional offsets of greater than 800 metres (Bruce, 1997). Sensitivity analysis examines uncertainty from output back to input, which is the reverse direction to the error propagation approach that provides a quality test for the input data and the model (Crosetto and Tarantola, 2001). However, these approaches assume certainty in the definition of an object or the conceptual model used to represent a phenomenon. Methods are needed to deal with the polythetic (defined by multiple attributes) and inexact nature of many natural geographic phenomena.

An approach that addresses problems associated with vagueness of definition is fuzzy set theory (Fisher, 1999). In classic set or Boolean theory an object is defined as belonging or not belonging to a class or set. This membership status is coded as values of 1 or 0 respectively (Fisher, 2000; Fisher, 1999; Robinson, 2003). In fuzzy set theory, membership is defined by a real number in a range from 0 (nonmember) to 1 (full member) (Bernhardsen, 2002; Cowell and Zeng, 2003; Fisher, 1999). Fuzzy set theory allows an object membership to a varying degree and to more than one set (Robinson, 2003). For example, in representing the presence of seagrass using classic set theory a seagrass patch may be assigned a value of 1 and an area of bare substrate a value of 0. The same patch represented using fuzzy set theory may have a value reflecting its high degree of membership in the class of seagrass (e.g. 0.85) and a low value indicating a small degree of membership in the class of sand (e.g. 0.1). The assignment of the membership function of a fuzzy set is subjective in nature but the process of assignment is not arbitrary (Robinson, 2003). A range of approaches for assigning grades of membership can be adopted when applying fuzzy sets in GIS (Robinson, 2003).

Fuzzy membership is further complicated when representing natural phenomena that change. Cheng (1999) presents a method to identify coastal features and their dynamics using fuzzy spatial extent (regions) extracted at different epochs. Application of fuzzy set theory for predicting coastal geomorphic hazard associated with sea-level induced recession and storm erosion has also been demonstrated for beach environments in SE Australia (Cowell and Zeng, 2003). Although not a complete solution to the problem of vagueness, an important strength of fuzzy set theory is that it allows partial implementation of the vagueness concept (Fisher, 1999; Robinson, 2003).

5.5 COGNITIVE DIMENSIONS OF UNCERTAINTY

Certainty in the spatial representation of reality has been contested in the broader spatial science community since the earlier stages of GIS (Burrough, 1986; Chrisman, 1982; Chrisman, 1984; Goodchild, 1989). Conceptualization of the nature of uncertainty in GIS has been documented (Fisher, 1999) and methods have been developed for the representation and handling of uncertainty (Agumya and Hunter, 2002; Veregin, 1999). In recent GIS literature there has been recognition of a shift in the perception of uncertainty. Rather than focusing on uncertainty as the consequence of empirical failings Couclelis (2003) argues for accepting uncertainty as an intrinsic property of knowledge. The challenge faced by users of coastal and marine GIS is how these insights on uncertainty translate to the decision-making process.

With the increasing flow of spatial data, facilitated through web-based data clearinghouses, Internet mapping and interoperability, there is potential for unchecked dissemination of associated data uncertainty. Attached metadata documents and national and international spatial data quality standards assist in addressing the unsuspected promulgation of errors and inherent bias. However, in order to accommodate uncertainty coastal and marine managers, planners and communities must convert these data qualifiers into guidelines that will form the basis of meaningful decisions. There is a cognitive dimension to the communication of uncertainty that must extend concepts accepted in the realm of GIS to ensure accessibility to the diverse community of end users. For example, inclusion of an uncertainty measure in the output of a coastal hazard risk model will result in risk being mapped along a continuum rather than an absolute (Zerger *et al.*, 2002). In contrast, emergency managers require clear guidelines for identifying and evacuating risk areas (Zerger *et al.*, 2002).

There is often a requirement from end-users of GIS derived coastal and marine information for unambiguous information that can be incorporated directly into decision structures. These user requirements need to be considered in any definition or representation of uncertainty if end-users are expected to recognise and adopt adequate precautionary approaches. Returning to the wetland boundary example presented earlier, depiction of the ‘zone of confusion’ surrounding each wetland boundary, although highlighting uncertainty, does not provide a clear basis for administrative decisions. Legislation relating to the environmental protection of a coastal wetland system may require precise definitions of spatial extent. For example, under the New South Wales Coastal Protection Act in Australia definition of the coastal zone includes all coastal rivers to the upstream

limit of mangrove plants. Although inclusive of a critical estuarine habitat, this definition does not include saltmarsh communities resulting in the occurrence of these habitats outside the jurisdiction of associated coastal management policies. Many approaches for the representation and handling of spatial uncertainty do not translate neatly into legislative frameworks designed to ensure the protection and ecological sustainability of coastal and marine environments.

Ensuring end-user awareness of spatial uncertainty is not restricted to coastal and marine managers, policy makers and other practitioners. There is a growing role for GIS in encouraging community involvement in environmental decision-making. Recent approaches for integrating local expertise and perceptions into a GIS framework have been described as GIS for participation (GIS-P)(Cinderby, 1999; Weiner *et al.*, 2002). An example includes the adoption of participatory GIS to capture local fisheries knowledge from a Newfoundland fishing community on the northeast coast of Canada (Macnab, 2002). These applications further highlight the cognitive aspect of spatial uncertainty. If results are presented without building the capacity of participating groups to understand limitations associated with the analysis and input data, conflict within communities may be exacerbated by the implementation of GIS-P (Cinderby, 1999). There is value in the use of fuzzy representations over distinct map boundaries in these circumstances. Exact boundaries may infer a level of positional precision that could potentially prompt dispute. Successful development of participatory GIS applications in the coastal and marine environment will require approaches that ensure concepts of uncertainty are communicated in an accessible way to the broader community.

Difficulties associated with incorporating uncertainty measures, fuzzy boundaries or relative values may depend on the temporal scale and the legislative framework of the decision-making process. There is greater capacity for longer-term coastal planning strategies to accommodate fuzzy results and translate these to precautionary decisions based on acceptable risk. However, faced with the immediate responsibility to approve or reject a development proposal under the threat of habitat loss a decision maker may require absolute values on which to base and defend, perhaps in court, their decision.

The exactness of legislative statements relating to the protection of coastal and marine environments is not suited to vague or ambiguous representations and often requires the conversion of fuzzy values into Boolean structures, forced classifications and precise boundaries. This need to simplify the world may reflect an underlying familiarity with the traditional cartographic paradigm of discrete entities. Perhaps there is a need to investigate approaches for accommodating flexible definitions in coastal and marine policies and associated legislative frameworks.

5.6 CONCLUSION

Uncertainty in GIS is inevitable; it is how this uncertainty is perceived and handled which is important. Fragmentation of the landscape into manageable spatial data models is required in many applications of coastal and marine GIS. During this process of abstracting reality the loss of detail potentially introduces levels of uncertainty that may have significant consequence in analysis and modeling

results. Compounded by this is the uncertainty that stems from the complexity of coastal and marine systems (Cowell and Zeng, 2003). Techniques exist that allow the quantification, assessment and communication of uncertainty in GIS applications as demonstrated by the range of coastal and marine case studies cited in this chapter. However, if widespread awareness of spatial uncertainty issues is to be achieved amongst the coastal and marine GIS community a greater understanding of users' cognition, perception and incorporation of uncertainty is needed.

5.7 REFERENCES

Agumya, A., and Hunter, G.J., 2002, Responding to the consequences of uncertainty in geographical data. *International Journal of Geographical Information Science*, **16**, pp. 405-417.

Anders, F.J., and Byrnes, M.R., 1991, Accuracy of shoreline change rates as determined from maps and aerial photographs. *Shore and Beach*, **59**, pp. 17-26.

Australian New Zealand Environment Conservation Council, and Task Force Marine Protected Areas. 2000, *Review of Methods for Ecosystem mapping for the National Representative System of Marine Protected Areas*, (Canberra: Environment Australia).

Bartlett D.J. and Bruce, E., 2002. Quality Control in Coastal GIS, In *An Evaluation of Progress in Coastal Policies at the National Level, A Transatlantic and Euro-Mediterranean Perspective*, edited by B. C. Sain, I. Pavlin and S. Belfiore. Dordtrecht, The Netherlands, Kluwer Academic Publishers (NATO Science Series).

Bernhardsen, T. 2002, *Geographical Information Systems: An Introduction*, (New York: John Wiley & Sons Inc.).

Blakemore, M., 1984, Generalisation and error in spatial databases. *Cartographia*, **21**, pp. 131-139.

Bruce, E., Cowell, P.J., and Stolper, D., 2003, Development of a GIS-based estuary sedimentation model. In *Coastal GIS - An Integrated Approach to Australian Coastal Issues* (Wollongong: University of Wollongong). 553pp.

Bruce, E.M., 1997, Application of Spatial Analysis to Coastal and Marine Management in the Shark Bay World Heritage Area, Western Australia. PhD thesis, University of Western Australia, Perth.

Bruce, E.M., Eliot, I.G., and Milton, D.J., 1997, A method for assessing the thematic and positional accuracy of seagrass mapping. *Marine Geodesy*, **20**, pp. 175-193.

Burrough, P.A. 1986, *Principles of Geographical Information Systems for Land Resources Assessment*, (New York: Oxford University Press).

Burrough, P.A., and McDonnell, R.A. 1998, *Principles of Geographical Information Systems*, (New York: Oxford University Press).

Capobianco, M., DeVriend, H.J., Nicholls, R.J., and Stive, M.J.F., 1999, Coastal area impact and vulnerability assessment: The point of view of a morphodynamic modeler. *Journal of Coastal Research*, **15**, pp. 701-716.

Caspary, W., and Scheuring, R., 1993, Positional accuracy in spatial databases. *Computers, Environment and Urban Systems*, **17**, pp. 103-110.

Cheng, T. 1999, *A Process-Oriented Data Model for Fuzzy Spatial Objects*, (The Netherlands: ITC Publications).

Chrisman, N.R., 1982, A theory of cartographic error and its measurement in digital data bases. In *Fifth Symposium on Computer- Assisted Cartography and International Society for Photogrammetry and Remote Sensing Commission*, Virginia, US, (Environmental Assessment and Resource Management) pp.159-168.

Chrisman, N.R., 1984, The role of quality information in the long-term functioning of a Geographic Information System. *Cartographica*, **21**, pp. 79-87.

Chrisman, N.R., 1989, Modeling error in overlaid categorical maps. In *Accuracy of Spatial Databases*, edited by Goodchild, M. F., and Gopal, S., (London: Taylor & Francis), pp. 21-34.

Cinderby, S., 1999, Geographic Information Systems (GIS) for participation: The future of environmental GIS? *International Journal of Environmental Pollution*, **11**, pp. 304-315.

Congalton, R.G., and Green, K., 1993, A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, **59**, pp. 69-74.

Congalton, R.G., and Mead, R.A., 1983, A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, **49**, pp. 69-74.

Conroy, M.J., Cohen, Y., James, F.C., Matsinos, Y.G., and Maurer, B.A., 1995, Parameter estimation, reliability, and model improvement for spatially explicit models of animal populations. *Ecological Applications*, **5**, pp. 17-19.

Couclelis, H., 2003, The certainty of uncertainty: GIS and the limits of geographic knowledge. *Transactions in GIS*, **7**, pp. 165-175.

Cowell, P.J., and Zeng, T.Q., 2003, Integrating uncertainty theories with GIS for modeling coastal hazards of climate change. *Marine Geodesy*, **26**, pp. 5-18.

Crosetto, M., and Tarantola, S., 2001, Uncertainty and sensitivity analysis: Tools for GIS-based model implementation. *International Journal of Geographical Information Science*, **15**, pp. 415-437.

de Vries, P.G. 1986, *Sampling Theory for Forest Inventory: A Teach-Yourself Course*, (Berlin: Springer-Verlag).

Fisher, P., 2000, Sorites paradox and vague geographies. *Fuzzy Sets and Systems*, **113**, pp. 7-18.

Fisher, P.F., 1999, Models of uncertainty in spatial data. In *Geographical Information Systems: Principles and Technical Issues*, vol. 1, edited by Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W., (New York: John Wiley & Sons Inc.), pp. 191-205.

Foody, G.M., 2003, Uncertainty, knowledge discovery and data mining in GIS. *Progress in Physical Geography*, **27**, pp. 113-121.

Frank, A.U., 1996, The prevalence of objects with sharp boundaries in GIS. In *Geographic Objects With Indeterminate Boundaries, GISDATA 2*, edited by Burrough, P. A., and Frank, A. U., (London: Taylor & Francis), pp. 29-40.

Freksa, C., and Barkowsky, T., 1996, On the relations between spatial concepts and geographic objects. In *Geographic Objects With Indeterminate Boundaries*, edited by Burrough, P. A., and Frank, A. U., (London: Taylor & Francis), pp. 109-121.

Goodchild, M.F., 1989, Modeling error in objects and fields. In *Accuracy of Spatial Databases*, edited by Goodchild, M. F., and Gopal, S., (London: Taylor & Francis), pp. 107-113.

Hamby, D.M., 1994, A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*, **32**, pp. 135-154.

Harvey, N., and Caton, B. 2003, *Coastal Management in Australia*, (Melbourne: Oxford University Press).

Heuvelink, G.B.M., 1999, Propagation of error in spatial modelling with GIS. In *Geographical Information Systems: Principles and Technical Issues*, vol. 1, edited by Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W., (New York: John Wiley & Sons Inc.), pp. 207-217.

Heuvelink, G.B.M., Burrough, P.A., and Stein, A., 1989, Propagation of errors in spatial modelling with GIS. *International Journal of Geographic Information Systems*, **3**, pp. 303-322.

Janssen, L.L.F., and van der Wel, F.J.M., 1994, Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing*, **60**, pp. 419-426.

Li, R., 2000. Data models for Marine and Coastal Geographic Information Systems. In *Marine and Coastal Geographical Information Systems*, edited by Wright, D., and Bartlett, D., (London: Taylor & Francis), pp.25-36.

Lodwick, W.A., Monson, W., and Svoboda, L., 1990, Attribute error and sensitivity analysis of map operations in geographical information systems: suitability analysis. *International Journal of Geographical Information Systems*, **4**, pp. 413-428.

Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W. 2001, *Geographic Information Systems and Science*, (Chichester, England: John Wiley & Sons Inc.).

Macnab, P., 2002, There must be a catch: Participatory GIS in a Newfoundland Fishing Community. In *Community Participation and Geographic Information Systems*, edited by Craig, W., Harris, T., and Weiner, D., (London: Taylor & Francis), pp. 173-191.

Morris, A., 2003, A framework for modeling uncertainty in spatial databases. *Transactions in GIS*, **7**, pp. 83-101.

Mount, R., 2003, The application of digital aerial photography to shallow water seabed mapping and monitoring - how deep can you see? In *Coastal GIS - An Integrated Approach to Australian Coastal Issues*, (Wollongong: University of Wollongong). 553pp.

Mumby, P.J., and Edwards, A.J., 2002, Mapping marine environments with IKONOS imagery: Enhanced spatial resolution can deliver greater thematic accuracy. *Remote Sensing of Environment*, **82**, pp. 248-257.

Robinson, V.B., 2003, A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. *Transactions in GIS*, **7**, pp. 3-30.

Sinclair, M., Townsend, N., and Barker, R., 2003, Survey operations for coastal zone management in New South Wales using Airborne Laser Hydrography. In *Coastal GIS - An Integrated Approach to Australian Coastal Issues*, (Wollongong: University of Wollongong) 553pp.

Skidmore, A.K., and Turner, B.J., 1992, Map accuracy assessment using line intersect sampling. *Photogrammetric Engineering and Remote Sensing*, **58**, pp. 1453-1457.

Van der Wal, D., and Pye, K., 2003, The use of historical bathymetric charts in a GIS to assess morphological change in estuaries. *Geographical Journal*, **169**, pp. 21-31.

Veregin, H., 1999, Data quality parameters. In *Geographical Information Systems: Principles and Technical Issues*, vol. 1, edited by Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W., (New York: John Wiley & Sons Inc.), pp. 177-189.

Von Meyer, N., Foote, K.E., and Huebner, D.J., 2000, Information quality considerations for coastal data. In *Marine and Coastal Geographic Information Systems*, edited by Wright, D., and Bartlett, D., (London: Taylor & Francis), pp. 295-308.

Ward, R., Roberts, C., and Furness, R., 2000, Electronic chart display and information systems (ECDIS): state-of-the-art in nautical charting. In *Marine and Coastal Geographical Information Systems*, edited by Wright, D., and Bartlett, D., (London: Taylor & Francis), pp. 149-161.

Weiner, D., Harris, T.M., and Craig, W.J., 2002, Community participation and geographic information systems. In *Community Participation and Geographic Information Systems*, edited by Craig, W. J., Harris, T. M., and Weiner, D., (London: Taylor & Francis), pp. 3-16.

Zadeh, L.A., 1965, Fuzzy sets. *Information and Control*, **8**, pp. 338-355.

Zerger, A., Smith, D.I., Hunter, G.J., and Jones, S.D., 2002, Riding the storm: A comparison of uncertainty modelling techniques for storm surge risk management. *Applied Geography*, **22**, pp. 307-330.

CHAPTER SIX

New Directions for Coastal and Marine Monitoring: Web Mapping and Mobile Application Technologies

Sam Ng'ang'a Macharia

6.1 THE VALUE OF COASTAL AND MARINE RESOURCES

Coastal and marine areas are ever increasing in value to the welfare of nations. These areas provide natural, social and economic functions that contribute to increased quality of life. The oceans are instrumental in determining climate that beneficially affects all life on Earth (Payoyo, 1994). Other natural functions include habitat for endangered species, species breeding and resting areas, water treatment, groundwater recharge and flood attenuation. Some social and economic functions include tourism, commercial and recreational fishing, oil and gas development, and construction (Eckert, 1979; Prescott, 1985; Gomes, 1998).

It is clear that coastal and marine areas are of vital importance to human life. Yet human terrestrial and marine activities have proven to have destructive effects on these areas. According to Canada's National Program of Action (CNPA) (2000) the major threats to the health, productivity and bio-diversity of the marine environment result from human activity in the coastal areas and further inland. Approximately 80 percent of marine area contamination results from land-based activities such as municipal, industrial and agricultural waste and run-off, in addition to the deposition of atmospheric contaminants resulting from human industrial activities (CNPA, 2000; Sanger, 1987).

There is a need for a wider dissemination of knowledge relevant to the importance of coastal and marine areas to the world's well-being, and a re-evaluation of societies' attitudes towards these spaces. Good coastal and marine governance (e.g. information dissemination, management, monitoring, etc.) is therefore a key factor in the sustainable use of these environments and will require an integrated, coordinated and equitable approach (Crowe, 2000). If governance is about decision-making and steering, then up-to-date, accurate, complete, usable information (which feeds into the acquisition of knowledge) is indispensable to governance. This is especially critical in the information age of rapid changes, interconnectivity, and globalization that have brought more information to more people making them acutely aware of the unsustainable nature of current social,

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

economic and political use of marine and coastal spaces (Juillet and Roy, 1999; Rosell, 1999; Miles, 1998).

Where informed decisions have to be made using real-time information there is a need for architecture that quickly disseminates information affecting coastal and marine resources. Accurate, up-to-date, complete and useful spatial information (on many levels) regarding the resources that currently exist, the nature of the environment within which those resources exist, as well as on the users of those resources is always a requirement for effective monitoring of coastal and marine areas. Information on (but not limited to) living and non-living resources, bathymetry, spatial extents (boundaries), shoreline changes, marine contaminants, seabed characteristics, water quality, and property rights all contribute to the sustainable development and good governance of coastal and marine resources (Nichols, Monahan and Sutherland, 2000; Nichols and Monahan, 1999).

6.1.1 Web Mapping and CMM Networks

There are several Coastal and Marine Monitoring (CMM) networks in the world. Most of them are involved in obtaining real-time quantitative indicators that impact on coastal and marine health such as water temperature, water level and meteorological conditions (wind speed and direction, temperature, barometric pressure), together with qualitative indicators such as visual images of the beach and nearshore. Others are involved in collecting quantitative and qualitative indicators of coastal environmental quality. The networks involve government, academic and environmental NGO institutions. They consist of huge repositories containing databases of archival and current material. They vary in scope - from local to regional to national and international networks. Their results are provided in synchronous and asynchronous fashion but increasingly, electronic means of communication are being used to provide information needed by the various interest groups.

The World Wide Web has had a tremendous effect on the way businesses communicate. Large amounts of information can be made available quickly and conveniently to anyone with Internet access and a web browser. The ability to distribute and view spatial information has quickly shifted from a desktop application (fat client) to a browser-based architecture (thin client). This latter architecture is referred to as thin client since the user only needs a web browser to access services and information on the web (Fitzgerald, 2000). With particular reference to spatial information, there has traditionally been the question of accessing the volumes of information, especially if it resides in several different geographical locations. The web lets a data provider make spatial information available to a wider audience. The data provider can therefore provide a virtually centralised repository of resources without having to change the physical location of the data. This prevents any problems that might arise from maintaining or updating duplicate data sources, such as limited space or corrupt data. The web therefore makes it easy to provide the most up to date spatial data (Fitzgerald, 2000).

It is therefore common practice to go to one of several websites dedicated to CMM and download (or interact with) information about a specific geographical location that one is interested in. Whether it is the "Enviromapper for Watersheds" interactive mapping application that provides an index of watershed indicators for US aquatic resources, or the "Scorecard" mapping application, which maps toxic chemicals released from facilities, institutions like the Environmental Protection Agency in the USA are increasingly providing these CMM indicators through their websites (United States, 2003). The US Department of Commerce's National Oceans and Atmospheric Administration (NOAA) and the Canadian Department of Fisheries and Oceans (DFO) provide a variety of information on their websites on oil and chemical spills, tides, marine weather, and fisheries. This information is superimposed on maps to give a geographical context to the information. The superimposition of information on maps – and the provision of such information on the web, together with limited GIS functionality – is what is referred to as web mapping, webGIS or interactive mapping.

6.1.1.1 Web Mapping Technologies

The emergence of web-GIS technologies is providing the catalyst for easier collaboration, integration and cooperation among organizations with a stake in good governance and sustainable development. This is done by providing an environment for data sharing and integration over the Internet, sometimes without organizations having to make any major changes to the structure and formats of the data they maintain. The full range of analytical capabilities available in most contemporary desktop GIS however is not available on the web browser (or WebGIS client) since they are built on the thin-client concept. To include more functionalities at the client end would seem to defeat the concept of the low cost and convenience of utilizing only a web browser to access spatial data.

Traditional mapping issues still pose challenges to web mapping. Some of these new technologies support different data formats (e.g. ESRI shape files, CARIS, MapInfo files etc.), projections, scales, datums, etc., with conversions and visualization being done "on the fly." Certain web-GIS technologies now facilitate the transmission, integration, visualization and analysis of spatial information stored in geographically dispersed locations. A user with permission to access the geographically dispersed data sets need only have access to a web browser in order to view, query, and analyse the data sets. In some instances however, some webGIS can only deal with data residing on one server.

6.1.1.2 A Web Mapping Example - CARIS Spatial Fusion™

Let us briefly describe a web-mapping technology in order to explain the underlying architecture. CARIS Spatial Fusion™ is a "web-mapping" technology that lets users integrate distributed data sources using a web browser. It is an Internet-based technology whose primary function is accessing, visualizing, and analysing heterogeneous, distributed data sources (Fitzgerald, 2000, Webmapper.com, 2000). Spatial Fusion™ combines the speed, convenience and simplicity of the Internet with the ability to read multiple data sources in their native format. Earlier versions of Spatial Fusion™ consisted of a customized Java

client and a number of Fusion Data Services. On the server side, Spatial Fusion™ was made up of the following components (CARIS, 1999, Fitzgerald, 2000):

- A Web Server: One must already be running on the network.
- The Orbix™ Runtime needs to be installed on every machine that hosts a Fusion Data Service. The Orbix™ Runtime lets the Spatial Fusion™ applet and the Data Services communicate across the Internet.
- Catalog Service: This service lists all of the available Fusion Data Services.
- Fusion Data Services: These services must be registered with the OrbixWeb™ Implementation Repository. Each service has an accompanying configuration file that contains the name used to register the service with the daemon and the location of the data source.
- Configuration Utilities: CARIS MapSmith™ and CARIS dbMaps™ are provided to help customize the display of CARIS, Oracle 8i Spatial, or Shapefile data.

Recent changes in Spatial Fusion™ Version 3.0 have removed the Orbix runtime component and replaced it with TAO ORB - an open source implementation of an Object Resource Broker using Common Object Request Broker Architecture (CORBA). In Spatial Fusion™, the Object Request Broker (ORB) is the broker for information from the java servlets to various data services. The servlet container is Tomcat developed by the Apache Software foundation to manage and invoke servlets when requested. The normal setup is to have a web server such as Microsoft Internet Information Services (IIS) conFigured to connect to Tomcat. When an action is performed, the Spatial Fusion™ applet sends a request to a servlet, which is handled by the Jakarta Tomcat Servlet container. The request is then sent by the servlet to the data services through the ORB (CARIS, 2003).

As far as the user is concerned, they simply download the Spatial Fusion™ applet from a web server. At that point, a user can easily open data from any fusion service they have access to, providing them with a secure and fully scalable environment (CARIS, 1999, Webmapper.com, 2000). In addition, CARIS Spatial Fusion Developer™ lets users customize the client, so that specialized applets can be rapidly built in a drag and drop environment, giving the ability to tailor applets for specific users (Fitzgerald, 2000).

Thin client web mapping applications such as Spatial Fusion™ have traditionally depended on “wired” infrastructure because large amounts of spatial information have to be sent from the server to the client. The need to have information provided on mobile devices such as PDAs and cell phones is however driving information dissemination towards a wireless infrastructure. But would there be a need for wireless access to CMM networks? Is the general public ready to embrace this technology? In the following section we address these questions.

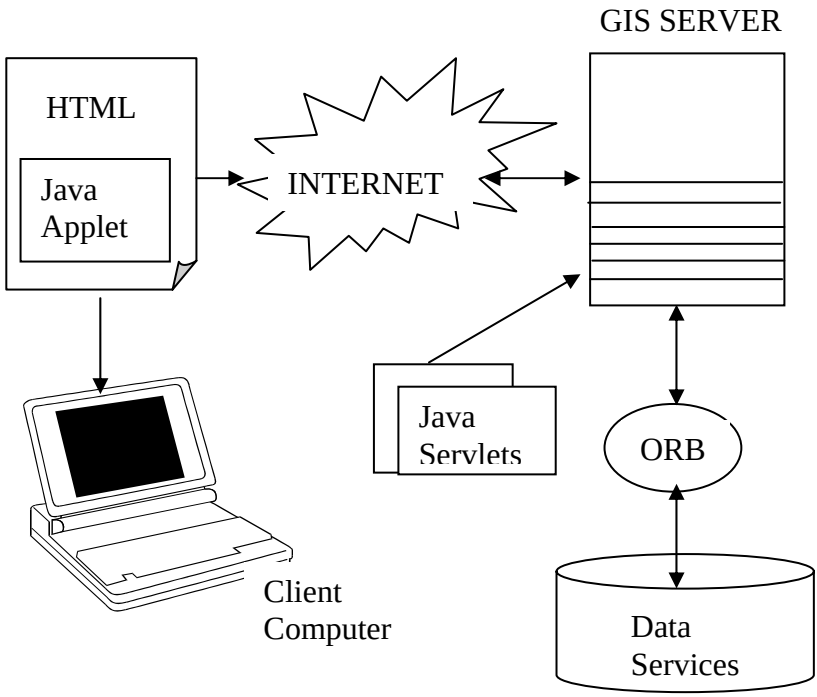


Figure 6.1 CARIS Spatial Fusion Architecture (from the CARIS Spatial Fusion Administrators Guide)

6.2 RE-ENGINEERING MARINE MONITORING NETWORKS

There exists in every jurisdiction a public policy issue associated with information collection (McLaughlin, 1991). This issue is related to the notion of providing ready and efficient access to information about resources, and rights associated with them, to all members of the community. The public is defining itself as an individual property owner – the property being the marine commons and its resources – and is realigning itself to be a primary decision maker. Most people are therefore interested in how resources are being utilized and whether resource use is being policed effectively.

Coastal and marine monitoring networks fulfill the role of providing information effectively. But they are also builders of communities of interest i.e. a forum where participants are able to interact on a specific topic. These online communities are valuable as they are increasingly providing forums for like minded individuals to participate in the decision-making process. With more citizens being involved in advocacy efforts and words like “stewardship of

resources" increasingly being used to rally the general public into participation in the decision-making process, provision of information about the space that citizens inhabit is becoming critical.

Sociologists have traditionally viewed this kind of participation as social capital – a relationship that can blur social inequality and enhance transactions that lead to capital accumulation and social growth. We are seeing the public being involved in a culture of participation, through these networks, although this is concentrated on citizen access to information. What is needed is the enhancement of citizen participation, to citizen-based leadership: where the citizens pro-actively seek information (and are able to receive it as it happens) rather than react to information that they come across. This is where wireless access to information becomes important.

6.2.1 Wireless Access to Information

Developments in Internet-enabled spatial data integration and analysis tools are now allowing decision-makers the opportunity to have access in real-time (or near real-time) to spatial information. The public is becoming more technologically "savvy" and is using newly available devices to access various kinds of information. What one finds online is also now available on mobile devices – and with evolving wireless technology we are seeing web-enabled mobile devices being used to integrate and retrieve geographic data as it is collected. Additionally, it makes sense for companies to focus on wireless Internet services given projections for future growth: the Yankee Group predicts that there will be more than one billion mobile devices worldwide by 2003. Some estimate that half of those will be web-enabled (Obeid, 2000).

This is not a new trend. Increasingly, we now have various types of information being sent to us via our mobile devices – allowing us to remain informed about various topics that we are interested in. Take the example of access to information about entertainment or restaurant locations, based on one's geographical location, delivered to a mobile phone. This represents a relatively new trend referred to as Location Based Services (LBS).

6.2.2 Location Based Services – The Driving Force

When this paper was written in 2001, the author had been following the evolution of killer applications in mobile mapping technologies – location based services (LBS). In addition to maps and directions, LBS allows customers to locate branches, service centres, stores or any points of interest near their home. A business can use LBS to locate its customers, manage its supply chain, and to implement its business rules. In fact, recently Microsoft has joined with AT&T Wireless to provide location-based services (Francica, 2002). LBS are now the driving force in wireless mapping applications.

In the coming years, Location Based Services (LBS) are expected to play a defining role in driving not only the mobile Internet but also mobile commerce as well. Location based services find the geographical location of a mobile device

and provide services based on that location. LBS applications automatically identify subscribers' locations from the wireless networks and combine this with other information to obtain maps and driving directions, perform proximity searches and find places and points of interest. Although there has been a slow rollout of LBS, Konish (2002) indicates that the slow economy is forcing wireless carriers to look for revenue-bearing services. Inevitably these are being found in making wireless devices location aware and providing location based services. Many LBS applications are already available especially in the field of asset tracking and fleet management (Sweeney, 2002). Other applications include automatic vehicle location and tracking, sale and marketing automation, consumer travel services, wireless call centre tracking, and location-based billing.

6.2.3 Providing Web Mapping to Mobile Devices

LBS can be provided in several different ways – from simple text messages providing a list of options e.g. restaurant names and locations – to a full webcentric interaction with the LBS service – involving GPS locations superimposed over interactive maps. This latter option is accomplished by using a dedicated protocol, the Wireless Application Protocol (WAP), which enables a mobile device to behave like a micro browser (Buckingham, 2000; Mobile Applications Initiative, 2001; Mobile Lifestreams Ltd, 2000). The Wireless Application Protocol (WAP) is an open, global specification that empowers users of wireless devices to easily access and interact with information and services themselves (Buckingham, 2000, Mobile Applications Initiative, 2001). This information is available through a data portal that allows the user to access services in a manner similar to that found on the Internet. The protocol however has certain limitations to it – for example, it only allows the use of sequential menus for users to browse for information. However, WAP is continually evolving to encompass greater functionality (Buckingham, 2000).

The Wireless Application Protocol incorporates a relatively simple micro-browser into the mobile phone. As such, WAP's requirement for only limited resources on the mobile phone makes it suitable for thin clients and early smart phones. WAP is designed to add value-added services by putting the intelligence in the WAP servers whilst adding just a micro-browser to the mobile phones themselves. Microbrowser-based services and applications reside temporarily on servers, not permanently in phones. The Wireless Application Protocol is designed for use with anyone's standard such as Code Division Multiple Access (CDMA), Global System for Mobiles (GSM), or Universal Mobile Telephone System (UMTS) and is aimed at turning a mass-market mobile phone into a "network-based smart phone" (Buckingham, 2000; Mobile Applications Initiative, 2001).

Wireless terminals incorporate a Handheld Device Markup Language (HDML) microbrowser that uses the Handheld Device Transport Protocol (HDTP). HDML is a version of the standard HyperText Markup Language (HTML) Internet protocol that is specifically designed for effective and cost-effective information transfer across mobile networks themselves. This protocol is used to link a mobile device terminal to a server that is connected to the Internet (or Intranet) where the information resides. The content at the site is tagged with

HDML tags that can be easily read and converted by the microbrowser in the handheld device themselves (Buckingham, 2000; Mobile Applications Initiative, 2001).

A user with a WAP-compliant phone uses the in-built micro-browser to make a request for information or service. This request is passed to a WAP Gateway that then retrieves the information from an Internet server either in standard HTML format or preferably directly prepared for wireless terminals using Wireless Markup Language (WML) themselves. In this paper, WML and HDML are used interchangeably. If the content being retrieved is in HTML format, a filter in the Wireless Application Protocol Server may try to translate it into WML. The requested information is then sent from the WAP Gateway to the WAP client, using whatever mobile network bearer service is available (Buckingham, 2000; Mobile Applications Initiative, 2001; Mobile Lifestreams Ltd, 2000).

6.2.4 Extending the Trend to CMM Networks

N'gang'a *et al*. (2001) provided a possible scenario that utilized the technological components mentioned in the preceding sections of this article. A summary of the scenario that had previously been proposed is as follows:

1. A marine event involving coastal or marine environments occurs. This event could be a oil-carrying tanker being involved in an accident, or as diverse as enforcement against banned activities or the enhancement of safety of navigation;
2. The event is detected by remote sensing means or by visual observation. These event sources are important as they establish how the event will enter into the marine event information circle. The marine event information circle is a term that has been coined by the author to represent the flow of information from the event source to the interested group;
3. If the event is detected by visual observation, the observer has the option of using a VHF radio to communicate with someone who has access to a web-mapping server (e.g. a Spatial Fusion™ data server). Alternatively, the user can use other means of event dissemination such as the mobile phone and Internet connectivity. The term "event content" refers to the information that is recorded regarding the event. In this work, this information is the location and the nature of the event;
4. The event is then registered with the event server, which in this case is analogous to a Spatial Fusion™ Data server. In particular, the data needs to be conFigured so that it can be accessed using the fusion services. An event coordinator (or moderator) must exist in the marine environment information circle to serve this purpose;
5. Transmission of information updated with the marine event is then accomplished using electronic means. An email list of a registered interest group can be used to contact the users of the marine environment information circle, or the information can be relayed to a website that the general user can access. Alternatively if there is a WAP server, the information can be relayed (with additional functionality) to members of

the registered interest group who are within geographic proximity of the event. These options allow the event server to provide alternative value-added services.

6. Information dissemination is accomplished using the wireless application protocol and existing GSM networks. In the near future GSM combined with GPRS or the third generation UMTS networks can be used for broadband access to marine events.

Once the information is served up to a user group, decisions can be made regarding implications of the marine event on existing policies, activities, or populations within (or adjacent to) the marine environment. Informed decisions are therefore made with available information about the nature and location of marine events.

6.2.5 What about Bandwidth Issues?

Spatial data involves the transmission of geometric as well as attribute information. It can therefore be argued that a substantial amount of data must be transmitted which implies that a large bandwidth must be available. There is a general consensus in the wireless industry (see Salvo, 2000; Mobile Lifestreams, 2000; Villalobos, 2000) that in the near future bandwidth will no longer be a problem especially when the third generation of mobile phone network becomes available.

We now are connected using the Global System for Mobiles (GSM) wireless network that provides data rates of 9.6 Kbytes/s. An ancillary service on the GSM network is the General Packet Radio Service (GPRS), which provides compression technology that improves data transmission rates to 100 Kbytes/s (Mobile Lifestreams, 2000; Villalobos, 2000). In the near future, the Universal Mobile Telephone System (UMTS) will provide 2 Mbytes/s (Villalobos, 2000). This will definitely change the way we use both wireless and Internet technologies.

Third-generation wireless communication (3G) such as UMTS will definitely play an important role in web based spatial visualisation in the near future. It is forecast that the mobile phone will become a portable terminal where the Internet, video and audio will be supported in a multimedia environment (Pinto, 2001). Therefore, it is expected that the WAP will adapt to the potential offered by this technology. There have been arguments that the WAP protocol is not a protocol for the future[1] due to its graphical limitations and access speed (see Salvo, 2000; Mobile Lifestreams, 2000; Villalobos, 2000); however, the protocol is still evolving and advances in technology and application are expected to eventually address these concerns.

The solution simply requires data connectivity from devices. The amount of data passed over the wireless network for the LBS applications is quite small and doesn't require the higher bandwidth allowed by 2.5G and 3G networks.

---

[1] Note that a competitor to WAP is the imode format, which uses compact HTML (cHTML) as opposed to WAP and is in use in Japan (Archey, 2002; Eurotechnology.com, 2003)

6.2.6 Some Examples

While there are several web mapping software that exist, only recently has the provision of web mapping services begun to move to mainstream mobile devices. Some of these products include (Geoplace.com, 2003;Webmapper.com, 2003)

- AutoDesk's MapGuide and Onsite – Onsite enables viewing of digital design data on Microsoft Windows CE-based handheld computing device.

- ESRI's ArcIMS and ArcPad – ArcPad provides database access, mapping, GIS, and global positioning system (GPS) integration to users out in the field via handheld and mobile devices.

- MapInfo's MapExtend and MapExtreme - *MapInfo*® *MapXtend*® is a developer tool for creating location-based applications running on wireless handheld devices.

- Microsoft Map Point Webservice – The MapPoint Web Service is the Microsoft® platform for location-based services (LBS) and can be used by mobile applications.

In Canada there are several researchers already working successfully on the kind of approach outlined in this paper. Canadian researchers Nickerson et al. (2003) have already provided a general-purpose architecture for real-time web access to mobile geospatial operations. The research demonstrated that their three-tier architecture, using Wireless Application protocol (WAP), along with eXtensible Markup Language (XML) tags embedded in a Uniform Resource Locator request, provided a sound basis for decision support of real time geospatial operations. Real time field activities were viewed by web users globally as well as stored in (and retrieved from) an online central server and database.

In Ng'ang'a et al., (2000) the authors had drawn attention to some advances in web mapping location-based services such as the automatic vehicle locator (AVL) application used at the city of Fredericton Police department[2]. This was part of the Wireless Public Safety Decision Support System (WPSDSS) for the city of Fredericton (Lunn, 2001). The application uses a cellular digital packet data wireless data network to transmit information about the location and status of police vehicles. The information is overlaid onto the City's GIS mapping layers and orthophoto imagery. While the application does not display information on mobile devices (like cell phones or PDAs), it is just a matter of time before it does.

On the CMM front in Canada, Fisheries and Oceans Canada, Environment Canada, the Defence Department, Transport Canada and the Canadian Space Agency are already collaborating on a program called Integrated Satellite Targeting of Polluters (I-STOP), to use a satellite operated by Radarsat International to identify ocean slicks (Calgary Herald, 2002; CCMC, 2002).

[2] Located in the province of New Brunswick, Canada

Canada's Radarsat satellite will be used to provide all weather, day/night images of the Earth's surface and give officials an indication of suspicious activity in a vast area that couldn't easily be covered by any other means, such as aircraft. The technology is only being used on the East Coast for now. If it proves successful, it could also be used on the Pacific coast. The Calgary Herald (2002) reported that the first hit came nine days into the project when the satellite spotted a spill off Newfoundland's south coast, near one of the province's most valued marine bird reserves. A plane was sent to the area off Cape St. Mary's and the ship was ordered back to port, where its operators now face charges under the Migratory Birds Act for allegedly creating a 116-kilometre-long, 200-metre-wide slick. While WAP technology is not being used in this program, one can see that soon enough there will be a situation where individuals interested in coastal and marine monitoring will have information delivered about this fragile ecosystems health to their mobile devices.

6.3 SUMMARY

This paper has begun by discussing the importance of information to coastal and marine governance and then expanded the discussion by outlining existing technologies that can be used to facilitate enhanced decision-making. In particular, the paper has followed developments in wireless application protocol (WAP), and associated technologies, and indicated how this trend can be adopted to disseminate marine information. This paper has then outlined an original scenario that the author had hypothesized, given the technology trend that was being observed in 2001 - and has concluded by giving real world examples of similar scenarios currently in effect.

This paper is not intended to provide an in-depth technological review of the wireless application protocol (WAP) or of advances in mobile applications. It is intended to provide an overview of trends in wireless technology and how they can be used to address the very real issue of disseminating information using commonly available devices to the interested public.

6.4 REFERENCES

Archey, C., 2002, Wireless Browser Standard. *Computer Technology Review*, **22**(10), pp.17.

Buckingham S., 2000, An Introduction to YES 2 WAP, http://www.gsmworld.com/technology/yes2wap.html#1,May.

Calgary Herald, 2002, *Canadian Government using Radarsat to Detect and Track Oil Spills*. October 4th.

Canada's National Programme of Action (CNPA), 2000, http://www.cc.gc.ca/marine/npa-pan/sum_e.htm. Accessed September 2000.

Canadian Centre for Marine Communications (CCMC), 2002, *Marine Information and Communications Technology News*, **13**(29), http://www.ccmc.nf.ca/ diffusion/issues/2002/issue_13,29.htm

CARIS - Universal Systems Ltd, 2003, CARIS Spatial Fusion™ 3.0 Users Guide, CARIS Spatial Components, (Fredericton: CARIS Universal Systems Ltd.).

Crowe, V., 2000, Governance of the oceans, *Short report on the Wilton Park Conference 586*, December 3–5, 1999.

Eckert, R. D., 1979, *The Enclosure of Ocean Resources: Economics and the Law of the Sea*, (Stanford: Hoover Institution Press).

Eurotechnology.com, 2003, *The unofficially independent imode FAQ*, http://www.eurotechnology.com/imode/faq-gen.html

Fitzgerald J., 2000, *CARIS Spatial Fusion: an Internet GIS. White Paper #7*, http://www.spatialcomponents.com/techpapers/fusion.pdf, (Fredericton: CARIS Universal Systems Ltd.).

Francica, J., 2002, SPECIAL REPORT: Microsoft & AT&T Wireless join forces in providing Location-based Services, *Directions Magazine*, July 31, 2002,

Friedheim, R.L., 1999, Ocean governance at the millennium: Where we have been - where should we go. *Ocean & Coastal Management*, **42**, pp. 747–765, http://www.directionsmag.com/article.php?article_id=238.

Geoplace.com, 2003, *Web Mapping Guide – Vendors*, http://www.geoplace.com/gr/webmapping/vendors.asp, Accessed July 2003.

Gomes, G., 1998, *The review of the NOAA National Coastal Zone Management Programs for Estuary and Coastal Wetland Protection*, http://www.oceansconservation.com/iczm/gomes.htm, Accessed September 2000.

Juillet, L. and J. Roy, 1999, Investing in people: the public service in an information age (reflections on the 1999 APEX conference), *Optimum*, **29** (2/3).

Konish, N., 2002, Hope rests on Location-based Services. *Wireless Systems Design*, **7**(10), pp.9.

Lunn, R., 2001, *The Use of Automatic Vehicle Locator (AVL) Technology at the City of Fredericton's Police Department*, Unpublished Report, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada.

Lutz, E. and M. Munasinghe, 1994, Environmental accounting and valuation in the marine sector. In *Ocean Governance: Sustainable development of the Seas*, edited by Payoyo, P.B., (Tokyo, New York, Paris: United Nations University Press).

Manning, E. *et al.*, 1998, Renovating governance: lessons from sustainable development. *Optimum*, **28**(3), pp. 27–35.

Miles, E.L., 1998, *The concept of ocean governance: Evolution towards the 21st century and the principle of sustainable ocean use*. In Proceedings of the SEAPOL workshop on ocean governance and system compliance in the Asia-Pacific context. Rayong, Thailand, 1997.

Mobile Applications Initiative, 2001, *GPRS & 3G Information*. http://www.gprsworld.com, May.

Mobile Lifestreams Ltd, 2000, *Data on 3G - An Introduction to the Third Generation*. http://www.mobileipworld.com/wp/positioning.htm, November.

Ng'ang'a S.M, A. Campos, M. Sutherland and S.E. Nichols, 2001. *Supporting Coastal and Marine Monitoring by Remotely Accessing Data Using Spatial Fusion and WAP*. In Proceedings of the 4th International Symposium on

Computer Mapping and GIS for Coastal Zone Management: CoastGIS Conference, Halifax, Nova Scotia, June 2001.

Nichols, S. and D. Monahan, 1999, Fuzzy Boundaries in a Sea of Uncertainties, The Coastal Cadastre - Onland, Offshore. In *Proceedings of the New Zealand Institute of Surveyors Annual Meeting*, Bay of Islands, NZ, Oct 9-15, pp. 33–43.

Nichols, S., Monahan, D., and Sutherland, M.D., 2000, Good governance of Canada's offshore and coastal zone: Towards an understanding of the maritime boundary issues. *Geomatica*, **54** (4), pp. 415–424.

Nickerson B.G., Shan, Y., and McLellan, J.F., 2003, Web Access to Real Time Wireless Mobile Geospatial Information. *Geomatica*, **57** (1), pp. 13–29.

Obeid, D., 2000, *How WAP Works*, Oracle Publishing Online Article. http://www.oracle.com/oramag/webcolumns/2000/index.html?howwapworks.html

Payoyo, P.B., 1994, Editor's introduction. In *Ocean Governance: Sustainable development of the Seas*, edited by Payoyo, P.B., (Tokyo, New York, Paris: United Nations University Press).

Prescott, J.R.V., 1985, *The Maritime Political Boundaries of the World*, (London, New York: Methuen).

Pinto, N.O., 2001, *UMTS*. Exame Digital nº4, Lisboa, Portugal.

Rosell, S.A., 1999, *Renewing governance: Governing by learning in the information age*, (Oxford: Oxford University Press).

Salvo, R., 2000, *Dossier Telecomonicações*, Fortunas & Negocios nº104, Lisboa, Portugal.

Sanger, C., 1987, Ordering *the Oceans: The making of the law of the sea*, (Toronto: University of Toronto Press).

Sweeney, D, 2002, LBS seeking its day in the sun. *Telecom Asia*, **13**(6), pp.44.

United States Environmental Protection Agency, 2003, *Interactive Web Mapping*. http://www.epa.gov/ceisweb1/ceishome/atlas/learngeog/interactivewebmapping.html

Villalobos, L., 2000, *Como aquecer o primeiro lugar*, Fortunas & Negocios nº104, Lisboa, Portugal.

WAPArch, 1998, WAP Forum 1998, *Wireless Application Protocol Specifications*, http://www.wapforum.org/what/technical.htm

WAP Forum, 2000, Wireless Application Protocol White Paper, *Wireless Internet Today 2000*, June, see http://www.wapforum.org/what/whitepapers.htm

Webmapper.co, 2000, CARIS Spatial Fusion product review. http://www.web-mapper.com/product/CARIS/spatial_fusion/review.html

WebMapper.com, 2003, Web Mapping Products. http://www.web-mapper.com/company/showproducts.cfm, Accessed July 2003.

CHAPTER SEVEN

Exploring the Optimum Spatial Resolution for Satellite Imagery: A Coastal Area Case Study

Chul-sue Hwang and Cha Yong Ku

7.1 INTRODUCTION

Researchers in geography are much interested in issues of scale and resolution, and the strong influence these may have on the results of analyses of spatial patterns (Harvey, 1968; Stone, 1972; Goodchild and Quattrochi, 1997). Issues of scale usually arise from not knowing the optimal unit size for a study area; while MAUP (the Modifiable Area Unit Problem), one of the fundamental research problems in geographical information science, can also be considered to be related to the understanding of the scale for spatial data (Openshaw, 1984; Fotheringham and Wong, 1991).

GIS and remote sensing have shown active interests in both the operational scale and measurement scale of data. Technological developments in GIS and remote sensing have led to the gradual detailing and elaboration of the measurement scale. In the field of remote sensing, for example, the 78m resolution Multi-Spectral Scanner (MSS) data of the 1970s has given way to 30m Thematic Mapper (TM) data in the 1980s, SPOT in the 1990s, and 1m IKONOS imagery in the 21st century. Such improvements in the measurement scale will allow for the new understanding of geographical phenomena that could not be measured at the previous scales.

Changes of the ground resolution of satellite imagery have modified the inherent characteristics of imagery due to the "scale effect" (Goodchild and Quattrochi, 1997). Therefore, researchers have to select the image with the optimum resolution among the satellite imagery at diverse resolutions, at which a specific geographical phenomenon can be understood more effectively. In order to explore the optimum resolution, the objective procedures or indices must be designed for the image characteristics at different resolutions. They should be able to represent the changes in image characteristics that occur at different resolutions.

The aim of the present study is to propose a procedure for identifying the optimum resolution for the study of coastal wetlands. These wetlands, as the point where the ecosystem of the land meets the ecosystem of the ocean, have unique

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

and diverse topographical features, giving rise to the formation of vegetation colonies, and exhibits diverse spatial distribution due to diverse environmental factors. Due to the dense vegetation and unstable topography of the coastal wetland, it is more effective to use satellite imagery to understand its characteristics, rather than through field investigation (Lyon and McCarthy, 1995). However, it is not necessarily effective to use the satellite imagery of higher ground resolution in the coastal area. For example, in the case of the salt marsh where colonies appear as clusters, the use of low-resolution imagery of the approximate size of the clusters is more effective (McNairn *et al.*, 1993).

In this vein, this study will attempt to find procedures and techniques for the optimal spatial resolution of satellite imagery in the study of coastal wetlands. To achieve this goal, it firstly selected the indices considering the following conditions: (1) the sensitivity of the indices, which change in response to the different resolutions, should be high enough to respond accurately to the changes in image characteristics; (2) the indices should be the ones that can provide the approximate estimate of the final result only with the minimal procedure; (3) the indices should reflect the classification accuracy of the attribute information. We then coarsened the resolution of the satellite data, to produce images at various resolutions, and observed the changes in the textural and attribute information characteristics of these images at various resolutions. As a result, the study was able to decide the operational scale, an optimum resolution in the use of satellite imagery.

7.2 SATELLITE IMAGERY CHARACTERISTICS BY RESOLUTION

7.2.1 Textural Characteristics of Satellite Imagery

The local variance method is frequently used in the search for the optimum resolution. Proposed by Woodcock and Strahler (1987), the local variance method measures the variance by placing a 3 x 3 moving window over the image data, to capture the textural characteristics of the image. However, local variance is limited by the fact that it relies on the overall variance of the image. In other words, the local variance measurements of one image cannot be compared to the results for another image. In addition, using this index makes it impossible to understand the spatial distribution of the variance within the image.

Cao (1992), in an attempt to calculate the fractal dimension of satellite imagery, compared and analysed the Isarithm method (Goodchild, 1980), the Variogram method (Mark and Aronson, 1984) and the Triangular prism method (Clarke, 1986). From the analysis, he found the Isarithm method most useful. In general, the fractal dimension *D* of satellite imagery appears as a real number between the 2D and 3D. As the complexity of the image's spatial structure increases, the value of the fractal dimension also gets higher. Within this principle, obtaining the fractal dimension gives a numerical representation of the satellite imagery characteristics by resolution.

In order to understand how the distribution characteristics of images change during the coarsening of resolution, the images of variances as well as the images of averages are needed. It is similar logic that we use the variance to make up the weak points in using the arithmetic mean (or average) to represent a data set. The variance of an image set provides a measure of the average squared deviation of a set of values around the mean. While the images of averages represent the image as one pixel, taking the arithmetic mean of the original pixels, the images of variances represent the dispersion characteristics or the variability of the original image during the coarsening of the resolution (Figure 7.1). During the resolution coarsening, if the heterogeneity between the pixels is large, the variance of the generated image will be large. Conversely, if the pixels are of similar value, the variance of the generated image will be small. Low variances are recorded for smooth and wide regions during the resolution coarsening, and high variances are recorded for narrow and complex regions.

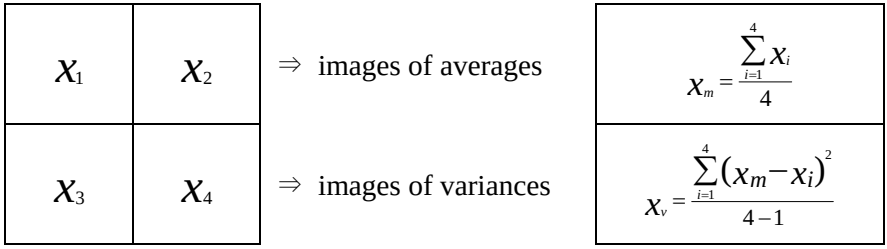


Figure 7.1. Calculating images of averages and images of variances.

Understanding the spatial distribution pattern of the images of variances derived from the process of the resolution transformation leads to the analysis of changes in image characteristics by resolutions. Therefore, the analysis of the spatial distribution of the images of variances can be used to understand the image transformation that follows the resolution coarsening. The spatial autocorrelation of the images of variances is selected to measure the spatial distribution in which the image characteristics transform. The spatial autocorrelation using Moran's *I* is generally appropriate to measure the distribution pattern for data with a continuous variable such as satellite imagery.

7.2.2 Measuring the Spectral Separability of the Training Set

Divergence is a common measurement to estimate the statistical separability of satellite imagery. Divergence is measured using the average of the statistics by class and the covariance matrix gathered through the supervised classification of the training set. However, since divergence has a tendency to increase abruptly as the distance between the classes increases, it is difficult to generalise divergence (Jensen, 1996).

The Jeffreys-Matusita distance (Matusita, 1966) is a value calculated by the exponential transformation of the Bhattacharyya distance that has been modified from divergence (Kailath, 1967; Wacker, 1971). This J-M distance is more widely

used than divergence for calculating the spectral separability of multi-variate classified classes because the J-M distance does not exhibit the abrupt increase in the spectral separability. However, complex calculations and a long processing time are some of the disadvantages of the J-M distance.

Mahalanobis distance (Mahalanobis, 1936) is a value that takes into account both the correlation and distribution features between the data in terms of Euclidean distances. Mahalanobis distance multiplies the Euclidean distance by the covariance matrix between classes. In contrast to the J-M distance, Mahalanobis distance requires simpler calculations and less processing time. In addition, it does not exhibit the abrupt increases in the spectral separability. For these reasons, Mahalanobis distance is the most widely used method for measuring the spectral separability.

7.3 PROCEDURE FOR EXPLORING THE OPTIMUM SPATIAL RESOLUTION

The optimum spatial resolution using satellite imagery is the one that provides good representation of the attribute information, and is referred to as the 'operational scale.' In order to select the image with the optimum resolution from satellite images at different resolutions, the objective procedures or indices must be designed for the image characteristics at different resolutions.

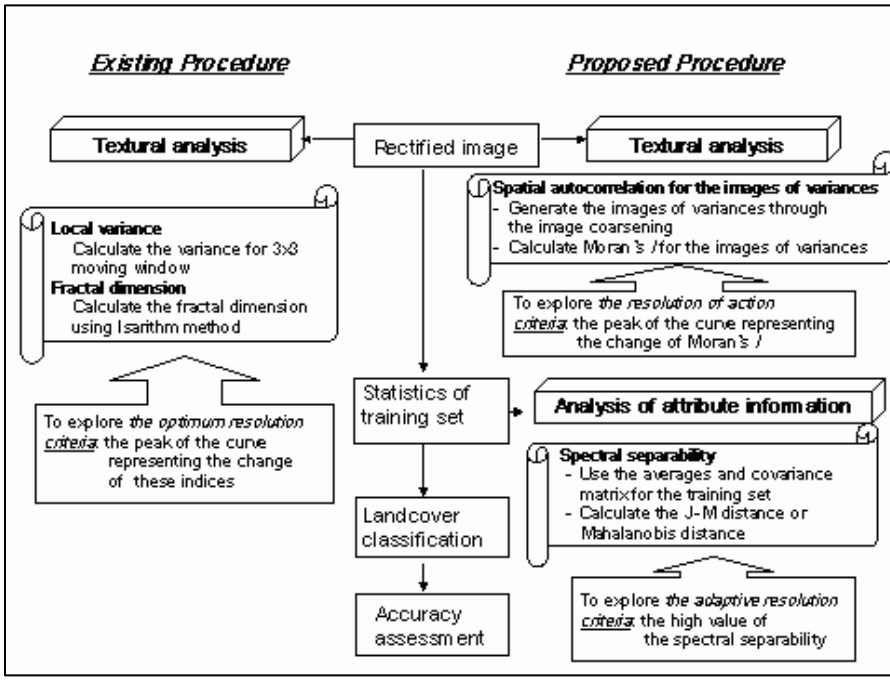


Figure 7.2 The proposed procedure to explore the optimum spatial resolution of satellite imagery.

However, it is unreasonable to rely on a single index to select the optimum spatial resolution. Indices have their own specific objectives to understand the characteristics of the imagery. Therefore, this study will attempt to combine the various indices and present a procedure for exploring the optimum spatial resolution (Figure 7.2).

Until now, existing studies on selecting the optimum resolution have focused only on the textural changes of the images at different resolutions. Local variance and fractal methods are particularly mentioned in such studies. However, because they only reflect the result without illustrating processing procedure of attribute information obtained from the images, they should not be used to understand the changes in attribute information.

This study will attempt to approach the exploration of the optimum resolution using satellite imagery using a two-step procedure. First, through the measurement of the image's textural characteristics, the resolution of action will be selected, and then, through the estimation of the spectral separability, the adaptive resolution for the attribute information will be chosen.

7.3.1 Textural Characteristics: Searching for the Resolution of Action

If the resolution of the original image is gradually scaled down to produce images at low resolutions, the image achieves a structure much like that of a pyramid. This method of the scale transformation is called the pyramid model (Figure 7.3). When producing an image at a low resolution, producing both images of averages and images of variances at the same time is more effective for the analysis of the satellite imagery. The images of variances are the ones that reflect the differences between the original images and the coarsened images during the resolution coarsening.

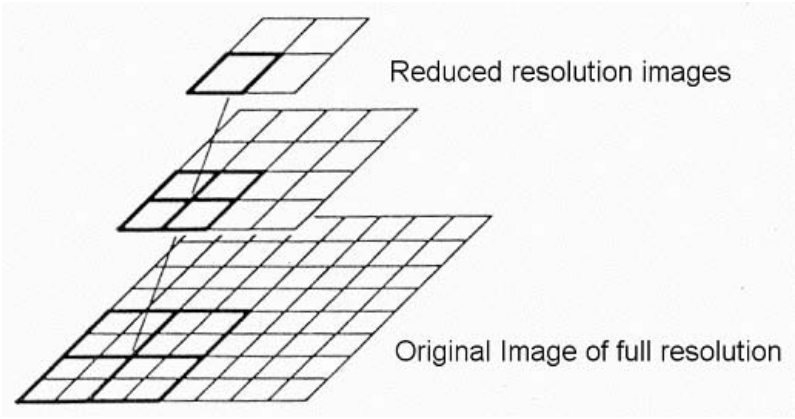


Figure 7.3 Construction of an image pyramid (source: Richards, 1993, p. 33).

- *Step 1*: Moran's *I* is calculated based on the images of variances. It can be said that the higher the Moran's *I* value, the higher the spatial autocorrelation will be and the spatial distribution pattern will then

concentrate. The concentrated spatial distribution pattern for the images of variances means that there is a large transformation in some regions following the resolution coarsening.

- *Step 2*: Graph the Moran's *I* values measured at each resolution.
- *Step 3*: Find out the changes in the Moran's *I* value to select the resolution of action. The resolution at the apex of the curve on the graph is the one at which the image information changes. In other words, the image information changes with respect to this resolution (i.e. the resolution at which the satellite image characteristics are changing) and lowers the classification accuracy.

7.3.2 Attribute Information: Searching for the Adaptive Resolution

In searching for the optimum spatial resolution to acquire the proper attribute information, it is effective to apply the measurement of the spectral separability of the training set (Markham and Townshend, 1981). The J-M distance and Mahalanobis distance are the indices that give precise representations of the spectral separability. The resolution where the J-M and Mahalanobis distances are high and the accuracy of the image classification is also high is the one that can best acquire the attribute information of the image. It can be called the *adaptive resolution*.

- *Step 1*: Select the training set in the satellite image.
- *Step 2*: Obtain the statistical calculations (average, standard deviation, minimum and maximum value, covariance between bands, etc.) for the image training set at each resolution and save them as a signature file.
- *Step 3*: Measure the spectral separability, J-M distance and Mahalanobis distance from the signature file.
- *Step 4*: Graph the J-M distance and Mahalanobis distance measured at each resolution.
- *Step 5*: Find out the changes in the J-M distance and Mahalanobis distance to select the resolution that best represents the image's attribute information. The point on the graph at which the indices abruptly increase is the resolution that best represents the image's attribute information. In particular, of the resolutions that are lower than the resolution of action, the resolution at which the spectral separability increases is determined to be the one where the classification accuracy increases. In other words, this resolution can be selected as the adaptive resolution (or the optimum resolution) that best represents the attribute information of the satellite image.

7.4 APPLICATION AND EVALUATION OF THE PROPOSED PROCEDURE

7.4.1 A Case Study Area and Data

The two-stage procedure to explore the optimum spatial resolution proposed in this study was tested by its application to Landsat TM imagery, that covers Suncheon Bay, a typical coastal area in a southern part of Korea (Figure 7.4). Suncheon Bay area has various land cover types such as coastal wetland (i.e., salt marshes, tidal flats, waterways, etc.), urban areas, agricultural areas, and forest.

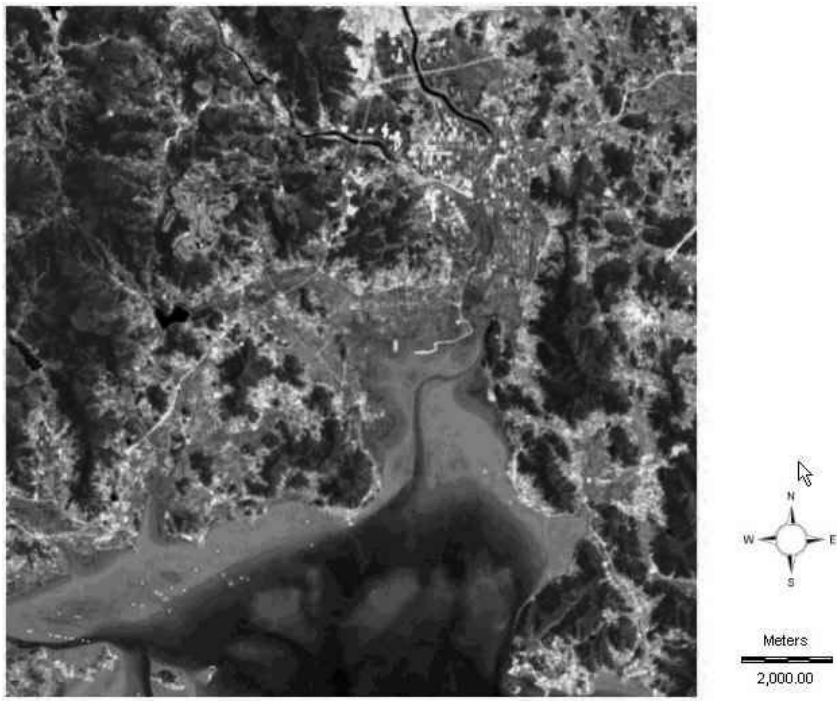


Figure 7.4 Landsat TM image of Suncheon Bay (original in colour).

7.4.2 Applying the Steps Proposed

7.4.2.1 Image Coarsening

To coarsen the resolution, the averages and variances of the adjacent pixels were calculated and then these values were used to produce the images of averages and variances. The images of variances were used to measure textural characteristics through the spatial autocorrelation analysis, and the images of averages were used to measure spectral separability. In this study, the 30m image was coarsened in 30m intervals up to a resolution of 480m, for the production of a total of 16 images (Figure 7.5).

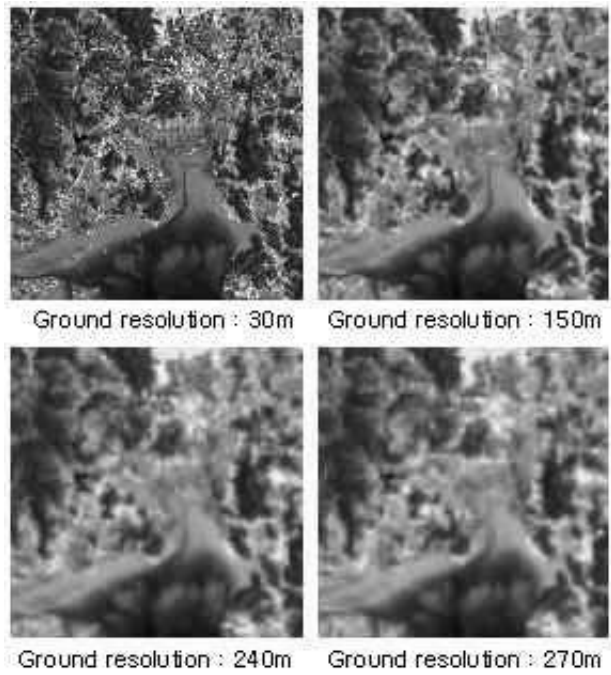


Figure 7.5 Images coarsened into various resolutions.

7.4.2.2 Step 1: Measuring the Textural Characteristics

By measuring the spatial autocorrelation of the images of variances, the following characteristics were discovered (Figure 7.6). In most bands, the Moran's *I* value gradually decreases to a resolution of 120m and then starts to increase at a resolution of 150m. However, this value then begins to decrease again at a resolution of 180m. In other words, the spatial autocorrelation rises abruptly at a resolution of 150m and then falls at later resolutions. Similar phenomena can be seen at resolutions of 300m and 390m. The increase in the Moran's *I* value means

that the spatial distribution of the variance is clustered at a certain region. It indicates that there is a large fluctuation in the specific region during the resolution transformation process. As follows, the point at which the Moran's *I* value increases and then decreases can be understood to be the resolution at which the variance drift is changing. In other words, the resolutions of 150m, 300m and 390m are the points at which a change is occurring in the spatial distribution pattern of the images of variances.

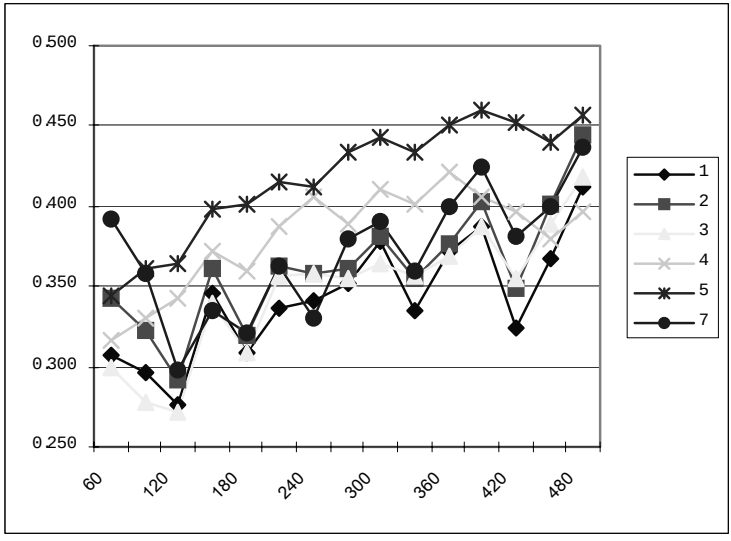


Figure 7.6 The changes of spatial autocorrelation by 7 bands.

In the case of the image of Suncheon Bay area, it can be said that a significant spatial change has occurred in the image as the resolutions changed from 120m to 150m and from 270m to 300m. Therefore, the resolutions of 150m and 300m are the points at which the image's textural characteristics change and can be interpreted as being the resolutions of action. Why does this happen? It is because there exists salt marsh in the study area that has a maximum breadth of 120m with its cluster having the maximum breadth of 280m. Although the image up to a resolution of 120m, which is the maximum size of the salt marsh, can reveal the features of the study area, it is difficult to portray the changes in image characteristics at lower resolutions. At a lower resolution, the salt marsh is portrayed as a cluster, but at a resolution of 300m, which is larger than the size of the cluster, the characteristics can no longer be displayed. It can be said that the characteristics of this study area are reflected in the spatial autocorrelation changes of the images of variances.

7.4.2.3 Step 2: Measuring the Spectral Separability

The land covering types of Suncheon Bay were classified into 6 groups. They were water body, tidal flat, forest, agricultural area, salt marsh and urban area. In addition, two training areas were selected from each land cover type. The information about training sets was saved into the vector formatted file allowing for the same application of the resolution image. The basic statistical information of the training set and the covariance between bands was calculated by every image at resolutions. Based on it, the J-M distance and Mahalanobis distance were measured.

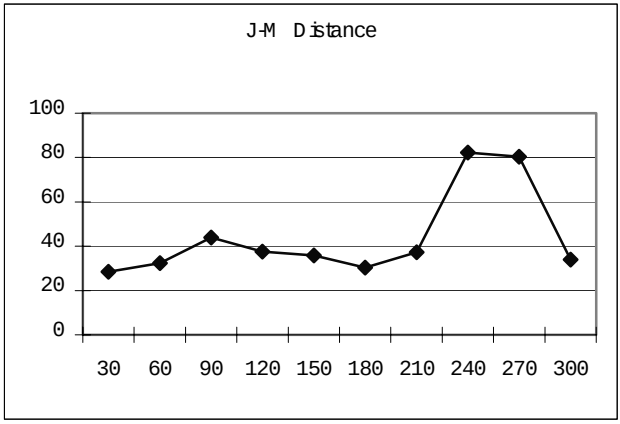


Figure 7.7 The changes of the J-M distance at various resolutions**.**

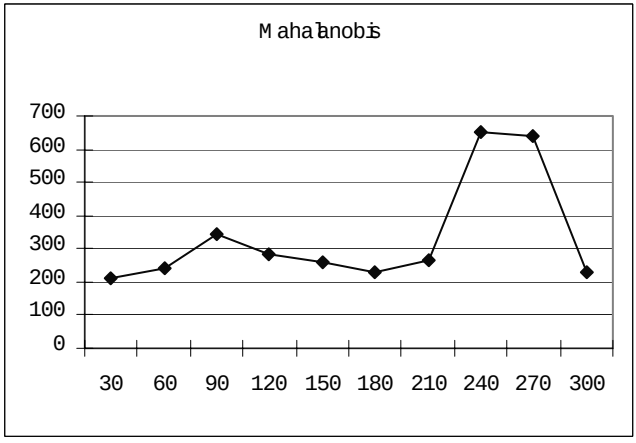


Figure 7.8 The changes of Mahalanobis distance at various resolutions.

Figures 7.7 and 7.8 show the changes in the J-M distance and Mahalanobis distance at each resolution, respectively. It is found that the spectral separability does not change up until a resolution of 150m, which was determined to be the resolution of action in the previous step, but that changes in spectral separability become evident at lower resolutions. In particular, the spectral separability increases at the resolutions of 240m and 270m. At these resolutions, the attribute information is clearly classified making these resolutions the adaptive resolutions (or the optimum resolutions) for representing the attribute changes of the study image.

7.4.3 Evaluation of the Procedure Proposed

7.4.3.1 Comparing Classification Accuracies of Images at Different Resolutions

In order to determine if the resolution of action and the adaptive resolution accurately reflected the image characteristics, the image of the study area was classified by land cover types and the classification results were then evaluated (Figure 7.9).

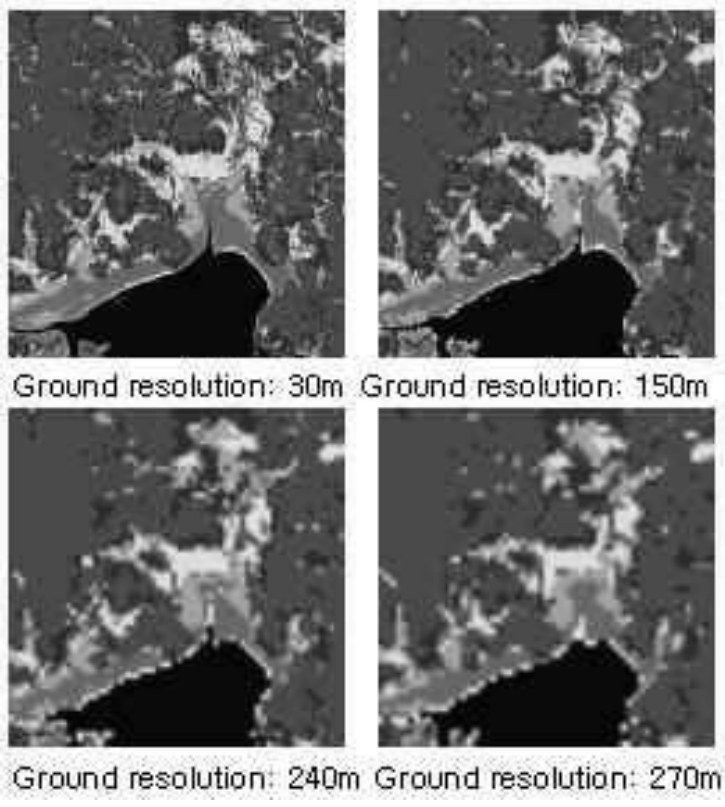


Figure 7.9 The land cover classification at various resolutions.

For the purposes of classification accuracy analysis, the topographic map, forest map and aerial photos of Suncheon Bay area were used to prepare a reference map in Arc/Info vector coverage format, which can be converted into raster data with various resolutions. The reference map had 4 classes, which were water body, tidal flat, agricultural area and forest. In order to measure the image classification accuracy by resolutions, a variety of resolution images were compared to the raster reference data for each resolution. An overlay analysis was conducted on the images at each resolution and the overall accuracies and Kappa Indices were calculated (Figures 7.10 and 7.11).

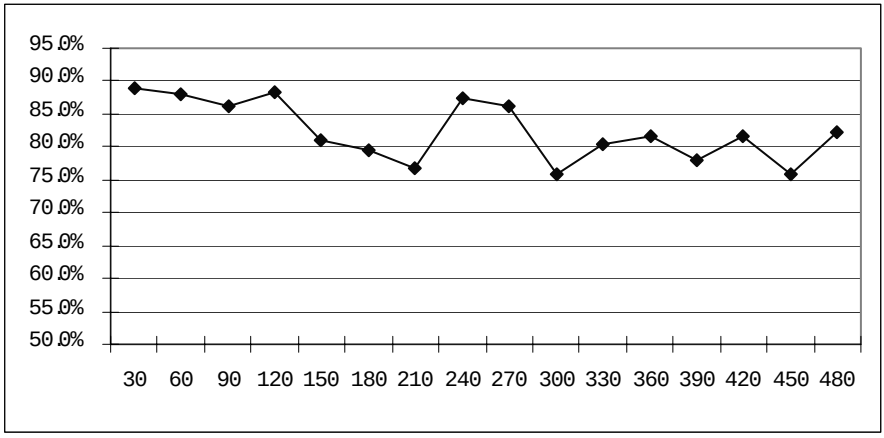


Figure 7.10 Changes of accuracy by spatial resolution.

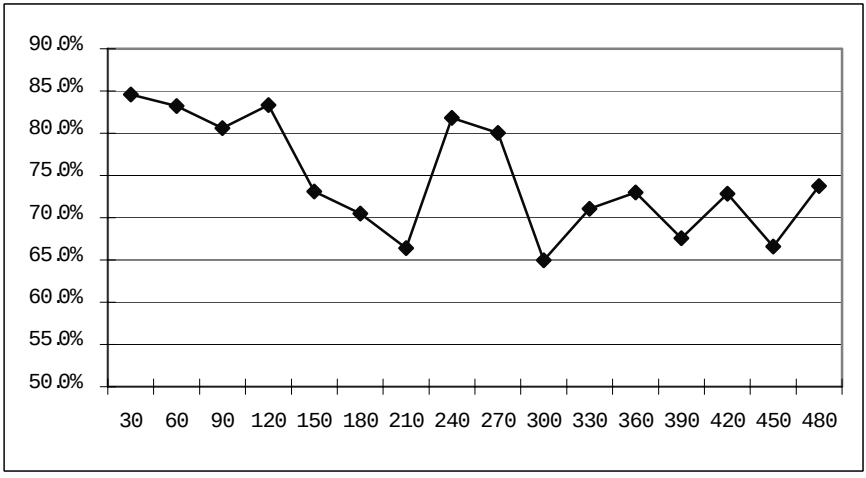


Figure 7.11 Changes of Kappa index by spatial resolution.

The changes of image classification accuracy by resolution were compared to the image characteristics revealed by the procedure for exploring the optimum resolution. The following results were discovered. First of all, there are sharp drops in accuracy at the resolutions of 150m and 300m. These resolutions could be the resolutions of action. They were obtained from the measurement of the spatial autocorrelation of the images of variances (see Figures 7.6 and 7.11). This result is expected to provide the basis for revealing the usefulness of the spatial autocorrelation measurement for the images of variances. Second, the classification accuracy increases at the resolutions of 240m and 270m, the adaptive resolutions to represent attribute information properly, which were selected using the method of the spectral separability measurement (see Figures 7.7, 7.8 and 7.11). In other words, the classification accuracy improves at the adaptive resolutions, which are found at resolutions lower than the resolution of action. In this vein, it can be said that the J-M distance and Mahalanobis distance can be used in the procedure to explore the optimum resolutions.

7.4.3.2 Correlation Analysis Between the Accuracy and the Index

Next, in order to evaluate the indices that reflect the image characteristics by resolutions, the correlation between these indices and the classification accuracies were analysed. Spearman's rank correlation analysis was more suitable than Pearson's product moment correlation analysis in this case because of a small amount of data.

Table 7.1 shows the correlation analysis between the 3 indices, measuring the textural indices of satellite imagery, i.e. local variance, fractal dimension and spatial autocorrelation of images of variances respectively, against the classification accuracy by resolutions.

Table 7.1 Spearman's rank correlation coefficient between the textural indices and the classification accuracies

	Local Variance	Fractal Dimension	Autocorrelation of the images of variances
Correlation	-0.273	-0.200	-0.717

Compared to the low correlations of the classification accuracies with the local variance and fractal dimension, the spatial autocorrelation index exhibits a higher negative correlation. The negative value means that the classification accuracy declines as the spatial autocorrelation index increases. For instance, according to the earlier analysis in this study, the sharp decline in classification accuracy seen at the resolution at the apex of the spatial autocorrelation is a direct reflection of this negative correlation (see Figure 7.6).

Table 7.2 shows the correlation analysis between the 3 methods for measuring the spectral separability, divergence, the J-M distance and Mahalanobis distance respectively, against classification accuracy.

Table 7.2 Spearman's rank correlation coefficient between the indices of the spectral separability and the classification accuracy

	J-M distance	Mahalanobis distance	Divergence
Correlation	0.636	0.588	0.042

Whereas the J-M distance and Mahalanobis distance show a relatively high positive correlation with the classification accuracy, the correlation of the divergence is very low. In other words, with the correlations near 60%, the J-M distance and Mahalanobis distance can be accepted as the indices of the spectral separability. However, for divergence, it does not represent the changes in attribute information by resolutions as an index of the spectral separability.

7.5 CONCLUSION

Analyses of geographical or spatial phenomena can yield different results depending on the spatial data scale selected by a researcher. The scale is also an important variable in research in remote sensing and geographical information sciences, which calls upon quantitative methodologies for the understanding and analysis of spatial processes and patterns. In particular, the spatial resolution determines the scale in such studies that utilise remotely sensed data. This resolution is a decisive factor in the interpretation of the geographical information contained in the satellite imagery.

The aim of the present study was to investigate the proper procedure to select the optimum resolution for satellite imagery. In the first step, the textural characteristics of the original image were analysed and the resolution of action was explored. At this point, Moran's *I* for the spatial autocorrelation of the images of variances was used. The resolution of action, the resolution at which the image characteristics change, was the one where the image's classification accuracy dropped sharply. By exploring the image's resolution of action, the resolution boundaries to maintain the image characteristics were found.

In the second step, the adaptive resolution was explored in order to extract the image's attribute information accurately. In this step, the indices of the spectral separability for the training set were used. It was ascertained that the J-M distance and Mahalanobis distance were the good indices for estimating the accuracy of attribute information. Therefore, the resolution where the J-M distance and Mahalanobis distance were high was the one with the high image classification accuracy.

The procedures for exploring the optimum spatial resolution proposed in this investigation have the following advantages over existing methods: First of all, whereas the procedure to select the optimum resolution in existing methods is vague, the method proposed here searches for the resolution of action first and

then offers clearer criteria. Second, it allows the exploration of the resolution that clearly represents the attributes. It can be said that the resolution at which the spectral separability increases is the one where the image's attribute information accuracy is high.

Lastly, this research must be followed by studies of more diverse land covering types, and images with various sensors. Further studies should be able to provide generalised rules that will serve as a clearer foundation for selecting the optimum resolution.

7.6 REFERENCES

Atkinson, P.M., 2001, Geostatistical regularization in remote sensing. In *Modelling Scale in Geographical Information Science*, edited by Tate, J.N. and Atkinson, P.M. (Chichester: John Wiley & Sons), pp. 237-260.

Bailey, A. and Gatrell, A., 1995, *Interactive Spatial Data Analysis*, (Essex: Longman).

Cao, C., 1992, Detecting the scale and resolution effects in remote sensing and GIS, Unpublished Ph.D. dissertation, Louisiana State University.

Cao, C. and Lam, N., 1997, Understanding the scale and resolution effects in remote sensing and GIS. In *Scale in Remote Sensing and GIS*, edited by Quattrochi, D.A. and Goodchild, M.F. (Boca Raton: CRC Press), pp. 57-72.

Clarke, K., 1986, Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. *Computers and Geosciences*, 12, pp. 713-722.

Cressie, N., 1993, *Statistical Analysis of Spatial Data*, (New York: Wiley).

Duda, R.O. and Hart, P.E., 1973, *Pattern Classification and Scene Analysis*, (New York: Wiley).

Fotheringham, A.S., Brunsdon, C. and Charlton, M., 2001, Scale issues and geographically weighted regression. In *Modelling Scale in Geographical Information Science*, edited by Tate, J.N. and Atkinson, P.M. (Chichester: John Wiley & Sons), pp. 123-140.

Fotheringham, A.S. and Wong, D., 1991, The modifiable area unit problem in multivariate statistical analysis, *Environment and Planning A*, 23, pp. 1025-1044.

Friedl, M.A., 1997, Examining the effects of sensor resolution and sub-pixel heterogeneity on spectral vegetation indices: Implications for biophysical modeling. In *Scale in Remote Sensing and GIS*, edited by Quattrochi, D.A. and Goodchild, M.F. (Boca Raton: CRC Press), pp. 113-140.

Goodchild, M.F., 1980, Fractals and the accuracy of geographical measures. *Mathematical Geology*, 12, pp. 85–98.

Goodchild, M.F. and Quattrochi, D.A., 1997, Introduction: Scale, multiscaling, remote sensing, and GIS. In *Scale in Remote Sensing and GIS*, edited by Quattrochi, D.A. and Goodchild, M.F. (Boca Raton: CRC Press), pp. 1-12.

Griffith, D.A., 1987, *Spatial Autocorrelation: A Primer*, (Washinton, DC: Asscociation of American Geographers).

Harvey, D., 1969, *Explanation in Geography*, (London: Edward Arnold).

Hwang, C.S., 1999, Alternative methods for assessments of DEMs' errors. *Journal of the Korean Society for Geo-Spatial Information System*, 7(2), pp. 23-34. (in Korean).

Hwang, C.S., Hwang, S.Y. and Lee, J.Y., 2000, Spatial data analysis and GIS. *Jirihakchong*, 28, pp. 1-12. (in Korean).

Jeffreys, H., 1948, *Theory of Probability*, (Oxford: Oxford University Press).

Jensen, J.R., 1996, *Introductory Digital Image Processing: A Remote Sensing Perspective*, (Englewood Cliffs: Prentice Hall).

Jupp, D.L.B., Strahler, A.H. and Woodcock, C.E., 1988, Autocorrelation and regularization in digital images: Basic theory. *IEEE Transactions on Geoscience and Remote Sensing*, 26, pp. 463-473.

Kailath, T., 1967, The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Theory*, COM-15, pp. 52-60.

Lam, N., 1990, Description and measurement of Landsat TM images using fractals. *Photogrammetric Engineering and Remote Sensing*, 56, pp. 187-195.

Lam, N. and De Cola, L., 1993, *Fractals in Geography*, (Englewood Cliffs: Prentice Hall).

Lam, N. and Quattrochi, D.A., 1992, On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Professional Geographer*, 44, pp. 88-98.

Lyon, J.G. and McCarthy, J. (eds.), 1995, *Wetland and Environmental Applications of GIS*, (Boca Raton: CRC Press).

Mahalanobis, P.C., 1936, On the generalised distance in statistics. In *Proceedings of the National Institute of Science in India*, Part A2, pp. 49-55.

Mark, D.M. and Aronson, P.B., 1984, Scale-dependent fractal dimensions of topographic surfaces: An empirical investigation with applications in geomorphology and computer mapping, *Mathematical Geology*, 11, pp. 671-684.

Markham, B.L. and Townshend, J.R.G., 1981, Land cover classification accuracy as a function of sensor spatial resolution. In *Proceedings of the 15th International Symposium on Remote Sensing of Environment*, pp. 1075-1090.

Matusita, K., 1966, A distance and related statistics in multivariate analysis. In *Multivariate Analysis*, edited by Krishnaiah, P.R. (New York: Academic Press), pp. 187-200.

McNairn, H., Protz, R. and Duke, C., 1993, Scale and remotely sensed data for change detection in the James Bay, Ontario, coastal wetlands. *Canadian Journal of Remote Sensing*, 19, pp. 45-49.

Moellering, H. and Tobler, W., 1972, Geographical variances. *Geographical Analysis*, 4, pp. 34-64.

Openshaw, S., 1984, *The Modifiable Areal Unit Problem*, Concepts and Techniques in Modern Geography 38, (Norwich: Geo Books).

Quattrochi, D.A., Emerson, C.W., Lam, N.S. and Qiu, H., 2001, Fractal characterization of multitemporal remote sensing data. In *Modelling Scale in Geographical Information Science*, edited by Tate, J.N. and Atkinson, P.M. (Chichester: John Wiley & Sons), pp. 13-34.

Park, E.J., Seong, J.C. and Hwang, C.S., 2003, A study of drought susceptibility on cropland using Landsat ETM+ imagery. *Korean Journal of Remote Sensing*, 19, pp. 107-115.

Stone, K.H., 1972, A geographer's strength: The multiple-scale approach. *The Journal of Geography*, 71, pp. 354-362.

Wacker, A.G., 1971, The minimum distance approach to classification, Ph.D. dissertation, Purdue University at West Lafayette.

Weigel, S.J., 1996, Scale, resolution and resampling: Representation and analysis of remotely sensed landscapes across scale in geographic information systems, Unpublished Ph.D. dissertation, Louisiana State University.

Woodcock, C.E. and Strahler, A.H., 1987, The factor of scale in remote sensing. *Remote Sensing of Environment*, 21, pp. 311-332.

Xia, Z.G. and Clarke, K.C., 1997, Approaches to scaling of geo-spatial data. In *Scale in Remote Sensing and GIS*, edited by Quattrochi, D.A. and Goodchild, M.F. (Boca Raton: CRC Press), pp. 309-360.

CHAPTER EIGHT

Visualisation for Coastal Zone Management

Simon R. Jude, Andrew P. Jones and Julian E. Andrews

8.1 INTRODUCTION

8.1.1 Coastal Management in the United Kingdom

The United Kingdom has an extensive coastline of over 12,000km in length that is formed by a number of environmental processes and today is subjected to a range of natural and anthropogenic pressures. Coastal management in the UK is underpinned by the development and implementation of Shoreline Management Plans (SMPs). Introduced in 1995 to provide long-term sustainable coastal defence policies and management objectives for sediment cells or sub-cells, SMPs are developed through co-operative discussions between the numerous organisations involved in managing the coastline (Purnell, 1996; Potts, 1999). SMPs can encompass a range of management options including 'do nothing', 'hold the line' of existing defences, 'advance the line' of existing defences or 'retreat the line' (Ash *et al.*, 1996). However, whilst they define the long-term management objectives, individual management schemes remain subject to economic and environmental appraisal as and when they are proposed.

Amongst the key challenges facing coastal zone managers are the need to widen public consultation and strengthen public participation during the selection of management options, and the requirement to improve the information dissemination process once decisions have been made. Shoreline Management Plans are complicated documents for those without prior technical knowledge of coastal processes, and the method in which they are prepared has been criticised for lacking adequate scope for public participation. It has been argued that this has led to suspicion amongst local communities regarding the beneficiaries of the plans (O'Riordan and Ward, 1999). Traditionally, SMPs and environmental and economic appraisals have only been disseminated to a limited number of organisations and interested individuals. The dissemination has generally been paper-based, and two-dimensional maps have been used to illustrate the plans. The

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

government realises that wider access to information contained within SMPs will be required in the future if they are to gain support for the plans, as the policies outlined in the SMP must be seen to be acceptable to the general public (O'Riordan and Ward, 1999; Potts, 1999). Indeed, a recent government review of the SMP process identified the difficulties associated with facilitating public participation as being very significant (MAFF, 2000). In line with this, the review called for innovative new communication techniques to be developed and incorporated into future SMP documents and dissemination programmes (MAFF, 2000).

The problems found in the United Kingdom are mirrored elsewhere. For example, the European Union Demonstration Programme on Integrated Coastal Zone Management recently noted that stakeholders should be more involved in the development and implementation of coastal management plans (CEC, 2000a). This view is now reflected in recent EU recommendations promoting participatory planning in coastal management and encouragement to develop systems that allow the monitoring and dissemination of coastal zone information (CEC, 2000b).

There is a clear need for the further development of new methodologies that will help enable interested individuals and organisations to be informed of shoreline management decisions in the most inclusive manner possible (Belfiore, 2000; King, 1999). Indeed, King (1999) has specifically called for the use of *electronic* methods to facilitate communication between coastal managers and the public, whilst many others have highlighted the need for research to exploit the potential of GIS in educating, promoting and involving the public in coastal planning and decision-making (Bartlett and Wright, 2000). Certainly, traditional GIS packages are already widely used by organisations involved in coastal management and these systems are frequently cited as one of the tools associated with best practice (e.g. Bartlett, 1994). However, GIS does not provide a universal solution despite its potential for assisting informed decision-making (O'Regan, 1996; Bartlett, 2000). Ultimately a traditional GIS and its output are oriented towards experts with knowledge of complicated terminology, as opposed to the layperson who often has the most to lose from management decisions. These limitations are compounded by the fact that coastal decision-makers are themselves often overwhelmed by the complexity of many GIS applications (Green, 1995). Consequently, the GIS based coastal management systems that have been developed are often simply employed to produce thematic maps of coastal areas for SMPs, and much of the potential of the technology remains unrealised.

8.1.2 VRGIS - a Possible Solution?

One technology with the potential to widen communication in shoreline management planning is Virtual Reality GIS (VRGIS). A VRGIS is in many aspects similar to a traditional GIS, but it encompasses Virtual Reality visualisations as a key output and interaction method. The virtual reality (VR) aspect of VRGIS has evolved mainly as an interface technology within which user interaction issues are of key importance. The more traditional GIS acts as a data storage and manipulation technology.

The important role of visualisation in environmental decision-support has been recorded by a number of authors who have highlighted the need to develop

such techniques to assist in the public presentation of complex environmental process models (Bishop, 1994; Bishop and Karadaglis, 1997). The recent development of VRGIS provides an opportunity to further develop public involvement in coastal zone management by providing the functionality to produce realistic virtual reality visualisations of different shoreline management outcomes. These may prove to be a significant advance on traditional methodologies. Using a case study of the north Norfolk coast in eastern England, this article reports on a research project that is developing an integrated VRGIS methodology for the assessment, visualisation and public communication of the environmental impacts of several proposed real-world management schemes.

8.2 METHODOLOGY

8.2.1 The Case Study Area – The North Norfolk Coastline

The north Norfolk coast is a relatively undeveloped low-lying barrier coastline that began to form in its current state around 6,000 to 7,000 years ago (Andrews et al., 2000). Because of its relatively undeveloped nature, the coastline has high scientific, economic and recreational value, reflected by the whole zone being protected by national and international legislation. Management of the coastline is complicated, with numerous statutory and non-statutory bodies involved in overseeing a wide range of sites including a number of nature reserves. The coastline has been studied widely and benefits from an extensive monitoring programme managed by the UK Environment Agency. The development of a first generation SMP began in 1993 and was published in 1996 (Environment Agency et al., 1996).

The SMP covers a very large area. Therefore, a number of smaller project-level study sites were identified through consultation with a range of statutory and non-statutory organisations involved in managing the coastline. In order to illustrate this work, a single scheme at Brancaster West Marshes is described here (Figure 8.1, see colour insert following page 164).

8.2.2 Brancaster West Realignment Scheme

With sea level predicted to rise by up to 88cm by 2100 (Houghton et al., 2001) there is considerable concern regarding the potential for future active management of the coastline because of its vulnerability to North Sea storm surges (Thumerer et al., 2000). A possible option to accommodate future rises includes allowing reclaimed freshwater marshes to revert back to their natural state, a process known as managed retreat, setback, or coastal realignment. Coastal realignment has triggered considerable concern and debate amongst the public (Clayton, 1995), although the European Habitats Directive does require such schemes to offset habitat losses by creating new habitats elsewhere along the coast.

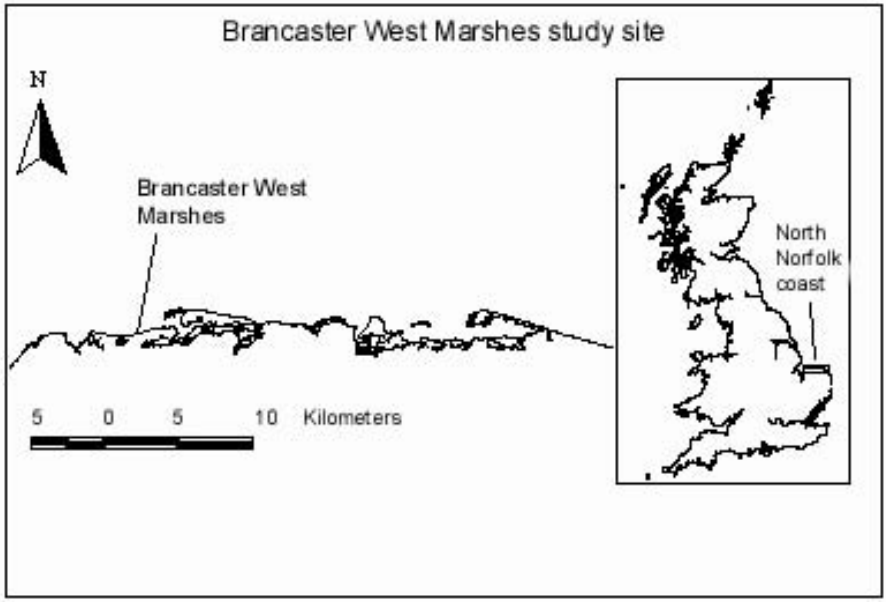


Figure 8.1 The location of the Brancaster West Marshes study site.

Brancaster West Marshes is a site currently under consideration for coastal realignment. The Marshes comprise approximately 40ha of freshwater grazing meadows forming a Site of Special Scientific Interest (SSSI) and a Special Protection Area (SPA) under the European Union Birds Directive (Tyrrell and Dixon, 2000). The site is flanked by earth flood embankments with its frontage protected by defences strengthening the natural dune frontage. The latter were constructed in 1978 to provide protection against storm surges but have degraded to such an extent that the Environment Agency has proposed a managed realignment scheme in which the frontage will be removed, with a new defence constructed 300m inland from the original location (Tyrrell and Dixon, 2000). The freshwater marshes to the north of the new defence will subsequently be allowed to revert to salt marsh.

The scheme has attracted considerable attention because it impacts a site protected under the Birds Directive. Furthermore, it could potentially interfere with the defences and frontage protecting the adjacent Royal Society for the Protection of Birds (RSPB) reserve at Titchwell. Additional complications arise from the privately owned defences belonging to the Royal West Norfolk Golf Club to the east, who plan to construct their own defence to protect their practice ground in response to the scheme.

8.2.3 Database Construction

An extensive GIS database was developed using ArcInfo and ArcView GIS packages (Table 8.1). Data was obtained from organisations involved in managing the coast, and supplemented a number of Ordnance Survey products including Land-Line.Plus, and Land-Form PROFILE. Where management plan data was unavailable in a digital format it was digitised, with permission, from management documents.

Database construction illustrated the difficulties associated with integrating data from a range of different organisations. Firstly, identifying data holdings availability was time-consuming and frustrating, with some organisations wishing to charge for data conversion. Secondly variations in the GIS software used by organisations required further complex conversion into a standard format.

Table 8.1 Sources of data.

Dataset	Supplied by	Provides
Aerial photography	Norfolk County Council and Natural Environment Research Council	Colour aerial photography of the site from 1988 and 2001
CASI	Natural Environment Research Council	5m resolution Compact Airborne Spectrographic Imager image with intertidal zone classification
Geology	British Geological Survey	Solid and drift geology
Landcover Map of Great Britain	Centre for Ecology and Hydrology	25m resolution landcover grid
Land-Form PROFILE	Ordnance Survey	10m DEM grid
Land-Line.Plus	Ordnance Survey	Large-scale vector data
LIDAR	Environment Agency	2m DEM grid
Shoreline Defence Survey	Environment Agency	Flood defence location, design and condition
Shoreline Management System	Environment Agency	Coastal monitoring data including beach profiles

8.2.4 Assessment Methodology

Assessments of future changes at the study site were made using information from a range of sources. Data on the coastline's evolution was provided from results of the Natural Environment Research Council's Land-Ocean Interaction Study (LOIS) together with historical mapping and aerial photography of the site. Furthermore, beach profiles collected by the Environment Agency's Shoreline Management System for the frontage over the last 10 years were incorporated into the GIS. This historical data was complemented with assessments of the potential impacts of future sea level rise on the site. Future sea level rise was calculated using the Model for the Assessment of Greenhouse-gas Induced Climate Change (MAGICC) (Hulme et al., 1995) for each of the latest emissions scenarios from the Intergovernmental Panel on Climate Change (IPCC). The results from MAGICC exhibited considerable uncertainty in predicted rates of future sea level rise. Therefore, for the purpose of the research, the assessments were based on the IPCC's IS92a scenario using the high sea level rise predictions calculated by MAGICC (86cm by 2100). The assessment of sea level rise also accounted for isostatic changes in the land level and used mean isostatic adjustment rates as predicted for eastern England by Shennan (1989).

The estimation of how the combined sea level rise and isostatic adjustment would affect the study site was undertaken in ArcInfo using an Arc Macro Language (AML) script that calculated a new Digital Elevation Model (DEM) for each year between 2000 and 2100. Sources of elevation data employed for this included 2m resolution LIDAR data provided by the Environment Agency and Ordnance Survey Land-Form PROFILE 10m resolution products. This methodology allowed the applicability of different sea level rise scenarios to be quickly assessed before they were used for creating VR visualisations.

8.2.5 Large-scale Data Preparation

Ordnance Survey Land-Line.Plus large-scale digital topological data was used for the basis of the detailed visualisation work. It was supplied in $1km^2$ titles at 1:2,500 scale in line-only format, and thus required conversion to a polygon topology to allow attribute data to be incorporated. All editing took place in ArcInfo GIS, where the tiles covering the study site were appended and cleaned to produce a line coverage of the study site. Unfortunately the data was produced in such a way that automated polygon topology generation methods that have been applied for urban areas such as those used by Lake et al. (2000) are unsuitable. This is because not all polygons are completely closed and, although the data is provided with seed points relating to polygons, they fail to cover every land parcel. Hence, where required, polygons were closed and seed points were generated manually.

Once the polygon coverage was created, attribute data was added. Attributes were assigned to each polygon from three separate data sources. For the intertidal zone, 5m resolution Compact Airborne Spectrographic Imagery (CASI) was available from the LOIS Project. This imagery was georeferenced and converted to a grid format using ERDAS Imagine before being used to assign attributes to the

polygons. Land-based polygons falling outside the intertidal CASI coverage were classified from aerial photography and landcover data obtained from the Centre for Ecology and Hydrology Landcover Map of Great Britain. In total, 38 different landcover classes were used.

Following construction of the polygon dataset, extensive fieldwork was undertaken to both ground-truth the coverage and to collect digital photographs for use during the production of the visualisations. Plans for the realignment scheme were obtained from the Environment Agency and were digitised to create a new coverage reflecting proposed future state of the site. The Environment Agency dataset was appended to Land-Line.Plus to allow evaluations of the areas of loss to be calculated, and for the individual environmental indicators to be assessed.

8.2.6 Production of Visualisations

Visualisations of the scheme were produced using two techniques. Firstly interactive visualisations were produced using ArcView 3D Analyst Extension to provide 'fly through' Virtual Reality Modelling Language (VRML) experiences, and secondly, static visualisations were produced by exporting the GIS results into World Construction Set, a photorealistic-rendering package. Whilst the virtual fly-through is the most common output of VRGIS, the experience it provides does not necessarily equate to the way members of the public are able to best perceive landscapes. Hence the two methodologies were chosen so as to allow an assessment to be made of their respective roles in widening public understanding of future coastal management schemes. All the visualisations were created using a 1GHz AMD Athlon-based PC with 512MB RAM running Windows 2000.

8.2.6.1 ArcView 3D Analyst

The production of visualisations using ArcView 3D Analyst involved the creation of a Triangulated Irregular Network (TIN) DEM using the Land-Form PROFILE data, over which the Land-Line.Plus polygon dataset was then draped to create a 3D scene. Sea defences and buildings were added as separate coverages, allowing them to be extruded from the DEM to produce 3D surface features. This process involved obtaining sea defence height information from a database of Environment Agency defences. The absence of true height information in Land-Line.Plus meant that, for buildings, a height of 3m per storey was assumed. Sea defence heights were extruded from Ordnance Datum whilst buildings were generated using the DEM as the base elevation. For the production of visualisations of the site following the realignment scheme, the DEM was reprofiled in those areas where change was deemed likely to occur, then it was converted to a TIN. Finally the Land-Line.Plus-derived coverage representing the future site state was used as the drape.

3D Analyst allows the attributes of the resulting visualisation to be queried, and facilities enabling the user to navigate around and zoom in and out of the 3D scene in real-time are provided. For the purpose of this project, static images of the environments were created. Each scene was also exported to a VRML file for use in any Web viewer.

8.2.6.2 World Construction Set

In contrast to ArcView, World Construction Set allows photorealistic visualisations to be generated, and has the advantage for many GIS users that ASCII DEMs and ArcView shapefiles may be imported and used as the basis for these visualisations. However, one drawback, at least for UK users, is that input data must be reprojected to decimal degrees.

The first stage in the production of the visualisations involved the importation of the DEM data into the package. World Construction Set permits terrain features to be generated using Terraffectors. Terraffectors create terrain from two-dimensional vector representations. Their advantage is that a cross-section profile of a study area may be created to represent real world features in which the ability to specify the height or depth of the feature in relation to the underlying DEM is implicit. Area Terraffectors, based on vector polygons, were used to generate representations of the site's sea defences from defence heights and cross-sections as provided by the Environment Agency Shoreline Defence Survey from 1999. They were also employed to create creeks and pools. Despite its poorer vertical resolution, the smoother Land-Form PROFILE data was chosen as the underlying DEM in preference to the Environment Agency LIDAR dataset because Terraffectors were found to work better with that source.

The second stage in visualisation generation involved importing the individual ArcView coverages for which new environment effects were to be created. By splitting the original landcover classification dataset into separate ArcView shapefiles for each landcover type, different features could be assigned different ground effects. For example, grasslands were given a surface texture representing long meadow grass. Later 3D vegetation and trees were included, although the limited selection of pre-generated foliage features and 3D objects in World Construction Set required the creation of new objects from photographs of the study sites. Finally, 3D buildings were added. Their addition involved a multi-stage process where the building polygons were converted from an ArcView shapefile coverage to a DXF file, allowing a package called SoftCAD to be used to create a 3D building. This enabled them to be to converted into a 3D object that was useable in World Construction Set.

During each stage of the visualisation generation process, preview renders were used to test the result of parameter changes and, where necessary, adjustments were made to each feature. Once a satisfactory result was achieved, final rendered images were produced using a number of virtual camera locations that best represented the change that would occur at the site.

8.3 RESULTS

The results of the assessment of the impacts of the scheme revealed that, although there will be a predicted loss of 10.2ha of freshwater grazing marsh following construction, this Figure will most likely be offset by an increase in salt marsh area of 6.3ha and 1.6ha for water channels and creeks. However, what is less certain from this analysis are the impacts on biodiversity that this change may bring about.

The results of the visualisations are provided in Colour Plates 8.1 to 8.4 (following page 164). In each case the same viewing location is used to illustrate

the changes at the site following construction of the realignment scheme. The immediate difference between the ArcView 3D Analyst and World Construction Set images is the level of detail, the former being more stylised in comparison to the latter. For example the ArcView 3D Analyst images do not contain the extensive colours, textures and 3D features found in the World Construction Set images. Likewise, the limited VRML functionality in ArcView 3D Analyst results in crude representations of the defences, produced by extruding the defence shapefiles, whilst the Terraffectors available in World Construction Set rendered their detailed cross-sections. This trade-off in detail does, however, have important implications for the rendering times for each method with the ArcView 3D Analyst 3D scenes being continually updated in real-time as the scene was explored, whilst the static rendered images of World Construction Set took approximately 1 minute to render a preview image so it is likely to be some time before explorable World Construction Set environments can be produced.

8.4 DISCUSSION

One of the key criteria for the choice of study area here was that extensive coastal monitoring data was available from a range of organisations. In particular, it was very important that large-scale digital data, plus detailed management plans and information, were available for the chosen study site. Without such information, it would not have been possible to produce assessments and visualisations of future environments that would stand up to public scrutiny. However, even here there were problems in gaining access to some key documents. For example, whilst the Environment Agency Regional Office housed the Shoreline Management System GIS, the project manager for the Brancaster West scheme was based at the local office. This led to difficulties when trying to locate up-to-date plans for the scheme.

The primary drawback associated with the methodology at present is the lack of widespread availability of large-scale digital data in a suitable format. Although Land-Line.Plus represents the best available source, the time-consuming editing required to produce a suitable polygon coverage may prove extremely costly within an organisational context. Despite this limitation, once the data was finally cleaned it lent itself extremely well to use in the visualisation software. Few difficulties were encountered during the GIS analysis stage of the work although problems were identified relating to the interoperability of data between ArcView and World Construction Set. The primary difficulty came from the requirement for ArcView shapefiles to be projected in decimal degrees as opposed to the UK National Grid. Further problems were also typified by the laborious process required to create and import buildings due to the need to convert files to suitable formats. Such findings reflect a general problem relating to the incompatibility of GIS and VR software formats (Williams, 1999).

The two methods presented in this paper were chosen to represent alternative techniques that may be suited to different purposes in coastal management. The 3D Analyst visualisations are more dynamic and interactive, allowing the user to query the results. They may be more suited to those directly involved in the decision-making process and who wish to obtain quantitative information from

them. Conversely, World Construction Set provides extremely realistic static images. These may be more suited to inclusion in management documents, such as environmental appraisals, where they could be used as an alternative to photomontages. With calls for SMPs to be distributed electronically (MAFF, 2000), both techniques lend themselves well to dissemination via the Worldwide Web. World Construction Set images could be displayed as scaleable bitmaps, whilst the virtual environments from ArcView 3D Analyst can be easily converted into VRML files, enabling the user to explore the policy impacts for themselves.

The two visualisation techniques do pose a number of questions that reflect the numerous challenges facing the application of VRGIS. The concepts of visualisations and virtual environments are relatively new and as such it has been argued that, until they are more widely used, knowledge of how best to design them will be lacking (Batty et al., 1998). This leads to the question of how the public, as opposed to experts, relate to visualisations and whether simple visualisations like those produced by 3D Analyst are more effective than detailed photorealistic visualisations at conveying complicated information? There are obvious difficulties when trying to compare the applications of visualisation presented here with research using high performance graphics workstations by some other authors (Bishop and Karadaglis, 1997). However, although improvements may be made using customised software on high performance hardware, any practical application in coastal management is likely to be constrained by costs incurred in software, hardware and training. This work has illustrated that effective visualisations may be created using off-the-shelf software running on a high-powered desktop PC.

It is obvious that visualisation techniques on their own will simply produce pretty images that are of little use if not employed properly by coastal managers. VRGIS will clearly only achieve its full potential for coastal management if it is integrated into the planning process (Zube et al., 1987; Lange, 1994). It could potentially be used from the beginning of the SMP or project planning stage, assisting communication between management organisations during the development of alternative options. Later during the decision-making process VRGIS has an obvious use in public consultation and participation. However, such advances will only be achievable if the underlying planning process is opened up to the public to allow their participation, which in the United Kingdom would require major changes to coastal planning legislation. What visualisation techniques should *not* do is, as Lange (1994) points out, be used to sell a particular scheme to the public. This would be a wasted opportunity; visualisations have the potential to *promote* the discussion of alternatives in initiatives to identify the optimal solutions to coastal management problems.

Virtual Reality GIS should not be seen as an immediate and universal solution to coastal management problems because at present there is a lack of research understanding concerning the methods that can be used and the effect of the different contexts in which they may be applied (Zube et al., 1987). As Williams (1999) notes, VRGIS has the potential to enhance conventional GIS but should not be viewed as a replacement. This research is, however, investigating the potential application of the assessment methodology and visualisation techniques in the SMP and planning process. Focus groups and participation seminars have

been undertaken to gain feedback both from coastal managers and planners who may adopt these methodologies. Further surveys are being organised with members of the public, who have the most to gain from the use of visualisation techniques as a means of communication and a participatory tool. In this way it is hoped to explore the potential difficulties of applying the methodology in more detail. A second strand of the work will also assess how the visualisations could be used in choice experiments to determine if they may be used in decision-making contexts such as that of extended cost-benefit analyses. It is hoped that this project will illustrate the potential of VRGIS in participatory coastal management.

In conclusion, we believe that this research highlights the potential of GIS and VRGIS as an integrated tool to assess, visualise and potentially value future coastal landscapes through the use of high quality, yet disparate, data related to management of the coastal zone. It also illustrates how VRGIS may communicate, educate, inform and involve the public and stakeholder groups in coastal management decisions by presenting information in a recognisable and understandable format. We firmly believe that the application of VRGIS can stimulate meaningful discussion and dialogue between groups traditionally associated with conflicting opinions. Coastal Zone Management can only benefit from its application.

8.5 ACKNOWLEDGMENTS

The authors would like to thank all of the organisations that provided input into the research and the Environment Agency, Ordnance Survey and Suffolk County Council for providing access to their GIS data. The research was completed whilst the lead author was in receipt of an ESRC/NERC Interdisciplinary PhD Studentship.

8.6 REFERENCES

Andrews, J.E., Funnell, B.M., Bailiff, I., Boomer, I., Bristow, C. and Chroston, N.P., 2000, The last 10,000 years on the north Norfolk coast - a message for the future? In *Geological Society of Norfolk 50th Anniversary Jubilee Volume*, edited by Dixon, R., pp. 76-85.

Ash, J.R.V., Nunn, R. and Lawton, P.A.J., 1996, Shoreline Management Plans: A case study for the North Norfolk coast. In *Coastal Management: Putting Policy into Practice*, edited by Fleming. (Thomas Telford), pp. 318-330.

Bartlett, D., 1994, GIS and the coastal zone: Past, present and future. *AGI Notes* (UK: Association for Geographic Information), pp. 30.

Bartlett, D.J., 2000, Working on the frontiers of science: applying GIS to the coastal zone. In *Marine and Coastal Geographical Information Systems*, edited by Wright, D.J and Bartlett, D.J. (London: Taylor and Francis), pp. 11-24.

Bartlett, D.J. and Wright, D.J., 2000, Epilogue, In *Marine and Coastal Geographical Information Systems*, edited by Wright, D.J. and Bartlett, D.J. (London: Taylor and Francis), pp. 309-315.

Batty, M., Dodge, M., Doyle, S. and Smith, A., 1998, Modelling virtual environments. In *Geocomputation: A Primer*, edited by Longley, P.A., Brooks, S.M., McDonnell, R. and MacMillan, B. (Wiley), pp. 139-161.

Belfiore, S., 2000, Recent developments in coastal management in the European Union. *Ocean and Coastal Management*, **43**, pp. 123-135.

Bishop, I.D., 1994, The role of visual realism in communicating and understanding spatial change and process. In *Visualisation in Geographic Information Systems*, edited by Unwin, D. and Hearnshaw, H. (Wiley), pp. 60-64

Bishop, I.D. and Karadaglis, 1997, Linking modelling and visualisation for natural resources management. *Environment and Planning B: Planning and Design*, **24**, pp. 345-358.

Clayton, K.M., 1993, Adjustment to greenhouse gas induced sea level rise on the Norfolk Coast - a case study. In *Climate and Sea Level Rise: Observations, Projections and Implications*, edited by Warrick, R.A., Barrow, E.M. and Wigley, T.M.L. (Cambridge: Cambridge University Press), pp. 310-321.

Commission of the European Communities, 2000a, *Communication from the Commission to the Council and the European Parliament on Integrated Coastal Zone Management: A Strategy for Europe* (Brussels: CEC).

Commission of the European Communities, 2000b, *Proposal for a European Parliament and Council recommendation concerning the implementation of Integrated Coastal Zone Management in Europe* (Brussels: CEC).

Environment Agency, North Norfolk District Council, King's Lynn and West Norfolk Borough Council, Ministry of Agriculture, Fisheries and Food and English Nature, 1996, *North Norfolk Shoreline Management Plan*: *Sheringham to Snettisham Scalp* (Mouchel Associates).

Green, D.R., 1995, User access to information: A priority for estuary information systems. In *Proceedings of CoastGIS'95*, University College Cork, Ireland, pp. 35-49.

Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van de Linden, P.J. and Xiaosu, D., eds., 2001, *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge: Cambridge University Press).

Hulme, M., Raper, S.C.B. and Wigley, T.M.L., 1995, An integrated framework to address climate change (ESCAPE) and further developments of the global and regional climate modules (MAGICC). *Energy Policy*, **23**. pp. 347-355.

King, G., 1999, EC Demonstration Programme on ICZM. *Participation in the ICZM processes: Mechanisms and procedures needed* (Hyder Consulting).

Lake, I.R., Lovett, A.A., Bateman, I.J. and Day, B., 2000, Using GIS and large-scale digital data to implement hedonic pricing studies. *International Journal of Geographical Information Systems*, **14** (6), pp. 521-541.

Lange, E., 1994, Integration of computerized visual simulation and visual assessment in environmental planning. *Landscape and Urban Planning*, **30,** pp. 99-112.

Ministry of Agriculture, Fisheries and Food, 2000, *A review of shoreline management plans 1996-1999 - final report March 2000*. A report produced for the Ministry by a consortium led by the Universities of Newcastle and Portsmouth.

O'Regan, P.R., 1999, The use of contemporary information technologies for coastal research and management - a review. *Journal of Coastal Research*, **12** (1), pp. 192-204.

O'Riordan T. and Ward, R., 1997, Building trust in shoreline management: Creating participatory consultation in shoreline management plans. *Land Use Policy*, **14** (4), pp. 257-276.

Potts, J.S., 1999, The non-statutory approach to coastal defence in England and Wales: Coastal Defence Groups and Shoreline Management Plans. *Marine Policy*, **23**(4-5), pp. 479-500.

Purnell, R.G., 1996, Shoreline Management Plans: National objectives and implementation. In *Coastal Management: Putting Policy into Practice* edited by Flemming (London: Thomas Telford), pp. 5-15.

Shennan, I., 1989, Holocene crustal movements and sea level changes in Great Britain. *Journal of Quaternary Science*, **4**, pp. 77-89.

Thumerer T., Jones A.P., and Brown D., 2000a, GIS based coastal management system for climate change associated flood risk assessment on the east coast of England. *International Journal of GIS*, **14** (3), pp. 265-281.

Tyrrell, K. and Dixon, M., 2000, *Brancaster West Marsh Engineers Report*. (Environment Agency).

Williams, N.A., 1999, Four-dimensional virtual reality GIS (4D VRGIS): Research guidelines. In *Innovations in GIS 6*, edited by Gittings, B. (London: Taylor and Francis).

Zube, E.H., Simcox, D.E. and Law, C.S., 1987, Perceptual landscape simulations: History and prospect. *Landscape Journal*, **6** (1), pp. 60-80.

CHAPTER NINE

Application of a Decision Support System in the Development of a Hydrodynamic Model for a Coastal Area

Roberto Mayerle and Fernando Toro

9.1 INTRODUCTION

Recent advances in numerical modelling of physical processes and field survey technology nowadays allow the development of numerical models with extensive data sets. As a consequence, model developers are facing new challenges to handle the increasing amount of data and its analysis.

Furthermore, model development and application concerning coastal areas are heavy time demanding tasks that need tools to assist the researcher. Most of the time they involve analysis of field measured data and its comparison with numerical model outputs. Even though there are good tools for specific tasks, there is a lack of them for integrating measured and modelled data.

An application of a Decision Support System (DSS) in the development of a hydrodynamic model for a coastal area is presented. Description of the DSS components, their interaction, and its GIS capabilities to handle spatial data are also given.

9.2 DEVELOPMENT OF HYDRODYNAMIC MODELS

A hydrodynamic model is the mathematical representation, with the assistance of numerical algorithms, of flow and wave situations to study them more effectively. However to obtain reliable information, development of numerical models must include evaluation of results. This process is time-consuming and involves comparison of numerical model outputs with measured data. Moreover, a unique model development procedure can be restrictive due to the large amount of variables involved in the physical process. The number of variables and their levels of significance depend on the objective of the study. In spite of these difficulties a general trend for the development of hydrodynamic models can be defined and is presented next.

Figure 9.1 shows the phases and steps in the development of a hydrodynamic model. The model development consists of several phases (i.e. sensitivity analysis, calibration, validation and application) in which the modelling experts fine-tune

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

the numerical and physical parameters to reproduce the field conditions. Each of these phases has several steps that share information with the other phases.

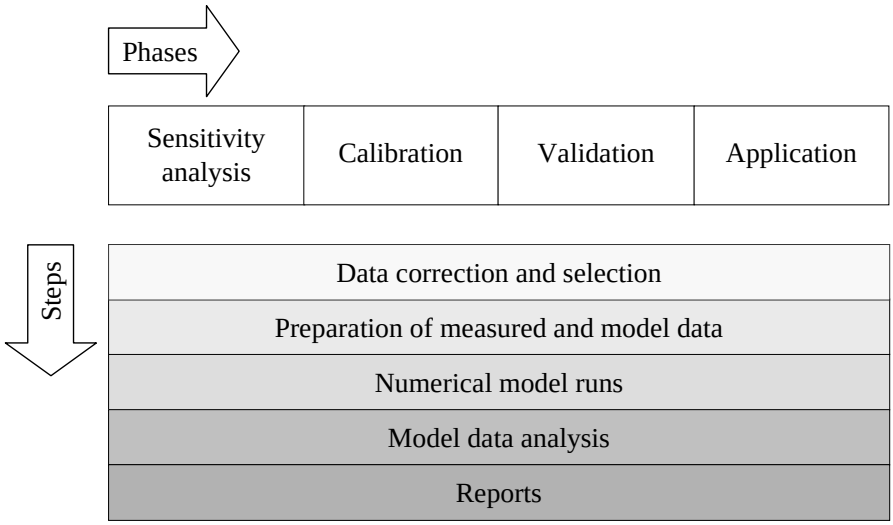


Figure 9.1: Phases and steps in the development of hydrodynamic models

The "*sensitivity analysis*" evaluates the response of the model to changes of the physical and numerical parameters. The numerical parameters are set and the physical parameters having an effect on the results are selected. Physical parameters are adjusted to represent the conditions in the studying area during the "*calibration*". During the "*validation*" the model ability in reproducing the field conditions, within certain level of accuracy, is checked. The model "*application*" is the last phase in the model development and is performed by the end users to make decisions. The validated model is then used in hindcasting, nowcasting and forecasting situations.

Regardless of the model development phase, each of them has several steps as indicated in Figure 9.1. The first one is the "*data correction and selection*" in which the measured data is corrected, selected and stored in a database. This is usually data that needs to be processed to fit the model formats. In situ measurements are in the form of time series at a certain location or recordings at a certain area for a given time. The "*preparation of measured and model data*" includes the creation of the model grid and definition of some default (but logical) numerical and physical parameters. The measured data preparation includes reformatting of the data to be used as initial conditions and boundary conditions for the model and to enhance easy comparison with the model results. "*Numerical model runs*" is the core of each phase and refer to determination of the change in the numerical model outputs caused by changes in input data. "*Model data analysis*" is the step in which the model results are analysed by means of the experience and judgement of the model developer, and compared with other model results or measured data. The data analysis is an iterative process with the model runs and gives feedback to the model developer whether or not to choose new

parameters to run the model again. This is a procedure in which software tools are needed to speed up and facilitate model development. Therefore, a "*Decision Support System*" (DSS) that assists the model developers is of vital importance. And finally, the "*reports*" document the work that has been done and summarises the experience and conclusions obtained. These written reports are used as reference in the future phases of the model development.

9.3 A DECISION SUPPORT SYSTEM FOR MODEL DEVELOPMENT

The handling of information for hydrodynamic model development is a difficult task as a consequence of different sources, types and formats of the data and the increasing amount of data involved in such activities nowadays. An efficient way to retrieve, analyse and store information accurately and speedily is crucial for the good performance in the model development process.

A Decision Support System (DSS) for development of hydrodynamic models is an interactive computer-based system intended to help modellers use data from measurements and models to simplify and speed up model creation. It helps to retrieve, summarise and analyse data. It comprises three integrated components: "*interface*", "*database*" and "*models*". Figure 9.2 shows the components of a DSS for development of hydrodynamic models, with the input and output data.

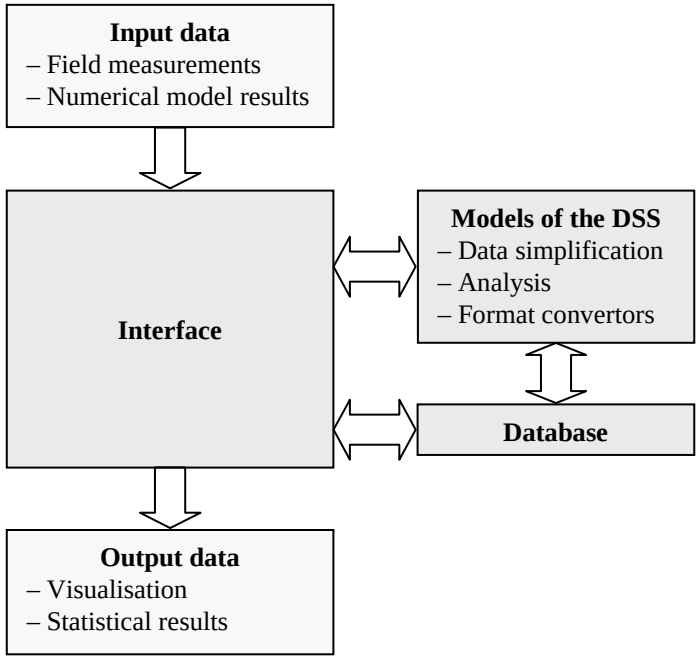


Figure 9.2 Components, input and output data of the DSS

The proposed DSS centres on the model development of depth-integrated (2DH) flow models with field data applied to coastal areas. The DSS has been tailored to handle primarily data such as water levels and current velocities.

The input data to the DSS, from measurements and numerical model outputs, is fed to the system through the interface and is stored in the database component of the system. The GIS application, as a DSS generator, is used to handling spatial data in conjunction with the database. The instructions to the system are given by the decision-maker through the interface that passes the type of analysis to be made by the model components of the DSS. Finally, the output data is presented in the form of statistical tables or visualised in plots and Figures.

The "*interface*" component is the part of the DSS that interacts with the user. It should be user-friendly and usually includes graphical components and it is then called a "*Graphical User Interface*" (GUI). The front end of the system is a graphical interface created customising ArcView 3.2a GIS software (ESRI, 1996).

The "*database*" is the collection of data that is organised so that its contents can easily be accessed, managed and updated. The size and effort spent in the database depend on the amount of available information. It is the data source for the DSS that processes and presents the information. The importance of the database in the system relies on the versatility for the selection of the data for analysis rather than in the storage capacity. In the case of the DSS, its capabilities are implemented using the GIS software generator, with tables connected under the concept of relational database.

The "*models of the DSS*" use the data provided by the database and allow the user a direct opportunity to explore the data relevant to the problem. The model components of the DSS take care of evaluating, analysing and correlating the information from different sources available in the database and reduce the information to make it easy for the user to understand the physical process. The distinction between the models of the DSS and the numerical models should be made clear here. Whereas the latter ones are software tools used to solve partial differential equations for the hydrodynamics in coastal areas, models of the DSS are software tools to link and analyse numeric output and in-situ measurements. The model components are mainly programmed as Matlab (The MathWorks, 2000) functions grouped in a software library that makes the DSS a versatile system open for new types of data analyses, new types of data sources and makes their maintenance easier.

Figure 9.3 makes a parallel between the model development steps in which the DSS assists the researcher and the DSS components. The DSS interface is used in all steps of model development and even in earlier stages during the data acquisition to visualise information. The database component is used all along the development steps and it is queried according to the model development phase that is being analysed. The data simplification, analysis and conversion are performed by the DSS models and is made during the data preparation to run the model and during the evaluation of the numerical model outputs.

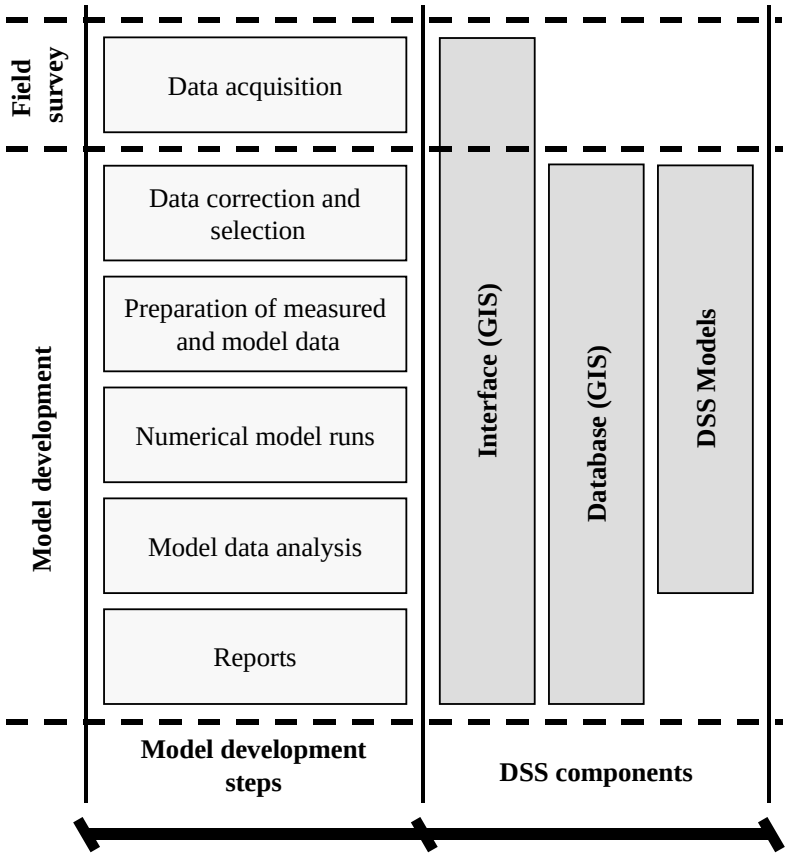


Figure 9.3 Model development phases and DSS components

9.4 APPLICATION

This section presents the application of the DSS during the development of a hydrodynamic model for an area on the German North Sea coast. The multiple possibilities of the DSS for data handling and its versatility to the needs of the user are shown.

The in situ measurements presented herein were collected as part of the project "Predictions of Medium Scale Morphodynamics" (Promorph) (Zielke et al., 1998) and were used for the development of the flow model (Palacio et al., 2003).

The domain of interest is the Meldorf Bight and the adjacent tidal channels covering an area of about 600km^2. Around 50% of the area of investigation is dry during the ebb phase and almost the whole domain is submerged during the flood phase. Figure 9.4 shows the area of investigation, model boundaries and location of in-situ measurements. The tidal channels have a maximum depth of about 20m. Tidal range is around 3.0 to 3.5m. The calibration of the flow model is carried out

using measured water levels at several gauge stations and velocity measurements over several cross-sections obtained by means of acoustical profilers.

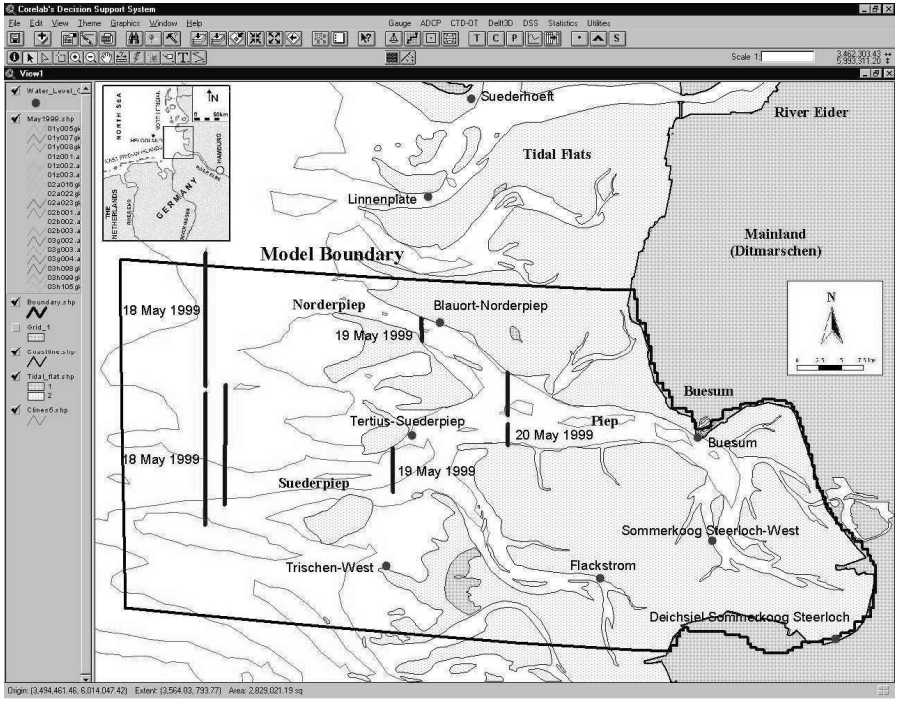


Figure 9.4 Area of investigation, model boundaries and in situ measurements

9.4.1 Data Acquisition

The acquisition of field data describing the model domain is necessary for the model development. Data may be collected from the relevant authorities, research institutes or measured in field campaigns. Here the following data sets were used:

- bathymetry in the tidal channels and tidal flat areas,
- continuously recorded water level measurements from gauge stations,
- atmospheric pressure and wind data,
- current velocity measurements.

Figure 9.4 shows the GUI for the study area with location of the water level gauges and some of the cross-sections selected for current velocity measurements.

An example of measured current velocity along a cross-section is shown in Figure 9.5. This type of plot can be animated when having several cross-section plots in a period of time.

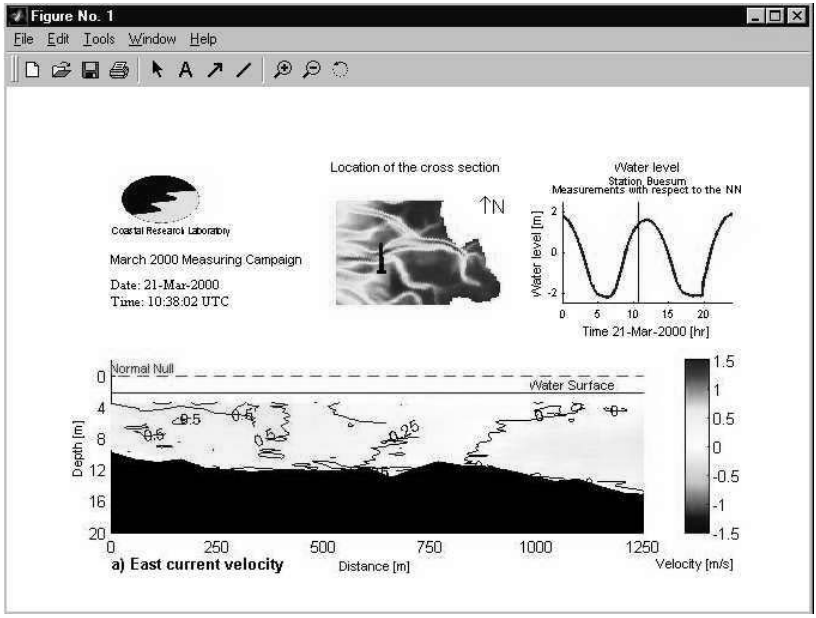


Figure 9.5 Measured current velocity transversal to a cross-section

9.4.2 Data Correction and Selection

Among others, the following corrections are incorporated with the collected measurements:

- the European time used during the recordings is referenced to UTC,
- referenced locations using differential GPS are projected onto the same reference system of the model grid,
- the depth during the measurements is usually referred with respect to the water surface. Due to the changes in the water surface in time accompanying the tide, all data is corrected and referred with respect to a fixed reference level (e.g. mean sea level) to allow comparison of data at different periods of time.

In addition to data correction and projection, some additional data is generated for later use with numerical model outputs; that is for the calculation of the depth-integrated velocity out of the 3D velocity measurements.

In situ measurements are stored in the DSS database and can be accessed by the user. The database tables are used in conjunction with the GIS map layers in such a way that the database can be queried selecting objects in the map views.

Data once entered in the DSS database are available for visualisation and selection for comparison and analysis during model development. The first step in the data selection is to display an inventory of the data showing which in situ measurements and the covered measuring period are available. The time interval considered for model development can be chosen as the time period for which simultaneous measurements of several of the required quantities are available. Or, on the other hand, it gives an idea of the missing field data in the selected interval.

Figure 9.6 shows the in situ measurements available in 2000 in the Meldorf area. The vertical axis shows the available magnitudes in terms of water levels and current velocities and the horizontal axis corresponds to the time axis. Such a Figure enhances detection of the time intervals when data sets are available for the model development. The periods selected for the model development are indicated in the Figure.

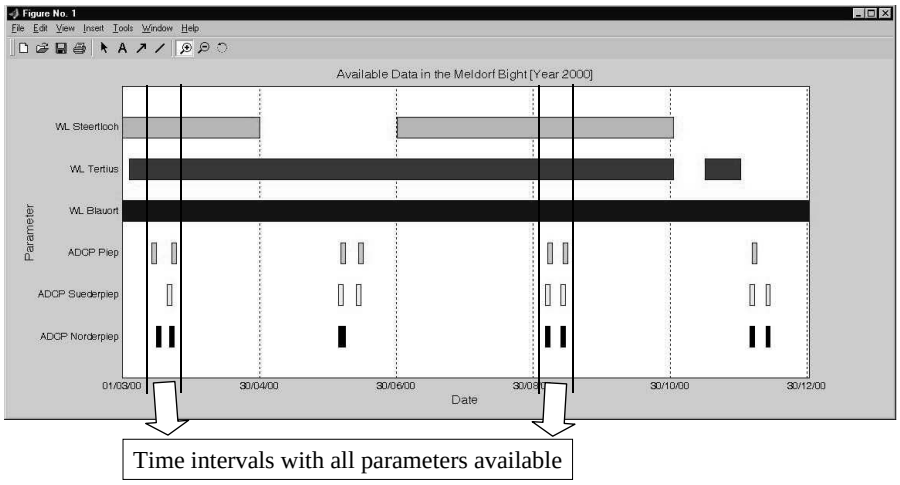


Figure 9.6 Available data in the domain area.

9.4.3 Preparation of Measured and Model Data

The model set-up, grid generation and bathymetry interpolation are done with the software tools of the hydrodynamic model.

A curvilinear grid, adjusted to the bathymetry of the domain, was created, covering an area of 600km^2, with around 35,000 cells and a resolution from 60m (inner part of the model close to the main tidal channels) to 180m (in areas close to the model boundaries). The model has three open boundaries (North, South and West). Along these boundaries the boundary conditions are specified as a function of imposed water levels. Since the bathymetry in the area was obtained from measurements done at different periods, one of the functionalities of the DSS is the selection of the appropriate bathymetry from different data sets. Besides it could also be used in updating the model bathymetry regularly.

The selection of the simulation periods is based on the intervals with available data sets that are going to be used as initial and boundary conditions for the model. These periods can be selected among the intervals shown in Figure 9.6. In the calibration and validation phases additional in situ measurements are needed for the comparison with the model results.

9.4.4 Numerical Model Runs and Model Data Analysis

Subsequently, the DSS assists the modeller in the statistical analysis and comparison of the measured data and/or the numerical model outputs depending on the model development phase being performed, i.e. sensitivity analysis, calibration, validation or application.

9.4.4.1 Sensitivity Analysis

Here the behaviour of the numerical model outputs with respect to the physical and numerical parameters is checked. The aim is to define the numerical parameters and to Figure out the physical parameters to be considered in the calibration of the model. The analysis is done at several locations all over the domain and periods of time that represent the main flow conditions; for instance, in case of tidal dominated areas, tidal phases and tidal cycles as well as periods with calm winds and during storms are considered.

For the Meldorf Bight Model the evaluation was made analysing the computed water levels and current velocities. For the comparison of the modelled water levels, the DSS makes an analysis of the variation of "*amplitude/phase*" with respect to several physical and numerical parameters. The differences in height (amplitude) and time (phase) between the peaks of two time series are computed. This analysis is used for harmonic magnitudes that have a cyclic behaviour, e.g. when evaluating water level data sets. The analysis of the water levels combines data from several locations and periods of time.

Figure 9.7 shows the GUI with a table that summarises the effect in the water levels for different time steps in the model set-up at several locations. The table includes for each location (record) and parameter (field) the number of records, mean, standard deviation and RMS-Error. The aim is to identify an appropriate time step that gives sufficient accuracy in the model results in order to minimise the computational time of the simulation. Different time steps were considered. A period of 30 days was used in the analysis to consider the effect of spring and neap tide. The selected time step is 0.50 min.

For the analysis of the data and the statistical computations the data can be organised in many different data sets. Figure 9.8 shows the DSS interface with the windows, in which the conditions for statistical analysis and selection of the data sets are set.

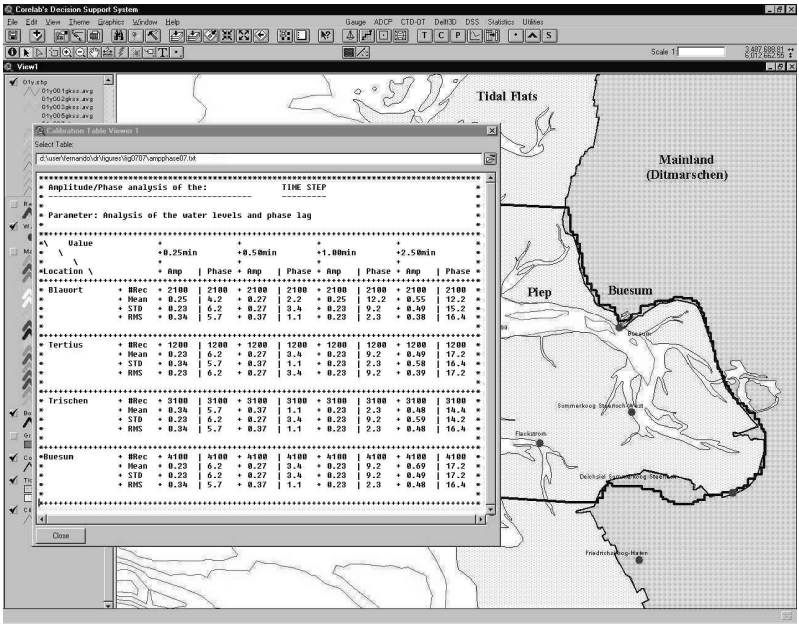


Figure 9.7 Summary of the sensitivity analysis of the time step

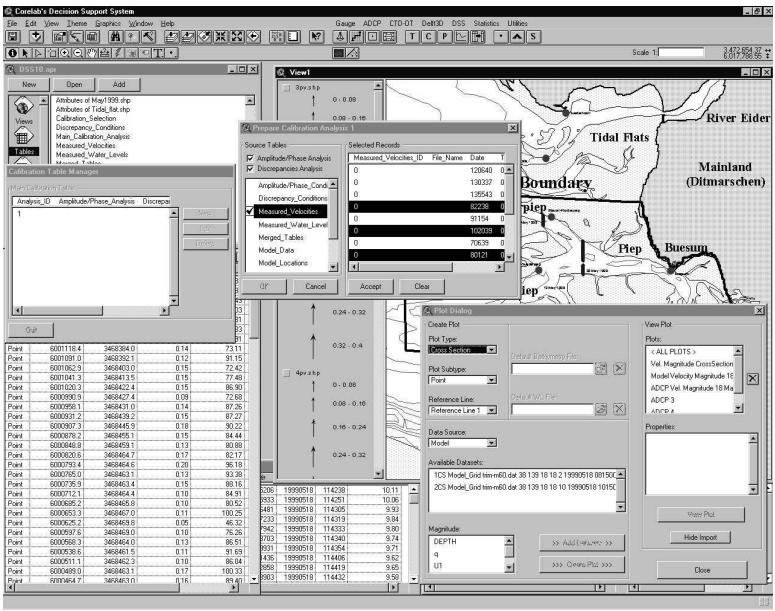


Figure 9.8 DSS interface with input data windows for the statistical analysis

9.4.4.2 Model Calibration

Physical parameters are adjusted to represent the conditions in the studying area during the model calibration. The evaluation was made analysing water levels and flow velocities. Only the calibration of the bottom roughness in terms of Chezy coefficients in the whole domain is presented. Bottom roughness with constant values of Chezy coefficient were considered.

For the comparison of the flow velocities, the DSS makes a "*discrepancies analysis.*" The *discrepancies analysis* computes the differences between the model results and measured data, and performs some statistics with these differences. This analysis can be performed organising the data by locations or velocity intervals (data sets are grouped in several classes according to the velocity magnitude from the minimum to the maximum velocities available in the data sets).

Figure 9.9 summarises the statistics of the analysis of depth integrated velocity for these three types of bottom roughness and several intervals of flow velocity. The table includes some values such as mean, standard deviation and value of the parameters RMS-Error and Relative Mean Absolute Error (RMAE) (van Rijn, 2001) for the flow velocity differences (modelled - measured) used in the analysis. The best option is obtained on the basis of these statistical parameters. In this case there are not significant differences between the three options; however, there are slightly better results with a Chezy coefficient of 60 $m^{0.5}/s$ for the bottom roughness.

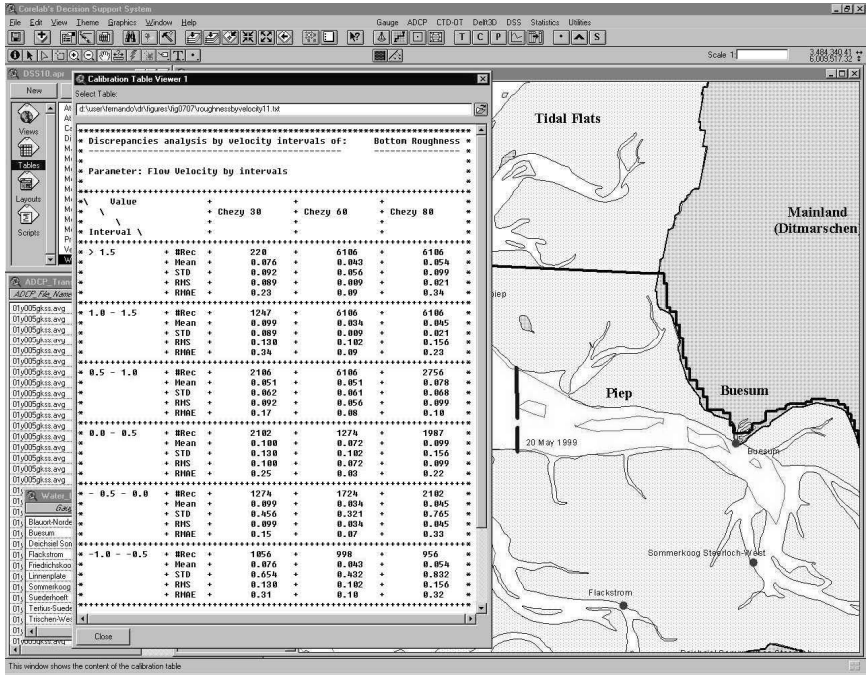


Figure 9.9 Summary of the model calibration of the bottom roughness.

9.4.4.3 Model Validation

During model validation, the model performance to reproduce the field measurements is determined. As in the model calibration, measured and modelled data are compared and therefore the same DSS tools can be applied. The data sets used in the validation are different from those used in the model calibration. The tools presented in section 9.4.2 are employed for selection of additional measurement data.

Figure 9.10 shows a comparison of modelled and measured flow velocity, integrated over the depth, at a cross section within the model domain.

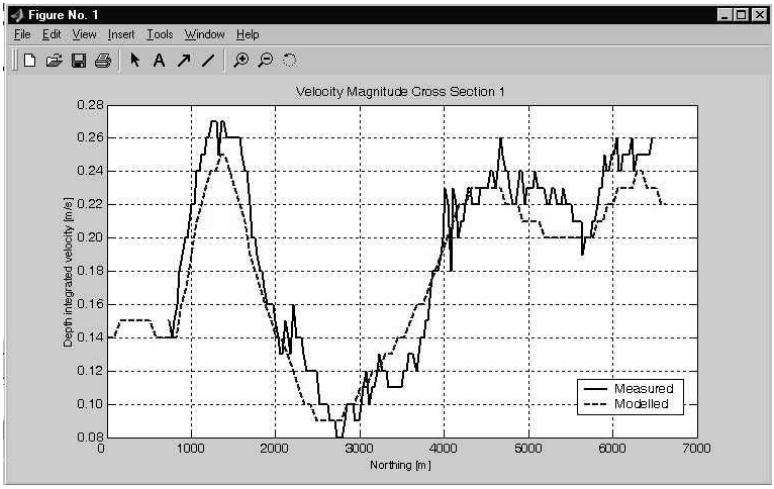


Figure 9.10 Velocities comparison during the model validation.

9.4.4.4 Application

The model, once validated, is used for forecasting flow conditions following several "if" scenarios; this allows to study a problem and possible engineering solutions (e.g. determine water levels for different wind conditions for the design of dikes). In this phase, all the DSS tools already described are available for handling existent and additional data that might be collected after the model validation.

9.4.5 Quasi-Real Time Validation

Another application of the DSS during the data acquisition might be the comparison of the measurements with the model results previously prepared. The steps involved in the quasi-real time validation (Mayerle and Toro, 2002) are as follows:

- prior to the measuring campaign, model simulations are carried out (for different parameters) for the period in question. For this, open sea boundary conditions on the basis of astronomical constituents can be used,

- the model results covering the measuring period are stored and brought to the field,
- during the cross-sectional measurements, the measured current velocities (using acoustical profilers) are transformed into depth-integrated values,
- comparisons between 2DH model results and measurements are done just after the completion of a cross-section,
- after the measuring campaign is achieved (12-13 hours covering one entire tidal cycle), the results taken at all the cross-sections and covering the entire measuring period are analysed and improvements in the model and/or the measuring strategy can be considered on the basis of the statistical analysis. Figure 9.11 shows the DSS in use covering a cross-section along one of the main tidal channels. Figure 9.11a shows the research vessel during the ADCP measurements. Figure 9.11b displays the resulting velocity field from the simulations. In Figure 9.11c a researcher using the DSS during the measurements is shown. Comparisons of modelled and measured current velocities, integrated over the depth, are shown in Figure 9.11d.

9.4.6 Reports

Usually a large number of graphs are created to visualise the model at several locations and certain periods of time. The DSS has been linked to MS Word and MS Powerpoint to summarise these graphs and document the conclusions of the work that has been done.

Figure 9.12 shows several plots of the flow velocities from the Meldorf Model during a tidal cycle already imported into a report. Animation of the behaviour of the velocity field during the tidal cycle can be created with this sequence of Figures.

9.5 CONCLUSIONS

Progress in field measurements and advances in numerical modelling are generating a large amount of data creating new challenges for its handling and analysis. This brings acutely the necessity of software tools to assist the researcher during development of hydrodynamic models.

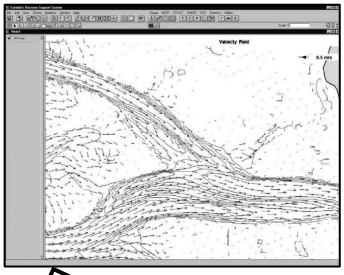
The use of GIS technology in the model development brings many new capabilities and commodities to the model developer for spatial data handling and easy interaction with a relational database. This chapter shows the application of GIS as a software generator to develop a DSS that assists the modeller in integrating measured data and model results for the development of hydrodynamic models.

The structure of the DSS gives flexibility to the researcher regardless of the methodology of model development and the criteria of analysis for comparing numerical model outputs and field measurement data. It has also flexibility for coupling alternative extensions for new data sources and types of statistical analysis. The DSS also allows the storage and processing of the measured data at

the end of a measurement campaign allowing the analysis of the data and to reprogram further measuring campaigns, if necessary.

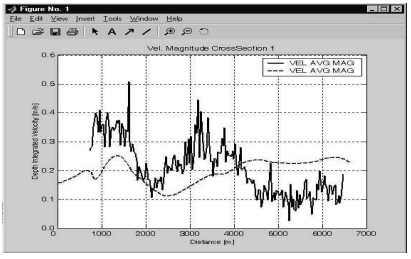
New technologies, like telemetrics, increase the possibilities of the DSS to use it in quasi-real model validation situations in which the model is prepared *a priori* and is brought to the field to compare the numerical model outputs with the measurements that are being taken. After the completion of a measuring campaign the results taken at all the cross-sections and covering the entire measuring period are analysed and improvements in the model and/or the measuring strategy can be considered on the basis of the statistical analysis.

a. ADCP measurements



b. Model Results

c. Use of the DSS on board



d. Comparison of measured and computed velocities

Figure 9.11 Online validation of the numerical model.

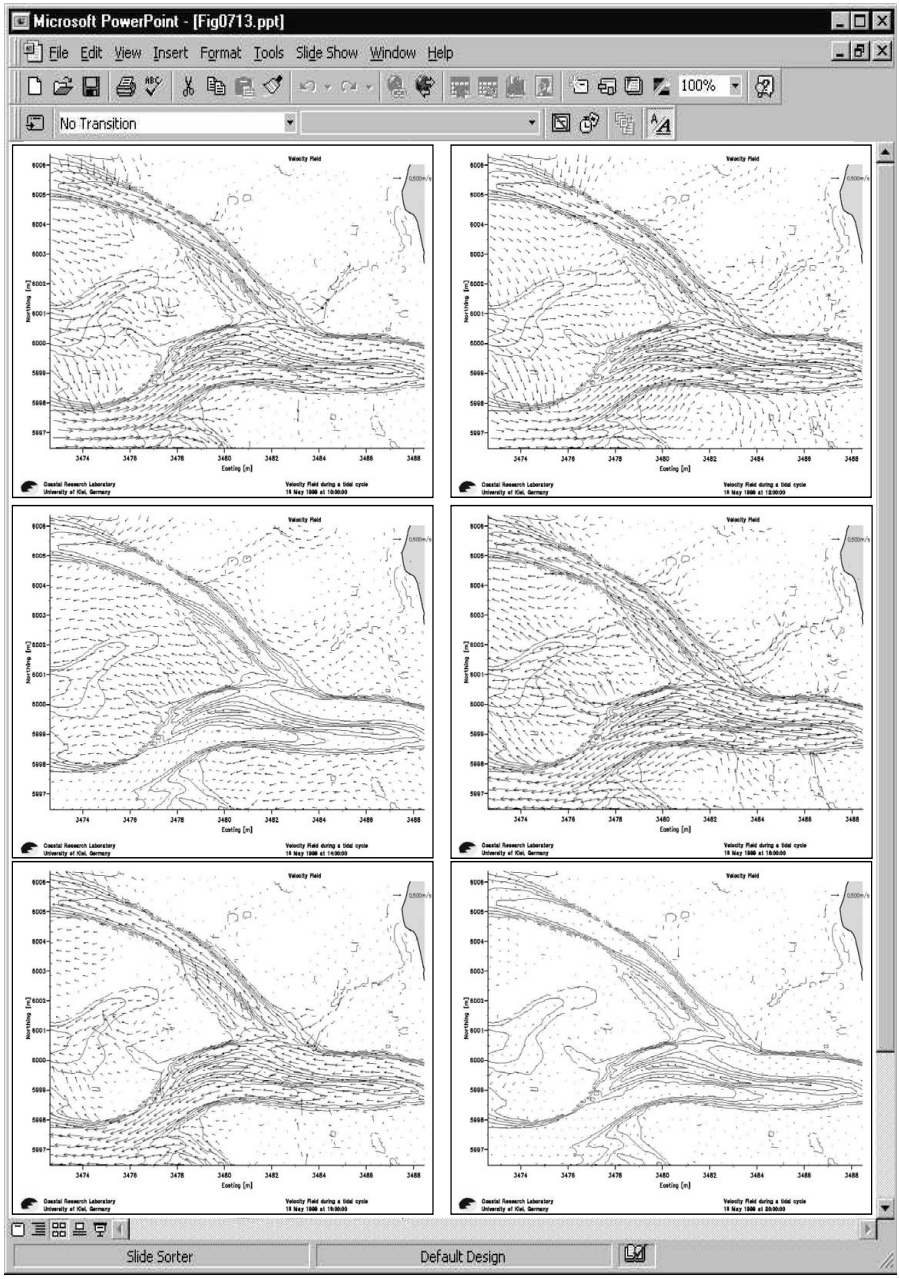


Figure 9.12 Velocity plots imported into a report.

9.6 REFERENCES

ESRI, 1996. Using ArcView GIS, ESRI, Redlands, California, 340pp.

Mayerle, R. and Toro, F., 2002. Effectiveness of a Decision Support System for Quasi-Real Time Model Calibration. Proceedings Hydro 2002, Kiel, Germany, October 2002. pp 497.

Palacio, C.A., Mayerle, R., and Toro, F.M., August 2003. Sensitivity Analysis for the Meldorf Bight Hydrodynamic Model. XXX Congress of the International Association of Hydraulic Engineering and Research (IAHR), Thessaloniki, Greece.

The Mathworks, 2000. Using Matlab, Version 6. The MathWorks, Inc., Natick, MA, USA.

Toro, F. and Mayerle, R., 2001. A Decision Support System for Enhancing Model Development and Application. Proceedings ESRI International User Conference 2001, San Diego, USA, June 2001.

Van Rijn, L.C., 2001, Application of profile and area models to the Egmond site on the time scale of storms and seasons; COAST3D Experiments 1998-1999. Report Z2394, Delft Hydraulics, Delft, The Netherlands.

Zielke, W., Mayerle, R., Gross, G. Eppel, D.P., and Witte, G. 1998. *Predictions of Medium Scale Morphodynamics – Promorph*. German Ministry of Education and Research (contract number 03F0262A/03F0262B), Berlin.

CHAPTER TEN

Decision-Making in the Coastal Zone Using Hydrodynamic Modelling with a GIS Interface

Jacques Populus, Lionel Loubersac, Jean-François Le Roux, Frank Dumas, Valerie Cummins, and Gerry Sutton

10.1 INTRODUCTION AND CONTEXT

There are many kinds of coastal water quality issues. They arise from various sources of inputs to the coastal zone that are either chronic or accidental. Pollution of seawater mostly results from a) point source and diffuse pollution originated from agricultural, industrial and urban activities, or b) pollution from maritime activities, e.g. waste oil as well as all types of toxic substances being dumped into the sea, including radioactive ones (Kershaw, 1997).

Toxic phytoplankton blooms are a new plague. Although they are not the direct result of human behaviour, they are probably linked to human activities, and particularly contaminated ballast water. These harmful disruptions have severe effects on shellfish stocks, entailing long production shutdowns. Problems also currently exist with open sewers discharging into places designated for shellfish production or into recreational waters (Boelens *et al.*, 1999).

After flooding, there is run-off pollution from agriculture and urban areas and nitrate concentrations and bacterial counts increase remarkably. The increase in nitrate concentrations can lead to dramatic eutrophication, whereas pathogenic bacteria can make aquaculture products hazardous to human health (Lees, 2000). Other activities or phenomena such as dredging, deep water sludge disposal or landfill seepage are concerns for water quality and marine living resources (Sullivan, 2001).

While hydrodynamic modelling results and the handling of geo-referenced information are becoming more readily available to coastal management stakeholders within GIS (Geographic Information Systems), there is still a lack of direct interfacing of model results with baseline mapping data (BASIC, 2001). This paper discusses solutions to bridge this gap and illustrates them with two case

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

studies where effective model outputs are being used for improved environmental management and decision-making.

10.2 HYDRODYNAMIC MODELLING BASICS

Mathematical models (Garreau, 1997) can solve geophysical fluid mechanics equations using a number of simplifying assumptions. In essence, knowledge of the bathymetry, wind tide and currents are used to predict water levels and concentrations of conservative elements over space and time.

The ability of these models to accurately and reliably reproduce the important features of complex systems has improved considerably along with rapid growth in understanding of the underlying physical phenomena, and the ability to quantify them in terms of valid mathematical formulations. The widespread and affordable availability of high-powered computing facilities has also contributed to this development, given the heavy processing loads involved in running all but the simplest models.

Confidence in the predictive capacity of models is governed by the quality of fit between modelled output data and real field measurements. Standard practice in building a typical coastal model entails a number of discrete steps, the first of which involves reconstructing the morphology of the area of interest in the form of a digital terrain model. The next step involves adding water to the system, and then setting it in motion through replication of vertical tidal fluctuations and horizontal wind friction. Together these forcing factors create gradients, which in turn drive horizontal current flows. A range of assumptions is usually made in order to simplify the system, and allow it to run effectively. It is recognised that models do not strive to exactly reproduce in detail all the features of a natural system; however the aim is to arrive at a situation where the model is capable of reliably reproducing the principal features through an iterative process of validation and calibration.

10.3 GIS FOR COASTAL ZONE MANAGEMENT

A Geographic Information System (GIS) is a computer-based information system used to digitally represent and analyse geographic features. It is used to input, store, manipulate, analyse and output spatially referenced data (Burrough and McDonnell, 1998). A GIS can be distinguished from database management systems or from visualisation packages through its specialised capability for spatial analysis. The use of GIS for coastal zone management has expanded rapidly during the past decade and references are numerous (Durand, 1994; Populus, 2000; Wright and Bartlett, 2000). For optimum efficiency, geo-referenced data should be properly stored in geo-databases built on spatial data model design (Prélaz Droux, 1995). Some of the greatest challenges currently faced by those handling coastal zone data are a) the land-sea interface, with different mapping references in both horizontal and vertical modes, b) water dynamics and the related temporal issues, and c) 3D display requirements.

10.4 TOOLS AND DATA

10.4.1 Technical details of regional and local models

MARS-2D is a bi-dimensional model using a finite difference method called ADI (Salomon, 1995). A broad regional model, extending between 40˚N and 65˚ N and from 20˚ W to 15˚ E with a 5 km grid, is used as a framework in which to embed further models of smaller extent for areas of interest along French Atlantic and English Channel coasts. Commonly, the embedding system has up to 5 levels allowing phenomena to be examined at resolutions from 1km down to 50m. This type of model is suitable for applications in areas whose waters are typically well mixed (e.g., coastal or mega-tidal areas). The model is designed to solve for tidal and wind driven currents and the transport of dissolved materials.

MARS-3D (Lazure, 1998) is a fully finite difference model, in both vertical and horizontal orientations, which uses a time splitting method based on MARS-2D for the barotropic mode. Good coupling between 2D and 3D modes has largely been achieved through the use of iterative methods. MARS-3D is used at resolutions ranging from 5 km over regional areas of epicontinental seas, down to 100 metres over detailed areas, narrow bays and estuaries. It is currently run operationally at the regional scale (with a 5 km mesh) on all French seaboards. Table 10.1 gives an overview of MARS model features.

Table 10.1 Comparison of MARS-2D and MARS-3D main features

Model	MARS-2D	MARS-3D
Area	English Channel, Bay of Biscay	English Channel, Bay of Biscay
Grid and time step	From 50m up to 10km	~ 5km, 5 – 20 minutes
Period of time	From days to decade	Year to decade
Applications	Tide currents, dissolved matter, salinity under homogenous conditions	Tide, currents, temperature, salinity, transport of dissolved matter
Type	2D Finite difference	3D Finite difference

10.4.2 The modelling system

A modelling chain has been developed at Ifremer over a period of many years. The system has a series of pre-processors and tools for the graphic display of results produced by the MARS computation kernel.

This mathematical model uses some simplified hypotheses to solve the equations that govern how marine currents and sea levels evolve. In order to function, the process requires an input water level along the edge of the area of interest. These boundaries are usually unknown locally, since they are dependent on tidal and weather conditions, which are themselves usually derived from modelling over a much larger area. Thus, the modelling process advances through the generation of a series of sequentially nested models. An initial general model covering a large area of the continental shelf and the Channel is followed by a succession of intermediate models of increasingly smaller scope, but higher resolution. Boundary conditions for the wide-area model are resolved using world tide models, into which modelled meteorological forcing factors have been assimilated. The modelling chain can be summed up as follows:

- A generated link calculates the position, extent and resolution of each sub-model, from the large-area model to the detailed high resolution model based on computational and hydrodynamic design criteria. Computational efficiency is optimised by maintaining a maximum resolution ratio (mesh size ratio) of between four and five between any two consecutively nested models. Hydrodynamic criteria are observed as far as possible in the design through avoidance of islands or zones with violent currents, although at present the system only works when model boundaries are strictly aligned to parallels and meridians. This link of the chain has a user-friendly graphic interface and generates a descriptive file of the entire nesting process.

- The second link in this computation chain is a software program which calculates an interpolated bathymetry for each nested model. The link also has a graphic interface developed in UNIRAS, which restores a depth for each calculation link in a file. The bathymetry used for the large-area model has been validated, and is essentially taken as fixed. However, it is updated on an occasional basis, as new information becomes available. The MARS-2D computation kernel is used in both case studies below, where prevailing tidal currents have a relatively homogenous vertical structure, providing a good approximation of the mean current fields pertaining in the study areas.

- Lastly, a range of graphic tools are used to display the resulting modelled outputs which are written in NETCDF format (Rew and Davis, 1990). NETCDF is a widely used self-documenting format, which also provided a suitable platform upon which the ArcView portal was subsequently developed.

10.4.3 Reference mapping data

Currently, coastal practitioners must refer to common baseline or reference data, i.e., primary data to which secondary (or more application-related) data will be subsequently linked (Allain, 2000). The coastline, bathymetric data, and major administrative boundaries are examples of such baseline data that could be readily

provided to a wider public under optimal conditions of accuracy, updating, and scalability to suit various needs.

However, it is noted that the bathymetry of inter-tidal areas, and other near-shore zones, is often less easily available or poorly defined. Such paucity of data is usually associated with the high cost of acquisition and restricted accessibility. Bathymetric data for the two study sites under consideration was available from the French hydrographic service at a scale of 1:50, 000.

ArcView™ was the main GIS platform used for the studies, within which the Spatial Analyst™ extension facilitated a number of operations on raster images such as recoding, resampling, changing extent, computing statistics and the use of algebraic image combination functions. Existing ArcView functionality also facilitated interactions between raster and vector data layers drawing on the attributes of file features attaching to vector data sets. However a major hurdle remains in the efficient handling of large numbers of raster data layers that typically accrue as the output from multiple model runs.

10.4.4 An Integrated GIS/model interface

The primary rationale behind the creation of a GIS/model interface was to allow a) consultation and display of results contained in the output files produced by the MARS hydrodynamic model; and b) extraction of these results to import them to ArcView. The concept was initially tested through the development of ModelView (Loubersac *et al.*, 2000), a prototype interface the functionality of which was successfully demonstrated in the case of hydrobiological contamination in the Bay of Marennes-Oléron, France.

As a further development, in order to broaden the scope of application a platform-independent stand-alone interface, MODELCON, was designed for use in conjunction with a range of GIS packages. In order to ensure optimum compatibility with standard software packages and other GIS (e.g. Excel, MapInfo™ etc.), MODELCONV extracts NETCDF files to a standard ASCII format. The MODELCONV interface was developed in JAVA, with Microsoft's Visual Studio 6 environment and the JAVA library to access NETCDF files (NETCDF JAVA version 2), ensuring maximum portability in anticipation of future use on Unix systems or via the Web.

10.4.5 A Geographic conversion module

An additional processing module was developed in order to address specific geoprocessing requirements beyond those available in the standard version of ArcView. This module operates as an ArcView extension and was implemented in Avenue. It enables the user to perform a range of geodesic processing operations on point, multipoint, polyline and polygon data, as well as 2-D and 3-D related measurements (pointM, multipointM, polylineM, polygonM, pointZ, MultipointZ, PolylineZ, PolygonZ).

This module also allows data to be projected or unprojected, i.e. as geographic co-ordinates (latitude, longitude), or projected Cartesian co-ordinates including 3-D (X, Y, Z). Other operations that are supported include:

- switching between geodesic systems: WGS 84, NTF and Europe 50;
- switching ellipsoids: Clarke 1880, IAG GRS 80 and Hayford 1909;
- switching projections: Lambert (various), Transverse Mercator (UTM).

10.5 CASE STUDY ONE: SHELLFISH PRODUCTION IN THE GOLFE DU MORBIHAN, SOUTHERN BRITTANY

10.5.1 Water quality issues in the Golfe du Morbihan

The Golfe du Morbihan is located in southern Brittany, France. Enclosing an area of 125 km^2, with many islands and extensive intertidal flats, the basin connects to the open ocean via a 1km wide channel. Its perimeter is highly indented, and is traversed by numerous streams and rivers. The main anthropogenic pressures related to two main cities, Auray (11,000 inhabitants) and Vannes (50,000 inhabitants), are augmented by very significant seasonal tourist populations. Important natural shellfisheries (*Venus* spp.) are commercially exploited, and the gulf also supports a valuable oyster farming industry. Both rely on suitable water quality being maintained within narrow sanitary limits. In general, water and shellfish quality criteria are set by the EU Shellfish Hygiene Directive (Lees, 1995). However, at national level slight differences are found in the interpretation of shellfish hygiene *E. coli* guidelines, specifically in stating what proportion of samples must fall under the concentration thresholds and in the treatments required prior to human consumption (BASIC, 2001). In France, these values are defined by the modified decree n° 94-340 of 28 April 1994, as shown in Table 10.2 below. Shellfish farmers must take different measures for depuration with respect to these categories.

Table 10. 2 French shellfish hygiene categories

Categories	**Level of contamination in faecal coliforms***			
	300	**1,000**	**6,000**	**60,000**
A	≥ 90%	≤ 10%	0%	
B	≥ 90%		≤ 10%	0%
C	≥ 90%			≤ 10%
D				> 10%

* per 100g shellfish

10.5.2 Microbiological modelling

Faecal coliforms (FC), of which *Escherichia coli* is a major component, are good indicators of bacteriological contamination levels in seawater and shellfish. EU health standards for recreational waters and shellfish harvesting zones are based on organism counts. These bacteria mainly come from rivers and streams which receive waste water from various sources including surface runoff and soil outwash, sewage treatment plant discharges (especially in heavy rainfall situations, when function is impaired by lower residence time) and unauthorised discharges. Other direct sources are storm water drain systems discharging directly into the sea and diffuse contamination in the vicinity of moored boats.

Coliform bacteria do not tolerate exposure to solar UV radiation, and thus have a limited life span in seawater. Their survival time will vary depending on their metabolic state and environmental conditions. Bacterial survival times are positively influenced by the following:

- lower winter temperatures which can slow down bacterial metabolism and extend survival based on slower rates of consumption of energy reserves;

- increased turbidity values, linked to levels of suspended solids. Turbidity has a dual effect, as a potential food source and as protection against solar UV (ultraviolet) radiation. Turbidity is usually higher in winter, due to greater discharge and sediment resuspension.

In general modelling applications, bacterial survival is represented by the term T90 (being time at which 90% of the bacteria will have disappeared), which assumes an exponential rate of decline in numbers. Values for T90 are normally established on an empirical basis, and those used in the current study were based on experience at local sites (Guillaud, 1997).

10.5.3 Simulation descriptions

In order to characterise coliform distributions under a range of environmental conditions, a series of MARS 2-D model simulations were undertaken in which the following parameters were varied:

- Tidal conditions. Simulations were made over a period of three weeks under realistic tidal conditions, i.e. of sufficient duration to capture both spring and neap cycles.

- Seasonal influence. This was investigated by varying discharge volumes, i.e. coliform flux.

- Weather conditions. Three sets of commonly prevailing conditions were selected, in accordance with wind statistics derived from data supplied by the French meteorological office. These were 1) zero wind (baseline condition), 2) westerly $8 m.s^{-1}$ and 3) north easterly $8 ms^{-1}$.

- T90s. A summer value of 10h, with 24h for winter was chosen in accordance with local measurements.

- The impact of major malfunctioning of water treatment plants in heavy rainfall periods. This was modelled by doubling the amounts of bacteria discharged over a 24-hour period, which is typical of what can occur during a summer storm. The objective was to investigate the impact of episodic events on coliform distribution, particularly with respect to existing zonation patterns.

The above scenarios were investigated through a combined series of seven simulations (two seasons under three different conditions, plus one exceptional situation type). In order to assess the validity of the simulations, the resulting coliform concentration distributions were then categorised according to EU shellfish farming criteria (Table 10.2). Model-derived zonation patterns were found to be in broad agreement with existing zonation plans based on both tissue sampling and water quality monitoring network samples.

10.5.4 Simulation results

Having established the overall validity of modelled results, priority was given to locating the zones likely to be subject to the highest contamination. These were found to be in upper reaches of the main rivers and in the bay of Vannes (Figure 10.1). Observations performed during model runs indicated that once steady state conditions are attained, the spring/neap tidal coefficient has little influence on subsequent coliform concentrations. The impact of a constant, moderate wind (8 m/s) on contamination plumes was imperceptible, both in summer and winter. In the summary of results, only the baseline “no wind” situation is referred to.

T90 was found to exert the greatest influence on coliform concentration, and when set to 24hours gave rise to the highest levels of contamination. This is reflected in the final model run for which a 24h T90 linked to winter discharge rates was chosen.

10.5.5 Comparison between the actual classification and a simulation

Figure 10.2 shows the results of zonation categories based on samples collected by the water quality monitoring network (Ifremer, 2003). Zonations are based around 30 such samples per site that are routinely collected over the course of each year at low water during each spring tidal cycle (theoretically least favourable situation). Dots indicate the locations of effluent discharge sources. Figure 10.2 clearly shows the main water body of the gulf officially ranked as category A, whilst the estuaries and their outlets (north and northeast) are in the B or D categories. It is notable that no estuarine areas have been zoned in the C category and have instead been allocated to category D as a precautionary measure in respect of the EU Shellfish Directive. This precautionary designation takes into account the obvious

risk of contamination near urban and port areas, as well as the impact of reduced salinity on mariculture products in upper estuarine reaches.

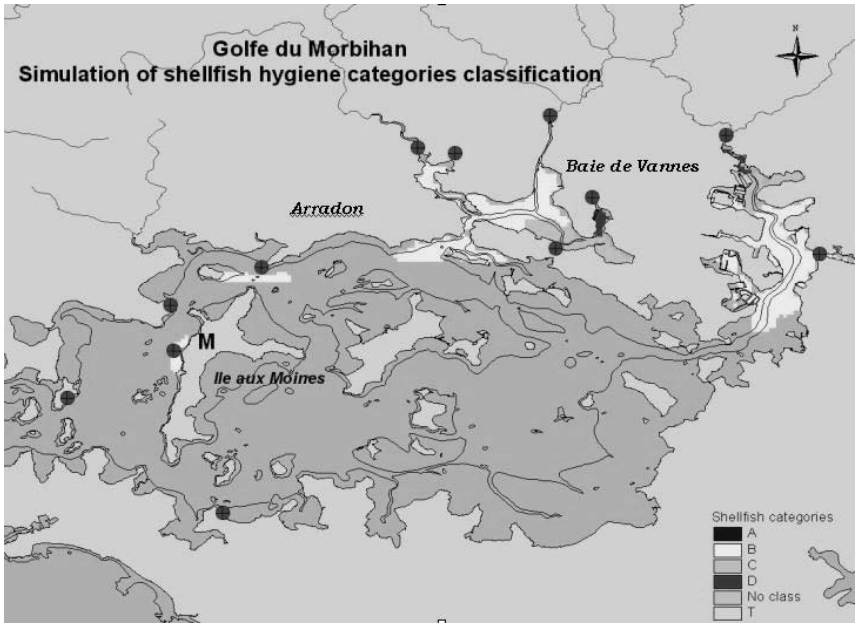


Figure 10.1 Screen capture from GIS showing the Golf du Morbihan. Coloured areas denote EU shellfish classifications based on simulated coliform concentration distributions

Comparing the results highlights the good consistency between simulation results and those obtained from monitoring network observations. Whilst the general configuration of zonation categories (Figure 10.1) based on simulations is broadly consistent with the official classification scheme (Figure 10.2), the former logically appears to give rise to a less conservative regime under which the estuarine areas in the north and northeastern gulf are designated as category C, rather than the precautionary official D designation.

This may be explained by the relative shortness of the period simulated (3 weeks), whereas the official designations are based on an annual monitoring cycle. Other reasons may be that the coliform flows used for the simulation did not include the annual maxima (in the case of water collection and treatment plant malfunctions). Furthermore the actual T90 may exceed the 24h value used in simulations, especially in naturally turbid upper estuarine areas, or as a consequence of winter storm induced sediment resuspension.

By comparing the Figures where the discharge points are identified by a dot symbol, we also note that the model, which minimised coliform concentration levels as seen above, reveals two B category zones in the outer Bay area. These can be seen in Figure 10.1, located to the north and to the west of the Ile au Moines (the island is indicated by an M on the map). Their coincidence with

effluent discharge points (dots) suggests the logical cause of reduced water quality in these areas. This result highlights the valuable insights that can be obtained through the use of realistic simulations based on numerical models in identification of localised water quality issues, which may then be addressed through the allocation of additional monitoring resources. In this case, priority was given to the northern most area (associated with effluent discharge Arradon) owing to its proximity to significant shellfish farming areas, resulting in the establishment of an additional hygiene monitoring station.

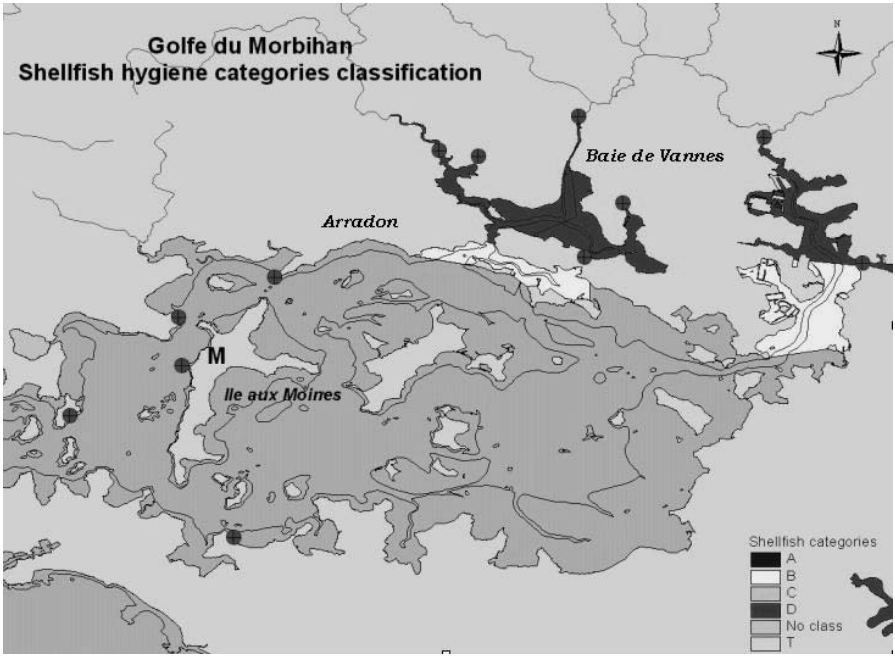


Figure 10.2 Screen capture from GIS showing the Golf du Morbihan. Coloured areas denote EU shellfish classifications based on measured coliform counts.

10.6 CASE STUDY TWO: THE WRECK OF THE *IEVOLI SUN*

10.6.1 Context of the case study

The Italian chemical tanker, *Ievoli Sun*, sailing from Rotterdam to Genoa sank around 9 am local time on 31th October 2000 in the central English Channel, approximately 9 nautical miles north of Les Casquets (Channel Islands) and 20 nautical miles west-north-west of the French Cap de la Hague off the north coast of Normandy (Figures 10.3 and 10.4).

Figures 10. 3 and 10.4 *Ievoli Sun* being towed a few hours before she sank (French Navy sources).

The vessel was carrying three chemical products: 3998 tonnes of styrene (Vinylbenzene), 1027 tonnes of Methyl ethyl ketone (MEK or 2-Butanone) and 996 tonnes of Isopropyl alcohol (Propanol-2). Although the latter two products are considered to be practically non-toxic for aquatic life, styrene is known to be hazardous for the marine environment and for humans. According to the Group of Experts on the Scientific Aspects of Marine Pollution of the International Maritime Organisation (GESAMP, 1989) styrene is bioaccumulable with a propensity to affect edible marine resources. Moreover, it is dangerous for human health when in contact with skin or when inhaled. It can cause major environmental problems due to its persistent irritant toxic odour, which can force the closure of beaches.

French Naval investigations in the aftermath of the sinking showed that the wreck was lying on her port side, and subsequent remotely operated vehicle (ROV) surveys confirmed that the hold was ruptured with a consequent release of styrene into the marine environment.

The wreck was located close to the Channel Islands, in an area of high biodiversity, and in close proximity to shipping lanes, cross-channel traffic, fisheries, aquaculture, and beaches. Neighbouring coastlines are rugged and highly indented, with many sensitive and vulnerable sites, and are under the jurisdiction of several different regional authorities, including those of the Channel Islands.

A rapid response team including experts in pollutant chemistry, ecotoxicology, modelling, fishery science, aquaculture and digital mapping was assembled within two hours of the sinking. The prime objectives of this group were to establish a baseline for environmental contamination, and devise a plan for coastal environmental monitoring in the subsequent hours. This meant being able to characterise the fate of pollutants in the water mass and identify sensitive areas and fish species likely to be affected. Thus the main requirements were (Loubersac *et al.*, 2002):

- to rapidly compile various forms of multi-thematic information from multiple sources;

- to provide simulations of pollutant dispersion at sea for several scenarios in the shortest possible time;

- to standardise these compilations and simulations and make them easy to display, communicate and update for simultaneous use across several decision centres.

10.6.2 Simulation of pollutant behaviour using MARS-2D

The MARS modelling suite (Salomon *et al.* 1995; Garreau 1997) was used to rapidly derive a 500m horizontal mesh hydrodynamic model centred on the wreck position. A 2-D horizontal model was adopted to simulate the local hydrographic regime, where dissolved substances were likely to become well mixed owing to the strength of prevailing tidal currents which are relatively consistent throughout the vertical extent of the water column. Actual wind fields supplied by the French meteorological office were also assimilated into the model.

Simulations were run to reproduce the evolution of instantaneous and residual tidal current fields, tidal trajectories and concentrations of the dissolved fraction of the spilled products. The following three scenarios were considered assuming maximal dissolution of the styrene in seawater: a) a single total discharge of 4,000 tonnes, b) simulation of 4,000 tonnes of the product leaking from the wreck over a period of 3.5 days, and c) simulation of continuous discharge of 20 $l.min^{-1}$ over 25 days (closely matching observed leakage rates).

10.6.3 Setting up a customised GIS

A GIS was established for the wreck site and surrounding maritime region, in which functionality was tailored towards handling and visually representing marine and coastal data. The main steps in this process were:

- Building a reference geographic data set (Allain *et al.*, 2000) in order to ensure stakeholder interoperability;
- Conversion of all data into decimal degree geographic coordinates, and establishing a common vertical reference frame for marine and land-based topographies;
- Adoption of a uniform map template for all data layers optimised for a standard working scale of 1:50,000.

10.6.4 Results

The *Ievoli Sun* sank in an area in which strong tidal currents and an irregular seabed give rise to intense vertical mixing. Under these conditions it is reasonable to assume that styrene escaping from the wreck became rapidly dispersed throughout the water column. A 3-D perspective view showing the spatial extent of dispersed styrene as it appeared after 10 days of simulated advection is illustrated in Figure 10.5. This simulation was based on scenario c) with the following assumptions: *i)* the matter coming from the wreck dissolved completely, *ii)* there was no chemical modification over time, and *iii)* an average current was responsible for the transport of these dissolved substances.

The image presented in Figure 10.5 was produced by merging modelled data with those from the GIS mentioned above, using ArcView 3D Analyst. This image corresponds to the situation on the 9th November at 2 am (10 days after the spill), taking into account the atmospheric forcing data provided by the French meteorological service. The concentration limits indicated on this Figure are 0.30 ppm (parts per million or mg/l) (darkest), 0.15 ppm (lightest), and 0.05 ppm (medium-grey) in the raised area.

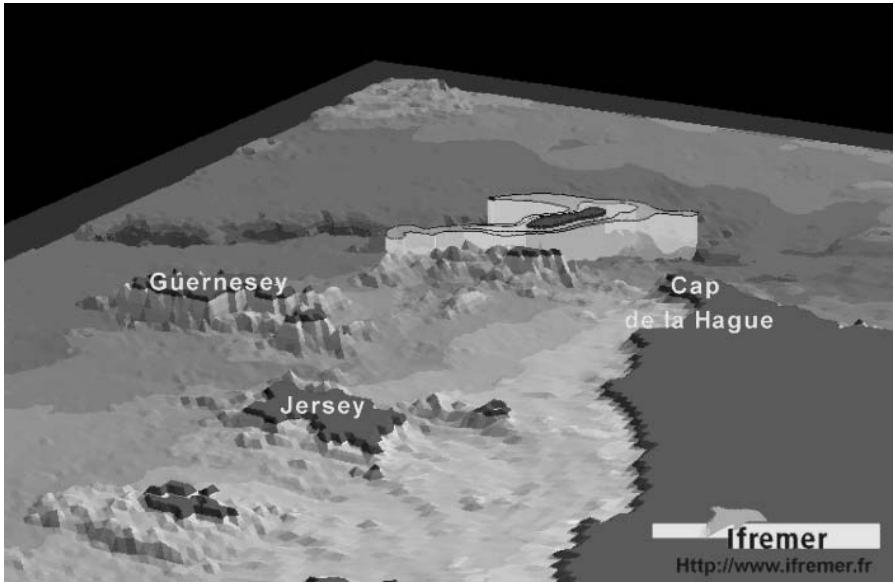


Figure 10.5 3D perspective representation of the plume of dissolved pollutant dispersion after 10 days.

Dispersion of a continuous discharge of 20 $l.min^{-1}$ was also simulated for the period between 30th October and 24th November, incorporating actual wind measurements up to the 19th November and forecasts for the remainder of the simulation period. Plate 10.1 shows the situation at the end of the simulation (24/11/00 at 0600) with contaminant concentration expressed in parts per million (ppm). While the bulk of the dilution patch is located along an axis running East North East – West South West across the wreck site, a distinct branch can be seen extending southward towards the island of Alderney. However, seeing the very low flow simulated, the concentrations of dissolved product in the water column remain extremely low, since the highest concentrations reached in the simulation did not exceed 0.056ppm. However it should be noted that the overall concentrations remained below 0.056ppm throughout the water column; such low values are however consistent with expectations given the low rate of discharge.

10.7 CONCLUSION AND PERSPECTIVES

An operational model of the Golfe du Morbihan was constructed, which successfully reproduced the hydrodynamic regime of the study area to within acceptable tolerances. The model has also been used to effectively simulate patterns of bacterial dispersion, the results of which were generally in accordance with officially established shellfish zonation patterns, based on the sampling regimes of local water quality monitoring networks. Time-series of computed concentrations for specific locations can be extracted from existing model runs, and can be selected to correspond with monitoring sites. The approach thus provides a means of optimising sampling programmes to make best use of available resources, and to improve responsiveness to episodic events with potential consequences for health.

The present arrangement and state of development of these modules allow a user to work effectively with a given number of grids generated as output from a limited number of modelled scenarios. Future developments aimed at automation would reduce the time needed to process and import results, while maintaining the traceability of actions run and all the metadata required to understand and analyse the simulation results. To this end, future efforts will be concerned with improvements in the following areas:

- Ensuring that model-run relevant metadata which is integral in NETCDF files is fully maintained throughout the processing chain;
- Transforming the tables that MODELCONV generates into Arcview grids;
- Implementing batch processing tools, both on the MODELCONV level and to select user-defined time steps, and on operations between variables to generate more synthetic results than in ArcView, and to import a series of files and automatic captioning for them.

The effectiveness of the *Ievoli Sun* pollution response can be attributed to the favourable platform upon which it was based. Pre-existing tools could be activated within 48 hours, facilitating rapid delivery of environmental maps incorporating hourly forecasts of styrene slick position and concentration. Prudent diagnoses of the potential effects on marine life could then be made in the light of various leakage scenarios. Integration was made possible by providing consistent geo-coded information on the land-sea interface, through a GIS interface closely connecting these data and simulations issued from digital models. Integration of consistent geo-coded information on the land sea interface with model simulated outputs within a GIS environment has resulted in the creation of an active interface for pollution response decision makers that incorporates up-to-the-hour scientific assessments.

A further logical extension of this work would be to jointly use hydrodynamic, sedimentological and biological modelling combined with GIS. Links with catchment models to allow regional-scale assessments of human activities (e.g. agriculture) and natural events (e.g. flooding) should also be considered. Planned future work will incorporate sedimentological and biological

process modules into the existing hydrodynamic modelling/GIS schema, thereby increasing the diversity of assessments that can be made. Critical coastal management issues including bacterial and viral contamination of bathing waters, occurrence and impact of harmful algal blooms, eutrophication, and definition of coastal water masses (e.g. in relation to the EU Water Framework Directive) could then be addressed.

It is anticipated that schema will ultimately be extended in order to incorporate links with catchment models, thereby allowing regional scale assessments of human impacts (e.g. agriculture) and natural events (e.g. flooding).

10.8 REFERENCES

Allain S., *et al.*, 2000, Données géographiques de référence en domaine littoral. In *"Geomatics and Coastal Environment"*, edited by Populus, J. and Loubersac, L. (Editions Ifremer/Shom), pp. 67-79.

Boelens R.G.V., Walsh A.R., Parsons A.P., and Maloney D.M., 1999, Ireland's Marine and Coastal Areas and Adjacent Seas: An Environmental Assessment. Quality status report. Prepared by the Marine Institute on behalf of the Departments of Environment and Local Government and Marine and Natural Resources, Dublin.

BASIC, 2001, A Scoping Study to Establish a Common Approach to Examining the Impact of Atlantic Arc Water Quality and Dynamics on Coastal Activity and Sensitive Marine Environments, Atlantic area INTERREG-IIC programme. Final report, (Cork, Ireland: CRC).

Durand H., *et al.*, 1994, An example of GIS potentiality for coastal zone management: Pre-selection of submerged oyster culture areas near Marennes Oléron (France). *EARSEL Workshop on Remote Sensing and GIS for Coastal Zone Management*. Delft, The Netherlands, 24 - 26 Oct.

Garreau P., 1997, Caractéristiques hydrodynamiques de la Manche. *Oceanis*, **23**(1) pp. 65-97.

GESAMP [Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection], 1989, The evaluation of the hazards of harmful substances carried by ships: (revision of GESAMP Reports and Studies n° 17). *GESAMP Reports and Studies*, 35: 44p., [11] annexes.

Guillaud J.F., Derrien A., Gourmelon M., and Pommepuy M., 1997, T90 As a tool for engineers: Interest and limits. *Wat. Sci. Tech.*, **35**(11-12), pp. 277-281.

Ifremer, 2003, http://www.ifremer.fr/envlit/surveillance/remi.htm (on-line. Accessed October 2003).

Kershaw, P.J., 1997, Radioactive contamination of the Solway and Cumbria coastal zone: The Solway Firth, *ECSA/JNCC*, pp 43-50.

Lazure P. and Jegou A.M., 1998, 3D modelling of seasonal evolution of Loire and Gironde plumes on Biscay Bay continental shelf. In 5ème Colloque International d'Océanographie du Golfe de Gascogne. *Oceanologica Acta*, **21**(2).

Lees, D.N., Nicholson, M., and Tree, J.A., 1995, The relationship between levels of *E. coli* in shellfish and in seawater with reference to legislative standards.

Lees, D.N., 2000, Viruses and bivalve shellfish. *Int. J. Food Microbiol.* **59**, pp. 81-116.

Loubersac L., Salomon J.C., Breton M., Durand C., and Gaudineau G., 2000, Perspectives offertes par la communication entre un modèle hydrodynamique et un SIG pour l'aide au diagnostic environnemental; caractérisation de la dynamique et la qualité des masses d'eaux . In *"Geomatics and Coastal Environment"*, edited by Populus, J. and Loubersac, L. (Éditions Ifremer/Shom), pp 173-185.

Loubersac *et al.*, 2002, Communication de l'information géographique maritime et côtière pour la gestion d'une crise environnementale. *Revue internationale de Géomatique*. **12**(3), pp. 355-371.

Populus J. and Loubersac L., 2000, (eds), CoastGIS'99: Geomatics and coastal environment. (Brest: Éditions de l'Ifremer).

Prélaz Droux R., 1995, *Système d'Information et Gestion du Territoire*. (Lausanne: Presses Polytechniques et Universitaires Romandes. Coll. META)

Garreau P., 1997, Caractéristiques hydrodynamiques de la Manche. *Oceanis*, **23**(1) pp. 65-97.

Rew, R. K. and G. P. Davis, 1990, NetCDF: An interface for scientific data access. *Computer Graphics and Applications, IEEE*, pp. 76-82, July 1990.

Salomon J.C., Breton M., and Guegueniat P., 1995, A 2D long term advection-dispersion model for the Channel and Southern North Sea. Special Issue MAST 52, part B and C. *Journal of Marine Systems*, **6**(5-6), pp. 495-528.

Sullivan, N., 2001, The relationship between the disposal site and the SAC in Falmouth bay, south-west England. *Unpublished Dissertation, Msc in Estuarine and Coastal Zone Management*.

Valuing our environment, 1999, *National Trust (UK)*.

Wright D. and Bartlett D., 2000, *Marine and Coastal Geographical Information Systems*. (London: Taylor and Francis).

CHAPTER ELEVEN

Towards an Institutional GIS for the Iroise Sea (France)

Françoise Gourmelon and Iwan Le Berre

11.1 INTRODUCTION

As far as coastal zones are concerned, many challenges faced by the scientific community and policy makers, planners and managers justify adoption of a transdisciplinary approach based on bio-chemical, geophysical and socio-economical factors (Burbridge and Humphrey, 1999). GIS are well-established computer-based systems for storing, retrieving, analysing, modelling and visualising the vast amounts of spatial data that may be collected by several providers (Fabbri, 1998). Nevertheless, the implementation of a coastal database is a complex process which requires institutional support to guarantee the multidisciplinary approach, the sustainability of the project in terms of raising funds and human resources, and to promote relationships with other institutions working on the same area (De Sède and Thiérault, 1996).

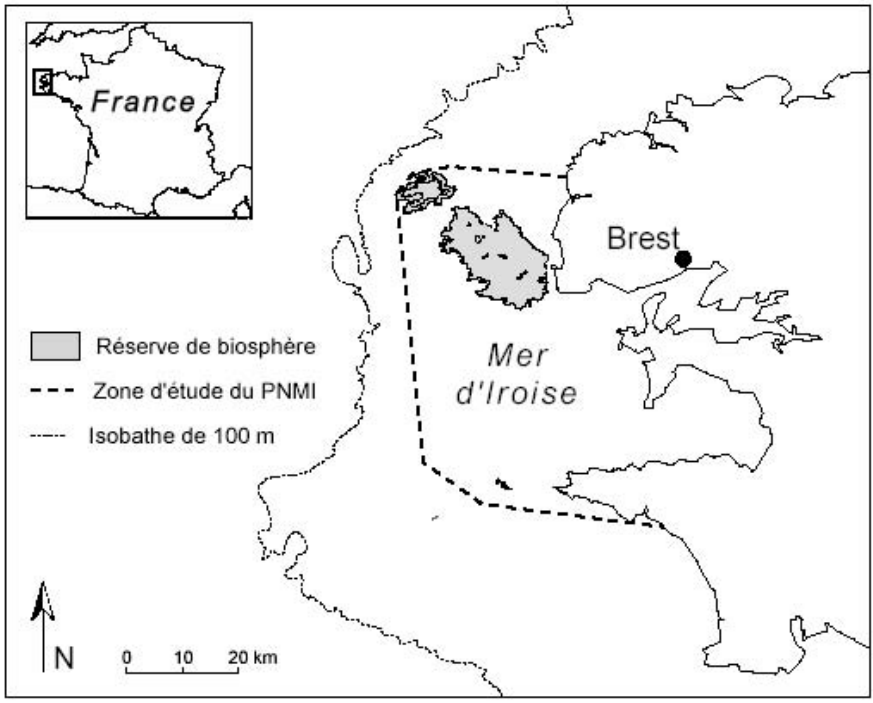
The major pollution caused by the *Erika* disaster, coupled with the catastrophic storms which reached the Atlantic coastal zone at the end of 1999, has led the French authorities to propose creation of a national coastal GIS. The implementation of such a project is complex, witnessed by the lack of any decision about which reference geographical data for the coastal zone to adopt (Allain *et al.*, 2000). In fact, coastal data are scattered among many organisations. They are produced for lots of purposes and therefore are available at various scales, typologies and formats. In spite of these difficult conditions, there are a number of smaller, independent GIS projects at work on the French coastal zone (ENR/OELM, 2000; Guillaumont and Durand, 2000).

This contribution describes the GIS implemented by Géosystèmes laboratory (CNRS, European Institute for Marine Studies) on the coastal zone of Finistère (western Brittany, see Figure 11.1) during the last ten years. In the beginning, the GIS was conceived as a support tool for monitoring and managing the Biosphere Reserve of Iroise, but today, it also serves as a powerful tool for carrying on integrated research on the coastal zone of the Iroise Sea, especially within the framework of the European Institute for Marine Studies (Institut Universitaire Européen de la Mer, IUEM). This latter institute brings together a number of

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

separate marine research teams from the University of Western Brittany, Brest, dealing respectively with oceanography, geology, biology, chemistry, geography, economy and law. The development from a GIS dedicated to scientific applications towards an institutional support is described below.

Figure 11.1 Study area



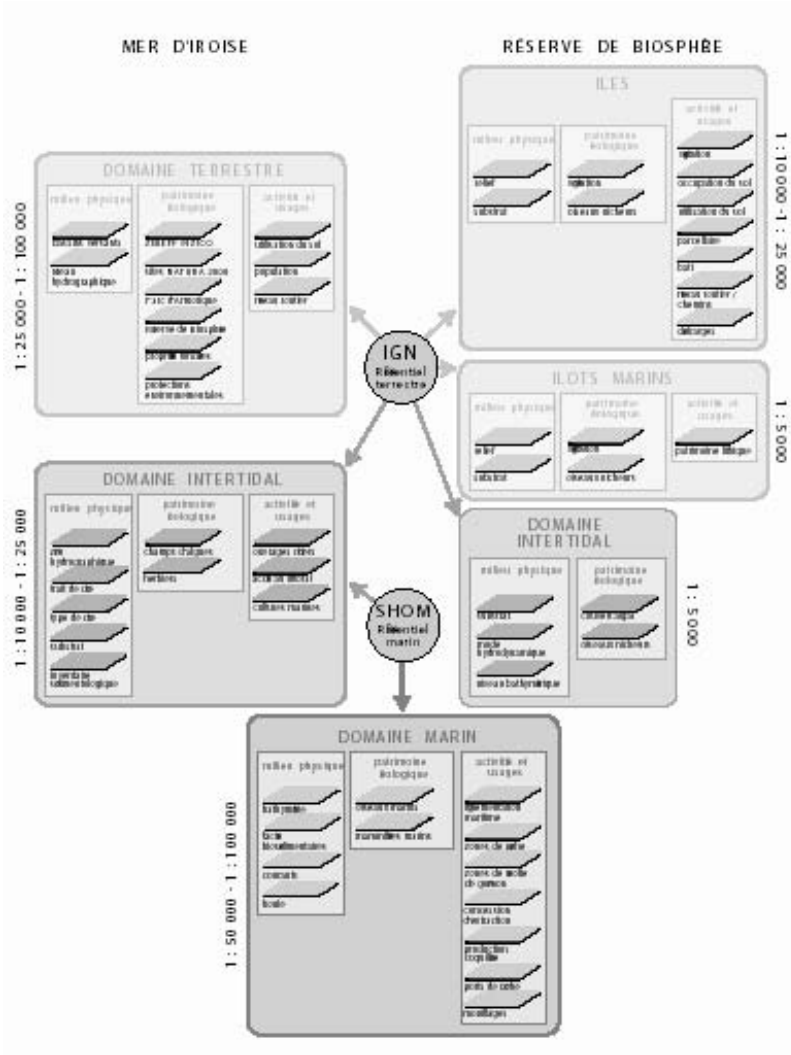
11.2 THE DATABASES

Nowadays the GIS supports two complementary databases, called *SIGouessant* and *BIGIroise*, that were originally developed with separate objectives and data (see Figure 11.2).

The *SIGouessant* database deals with the Mer d'Iroise Biosphere Reserve at a local scale. The biosphere reserve label is allocated to representative terrestrial and coastal areas by the UNESCO Man and Biosphere (MAB) programme. The conception of such an area is based on three main functions: conservation of biodiversity and landscapes; sustainable use in regional units; and logistic support for research, monitoring, education, information and involvement of the local population. In this context, many thematic studies have been performed on the Mer d'Iroise Biosphere Reserve to increase knowledge of dynamic terrestrial and marine processes, and to study the aftermath of human activity such as tourism, management and fishing on biodiversity. The objectives are to provide a basis for land management recommendations as well as for wildlife management

programmes. The *SIGouessant* database has been developed since 1990 as a support to the long term ecosystemic approach, and to provide data for scientific and management investigations (Gourmelon *et al.*, 1995). The geographic reference data is provided by orthophotographs produced by the French National Geographical Institute (IGN - Institut Géographique National). The thematic layers describe physical, natural and social parameters with the same classification systems and scales, collected at various dates thanks to continued monitoring.

Figure 11.2 Database organisation



The *BIGIroise* database is an extension of *SIGouessant* to the whole coastal area of the Iroise Sea. Its initial aims were to produce a synthetic environmental

mapping through combination of multiple data sets available for the coastal zone (Le Berre *et al.*, 2000). The database uses marine geographical information produced by the French hydrographic service (SHOM – Service Hydrographique et Océanographique de la Marine), and is built up according to the environmental planning methods developed and proposed by the Unesco MAB committee (Journaux, 1985). After the collection of existing digital data, and the digitisation of other data sources (atlases, maps, etc.), 34 thematic layers, concerning physical, biological and socio-economical parameters for marine and terrestrial areas of the Iroise Sea have been integrated in the GIS (see Figure 11.2).

After ten years of functioning, this GIS has become an adopted tool for data inventory, environmental analysis and decision-making, which have been described by Crain and Mc Donald (1984) as three development stages of a GIS.

11.3 OVERVIEW OF IROISE GIS APPLICATIONS

11.3.1 State of the knowledge

After the development of a set of tools for data processing and analysis, using techniques such as association, combination and generalisation, the resulting databases have been used to compile a baseline assessment of the state of knowledge regarding the marine and terrestrial environments of the Iroise coastal area. This has allowed the production of an atlas for public communication (Gourmelon *et al*., 1995), and the creation of a synthetic map showing the potential conflicts of interests in the Iroise Sea (Le Berre, 1997). These documents are used to support discussions among the stakeholders.

However, this inventory of existing knowledge also shows that the quality of available data is heterogeneous — especially in terms of exhaustiveness and age — though strong efforts to collect and structure data on the Iroise Sea have been made by a number of scientific and institutional organisations. One of the main difficulties lies on the compatibility of marine and terrestrial geographical reference data. Although the compatibility is essential for database coherence, we are still waiting for a national consensus (Alain *et al*., 2000) on this issue. Other difficulties come from the lack of knowledge about the marine environment. The structure and the functioning of the Iroise Sea ecosystem remains barely understood. Moreover, except for some legal aspects, scarcely any geographical information dealing with human activities is available.

GIS offers many functions that may be used to bypass these problems: For example, the database can help define sampling strategies to be used for collecting additional marine biological data; and it may also provide relevant data for the implementation of theoretical models of ecological population distribution, based on the applications developed for the US Gap Analysis project (Davis *et al.*, 1990). The population models used for Iroise Sea are based on physical parameters such as bathymetry, sedimentology, and hydrodynamic features (marine currents and waves) and have been tested successfully on seaweeds and on the bottlenose dolphin (*Tursiops truncatus*, see the next section, below). Furthermore, integration

with remote sensing allows the production of synoptic and multi-temporal data that are particularly useful for marine studies, because of the variety of the sensors used for recording parameters such as water colour, sea surface temperature, shore morphology, etc. (Van Zuidam *et al.*, 1998).

11.3.2 Environmental issues

GIS is now widely used for applications relating to terrestrial studies and management issues. In landscape ecology especially, the capabilities of GIS are successfully used to perform spatial and statistical analysis of many environmental components, and for modelling real-world processes (Haines-Young *et al.*, 1993). In the islands and the islets of Iroise Sea, many similar GIS applications are concerned with exploring relationships between vegetation and land-cover changes, fauna (rabbits and nesting birds) and human activities. For example, land-cover and land-use changes of Ushant island over a 150-year timeframe have been studied, using field studies, aerial photographs and the 19th Century land registry (Gourmelon *et al.*, 2001) as primary sources of data. Over this period, the island underwent a drastic transformation from rural landscape to extensive shrubland. Only traditional extensive sheep breeding is actually maintained. Within a GIS analysis, the relationships between sheep grazing and land-cover have been established, and scenarios of land-cover potential related to changes in the intensity of sheep grazing produced. The scientific results have provided an objective framework for further assessment of fallow land management.

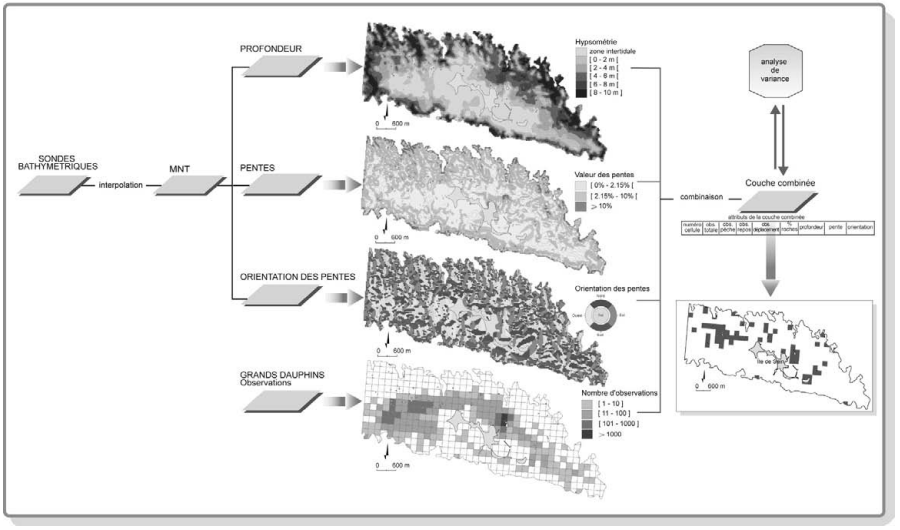


Figure 11.3 Investigating factors influencing spatio-temporal distribution of bottlenose dolphin (*Tursiops truncatus*)

In contrast to these successes on-shore, the implementation of GIS is more difficult in the marine environment due to the peculiarities of maritime space-time

scales, which tend to be generally wider, more complex and less well-known than those of terrestrial ecosystems. A regional approach to marine environments is often required, and data collected at several spatial and temporal scales need to be integrated, for example data relating to the dynamic component of biological and hydrological parameters that may occur daily (like tides), seasonally or annually (like lunar and biological cycles), or that may only be noticeable over several years and decades, such as long term changes (Holligan, 1994). This complexity implies a need to link the GIS to models for the production of dynamic data, i.e. hydrodynamic modelling, and for the analysis of the functioning of the whole system and its subsystems (Capobianco, 1999).

As has already been mentioned above, within the Iroise Sea one area of data analysis aims to investigate the environmental factors that may influence the spatio-temporal distribution of the bottlenose dolphins (see Figure 11.3). In France, the increasing interest in natural heritage has encouraged development of resource management efforts that maintain a high degree of biodiversity. In this context, a Marine National Park project has been proposed for the Iroise Sea, which would include both the areas of coastal bottlenose dolphins. A spatial and temporal approach to the dolphin habitats is required to define a conservation strategy based on habitat features, more than on spatial distribution of groups. Such an approach has been developed in the United States through the Gap Analysis Program conducted by the U.S. Geological Survey (Davis *et al.*, 1990), and is being adapted for the Iroise Sea. Within this wider framework, a more localised study is focusing on a particular group of dolphins observed around Sein Island, and which has been monitored since 1992 (Gourmelon *et al.*, 2000). The annual and seasonal distributions of this dolphin group reveal some preferential sites inside its vital area. Investigating the factors leading to this perceived spatial determinism requires that spatio-temporal parameters be considered, and is thus well-suited to GIS analysis. Meanwhile, at a more regional spatial scale, a classification of bathymetric parameters has been performed to produce a sea-floor stratification that may be compared to observations of bottlenose dolphin distributions (Le Berre *et al.*, 2002). At both scales, local and regional, strong relationships appear between bottlenose dolphin distribution and topographic sea-floor features. The GIS is subsequently used for dolphin habitat suitability modelling, based on logistic regression. This methodology provides a means of mapping the dolphin distribution, and providing useful information for the future National Marine Park. Nevertheless, distribution of prey with depth of slope and hydrodynamic activity is probably of prime influence. The next step of this research will be the introduction of physical and biological modelling in the GIS. The synergy between numerical models and GIS is necessary for the production of dynamic data useful in such an investigation.

11.3.3 Contribution to integrated coastal zone management

A large part of the Iroise coastal zone belongs to protected areas of various types (natural reserves, hunting reserves, biosphere reserve, proposed national marine park, etc.) for which there is a strong need for scientific long-term monitoring, and also of concrete strategies and policies for management. In such contexts, maps are

essential, both for decision-making by field staff, and also for information communication aimed at the stakeholders involved (decision-makers, managers, scientists, etc.). The integration of data from various sources within a consistent framework, as provided by a GIS, improve their accessibility and their availability. GIS also provides efficient functions for rapid production of thematic or synthetic maps and other visualisations, designed for different purpose. Above all, GIS shows great potential for the design and the production of good quality maps tailored to specific uses, such as decision-support during a specific critical event such as an oil-spill, particularly if the mapping process is automated. For instance, in the environmental monitoring programmes implemented for the Iroise sea, the GIS is used for the preparation of sampling strategies, for the detection of spatial changes, for the assessment of future evolution, and for impact simulations.

The assessment and simulation of impacts or changes that affect coastal areas are generally made within the scope of management strategies and policies (Capobianco, 1999). Spatio-temporal modelling, coupled with geographical databases may be very efficient for the production of realistic syntheses of an ecosystem's evolution after different disturbances. In such an integrated and multidisciplinary context, the design of operational tools for decision support depends on the upholding of links between numerical geographical information and other databases, on the integration of spatio-temporal modelling and on the production of user-friendly interfaces that facilitate access to information.

In order to test the ability of *BIGiroise* as an operational tool, the geographical information available for the tidal zone of a limited part of the Iroise Sea was used as a test bed for the development of a model dedicated to oil spill contingency (Le Berre, 1999). It takes into account the relevant data for assessing the sensitivity of coastal areas in order to prepare and to plan the strategies of intervention and to provide useful information for the monitoring of damaged sites (see Figure 11.4). In a practical way, shore sensitivity is assessed through morpho-sedimentary, ecological and socio-economical indices (Michel and Dahlin, 1997) but, for the production of a relevant information for operational needs, vulnerable areas exposed to the pollution risk must be identified. The assessment of risk requires that meteo-oceanic and physico-chemical features of the pollutant be taken into account. This can only be achieved within a spatio-temporal model. In this operational context, the GIS may provide information, especially maps, useful for helping decision-making, but this is only a part of a more comprehensive facility that should be implemented. An efficient system for such purposes requires significant extensions such as oil spill drift models, links with thematic databases (i.e. about pollutant composition) and the production of users interfaces (Howlett *et al.*, 1997).

11.4 PLAN FOR AN INSTITUTIONAL COASTAL ZONE INFORMATION SYSTEM

The research programme described herein demonstrates the contributions and the limits of the GIS implemented for the Iroise Sea. The major issues are now its timelessness and the improvement of its ability to produce a better knowledge of environmental and socio-economical processes in the aim to provide useful

information for decision-making. It implies in a next step an integration into a more efficient framework (see Figure 11.4). An integrated environmental coastal zone management system, for assessing coastal zone changes and functioning, should integrate remote sensing, modelling, GIS and GIS-based decision support systems (Van Zuidam *et al.*, 1998).

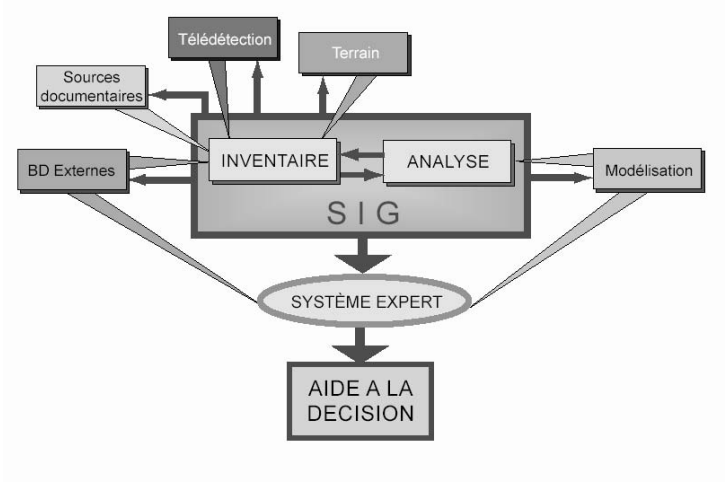


Figure 11.4 Development of an operational GIS

11.4.1 Context and objectives

The GIS developed for the Iroise Sea is an essential component of the infrastructure established to support environmental research programmes conducted by the "Géosystèmes" laboratory. The GIS is used for scientific analysis, to improve the protocols for field data collection, and in some cases to provide decision-making information such as maps and statistics. Due to its environmental characters, the Iroise Sea is the core area of most of these research programmes, sometimes spanning several decades. A proposal has been submitted by the component research units within the laboratory, to the national authorities, to establish a Coastal Domain Observatory with the aim of recording and monitoring long-term changes in a western European coastal area strongly influenced by both climate fluctuations and anthropogenic impacts. The proposal is currently undergoing evaluation for possible funding. In this context, a Coastal Environment Information System (Système d'Information des Environnements Côtière, SIEC) connected with an intensive computing centre is planned, to reinforce the synergy of the Institute, especially by providing facilities for gathering remote sensing data, image processing, GIS and dynamic modelling.

This common service will have five main functions:

- **Collection of data and integration of existing datasets** produced by the IUEM, concerning physical, biological and anthropogenic parameters, and creation of a consistent GIS database, based on the geographical reference information approved by the French National Committee for Geographical Information (Conseil National pour l'Information Géographique, CNIG, see http://www.cnig.fr);
- **Develop and maintain global data exchange conventions** in order to make geographical reference information produced by organisations such as IGN, SHOM, Ifremer or Météo-France available to the IUEM research teams;
- **Information production** by the digitisation of maps, or by data processing of numerical data (interpolation, image analysis, etc.);
- **Metadata production** for supplying a catalogue to assist internal and external data exchange and diffusion between data producers (laboratories, institutions) and users (laboratories, marine professionals, policy and decision-makers);
- **Education and training** of the IUEM staff and students in Geographical Information Sciences.

The feasibility of this joint project has been tested from an inventory of the whole data produced in the IUEM research labs.

11.4.2 The first stage: data dictionary and metadata production

Within the IUEM, the needs for data cataloguing are both external and internal. For internal uses, the maintenance of an up-to-date description of datasets is a guarantee of its long-term viability, and may improve collaboration and exchanges between research teams as it avoid redundancies in research projects. As a research institute, IUEM is an important data producer and user, and has a strategic need for being identified as a data provider by different kinds of external users, especially in a context of the growth of data and information communication across the World Wide Web. The Institute's data broadcasting policy within the SIEC implies a strong compliancy with existing metadata standards, in order to guarantee their robustness and encourage their exchange (Bartlett *et al.*, 2000).

At present, two major metadata standards are used by the international community of geographic information providers: the European standard (CEN/TC 287) and the International standard (ISO/TC211). Correspondence between these two is not perfect, especially in terms of data ownership rights, but the general hierarchical structure of the data descriptions allows the conversion from one standard to the other. As the contribution and established collaborations of the IUEM to international research programs takes place mainly at the European level, the European standard has been adopted for SIEC. The metadata are integrated and managed with Report V2.0, a CEN/TC287-compliant metadata software system developed by the French Ministry of Equipment. At the laboratory level, although the geodatabase was developed initially with ArcInfo software (where ArcCatalog is compliant with the ISO standard), Report is also used for producing and

managing the metadata catalogue until such time as ArcCatalog can take into account the European metadata standard.

In fact, the conversion of standardised metadata to specific formats stays unavoidable when supplying information to Web geodatabase repertories or clearinghouses such as Bosco (http://www.bosco.tm.fr, for French coastal data) or EDMED (http://www.sea-search.net/edmed/welcome.html for data produced during MAST research programs) for example. The need to be able to undertake this kind of specific task will be taken in account when developing the SIEC structure.

11.5 CONCLUSION

The experience acquired after ten years of coastal GIS development shows that even if the databases are useful in many scientific and management applications, their prospects depend on a better standardisation of geographical reference data and their metadata, and on a better integration of the GIS into environmental coastal zone management systems. In the future, data collected and used in Brest will be transferred into the SIEC, a coastal environment information system planned at the IUEM level. In this long-term framework, specific funding and a proper organisational context will be allocated to this information system. Currently, the inventory of data available to the concerned laboratories, and the metadata production, are continuing. The next steps of this joint project will concern the design and the implementation of this multidisciplinary coastal GIS (Muller *et al.*, 2000). From specific scientific applications, the GIS now has to evolve towards an institutional system designed and developed by a team based on future users and data providers.

11.6 ACKNOWLEDGEMENT

The authors would like to pay tribute to the late François Cuq, Director of the Géomer (Géosystèmes in the past) CNRS laboratory and of the Coastal observatory of the European Institute for Marine Studies, who died in May 2003, for his strong involvement in the birth and the growth of a GIS team of geographers in Brest, and for his huge contribution to the development of applied geomatics to coastal zone management in France.

11.7 REFERENCES

Allain, S., Guillaumont, B., Le Visage, C., Loubersac, L., and Populus, J., 2000, Données géographiques de référence en domaine littoral. In *Coastgis'99: Geomatics and coastal environment*, edited by Populus, J. and Loubersac, L. (Brest: IFREMER-SHOM), pp. 67-79.

Bartlett, D., Fowler, C., Longhorn, R., Cuq, F., and Loubersac, L., 2000, Coastal GIS at the turn of the century. In *Coastgis'99: Geomatics and coastal*

environment, edited by Populus, J. and Loubersac, L. (Brest: IFREMER-SHOM), pp. 307-318.

Burbridge, P. and Humphrey, S., 1999, On the integration of science and management in coastal management research. *Journal of Coastal Conservation*, **5**, pp. 103-104.

Capobianco, M., 1999, On the integrated modelling of coastal changes. *Journal of Coastal Conservation*, **5**, pp. 113-124.

Crain, I.K. and Mc Donald, C.L., 1984, From land inventory to land management. *Cartographica*, **21**, pp. 40-46.

Davis, F.W., Stoms, D.M., Estes, J.E., Scepan, J., and Scott, J.M., 1990, An information systems approach to the preservation of biological diversity. *International Journal of Geographical Information Systems*, **1**, pp. 55-78.

De Sède, M.H. and Thiérault, M., 1996, La représentation systémique du territoire: un concept structurant pour les SIRS institutionnels. *Revue internationale de Géomatique*, **1/1996**, pp. 27-50.

ENR/OELM, 2000, *Diagnostic de territoire de la Côte d'Opale*. Publication Espace Naturel Régional / Observatoire de l'Environnement Littoral et Marin, http://www.enr-littoral.com.

Fabbri, K.P., 1998, A methodology for supporting decision-making in integrated coastal zone management. *Ocean & Coastal Management*, **39**, pp. 51-62.

Gourmelon, F., Bioret F., and Le Berre, I., 2001, Historic land-use changes and implications for management of a small protected island. *Journal of Coastal Conservation*, **7**, pp. 41-48.

Gourmelon, F., Liret, C., and Bonnet, M., 2000, Approche géomatique de l'habitat grand dauphin en mer d'Iroise. In *Coastgis'99: Geomatics and coastal environment*, edited by Populus, J. and Loubersac, L. (Brest: IFREMER-SHOM), pp. 186-197.

Gourmelon, F., Bioret, F., Brigand, L., Cuq, F., Hily, C., Jean, F., Le Berre, I., and Le Demezet, M., 1995, *Atlas de la Réserve de Biosphère de la Mer d'Iroise : inventaire numérique des milieux terrestres, intertidaux et marins* (Hanvec: Parc Naturel Régional d'Armorique).

Guillaumont, B. and Durand, C., 2000, The integration and management of regulatory data in a GIS: an applied analysis of the French coasts. In *Coastgis'99: Geomatics and coastal environment*, edited by Populus, J. and Loubersac, L. (Brest: IFREMER-SHOM), pp. 269-283.

Haines-Young, R., Green, D.R., and Cousins, S.H., 1993, *Landscape ecology and GIS*, (London: Taylor & Francis).

Holligan, P.M., 1994, *Land Ocean Interaction in the Coastal Zone (LOICZ) : Implementation Plan*. IGBP.

Howlett, E., Anderson, E., and Spaulding, M.L., 1997, Environmental and geographical data management tools for oil spill modelling applications. In *Proceedings of the 20th Arctic and Marine Oilspill Program (AMOP)*, Environment Canada, (Vancouver), pp. 893-908.

Journaux, A., 1985, *Cartographie intégrée de l'environnement un outil pour la recherche et pour l'aménagement*. MAB-UNESCO, ed. UGI.

Le Berre, I., 1997, *Réserve de biosphère de la Mer d'Iroise : carte de synthèse Conseil Général du Finistère*. MAB-UNESCO.

Le Berre, I., 1999, *Mise au point de méthodes d'analyse et de représentation des interactions complexes en milieu littoral.* Thèse de Doctorat, Géographie, Brest, UBO.

Le Berre, I., Gourmelon, F., and Liret, C., 2002, Modélisation bathymétrique de la mer d'Iroise, application à l'étude du grand dauphin côtier. *Revue internationale de Géomatique*, 12/2002, pp. 337-354.

Le Berre, I., Meyrat, J., and Pastol, Y., 2000, Application des données hydrographiques à l'étude synthétique de l'environnement côtier: exemple d'un SIG sur le littoral du Finistère (France). In *Coastgis'99: Geomatics and coastal environment*, edited by Populus, J. and Loubersac, L. (Brest: IFREMER-SHOM), pp. 233-244.

Michel, J. and Dahlin, J., 1997, *Guidelines for developing digital ESI atlases and databases*. NOAA.

Muller, F., Donnay, J.P., De Cauwer, K., Schwind, L., Devolder M., and Scory, S., 2000, Design of an oceanographic database. In *Coastgis'99: Geomatics and coastal environment*, edited by Populus, J. and Loubersac, L. (Brest: IFREMER-SHOM), pp. 124-133.

Van Zuidam, R.A., Farifteh, J., Eleveld, M.A. ,and Cheng, T., 1998, Developments in remote sensing, dynamic modelling and GIS applications for integrated coastal zone management. *Journal of Coastal Conservation*, **4**, pp. 191-202.

CHAPTER TWELVE

Cultural Intermixing, the Diffusion of GIS and its Application to Coastal Management in Developing Countries

Darius Bartlett and R. Sudarshana

12.1 INTRODUCTION

Whether viewed as system or as science, the origins and development of GIS are essentially rooted in Western (largely "Anglo-Saxon") geographies, sciences, and technologies. Indeed, Coppock and Rhind are even more specific, and point to the "dominant contribution of North America to the development and implementation of GIS up to the mid- and late-1980s" (Coppock and Rhind, 1991). In addition, the overwhelming majority of international scientific journals and educational texts in the discipline similarly originate in Europe and North America. They are also mostly anglophone, although recent years have seen the gradual emergence of a corpus of GIS literature in non-English, though still predominantly European, languages (e.g., for the French language, see Collet, 1992; Didier and Bouveyron, 1993; Pantazis and Donnay, 1996; Pornon and Hortefeux, 1992; as well as references in Bourcier, this volume, Populus *et al.*, this volume, and Gourmelon and Le Berre, this volume).

While the origins of GIS and most other Information and Communications Technologies (ICT) lie in the West, they are increasingly being applied to other parts of the world, and their diffusion may be seen as a step towards the eventual creation of global information infrastructures. This transfer of GIS and ICTs, as well as the know-how to apply them, is generally heralded as a "good thing" (Nag, 1987; Taylor, 1991; Hastings and Clark, 1991; Yeh, 1991; Salem, 1994; Alhusein, 1994; Metz *et al.*, 2000; Nwilo, this volume), and a growing corpus of literature attests to the benefits that can accrue to receiving countries from the acquisition and application of spatial information technologies.

Less well-publicised is the potential for cultural and other impacts on non-western societies as these tools and technologies become more widespread. In this paper, we explore some of the issues that we believe merit closer study in this

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

regard, drawing by way of example on the use of GIS for coastal zone management in the Indian sub-continent.

12.2 GIS DIFFUSION TO THE NON-"WESTERN" WORLD

The process of technology transfer is arguably as old as humanity itself. In the past fifty years, however, it has gradually gained increasing importance as an element in the quest for sustainable development, reduction of vulnerability to environmental hazard, and the opening up of new markets. In short, technology transfer may be seen as a cornerstone of globalisation.

Based on our own experiences in a number of technology and know-how transfer initiatives, as well as a critical review of the literature, we have come to the conclusion that much technology transfer, particularly from north to south, and from west to east, appears often to be predicated on a number of important, though often untested assumptions.

1. **Technology transfer is driven primarily by philanthropic notions of serving "the public interest."** In reality, much transfer of technology is driven by very different forces, such as the marketing efforts of large (mostly Northern) corporations, including GIS vendors, or the interests of donor governments seeking to obtain or retain political power and influence, pursue military strategies, influence global geopolitics, etc. These latter forces are rarely motivated by altruism, and may frequently seek to attach political or other conditions to the transfer: Alhusain (1994), for example, argued that the implementation of GIS in many developing countries is "still in its primitive stages… due to many wrong practices common in these countries… One of the most serious practices in these countries is that many of them - with some exceptions - are non-democratic societies." In order to help overcome these perceived deficiencies, he advocates that the transfer of GIS and other technologies should be explicitly linked to political change: "International organisations such as ISPRS and the UN… can play very important role in encouraging the process of transferring the IGIS (Integrated GIS) technology to developing countries … *and, particularly, the democratic practicing [sic] within these societies should be highly appreciated and encouraged*." (Alhusain, 1994, emphasis added).

2. **The notion that "new" technology equals "better" technology** and, linked to this, an implicit equation suggested between acquisition of technology and generally improved quality of life. This can lead to the acquisition of technologies becoming seen as a desirable and worthwhile goal in its own right, rather than as a means to an end. Thus the allure of GIS may be as much based on the wish to acquire a high-visibility "status" product, linking the owner with the developed world, as well as being a route towards accelerated national development. As Apfell-Marglin (1996) explains it, "The First World is 'developed' and the Third World is 'developing' or 'underdeveloped'. In these phrases the *telos* of development stands revealed and the superior results are there already, luring everyone 'forward.'" Particularly in recent years, uneven diffusion and take-up of spatial information systems, as well as the

Internet, and other Information and Communication Technologies, have led to the expression of fears among sections of the developing world that the technology "connects people who mostly belong to developed nations [and] excludes the major population from the developing nations" and even that "the developing countries will suffer from unequal distribution of scientific knowledge" and might thus be deprived of opportunities for development (Shanmugavelan, 2000). This fear of "being left behind" adds to the pressures on decision-makers in developing countries to embrace new technologies.

3. **Technology transfers will, indeed should, only consist of one-way (generally North-South or West-East) flows.** As currently practiced, the diffusion of GIS to the developing world exemplifies and reinforces this belief. However, there is in practice no reason why this should be the only pathway available, and there may be circumstances in which a reverse flow – that is, a transfer of ideas, technologies and methods, from South to North or from East to West - could be not only possible, but perhaps even desirable. The countries of the South may be less developed than those of the North, but they are not totally lacking in their own expertise, wisdom and ways of doing things: as the late president of Tanzania, Julius Nyerere, argued in a 1968 speech, "Knowledge does not only come out of books… We have wisdom in our own past and in those who still carry the traditional knowledge accumulated in that tribal past…we would be stupid indeed if we allowed the development of our economies to destroy the human and cultural values which African societies had built up over centuries" (quoted in Swantz and Tripp, 1996).

4. **Delivery of knowledge and capabilities to the recipient country is sufficient**, and subsequent diffusion of these tools or techniques within the country can be left to take care of itself. In practice, greater focus on internal diffusion is needed, especially from the often more cosmopolitan and progressive urban centres to the traditionally more conservative, and particularly undeveloped small towns and rural settings.

These observations, we believe, carry potentially important implications for the manner in which GIS is diffusing to, and being applied to assist coastal management tasks within, particularly developing world countries. The two areas where we believe these issues to be especially relevant are in the formalisation and representation of conceptual models of the coast within GIS databases; and in the incorporation of GIS within the decision-making process. Each of these will be examined in more detail in the remainder of this paper.

12.3 CONCEPTUAL MODELS AND REPRESENTATION OF SPACE

GIS consists of a nested and interlinked series of conceptual, logical and data models (Peuquet, 1984) that unite the human user, the computing hardware and software, and the data and operations being performed, into a more-or-less integrated and synergistic whole. The resulting system is a metaphor which we

hope embodies all the elements and properties of what we perceive as important in the real world, while discarding the unimportant (Bartlett, 2000).

Measurement, description and apportionment of land appears to have always been an important feature of human society, to the extent that spatial thinking is probably a fundamental element of human intelligence (Chrisman, 1997). The *representation* of geographic information in some (carto)graphic form may also be seen in the histories of civilisations the world over (e.g. Noble, 1981; Thrower, 1996; Burrough and McDonnell, 1997; Godlewska, 1997; Harley and Woodward, 1989; Wilford, 1981). The Chinese, under Yü Kung, and the Greeks, under Anaximander, were reputedly making detailed maps and surveys as early as the sixth century B.C. (Thrower 1996). The diffusion of Buddhism from northern India into China was also taking place at about this time, and there were regular trading links and exchanges between China, India and the Roman Empire along the "Silk Road", as well as perhaps by sea. It may be assumed that much of the early European cartographic thinking that was emergent at this time was derived from, or at least influenced by, eastern philosophies and ideas.

The intervening centuries, however, have seen the ascendancy of western map-making. This cartography, while it almost certainly borrows unconsciously from other traditions, is nonetheless rooted firmly in western science, western world-views and western ideas of effective information communication. The net result has been the effective globalisation and hegemony of a cartography based on Euclidean distances, Cartesian space, the conceptual separation of time and space, and cartographic design rules based on western aesthetics. It is this cartography, enhanced through the application of Boolean logic and implemented through the medium of binary arithmetic, that we find embedded in the representation of geographic space within current GIS. These European (and, partly by derivation North American) scientific-cartographic paradigms have thus become the *de facto* worldwide standard for geographic data modelling, to the effective exclusion of all possible alternative models (Bartlett, 2000).

12.3.1 Geography, GIS and metaphysics

While the metaphors and paradigms embedded in GIS are thus western in origin, the applications of GIS are increasingly extending to other cultures and traditions. Little attention appears to have been given, until now, to the implications underlying this essentially top-down convergence of western technologies with non-western thinking and modes of practice.

White (1967) has observed that human ecology is "deeply conditioned by beliefs about our nature and destiny - that is by religion," while Robinson (1982, quoted in Singh, 1999) reminds us that, "geographers and most other researchers, consciously or unconsciously, have tended to assimilate a scientific perspective which in turn is based partly upon certain metaphysical assumptions concerning the nature of relationship between man and environment." In practice we implicitly know that western perspectives on science, which are largely based on Judaeo-Christian ethics and the application of Newtonian physics and mechanics, are merely subsets of an enormous range of possible world views and approaches to human ecology. There are fundamental and very important differences in the "metaphysical assumptions" and views of human-environment relationships,

between the major religions and cultures of the world. Ipso facto, alternative world views and mental maps must also exist although, given their diverse cultural and geographic origins, these "maps" may take very different form and appearance from the types of cartography more usually encountered in the West. For example, if the late Bruce Chatwin's hypotheses are to be believed, even the Aboriginal "songlines" may represent a set of "musical maps" of the Australian continent, and a mapping of space relayed through oral tradition rather than via any graphic medium (Chatwin, 1988).

As is well known, coastal zone management remains an area of potential GIS application where major challenges are encountered. Many of these difficulties are due to pervasive inherent weaknesses in the conceptual and data models used by these systems, that mitigate against effective multi-scalar representation of highly dynamic objects with indeterminate boundaries. Although progress is being made towards resolving these issues, it is perhaps worth asking whether any solutions or insights to better representation of coastal space might be found through exploring alternative, non-western perspectives and conceptual models, and seeking to embody these within our computer representations? Such an exploration would require knowledge of computer technology to be interfaced with concepts and methods drawn from the social sciences including, perhaps, cultural and social anthropology. To the best of our knowledge, no such collaboration between these very distinct disciplines has yet been undertaken.

The Indian sub-continent would appear to be a particularly promising place to begin such research. As well as being an ideal "laboratory" for testing and further refining any leads that might suggest themselves, due to the extent, diversity and importance of the coastal zone, in India the predominant Hindu-Buddhist philosophies interface with western, Judaeo-Christian traditions, and the products of these, in virtually every sector and aspect of daily life. India is also a land of strong contrasts between traditional, frequently feudal-based rural societies on the one hand and, on the other, a burgeoning Information Technology sector – including its own space programme and orbiting remote sensing platforms – to match anything in the West.

12.4 THE APPLICATION OF GIS TO COASTAL ZONE MANAGEMENT

One of the primary reasons for wishing to model coastal space in a GIS framework is so that the technology may be used for enhanced decision-making and territorial management. Humanity has always had a close relationship with the coast, with at least 40% of the world's population now living within the coastal zone (Carter, 1988; Bartlett, D.J. and Carter, R.W.G. 1990). The oceans, and especially the continental margins, provide resources, living space, industry, and the locale for a growing leisure and tourism activity, all of which may contribute significantly to the social and economic development of society. The coastal zone may also be an important milieu for confrontation and intermixing of cultures, particularly as a growing global population competes for ever scarcer resources, and as globalisation brings yet more "exotic" coastal regions – often located within the developing world – into easy reach of leisure- or adventure-seeking holiday-makers, especially from the more affluent nations of the world.

Finally, the coastal zone is one of the world's more hazardous regions in which to live and work (Carter, 1988). Specific dangers include the risks of flooding and shoreline erosion, which are themselves often caused or exacerbated by extreme weather events (hurricanes, cyclones) or geotectonic disturbances (e.g. tsunami arising from earthquakes) (Burton *et al.*, 1993; Smith 1996). On top of these lie the indirect impacts and second-order consequences of disaster, arising through disruption of economic and social systems. The nature of these consequences may vary between urban and rural areas, as well as between developed and less-developed countries but, wherever they occur, their effects can be far-reaching and severe (Smith, 1996). Continued growth of world population, allied to recent warnings about possible near-future changes in global climate and accompanying rises in sea-level, increases in storm frequency and impact, etc., suggest the coast is likely to become more, rather than less, hazardous in the future (e.g. Titus, 1987; Carter, 1988; Devoy, 1992).

12.4.1 Western approaches to coastal management

The challenge of coastal management is to reconcile human and non-human uses of the shore. Unfortunately, this goal has more often remained elusive than achieved, prompting Carter to remark that humanity "has an uneasy relationship with the coast. Throughout history, we have tried to ignore it, adjust to it, tame it or control it, more often than not, unsuccessfully" (Carter, 1988). Until comparatively recently, coastal management in the West (and in those countries influenced by western thinking) has tended towards increasing reliance on engineering methods, and the application of various technologies – including GIS – in order to resolve specific problems.

This technocentric approach finds several echoes in Western (Judaeo-Christian) approaches to nature generally, and traditionally sees nature/the world as being provided for human use. It is based on an essentially linear view of time, and on reductionist thinking where the progress of science is directed towards ever deeper understanding of ever smaller parts of the picture, with comparatively little attention being given to the whole (Koestler, 1969). We find that this philosophy of "nature to be commanded" has impacted on the terminology used to define and delineate coastal management objectives with, for example, recurring use of the language of conflict and expressions such as "protecting the coast", "armouring the line of the shore", "defending the beach", and even "reclaiming land" (when, in many cases, what is being referred to is not a re-taking of that which was once lost to the sea, but instead a carving out and colonisation of new territories we never had in the first place!). Carter conveys the flavour of this ethos very well when he quotes Sir John Rennie's Presidential address to the Institute of Civil Engineers, London, in 1845: "Where can Man find nobler or more elevated pursuits than to interpose a barrier against the raging ocean (Carter, 1986)?"

While this particular sentiment was expressed one and a half centuries ago, it is still encountered regrettably frequently to the present day, and inevitably makes its way into the strategies, algorithms and models adopted when applying GIS to the coast. In the West, most coastal management problems are likely to be articulated in the language as alluded to above; the questions for which answers are sought will be framed by the assumptions and beliefs espoused by the society

the coastal manager is part of; and the science policies adopted in order to resolve the situation will likewise be determined largely by the cultural and institutional contexts. Thus, traditional applications of GIS to coastal management have likewise drawn heavily on civil engineering and CAD-based approaches, the goal of which is most frequently to "stabilise" the line of the shore. Social, cultural and, indeed, environmental aspects of the problem have often tended to take a subsidiary role.

12.4.2 Eastern approaches to coastal management

In the East, and especially in India, the situation is much more complex. Many Asian countries were under European (in the case of India mainly British) rule for a few centuries, and the pre-existing feudal system of resource use was confronted with bureaucratic administration based on European principles. Thus we see a convergence of "East" and "West". Under colonial influences, management decisions were made, and resources categorised, largely on the basis of revenue unit polygons. Such units have spatial expression, and are relatively easy to represent in a traditional GIS. However, operating according to a completely different geography, the feudal system still exists or has re-emerged (in some places in the form of Mafioso-style organised crime) in many parts of post-colonial India, and decisions concerning the fate of a resource are now no longer expressed as per-revenue units that are easily interfaceable with GIS. Debt repayment systems based on lines of bartering, species segregation, accounting of fish catch instead of geographic accounting, *etc*., are all local values that stand distinctly different in comparison to the pattern of accountability in the west. The complexity of applying GIS to such management contexts lies in the fact that the local governments attempt to keep records in the Western pattern, while resource management is grossly influenced by indigenous, informal and feudal beliefs inaccessible to the procedures of modern information systems.

One of the basic assumptions underpinning both Western-style coastal management, and also implementation of a GIS, is the laminar structure of environmental management process. This philosophy of "layers" projects the nature of the problem, as well as the knowledge, decisions and actions needed to solve it, in a stratified manner. Within GIS, these "layers" are most frequently recognised in the thematic stratification of information. They also, however, find expression in the institutional contexts that frame uses of the technology. The layers in an organisation are generally distinguished by expertise, experience, hierarchical influence and a divisional or departmental approach to achieving work objectives. They can only be established and interlinked by the social process of *co-operation*. But whereas co-operation is a frequently encountered social value of Western society, *subjugation* and conflict at a human level are the dominant social values of the East, arising from the roots of the civilisations concerned.

This carries important implications for the use and transmission of information, since the co-operative laminar approach depends more on mechanical and modular interpretation of information through an iterative process, while the subjugative approach lends itself to single-window interpretative mode based more on human bias, instincts and an individualistic ethos. Hence, it may be argued that

machine-based knowledge and decision-making will always find less acceptability in the East in comparison to the West. This applies in particular to the process of automated information extraction, e.g. by GIS, which finds broad acceptability in the Western mind whereas manual charting and information extraction methods are much more common in the East. Again, in the latter, responsibility for the validation of information and interpretation is hierarchical, and finds its acceptance once more in the unquestioned or less questioned 'subjugation' social ethos.

As was seen above, human affairs in Asian society tend to be based strictly on a feudal approach to responsibility and authority, and thus based on an ethos of domination by one group or individual over another. When it comes to human-nature interactions, however, a somewhat different ethos prevails. Here, when compared to Judaeo-Christian philosophies, Eastern traditions appear to give far greater emphasis to a fundamental interconnectedness of things; they place great stress on the importance of cycles (and also see time as a circular rather than a linear concept); and teach a need to accept and adjust to change rather than resist it. In short, they would appear to be much more inherently compatible with the goals of sustainability and integrated coastal zone management.

Assuming that the GIS has to cater to the process of coastal decision-making, such issues again become important, and it is once more necessary to debate on the background social realities of decision-making. We know that decisions are either made on a Long Term Perspective (LTP) or with a Short Term Perspective (STP). LTP aims at qualitative values in deliverables and is prompted by slowly emerging needs, with fewer and diminishing contingency methods built in to the system of delivery. However, STP has the urge and need to cater rapidly and in a quantitative manner, since speed and precision of delivery may be critical to sustainability and survival. These contrasting perspectives are also discernible between East and West: Western societies tend to be more stable and secure, and hence can afford both the resources and the time required for longer-term planning. In less-developed countries, however, the concerns and the priorities are usually very much more immediate and focused on issues of survival and daily sustenance. These will inevitably have implications for the design of appropriate information systems. Hence, system modelling and both the design and implementation of a GIS needs to be very meticulous and detailed under the LTP mode while it only has to be at a practically manageable level to yield the effective STP that is vital to many Eastern societies. It is also a reality to work with STP in many politically less stable social systems. Such being the case, there is a need to revisit basic GIS principles and practices for the East, to cater to faster but less precise (i.e. gross) decision-making.

12.5 CONCLUSION

Ultimately, integrated coastal zone management should be *informed* coastal zone management. This evidently requires access to appropriate, timely and reliable data and information which, in turn, suggests an important role for GIS and other suitable information technologies. However, it is worth bearing in mind the cautionary message of Henry Nix who, in his Keynote address to the 1990 AURISA conference, pointed out that:

> *"Data does not equal information; information does not equal knowledge; and, most importantly of all, knowledge does not equal wisdom. We have oceans of data, rivers of information, small puddles of knowledge, and the odd drop of wisdom"* (Nix, 1990)

If we truly desire to know the coast, and manage its resources wisely, we urgently need to look beyond the confines of our current technologies and our scientific (and social) paradigms.

We do not intend to suggest here that Indian, or for that matter any other, cultures and philosophies hold *all* the answers to solving current problems of coastal GIS. Nor do we wish to suggest that GIS and other "Western" technologies have no role to play in the developing world. Instead, it is the intention of this paper to argue for a greater receptivity to other ideas and approaches; a greater willingness to step outside the boundaries of western-dominated methods of science; and finally, a greater appreciation that introduction of GIS as part of wider technology transfer programmes needs to be carefully attuned to the cultural and other social contexts of the receiving communities. We believe that if the diffusion of GIS around the world is conducted in a spirit of genuine collaboration and desire to share experiences among and between cultures, then nature, as well as humanity, is likely to be the richer for the endeavour.

12.6 ACKNOWLEDGEMENTS

This paper evolved out of discussions between the authors that took place during and subsequent to a workshop on "Subtle Issues in Coastal Management" (Sudarshana *et al.*, 2000), which was held in Dehra Dun, India, in February 2000. Darius Bartlett gratefully acknowledges generous funding from the Embassy of India, Dublin, which was crucial in enabling his participation in this workshop.

12.7 REFERENCES

Abdel-Kader, 1994, Requirements for Implementing Regional GIS Systems to serve Developing Countries, in *Proceedings of an International Workshop on Requirements for Integrated Geographic Information Systems.* 2–3 February (New Orleans, Louisiana, USA), pp. 75–87.

Alhusain, O., 1994, Implementation of Integrated Geographic Information Systems in Developing Countries - A Special Need, in *Proceedings of an International Workshop on Requirements for Integrated Geographic Information Systems.* 2–3 February (New Orleans, Louisiana, USA), pp. 89–94.

Apfell-Marglin, F., 1996, Introduction: Rationality and the World. In *Decolonizing Knowledge. From Development to Dialogue*, edited by Apfell-Marglin, F. and Marglin, S.A., (Oxford, England: Clarendon Press).

Bartlett, D.J., 2000, Cultures, Coasts and Computers: Potential and Pitfalls in Applying GIS for Coastal Zone Management, In *Hidden Risks: Subtle Issues in Coastal Management*, edited by Sudarshana, R., Mitra, D., Mishra, A.K., Roy,

P.S. and Rao, D.P., Dehra Dun, India: Indian Institute for Remote Sensing (National Remote Sensing Agency) pp. 33–44.

Bartlett, D.J. and Carter, R.W.G., 1990, Seascape ecology: The landscape ecology of the coastal zone. *Ekologia (CSFR),* **10,** pp. 43–53.

Burrough, P. and McDonnell, R., 1997, *Principles of Geographical Information Systems,* (Oxford: Oxford University Press).

Carter, R.W.G., 1988, *Coastal Environments,* (London: Academic Press).

Chatwin, B., 1988, *The Songlines,* (Harmandsworth, Essex: Penguin).

Chrisman, N.R., 1987, Design of information systems based on social and cultural goals. *Photogrammetric Engineering and Remote Sensing*, **53**, pp. 1367–1370.

Collet, C., 1992, *Systèmes d'information géographique en mode image,* (Lausanne: Presses Polytechniques et Universitaires Romandes).

Didier, M. and Bouveyron, C., 1993, *Guide économique et méthodologique des SIG,* (Paris: Hermes).

Godlewska, A., 1997, The idea of the Map. In *Ten Geographic Ideas that Changed the World*, edited by Hanson, S., (New Jersey: Rutgers University Press), pp. 15–39.

Harley, B. and Woodward, D. (eds.), 1989, *History of Cartography*, (Chicago: University of Chicago Press).

Koestler, A., 1969, Beyond atomism and holism – the concept of the holon. In *Beyond Reductionism: New Perspectives on the Life Sciences (The Alpach Symposium)* edited by Koestler, A. and Smythies, J.R., (London: Hutchinson).

Metz, B., Davidson, O., and Van Wie, L.L. (eds.), 2000, *IPCC Special Report on Methodological and Technological Issues in Technology Transfer*, (Cambridge, UK: Cambridge University Press), in prep.

Nix, H., 1990, *A national geographic information system - an achievable objective?* Keynote address, AURISA.

Pantazis, D.N. and Donnay, J.-P., 1996, *La Conception de SIG: Méthode et formalisme,* (Paris: Hermes).

Peuquet, D., 1984, A conceptual framework and comparison of spatial data models. *Cartographica,* **21**, pp. 66–113.

Pornon, H. and Hortefeux, C., 1992, *Systèmes d'Information Géographique pour petites communes: Guide méthodologique,* (Paris: Les Ed. du STU).

Salem, B.B., 1994, Scientific Applications of GIS in Egypt: A commentary on the current status. In *Proceedings of an International Workshop on Requirements for Integrated Geographic Information Systems.* 2-3 February (New Orleans, Louisiana, USA), pp. 29–34.

Shanmugavelan, M., 2000, IT in developing nations. *Intermedia* **28**(1), (International Institute of Communication, UK).

Singh, R.P.B., 1999, Nature and cosmic integrity: A search in Hindu geographical thought. In *Nature and Identity in Cross-Cultural Perspective,* edited by Buttimer, A. and Wallin, L. (Dordrecht, the Netherlands: Kluwer) pp. 69–86.

Smith, K., 1996, *Environmental Hazards. Assessing Risk and Reducing Disaster,* (London: Routledge).

Sudarshana, R., Mitra, D., Mishra, A.K., Roy, P.S., and Rao, D.P., 2000, *Subtle Issues in Coastal Management,* (Dehra Dun, India: Indian Institute of Remote Sensing).

Swantz, M.-L. and Tripp, A.-M., 1996, Development for 'Big Fish' or for 'Small Fish'?: A study of Contrasts in Tanzania's Fishing Sector. In *Decolonizing*

Knowledge. From Development to Dialogue edited by Apfell-Marglin, F. and Marglin, S.A., (Oxford, England: Clarendon Press).

Thrower, N.J.W., 1996, *Maps and Civilisation. Cartography in Culture and Society*, (Chicago: University of Chicago Press).

White, L. Jr., 1967, The historical roots of our ecological crisis. *Sciences*, **155**.

Wilford, J.N., 1981, *The Mapmakers*, (New York: Random House).

Wright, D. and Bartlett, D.J. (eds.), 2000, *Marine and Coastal Geographical Information Systems*, (London: Taylor & Francis).

Yeh, A.G.O., 1991, The Development and Application of Geographic Information Systems in Urban and Regional Planning in the Developing Countries. *International Journal of Geographical Information Systems*, **1**, pp. 5–27.

CHAPTER THIRTEEN

The Use of GIS to Enhance Communications of Cultural and Natural Resources and Contamination

John A. Lindsay, Thomas J. Simon,
Aquilina D. Lestenkof, and Phillip A. Zavadil

13.1 INTRODUCTION

Five Pribilof Islands, volcanic in origin and remotely located in the Bering Sea, are home to Aleuts and the breeding grounds to 70% of the world's northern fur seal population and numerous seabird species. The once uninhabited islands were first occupied by the Russians in 1786 and later became controlled by the U.S. Government. At the time of the Russian incursions into the Aleutian Islands, some have estimated the Aleut population at 15,000-18,000. First Russia and then the U.S. relied on the forced labour of Aleuts relocated to the Pribilof Islands from the Aleutian Islands chain to harvest sea otters, fur seals, and arctic fox. From their first arrival and through much of the 20th century, *Unangan* (people) or Pribilof Aleut people relied on the islands' and Bering Sea's natural resources, including fur seals, sea lions, whales, arctic fox, walrus, sea bird species, and a variety of plants for customary traditional purposes and subsistence. Even today natural resources are vital to the survival interests of *Unangan* on the Pribilofs. *Unangan* of St. Paul and St. George Islands are the world's single largest ethnic Aleut community whose world population approximates 3,200 individuals. Fur seal, endangered Steller sea lion, introduced reindeer, halibut, crab and other marine invertebrate species, and plants, such as crowberry, continue to play a significant role in customary traditional practices, economic development, and the maintenance of cultural and ecological harmony.

Settlement terms under the Alaska Native Claims Settlement Act (ANCSA) of 1971 required the National Oceanic and Atmospheric Administration (NOAA), as the most recent of the former federal land managing agencies for these islands, to transfer more than 95% of the land area to the local Aleuts. A 1976 Memorandum of Understanding and a 1984 Transfer of Property Agreement (TOPA) incorporated the details of the property transfer between NOAA and various local entities. The settlement and subsequent legislations including the Pribilof Islands Environmental Restoration Act of 1995 and the Pribilof Islands Transition Act of 2000 required NOAA to restore the islands' environmental

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

integrity compromised by U.S. Government activities supporting its commercial fur sealing enterprise. Environmental concerns included numerous releases of petroleum fuel products associated with the disposal of used oil, overfilling of storage tanks, corrosion of storage tanks, pipelines, and barrels, as well as landfills for household wastes, construction and demolition debris, scrap metal and junked vehicles, boats, barges, and aircraft. In addition, military activities during World War II contributed to soil and groundwater contamination. More than ninety sites on the two inhabited islands, St. George and St. Paul, required evaluation and potential restoration in order to complete the land transfer. The islands' remoteness, approximately 2,200 statute miles from NOAA's base of operations in Seattle, Washington requires a minimum of eight hours in travel time. Weather extremes involving fog, snow, and wind frequently cause flight cancellation. Improving communications through such multimedia technologies as Internet, GIS, and video were identified early as absolutely critical to the restoration project's success.

In the context of public land transfer to private sector entities and the restoration of lands and structures located within a National Historic Landmark and District, the National Historic Preservation Act (NHPA) further mandates NOAA to mitigate for these actions. Mitigation efforts at a minimum will provide a historical legacy associated with the federal ownership period for the benefit of future generations. The legacy is expected to include a book, video documentaries, and a detailed Geographic Information Systems (GIS) project. This chapter focuses on the GIS project. The NOAA Pribilof Islands Environmental Restoration Project Office (PPO) entered into collaboration with the St. Paul Tribal Government's Ecosystem Conservation Office to build a GIS project that included not only extant conditions, but also traditional cultural and natural resource features. This chapter presents an ongoing project that began in 1999, the conclusion of which is not expected for several more years.

13.2 GIS APPLICATION

The environmental restoration project began nearly two decades ago. Numerous written reports included site locations and analytical results from hundreds of soil samples. Unfortunately, site boundaries and sample point locations have not been easy to recover because of poor referencing and a relatively monotonous environment that becomes overgrown with waist high vegetation or eroded by violent winds. A decision was made at the outset to map all features using differential global positioning systems (DGPS) and a GIS. This approach would allow the relocation of sites and sampling points in years distant if necessary, as well as aid in other efforts, such as defining the vertical and horizontal extent of contamination, and the calculation of volumes of soils requiring excavation. The PPO also recognized the importance of mapping historical, natural, and cultural resource features, such as gravesites, to aid the communities in their future management of the islands, and subsequently to comply with the NHPA and the Coastal Zone Management Act.

Alaska's high latitudes and the islands' remoteness challenge GIS application. Normally available geo-referenced base maps, such as topographic quadrangles and nautical navigation charts, are scarce, small scale, or nonexistent.

Early project acquisition and prosecution of DGPS data acquired with Trimble Pro XRS and corrected with post processing layered over existing electronic versions of nautical charts revealed positional errors requiring exhaustive research to rectify.

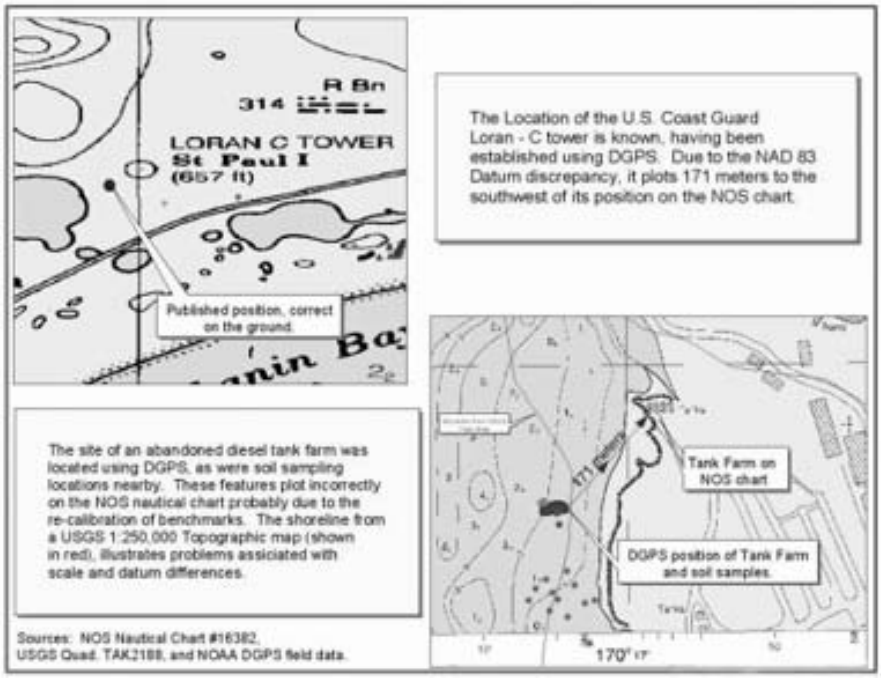


Figure 13.1 Errors in previous mapping projects.

The project resurveyed the two principal islands, St. Paul and St. George, to first order following National Geodetic Survey protocols. The results revealed a transcriptional error in an earlier version of the cartographic process that translated into an approximate 180-meter error (Figure 13.1). New benchmarks were installed on the islands, and NOAA/NOS issued new nautical navigation charts in March 2001. The chart scale, however, was still too small (1:50,000) for practical use. The NOAA project entered into a cost share collaboration with the US Geological Survey to provide Digital Elevation Models (DEMs) at 1:24,000 scale, topographic quadrangles at 1:25,000 scale with 10-meter contour intervals, and Digital Ortho Quarter Quads (DOQQ's) at 1:12,000 scale using pre-existing (1993) colour aerial photography. In an effort to acquire recent high-resolution imagery, NOAA contracted with a satellite vendor to provide 1-meter panchromatic and 4-meter multi-spectral imagery. The PPO is considering the application of interferometric synthetic aperture radar (IFSAR) imagery of the islands for use in establishing hydrographic controls to further improve island charting and to monitor long-term Aeolian erosion processes that have a potential to impact current restoration efforts. These products, when combined with historical maps and project data, will provide a comprehensive view of the islands and their changes over time.

The Pribilof Islands GIS mapping project uses ESRI's ArcView version 3.2. Federal Geographic Data Committee compliant metadata allows rapid data searches. The project focuses on several coverages and features (Table 13.1). As noted above, base maps include NOAA navigational charts as well as airphoto- and satellite-derived DEMs. Historical maps acquired from various sources including libraries and published reports dating back to the late 1800's provide a reference for changes occurring over more than a century. Changes of interest include fur seal, sea lion, and walrus rookery and haul out locations, as well as building and off-road trail locations. The maps are scanned at 1200 dpi and overlaid on registered and rectified base maps. The PPO uses Trimble Total Station RTK (real time kinematic) 5700 DGPS survey equipment to provide real time horizontal spatial accuracy of ±1 cm and vertical accuracy of ±2 cm. Acquired data is used for contouring and creation of 3-D models of landfill activities, excavations, contaminated soil stockpile volumes, and potentially for plotting seabird nests on cliff faces combined with laser spotting technology.

Table 13.1 Pribilof Islands GIS coverages and features

Coverage	Features
Geographic	Islands and land masses
Geologic	Soils, stratigraphy
Chemical	Soil and water
Demographic	Buildings, roads, debris sites; historic and current video and still images
Topographic	Cinder cones, landfill relief, DGPS ground surveys, aerial and satellite imagery
Biological (fauna and flora)	Fauna, flora, rookeries, haul outs, migration patterns, historic and current video and still images
Oceanographic	Bathymetry, currents, ice pack
Archaeological	Habitations, burial sites, historic and current video and still images

The authors' previous experience with marine GIS projects (e.g. Lindsay *et al.,* 1999; Butman and Lindsay, 2000) revealed the robustness of ArcView GIS. Using hotlinks the GIS project is enabled to include multimedia features, such as still, raster (laser line scan), and video imagery and sound recordings. The PPO uses Canon XL-1 digital and Panasonic 720p High Definition (HD) electronic media cameras and digital still cameras to document field activities. HD is down-converted to standard definition. Standard Def and digital videotape postproduction relies on MacIntosh G4 editing suites with Final Cut Pro software and DVD Studio Pro authoring software. Real GIS data are imported into a Hewlett Packard X4000 with XSI animation software to create textured, lighted, and animated production quality 3D models for viewer friendly but scientifically accurate renditions of real data. Historical video on Betacam and VHS tapes are converted to digital format using the same hardware and software. The GIS project itself and video mini-documentaries are placed onto DVD (digital video disk) using super drives. (Please note that any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.)

13.3 ENVIRONMENTAL RESTORATION

The U.S. Government initiated environmental restoration activities on the Pribilof Islands in 1983 following the final decision to transfer island lands from the public to the private sector. Former operations managers, residents, and historical reports aided in the identification of contaminant and debris sites. From a mapping perspective the primary emphasis of the environmental restoration activities on the Pribilof Islands has been on the locations of land features, boundaries, and the acquisition of soil and water chemical contaminants data. Demographic land feature focus has been on ownership, including federal land tracts and parcels targeted for transfer under the ANSCA and TOPA, topography, cultural uses associated with waste disposal practices, debris and contaminated waste site boundaries, building locations, utility locations (water, electrical, fuel transfer, cable, telephone and television lines), and grave site locations. Pribilof Islands' soil and groundwater chemistry has typically been evaluated at sites where spills or releases of petroleum products were suspected or known to have occurred, such as landfills and building sites.

A significant commercial infrastructure was built on St. George and St. Paul Islands to harvest and cure northern fur seal skins between 1872 and 1984. Numerous buildings were constructed for brine and salt curing, barrel making, skin packing, and seal by-product production, such as oil and animal meal. Other structures included electrical generation plants, gasoline stations, a radio station, fuel storage and transfer, commercial store, storage and utility buildings, a school, health clinic, quarries, and roads. As the value of the harvesting grew, so did the sophistication of the infrastructure and the use of petroleum based fuels for electrical generation and motorized vehicles. Storage tanks of various sizes ranging from about 500 gallons to 200,000 gallons and fuel lines serviced the islands' energy needs. Until about 1970, these tanks were filled from 55-gallon barrels brought ashore by bidarrahs, a native rowing craft with Russian influenced origins that were covered by sea-lion skins, and eventually by fuel barges with floating transfer lines.

Environmental restoration efforts prior to 1999 lacked either resolve and/or management experienced in restoration; consequently their efforts resulted in restoration of readily visible debris that included six thousand tons of surface debris, thousands of barrels either empty or containing used oil, and some underground storage tanks. Contaminated soils, groundwater, and landfills were left virtually untouched. In addition, initial efforts relied on conventional surveying and mapping techniques that archived features on paper maps without providing adequate geo-references. Subsequent field activities failed to recover site boundaries and soil sampling points. Short field seasons punctuated by years of inactivity resulted in as many as four characterizations at some sites in this treeless and knee-high grass environment in part due to the lack of geo-referencing and unique visual cues such as building foundations, stressed vegetation, soil staining, or surface debris. In an effort to capture these data, the PPO scanned existing paper maps, periodic air photos dating back to 1948, geo-referencing these when possible, and including them in the GIS project along with textual and tabular information. Since 1999, the PPO has relied on DGPS survey systems to log data points and vectors, to catalogue site features, and to relocate sites and sampling points. For the first time Pribilof Islands data was entered into a GIS

project that permitted near real-time project map creation and distribution. Maps were spontaneously created in the field office for planning and discussion purposes. Subsequent maps demonstrated the extent of soil and groundwater contamination, and served to increase the efficiencies of removal action decisions.

For example, diesel fuel contaminated sites are first verified by laboratory soil chemical analyses. Using GIS created maps depicting analytical results, and site historical and current features as guides, a removal or excavation is initiated within the bounds defined about these analytical results. These analyses seldom define within several meters the horizontal and vertical extent of contamination. Therefore, following initial excavation of contaminated soils, the project relies on thin layer chromatography (TLC) soil screening to verify whether remaining soils have diesel fuel contamination. The project has determined that field applied TLC based on standards offers a high probability (>95%) of accuracy within a sliding scale (e.g. 0-100 mg/kg, 100-250 mg/kg). TLC samples and their horizontal and vertical positions can be collected, processed, and the results entered into a GIS within approximately two hours. The sample results are overlaid onto registered and rectified historical air photo(s) depicting the historical structures, such as buildings, fuel tanks and fuel lines linked as sources of the contamination but long since dismantled. The composite photomap allows the cleanup team to substantiate the TLC results in context with the extent of the excavation and likely sources. If the map depicts a sample result with likely low levels of contamination, and no historical source was within close proximity to the spot, the cleanup team has received greater certainty about stopping further excavation. When excavation is considered complete, additional soil samples are taken for fixed laboratory analysis. The photomaps created during these field exercises can be displayed on laptop computer screens, printed on photographic papers, projected onto walls and plasma screens, and transmitted via Internet to far off locations.

High resolution (sub-meter scale) topographic maps of quarries and mineral stockpiles are created with the DGPS and GIS. On the islands, the project is required to purchase sand, scoria, and rock for use as backfill in excavations of contaminated soils, making haul-roads over the tundra, and capping landfills. Owners of the mineral rights require the purchaser to conduct pre- and post-surveys of the quarries to first approve and then verify quantities removed to establish a cost or royalty. The projects survey grade GPS in combination with the GIS are used to create the survey maps.

As mentioned above, the driving force behind the environmental restoration is the land transfer from the American government to local entities. These lands in the form of lots, parcels, and tracts were initially surveyed more than two decades ago using traditional surveying techniques by a separate federal agency. The available plats are in paper form and they are not easily applied in discussions involving numerous individuals and multiple issues that may involve structures, and contamination of soil and groundwater. In addition some shoreline changes have occurred since the original surveys. The project scanned and registered best fits of these plats and entered them into the GIS project for use as overlays with air and satellite photos. These enhanced maps from the original plats, hotlinked to still or video imagery, greatly improved communications between on- and off-island officials.

13.4 NATURAL RESOURCE NAMING CONVENTIONS

On and around the Pribilof Islands, marine mammals and seabirds dominate the environment. Only three terrestrial mammals are indigenous to the Pribilofs; they include the arctic fox, the black-footed lemming on St. George, and the Pribilof shrew on St. Paul. While English naming is the predominant convention followed today, some *Unangan* elders still refer to various species and/or their life stages in either *Unangam tunuu* (our people's language) or Aleut language, or Russian. For example, the arctic fox is called *aygagux* in *Unangam tunuu*. Fur seals are identified by sex and life stage in one or more languages. GIS mapping features can incorporate the various naming conventions applied to these species as appropriate.

13. 5 LOCATION NAMES

Present Pribilof Islands location names are also a mixture of *Unangam tunuu*, Russian, and English. For example on St. Paul Island, Lukanin Bay and beach, also spelled Lukannon, was named after the *promyshlenik* [fur hunter] who along with another, Kaiekov, killed 5000 sea otters during the first year of occupation in that area. Zolotoi Beach is literally interpreted as "Golden Sands" (Martin, 1960). Other present-day place names demonstrate American influence. On St. Paul, Hutchinson Hill acquired its name after Hayward Hutchinson, the first American to secure the commercial rights to the islands' fur seals.

The northern fur seal is the most intensively studied mammal on the islands. Records and maps of rookery locations, and haul out areas date back to the Russian period. Henry W. Elliot and others (*cf.* Elliot, 1881; Townsend, 1896; Osgood *et al.*, 1915) provided some of the earliest American rookery and haul out area maps. Rookery names as with other points of interest varied over time. For example, today's Northeast Point rookery on St. Paul was known during the Russian period as Novastoshnah Rookery while in *Unangam tunuu* it is *Chaxax*.

During NOAA's restoration activities the Aleut community expressed desire for greater recognition for their contribution to the islands' history. Although the literature is not devoid of *Unangam tunuu* place names, such names are uncommon on maps. Leaders on both islands had already begun to ascribe *Unangam tunuu* (Aleut place names) but official maps were wanting. NOAA enlisted the assistance of the U.S. Geological Survey (USGS) to procure digital aerial mosaics of the islands for the construction of base maps and the first large scale (1:25,000) topographic relief maps of the islands. These maps were released as a publication in March 2003. NOAA and USGS encouraged the local residents to document *Unangam tunuu* on these topographic maps and the USGS worked to secure their official status by having the place names reviewed and accepted as Geographic Names by the Alaska Historical Commission.

13.6 POPULATION DISTRIBUTIONS

While the Russian discoverers, beginning with the navigator Captain Gerassim Pribylov, may have recorded various marine mammal population statistics,

locating such records will require more concerted efforts than we have so far undertaken. But early American records clearly indicate that many of the aforementioned species at least once flourished on the islands.

Pack ice is said to influence the comings and goings of a variety of species, but most notably, walrus, arctic fox, and polar bear. Mastodon bones and molars have been recovered from lava tubes on the islands, suggesting either arrival *via* land bridge or pack ice. Purportedly in 1835, long lasting pack ice around St. Paul prevented haul out by pregnant fur seals, which resulted in the loss of both females and potential pups (Martin 1960). The pack ice in the early 1960's is reported by locals to have brought a polar bear to St. Paul, as well as several other times in the past. During 2000, pack ice was responsible for the occurrence of 6 to 7 arctic foxes on Walrus Island, a significant seabird rookery. These foxes are largely responsible for depopulating sea bird colonies, especially murres, on this small rocky island. Presently, NOAA has pack ice distributions in GIS format dating back to 1972. Ice pack incursions about the Pribilofs are being included in the GIS Project. More detailed information can be obtained through a weblink to NOAA's National Environment Satellite Data and Information Service's National Ice Center (www.natice.noaa.gov).

As noted elsewhere, fur seals received the attention of both the Russians and the Americans. Official duties of government agencies from both countries included mapping rookeries and haul out areas and maintaining various population statistics. With the passage of time, the influence of humans and the alteration of habitat has resulted in rookeries being modified or lost. For example, Webster's Point and the Salt Lagoon Rookeries are no more. GIS layering of historical areas is useful for the study of such changes, and the project also takes into account location names as ascribed over time within and between cultures.

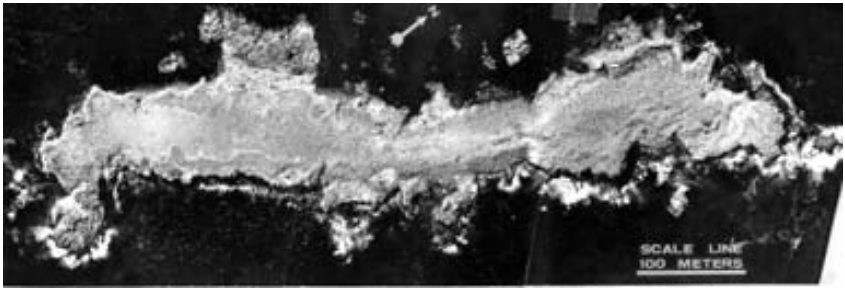


Figure 13.2 1948 aerial survey of Walrus Island.

The NOAA National Marine Mammal Lab has accumulated numerous aerial photographs of seal rookeries on St. Paul. The PPO scanned, registered, and rectified images from 1948, 1967, and the 1980's and included them in the GIS project.

Elliott noted (1875) only "a few sea-lions" bred on Walrus Island, although thousands could be seen among the fur seals at Northeast Point on St. Paul. Osgood *et al*. (1915) reported only two breeding rookeries: one of these was at Sea Lion Neck on St. Paul and the other near Garden Cove on St. George. But they also noted major haul out areas on all five Pribilof Islands. An aerial survey in 1948 found 1,258 adult sea lions on Walrus Island, but the resolution was

insufficient to determine the number of pups if any. The aerial photos have been registered and rectified for inclusion into the GIS project. During a rare visit to Walrus Island in 2000, our contingent observed 109 sea lions of which 11 were pups. During the same period fewer than 10 sea lions were observed on Otter Island. Today, despite the significantly reduced numbers, Walrus Island is the major sea lion rookery in the Pribilofs. Eleven species of seabirds nest on the Pribilof Islands. The red-legged kittiwake, as one example, occurs only in the Bering Sea and it has major rookeries on the Pribilofs. Only one shore bird, the rock sandpiper, breeds on the Pribilofs. The U.S. Fish & Wildlife Service monitors seabird nesting on the islands, but no electronic maps are available to date.

13.7 CULTURE AND SUBSISTENCE

Veltre and Veltre (1981) undertook the first focused comprehensive evaluation of the subsistence natural resources on the Pribilofs, although much of their work was acquired from historical documents dating back to the 1800's. They point out that the sealing industry "had significant ramifications on subsistence in the Pribilofs." Table 13.3 provides a listing of some representative natural resources on or about the Pribilof Islands and the customary traditional uses of each.

Two species have been virtually eliminated from the Pribilof Islands, the Pacific walrus and the sea otter. Neither animal apparently played much of a subsistence role on the Pribilof Islands in part because of their demise by the promyshleniks who sought the valuable ivory and pelts, respectively. Their historical haul out areas are included in the GIS project.

A third species, the Steller sea lion, is close to extinction on the Pribilof Islands as discussed above. Scans of historical maps depicting sea lion rookeries and haul out areas are layered onto the GIS project and verified observations of sea lions on shore will be logged into the project. Today, the sub-adult male sea lion constitutes a subsistence resource and is hunted primarily from the shore. Some hunting does occur on the water at St. George Island, however. Sea lion are hunted for their meat, blubber, internal organs, and fore flipper, although in earlier years the animal provided several customary traditional uses (Table 13.2). Historically, the dried flesh and hides of sea lions along with fur seals has long been a staple on the Pribilofs. The sea lion provided essential meat and clothing, as well as covering for the bidarrah.

The northern fur seal continues to play a major subsistence role on the Pribilof Islands with several hundred being killed on St. Paul and St. George each year. Until recently, the U.S. Government allowed a subsistence take of 2,500 young bachelor males to be harvested each year. Today, the Tribal Governments on each island and the NOAA National Marine Fisheries Service (NMFS) are co-managing the subsistence harvest of fur seal. Locations authorized for subsistence sealing by year and favourite hunting areas are being included in the GIS project. Several recent seal midden locations are being mapped as well.

Links from the GIS maps are used to display more detailed information on the food subsistence value offered by the northern fur seal. A link from the map to a photograph (Scheffer 1948) that laid out the various body parts provides even more detailed information (Figure 13.3).

Table 13. 2 Species list using either Aleut or English name found on or about the Pribilof Islands and their customary traditional use.

Subsistence Resource	Customary Traditional Use
qawax	Hides: baidarkas and bidarrahs (skin boats) Intestines: waterproof garments Throats: boot tops Fore-flipper palms: boot soles Stomachs: oil containers Whiskers: traded to Chinese Sinews from back: "thread" for sewing Gall bladder: healing sores Carcass, liver, lungs, lard: food (Elliott, 1881)
itxayan	Carcass: food
aygagux	Skin: sold for it fur value
kagayax	Eggs: food Carcass: food
chuchiigix	Carcass: food
sikiita	Skins and feathers: parkas Eggs: food Carcass: food
saquudax	Plant stem: food
qayux	Berries: food
Salmonberry	Berries: food
aguganax	Roe: food
waygix	*Body: food*

Today, as in the past, many seabirds have customary traditional uses. Millions of seabirds flock to the islands each year for breeding. Henry W. Elliott was the first American on the Pribilof Islands that accounted for their natural history and cultural subsistence through prolific writings extending over 20 years. He noted that large flocks of *chuchiigix* (*Aethia pusilla*) arrived in the spring and were harvested by *Unangan* with hand scoop nets. *Chuchiigix*, known as the least auklet, make nests in crevices among boulders. *Chuchiigix* are still hunted today by using a long bamboo pole to strike the small birds out of the sky (bamboo often washes up on the shores of the islands). Veltre and Veltre (1981) point out that this is one of the few species that can be hunted by children. Elliott (1881) reported that least auklets were replaced in late July by large flocks of red-legged turnstones that were also eaten by *Unangan*. Unlike the *chuchiigix*, the turnstones bred elsewhere, but they stopped on the Pribilofs to feed on the "flesh flies and their eggs, which swarm over the killing grounds." The killing grounds are not

nearly as extensive as they once were, but whether the abundance of turnstones on the islands was affected is not known.

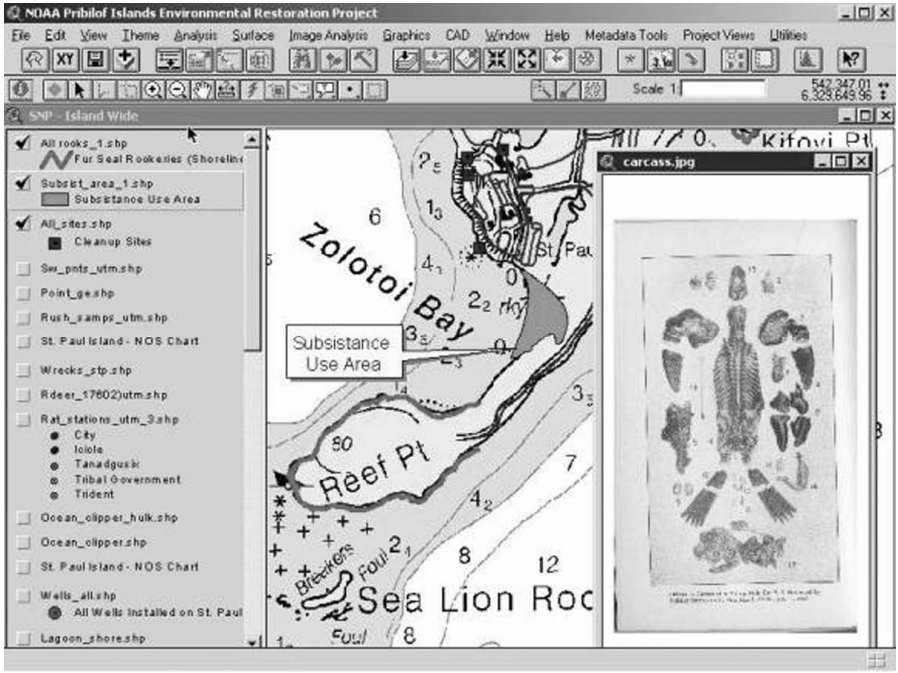


Figure 13.3 Example of hyperlinked media accessible from within the GIS: body parts and subsistence values of a northern fur seal.

Egg gathering was a major annual activity on the Pribilofs, although it is much less so now. Primary egg gathering areas included the steep rocky bluffs and the tabletop flat Walrus Island. According to Elliott, the primary egg-gathering season on Walrus Island was June-July (Elliott 1875). Historical egg gathering locations are included in the GIS project.

The Pribilof Islands are treeless, except for a low-to-the-ground-willow species. Several plant species afforded both cultural and subsistence value. For example, Elliot (1875) commented that only two berry plants, *Empetrum nigrum* and *Rubus chamemorus*, are readily harvested by the Aleuts during the end of August and first week in September. These species are still much sought after and often families spend summer afternoons harvesting the berries. *Saquudax* (wild celery) seemed to have had more importance in the past than it does now. A silent 1930's black and white film archived in the Oregon Historical Society shows Aleut children harvesting *saquudax* and demonstrating how it is eaten. A digitized version can be viewed via a hotlink when the distribution plots for the plant are active in the GIS project.

Heretofore, little documentation logged by the U.S. Government, Aleut community, or Russian Orthodox Church is known to exist as to the burial of deceased islanders. Some gravesites are marked, while many are known only

orally and most are not recalled at all. As a community service and as part of the interest in historical documentation, NOAA received permission from the local churches to develop a GIS project of the cemeteries. Using a survey grade GPS, more than 700 gravesites on St. Paul and St. George Islands were entered into the GIS project. Digital images of each location along with individual statistics (individual names and photo, birth and death dates, and short biographical sketches as available) are linked to the geographic locations. Printed hard copy maps were provided for use by the local churches and municipalities. Using DVD authoring software, the GIS project enhanced with animation and Aleut choir music was transcribed onto a DVD. NOAA provided the respective DVDs to the local church councils. In turn, the St. Paul Island council is relying on the Tribal Government for long-term maintenance of the GIS project. Recently, the Russian Orthodox Church Diocese of Alaska began promoting this GIS application throughout the Diocese.

13.8 LONG TERM NATURAL RESOURCE MANAGEMENT

Currently, Tribal Government of St. Paul's Ecosystem Conservation Office (ECO) is using GIS in combination with traditional ecological knowledge for use in environmental and resource management. Due to the increased human interactions with the Pribilofs' Bering Sea environment, the need to monitor and record these interactions and their impacts on St. Paul and neighbouring islands has become vital. Monitoring of commercial and customary traditional activities, the island ecosystem (i.e. numerous plant and animal species, lands, and environmental changes), in addition to ecological disaster prevention planning are essential in safeguarding customary traditional practices and for protecting and conserving the health of the islands. Critical response planning to prepare the community to effect immediate responses to future ecological perturbations is also vital as competition increases between traditional and commercial interests for the dwindling Bering Sea fishery resources.

The impacts associated with regional oil spills, off-road all terrain vehicle trailbreaking, and commercial fishing along with knowledge about decreasing populations of northern fur seals, Steller sea lions, and various seabird species serve to highlight a need for a balanced harmony between the subsistence users and other interests. Harmonious relations can be achieved through communication, planning, negotiation, and mutual work towards an understanding and respect for differences among groups. To this end, ECO has worked on ongoing development of *Tanalix Amgignax* (Island Sentinel/Guard) Program in coordination with the Pribilof Islands Stewardship Program, PPO, NMFS, and USFWS. This program includes the following elements:

- Promote "caretaking" and stewardship of the Pribilof Islands' many life systems.
- Provide a central forum that actively deals with environmental issues.
- Monitor activities that have potential adverse implications on this area of the Bering Sea environment.
- Promote a proactive position on behalf of all life systems indigenous to the Pribilof Islands Bering Sea area for immediate monitoring.

- Exercise the Aleuts' inherent right to be responsible for the islands' environment.
- Share caretakership concerns and activities with others, expecting visitors to be on their "best behavior" while visiting St. Paul Island and the surrounding sea.
- Promote environmental and cultural awareness.

Some specific *Tanalix Amgignax* activities that also involve GIS include: preventing rats from coming to the island by actively setting and maintaining rat traps in and around the harbour area and sanitary landfill; monitoring the islands' shoreline and interior for natural process impacts (weather and wildlife) and disturbances, such as vegetative changes by overgrazing reindeer, shoreline erosion, and derelict fishing gear (i.e. nets, buoys, line, etc.) that wash up on shore; removal of fishing gear from around fur seals' necks; and sharing the importance of taking care of the Bering Sea environment with the community and visitors alike. Rat trapping locations are entered into the GIS project. If and when any rats are captured this information will be similarly logged into the project. GIS mapping of derelict fishing gear not only provides information to managers for directing shoreline cleaning up efforts, but also documents the periodicity of cleanup efforts along shorelines, the relative quantities and types of debris recovered, and identifies areas where oceanographic conditions work to focus debris.

Tanalix Amgignax Program is a prime example of the use of GIS in bringing together traditional ecological knowledge and "western science." NOAA's contribution to the ECO GIS project will foster the dialogue among the many interests that the *Tanalix Amgignax Program* seeks to bring together.

13.9 CONCLUSIONS

GIS application in the Pribilof Island Environmental Restoration Project afforded insights and opportunities not contemplated at the outset of the project. The use of DGPS technology to simply map contaminated and debris site locations, and sampling points led to numerous significant contributions. One of the more routine contributions involved using JPEG and TIFF files cut from the GIS project to make maps for documents on topics such as site investigations, corrective action plans, corrective action reports, and requests to oversight regulators for no further action. Among other significant contributions came new nautical charts with improved near-shore confidence by local fishers in GPS navigation systems; map overlays of historical northern fur seal and sea lion rookeries for comparison with current and ever changing population and environmental conditions; estimates of shoreline erosion rates; improved control management of shoreline marine debris; near real time mapping of contaminant excavations allowing for on the ground decisions; sub-meter scale topographic maps to assist engineers in landfill closeout design and calculating royalty payments from sand and scoria (used as excavation backfill and capping solid wastes) mining; improved communications among officials who lacked on-island experience yet hold responsibility for land transfers following cleanup; and sorting through conflicting boundary descriptions arising from a variety of official documents over an approximate 25-year period.

A collaborative spirit between the government and Aleut entities grew out of training opportunities provided by NOAA in DGPS and GIS, access to NOAA DGPS equipment for ECO Projects, and contributions of GIS software to the City of St. Paul and ECO. Subsequent collaborative geo-spatial projects further improved overall communications. DGPS and GIS enabled local Aleuts to express their own cultural heritage by applying Aleut location names to the first ever large-scale topographic quadrangles of the islands. In terms of community relations, the most significant contribution afforded by DGPS and GIS came from the mapping of the islands' gravesites. Unwittingly, this single DGPS/GIS exercise offered up as a community service in context with a federal mandate to preserve historical information provided the strongest bond and trust between the islands' Aleut population and U.S. Government agents.

Acquisition of precision geo-spatial data in real time is one of many challenges in the higher latitudes. Geodetic control points, such as those offered by Continuously Operating Reference Stations (CORS), are far and few between in America. Alternatively, one can, as done in this project, survey in a base station to provide short-range control. For long-term applications base stations must be maintained against vandalism, climatic events, and curious animals such as foxes that chew through cables. Repeater stations serve to overcome multipath interference such as caused by hilly terrain and buildings. Once these obstacles are overcome, real-time DGPS in combination with GIS provide the investigator with tremendous power to work multiple field situations near real time as well as providing for long-term documentation.

13.10 REFERENCES

Butman, B. and Lindsay, J., 2000, *A Marine GIS Library for Massachusetts Bay: Focusing on Disposal Sites, Contaminated Sediments, and Sea Floor Mapping.* U.S. Geological Survey Open-File Report 99-439.

Elliott, H.W., 1875, *A Report Upon the Condition of Affairs in the Territory of Alaska* (Washington: Government Printing Office).

Elliott, H.W., 1881, *The Seal Islands of Alaska*, Bulletin Tenth Census of the United States. Dept. of the Interior (Washington Government Printing Office).

Lindsay, J.A., Simon, T, Graettinger, G., and Bailey C., 1999, A Global Offshore Hazardous Materials Sites GIS (GOHMS-GIS). In *Proceedings of Coastal Zone '99 Conference*, San Diego, California, pp. 392-394.

Martin, F., 1960, *Sea Bears: The Story of the Fur Seal.* (Philadelphia: Chilton Company).

Osgood, W.H., Preble, E.A., and Parker, G.H., 1915, The Fur Seal and Other Life of the Pribilof Islands, Alaska, in 1914. *Bulletin of the Bureau of Commercial Fisheries*, Vol. XXXIV, 1914; Document No. 820 (Washington: Government Printing Office).

Scheffer, V.B., 1948, Use of Fur-Seal Carcasses by Natives of the Pribilof Islands, Alaska. *The Pacific Northwest Quarterly*, **39** (2), pp. 131-132.

Townsend, C.H., 1896, Condition of Seal Life on the Rookeries of the Pribilof Islands. In *Reports of Agents, Officers, and Persons, Acting Under the Authority of the Secretary of the Treasury, Condition of Seal Life on the Rookeries of the Pribilof Islands and to Pelagic Sealing in Bering Sea and the North Pacific*

Ocean in the Years 1893-1895. 54th Congress, 1st Session, Senate, Document 137, Part 2 (Washington: Government Printing Office).

Veltre, D.W. and Veltre, M.J., 1981, *A Preliminary Baseline Study of Subsistence Resource Utilization in The Pribilof Islands*. Alaska Department of Fish and Game, Division of Subsistence (Contract 81-119) Technical Report Number 57.

CHAPTER FOURTEEN

GIS Applications In Coastal Management: A View from the Developing World

Peter C. Nwilo

14.1 INTRODUCTION

Geographic Information Systems have revolutionised the way spatial data/information is acquired, stored, managed, and displayed. In the past the acquisition of spatial data was done mainly with analogue theodolite and levels, and the recording and storage of these data were done in analogue manner. Data management was a very cumbersome exercise and the display was in the form of paper maps. Recent developments in information and digital technologies, communication and satellite technology, and improved computing speed and computer hard disk space, have impacted tremendously the way spatial data is acquired, stored, managed, and displayed. The development has created a revolution in spatial information technology in a way never experienced before. Today, virtually every human activity is spatial. Some schools of thought stipulate that in countries such as the United States and the United Kingdom as much as 75-80% of all the activities is spatial. There is correlation between geospatial information technology development and economic development. Developed countries of the world are associated with high spatial information technology development. Conversely, poor economic development is associated with low spatial information technology development as witnessed in developing countries of the world.

The development of Geographic Information Systems and associated technologies has created and is still creating employment opportunities all over the world. It has also led to changes of curricula of training institutions, and new training opportunities are being created. Transfer, sharing access to and development of spatial data standards have become major issues. We now discuss spatial data infrastructure as a major requirement for development, as has long been done with regards to other infrastructures such as roads, rail and waterways, electricity, and water supplies.

Geographic information has been employed in several fields such as management of disasters, vehicle tracking systems, forestry, utilities management, oil exploration, environmental management, health management, census, and

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

mineral resources development, governance (EIS Africa, 2002) and – the focus of this paper – coastal management.

14.1.1 Coastal Areas

The coast is an area that witnesses a substantial amount of physical and economic activities. It is a zone of great significance in the sense that intense agriculture, business and recreational activities take place here. Settlements have been established in these areas and international trade and communication originated in the coastal fringes (Nwilo, 1995; Nwilo *et al.*, 1995). The coastal zone houses about two thirds of the world's population and this percentage is expected to increase to 75% by the year 2020 (UNEP, 1992; Hanson & Lindh, 1993). Most of the important cities of the world, such as Alexandria, Lagos, London, Rotterdam, Shanghai, Tokyo and Venice, are located within the coastal environment, while two thirds of the world's cities having a population of over 2.5 million are within the coastal zone (Borrego, 1994). The coastal environment is very rich in both living and non-living resources. For example, a lot of crude oil and gas exploration and exploitation take place with the coast.

Wetlands are known to be a very productive ecosystem. Most of these wetlands are located in the coastal areas. They are the sanctuaries to many global endangered species and also can assist in dampening the effect of flooding from normal sea flooding, storm surges and sea level rise. Fishes that are caught in the upwelling zone of the Canary Island coast are hatched within the mangrove ecosystem of the Niger Delta and later migrate to the upwelling area. Similar relationships are found in most of the world's major fishing waters.

The coast is therefore a very important zone. Having discussed the importance of the coastal areas, it is now necessary to define the 'coastal zone.' Bird (1967; 1985 & 1993) defines the coast as a zone of varying width including the shore and extending the crest of cliff, the head of a tidal estuary, or the solid ground that lies behind the coastal lagoon, dunes and swamps. Komar (1976) on the other hand defines the coastal zone to include the littoral zone and extending further inland to include the sea cliffs, any marine terraces, dune fields and so on; the seaward limit is limitless. For most studies the coastal zone refers to that zone that is affected by what happens on the sea and on land. The coast is therefore very important for the existence of human life.

14.1.2 Integrated Coastal Area Management

Integrated Coastal Areas Management is defined as a dynamic process in which a coordinated strategy is developed and implemented for the allocation of environmental, socio-cultural and institutional resources to achieve the conservation and sustainable use of the coastal area (Coastal Area Management and Planning Network, 1989; Nwilo, 1995). Planning for sustainable resource management is based on weighing priorities, translating these priorities into policies and finally defining goals, identifying responsibilities for each step and establishing a time frame for action and review (Nwilo, 1995). There is no one

'right' way to manage coastal areas. However the design of a coastal management programme or policy for a nation or a region should take into consideration the legal instrument for the implementation of the program as well as the instrumental frame work for policy implementation (Emovon, 1991).

In order to practice effective coastal management, planners need to understand the way the natural environment and human activities are interconnected to form a system. Key aspects of the system include the following information themes:

i. Biological: This includes type and extent of ecosystem, primary productivity, specie diversity and abundance, nursery grounds and life cycles;
ii. Physical: This includes topography, geology, temperature, salinity, nutrients, tides, sea level and current, meteorology, sediment types and distribution, flooding and erosion/ accretion;
iii. Socio-economic: This includes human population distribution and growth, economic activities and land use;
iv. Legal and Institutional: Land tenure system, resource use rights, relevant laws and regulations, responsible agencies and availability of financial and human resources (Borrego, 1994).

Each of these activities is greatly influenced by activities within and beyond the coastal zone. It is for this reason that the resolution of conflicts in the use of coastal resources requires a broad perspective on the environmental process and interaction among human activities.

14.2 THE GUINEA CURRENT LARGE MARINE ECOSYSTEM PROJECT

14.2.1 The Concept of Large Marine Ecosystem Project

Large Marine Ecosystems are regions of the ocean space encompassing coastal areas from river basins and estuaries to the seaward boundaries of the continental shelves, and the outer margins of the world's current system. These marine regions are considered in terms of their ecological unity based on their distributive bathymetry, hydrographic, productivity, and trophic linkages (Sherman *et al.*, 1993; UNIDO, 2002). The world's 64 LMEs are the most productive ecosystems, and together produce 95% of marine fishery biomass yields. It is also in the LMEs that most of the global ocean pollution, fisheries over-exploitation and coastal habitat alteration take place.

The LME ocean management approach focuses on the sustainable development of the ocean resources and requires a paradigm shift from a small spatial scale to a larger one and from a short-term to a long-term perspective. It requires countries bordering the LME to set priorities on how to tackle their common transboundary issues. Since the early 1990s, the developing countries have approached the Global Environment Facility (GEF), the United Nations Industrial Development Organisation (UNIDO) and other United Nation implementing agencies for technical and scientific assistance in restoring and

protecting their coastal and marine ecosystem. GEF agreed to provide guidance and funding in addressing these issues within the framework of sustainable development. It recommended the use of the Large Marine Ecosystem (LME) concept and their contributing freshwater basins as the geographic focus for addressing the issues (UNIDO, 2002).

14.2.2 The Guinea Current Large Marine Ecosystem Project Implementation

The Guinea Current Large Marine Ecosystem Project is funded by the Global Environment Facility (GEF) and managed by the United Nations Industrial Development Organisation (UNIDO). The phase I of this project was started in 1995 and ended in 1999 with the countries of Cameroon, Nigeria, Benin Republic, Togo, Ghana and Côte d'Ivoire participating. Apart from GEF and UNIDO, some form of scientific and technical support came from the United Nations Education and Scientific Cultural Organisation (UNESCO), other UN agencies, and the US National Oceanic and Atmospheric Administration (NOAA). The participating countries provided in country support in the form of logistics such as provision of offices and transportation vehicles. The execution of the project was in form of modules. The modules that were covered in the project included Integrated Coastal Areas Management (ICAM), Geographic Information System (GIS), Mangroves, Industrial Pollution, Policy Issues, Fisheries, Coastal Erosion, Plankton and other pollution. The structure made provision for a National Project Director in each of the participating countries and a National Expert for each of the modules. All the national experts were reporting to an International Expert who co-ordinated each of the modules. The project had a Project Coordinator who was based in Abidjan, Côte d'Ivoire. The office of the Project Coordinator also served as the Secretariat for the project. In each of the participating countries, there was a Focal Point Agency and Focal Point Institution.

In Nigeria, the now-defunct Federal Environmental Protection Agency was the focal point agency, while the Nigerian Institute for Oceanography and Marine Research was the focal point institution. There were other participating institutions such as the University of Lagos. In the other participating countries, the focal point agency was the Ministry of Environment. Efforts were made in the project to collaborate with relevant institutions.

The second phase of the project is bringing in an additional 10 countries. These include Guinea Bissau, Guinea, Sierra Leone, Liberia, Congo Brazzaville, Democratic Republic of Congo and Angola. These countries, in addition to the original six, are in partnership activities for sustainable development of the Guinea Current Large Marine Ecosystem (GCLME). The addition of these to the original six countries meant that the full extent of the Gulf of Guinea Current Large Marine Ecosystem is covered. This stage of the project will commence in January 2004 (UNIDO, 2002).

All participating countries are making contributions to a Transboundary Diagnosis Programme (TAD) and a Strategic Action Programme (SAP) for the project (UNIDO, 2002).

14.2.3 The ICAM Module

The essence of the Guinea Current Large Marine Ecosystem project is the sustainable management of the living resources of the Gulf of Guinea region. The region was experiencing a serious environmental stress due to urbanisation, pollution from industries and domestic wastes, population explosion, oil spillage, deforestation, erosion, and over-fishing. The fish stocks were depleting at a fast rate and the socio-economic lives of the inhabitants are adversely affected. The lagoons that dot the coast from Côte d'Ivoire to Nigeria had become heavily polluted. Most of the problems highlighted above are transboundary in nature. There was therefore a need to reverse the situation and address these problems from a regional perspective rather than from a national perspective.

Under the module, an experienced ICAM professional prepared guidelines for implementation of ICAM in the countries. The National Experts prepared a country profile for each of the six participating countries in the first phase. The country profile is a baseline of the situation in each of the participating country's coastal area. The information used in preparing the country profile was obtained from existing literature and from relevant institutions. This was followed by a three-day workshop where stakeholders discussed the profile and made recommendations where necessary. A final copy of the country profile was made after effecting the necessary corrections from stakeholders.

An outcome of the workshop was the setting of the National Integrated Coastal Area Committee for each of the participating countries. Another success of the workshop and activities of the Guinea Current LME phase 1 is the creation of awareness on the importance of Integrated Coastal Area Management: prior to the Guinea Current LME project, there was no National ICAM Committee in any of the six participating countries.

As part of the efforts to assess the health of the Guinea Current LME, a fish trawl survey was carried out from Côte d'Ivoire to Cameroon. Results from the survey confirmed the fear that the fish stock, together with the plankton, was being depleted.

Figure 14.1 shows the Gulf of Guinea Region while Figure 14.2 shows the countries that took part in the first phase of the project

Figure 14.1 The Guinea Current Region

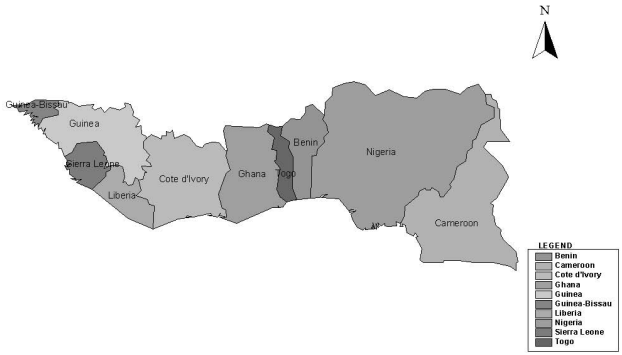


Figure 14.2 Countries that participated in the GCLME Project, Phase I

14.2.4 Application of GIS to The Project

As was the case in the Integrated Coastal Area Management Module, there was a National Expert for the GIS Module in each of the participating countries in phase I of the project. The author served in this position for Nigeria. The project started in 1995 but the National experts were engaged in 1997 for a two-year period. The contract was on part time basis.

The major goal of the GIS application in the project was to use GIS as an information and decision support system. Decision-makers are not always in a position to read all the volumes of materials that are passed to them. The graphic displays that are provided by GIS will invariably reduce the time that would have been spent in reading several volumes of materials, and assist in taking decisions that affect a project quickly. For example, it is easy to visualise the way a mangrove ecosystem of a region is changing over time through graphical displays.

14.2.5 Capacity Building for the GIS Module

The project provided for National Experts to attend Training workshops and conferences. A Project Expert Committee Workshop was organised by the Project Coordinator's office in Abidjan in April 1997 for all the National GIS Experts. This workshop touched on GIS application in the project and in particular the issue of metadata. Since the project is regional in nature the issue of harmonisation of different mapping datums in the region was discussed extensively. The National Experts provided information on institutions using GIS in the different participating countries. It was also a forum for experts to get to know each other. Another GIS Expert meeting took place in Abidjan between 29-30 September 1998 to deliberate on issues affecting the project.

14.2.6 SIMM Coast Workshop

SIMM is an acronym for sustainable integrated management and so SIMM Coast stands for sustainable integrated management of Coast. This workshop was organised by Prof. McGlade, formerly of the Centre for Marine Research, University of Plymouth, United Kingdom. It was essentially on the application of artificial intelligence in integrated coastal management. All the GIS National Experts were sponsored by UNIDO to participate in the workshop. This workshop took place at the Food and Agricultural Organisation office in Accra, Ghana between July 30th and 31st, 1998.

14.2.7 The Accra Declaration

The first meeting of the Ministerial Committee of the Gulf of Guinea Large Marine Ecosystem (GOG-LME) Project took place in Accra, Ghana, on 9th and 10th of July, 1998. Five Ministers with responsibility for the Environment respectively in Benin, Cameroon, Côte d'Ivoire, Ghana and Togo and the Director General/Chief Executive of the Federal Environmental Protection Agency of Nigeria attended the meeting.

After the deliberations of the ministers, which were based on extensive and substantive preparations, the Committee of Ministers adopted the Accra Declaration on Environmentally Sustainable Development of the Large Marine Ecosystem of the Gulf of Guinea. Part of the decision reached was that efforts shall be made to initiate, encourage and work synergistically with current and/or programmed national and international programmes on integrated coastal zone management in the region. The national concerns of flooding, and pollution caused by hydrocarbons, toxic chemical products, fisheries productivity and over-exploitation and, above all, coastal erosion, call for the special attention of donors. Also, data and information networking between the GOG-LME countries should be improved. National and Regional databases on the coastal and marine environment should be established using the Geographical Information System (GIS) to support decision-making, to be available to all users.

14.2.8 Results from the Phase I of the GCLME Project

As part of the success story of the GCLME project, guidelines for integrated coastal areas management and planning have been developed for Benin, Cameroon, Côte d'Ivoire, Ghana, Nigeria and Togo. Country coastal profiles were similarly developed and major issues and options for national and regional actions were identified.

The Gulf of Guinea Large Marine Ecosystem (GOG/LME) project has organised and facilitated a series of activities that have enabled each country to take the appropriate steps toward the adoption of Integrated Coastal Areas Management Plans, including strengthening legal and policy frameworks. In addition, these plans and policies are available for incorporation into National Environmental Action Plans. ICAM has proven to be the most effective way to involve the grassroots stakeholders in a meaningful way and could be the cornerstone for the GOG/LME potential Phase II Project.

Indications that member countries will embrace and make this effort sustainable are beginning to appear. First, National Steering Committees have been established by legal instruments. Second, the individuals who are participating in the programme range from top Ministry officials to citizens in NGOs and CBOs, indicating the existence of multilevel support.

An important parallel development with the effort to put National ICAM Plans in place has been the process of establishment of information management and decision-making systems such as Geographic Information Systems (GIS) at both the national and regional levels. ArcView 3.1 GIS software was chosen as the GIS platform for the project. It is hoped that in the second phase of the project, ArcGIS 8.3 will be the platform.

Some of the maps, which were developed from existing information in Nigeria during the execution of the project, are shown in Figures 14.3 and 14.4.

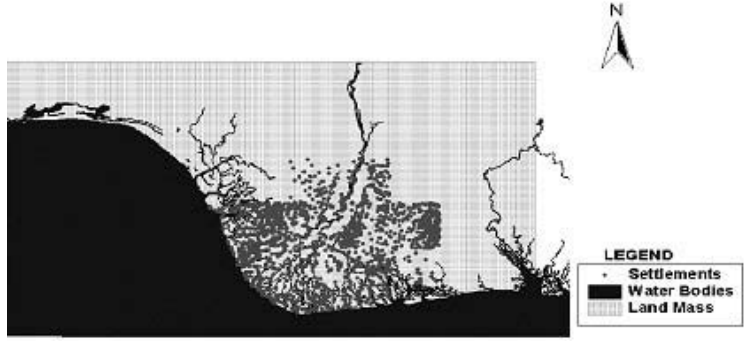


Figure 14.3 Settlements in the Niger Delta

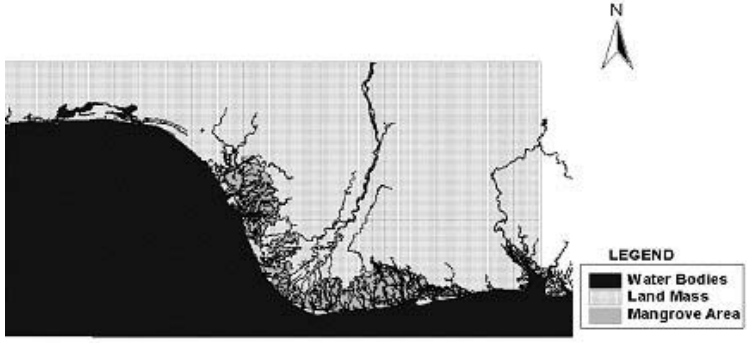


Figure 14.4 Mangrove Ecosystem in the Niger Delta

14.3 OTHER COASTAL AREAS PROJECTS IN AFRICA

14.3.1 Niger Delta Environmental Survey

The Niger Delta is an environmentally sensitive and fragile region owing to its peculiar natural physical setting and its distinctive ecological features and functions which over the years have drawn the attention of conservationists especially on the need for sustainable development and the protection of its biodiversity (NDES, 1997). Considerable changes are occurring in the ecological environment and in the socio-economic setting of the Niger Delta as a result of both natural and another anthropogenic transformations that include: upstream dam construction, coastal zone modification, urbanisation, deforestation, agriculture, fishing, industrial development, population pressure, and crude oil exploration and exploitation. The Niger Delta is therefore under increasing pressure from rapidly deteriorating ecological and economic conditions, social dislocation and tension in communities, problems that are not being addressed by current policy and behaviour patterns.

It is for this reason that the Niger Delta Environmental Survey (NDES) was set up. The NDES, in concert with communities and other stakeholders, undertook a comprehensive survey of the Niger Delta. It established the causes of ecological and socio-economic changes over time and induced corrective actions by encouraging relevant stakeholders to address specific environmental and related economic problems aimed at improving the quality of life or the people and achieving sustainable development in the region. Part of the objectives of the survey was to generate data and information on the Niger Delta including the establishment of a Geographic Information System (GIS). The data and information can be used in formulating strategies and plans for effective natural resource management towards the sustainable use of resources, in order to protect the environment and the livelihood of peoples of the region (NDES, 1997).

Analogue maps of the Niger Delta at scale of 1/25,000 were digitised and updated with remotely sensed images. The map information is now in a GIS environment, and is used in decision-making for the planning and management of the Niger Delta.

14.3.2 Secretariat for Eastern African Coastal Area Management (SEACAM) Project

SEACAM means secretariat for Eastern African Coastal Areas Management. It was set up in August, 1997 in Maputo, Mozambique, with the purpose of accelerating the implementation of integrated coastal zone management in the region as put forth in the Arusha Resolution of 1993 and Seychelles statement of 1996 on ICZM. The Secretariat is hosted by the Ministry of Environmental Affairs for Mozambique.

The objective was to assist Eastern African Countries to implement ICZM following up the Arusha resolution and the Seychelles statement on capacity building and information dissemination. It was supported by the Government of Finland, Danish International Development Assistance (DANCED), the World Bank and UNESCO. A number of other donors provide important support to select activities that match their programme areas.

In 1996, the Western Indian Ocean Marine Science Association organised an expert and practitioners workshop in Taga, Tanzania on integrated coastal area management, which provided an input into the second ministerial policy conference on integrated coastal and zone management in Eastern Africa and Island States held in Seychelles, October, 1996. It is not certain whether GIS was used in this project.

14.3.3 Lake Victoria Environmental Management Project

For decades, East African scientists, policy makers, stakeholders and others around the world noted the escalating environmental degradation of Lake Victoria, which is an international water body. It was observed that increased population is a major contributing factor. The total population of the riparian communities currently stands at 30 million.

As a result of the degradation the riparian countries to Lake Victoria, which are Kenya, Tanzania and Uganda, initiated discussions immediately after the United Nations Conference on Environment and Development (UNCED) to broaden regional cooperation in environmental and social issues affecting Lake Victoria Basin. This led to the signing of a triplicate agreement in 1994. The project attracted financial support from the International Development Association (IDA) and Global Environmental Trust Fund (GEF) to the tune of US $70.00 million. The participating countries agreed to contribute $7.2M as counterpart funds. Apart from other achievements of the project, mapping of the present land use/land cover and soil erosion hazard in the Lake basin is almost completed. Data on pollution loading are being collected from selected micro-catchments and provide useful information on soil nutrients and water losses. Human capacity is

being developed. An important component of the project is community participation in the preparation and implementation of the project. It is not clear whether GIS was used as an information management tool but it is obvious that there was mapping which could, if still in analogue format, be converted to digital format and managed in a GIS environment.

14.3.4 Information Technology for Coastal Zone Management of the Nile Delta

This is an initiative for development and implementation of an integrated coastal analysis and monitoring system. A workshop was organised between 7 – 10 April 2003 aimed at addressing issues such as coastal erosion, pollution, monitoring and fisheries management. It will serve as a model for information system implementation in the developing world context, where the need for operational coastal management is often great but for which information infrastructure and knowledge base are frequently lacking.

14.3.5 Eastern African Coastal Resource Maps Project

The Eastern African Coastal and Marine Environment Resources Database and the Atlas was founded within the framework of the 1985 UNEP brokered Eastern African Action Plan for the protection, management and development of the Marine Environment of the Eastern African Region (EAF). The countries of Comoros, La Réunion, Kenya, Madagascar, Mauritius, Mozambique, Seychelles, Somalia and Tanzania participated in the project.

The project was initiated in 1993 by the Ocean and Coastal Area Programme Activity Centre (OCA/PAC) in co-operation with the Global Activity Centre (GRID/PAC-Nairobi). The project was scheduled to run for five years. The first phase of the project only focused on Kenya while Phase II and III (1995-1996) concentrated on Comoros, Mozambique, Seychelles and Tanzania. Phase III and IV (1996-1997) concentrated on Madagascar, Mauritius, La Réunion and Somalia. Phase V is intended to round up all activities at the National Level.

The production of the digital database led to wiser use of resources and helped reduce wastage of non-renewable resources. The main objective of the project is to collect existing information on coastal resources of the region and to summarise the same in country map sheets managed in a GIS environment (Okemwa, 1995).

Scientists from the participating countries took part in the collaboration of the information and the development of the database. Information in the database covers parameters including but not limited to climate, hydrology, oceanography, coastal types, geomorphology, geology, tourist infrastructure, important ecosystems such as mangroves systems, wetlands, estuaries, rangelands and mineral resources.

The maps will be useful in studying the vulnerability of the coastline and indeed as a versatile tool in making strategic intervention in cases of oil spills.

14.4 CONCLUSIONS

Various coastal management projects in Africa have been discussed above. It is observed that most of these projects were conceptualised and executed by the developed countries, mainly European countries, while the African scientists only played supportive roles. Two projects in the opinion of the author stand out as clear exceptions to this situation. These are the Guinea Current Large Marine Ecosystem project and the Niger Delta Environmental Survey.

As earlier stated, the Guinea Current Large Marine Ecosystem project is funded by the Global Environment Facility and managed by the United Nations Industrial Development Organisation. Initially, the project was conceptualised by the governments of Nigeria, Ghana, Togo, Cameroon, Benin and Côté d'Ivoire. It was later realised that there was a need to include all the countries bordering the Guinea Current and this increased the number of participating countries to sixteen.

The other project, the Niger Delta Environmental Survey, was conceptualised by the oil prospecting companies in Nigeria. Its mission statement is: "In concert with communities and other stakeholders to undertake a comprehensive environmental survey of the Niger Delta, establish the causes of ecological and socio-economic changes over time and induce corrective action by encouraging relevant stakeholders to environmental and related socio-economic problems identified in the survey to improve the quality of life of the people and achieve sustainable development in the region" (Niger Delta Environmental Survey, 1996). Incidentally, these two projects are in Nigeria.

There is therefore a need to involve the African scientists and researchers from the conceptualisation stage of projects on the Continent. This is believed to be more beneficial to these countries and their scientists, particularly in the area of capacity building.

14.5 REFERENCES

Bird, E.C.F., 1967. *Coasts: An Introduction to Systematic Geomorphology*, (Cambridge, Mass.: The MIT Press).

Bird, E.C.F., 1985. *Coastline Changes. A Global Review*, (Chichester, England: John Wiley and Sons).

Bird, E.C.F., 1993. *Submerging Coasts: The Effect of a Rising Sea Level on Coastal Environments*, (Chichester, England: John Wiley and Sons).

Borrego, C., 1994, Sustainable Development of Coastal Environment: Why is it Important? In *Littoral 94*, edited by de Cavellio, S. and Gomes, V., pp 11-23.

Coastal Area Management and Planning Network, 1989, *The Status of Integrated Coastal Zone Management: A Global Assessment*, Workshop on Coastal Zone Management, Rosenteil School of Marine Sciences , University of Miami, USA.

EIS Africa, 2002, *Geoinformation Supports Decision-Making. An EIS Position Paper*

Emovon, E.U., 1991. National Science and Technology Policy and the Nigerian Environment. In *The Making of the Nigerian Environmental Policy*, edited by Aina, E.O.A. and Adedipe, N.O., pp 71-78.

Hanson, H., and Lindh, G., 1993, Coastal Erosion - An Escalating Environmental Threat. *Ambio*, **22**, (1), pp. 189-195.

Komar, P.D., 1976, *Beach Processes and Sedimentation*, (New Jersey: Prentice Hall).

NDES, 1996: Terms of Reference

NDES, 1997: Final Report Phase 1, Volume 1: Environmental and Socio-Economic Characteristics.

Nwilo, P.C., 1995, *Sea Level Variations and the Impacts along the Coastal Areas of Nigeria*, Ph.D Thesis, University of Salford, Salford, UK, 229p (Unpublished).

Nwilo, P.C., Onuoha A.E., and Pugh Thomas, M., 1995, *Monitoring Sea level/Relative Sea Level Rise in a Developing Country - The Nigerian Experience*, (Intergovernmental Oceanographic Commission).

Sherman, K., Alexander, L.M., and Gold, B.D., 1993, *Large Marine Ecosystem: Stress, Mitigation and Sustainability*, (Washington, D.C.: AAAS Press).

UNESCO, 1995, *Workshop Report No. 105 Supplement: International Conference on Coastal Change 95*, Bordeaux, France.

UNEP, 1992, *The World Environment 1972 - 1992*, edited by Tolba, M.K. and El-Kholy O.A., (London: Chapman & Hall).

CHAPTER FIFTEEN

High-Resolution Elevation and Image Data Within the Bay of Fundy Coastal Zone, Nova Scotia, Canada

Tim Webster, Montfield Christian, Charles Sangster, and Dennis Kingston

15.1 INTRODUCTION

The Applied Geomatics Research Group (AGRG) is a component of the Centre of Geographic Sciences (COGS) located in the Annapolis Valley, Nova Scotia. Its mandate is the application of geomatics technology for environmental research within Maritime Canada. In the fall of 1999 and summer of 2000 a large data acquisition campaign was initiated to collect high-resolution elevation and other remotely sensed image datasets along the Bay of Fundy coastal zone (Figure 15.1 and colour insert following page 164). The purpose of the research was to evaluate their effectiveness in obtaining critical information about the coastal zone, particularly data to be used to assess flood-risk potential associated with storm surge events. Global mean sea level has been increasing between 0.1 and 0.2 meters per century. With increasing greenhouse gases, sea level rise is expected to accelerate and the Intergovernmental Panel on Climate Change predicts that global average sea level may increase by 0.09 to 0.88 meters by 2100, placing the lives and property of an estimated 46 million people at risk (Houghton et al., 2001). The Bay of Fundy is no exception: relative sea-level is rising in this region by an estimated rate of 2.5 cm per century and many coastal areas are becoming more susceptible to flooding from storm events (Stea, Forbes, and Mott, 1992). In addition to sea level rise, storm surge and ocean waves are also factors at the coastline and are carried to higher levels on rising mean sea level. Storm surge in general is defined as the algebraic difference between the observed water level and the predicted astronomical level as one would find in tide tables. With possible increased storminess associated with climate change, the next 100 years will probably see more frequent flooding of coastal zones, and an increase in erosion of coastal features. With the recent increase in the spatial resolution of geomatics data available, both multispectral imagery (Ikonos, Quickbird, CASI) and high accuracy elevation data, landuse planners and policy makers now have access to the information required to manage the coastal zone.

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

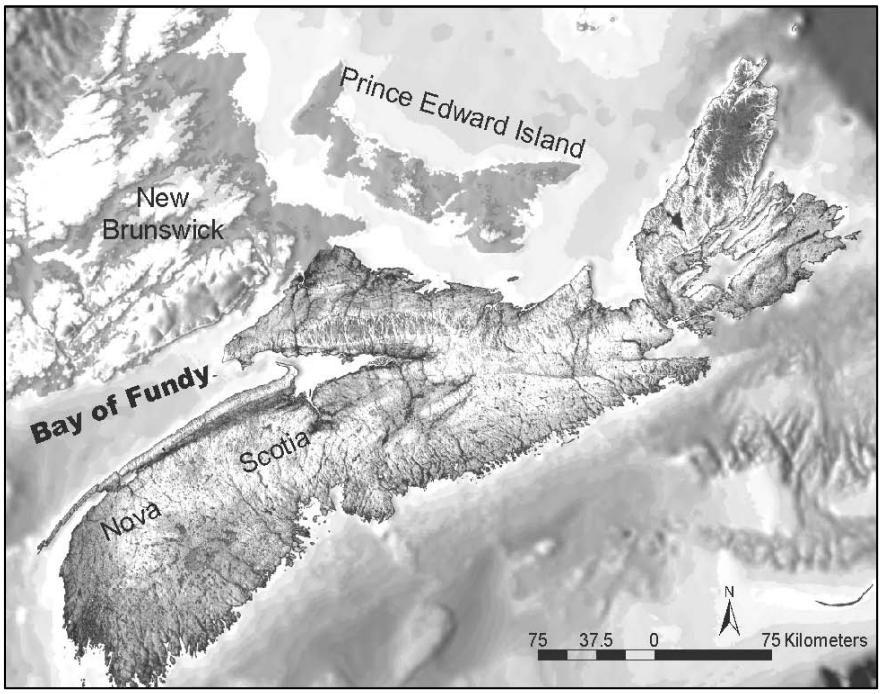


Figure 15.1 Location map, Bay of Fundy and Annapolis Valley (between Bay of Fundy and Nova Scotia label), Nova Scotia, Canada. This image is made up of a Radarsat S-7 mosaic for Nova Scotia, merged with a colour shaded relief map for the rest of Maritime Canada.
Radarsat data © 1996, Canadian Space Agency.

LIDAR (Light Detection and Ranging) technology has been employed for a number of years in atmospheric studies (e.g. Post et al., 1996; Mayor & Eloranta, 2001) and as an airborne technique for shallow bathymetric charting (e.g. Guenther et al., 2000), although cost was initially an impediment to widespread acceptance for the latter purpose. The technology can also be used to image the land and water surface (Hwang et al., 2000), as was done in the present study. Terrestrial LIDAR applications have been demonstrated in forestry (Maclean & Krabill, 1986), sea-ice studies (Wadhams et al., 1992), and glacier mass balance investigations (Krabill et al., 1995, 2000; Abdalati & Krabill, 1999). A general overview of airborne laser scanning technology and principles is provided by Wehr and Lohr (1999). Applications to coastal process studies in the USA have been reported by Sallenger et al. (1999), Krabill et al. (1999), and Stockdon et al. (2002), among others. Preliminary trials in Atlantic Canada were reported by O'Reilly (2000) and subsequent experience was described by Webster et al. (2001, 2002, 2003) and McCullough et al. (2002). Most of the coast of the conterminous USA has now been mapped using this technology (Brock et al., 2002). A comprehensive review of the theory and applications of Digital Elevation Models (DEM) covers both terrestrial and marine LIDAR as well as other technologies

used for DEM construction such as IFSAR – interferometric airborne synthetic aperture radar – is given by Maune et al. (2001).

Two different data providers were contracted to acquire LIDAR for the region. This chapter will discuss the details of the LIDAR systems, mission planning, data validation, and processing. The LIDAR DEM was then used in generating flood-risk maps associated with storm surge events along the coastal zone. This information has been passed to the local planning commissions to aid in management of the coastal zone. Another outcome of the project is that one of the data providers has significantly improved their acquisition system and approach to quality assurance when collecting LIDAR data.

15.2 DATA ACQUISITION AND PLANNING

Data that were acquired during the fall of 1999 were airborne polarimetric synthetic aperture radar (PSAR) data from the Convair 580 aircraft operated by Environment Canada, and satellite imagery from Radarsat-1, and Landsat-7. The polarimetric SAR was a simulation of Radarsat-2, Canada's second earth observation satellite planned for launch in 2005. Data acquired in the summer of 2000 campaign included high-resolution airborne data from the Compact Airborne Spectrographic Imager (CASI), and LIDAR systems, and more satellite imagery from Ikonos, Radarsat-1, and Landsat-7 (Table 15.1).

Table 15.1 Types of data collected during 1999, 2000 summers.

Data Type/Sensor	Resolution (m)	Attribute
PSAR	6	Polarimetric SAR signal, C-Band 5.6 cm, HH, VV, HV polarizations
Radarsat	12 (variable)	C-Band 5.6 cm, HH polarization, variable incidence angle
Landsat	15 pan, 30 mss	Visible, near and mid-infrared imagery
Ikonos	1 pan, 4 mss	Visible and near infrared imagery
CASI	2 (variable)	Visible and near infrared imagery
LIDAR	2 (variable)	Elevation (ground and non-ground)

Polarizations: HH – horizontal transmit, horizontal receive, VV – vertical transmit, vertical receive, HV – horizontal transmit, vertical receive.

The study area consists of the Annapolis Valley region of Nova Scotia, located on the southeast shore of the Bay of Fundy (Figure 15.1). LIDAR and CASI coverage consisted of the entire length of the valley and coastal zone. The satellite coverage was concentrated in the Annapolis and Minas Basins (Figure 15.2, see colour insert).

15.2.1 Acquisition Planning Issues

The Bay of Fundy is famous for its great tidal range, up to 13 m in this area. For such a large study area the timing of the tides vary by approximately 1 hour between Digby, within the Annapolis Basin, and the Minas Basin. Because of the variability of tide times within the study area, three locations were used to predict tide times and sites: Digby, Margaretsville, and the Minas Basin/Cape Blomidon. Acquisition of remotely sensed data at low tide has several applications including: 1. validation of tidal models; 2. determination of inter-tidal slope from derived elevations from the "waterline method" i.e. knowing the water depth at the time of image acquisition allows the land/water line to be used as a topographic isoline; 3. morphological and biological classification of the inter-tidal zone. Tide predictions were acquired for each port via the Internet (http://tbone.biol.sc.edu/tide/sitesel.html).

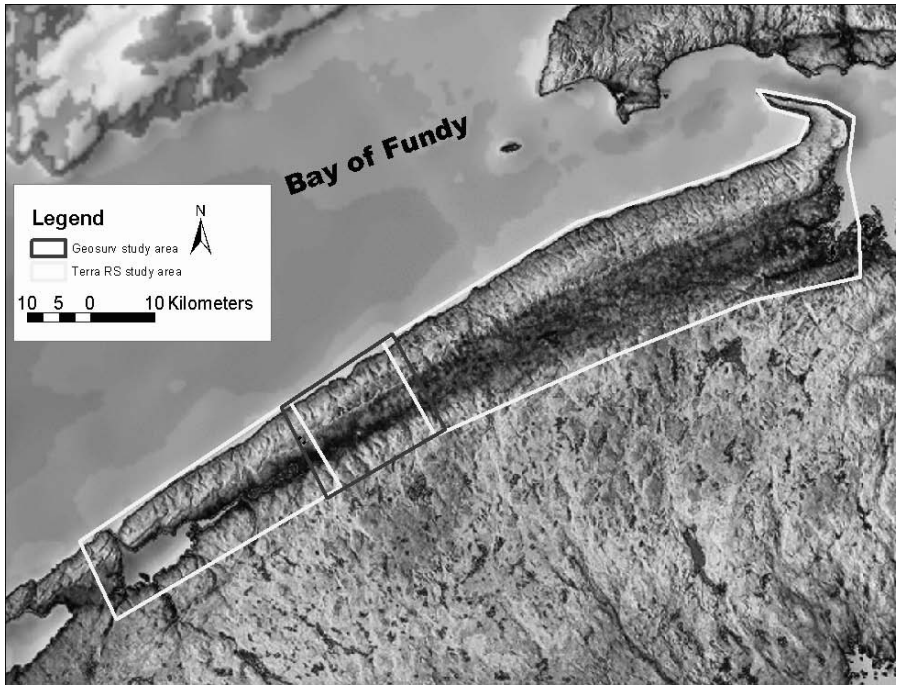


Figure 15.2 Location map of Annapolis and Minas Basin showing LIDAR, and airborne CASI coverage. Minas Basin is located in the upper right of the study area, and Annapolis Basin is located in the lower left of the study area. Radarsat data © 1996, Canadian Space Agency.

Radarsat-1 standard mode 2 scenes and a Landsat 7 scene were acquired near low tide conditions. The Ikonos satellite, which is owned and operated by Space Imaging, is capable of acquiring 1 m panchromatic and 4 m multispectral (3 visible and 1 near infrared band) imagery. Ikonos imagery were acquired near low tide by determining the date of low tide near 11:30 am local time (sun synchronous orbit

pass time) and requesting image acquisition from 2 days prior to the low tide's date to 2 days after the low tide's date. This allowed for variable weather conditions (clouds or overcast) and the fact that tide times advance approximately 45 minutes each day. Ikonos orders were sent through a local distributor. In addition to the satellite coverage, the coastal areas were imaged twice with the CASI sensor, once at high tide (3 m resolution) and once at low tide (1 m resolution) (Figure 15.3). The LIDAR survey area was divided into three regions, two flown by Vendor A, and one by Vendor B (Figure 15.2). The coverage for each company overlapped to allow a comparative analysis of the data from each company. All coastal areas were flown near low tide in order to acquire detailed inter-tidal topography. The in-land vegetation state in mid-July is at maximum leaf cover. Both LIDAR providers assured the AGRG that canopy penetration was still possible and that a significant number of laser hits would make it through the canopy to the ground. The data accuracy and data specifications for the LIDAR can be found in Appendix 15.1. The accuracy specifications are discussed in more detail in section 15.4. Details on each LIDAR system are presented in sections 15.3.1 and 15.3.2.

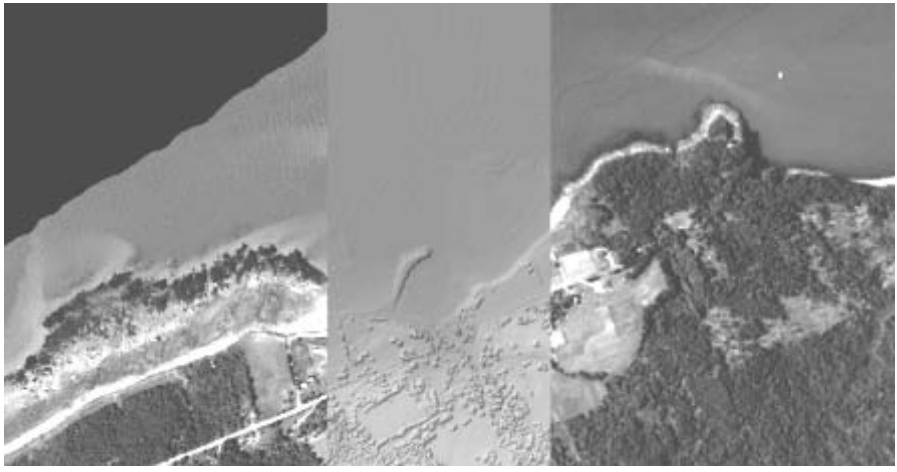


Figure 15.3 Mosaic of 1 m CASI at low tide (left image), 2 m LIDAR DSM at low tide (centre image), and 3 m CASI at high tide (right image) for Port Lorne along the Bay of Fundy. Overall image is approximately 4 km across.

15.2.2 Data Acquisition Issues

LIDAR is an active system providing its own pulse of near-infrared laser radiation and recording the reflected signal; thus it is not dependent on cloud free weather conditions as is the case for traditional aerial photography. However, rain or fog would cause the LIDAR survey to be delayed because the radiation cannot penetrate dense cloud or fog and therefore could not hit the ground. As a safety measure to ensure the laser is at a significant distance from the target, thus the power levels of the laser are not harmful to human eyes, the system will

automatically shut down if a laser return is detected to be at too close a range. For example, if the laser were to reflect off of an underlying cloud, high fog, or rain the system would automatically shut down. The LIDAR systems used in this project typically flew at a relatively low altitude, ranging from 300 m to 800 m, thus were often below any clouds. The ground activities for the LIDAR survey consisted of two tasks: 1. provide the aircraft with precise carrier phase GPS base station observations, and 2. collect precise elevation profiles to assist in data validation. Vendor B LIDAR area was estimated at 3 collection days and the Vendor A LIDAR area was estimated between 5 and 7 days.

15.3 LIDAR TECHNOLOGY

The terrestrial LIDAR system consists of an aircraft equipped with a GPS, attitude sensor and active near-infrared laser source and sensor. As the plane advances along the flight path the laser is fired, the pulse is directed toward the ground by an oscillating mirror and the reflected signal from the ground is recorded. By measuring the time it takes for the laser to reach the ground target and return, the range to the target can be accurately determined. The Time Interval Meter (TIM) records the time the pulse is transmitted and when the pulse is returned as well as the angle of the scanning mirror. This information, in combination with differential precise-code GPS and attitude measurements (e.g. pitch, yaw, and roll correction), is used to determine the height of the terrain relative to the ellipsoid. The LIDAR sensors available for this survey could record first or last returns only, while new generation sensors record multiple returns and the intensity of the signal. The Vendor B Optech ALTM 1020 system could record either the first or last laser return. The last return was specified since there is more chance of obtaining ground returns in this mode and a ground DEM was most desirable for our study. The resultant data from a LIDAR survey consists of a series of point location (laser hits) with associated heights above the ellipsoid. Since the ellipsoid based on the World Geodetic System of 1984 (WGS84) is a smooth mathematical surface which approximates the earth, a transformation is used to convert the heights relative to the geoid which is an equipotential surface based on the earth's potential gravity field. For this region, the transformation is based on a Geoid/Ellipsoid separation model known as HT1_01 that allows the data to be referenced to the Canadian Geodetic Vertical Datum of 1928 (CGVD28) defined by the Geodetic Survey of Canada of Natural Resources Canada. With heights now related to the geoid, termed orthometric height, they relate to approximate mean sea level. Since the geoid model is continually being refined, we requested the LIDAR data to contain both ellipsoid heights and orthometric heights (CGVD28). Therefore when new geoid models are released in the future, one can easily recompute the orthometric heights from the ellipsoid heights using the latest separation model.

15.3.1 Vendor B LIDAR Survey

Vendor B used two base stations for aircraft GPS control, the COGS permanent base station and a mobile station set up in the center of the study area to be surveyed. The distance of the survey aircraft from the base station should not exceed 40 km generally, because the GPS error increases with distance from the base station. The location of this central point was calculated based on observations between there and the COGS base station located at the eastern edge of the block (Figure 15.2). In addition to the LIDAR operator, Vendor B had two people on the ground collecting precise elevation values to be used to validate their LIDAR results. In addition to Vendor B, COGS also collected Real Time Kinematic (RTK) GPS elevation values within a limited radius of COGS in Lawrencetown due to the required line of sight radio link between the roving GPS unit and the COGS broadcasting base station. Vendor B collected both fast static point observations and kinematic locations from their vehicle, which required post processing.

Vendor B used a twin engine Piper Navajo airplane that they based out of the Digby airport. Trimble technology similar to that of COGS was used for their GPS collection, thus making data sharing very easy. The attitude information of the aircraft was collected using an Applanix POS Inertial Measurement Unit (IMU). The laser unit from Optech was housed in the camera mount within the aircraft. An Optech ALTM1020 sensor was used in the survey operating at a 5000 Hz laser repetition rate and a 15 Hz scan rate for the mirror. The area was flown at an altitude of 800 m above ground level (AGL) with a scan angle of 18 degrees, producing a swath width of approximately 520 m with raw laser point spacing every 3 m. At this altitude the laser beam had a footprint diameter of 25 cm. This sensor can capture first or last laser returns. For this survey it was set to collect the last return information, thus increasing the probability to hitting the ground in vegetated terrain. In addition to the LIDAR, a digital video camera also collected nadir looking video to be used later to assist in interpreting the laser returns and separating them into ground and non-ground hits.

Vendor B started their survey on July 6 and ended on July 13. They were delayed by some bad weather days, many due to rain and air turbulence. The study area consisted of approximately 64 flight lines, oriented parallel to the coast with two lines running transverse to the coast to be used to cross check the data. After processing the LIDAR data back in Ottawa they detected a problem with some elevation data that could not be resolved. As a result they re-flew several lines on September 1.

15.3.2 Vendor A LIDAR Survey

The Vendor A crew consisted of two people, one on the ground and the LIDAR operator. They used a Bell Ranger helicopter on the belly of which they mounted a pod containing the laser and video camera. Vendor A used Ashtech technology for their GPS collection thus requiring the COGS Trimble GPS data to be translated into an intermediate RINEX file format. The attitude information of the aircraft was collected using a Litton Inertial Reference System (IRS). An IRS is similar to an IMU in function, also known as an Inertial Navigation System (INS). Vendor

A's area of coverage was significantly larger than that of Vendor B and they used the Waterville airport as their base for the eastern block and the Digby airport for the western block. GPS control for Vendor A's local base stations, located at the two airports, was brought in from the COGS base. Vendor A's quality control strategy consisted of firing laser pulses at the base station antenna and comparing the LIDAR-determined height with that derived from GPS observations. They used the permanent COGS base station as a back up for the western extent of the eastern block. For the western block they established another mobile base station location over a provincial High Precision Network (HPN) point. This point is part of the geodetic GPS control network and is located near the Digby runway. Vendor A used a commercial first return laser unit operating at 1047 nm wavelength with a laser pulse repetition of 10,000 Hz and a 10-15 Hz scan rate for the mirror. The area was to be flown at an altitude of 600 m with a scan angle of 50 degrees, producing a swath of approximately 600 m with raw laser point spacing every 3 m. At this altitude the laser beam had a footprint diameter of 1.8 m.

Vendor A started their survey on July 11 and ended it on August 31 for the Nova Scotia study. During the data collection in the eastern block they experienced a power loss problem with the laser. The source of the problem could not be determined and it resulted in less penetration through the vegetation canopy. To remedy the situation they changed the flying altitude from 600 m to 300 m. This also affected the line spacing, since a lower altitude results in a smaller swath of coverage. With this system the relationship is roughly one to one i.e. at 600 m altitude a 600 m swath can be imaged. This caused much less area to be covered than the original estimated time, which in turn led to another problem of pilot fatigue causing more delays in addition to bad weather delays. The loss of power issue was eventually partially resolved by increasing the gain setting on the laser, thus allowing the aircraft to fly at the original altitude of 600 m.

15.4 GIS LIDAR PROCESSING AND VALIDATION

Raw LIDAR data may contain many erroneous hits. Collectively, these are known as 'noise.' Noise is usually filtered out by the vendor prior to delivery of the data using an algorithm. The data set will also contain hits from any number of surfaces (trees, buildings, roads, water, etc.). The vendor will also filter and classify the LIDAR point cloud into ground and non-ground hits (e.g. vegetation, buildings) prior to distribution. Thus two separate files may exist for a single 'tile' of LIDAR data. One file will contain only the ground hits, and one will contain points for all surfaces but the ground. It is important when working with LIDAR data to be aware of the composition of data sets being used. It is also prudent to use some sort of visualization software or technique to check for erroneous points that have escaped the noise filtering process or the ground/non-ground separation process. The LIDAR data was delivered on CD separated into 4 km by 4 km tiles. For both vendors tens of erroneous points per tile were identified by using a threshold of expected height values between –20 and 300 m. These erroneous points were probably a result of laser hits reflecting off of suspended aerosols. Each tile contained approximately 3 million LIDAR points and was in ASCII format. The survey specifications required an average point spacing, or "hit,"

every 3 metres. The ASCII files were imported into ARC/Info GIS and converted into point coverages. Triangular Irregular Network (TIN) files were built from point coverages of the ground hits. The large study area was processed using multiple tiles in overlapping sections to construct the TINs. The TIN files using ground only points were then gridded to produce a DEM at 2 m resolution using both a 5^{th} order polynomial fitting and a linear interpolation method. In the case of anthropogenic waterfront structures such as wharfs and breakwaters, the linear interpolation method is more suitable, while in rural areas, the quintic approach is more suitable to smooth the data. A Digital Surface Model (DSM) was generated from all of the LIDAR data, ground and non-ground points combined, using a linear interpolation method. A DSM differs from a DEM because it takes into account the height of the vegetation and buildings, while a DEM represents the "bald earth." The overlapping grids were then used to construct a seamless mosaic of the LIDAR surfaces, both a DEM and DSM.

The validation of the LIDAR involved both a comparison of benchmark GPS points to proximal LIDAR points and to the DEM. Two areas of the quality assurance testing involved the investigation of the spatial distribution of the LIDAR points and vegetation penetration, and the vertical accuracy of the LIDAR data utilizing GPS points. The horizontal accuracy of the data was assessed visually by comparing LIDAR map products to the 1:10,000 Nova Scotia Topographic Database.

15.4.1 Vendor B LIDAR Validation

Validation of the vertical accuracy of the Vendor B LIDAR was accomplished by measuring the differences between high precision GPS points and corresponding LIDAR points and by comparing GPS with the interpolated cells of the raster surface constructed from the LIDAR points. This was accomplished through differencing and linear regression of both the GPS elevation values and the LIDAR elevation data.

The first step in the validation process was to obtain quality benchmark data from GPS surveys. To compare the GPS points with the LIDAR surface, GPS accuracy needed to be equal to or better than that of the LIDAR points. This meant that only carrier-phase GPS could be used. This is the most accurate and expensive type of GPS. In the more common differential code GPS it is the code component of the GPS signal that is used to determine the position of the receiver. This code signal has a cycle width that is equivalent to a length of roughly three hundred meters. It is modulated onto a carrier wave that is approximately 20 cm in length. It is this carrier wave that is used to determine the position of the receiver during carrier phase GPS. The shorter carrier wavelengths provide a much finer degree of positioning than the 300 m cycle width of the code signal. This allows a positioning accuracy of within a decimeter using real-time and fast static techniques, or within less than a few centimeters with standard static techniques. Differential code GPS could only provide 1.5 – 0.5 m accuracies that would be insufficient for LIDAR validation. The GPS data used for this validation were collected using the real-time, static and fast static methodologies.

The validation points were compared to the LIDAR surface by a point-in-raster overlay. The validation points were intersected with the LIDAR interpolated

ground surface (DEM). The elevation values from the selected LIDAR surface cells were then joined to the validation point's attribute table. At this point the difference between the height values of the validation points and the corresponding LIDAR surface was calculated. In both cases the elevations used were orthometric heights derived from the CGVD28 geoid model. The results of a linear regression between the ground LIDAR surface and the validation points shows a very high correlation coefficient of 99.999%, with a standard error of 0.1869, indicating a very strong agreement between the elevation values from the two data sets.

Table 15.2 contains statistics on the elevation difference values between the LIDAR and the validation points. The vertical specifications in the contract were defined in the following terms: "[t]he vertical accuracy will be within an average of 15 cm of the measured GPS points, and 95% of the data points will not exceed 30 cm in vertical accuracy." As can be seen in the summary table of statistics, all specifications were met. One must be aware of issues that arise from the classification process that separate ground from non-ground points. For example, highway overpasses and bridges may be classified as non-ground points. Therefore any GPS points on and approaching such structures should not be used to compare with the DEM surface. The research group has acquired a Leica System 500 RTK GPS unit in 2003 and has since augmented the GPS checkpoints for this area. An analysis of these new data confirms that the vertical specifications were met.

Table 15.2 Summary of vertical validation of Vendor B LIDAR ground DEM.

Statistics	**Elevation Difference Values (m)**
Mean (specification 1)	0.13
Standard deviation	0.10
Minimum	0.00
Maximum	0.40
Median	0.10
Mode	0.10
Count	934
No. points <= 0.30 meters	915
Percentage points equal to or less than 0.30 meters (specification 2)	97.96%
Are specifications met?	Yes

15.4.2 Vendor A LIDAR Validation

A similar procedure was used to validate the Vendor A data, although it covered a much larger area. After an examination of the LIDAR point files, a distinct pattern was visible where there were no points, and thus no LIDAR returns, for the buildings or roads. This is because the roads and building rooftops are asphalt and low reflectivity to the near-IR laser. This lack of LIDAR return relates back to the

power loss experienced in the LIDAR unit during the survey. The lack of returns on the road made it difficult to check the precise relationship to the Nova Scotia Topographic Database but did allow us to determine that the road network generally corresponds to the lack of LIDAR returns along the road.

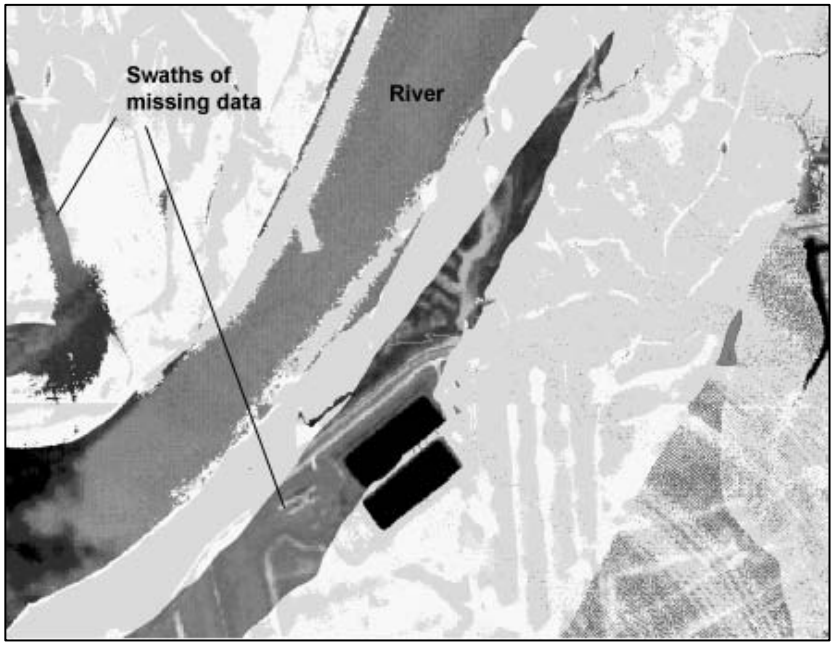


Figure 15.4 Vendor A LIDAR points (darker grey = ground, lighter grey = non-ground) over an Ikonos image for the Grand Pre dyke lands. Notice the missing swaths of data through the centre of the image and to the west. As a result of this missing data the dyke is not represented. Also the classification of ground and non-ground is not correct for the dyke area. Includes material copyright Space Imaging, LCC.

The other problem encountered with the point distribution along the coastal zone involved gaps in the LIDAR coverage. This was probably caused by the adjustment in flying altitude as a result of the power loss issue mentioned earlier. An example of this situation is shown in the area of Grand Pré near the Cornwallis River in the Minas Basin (Figure 15.4). The LIDAR hits for ground and non-ground are overlayed on the orthorectified Ikonos image acquired at low tide. Two areas of missing data swaths are evident, one in the central part of the image that includes the dyke land surrounding the river and the other in the western extent of the image. As a result of this area not being imaged during the LIDAR survey there is a lack of points; thus the ground and non-ground surfaces interpolated from the points do not accurately represent the presence of the dyke area, a critical barrier to coastal flooding. This results in a surface that is not accurate for flood risk modeling. In this case, the LIDAR data were supplemented by adding points manually in order to construct a valid surface model for this area. The Ikonos image was used as a guide for placing artificial points at the top and bottom of

both sides of the dyke feature in the areas of missing LIDAR data. Elevation values for the artificial points were based on values obtained from the LIDAR for the top and bottom of the dyke. Also, LIDAR points along the dyke misclassified as non-ground were selected as ground. This is an example of the type of visual quality inspection that should be done in such a project prior to using any derived data for decision-making.

Similarly to the Vendor B data the vertical accuracy of the LIDAR was examined in two ways: 1. comparing LIDAR point values close to the validation points, and 2. comparing the gridded LIDAR ground surface to validation points. There were not as many GPS validation points initially available for the Vendor A data set as for the Vendor B data set. However, three HPN points were compared to the LIDAR ground returns and statistics measured. LIDAR points within 10 m of the HPN points were selected for the comparative analysis. In general the differences between the average orthometric height of the nearest LIDAR hits and the HPN points were between 1 m and 3 m (Figure 15.5). The variance of the LIDAR hits associated with each HPN point were small and could not account for the differences between heights measured with LIDAR and the HPN network. It should be noted most HPN locations are near roads and the lack of LIDAR hits on and near roads does not facilitate a perfect correspondence of LIDAR hits to HPN points. Concurrent to this study, Vendor A also surveyed an area in Prince Edward Island for a Climate Change Action Fund project to derive flood risk maps. An offset of 0.9 m between GPS points and the LIDAR data was detected in that study (Webster et al., in press).

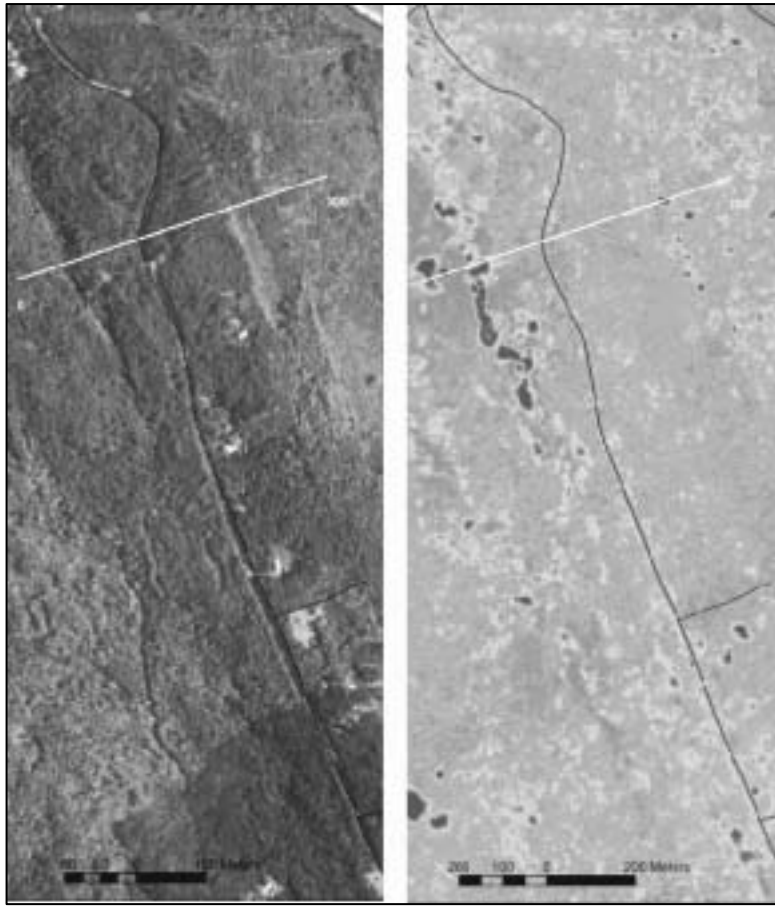


Figure 15.5 Comparison between Vendor B (2000) and Vendor A (2003). Left image is a digital true colour composite (converted to grey for this volume) CASI image; next is the difference map between the Vendor B and Vendor A 2003 DEMs. The main differences occur in the densely vegetated stream valley, and the small mounds (bumps) in the forest region.

As a result of this vertical bias in the LIDAR data, a follow-up GPS campaign was completed in the summer of 2001 by AGRG in the Annapolis Valley and shared with Vendor A in order to understand the nature of the problem. In this case GPS points were collected uniformly across the entire LIDAR swath in open flat areas. This was done to determine if the vertical bias was related to the scan angle of the system. It was finally deduced that the error resulted from the calibration of the system. An important aspect of a LIDAR calibration is the determination of any errors in the raw laser ranges. To properly calculate this value, multiple passes of varying altitudes must be made over known points. Since this aspect of the LIDAR calibration was not properly undertaken in the 2000 survey, an error of approximately 0.9 metres was introduced into the survey

(Webster et al., in press). Since calibration passes were made at an altitude of approximately 300 metres difference from the actual survey, the elevations matched at the calibration sites but did not match over the survey area. Comprehensive tests following the survey showed the need for a laser range bias and a laser range scale factor correction. Several lines of the LIDAR data in the survey area have since been re-processed by Vendor A to confirm these parameters.

15.4.3 Vendor A Re-fly of the LIDAR

As a result of the vertical offset and the limited vegetation penetration issues, Vendor A agreed to re-fly the area in the spring of 2003 at no additional costs. Since the 2000 survey there have been several improvements to their system and deployment strategy as a result of this experience (Shreenan, R. personal communications, 2003). A crucial step to any successful LIDAR survey is to conduct daily calibration passes over objects with known coordinates. Objects with long, straight elevated edges such as bridges and buildings are particularly favoured since they provide a check for both horizontal and vertical alignment. As mentioned earlier, calibration passes for this particular LIDAR survey entailed daily passes over the GPS base station. Although calibration parameters were adequately defined from this information the introduction of elevated man-made features allows a refinement of the calibration computation. This process has a direct impact on the overall accuracy of the survey and Vendor A has since adopted this practice.

The accuracy of a LIDAR system is only as good as the weakest component of the system. Since the 2000 survey, Vendor A has integrated a new Aplannix IMU into their system, which significantly improves the measurement of the roll, pitch and heading components of the aircraft over the previous sensor. The introduction of the new IMU improves the resolution of the aircraft attitude by a factor of five times for the roll and pitch and ten times for the heading (in comparing manufacturer specifications). Equally as important, the components of the IMU have been combined with the navigation sensors (GPS) for the purpose of regulating the IMU errors to those of the navigation sensors. This is commonly referred to as GPS Aided Inertial Navigation Systems and has greatly increased the overall accuracy of the system.

A final step to ensuring that an accurate product has been captured is to compare the processed LIDAR data to points that have been captured at a higher accuracy than the LIDAR. This practice has been adopted by Vendor A and uses the U.S. Federal Geographic Data Committee (FGDC) Geospatial Accuracy Standards (Parts 1-3) (www.fgdc.gov/standards/status/textstats.html) as a guideline for determining overall accuracy of all LIDAR surveys.

Another important step to any successful LIDAR survey is the accurate extraction of features from the LIDAR point cloud. The ability to perform this task accurately and reliably has a direct bearing on the overall accuracy of the survey. As a result of this survey, Vendor A embarked on developing and purchasing new software to reliably handle these tasks. The proprietary software used to process the collected dataset presented in this paper was extremely effective for single line corridor mapping but lacked the algorithms for processing multiple lines of

LIDAR. Also, the software lacked the visualization tools required to adequately detect outliers. The software now employed handles all of these issues as well as provides a host of other processing options including: 1. the removal of large outliers, thus improving the ground extraction algorithm's ability to accurately define ground hits; 2. the LIDAR data is processed according to neighbouring hits rather than the sequence of capture; 3. new hardware which enables tens of millions of points to be processed simultaneously rather than just millions of points; 4. several new visualization techniques have been implemented to help review the processed data and detect anomalies in the ground surface; and 5. video and digital mosaics are now used extensively to help review classified data.

Other improvements in the system include the introduction of a collimator attached to the end of the laser head. This device decreases the divergence of the laser beam to better than 0.3 mrad. This enables weaker reflective signals such as black top pavement and roofs to be captured because of the increased concentrated signal coming from the laser. Additionally, more penetration through the treetops is accomplished with the less divergent beam and therefore more ground hits captured. The foot print diameter in 2000 was 1.8 m, compared to 0.18 m in 2003. This was especially critical to reach the ground with a first return system.

In additional to improvements in the equipment, new field procedures have been implemented to include better field quality control to ensure no gaps exist between LIDAR flight lines. Gaps in the laser data can easily be identified on a daily basis, which can then be re-flown. Also, the laser data storage device has been reconFigured and repackaged which helps to eliminate data loss caused by the vibration of the airframe.

In the spring of 2003 prior to leaf-on conditions, Vendor A re-flew the eastern block at no additional costs and plans are in place to re-fly the western block (Figure 15.2) with a first and last multiple return LIDAR system in the fall of 2003. The preliminary data collected in the spring of 2003 has met the vertical specifications of 30 cm and is a significant improvement in terms of point density, returns from low reflectors (asphalt), and ground penetration in vegetated areas.

15.4.4 Comparison of LIDAR surfaces

When the original LIDAR areas were selected for each data provider, an area of overlap between the regions was maintained in order to compare the quality of the resultant DEMs between the two providers. For each data provider a ground DEM was constructed from the LIDAR points. This included Vendor B (2000), and Vendor A (2000 and most recently 2003 data). The ground surface was based on the construction of a TIN and quintic interpolation to a 2 m DEM grid.

The area of overlap is south of Lawrencetown and consists of a mixture of dense forest within a stream valley to cleared areas on the hilltop. The resulting DEMs are contrasted in Colour Plate 15.1 (following page 164). The problem with the original DEM derived from Vendor A is apparent by the triangular facets. This is a result of few LIDAR shots making it to the ground; therefore the DEM surface is not an accurate representation of the terrain. However, the DEM from the 2003 data that Vendor A re-flew is very similar to that of Vendor B. In fact, many of the small "bumps" on the landscape present in the Vendor B DEM have been removed in the Vendor A 2003 data. Although not verified in the field yet, these bumps are

interpreted to probably represent dense brush or other vegetation, based on their size, and do not truly represent the ground elevation. This highlights the improvements of the ground/non-ground classification algorithms from 2000 to 2003, thus enabling the production of a more accurate DEM.

15.5 COASTAL APPLICATIONS OF TERRESTRIAL LIDAR

Although much of this paper has been concerned with operational acquisition and validation of the LIDAR, end applications have been developed for the coastal zone concentrating on flood simulation modelling, inter-tidal slope calculations and merging the terrestrial LIDAR with bathymetry. For the flood-risk mapping, a similar methodology as in the studies by Webster et al. (2001, 2003, in press) have been employed.

Floodplain boundaries are important to emergency management personnel, insurance agents, and developers, as well as ordinary citizens. Flat floodplains are also attractions for further developments. Existing floodplain maps mainly derived from coarse elevation data do not depict actual floodplain boundaries. Major uses of LIDAR data include developing hydrological and floodplain models, telecommunications planning and analysis, transportation assessment, urban planning and natural resource and forest management.

For this study, the Minas Basin area was used because of the increased population and infrastructure that could be vulnerable to coastal flooding. The derived maps were passed to the municipal and town planning authorities to be used to assist in land development policies. In the absence of current land use vector data, Ikonos satellite imagery for the study area was classified to derive land cover types. The imagery was orthorectified using 1:10,000 road vectors and 5 m DEM from 1:10,000 contours. The land cover types derived through a maximum likelihood supervised classification were intersected with the flood levels derived from the LIDAR DEM to determine the percentage of affected areas.

The LIDAR ground points were used to construct the DEM surface. While delineating inundation areas on the derived DEM, some problems were discovered that related to the dykes not being properly represented. As mentioned earlier, some points along the dyke had been removed from the ground file. To depict the actual ground surface, it was necessary to construct a DEM with dykes. The dykes were extracted from the non-ground file by using the coded topographic base map vector database in combination with the Ikonos imagery. These points were in turn used to construct a new surface. As described earlier, where the LIDAR swath was absent points for the dyke and ground area were manually inserted. The heights were obtained by enquiring the LIDAR points adjacent to the missing areas.

Storm surges are the major cause of flooding in coastal areas. Storm surge is an abnormal rise in sea level accompanying a hurricane or other intense storms, and whose height is the difference between the observed level of the sea surface and the level that would have occurred in the absence of the storm. A storm surge at a low tide might go unnoticed but the storm surge at the highest tide might cause heavy coastal destruction. Therefore, the following calculations are performed for surges associated with high tides only.

Table 15.3 Height in meters above Chart Datum at Hantsport, Minas Basin. Source: Canadian Hydrographic Service Chart, HW – high water, LW – low water.

Large Tides		**Average Tides**		**Mean Sea Level (MSL)**
Higher HW	*Lower LW*	*Higher HW*	*Lower LW*	
14.8	-0.1	13	1.2	7.1

Tide levels and land elevation must be referenced to common datum in order to compute coastal flooding inundation areas. Because the LIDAR data and derived DEM are referenced to the geoid (CGVD28-MSL), and the tide levels are referenced to chart datum, which is typically the lowest elevation the largest tide will reach, the tide levels must be converted to reference a mean sea level datum (CGVD28). Thus the chart datum elevation of 7.1 m is subtracted from all the tidal water levels in order to relate them to MSL. Table 15.3 provides the original tidal information derived from the local chart, and table 15.4 shows the computed tidal levels in reference to mean sea level.

Table 15.4 Height in meters relative to Mean Sea Level (MSL) Datum, Hantsport, Minas Basin.

Large Tides		**Average Tides**		**Mean Sea Level**
Higher HW	*Lower LW*	*Higher HW*	*Lower LW*	
7.7	-7.2	5.9	-5.9	0

As can be observed in the above table, the highest tide in Minas Basin is 7.7 m above MSL. The average storm surge measured from all recording stations in the Bay of Fundy is 0.6 m for a one-in-twenty year event and 1.2 m for a one-in-one hundred year event. Surge levels were computed based upon these predictions. Table 15.5 provides the water levels predicted by combining tidal levels and storm surge levels referenced to MSL.

Table 15.5 Storm surge values. Probable return times of storms, one-in-20 year event, one-in-one hundred year event. Source : CHS Chart

Return Period	**Highest tide (m)**	**Storm Surge (m)**	**Resultant water level above MSL (m)**
Twenty-year flood	7.7	0.6	8.3
Hundred-year flood	7.7	1.2	8.9

Dykes have been constructed to protect low-lying coastal agricultural areas from tidal waters. The average dyke height ranges around 8.5 m. This means that the low-lying areas beyond the dykes would be flooded only if the tide level was

greater than 8.5 m. It could be deduced that low-lying coastal farmlands are well protected by dykes for floods with a storm return period of twenty years. The dykes are insufficient to stop flooding in a one-hundred-year storm event occurring at the largest tide of 8.9 m water level. These water levels do not take into account relative sea-level rise, which in this region is estimated at 25 cm per century (Stea, Forbes, and Mott, 1992). Therefore, for this project, a 9 m flood level was selected as the extreme limit to derive flood risk maps. The town of Kentville, located near the coast of the Minas Basin, has also used this level for their Municipal Planning Strategy maps.

Flood risk maps were generated using off-the-shelf GIS software tools rather than using flood models such as Mike11. The DEM used in the project has orthometric height values (MSL). The procedure of extracting inundation areas for various flood levels consists of selecting all the cells whose elevation (cell value) is less than or equal to that particular flood level. In order to determine the spatial extent of the flooding associated with a storm surge event, 10 flood levels at 0.5 meter increments were selected with the intention of building an animation. These were 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, and 9 m above MSL. An Arc Macro Language (AML) script was written for generating flood level grids where cells with values below the flood level are selected resulting in a binary map of flooded areas. In order to check for continuity between the flooded areas and the source of the water, in this case the bay or estuary, the grids were converted to polygons to permit selection of contiguous areas. For an area to be flooded, it must satisfy two conditions; the area must have elevation less than or equal to a surge level, and it must be connected to the water source, for example the Cornwallis River (Figure 15.6).

The process does not consider the effects of culverts and bridges due to lack of data of such infrastructure. However, it is known that many of the culverts and control gates of the smaller estuaries have one-way values to allow water to drain to the bay and control water flowing upstream. The flood levels of a storm return period of twenty (0.6 m, 8.3 m above MSL) and one hundred years (1.2 m, 8.9 m above MSL), with no consideration of sea-level rise, were overlaid on the Ikonos imagery to visualize the areas possibly affected. The calculated surge level is 8.3 m above mean sea level for a storm return period of twenty years and 8.9 m for a storm return period of hundred years. The flood levels generated in this study have been used for GIS overlay analysis to produce reports on the land cover and properties that could potentially be affected by a flood (see example, Colour Plate 15.2). The flood levels have been used to generate a series of visualizations including perspective views and flood animations. A report on the methodology, data processing, and results of the overlay analysis was given to the Kings County Planning Commission to be used for future recommendations.

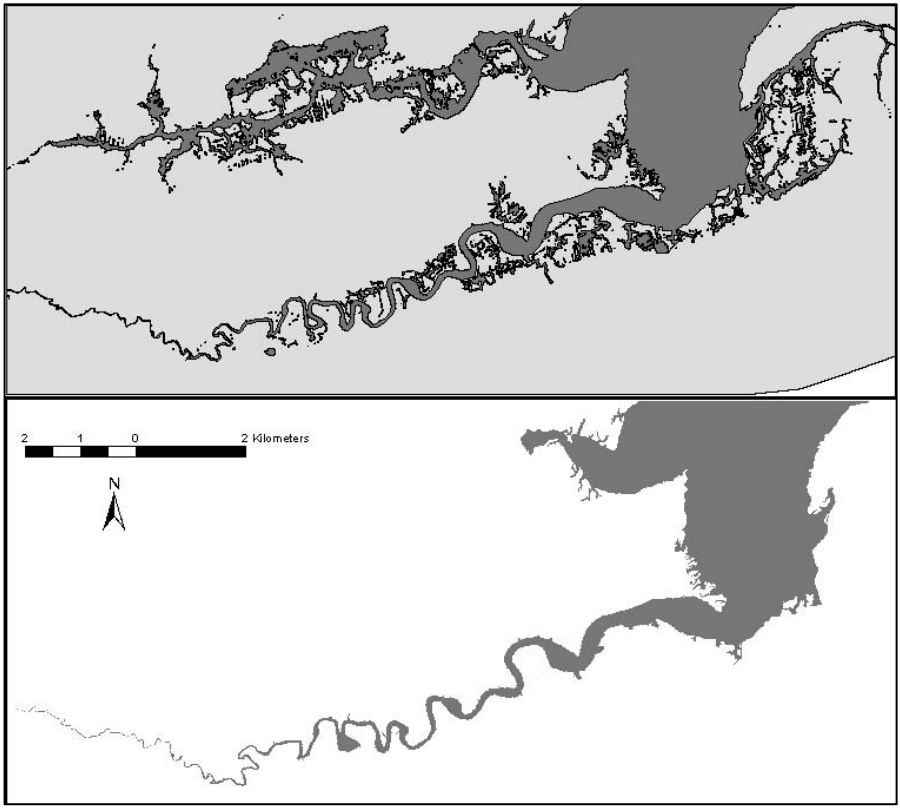


Figure 15.6 Example of how the flood-risk polygons are generated. Top image is the result of coding all cells less than or equal to 4 m. Bottom image shows only the areas that have connection with the bay or estuary. Flood level is 4 m above MSL; average high tide is 5.9 m above MSL.

15.6 CONCLUSIONS

The integration of high-resolution imagery and DEMs derived from LIDAR has several applications within the coastal zone. Here we have discussed the application of these technologies for developing flood-risk maps associated with possible storm surge events. LIDAR is a relatively new technology capable of providing dense elevation points of high precision and accuracy. It is important, however, for users of the data to be sure the data has met the accuracy specifications and independent validation data should be collected for this analysis. The validation of the LIDAR data indicated that the information from Vendor B met the specifications, while that from Vendor A did not for the 2000 survey. However, we worked with Vendor A to determine the cause of the problem and they have re-flown the area at no additional costs in 2003. The preliminary results of the new Vendor A data meet vertical specification of within 30 cm of measured GPS locations. In the case of both data providers, the separation algorithms removed features that the end user may want in the ground datasets, for example

highway overpasses and dykes. The end user must carefully inspect the LIDAR point data both for erroneous points, missing swaths of data, and for areas where the points have been misclassified into the ground or non-ground categories prior to using a gridded surface for decision-making. However, these problems often quickly reveal themselves after a surface has been constructed and a colour shaded relief has been made for visual inspection. The Bay of Fundy is unique because of the large tidal range, thus allowing the inter-tidal zone to be imaged with the LIDAR system. Wet smooth mud surfaces act as a spectacular reflector and often a low percentage of LIDAR hits return to the sensor, especially near the edges of the swath as one move away from nadir angles. Thus the overlap between flight lines should be increased to ensure laser returns in such environments.

The LIDAR data has been demonstrated to be applicable to determine areas at risk for coastal flooding associated with storm surge events. Once flood inundation boundaries have been calculated they were overlaid with the other infrastructure and property maps that exist for the municipality to calculate possible economic impacts of such events. High-resolution satellite imagery has been used to update the current land cover maps that were used in the overlay analysis with flood inundation areas. These types of data and information products are critical for land use planners and policy makers in order to manage the coastal zone effectively. This management could include restricting development in areas of high flood risk or mediation to try to minimize the impacts of such events. Other applications of these data for the coastal zone that have not been discussed in this paper but have been examined by the research group include generating slope maps of the inter-tidal zone, validating local tide models for the Minas Basin, and the generation of a seamless terrain model merging the land elevations and bathymetry for the area.

15.7 ACKNOWLEDGEMENTS

The processing of this data and validation examination has been a team effort at the AGRG. Team members include: Steve Dickie, Montfield Christian, Frances MacKinnon, Jamie Spinney, Charles Sangster, Michael Palmer and Trevor Milne. We would like to thank all of them and the other AGRG graduate students who have contributed for their dedication, hard work, and diligence. Also thanks to Dr. Robert Maher and David Colville, research scientists at the AGRG, for review of the manuscript and helpful discussions. We would like to thank Andy Sherin, Darius Bartlett and especially Jennifer Smith for the invitation to include this work in the book and assistance in preparing the manuscript. We are grateful to the staff and field crews from both Geosuv (now Mosaic Mapping) and Vendor A, especially Roger Shreenan of Vendor A, along with all of the other data providers including Hyperspectral Data International, GeoNet, and Radarsat International. We are grateful to the Nova Scotia Geomatics Center and the Municipality of Kings County for providing the base and property information for the study. Funding for the data collection was supported by a research grant from the Canadian Foundation for Innovation of Industry Canada.

APPENDIX 15.1

Terrestrial LIDAR Data Specifications for both vendors.

The LIDAR data will be supplied to the Applied Geomatics Research Group of COGS. The data will be in ASCII format, one row per elevation data point, organized in five columns as follows:

UTM Easting (WGS84)
UTM Northing (WGS84)
Height above Ellipsoid (WGS84)
Orthometric Height above Geoid (HT1_01)
GPS time

Three files will be delivered in this format on CDROM, containing

all data points,
ground points,
non-ground points.

The raw data files and airborne down-looking video will also be provided to the clients in care of COGS in formats they are capable of reading. Appropriate quality control information such as aircraft attitude and speed, quality of GPS positioning, and post-processing error data will be appended to each data point in one or all of the data files.

GPS control points will be occupied within a suitable range of the survey areas to achieve this accuracy and precision. Data will not be collected during increased ionisphereic activity which could degrade the GPS signal. Data will be collected under suitable conditions for video collection, under adequate light with no obscuring cloud beneath the aircraft. Data for coastal inter-tidal areas will be collected within 1 hour of lowest daily daylight tide levels in the survey area of Nova Scotia and within 2 hours of lowest daily tide in the Charlottetown, and North Rustico PEI survey areas.

All laser height, video, GPS, aircraft attitude and speed, and other quality control data will be delivered to the clients in care of COGS.

Full ownership of the data and information including working papers and reports will reside with COGS for data collected in Nova Scotia and the HM The Queen in Right of Canada (as represented by the Minister of Natural Resources) for data collected in PEI.

15.8 REFERENCES

Abdalati, W. and Krabill, W.B., 1999, Calculation of ice velocities in the Jakobshavn Isbrae area using airborne laser altimetry. *Remote Sensing of the Environment*, **67**, pp. 194-204.

Brock, J.C., Wright, C.W., Sallenger, A.H., Krabill, W.B., and Swift, R.N., 2002, Basis and methods of NASA airborne topographic mapper LIDAR surveys for coastal studies. *Journal of Coastal Research*, **18**, pp. 1-13.

Chagnon, R., 2002, Sea-ice climatology. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 5, 24 p. (on CD-ROM).

FGDC ([US] Federal Geographic Data Committee), 1998, Geospatial Accuracy Standards, Parts 1 to 3 (www.fgdc.gov/standards/status/textstatus.html, documents FGDC-STD-007.1, FGDC-STD-007.2, FGDC-STD-007.3).

Forbes, D.L. and Manson, G.K., 2002, Coastal geology and shore-zone processes. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 9, 84 p. (on CD-ROM).

Forbes, D.L., Shaw, R.W., and Manson, G.K., 2002, Adaptation. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 11, 18 p. (on CD-ROM).

Guenther, G.C., Brooks, M.W., and LaRocque, P.E., 2000, New capabilities of the SHOALS airborne LIDAR bathymeter. *Remote Sensing of the Environment*, **73**, pp. 247-255.

Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Kinden, P., Dai, X., Maskell, K., and Johnson, C. I., 2001, Eds. IPCC WG1. "Summary for Policy Makers", Climate Change 2001: The Scientific Basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change. (Cambridge and New York: Cambridge University Press)

Hwang, P.A., Krabill, W.B., Wright, W., Swift, R.N., and Walsh, E.J., 2000, Airborne scanning LIDAR measurement of ocean waves. *Remote Sensing of the Environment*, **73**, pp. 236-246.

King, G., O'Reilly, C., and Varma, H., 2002, High-precision three-dimensional mapping of tidal datums in the southwest Gulf of St. Lawrence. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 7, 16 p. (on CD-ROM).

Krabill, W.B. and Martin, C.F., 1987, Aircraft positioning using global positioning system carrier phase data. *Navigation*, **34**, pp. 1-21.

Krabill, W., Abdalati, W., Frederick, E., Manizade, S., Martin, C., Sonntag, J., Swift, R., Thomas, R., Wright, W., and Yungel, J., 2000, Greenland Ice Sheet: high-elevation balance and peripheral thinning. *Science*, **289**, pp. 428-430.

Krabill, W.B., Thomas, R.H., Martin, C.F., Swift, R.N., and Frederick, E.B., 1995, Accuracy of airborne laser altimetry over the Greenland ice sheet. *International Journal of Remote Sensing*, **16**, pp. 1211-1222.

Krabill, W.B., Wright, C., Swift, R., Frederick, E., Manizade, S., Yungel, J., Martin, C., Sonntag, J., Duffy, M., and Brock, J., 1999, Airborne laser mapping

of Assateague National Seashore Beach. *Photogrammetric Engineering & Remote Sensing*, **66**, pp. 65-71.

Maclean, G.A. and Krabill, W.B., 1986, Gross-merchantable timber volume estimation using an airborne LIDAR system. *Canadian Journal of Remote Sensing*, **12**, pp. 7-18.

Manson, G.K., Forbes, D.L. and Parkes, G.S., 2002, Wave climatology. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 4, 31 p. and 1 attachment (on CD-ROM).

Maune, David, F., 2001, Digital Elevation Model Techniques and Applications: The DEM User Manual. Edited by David F. Maune. *American Society of Photogrammetry and Remote Sensing.*

Mayor, S.D. and Eloranta, E.W., 2001, Two-dimensional vector wind fields from volume imaging LIDAR data. *Journal of Applied Meteorology*, **40**, pp.1331-1346.

McCulloch, M.M., Forbes, D.L., Shaw, R.W., and the CCAF A041 Scientific Team, 2002, *Coastal impacts of climate change and sea-level rise on Prince Edward Island.* Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, xxxiv + 62 p. and 11 supporting documents (on CD-ROM).

Milloy, M. and MacDonald, K., 2002, Evaluating the socio-economic impacts of climate change and sea-level rise. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 10, 90 p. (on CD-ROM).

O'Reilly, C., 2000, Defining the coastal zone from a hydrographic perspective. *Proceedings, Workshop on risk assessment and disaster mitigation: Enhanced use of risk management in integrated coastal management*. International Ocean Institute, Bermuda. *Backscatter*, April 2000, pp. 20-24

O'Reilly, C.T., Forbes, D.L., and Parkes, G.S., 2003, Mitigation of coastal hazards: adaptation to rising sea levels, storm surges, and shoreline erosion. *Proceedings, 1st Coastal, Estuary and Offshore Engineering Specialty Conference*, Canadian Society of Civil Engineering, Moncton, NB, CSN-410-(1-10), *in press*.

Parkes, G.S. and Ketch, L.A., 2002, Storm-surge climatology. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 2, 87 p. (on CD-ROM).

Parkes, G.S., Forbes, D.L., and Ketch, L.A., 2002, Sea-level rise. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 1, 33 p. and 5 attachments (on CD-ROM).

Post, M.J., Grund, C.J., Weickmann, A.M., Healy., K.R., and Willis, R.J., 1996, A comparison of the Mt. Pinatubo and El Chichon volcanic events: LIDAR observations at 10.6 and 0.69 mm. *Journal of Geophysical Research*, **101**(D2), pp. 3929-3940.

Sallenger, A.B., Jr., Krabill, W., Brock, J., Swift, R., Jansen, M., Manizade, S., Richmond, B., Hampton, M., and Eslinger, D., 1999, Airborne laser study

quantifies El Niño-induced coastal change. *Eos, Transactions, American Geophysical Union*, **80**, pp. 89-92.

Stea, R.R., Forbes, D.L., and Mott, R.J., 1992, Quaternary Geology and Coastal Evolution of Nova Scotia. Field Excursion A-6: Guidebook, Geological Association of Canada, Mineralogical Association of Canada.

Stockdon, H.F., Sallenger, A.H., List, J.H., and Holman, R.A., 2002, Estimation of shoreline position and change using airborne topographic LIDAR data. *Journal of Coastal Research*, **18** (3), pp. 502-513.

Shreenan, R., 2003, Personal communications. Research scientist with Terra Remote Sensing Inc.

Thompson, K., Ritchie, H., Bernier, N.B., Bobanovic, J., Desjardins, S., Pellerin, P., Blanchard, W., Smith, B., and Parkes, G., 2002, Modelling storm surges and flooding risk at Charlottetown. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada, Open File 4261, Supporting Document 6, 48 p. (on CD-ROM).

Wadhams, P., Tucker, W.B., III, Krabill, W.B., Swift, R.N., Comiso, J.C., and Davis, N.R., 1992, Relationship between sea ice freeboard and draft in the Arctic Basin and implications for ice thickness monitoring. *Journal of Geophysical Research*, **97** (C12), pp. 20325-20334.

Webster T.L., Dickie, S., O'Reilly, C., Forbes, D.L., Thompson, K., and Parkes, G., 2001, Integration of Diverse Datasets and Knowledge to Produce High Resolution Elevation Flood Risk Maps for Charlottetown, Prince Edward Island, Canada. In *CoastGIS2001 unpublished proceedings*, Halifax, Nova Scotia, Canada.

Webster, T.L., Forbes, D.L., Dickie, S., Colville, R., and Parkes, G.S., 2002, Airborne imaging, digital elevation models and flood maps. In *Coastal impacts of climate change and sea-level rise on Prince Edward Island*. Edited by Forbes, D.L. and Shaw, R.W. Geological Survey of Canada Open File 4261, Supporting Document 8, 36 p. (on CD-ROM).

Webster, T.L., Dickie, S., O'Reilly, C., Forbes, D., Parkes, G., Poole, D., and Quinn, R., 2003, Mapping Storm Surge Flood Risk using a LIDAR-Derived DEM. In *Elevation, a supplement to Geospatial Solutions and GPS World*. May.

Wehr, A. and Lohr, U., 1999, Airborne laser scanning – An introduction and overview. *Journal of Photogrammetry and Remote Sensing*, **54**, pp. 68-82.

CHAPTER SIXTEEN

Mapping and Analysing Historical Shoreline Changes Using GIS

Courtney A. Schupp, E. Robert Thieler and James F. O'Connell

16.1 INTRODUCTION

Understanding coastal change has become increasingly important to the 78 coastal communities in Massachusetts. About half of Massachusetts' 6.2 million residents now live in the coastal zone, and the coastal population is growing rapidly. Erosion threatens beachfront houses and development along much of the state's 2400 km of shoreline, creating a critical need for the government and the public to have accurate, up-to-date information on shoreline change. Prior to this shoreline change update, however, Massachusetts' existing historical shoreline change database had not been updated since 1978. New maps and data displaying long-term and more recent trends of shoreline behaviour increase the capability for sound decision-making and enhance public awareness of coastal change in Massachusetts.

To produce the necessary maps and data, we compiled historical shoreline positions in ArcInfo from a variety of map and aerial photograph sources. We generated 1:10,000 scale maps of the Massachusetts coast that display historical shorelines, rates of change, locations of rate-of-change measurements (transects), and orthophotographs. We also enhanced a Microsoft Access database created by Van Dusen (1996) that contains transects, dates, and rates of change; it can be used to perform spatial queries and to compile regional statistics and trends. The Massachusetts Office of Coastal Zone Management (MCZM) will distribute these maps, along with the data tables and a manual describing how to use the information, to regional government offices and coastal communities. These products will also be made available to the public through the Internet.

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

16.2 METHODS

16.2.1 Data Sources

Previous projects digitised and assembled historical shoreline data for the Massachusetts coast from 1842 to 1978 into ArcView GIS format (Benoit, 1989; O'Connell, 1997). Those projects obtained shorelines from six different data sources: 1) National Ocean Service (NOS) topographic maps (T-sheets), 2) NOS hydrographic maps (H-sheets), 3) Federal Emergency Management Agency (FEMA) Flood Insurance Study topographic maps, 4) printed orthophotographs, 5) aerial photographs, and 6) digital orthophotographs. The early shorelines (1842 to 1950) were digitised exclusively from NOS T- and H-sheets. Shorelines from the 1970s were compiled by digitising FEMA topographic maps, printed orthophotographs, and aerial photographs. These early data sets were digitised and placed into a GIS-compatible format using the Metric Mapping System (Clow and Leatherman, 1984; Benoit, 1989).

The aerial photography used to generate the 1994 orthophotographs has a nominal scale of 1:48,000 and was taken by the National Ocean Service in September and October of 1994. MCZM provided full-colour digital mosaicked orthophotographs that have a resolution of one meter per pixel. We used these geographically registered orthophotographs to digitise the 1994 shoreline directly within ArcView GIS software.

There are several possibilities for identifying a high-water shoreline. In most locations, we delineated the shoreline using the high-tide wrack line. However, due to the range in geomorphology along the Massachusetts coast, other delineations were sometimes more appropriate, such as the wet/dry interface, algal lines along rocky shores, vegetation changes in salt marshes, and the interface between seawalls and open water. Photographs of some areas, such as bleached-out sandy beaches, were difficult to interpret visually. In these cases, field checks and historical shorelines aided interpretation of the orthophotographs.

To verify the accuracy of orthophotograph rectification, we selected control points along the Massachusetts shore at easily recognizable sites, such as building corners and street intersections. We located these sites in the field and recorded the DGPS coordinates, which we then compared to orthophotograph coordinates. Results show that the orthophotographs comply with National Map Accuracy Standards (U.S. Bureau of the Budget, 1947).

Due to the variety of aerial photograph data sources and analytical techniques used to compile the shoreline positions described here, there are a number of potential sources of error that affect the accuracy of the shoreline positions shown on the shoreline change maps. Since much of the historical database had not been checked previously for geographic accuracy, the historical database was subjected to quality control review. The relatively high geographic accuracy and photographic detail of the 1994 orthophotographs, as evidenced by the accurate display of roads, buildings, and shoreline structures, allowed the identification of errors in the pre-1994 shorelines. Although the pre-existing data set was represented to comply with National Map Accuracy Standards, there are inevitably errors in any large spatial data set that more accurate data bring to light.

Our analysis resulted in the removal of approximately 43 kilometres of poor-quality data that could have contributed to erroneous interpretations of shoreline change.

Analysis of the various sources of error suggests that the individual shoreline positions in our completed data set are generally accurate to within +/- 8.5 meters. Rates of shoreline change derived from these shorelines have a resolution of +/- 0.12 meters/year (Thieler et al., 2001).

16.2.2 Generating the 1994 Shoreline

We used the line tool in ArcView 3.2 to delineate shoreline segments for each orthophotograph at a scale of >= 1:2,000. After saving shapefiles of the 1994 shoreline segments for each orthophotograph, we used the ArcView Geoprocessing Wizard extension to merge all segments into a single shapefile containing the 1994 shoreline. We added the attributes "length" and "year" to the attribute table before using the Geoprocessing Wizard again to merge the 1994 shoreline with the shapefile of historical shorelines.

Some areas had highly migratory shorelines, such as barrier islands where the landward migration over the past 150 years exceeded the width of the island. To ensure that seaward shorelines would only be compared to other seaward shorelines, and not to any landward shorelines, it was necessary to create two edited versions of the shoreline coverages. One coverage included only the seaward edge of migrating areas (for example, only the open-ocean side of a barrier island); the complementary coverage included only landward coasts (for example, only the bay side of a barrier island). This method enabled us to examine barrier island migration in addition to shoreline change.

16.2.3 Preparing the Workspaces

We used a suite of Arc Macro Language (AML) programs and a modified version of programs in the Digital Shoreline Analysis System (DSAS; Danforth and Thieler, 1992; Thieler and Danforth, 1994; Van Dusen, 1996) to pinpoint shoreline positions, cast orthogonal transects along the shoreline, compute a linear regression of the rate of shoreline change at each transect location, collect all of the rates and statistics into an INFO database, and create 1:10,000 scale maps. To achieve this it was necessary to establish a shore-parallel baseline and to create sequentially numbered workspaces containing data for each coastal area.

To establish a point of reference for shoreline change measurement, we used the *buffer* command to draw baseline segments 50 meters landward of and parallel to the general trend of the shorelines. We used a combination of the ArcEdit commands *unsplit, grain*, and *spline* to modify the baseline, placing vertices at 20 meter intervals. Van Dusen (1996) described a similar approach using the *spline, generalize*, and *densify* commands.

We saved each baseline segment, typically about 7 kilometres long, in its own workspace along with a coverage containing the corresponding shorelines. Using polygon coverages, we clipped the larger shoreline coverages, thereby creating smaller coverages to reduce AML processing time.

Once all of the completed shoreline and baseline coverages were saved in the appropriate workspaces, we ran an AML to format the data for the transect-casting and rate calculation software. In each workspace, the AML created a frequency file listing of all of the shoreline years in that area. The AML then used this file with the *reselect* command to select arcs by their year attribute before using the *build* and *ungenerate* commands on the coverage. It also used the frequency file to create a file containing the date for each shoreline. It created formatted text files for the baseline and all shorelines using the generate files.

16.2.4 Digital Shoreline Analysis System

As described by Van Dusen (1996), the basic software used to determine shoreline rates-of-change was a modified version of the DSAS, developed by the U.S. Geological Survey (Danforth and Thieler, 1992). This software is comprised of two C programs. Using the formatted text files, the first program applied a baseline-vertex approach to cast orthogonal transects from each vertex on the baseline to the most seaward shoreline (Figure 16.1 and colour insert following page 169). The program also created a file that contained both the coordinates of points at which the transects crossed each shoreline, and the dates associated with each of those shoreline points. A second C program used this file to compute the rates of change between shorelines at each transect location and to collect all of the rates and statistics into an INFO database. The resulting output file includes four statistical measurements for each transect: the end point rate, average of rates, linear regression rate, and jackknife rate.

These rates of shoreline change (Dolan et al., 1991) were calculated by measuring the differences between shoreline positions through time along a given transect. The end point rate was calculated by dividing the distance of shoreline movement by the time elapsed between the oldest and the most recent shoreline. The average of rates method (Foster and Savage, 1989) involved calculating separate end point rates for all combinations of shorelines and then taking the average of all end point rates. A linear regression rate-of-change statistic was determined by fitting a least squares regression line to all shoreline points for a particular transect; the rate is the slope of the line. The jackknife rate was implemented as an iterative linear regression which calculates a linear regression fit to shoreline data points with all possible combinations of shoreline points, leaving out one point in each iteration. The slopes of the linear regression lines were averaged to yield the jackknife rate.

16.2.5 Compiling the Data

A suite of four AMLs formatted the output files from the two C programs for the maps and database. The first AML created a coverage of the transect lines with associated shoreline change rates. To do this, it built the transect coverage as lines with the *build* command and added formatting symbols and values for the baseline and transect numbers to the attribute table for use in labelling transects on the maps. In order to determine angle and direction of the lines, it then built line and node topology and added coordinates of the transect nodes to the coverage's node

attribute table. Next, it imported the data from files created by the rate-calculation program into the ArcInfo tables and established a relate in the workspace for the next step. After joining the area files to the transect attribute table for that area, the AML brought the related information and arcs to a new appended coverage.

At this point, all of the necessary map coordinates and rate-of-change statistics were calculated. A second AML made a database for export to Microsoft Access. This AML employed the *append* command to create a new coverage containing the transects for the area workspaces. It then used the *tics* option with the *append* command to renumber transects in the appended coverage with unique, sequential ID values. After building the line coverage, the AML added a new item to the attribute table to allow reselection of every other transect. It then deleted alternate transects, creating a 40 meter interval for display on the printed maps, and renumbered the transects for labelling in the database and on maps. It also utilized the *additem* and *join* commands to add the town names from a separate coverage to the transect coverage. The final routine sorted the transect coverage by baseline and transect numbers.

We used ArcMap to remove manually the poorly located transects that crossed inlets and jetties. After the manual edits, a third AML added items to the attribute table of the transect coverage and selected the attribute table in ArcTables. It used the ArcEdit command *unsplit* to combine the identification arcs in the transect coverage and then built the transect line coverage. The AML then sorted the transects by transect number in ArcTables. After using the *copyinfo* command on the transect attribute table to create a data file, the AML used *dropitem* to delete attributes such as TNODE that are unnecessary for a user database. It then sorted this data file by transect number and used the *infodbase* command to create a database file. A final AML created the last files needed for the database by indexing the transects according to which map would display them. Using the *infofile* command, it created a DAT file for each map containing the transect number, town name, and town identification number. Finally, it used the *infodbase* command to copy this information into DBASE III+ files.

We imported the INFO database into Microsoft Access and converted it to a standalone application to improve users' ability to perform spatial queries and to compile regional statistics and trends. The database forms, reports, and macros were set up for an interactive display; only the tables needed to be replaced with the new database files and the transect coverage database. We used a macro to update the database with the new statistics and to calculate mean rates and bounding shoreline dates for each transect.

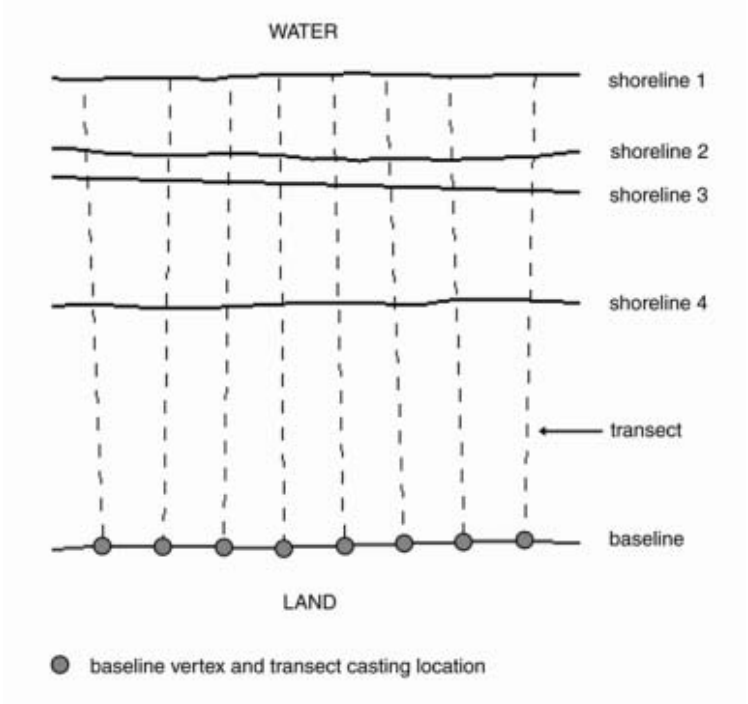


Figure 16.1 The transect casting scheme used to determine locations for the measurement of shoreline rates of change.

16.2.6 Map Production

The final project goal was to generate maps displaying shoreline change. We used an AML to format the maps with a legend and other explanatory information. This AML used the *reselect* command to draw shorelines with a different colour for each era. It also used the command *relate add* to correlate the files created by the rate-calculation program to the existing relate environment, which was saved as an INFO file. The *readselect* command selected every other transect in the writefile of selected transects, and the ArcPlot commands *arcendtext* and *leadertolerance* allowed the transects to be labelled with the associated identification number and linear regression rate of change.

16.3 RESULTS

The series of programs described above produced an Access database that can be searched by map number, community name, or transect number. Transect statistics can be listed in a table or graphed in histograms. The tables and histograms can either be printed or viewed on screen (Figure 16.2 and Figure 16.3).

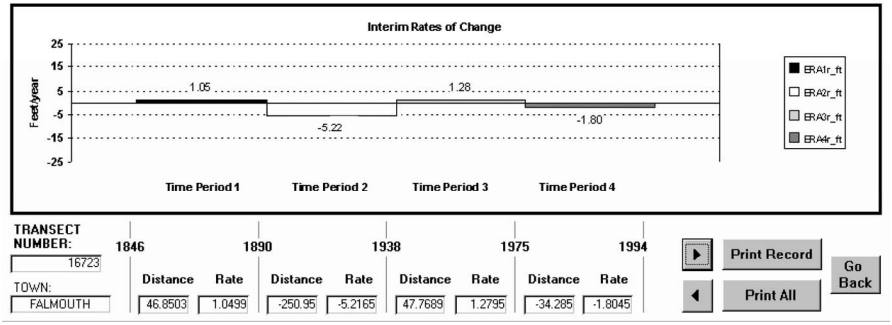


Figure 16.2 Sample histogram depicting interim rates of change.

MASSACHUSETTS HISTORIC SHORELINE CHANGE PROJECT — Transect Statistics for FALMOUTH

Transect Number:	First Time Period: 1846		Second Time Period: 1890		Third Time Period: 1938		Fourth Time Period: 1975		Net (overall) Time Period: 1994	Summary Statistics
	Distance*	Rate**	Distance*	Rate**	Distance*	Rate**	Distance*	Rate**	Net Distance Change*	Long Term Annual Shoreline Change Rate~*
16719	94.32	2.13	-443.11	-9.22	-94.68	-2.56	-82.84	-4.36	-526.31	-4.17
16720	-168.80	-3.84	-196.72	-4.10	-90.91	-2.46	-79.00	-4.17	-535.43	-3.58
16721	-191.40	-4.36	-192.29	-4.00	-46.85	-1.28	-113.98	-6.00	-544.52	-3.48
16722	-137.17	-3.12	-212.04	-4.43	-20.51	-.56	-104.43	-5.51	-474.15	-3.12
16723	46.85	1.05	-250.95	-5.22	47.77	1.28	-34.28	-1.80	-190.62	-1.57
16724	33.04	.75	-241.08	-5.02	10.01	.26	-63.68	-3.35	-261.71	-2.00
16725	29.53	.66	-232.15	-4.82	6.10	.16	-57.51	-3.02	-254.03	-1.94
16726	24.28	.56	-211.68	-4.40	-6.82	-.20	-59.84	-3.15	-254.07	-1.90
16727	-5.02	-.10	-177.59	-3.71	-24.54	-.66	-46.98	-2.46	-254.13	-1.87
16728	-17.68	-.39	-162.40	-3.38	40.88	1.12	-72.90	-3.84	-212.11	-1.44

Statistics for each interim time period show the distance difference and the derived rate of change. Summary statistics show the overall derived Long Term Annual Shoreline Change Rate.

* units in feet, ** units in feet/year, ~ determined using linear regression.

Go Back | Print

Figure 16.3 Sample transect listing produced by the Access database.

The programs also produced 91 ArcPlot files. The area maps of the Massachusetts coast are at a scale of 1:10,000 and display historical shorelines, transect locations, and long-term rates of change overlaid on the 1994 orthophotographs (shown in Colour Plate 16.1, following page 164).

The resulting database and maps include a total of 30,354 transects spaced at 40 meters along the coast of Massachusetts. Linear regression rates indicate erosion at 68% of the transects, accretion at 30% of the transects, and no net change at 2% of the transect locations.

Both the highest and lowest linear regression rates, 15.2 meters per year of accretion and 12.9 meters per year of erosion, occur on Nauset Beach in Chatham, where the barrier island shoreline faces the open Atlantic Ocean.

Several coastal areas have nearly the same number of accreting areas as eroding areas, such as Hingham (47% eroding, 48% accreting) and Hull (50% eroding, 46% accreting). Both of these areas are located within 40 kilometres of Boston, are highly populated, and have a number of seawalls holding the shoreline in place.

Eroding transect locations predominate in Plum Island (74% eroding, 24% accreting), Truro (83% eroding, 15% accreting), Wellfleet (81% eroding, 18% accreting), and Barnstable (73% eroding, 25% accreting). All of these areas have exposed sandy beaches.

The number of accreting transect locations exceeds the number of eroding transect locations in Harwich (36% eroding, 63% accreting), which is shielded by Monomoy Island to the southeast, and Manchester (40% eroding, 53% accreting), which is characterized by rocky cliffs interspersed with stretches of sandy beaches. The maps and database serve many purposes and interests. For example, state and local government agencies use the historical rate-of-change data to manage and regulate coastal development. Developers and potential homebuyers can examine the maps to determine the stability of beachfront property. Residents can view post-storm erosion in the context of long-term shoreline variability. This information can also aid in evaluating past management efforts, such as the emplacement of seawalls, jetties, and groins that have influenced shoreline change.

MCZM distributes these maps, along with the database and a manual describing how to use the information, to regional government offices and coastal communities. These products will also be made available to the public on the Internet.

16.4 SUMMARY

This project addresses the increasing need to measure and document coastal change by combining established shoreline evaluation techniques with improved GIS applications. The resulting maps and database comprise a current and accessible source of information on coastal change in Massachusetts, including historical shorelines and shoreline change statistics. Coastal managers in Massachusetts can continue to benefit from the database by updating the database with new shorelines and rerunning the suite of AMLs and C programs that calculate change and produce maps.

Academics, consultants, citizen groups, developers, and government agencies can use the map and data products described in this paper for a variety of purposes and interests. The database and maps have the potential to improve decisions concerning coastal management, residential and commercial development, and coastal research by making shoreline data comprehensible and available to a wide audience. The maps displaying shorelines and rates of change are visually appealing and quickly understood, particularly when overlaid on orthophotographs. More detailed information is readily available in an interactive database. If further modifications are desired, new modules can be incorporated into the current program and database structure. These methods and products are an excellent resource for anyone interested in coastal change, and they serve as a solid foundation for future applications.

16.5 ACKNOWLEDGEMENTS

The Massachusetts Shoreline Change Project was funded by the Massachusetts Office of Coastal Zone Management pursuant to a grant from the National Oceanic

and Atmospheric Administration, Award NA970Z0165. Susan Snow-Cotter, Steve Mague, Diane Carle, and Rebecca Haney of CZM offered many valuable suggestions for the development of the maps and users guide. Bill Danforth of the USGS provided computer support. Technical reviews by Laura Moore and Larry Poppe improved the clarity and content of the paper.

16.6 REFERENCES

Benoit, J. R., ed., 1989, *Massachusetts Shoreline Change Project*. (Boston: Massachusetts Office of Coastal Zone Management), pp. 19, appendices.

Clow, J.B., and Leatherman, S.P., 1984, Metric mapping: An automated technique of shoreline mapping. In *Proceedings, 44th American Congress on Surveying and Mapping*, American Society of Photogrammetry, pp. 309-318.

Danforth, W.W., and Thieler, E.R., 1992, Digital Shoreline Analysis System (DSAS) User's Guide, Version 1.0. *U.S. Geological Survey Open-File Report No. 92-355*. (Reston, Virginia: U.S. Geological Survey), pp. 42.

Dolan, R., Fenster, M.S., and Holme, S.J., 1991, Temporal analysis of shoreline recession and accretion. *Journal of Coastal Research*, **7** (3), pp. 723-744.

Foster, E.R., and Savage, R.J., 1989, Methods of historical shoreline analysis. In *Coastal Zone '89, Proceedings of the Sixth Symposium on Coastal and Ocean Management*. (New York: ASCE), pp. 4420-4433.

O'Connell, J.F., 1997, Historic shoreline change mapping and analysis along the Massachusetts shore. In *Coastal Zone '97, Proceedings of the Tenth Symposium on Coastal and Ocean Management*. (New York: ASCE).

Thieler, E.R. and Danforth, W.W., 1994, Historical shoreline mapping (II): application of the Digital Shoreline Mapping and Analysis Systems (DSMS/DSAS) to shoreline change mapping in Puerto Rico. *Journal of Coastal Research*, **10**, pp. 600-620.

Thieler, E.R., O'Connell, J.F., and Schupp, C.A., 2001, *The Massachusetts Shoreline Change Project: Technical Report 1800s to 1994*. U.S. Geological Survey Administrative Report. (Woods Hole, Massachusetts: U.S. Geological Survey), pp. 36.

U.S. Bureau of the Budget, 1947, National Map Accuracy Standards. (Washington: U.S. Government Printing Office).

Van Dusen, C., 1996, *Vector based shoreline analysis*. Unpublished report. (Boston: Applied Geographics, Inc.).

CHAPTER SEVENTEEN

GIS for Assessing Land-Based Activities that Pollute Coastal Environments

J.I. Euán-Avila, M.A. Liceaga-Correa, and H. Rodríguez-Sánchez

17.1 INTRODUCTION

According to Heathcote (1998), the development of workable management options (structural and non-structural actions) in a watershed requires the identification of all point and non-point sources of pollution. Discharge of effluents from industrial, urban and sewage treatment plants where a pipe or diffuser outfalls into a water body are called "point sources" of contamination. Another type of source, called a "non-point source" (NPS), is described as the diffuse drainage of rainwater from urban, industrial and agricultural lands that can introduce nutrients, pesticides, and metals to water bodies. Non-point sources are some of the more serious forms of pollution and the effects are often less obvious than those of point sources (Abel, 1998). Clapman *et al*. (1998) indicated that agricultural impacts on ground and surface water quality are more significant than other land use impacts because of their large aerial extent compared to other human land uses. Lack of information related to agricultural practices may lead to overuse of fertilizers and herbicides, and deforestation with serious impacts on soil erosion and water quality.

The Yucatan Peninsula is blessed with large freshwater reserves, bays and coastal lagoons and an exclusive economic zone of 200,000 km^2. However, the maintenance of the water quality seems to be an enormous challenge for the State and other interested groups considering the unfulfilled basic needs of a large part of the population, rates of population growth, immigration, and lack of an integrated approach to the management of the resources and coastal areas. In June 1996 a massive kill of 20,000 fish of the species *Arius felis* was reported in the Bay of Chetumal, State of Quintana Roo (SEMARNAP, 1996). Studies reported harm to their organs and accumulation of PCBs, organochlorine insecticides and polyaromatic hydrocarbon (Noreña-Barroso, 1998). Concerns exist that increasing loss of water quality may have adverse effects on mangroves and coral reefs in an area where the second most important barrier reef, the Mesoamerican reef, and the sanctuary of the manatee *Trichechus manatus* are drawing international attention.

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

Agricultural activities introduce diffuse pollution to watercourses, aquifers, lagoons and estuaries in the form of sediments, nutrients, pesticides, viruses, salt and other toxins which affect aquatic organisms. Pollution cause-and-effect relationships are complex and the need to find practical tools to generate useful information for decision-making may be addressed with models that introduce expert knowledge. Within this framework, this exercise attempts to rank the agricultural lands according to several factors that may contribute to water contamination on the Mexican side of the Rio Hondo watershed near the Othon P. Blanco municipality border with Belize.

17.2 GEOGRAPHICAL SETTING

17.2.1 The Yucatan Peninsula

The Yucatan Peninsula in México has mainly sub-surficial water dynamics driven by its geological and topographic nature: there are few rivers or lakes in the area. Exceptions are the Candelaria and Champotón rivers in the zone of the Términos Lagoon in Campeche, and the Hondo River in the State of Quintana Roo near the border with Belize. Yucatán has no rivers; however underground discharge in the coastal zone has been estimated at 9.7 million m^3 per year (CNA, 1998). The main features of the coastal areas are bays and lagoons, which are distributed throughout the three states: Términos Lagoon in Campeche; Celestún, Dzilam and Rio Lagartos in Yucatán; and Chetumal, Ascensión and Spiritu Santo bays in Quintana Roo. Surrounding these water bodies, other wetlands cover an area of approximately 8000 km^2, thus forming an important part of the coastal ecosystem (CNA-UNU/RIAMAS, 2000).

Due to its geographical location between the Caribbean Sea and the Gulf of Mexico, the region is influenced by severe hydrometeorological phenomena. The climate of the region is semi-arid in the coastal zone of the north part of the Peninsula and warm with variation of dry to humid in the rest of the peninsula. The mean annual temperature is 26ºC. There are two main seasons in the regional climate: the “rainy season,” including extreme phenomena such as hurricanes and tropical storms from May to October; and the "winds of the north” season, from November to April. The region receives abundant but uneven rainfall with mean annual precipitation ranging from 1600 mm in the southeast to 500 mm in the north. Mean annual evaporation is around 1.78 mm (CNA-NU/RIAMAS, 2000).

Human activities in the Yucatán Peninsula are related to agriculture, livestock production, tourism, fishing, oil production, and transportation, and recently to a large number of *maquiladoras*. Fertilizers, pesticides and metal residues have been found in coastal waters and the aquifer. These negative effects on the environment have been reported in several locations (Pacheco and Cabrera, 1996, Benitez and Bárcenas, 1996, Ortiz and Sáenz, 1997, Noreña-Barroso *et al*., 1998, CAN, 1998, Herrera-Silveira *et al*., 1998). Coastal resources in the Yucatan Peninsula provide increasing opportunities for economic development in a large number of traditional fishing communities. Lack of awareness of the deleterious effects of land-based activities on water quality may lead to a reduction in

biodiversity and esthetical quality of the landscape, thus putting at risk the continuity of fishing, tourism, and recreational activities.

17.2.2 The Municipality of Othon P. Blanco

Othon P. Blanco is a municipality located in the State of Quintana Roo in the southeast part of the Yucatan Peninsula (Figure 17.1). It has an area of 619,799 hectares, 161,226 (26%) of which are occupied by agricultural use, 129,396 (20.9%) by natural and cultivated grass, 325,155 (52.5%) by forest and 4,021 (0.6%) by other uses.

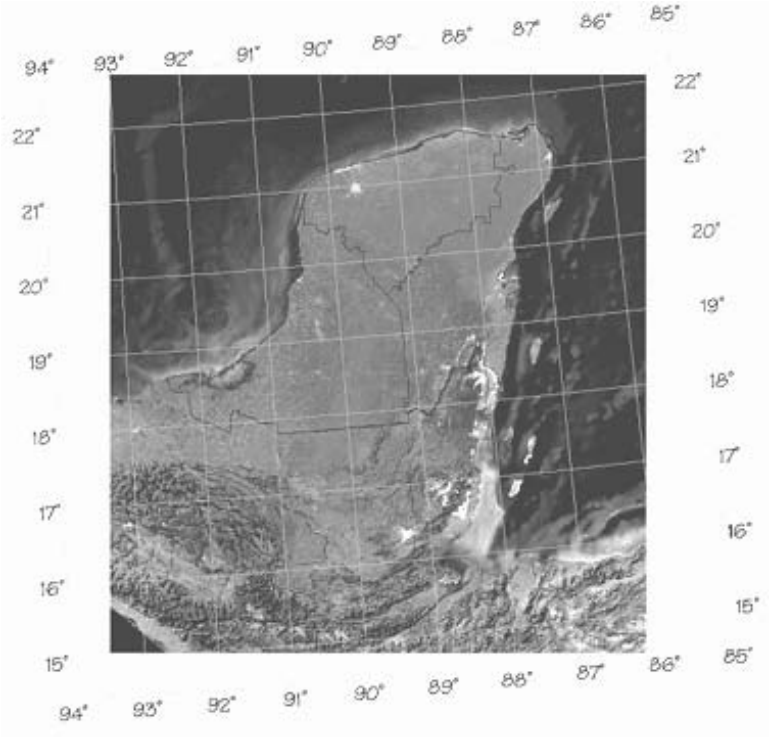


Figure 17.1 Study area in the Yucatan peninsula, Mexico. Municipality of Othon P. Blanco in the state of Quintana Roo. Main hydrological features are the Hondo River and the Bay of Chetumal.

The population in 1990 was 172,563 with 53.5% less than 19 years of age. The municipality has 437 rural and urban centers with 45% of the population living in 436 towns with less than 5,000 inhabitants. Close to 50% of the population are immigrants. Of the population 15 years or older, 17.4% completed elementary school (INEGI, 1991a). Main crops by cultivated area include: corn (19,196 hectares), sugar cane (16,000), chili (5,151), and beans (1, 391) (INEGI, 1993). Production in 1991 was estimated at 12,000 tonnes for corn, 785,000 for sugar

cane, 26,000 for chili, and 84 for beans. Other crops cultivated in the area are orange, coconut, and banana (INEGI, 1994).

Some of these agricultural lands can be found in the watershed of Chetumal Bay. The Bay is 67 km in length and 20 km wide, and receives freshwater from the Hondo River and Guerrero Lagoon, causing it to exhibit estuarine characteristics. Man-made channels and other tributaries close to agricultural lands are linked to the Hondo River. The most dominant soil types, Rendzinas and Litosoles, cover 70% of the Peninsula and 85% the study area (INEGI, 1985).

17.3 MODEL AND DATA LAYERS

17.3.1 Model

The selected model attempts to rank agricultural lands according to the potential menace they represent to water quality. The model is a multi-criteria evaluation method provided by the IDRISI software, which combines several layers as the criteria to form an index of evaluation (Eastman, 1999). Relevant factors in the process of NPS pollution assessment (being those for which an estimate was available) were: amount of agrochemical inputs, slope, proximity to surface water, and distance to aquifer. Given the natural continuity in factors, ratio layers and a weighted linear combination of them can provide an index for ranking. Their mathematical representation is as follows:

$$S = (\Sigma\, w_i * x_i) * \Pi\, c_j \qquad (17.1)$$

where s= NPS index, w_i= weighting factor i, x_i= factor i, c_j= constraint j

Factors must be normalized according to the accuracy of our knowledge of their respective ranges of impact, as well as the ways in which they behave at different scales. The procedure used, which is based on fuzzy sets, is also provided by the IDRISI program. Finally, the model allows a weight to be assigned for the relative contribution of each factor. Criteria called Analytical Hierarchy Process (AHP), based on a pair-wise comparison of factors along a continuous rating scale, can derive weighs by calculating the principal eigenvector of the created matrix (Eastman, 1999).

17.3.2 Data layers

17.3.2.1 Geographic location of agricultural lands

A good estimation of the location of agricultural land is a basic prerequisite for assessing NPS pollution. A two-band WIFS image with 180 x 180 m spatial resolution acquired in 1999 was used to estimate the location and extent of the

agricultural lands in the Othon P. Blanco municipality (Figure 17.2a). A supervised classification was conducted to estimate the total cultivated land in the area. Census data from the AGROS system elaborated by the INEGI provided crop information for areas called Basic Geo-statistical Area (AGEB in Spanish) (INEGI, 1996). Each AGEB is a well defined polygonal (vector) area with a link to a data base (Figure 17.2b). These two layers were used to estimate the location of cultivated areas by AGEB.

17.3.2.2 Agrochemical practices

The quantification of chemicals used per unit area and frequency of use was another factor in the analysis. A survey was conducted in September 1999 for the purpose of estimating the quantities of fertilizer and pesticides used per crop each year. A total of 97 farmers were interviewed in five agricultural towns: 1) Nicolás Bravo, 2) Palmar, 3) Pucté, 4) Sergio Butron, and 5) Morocoy. The questionnaire was divided into three sections to learn about fertilizer, insecticide, and herbicide practices per crop. Each section focuses on names, dosage, and frequency of applications of agrochemicals by crop per hectare per year. Substances were normalized according to thresholds determined through consultation with experts in the field.

17.3.2.3 Digital elevation model and slope

Slope is a well known factor that favours the movement of the substances on the terrain surface. Slope and aspect are the main factors that determine velocity and direction of the overland flow during storms. In areas with long slopes, the capacity of vegetation to reduce erosion is diminished. Slope was computed from a DEM of the area using the TNT software (Figure 17.2c). The DEM was produced by INEGI in a scale of 1: 250,000. Two archives, E1604 and E1607, were mosaiced to cover the studied area.

17.3.2.4 Proximity to surface water

Distance from agricultural activities to water bodies is also a factor that may facilitate contaminants reaching water courses, ponds, or estuaries. Proximity maps were constructed based on water features identified on band 2 of the WIFS image as well as those features found on topographic maps E16-4-7 at a scale of 1:250,000 produced by INEGI. These two layers were combined to generate a reference object for computing proximity to surface water (Figure 17.2d).

17.3.2.5 Proximity to groundwater

In karst formations (carbonated rocks) such as the Yucatan Peninsula in Mexico, precipitation tends to infiltrate rapidly because of the high number of fractures and solution cavities in the massif. Depth to the aquifer was estimated from a regression model using 15 well depths provided by the Regional Office of the *Comision Nacional del Agua* (CNA) and their corresponding elevation in the DEM (Figure 17.2e).

17.4 RESULTS

17.4.1 Agricultural lands

Agricultural areas detected from the satellite image were defined based upon the 1991 census. These areas were mainly devoted to growing sugar cane, corn, beans and *jalapeño* chili. The statistical data from AGEBS was proportionally assigned to current agricultural lands for a better spatial location of the cultivated areas (Figure 17.2b). As an example, Figure 17.2b shows, by AGEBS, the percentage of the total cultivated surface with sugar cane in 1991. The percentage for corn, beans and *jalapeño* chili was related to the area in a similar fashion.

17.4.2 Agricultural practices

Data from the five sampled regions indicated that farmers are working those lands for 11 years on the average (the interval is from 1 to 30 years with a standard deviation of 7 years, out of 93 valid cases). The mean value of the farmers' age was 42 (the interval was from 18 to 74, and the standard deviation was 12). The number of years each farmer had spent at school had a mean value of 4 (from 0 and 14, with a standard deviation of 3). These farmers reported applying agrochemicals to their primary crops (sugar cane, corn, *jalapeño* chili, and beans) during cultivation. Sugar cane is one of the main crops in the area and it received a large number of agrochemicals (more than 25 commercial products were identified). Given the large number of products, only those with a high frequency of use in each category (fertilizers, insecticides and herbicides) were selected for analysis. From these selected products, active substances were estimated giving the total applied quantity per hectare per year (an example is provided in table 17.1 for sugar cane). A layer with total phosphorous used in one year in kg/ha was calculated from four top selected crops with values of 70.25 kg/ha for sugar cane, 18.8 kg/ha corn, 40.19 kg/ha *jalapeño* chili, and 15.21 kg/ha beans each year. Similar computations were used to estimate nitrogen in the fertilizer category. In the case of herbicides substances such as 2,4D, paraquat, and ametryne were estimated. Finally, estimated substances in the insecticides category were metamidophos, chloropiriphos, and monocrotophos.

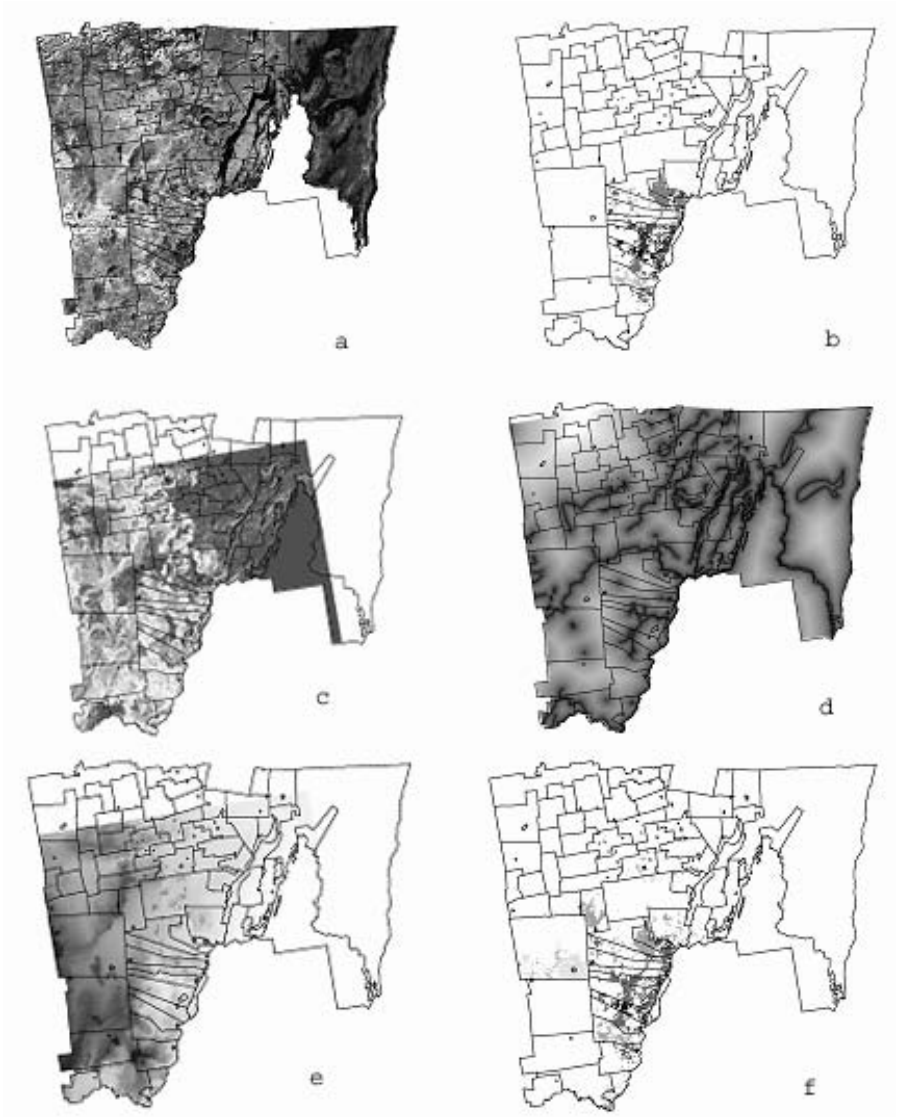


Figure 17.2 (a) WIFS image band 2, (b) cultivated areas for sugar cane and AGEBs, (c) slope from DEM, (d) proximity to surface water, (e) distance to aquifer, and (f) rank of nutrients (P and N).

17.4.3 Slope

Figure 17.2c shows the magnitude in percentage of slopes calculated from the DEM. Slopes in the area range from 0% to 48% with a mean value of 1.3% and a standard deviation of 2.5%. Elevation in the area ranges from 0 to 300 m. Slopes

on agricultural lands had a mean value of 1.84% and a standard deviation of 2.98%. These data were standardized from 2 to 10% increasing.

Table 17.1 Agrochemical substances used in sugar cane cultivation.

Substance	% of users	Mean annual number of applications and std. dev.	Mean quantity and std. dev. per application	Unit	Total amount of active substance
PK al 17%	44/60	1.40 - 0.92	317 - 089	Kg/ha	55.25
PK al 15%	14/60	1.30 - 0.61	332 - 130	Kg/ha	15.00
					70.25
N al 17 %	44/60	1.40 - 0.92	317 - 089	Kg/ha	55.25
N al 15 %	14/60	1.30 - 0.61	332 - 130	Kg/ha	15.00
N al 46 %	34/60	1.20 - 0.6	153 - 071	Kg/ha	47.84
					118.1
Ametryne 39.2%	14/60	1.10 - 0.28	2.8 - 0.86	l/ha	0.29
Ametryne 25%	24/60	1.50 - 0.77	3.9 - 1.2	l/ha	0.387 0.677
2,4-D 49.4%	31/60	1.20 - 0.66	2.3 - 1.1	l/ha	0.73
Monocrotophos 44.24%	17/60	1.90 - 1.00	1.9 - 2.6	l/ha	0.45

17.4.4 Proximity to surface and ground water

Proximity to surface water is shown in Figure 17.2d. Features included in the surface water category were rivers, ponds and channels. Results indicated a mean distance of 3.6 km, standard deviation of 3.2 km and mode of 0.9 km. The largest distance was 17 km. These data were standardized from .5 to 5 km decreasing. Finally, the result of the regression model for distance to the aquifer is shown in Figure 17.2e. Regressed elevation and depth measurements resulted in the following equation: aquifer depth = $0.0029h^2 + 0.2891h + 3.8161$ (h = elevation in the DEM), $r^2 = 0.97$ for n = 15. Depths ranged from 3.8 m to 211 m with a mean value of 42.5 m, standard deviation of 37.9 m and mode 3 m. Data was standardized from 1 to 15 decreasing.

17.4.5 Model results

Distribution of the S index for fertilizers, herbicides, and insecticides was calculated as a potential threat to water quality. An example, based upon nutrient load, is illustrated in Figure 17.2f. Parameters to compute the S index for standardized factors were selected from the observed data range of factors in the area. Equal weights were assigned to factors. With these parameters, the NPS index for fertilizers (phosphorous and nitrogen) had values from 1 to 204, a mean

value of 82 and a standard deviation of 44. Higher values of the index were found in the areas of Pucté and Palmar where sugar cane is the main crop. In the vicinity of Morocoy and Sergio Butron medium values were recorded, and the smallest values could be found in the area of Nicolas Bravo. The NPS index for herbicides (2-4D, paraquat, and ametrine) had values from 1 to 155, with a mean of 50 and a standard deviation of 32. Large tracts of Pucté and Palmar displayed higher values, while Sergio Butron displayed medium values and Morocoy and Nicolas Bravo the smallest. Finally, for insecticides (monocrotofos, chloropyrifos, and metamidofos) the NPS index ranged from 1 to 124, with a mean value of 40 and a standard deviation of 23. In this case, large areas of Pucté and Palmar displayed higher values, Sergio Butron and Morocoy displayed medium values, and Nicolas Bravo showed the smallest. The analysis suggest that Pucte and Palmar are major potential areas of threat to water quality as a result of nutrient, herbicide and insecticide inputs, while the importance of Morocoy, Sergio Butron and Nicolas Bravo in this respect changes according to the type of agrochemical under analysis.

17.5 CONCLUSIONS

An exercise in the use of GIS tools was conducted to characterize agricultural NPS pollution in the Yucatan Peninsula. Standard GIS processes and digital data from satellite, maps, and other digital products available in Mexico allowed ranking of agricultural lands in the Hondo river watershed according to the threat posed to water quality by the use of agrochemicals on these lands. Standard procedures available in the TNTmaps and IDRISI software packages were used to feed and run a multi-criteria model for decision-making. Three indices suggested that two out of the five agricultural study areas, where sugar cane is the main crop, have a large potential and non-homogeneous threat to water quality as indicated by the NPS index. The potential contribution of the other agricultural areas varies when examined by category (nutrients, herbicides, and insecticides). Increasing water contamination and limited resources in developing countries can be better allocated when knowledge of the co-occurrence of practices and terrain features is integrated into models for the assessment of the potential effects of human activities on water resources. As new knowledge is integrated these scenarios can be fine-tuned with more precise data and expert knowledge, providing improved spatial information to assist decision makers in the selection of appropriated management actions in the arena of NPS pollution.

17.6 ACKNOWLEDGMENTS

We thank the following people for their many contributions: J. Acosta, H. Hernández, G. Mexicano, Ricardo Rodríguez, Dr. Jorge Alvarado, P. I. Caballero, and students from the Centro de Estudios Tecnológicos del Mar (CETMAR). This research has been supported in part by Secretaría del Medio Ambiente Recursos Naturales y Pesca (SEMARNAP).

17.7 REFERENCES

Abel, P.D., 1998, *Water Pollution Biology*. (London: Taylor & Francis).

Benites, J.A. and Bárcenas, C., 1996, Sistemas fluvio-lagunares de la Laguana de Términos: hábitats críticos susceptibles a los efectos adversos de los plaguicidas. In *Golfo de México, contaminación e impacto ambiental: diagnóstico y tendencias*, edited by Botello, A.V., Rojas Galaviz, J.L., Benítes, J.A. and Zárate Lomeli, D. *EPOMEX* Serie Científica No. 5.

CNA, 1998, *Diagnóstico para la región XII, Península de Yucatán*, edited by Gerencia Regional de la Península de Yucatán de la Comisión Nacional del Agua.

CNA-UNU/RIAMAS, 2000, *El Proceso Final para el Cuidado y Manejo Responsable del Recurso Agua en la Península de Yucatán*. Reporte final, Comisión Nacional del Agua, pp. 2–6.

Eastman, J.R., 1999, *Guide to GIS and Image Processing, Vol. 2*. (Massachusetts: Clark Labs).

Heathcote, I.W., 1998, *Integrated Watershed Management: Principles and Practice*. (New York, Toronto: Wiley).

Herrera-Silveira, J.A., Ramírez, R.J., and Zaldivar, J.A., 1998, Overview and characterization of the hydrology and primary producer communities of selected coastal lagoons of Yucatán, México. *Aquatic Ecosystem Health and Management*, **1**(3–4), pp. 353–372.

INEGI, 1985, *Carta Edafológica 1:250 000*.

INEGI, 1991, Quintana Roo. Resultados Definitivos. Tabulados Básicos. *XI censo General de Población y Vivienda de 1990*, pp. 1–37.

INEGI, 1993, Othón Pompeyo Blanco, Estado de Quintana Roo. Cuaderno Municipal, Gobierno del Estado de Quintana Roo, edited by INEGI and H. Ayuntamiento Constitucional del Estado.

INEGI, 1994, Quintana Roo. Panorama Agropecuario. *VII censo agropecuario 1991*, pp. 23–33.

INEGI, 1996, AGROS, Información Censal Agropecuaria.

Noreña-Barroso, E., 1998, Contaminantes orgánicos y sus efectos a nivel histológico en bagres *Ariopsis assimilis* de la Bahía de Chetumal, Quintana Roo, México, Tesis de Maestría, CINVESTAV, Unidad Mérida, Yucatán, México.

Noreña-Barroso, E., Zapata-Perez, O., Ceja-Moreno, V., and Gold-Bouchot, G., 1998, Hydrocarbon and organochlorine residue concentrations in sediments from Bay of Chetumal, Mexico, *Bull. Environ. Contam. Toxicol.* **61**, pp. 80–87.

Ortiz, M.C. and Sáenz, J.R., 1997, Detergents and orthophosphates inputs from urban discharges to Chetumal Bay, Quintana Roo, México. *Bull. Environ. Contam. Toxicol.* **59** (03), pp. 486–491.

Pacheco, J. and Cabrera, A., 1996, Efecto del uso de fertilizantes en la calidad del agua subterránea en el Estado de Yucatán. *Ingeniería Hidráulica en México*. **11**(1), pp. 53–60, enero-abril.

SEMARNAP, 1996, Contaminación química en la Bahía de Chetumal, *Boletín Caribe*, edición de julio, Secretaría de Medio Ambiente y Recursos Naturales.

CHAPTER EIGHTEEN

Applying the Geospatial Technologies to Estuary Environments

David R. Green and Stephen D. King

18.1 INTRODUCTION

Around the World the value and vulnerability of estuaries has long been recognised. The National Estuary Study of 1969 in the U.S. is just one such example (http://www.inforain.org/mapsatwork/oregonestuary/Oregonestuary_page 4.htm). In the UK, initiatives by English Nature (EN) and Scottish Natural Heritage (SNH) to develop new approaches to estuary management in England and Wales, and Scotland respectively, led to Estuary Management Plans (English Nature, 1993). In Australia there have also been many similar initiatives to develop sustainable management of estuarine environments increasingly coming under pressure from tourism and industry (NLWRA, 2002) and, more recently, in the UK, from the impact of offshore windfarms e.g. Robin Rigg in the Solway Firth (Anonymous, 2003).

Green (1994; 1995) proposed the idea of GIS-based estuary information systems in the UK as a means by which it would be possible to collect, store, analyse and display spatial data and information to aid in estuary management. At the time, GIS was just beginning to develop into a practical tool for environmental applications in coastal zone management. In the intervening years, a great deal of interest has been shown in the development of GIS and the related geospatial technologies to aid in environmental monitoring, mapping, modelling, and management. Rapid developments in information technology (IT), including the Internet and the related technologies, have also led to the more widespread use of geospatial data and information for environmental applications by coastal managers and practitioners. Such developments have been responsible for providing the basis for access to data and information for management and public participation exercises. Increasingly, decision support systems (DSS) and information systems are also being used to support data and information requirements for coastal management.

With the continuing evolution of Information Technology (IT), the collection of, and access to, both data and information have rapidly been extended to the use of mobile technologies. These include Personal Digital Assistants (PDAs), Wireless Application Protocol (WAP)-enabled mobile phones, Global Positioning Systems (GPS), digital still and video cameras, and mobile GIS products such as ESRI's ArcPad, PocketGIS (http://www.posres.com), HandyGIS

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

(http://www.handgis.com) and FastMap (http://www.surveysupplies.com). Field data collection and access to information, via remote uploading and downloading, sending email attachments (e.g. photographs taken with a digital still camera or camera accessory for a mobile phone), and accessing the Internet using wireless technology are now all providing new opportunities to collect, work and interact with spatial data 'on-the-fly' (e.g. Vivoni and Camilli, 2003).

Alongside the hardware and software developments there have also been a number of new and important airborne and satellite-borne sensors. These have provided new sources of finer spectral and spatial resolution data to assist in the monitoring of coastal and estuarine environments e.g. CASI (Compact Airborne Spectrographic Imager), whilst LiDAR (Light Detection and Ranging) data provides unique high-resolution height data. The launch of ENVISAT in 2002 by the European Space Agency (ESA) for dedicated environmental monitoring, and European projects such as COASTWATCH (http://www.coastwatch.info) are also increasing awareness of the role of EO data for coastal monitoring in the context of GMES (Global Monitoring of Environment and Security).

This chapter presents an overview of some of the recent developments in the application of the geospatial technologies to estuary environments as the basis to support growing requirements for data and information in the context of environmental monitoring, mapping, and management. The role of some of the different geospatial technologies for environmental data collection, processing and access to information is examined. To conclude, some future developments are discussed. The chapter is illustrated using a number of examples drawn from Europe (including the UK), North America, and Australia.

18. 2 THE ESTUARY ENVIRONMENT

Estuaries are very well studied features of our coastal environment. A great deal of research has been undertaken into the physical processes operating in an estuary including sediment movement, water circulation, pollution and erosion (e.g. Stapleton and Pethick, 1996; Stove, 1978; Townend, 2002). The use of modelling techniques has been widely applied in an attempt to increase our knowledge and understanding about the processes that are active (e.g. Hinwood and McLean, 2002). Ecological studies of estuarine environments have included investigations into coastal habitat, macro-algal weedmats, and bird population distribution (e.g. Jernakoff *et al.*, 1996; Young *et al.*, 2000; Lewis and Kelly, 2001; Ripley *et al.*, 2002).

Estuaries are complex and highly productive ecosystems (among the most productive on Earth) supporting a wide range of habitats and species, extending from the river's upper tidal limit to the sea, and provide a constantly changing environment where sea and fresh water mix. They provide:

- Habitat (live and feed)
- Nursery Areas (reproduce and spawning)
- Productivity (abundance and biodiversity)
- Water Filtration (fresh and salt marshes)
- Flood Control (buffer to flood waters, dissipate storm surges)

- Erosion Prevention and Stabilisation (grasses and plants)
- Cultural Benefits (recreation, scientific knowledge, education, aesthetic value)
- Economic Benefits (natural resources used for recreation, industry, fishing, tourism)

(http://inlet.geol.sc.edu/nerrsintro/nerrsintro.html and http://www.epa.gov/owow/estuaries/about1.htm)

Many different habitat types, for example, are found in and around estuaries (http://www.epa.gov/owow/estuaries/about1.htm) including:

- Shallow open water
- Freshwater and salt marsh
- Sandy beaches
- Mud and sand flats
- Rocky shores
- Oyster reefs
- Mangrove forests
- River deltas
- Tidal pools
- Sea grass and kelp beds
- Wooded swamps

These are home to an abundance and diversity of wildlife including:

- Shore birds
- Fish
- Crabs and lobsters
- Marine mammals
- Clams
- Shellfish
- Marine worms
- Sea birds

(http://www.epa.gov/owow/estuaries/about1.htm)

Estuaries are also multi-value resources to society (NWLRA, 2002). They provide a whole suite of human resources, benefits and services (http://www.epa.gov/owow/estuaries/about1.htm). Such environments support commercial, traditional, and recreational fisheries, tourism, and wilderness experiences. They have also long been a source of inspiration for the human mind through aesthetics, art, and poetry. Since early times estuaries have been places for human settlement, providing sheltered harbours, transport and trade routes and natural resources for industry (http://www.iwight.com). Some are still relatively sparsely developed whilst others have become densely populated areas.

Not surprisingly the wide range of potential uses results in many conflicts, and therefore necessitates a management strategy. Each estuary has its own distinct characteristics, susceptibility to impacts, and management requirements. An

increasing number of national and international designations, policy and bylaws also make management of an estuary environment potentially very complicated. Worldwide there is now a wide range of organisations with statutory responsibilities on land, sea, and across the intertidal zone. In addition, there are many coastal and estuary management plans (http://www.iow.gov.uk) designed to provide a more holistic approach to sustainable management of coastal areas.

As local, national and international interest in estuaries has grown, so too have the sources of information available to managers and the public. Besides documents such as estuary management plans, there are now many local community groups who seek to involve and inform the public about estuaries. In the UK, for example, there are a number of organisations who look after the interests of the local estuary communities; for example, the Thames Estuary Partnership, the Solway Firth Partnership, the Moray Firth Partnership, and the Forth Estuary Forum to mention but a few. Besides providing documentation in the form of management plans, papers and publicity, these groups are also increasingly actively involved with local communities through the development of websites, online mapping and various local activities e.g. litter monitoring, habitat surveys, field visits and seminars.

In other parts of the world, for example the US and Australia, there are similar approaches to estuary management. For example, in Australia: the site http://www.dlwc.nsw.gov.au/care/water/estuaries/Inventory/Index_Geogr.html, and in the USA the sites http://www.spn.usace.army.mil/bmvc/baylink.html and http://www.estuaries.gov/about/programs.html provide lists of national programs and organisations involved in estuaries and resources. A number of recent projects have also focused on the status of Australian estuaries. For example, Australia's Near Pristine Estuaries: Assets Worth Protecting (NLWRA, 2002) has made a preliminary classification of all of Australia's estuaries based on their 'condition.'

18.3 ESTUARY MANAGEMENT

Estuaries are clearly very complex, interesting, and dynamic components of coastal zones around the World. They provide an interface between freshwater and saltwater environments, one that is constantly changing. They are also fragile environments and easily affected by human activities whether it be settlement, industry or tourism. Changes, however small, can be very harmful to the survival of an estuary (http://estuaries.gov/about/aboutestuaries.html).

With growing numbers of people settling and utilising estuaries, there are many activities that are now considered to endanger the survival of estuaries; such as dredging, the infill of flats and marshes, pollution, reconstruction of shorelines for housing, transportation, and agricultural needs. Also, marinas and tidal barrages can lead to dramatic changes in water circulation, erosion, and sediment movement. The effects of changes that have taken place have been unsafe drinking water, closure of beaches and shellfish beds, the development of harmful algal blooms, unproductive fisheries, loss of habitat and wildlife, and fish kills, as well as a host of other human health and natural resource problems (http://www.epa.gov/owow/ estuaries/about1.htm). Addressing such problems involves the use of planning, protection measures, and the implementation of

management strategies. The protection of estuaries from detrimental change is considered vital in order to preserve natural 'beauty and bounty,' as well as to sustain livelihoods that depend upon fishing and tourism. The designation of protective zones, use of patrols and enforcement can be used, whilst management, for example, can help to restore estuaries (http://www.estuaries.org).

But estuaries are often difficult areas to manage successfully because of all the conflicting interests that exist. Whilst local partnerships between communities, industry and government can be, and have been, very helpful in addressing some of the problems facing estuarine environments (for example, the Scottish Firths initiative e.g. the Forth Estuary Forum) the many competing interests are often very difficult to accommodate satisfactorily. Ideally for management to be successful requires the identification of a lead organisation at a higher level. National policy in Australia, for example, is still deemed essential despite the local and regional initiatives. But, the piecemeal approach to coastal management that seems to have developed around the World has also led to many problems. As noted by Brown (1995), "the current coastal management system is the cause of the failure to redress the problem of the coastline. This management system is characterised by fragmentation of responsibility between spheres of governments, among different professional areas of expertise, and across different social, economic and environmental interests. It is indeed difficult for the existing systems of coastal management to become part of the solution to coastal degradation while it remains part of the problem."

18.4 GEOGRAPHY AND THE GEOSPATIAL TECHNOLOGIES

Geospatial technology has the potential to assist in the future management of estuarine environments in a number of ways. Brown (personal communication, 2003) considers GIS, for example, to have multiple roles including record keeping, monitoring change, advising scientists, and most crucially (and currently the least developed) for demonstrating relationships between biophysical systems and socio-political options. The rationale for using geospatial technologies in a coastal environment is also clearly highlighted by Fabbri (1998) in the following quote:

"Given the complexities of coastal systems and the multidisciplinarity required for sustainable coastal development, computerized systems are necessary for the integration and distribution of vast amounts data and expert knowledge. They are also vital for performing analyses to aid decision makers in their difficult task of proving optimal and compromise coastal management solutions"

In this chapter, 'geospatial technologies' are considered to include remote sensing (aerial photography, airborne and satellite data and imagery), Geographical Information Systems (GIS), Global Positioning Systems (GPS), mobile computing e.g. portable computers and Personal Digital Assistants (PDAs), WAP-enabled mobile telephones, digital still and video cameras, portable data storage and compression, and the Internet. Together these technologies are all rapidly becoming increasingly useful tools for a wide range of environmental projects that have a requirement for data collection and access to digital spatial data and

information. These technologies, comprising both computer and other hardware and software, enable us to undertake:

- Data collection
- Data processing
- Data analysis
- Modelling and simulation
- Visualisation
- Communication and networking
- Provision of distributed data services
- Remote data access

Remote sensing (airborne and satellite sensors), for example, offers data collection and monitoring capabilities for estuarine environments at a number of different spatial, spectral, and temporal resolutions. This ranges from simple low-cost data acquisition platforms such as kites and balloons, model aircraft, and helicopters (Green *et al.*, 1998), and microlights for small-scale project work using 35mm cameras, digital cameras, and video, to the use of CASI and LIDAR, and satellite-based sensors on platforms such as IKONOS, ERS (European Remote Sensing Satellite), ENVISAT (Environment Satellite) and RADARSAT. Taken together these offer a powerful source of environmental data and information about estuary environments providing opportunities to derive up-to-date information on, for example, habitat type, algal weedmats and algal blooms, water depth, and sediment movement.

Rapidly decreasing costs of both hardware and software, together with improved usability and functionality, now offer end-users a wide range of tools that, when integrated, provide a practical framework for both acquiring and accessing data, storage, processing, analysis, modelling, and visual display. Recent developments in user-interfaces to both hardware and software have also made the geospatial technologies much easier to use than in the past. Likewise, developments in infrared (IR) and Bluetooth technologies have revolutionised communication links, facilitating the movement of large datasets from one location to another e.g. local to remote and vice versa. All of these developments now provide significant opportunities to interact with geospatial data in a laboratory and a field environment, the latter providing a means by which it is possible to access data and information 'on the fly.'

As these technologies have become cheaper and more usable, so they have become more accessible to a wider and more diverse end-user community including: academics who want to collect data in the field as part of a research project; commercial users working on contracts; ecologists, botanists, and biogeographers who need to gather data at a field study site; students collecting data as part of a dissertation; and schoolteachers and pupils wishing to undertake practical fieldwork as part of a curriculum requirement, class, or to aid in the development and creation of a virtual field course. All of these different categories of people are potential contributors to the well-being of an estuary, the very basis of the UK estuary fora.

Geospatial technology is rapidly becoming all-pervasive and there is growing interest in the potential that it can offer in terms of mobile access to data and

information. Already there is much 'geospatial' data and information available over the Internet, and together with developments in interface technology and the capability for greater end-user interactivity, the Internet is now providing access to powerful new tools that allow the delivery of online maps, animation, and virtual reality, not only on the desktop but now also in the field. For example, a combination of a PDA and a mobile phone will allow remote access to the Internet. Map data collected in the field using a GPS can be captured on a PDA and processed using software such as PocketGIS. Maps can be downloaded locally to a PDA, updated in the field and, with the aid of a mobile phone, uploaded into a remote database for storage, querying, retrieval, analysis, and display.

18.5 EXAMPLE GIS APPLICATIONS IN ESTUARY MANAGEMENT

The following are some examples of the different ways in which the geospatial technologies can and are currently being used in relation to an estuary environment:

18.5.1 Artificial Reef Siting

Artificial reefs have become increasingly popular as the means to attract fish populations for a variety of purposes, one of which is commercial fishing sport. These usually take the form of large metal structures, e.g. old boat hulls or drilling rigs, that are sunk thereby forming an 'artificial reef,' an ideal fish habitat. Where such structures are to be sited in shallow waters, close to the coast, and in areas in which there are many other potential users of the coastal area or zone, care must be taken in the siting process. The process of siting can be undertaken manually using relevant geographical information and siting criteria. An alternative approach is to use GIS to achieve the same task, with the added benefit of being able to undertake the process more quickly and efficiently, using more geospatial information, and the bonus of being able to generate a number of different siting scenarios. This type of siting exercise is a relatively simple example of using GIS in an analytical capacity. It involves the creation of an environmental database, comprising a number of raster and vector datasets, the selection of suitable environmental datasets according to certain criteria, the use of GIS functionality e.g. buffering, and the use of multi-criteria overlay analysis to create a composite in which only those areas that fulfil all the siting criteria are shown, those where artificial reefs might be located. An example of using this approach is a study in Moray Firth in the northeast of Scotland (Green and Ray, 2002; Wright *et al.*, 1998).

18.5.2 Habitat Management

The Nature-GIS (http://www.gisig.it/nature-gis/) project began in 2002 to examine the different ways in which geographic data and information are currently being used for protected areas management. Protected Areas Management includes Marine Protected Areas (http://www.ukmarinesac.org.uk) within estuaries.

Involving a total of nineteen European partners, the project also has as one of its aims the development of a protected areas geospatial data model (Green *et al.*, 2003). The data model aims to help standardise the way in which protected areas data is collected and stored for display and analysis with geographic data-handling tools, and subsequently used in protected areas management.

18.5.3 Erosion

Coastal erosion is of growing concern around the World. A European project, EUROSION (http://www.eurosion.org), has recently begun to provide documentation and information about coastal erosion along the coastlines of Europe. One particular area of concern is erosion in estuarine environments. Typically estuaries are sheltered areas of the coast protected from the force of the sea. However, the process of erosion is still active within such an environment. Consideration is being given to coastal erosion management, including some cases of estuaries in the Netherlands and Portugal. (Further details can be found at http://www.eurosion.org/project/reports.html.)

18.5.4 The Siting and Management of Yacht Marinas

In the UK, yacht and boat marinas have become increasingly popular coastal developments to cater for the growing leisure and tourism industry (Sidaway, 1991). Many of the estuaries around the UK coast are now home to both small- and large-scale marina developments. GIS can be used in a similar way to the artificial reef-siting example discussed above, to assist in the siting of a new marina. The GIS utilises a number of geospatial datasets and marina siting criteria to undertake an environmental impact assessment (EIA) and subsequently to isolate the ideal or optimum geographical location(s) for a marina. GIS can also be used as an effective data management tool for the marina and its assets. Already there are a number of commercial marina management tools available e.g. The Marina Program (http://www.ccmarina.com/index.htm). Although many largely are extended mapping toolboxes, the addition of a database in which to store the attributes describing the marina infrastructure and the location and characteristics of the boats in the marina provides an example of the functionality of the GIS toolbox. Mobile GIS hardware (e.g. Compaq iPAQ) and software (e.g. PocketGIS) and Global Positioning Systems (GPS) also provide considerable potential to enhance a marina database with information about the features that are not available from traditional data sources such as maps e.g. boat locations. Online GIS and mapping tools can provide the means to develop interactive retrieval and display systems for marina users, even extending to public information systems.

18.5.5 Weedmat Mapping

The problem of weedmats is common to many estuarine environments throughout the World. In estuary catchments that are predominantly agricultural, and where

there are areas of shelter, nutrient enrichment, and the growth of macro-algal weedmats may be common threats. There are also many other factors besides pollution that can affect the location and spatial distribution of weedmats, including sediment type and distribution as well as water flow. Macroalgae such as *Enteromorpha*, *Ulva* and *Chaetomorpha* typically form vast 'mats' that lie on the surface of estuarine mud- and sand-flats. These mats can either be very fine coverings (thin) or several centimetres deep (thick).

Remote sensing, most often aerial photography (panchromatic, colour, and colour infrared (CIR)) because of its spatial and temporal resolution, and more recently hyperspectral data (because of its finer spectral resolution), has been used to monitor and map weedmats with varying degrees of success. An example of this is the recent work undertaken on the Ythan Estuary in the north east of Scotland to establish a practical methodology for deriving weedmat area over time (Green and King, 2002; Orr, 2003).GIS can be used to undertake simple temporal differencing using overlay techniques to estimate change in weedmat area and distribution over time, as well as to establish area estimates for weedmat coverage using additional functionality contained in ArcView extensions such as Xtools (http://www.odf.state.or.us/DIVISIONS/management/state-forest/XTools.asp). In the longer term, the data may form the basis of an Ythan Estuary GIS (desktop and online using ArcIMS), which seeks to establish an environmental database comprising remotely sensed imagery, Ordnance Survey (OS) digital map data, digital hydrographic bathymetry from the United Kingdom Hydrographic Office (UKHO) and other digital datasets e.g. bottom sediment distribution, bird count data and distribution, and water quality data.

18.5.6 Search and Rescue (SAR)

In coastal search and rescue operations access to geographical data and information is vital. As well as occurring along the coast and out to sea SAR operations also occur in estuarine environments.

Information may be required during both the search and the rescue stages. Whilst some relevant data and information may be available e.g. in the form of maps and charts, much of the baseline data and information is currently not available at a scale that is useful to a SAR operation. Increasingly it has been recognised that such information should ideally (a) be available in a digital format, (b) take advantage of the digital communications networks, (c) embrace the Internet and online Geographical Information Systems (GIS) technology and (d) comprise sources of information at many different scales, the most important of which is local knowledge (Figure 18.1).

The acquisition of local knowledge and information can be greatly enhanced through the use of the mobile geospatial technologies that are now available, integration of which can be developed into a decision support system framework for coastal and marine search and rescue operations. For example, a wide range of mobile technologies can be brought together to provide a basis for gathering, processing, displaying, and communicating geospatial data and information in real-time. Hardware in the overall IT equation might include a PDA with PocketGIS for detailed mapping of an access footpath to a beach; a mobile phone to upload

the datafile to an online GIS e.g. ArcIMS; an on-shipboard PC with a wireless access to the Internet, accessing the ArcIMS-based online decision support system; a digital camera that sends by email on a mobile phone a picture of the coast from the seaward side to the local Coastguard either onboard a ship or to a rescue team on the cliff top; sonar information that can be uploaded to an ArcIMS system and combined with e.g. an Ordnance Survey digital coastline; a video clip or panoramic photograph of a section of the cliffline that can be zipped and sent as an attachment by email. Additional components of the system may also include digital or paper fax machines, which could, for example, be used to communicate an annotated sketch map or photograph to someone on land, on a lifeboat or inflatable, or at an RNLI (Royal National Lifeboat Institution UK) or Coastguard station.

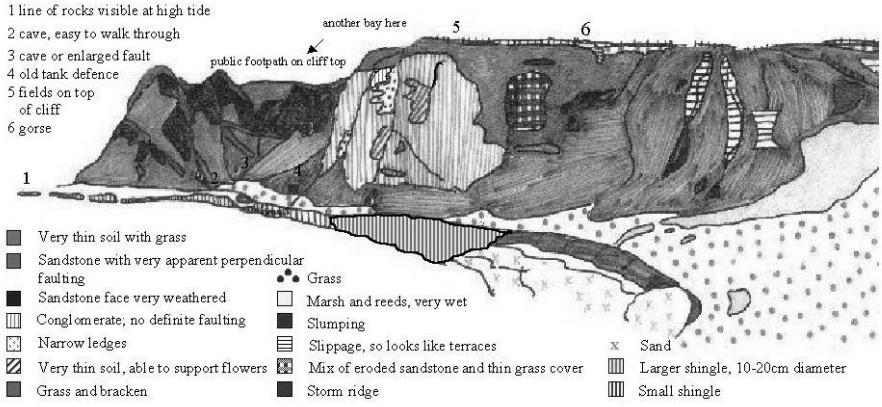


Figure 18.1 A field sketch of a section of coastline (courtesy of Joanna McDonald, MRI)

18.5.7 Participatory Planning

One outcome of the work reported earlier in this chapter about weedmat monitoring and mapping was to input the resulting information into a GIS, as a means to deliver information to a wider audience in the context of environmental awareness and participatory planning. Whilst geographic datasets can be viewed in a desktop GIS environment, not everyone has access to GIS software. An alternative is to place the datasets on CD to disseminate the information to the end-user. A more effective approach is to distribute the datasets with a GIS viewer such as ESRI UK's MapExplorer (http://www.esriuk.com). MapExplorer is freeware that allows the user to access GIS datasets for viewing and output, as well as providing limited GIS functionality such as zoom, pan, overlay, and measurement tools. MapExplorer is a 'cut down' GIS package with navigational functionality allowing an end-user to open and explore the information. GIS tools have also been developed for the Internet. The Internet is an ideal medium in which to deliver maps and images because it is a familiar interface to many people. ERSI's ArcExplorer (http://www.esri.com/arcexplorer), for example, can be used

to connect to an Internet webserver. Online mapping and GIS software provides an ideal mechanism to involve the public in the environmental planning process through simple interactive interfaces to visualise geospatial data and information in the form of maps and imagery.

18.5.8 Modelling

Developing an improved knowledge and understanding of the complexity of an estuary can often be aided through the use of modelling techniques and tools. Whilst often abstract, such models nevertheless help to provide insight into the physical and biological processes active in such an environment. One such model is the Forth Estuary Modelling System (FEMS). FEMS was developed as part of a Forth Estuary Forum (FEF) Flagship project. The aim was to take an ecological model and present this to the wider community in a form that would help to educate people of all ages about ecological processes active in the Forth Estuary. It comprised a Pelagic Submodel; Benthic Submodel; Fish Submodel; Sea Birds/Mammals Submodel; and a Human Influences Submodel. The model was written using an interactive modelling toolbox, STELLA. Initially distributed on a CD-ROM complete with HTML documents, FEMS was designed to be accessed as an educational tool from the Forth Estuary Forum website, facilitating updates as the model evolved (Green and King, 1999).

18.5.9 Education and Training

One application of GIS technology, either as a desktop or web-based system, is to provide the basis for helping to aid in education and training programmes about estuaries and estuary management. Access to data and information stored in a GIS using standard information retrieval and query tools is a simple way for people to retrieve, display, and communicate geographical information, most often in the form of an interactive map. Simple querying of a spatial database and visualisation in the form of a chart or a map can often help to raise awareness and educate people. Additionally, provision of simple end-user interfaces to the information e.g. using ArcView or ArcIMS on the Internet increases the power of the 'map' to communicate spatial information. Toggling map layers on or off can help to communicate spatial patterns and reveal spatial relationships. GIS mapping can also provide a simple data exploration tool.

GIS can also be used as an effective training environment to learn about estuarine environments. The GIS toolbox provides the means to integrate geographical datasets, to examine patterns, explore data, establish relationships, as well as to visualise data and information in a variety of different ways e.g. 2D, and 3D. Much of the output resulting from simple database queries, statistical summaries, analysis, and visualisation exercises provides a very effective form of communication that can be used as input to documents, reports, slide shows and presentations.

Many of the estuary websites now becoming available are primarily designed to provide access to both data and information. They usually comprise webpages,

downloadable documents, datasets, software and sometimes even models. These help to maximise access to information for a wider community. In some cases, these websites are taking advantage of many of the software tools that accompany the Internet including, for example, online mapping and GIS. Although most online mapping systems currently have relatively limited GIS functionality, they nevertheless provide significant potential to allow end-users to view information in the form of text and images. Some also include access to remotely sensed imagery (airborne and satellite) that can be viewed as 'quicklooks' or in some cases takes advantage of image compression software such as Lizardtech's MrSID, or ERMapper's ECW format, to deliver high-resolution images rapidly across the Internet.

18.6 INTERNET-BASED INFORMATION SYSTEMS

18.6.1 GDSPDS - Solway Firth

In 1999, a research project began at the University of Aberdeen to develop an Internet-based coastal information system. The aim of the project is to prove the concept of the need for and the capability to deliver geospatial information (digital images, documents, maps and remotely sensed imagery) to the coastal manager via the Internet, using the Solway Firth as a case study area. The Geo-information Decision Support Processing and Dissemination System (GDSPDS) is capable of serving GIS data, digital photography and video as well as remotely sensed imagery via the Internet.

The Solway Firth GDSPDS is built around Windows NT Server 4.0 using Internet Information Server 4.0 as the web server. In addition, ESRI ArcIMS 4.0 is used to deliver online mapping capabilities, and ER Mapper's Image Web Server (IWS) version 1.7 and LizardTech's Content Server are used to deliver remotely sensed imagery. The imagery is compressed using either ER Mapper's Enhanced Compressed Wavelet (ECW) technology or LizardTech's MrSID Geospatial Encoder (v1.4). Both of these compression technologies allow remotely sensed imagery to be both stored more easily (for example, a colour air photo mosaic of 178.7Mb can be compressed to 3.2Mb in ECW (30:1) and 3.11Mb in MrSID format (30:1)), and also served rapidly across the Internet to any PC with a web browser. In the case of ECW files, a relatively small plug-in must be installed to view the imagery. MrSID imagery can be delivered either via a plugin viewer or as a series of JPEG images. The imagery is stored in an MS Access database which is searchable using a web-based form. A number of basic search queries have been developed to try and help guide the user to the correct imagery for their specific management problem or application.

As well as being accessible on a desktop computer, the GDSPDS can also be accessed via a PDA connected to the Internet using a mobile phone. Remotely sensed images can be viewed in the web browser on the PDA, while ESRI's ArcPad software can be used to access and retrieve maps from the ArcIMS map service. This enables a coastal manager to go into the field and access geospatial

information to use for navigation and for carrying out temporal comparisons as well as completing updating tasks.

The GDSPDS is a means of more easily disseminating geospatial information to a wider audience without the need for each user to have expensive GIS software on their own computers. It provides support for data collection and updating, decision-making and report writing as well as greater interaction between coastal managers and users (Figure 18.2).

Figure 18.2 The Solway Firth GDSPDS. Viewing colour aerial photography via the Internet on a PDA (Aerial Photography Copyright © English Nature 1997)

18.6.2 Interactive Mapping using the Oregon Estuary Plan Book

Inforain has developed the Oregon Estuary Plan Book (http://www.inforain.org/interactivemapping/). Interactive maps from the Oregon Estuary Plan Book have been designed to help planners and decision makers in the evaluation of estuaries. The maps comprise several data layers allowing users to explore specific management units and habitat types. There is a base map layer together with other layers comprising information on dredge material disposal sites, estuary habitat, mitigation and special mitigation sites, estuary management units, shoreline management units, and areas of significant habitat (http://www.inforain.org/dlcd.asp). Datasets for download are stored in a compressed format. Provision of an image record is especially useful for

interpretation and communication to the layperson, either in a raw or processed format.

18.6.3 FREMP GIS: An Estuary Management Tool

FREMP (Fraser River Estuary Management Program) uses a GIS as a tool to promote integrated decision-making in the estuary (North and Beckmann, 1999). It uses orthophotos as a base layer and a number of thematic data layers including habitat inventory to assist in environmental recommendations. Initially FREMP did not use GIS software, but instead utilised aerial photographs and digital databases. Adoption of GIS, however, improved spatial data organisation and integration and provided wide access to data and information. It also provided a communication tool. Datasets comprised habitat inventory and shoreline classification maps, digital photographs, development projects, basemaps from FREMP and other external sources. In addition, metadata was provided. The interface to FREMP is an extension of the standard ArcView software but designed to be more user-friendly and including added functionality.

18.6.4 Virtual Estuary

This is a comprehensive online system for retrieving, analysing and integrating data and information related specifically to the Hudson Estuary in the USA. It has been targeted at a wide range of end-users from the scientist, to the decision-makers, and the general public. A joint project of Columbia University, COAST (Clean Ocean and Shore Trust) and a range of other institutions and agencies with a direct interest in the estuary. An online mapping system (using ESRI's ArcIMS) is also part of the Virtual Estuary project and is designed allow the creation of 'on-the-fly' maps from the estuary databases (http://www.nynjcoast.org/ Hudson_Estuary/hudson_estuary_.html).

18.6.5 OZEstuaries

OZEstuaries (http://www.ozestuaries.org) is an Australia-wide database that enables managers and stakeholders to develop a greater understanding of their own estuary, and where it fits into the national picture. It facilitates comparison, provides information on estuarine form and function, identifies potential threats, provides information on environmental indicators, delivers the National Land and Water Resources Audit (NLWRA Estuary Assessment, 2000) condition assessment, and provides a single point of entry for information on Australia's estuaries. This will be enhanced over time with the help of scientists, managers and stakeholders, with the provision of additional data, the addition of a GIS front end, estuary models and feedback from end-users.

18.6.6 PIVOT

In the USA, the National Estuary Program of the Environmental Protection Agency links community to agencies. PIVOT (Performance Indicators Visualization and Outreach Tool) provides an interactive graphics and maps toolbox to help users better understand the issues and visually track the National Estuary Program's progress toward achieving its habitat restoration goals. Provision of a map that combines spatial relationships of factors to priority issues and the management actions designed to address those issues is deemed to be a powerful tool as it helps to raise questions about the effectiveness of management actions (http://www.epa.gov/owow/estuaries/pivot/overview/intro.htm).

18.6.7 DEMIS

DEMIS (Dynamic Estuary Management Information System) is part of the Oregon Coastal Management Program and is designed to provide a useful information depot for both traditional and digital information for the estuaries of Oregon (http://www.lcd.state.or.us/coast/demis/demisfaq.htm). It provides access to data, GIS layers, metadata, estuary plan book information, as well as a CD. DEMIS has arisen in response to the problems associated with the current fragmentation of information. It has been designed to help improve access to information; establish a GIS (aggregate, standardise, distribute digital data); accumulate and analyse data for estuary restoration; incorporate new data into an information management system; and to maintain timely and efficient updates.

18.7 SUMMARY AND CONCLUSIONS

Powerful information technology tools now support the development, operation and use of estuary information systems of various different types, providing considerable potential to aid in estuary management. Information technology has greatly enhanced the data-into-information pathway, providing the means and the tools to collect more data, to process it more rapidly, and to communicate the results of analyses more effectively to a wider community. In recent years, rapid developments in the geospatial technologies have greatly enhanced the collection of spatial data, information processing, modelling, and visualisation of information in new, interesting, and informative ways. The capability to gather higher resolution data from remote sensing, to process, integrate and visualise environmental data within a GIS, and the use of GIS as a data handling tool for spatial analysis and modelling has provided the basis for studying the environmental processes operating in an estuary. Online GIS and decision support tools are also empowering the policy and decision-maker, as well as the coastal manager and practitioner.

In the future the practical application of the geospatial technologies will grow considerably providing tools that will aid the coastal manager and practitioner in the workplace. Some of the important issues that will govern the greater use of this technology in the future, however, will be end-usability of the hardware and

software, integration, and functionality. Moreover, data availability, data standards, data costs and copyright will also largely determine how much of the potential of geospatial technology can be practically implemented and realised in the workplace.

18.8 REFERENCES

Anonymous, 2003, Wind Turbines – I Can Hardly See Them! *Times and Star*, Friday March 21st 2003.

Brown, V.A., 1995, *Turning the Tide: Integrated Local Area Management for Australia's Coastal Zone*, Department of Environment, Sport and Territories, Canberra (second printing).

Brown, V.A., 2003, Personal Communication.

English Nature, 1993, *Estuary Management Plans - A Co-Ordinator's Guide. Campaign for a Living Coast.*

Fabbri, K.P., 1998, A Methodology for Supporting Decision-making in Integrated Coastal Zone Management. *Ocean and Coastal Management*, **39**, pp. 51-62.

Green, D.R., 1994, Using GIS to Construct and Update and Estuary Information System. *Proceedings of Management Techniques in the Coastal Zone. University of Portsmouth*. 24th-25th October, pp. 129-162.

Green, D.R., 1995, User-Access to Information: A Priority for Estuary Information Systems. *Proceedings of CoastGIS'95. International Symposium on GIS and Computer Mapping for Coastal Zone Management*. University College Cork, Ireland. February 3rd-5th, pp. 35-60.

Green, D.R. King, S.D., and Morton, D.C., 1998, Small-Scale Airborne Data Acquisition Systems to Monitor and Map the Coastal Environment (D.C. Morton and S.D. King) *Paper/Poster Paper for the Marine and Coastal Environments Conference* - San Diego, U.S. (October 1998). (Paper published in Proceedings. Volume 1: 439-449).

Green, D.R. and King, S.D., 1999, *Forth Estuary Modelling System (FEMS).* CD-ROM.

Green, D.R. and King, S.D., 2000, *Practical Use of Remotely Sensed Data and Imagery for Biological and Ecological Habitat Monitoring in the Coastal Zone.* Report/Monograph, (JNCC/English Nature).

Green, D. R. and King, S.D., 2003, Applying Geospatial Technologies to Weed Mat Monitoring and Mapping: The Ythan Estuary, NE Scotland. Special Session. *Proceedings of the Association of American Geographers*. March 5-9, New Orleans, Louisiana, USA.

Green, D.R., King, S.D, Annoni, A., Margoulies, S., and Humblet, J-P., 2003, Nature-GIS: Developing a Pan-European Geospatial Data Model for Protected Areas (PEPAN). Paper presented to 6th AGILE Conference on Geographic Information Science, 24-26 April, Lyon, France.

Green, D.R. and Ray, S.T., 2002, Using GIS for siting artificial reefs – Data issues, problems and solutions: 'Real World' to 'Real World', *Journal of Coastal Conservation*, **8** (1), pp. 7-16.

Hinwood, J.B. and McLean, E.J., 2002, Modelling the Evolution of an Estuary. In, *Proceedings of Littoral 2002 – The Changing Coast*, edited by Veloso Gomes,

F., Taveira Pinto, F., and Das Neves, L., EUROCOAST/EUCC. Porto, Portugal. September 22nd -26th. pp. 245-251.

Jernakoff, P., Hicl, P., Ong, C., Hosja, W., and Grigo, S., 1996, Remote Sensing of Algal Blooms in the Swan River. *CSIRO Marine Laboratories Report 226. CSIRO Blue-Green Algal Multi-Divisional Program.* CSIRO, Australia, March 1996.

Lewis, L.J. and Kelly, T.K., 2001, A Short-Term Study of the Effects of Algal Mats on the Distribution and Behavioural Ecology of Estuarine Birds. *Bird Study*. **48**, pp. 354-360.

North, S.M. and Beckmann, L., 1999, FREMP GIS: An Estuary Management Tool. *Proceedings of GIS'99: GeoSolutions in a Coastal World - Integrating Our World.* Vancouver. Thirteenth Annual Conference on Geographic Information Systems. March 1st-14th, pp. 125-129.

NLWRA, 2002, Australia's Near Pristine Estuaries - Assets Worth Protecting. http://www.nlwra.gov.au.

Orr, G., 2003. A Study of the Ythan Estuary. *Unpublished Undergraduate Dissertation*. University of Dundee, Scotland, UK.

Ripley, H., Jones, W., and Scheibling, R., 2002, Mapping Invasive Species: Assessing the Impact of Species Invasions on Coastal Ecosystems Using Remote Sensing. *Backscatter*. Fall/Winter 2000. pp. 8-12.

Sidaway, R., 1991, *A Review of Marina Developments in Southern England. Report.* Royal Society for the Protection of Birds (RSPB) and the World Wildlife Fund UK (WWF-UK).

Stapleton, C. and Pethick, J., 1996, Coastal Processes and Management of Scottish Estuaries. III: The Dee, Don and Ythan Estuaries. *Scottish Natural Heritage Review*, **52**.

Stove, G.C., 1978, The Hydrography, Circulation and Sediment Movements of the Ythan Estuary. *Unpublished PhD Thesis*. University of Aberdeen.

Townend, I., 2002. Identifying Change in Estuaries. In, Veloso Gomes, F., Taveira Pinto, F., and Das Neves, L. (Eds). *Proceedings of Littoral 2002 – The Changing Coast*. EUROCOAST/EUCC. Porto, Portugal. September 22nd -26th. pp. 235-243.

Vivoni, E. R., and Camilli, R., 2003, Real-Time Streaming of Environmental Field Data. *Computers and Geosciences*, **29,** pp. 457-468.

Wright, R., Ray, S., Green, D. R. and Wood, M., 1998, Development of a GIS of the Moray Firth (Scotland, UK) and its application in environmental management (site selection for an artificial reef), *The Science of the Total Environment*, **223** (1), pp. 65-76.

Young, D.R., Cline, S.P., Specht, D.T., Clinton, P.J., Robbins, B.D., and Lamberson, J.O., 2000, Mapping Spatial/Temporal Distributions of Green Macroalgae in a Pacific Northwest Estuary Via Small Format Color Infrared Aerial Photography. *Proceedings of the Sixth International Conference on Remote Sensing for Marine and Coastal Environments*, Charleston, South Carolina, USA. 1st -3rd May, Veridian ERIM International, Ann Arbor, Michigan, pp. II-285-II-291.

CHAPTER NINETEEN

A Territorial Information System (TIS) for the Management of the Seine Estuary – Environmental and Management Applications

Jean-Côme Bourcier

19.1 INTRODUCTION

Located at the geographic, sociological, economic, and environmental convergence of multiple interest groups, the estuary of the Seine (Normandy, France – Figure 19.1) presents itself as a coastal zone subject to divergent, if not contradictory, anthropogenic pressures. In fact, its position vis-à-vis its ecosystems gives it rare and specific functions and its location within geographic space confers strategic advantage. The urbanisation and level of primary industry of the region must develop in directions that respect an ecological heritage, and an exceptional landscape, that is subject to the attention of local communities as well as local, state and European levels of administration. Furthermore, the existence of various pollutants, risks from technological hazards, the exploitation of natural resources, an upstream/downstream divide, a disparity between left and right banks, and multiple activities, etc., are just some of the key issues in this unique space.

In this context, geographical information systems (GIS) present a wealth of applications as tools capable of collecting, harmonising and analysing data about the estuary coastline in order to monitor the health of the region and provide summary information as an aid to decision-making. Harnessing these technological, methodological, and operational tools enables coastal management problems to be defined and resolved in an holistic and consensual manner. They could also lead to a new approach to management and a remediation, by providing a better understanding of coastal processes, functioning and change; an approach more attuned to the natural and human complexities of a geographic space that encompasses all levels of organisation, and which has set sustainable development as primary goal.

These distinguishing features justify the establishment and use of a GIS specific to the Seine estuary within the CIRTAI laboratory at the University of

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

Le Havre (UMR 6063 CNRS – Research's group CIRTAI). More specifically, it consists of a territorial information system (TIS) tied to a multidisciplinary and multi-thematic, geographically referenced database designed to improve understanding of the processes and functioning of the estuary in order to support operational and participatory management practices. Within the range of spatially referenced information systems, (Prélaz-Droux, 1995), this territorial information system sits mid-way between the two end-points of land information system and geographical information system as defined by the author: it exceeds the institutional limits of a land information system, and applies the potential of a GIS in its ability to support development of land and environment in a complex geographic space that requires management at multiple scales and involving multiple actors.

The aim of this article is to demonstrate a particular application of the TIS, as an observatory of the natural heritage of the Seine estuary. Within this application, different levels of information have been integrated and analysed in order to provide qualitative and quantitative assessments of this heritage by means of territorial diagnostics.

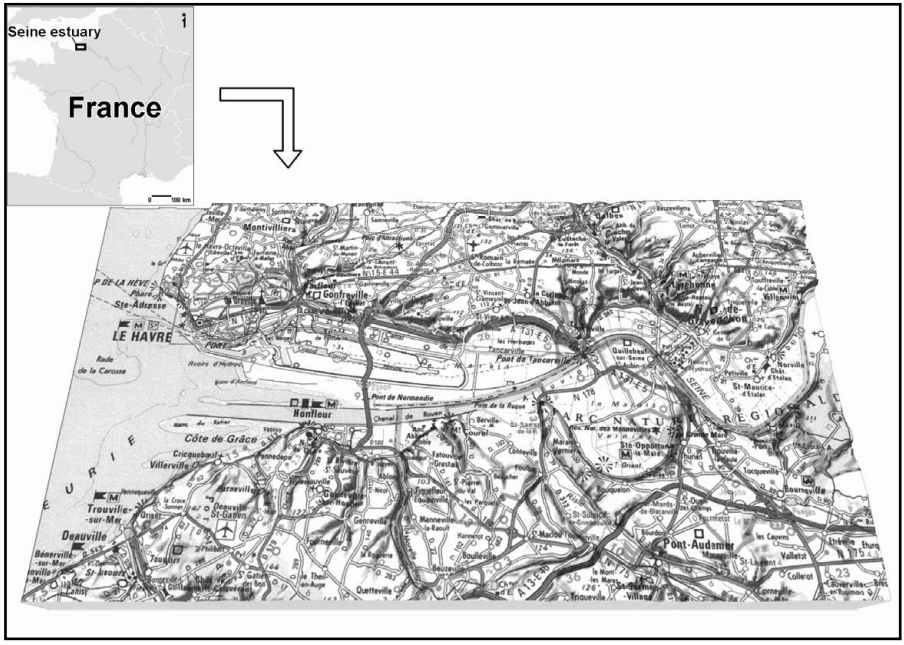


Figure 19.1 The Seine estuary (Normandy, France).

19.2 ON THE NEED FOR AN OBSERVATORY OF THE NATURAL HERITAGE IN THE SEINE ESTUARY

As a meeting point for terrestrial, fluvial and marine environments, the Seine estuary integrates a number of biologically productive and varied ecosystems. It is

an integral element in the routeway for many species of migrating birds and provides shelter and nutrients for fisheries and shellfish farming as well as for certain invasive species.

The estuary basin is characterised by the juxtaposition of alluvial plains and calcareous plateaux, giving rise to an extraordinarily diverse range of humid-zone landscape types that act as sponges and filter-cleaners for the environment.

Nowadays, the natural heritage of this estuary is subject to the impacts of numerous activities and processes at the local scale (industrialisation, urbanisation, agriculture, resource exploitation) as well as nationally (it is the outlet of a drainage basin within whose borders is found 30% of the French population and 40% of the national economic activity). This convergence of interests, and the conflicts that arise, raises the problem of how to conserve natural spaces whose resources are limited and whose equilibrium is fragile.

Management of this territory must therefore ensure balanced development, taking into account both scientific data and socio-economic factors (Chappuis, 1993), at a scale that is essentially local but which also has national and international elements (Bailly, 1995). Every stage in the process involves nature and the landscape, and it is often necessary to draw on a complete inventory of the natural and landscape heritage in order to steer decision-making in appropriate directions whenever choices present themselves.

In order to address these requirements, the Seine Estuary TIS acts as an observatory of natural heritage, and an intelligent tool for providing territorial and environmental indicators needed to manage this unique coastal system. GIS offers technologies with potential for better planning, management and analysis of our environment (Bryant, 1993); for tackling the integrated, holistic and rational management of space; for pooling necessary information; and for overcoming institutional limitations, to allow all levels of environmental organisation (hydrological systems, landscape systems, drainage basins, etc.) to be taken equally into account.

The objective is to make available to environmental managers and decision-makers a suite of key indicators that will allow better understanding of disparities between left and right banks of the river, conditions of flood and low water, competition and compatibilities of estuarine biota, etc., and ultimately allow the delimitation of zones of strong environmental vulnerability. This will also help bridge the gaps between scientific research and real-world practice and, more precisely, between those scientists, technicians, professionals, and politicians who are involved in strategic and political decision-making.

The territorial information system for the Seine estuary aims to bring the many facets of the same geographic reality into a common framework. Thus organised and structured, all available data for a given location may be integrated as part of a single analysis, in order to better define environmental objectives, and to provide support for decisions relating to territorial planning and management within the estuary.

19.3 FROM DATA COLLECTION TO DEVELOPMENT OF ENVIRONMENTAL DIAGNOSTICS

This TIS for the Seine estuary exploits a multi-source and multi-thematic geographically referenced database, presently consisting of more than 80 layers of data in vector format, standardised on the Lambert conformal conic projection and georeferenced on the Lambert zone 1 coordinate system (the local system for the north of France). Data in raster format are used to provide rapid enrichment and updating of the database, according to immediate needs. These data layers are all subject to a number of quality-control parameters (Laurini and Milleret-Raffort, 1993) including accuracy, precision, and timeliness (Rouet, 1991), and each layer is the subject of a file of metadata (Bourcier *et al.*, 1998) to ensure that it is used appropriately and optimally. A dynamic link has been developed, so that the metadata may be consulted directly via the user interface of the GIS software.

The operational use of the TIS requires, *inter alia*, data to be updated regularly, since they describe a territory that is subject to strong spatio-temporal evolution.

Data relating to the natural heritage of the Seine Estuary are collected on a regular basis by various groups: the regional Ministry of Environment for Upper and Lower Normandy, the Department of Hunting, an organisation for the conservation of wetlands, the Regional Nature Park of Brittany, Nature Park, etc. This demands an effective partnership, in order to respond to the specified quality criteria. Initially, exchange of these data was via hard-copy paper media (maps, cadastral plans, etc.), which implied heavy, time-consuming digitisation overheads. More recently, all the data-providing partners have opted for the advantages of digital technologies (GIS platforms, CAD, computer-assisted cartography), which has enabled much more efficient implementation of the system.

An environmental geodatabase is by definition multidisciplinary and multi-territorial; its quality and relevance depend on, among other factors, the abilities of the GIS team to manage interdisciplinary conflict, institutional compartmentalisation and protectionism, hierarchical management "pyramids" and cultures of regulation, etc.; and especially possessive attitudes and the equation of information with power, leading to a reluctance among some players to share information with others.

19.3.1 Environmental heritage diagnostics in the Seine Estuary

As has been discussed above, all management practices have environmental implications, and therefore require access to a complete inventory of the natural heritage. This essential knowledge of the natural resources of a region needs no further justification.

The natural heritage value of the Seine estuary is recognised through the implementation of a number of statutory instruments of national or European origin: the Ecological Fauna and Flora Inventory ("ZNIEFF"), Nature Reserve, Special Areas of Conservation (SACs), Biotope conservation, etc. To these are added other protective measures, such as the purchase of land.

It seems evident that this accumulation of regulations makes the direction of management projects even more difficult, especially since many of these provisions lack clarity, are obsolete, are contradictory and of uncertain relevance, or are imprecise and subject to multiple, subjective interpretations.

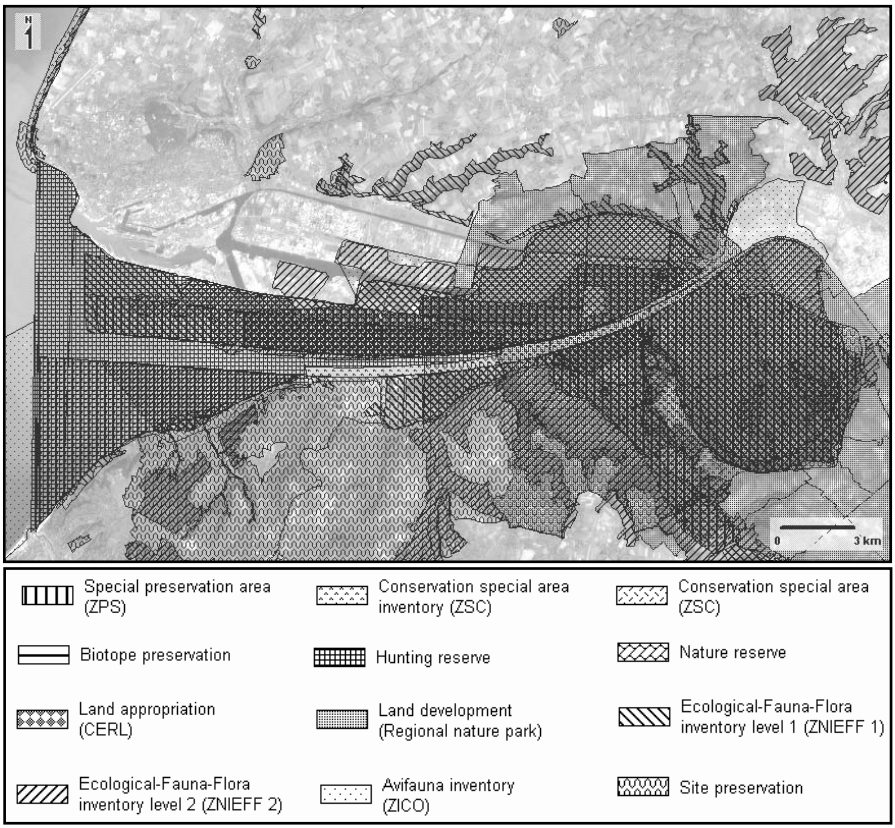


Figure 19.2 Spatial concentration of environmental conservation in the Seine estuary.

Superimposition of all these layers demonstrates visibly the multiplicity of initiatives and regulatory measures that currently apply to this territory. We may also note several important geographical overlaps between these various measures, rendering their representation in cartographic form particularly difficult.

Overall, such a representation implicitly demonstrates the high heritage value of this estuary, but the resulting legislative overload demonstrates clearly the need for harmonization of French Environmental law. French estuaries present a major challenge for the law, which must find ways to address their hybrid character and their multifunctional nature. Analysis of the legal treatment accorded these spaces demonstrates the weak acknowledgement in national law of their distinctive nature and, indeed, a strong tendency to promote their development, despite the accumulated knowledge of their ecological interest and their vulnerability (Auger and Verrel, 1998).

In terms of spatial analysis, it is necessary to go beyond simple superimposition of data layers, to examine the spatial distribution of particular entities and to synthesise the information they present. The aim is to create a new layer of information that compares, aggregates and analyses various perspectives on the region, so as to present alternative views of reality, for better environmental comprehension, decision-making and management. The complexity of the real world is thus rendered easier to read and communicate.

Thanks to the topological functionality of GIS, an expansion of the various layers of information relating to this natural heritage has enabled a "quantitative synopsis of protection measures" to be drawn up. This spatial analysis facilitates easy identification and classification of zones according to the sum of the protective measures that apply to them (Figure 19.3).

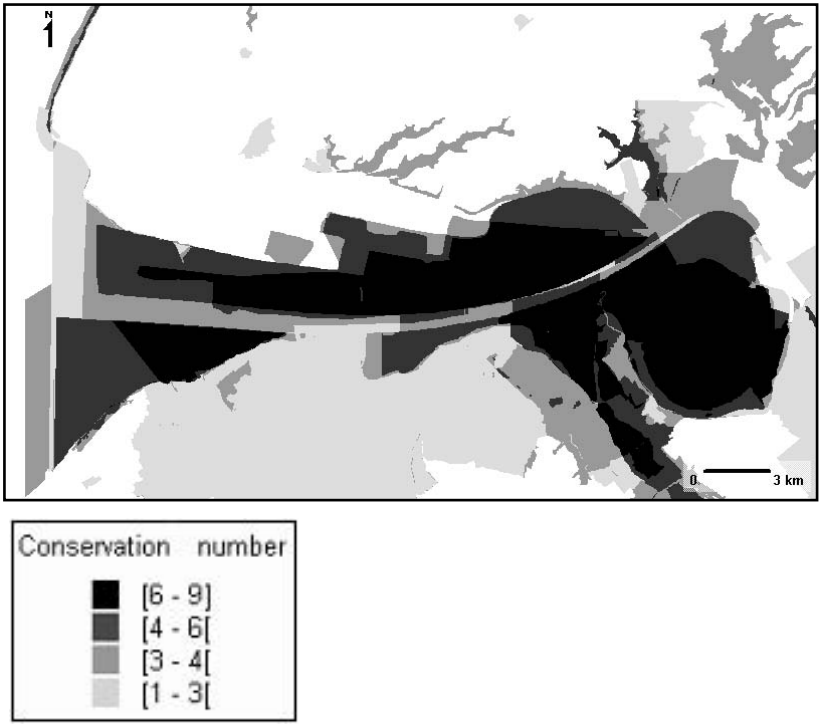


Figure 19.3 Quantitative synopsis of nature conservation in the Seine estuary.

While a diagnostic utility of this type is visibly effective at demonstrating the multiple protective measures, it also underscores the relevance of the known natural heritage of the Seine estuary, and explicitly reveals the conservation importance of the intertidal areas and of the flood plains. These latter are wetlands of acknowledged ecological importance, according to the National Programme of Research on Wetlands (Bourcier *et al.*, 1999). The TIS allows real-time spatial and thematic analyses, such as calculating the type, number and surface area of

biotopes for each of the identified sectors, thereby allowing optimal planning and management for these environments.

Still within the perspective of generating useful indicators for stakeholders and decision-makers, the foregoing increase in topological data is further enriched by the addition of information relating to the level of legal protection accorded to these zones. This leads to a "qualitative synopsis of the degree of protection of the natural environment", which identifies and classifies different zones according to a cumulative index of conservation (Figure 19.4).

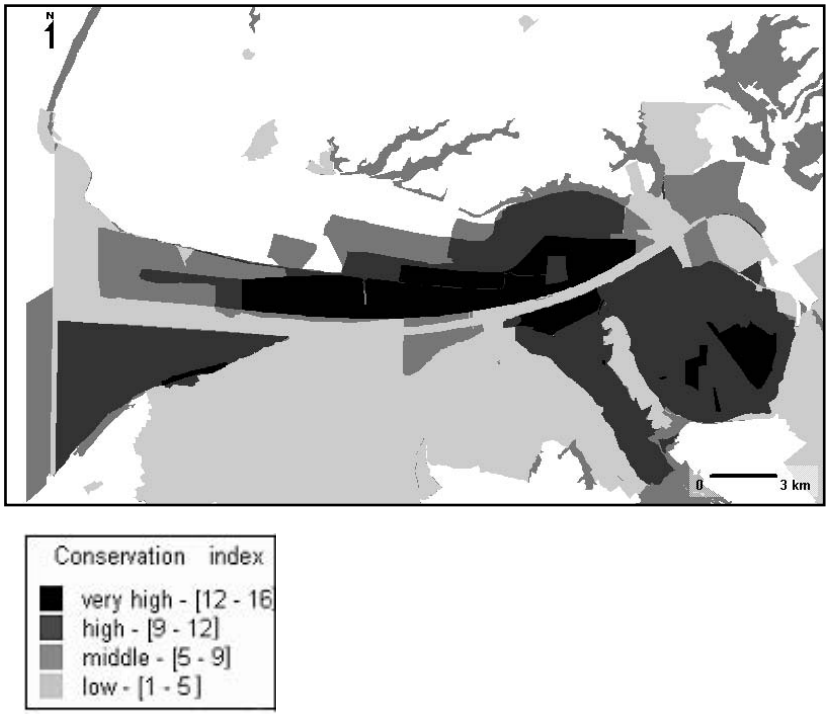


Figure 19.4 Qualitative synopsis of nature conservation in the Seine estuary.

The most protected zones (very high value) are located mainly in the alluvial plains, which underscores the ecological importance of these habitats.

This territorial index usefully complements the preceding map, by conveniently summarising the environmental constraints of the estuarine lands, providing indispensable information when territorial management projects or planning exercises are undertaken.

Achieving an optimal management of the natural heritage, at all levels of organisation, will probably require a greater harmonisation of the many superimposed levels (local to international) of environmental protection, in legal as much as in economic terms.

Additionally, this harmonisation offers the advantage of representing environmental constraints as homogenous zones that transcend administrative

boundaries, and which are thereby more appropriate to holistic understanding of the environment from the global to the estuarine scale.

From the perspective of optimal management of the natural environment, would it not be useful to avoid the heavy imposition of all these different protected zone boundaries and, perhaps, rely on European and international commitments (e.g. the Natura 2000 network)? These levels of administration could provide powerful tools for protection through regulatory and financial instruments.

19.3.2 An environmental impact of coastal management in the Seine Estuary

The usefulness of the spatial analytical functions of the TIS may equally be demonstrated in the assessment of the impacts of traffic noise on the bird life of the estuary: by combining different of layers of information which, at first sight, have no apparent link between them, we can demonstrate the complexity of the real world and, more particularly, the influence of human activities on the natural environment (Figure 19.5).

The alluvial plain north of the Seine estuary is also subject to conflicts between natural environment and economic development, the resolution of which calls for sustainable development strategies to be drawn up. In this context, spatial analysis allows vehicular traffic levels to be correlated with the natural habitat of the Spoonbill (*Platalea leucorodia*, a migratory bird, see Figure 19.6), in order to measure the potential impact of noise pollution on the distribution of this bird. The routeways thus identified are of particular interest when they lie adjacent to the Natural Reserve of the Seine Estuary, liable to be directly linked to Port 2000, the future expansion strategy for the current commercial port.

Statistics on the movements of vehicles (light and heavy freight) were obtained from the authorities of the three departments concerned, and were used to calibrate the “road traffic” data layer.

A noise level envelope was calculated, based on traffic levels of 5000 vehicles per day (all vehicles, travelling in both directions), which corresponds to a noise level of 70 dB SPL. The TIS then allows a buffer polygon to be created around selected entities, and this can be intersected with the Spoonbill habitats to create an output layer that precisely identifies those areas where noise disturbance could be a potential problem. By simple querying, it is easy to obtain the location of each area, as well as their spatial attributes (length, width, perimeter, area, etc.).

Figure 19.5 ***Platalea leucorodia*** (Spoonbill: a migratory bird).
(© Cybernat - Le Havre University).

Figure 19.6 Effect of road traffic noise on *Platalea leucorodia* area.

This type of spatial analysis can also be applied to other domains (pollution of air or soil, etc.), by developing specialised algorithms for use in modelling impact analysis or simple diagnosis and visualisation. In the final analysis, the GIS tools have demonstrable relevance to aiding understanding, decision-support and management of environmental resources in a territory such as the Seine Estuary. By means of regional indicators, they can contribute to more holistic approaches to territorial management operations.

19.4 CONCLUSION

The concept of environmental management demands engagement with the interface between society and nature, since the multiple activities (often concurrent and mutually conflicting) that take place within a region are governed by economic and social paradigms.

Even if environmental considerations are nowadays better taken into account during processes of economic and social decision-making, they are still often only brought in towards the end of the process, while the various actors may frequently lack important elements needed to establish priorities and anticipate problems. Current practices and procedures, at different scales, are poor at anticipating environmental consequences of territorial management, due to a lack of appropriate information technologies and, equally, of suitable conceptual tools.

Spatial management of an estuary such as that of the Seine can only be successful and balanced if suitable structures are established to arbitrate effectively and equitably between the many pressures that currently lead to conflict.

Sustainable development is frequently advocated as a solution to these problems, whereby a dynamic balance is sought between economics, ecology and the needs of society, at local levels of organisation but also compatible with national and international levels.

Implementation of sustainable development policies requires a global approach and interdisciplinary vision, so that policies can lead to improved decision-making. In this context, the TIS for the Seine estuary has many potential applications. It serves as a tool capable of collecting, organising, integrating, and analysing information about the environment, from which territorial indicators and syntheses may conveniently be obtained to support decision-making, and to help solve problems of regional management and development.

In particular, it allows the properties and geographic distribution of natural resources or areas of interest to be defined; it improves the compatibility and interoperability of diverse datasets; and it facilitates spatial analyses of environmental impacts that would otherwise be overlooked or discarded for reasons of cost or procedural complexity.

This TIS constitutes a relevant tool because it has as its first objective the representation of a region, in this way improving our understanding and permitting the description and analysis of natural and human phenomena produced by the landscape. Since it lies at the junction of several disciplines, this TIS brings convergence and standardisation to the description and representation of space. It aims to bring coherence to all the many aspects of the same reality and, structured accordingly, data relating to the same location can be integrated within a single analysis, in order to better define environmental objectives and provide support for arguments or reasoning concerning regional development and planning within the estuary. The TIS has thereby enabled information relating to the natural heritage of the Seine estuary to be related, aggregated, and synthesised, in order to create territorial indicators of use for understanding, decision-making and management of the region.

Nonetheless, “mastering GIS, and developing robust methods of analysis, implies the ability on the one hand to assimilate elements of different terminologies, and the concepts associated with them, so as to permit truly

multidisciplinary dialogue; while, at the same time, allowing questions to be formalised in terms that are compatible with the various elements of the information system" (Meillet, 1999).

Finally, it remains the case that these techniques carry limitations, even though they allow the integration of multiple sources of data. The user must always understand the consequences of their implementation, and minimise possible errors due to data quality, problems linked to the treatment of data, or due to sensitivities inherent in the selected methodologies (Marois, 1993). In other words, their use within the process of large-scale choice and decision-making requires those concerned to have a high degree of confidence in the relevance and reliability of the data that are exploited. The quality of data and of their interpretation thus prove to be the crucial *sine qua non* of any process of decision-making.

19.5 REFERENCES

Auger, Ch. and Verrel, J.-L., 1998, *Les estuaires français – Évolution naturelle et artificielle*, (Plouzané: IFREMER).

Bailly, A., 1995, *Les concepts de la géographie humaine*, (Paris: Masson).

Bourcier, A., Bourcier, J.-C. and Pouchin, Th., 1999, Appréhension de la cinématique paysagère par télédétection spatiale en estuaire de Seine: contribution au programme national de recherche sur les zones humides. In *La télédétection en francophonie: analyse critique et perspectives*, Collection Actualité Scientifique, (Montréal: Agence Universitaire de la Francophonie), pp. 215–223.

Bourcier, J.-C., Bourcier, A. and Pouchin, Th., 1998, *L'Observatoire des zones humides - Zones humides de l'estuaire et des marais de Seine: structure, fonctionnement et gestion*, (Le Havre: Université du Havre).

Bryant, C.R., 1993, Les SIG au service de l'aménagement et du développement communautaire. *Revue de Géographie Alpine*, **9**, (Grenoble: Université Joseph Fourrier), pp. 63–66.

Chappuis, J.-B., 1993, *Protéger la nature – Guide pratique de la protection de la nature et du paysage*, (Lausanne: Delachaux et Niestlé).

Laurini, R. and Milleret-Raffort, Fr., 1993, *Les bases de données en géomatique*, (Paris: Hermès).

Miellet, Ph., 1999, *SIG pour la gestion et l'aménagement urbain*, (Paris: Ministère de l'Équipement, des Transports et du Logement).

Marois, C., 1993, Environnement et Systèmes d'Information Géographique. In *Revue de Géographie Alpine*, **9**, (Grenoble: Université Joseph Fourrier), pp. 11-14.

Prélaz-Droux, R., 1995, *Système d'information et gestion du territoire: Approche systémique et procédure de réalisation*, (Lausanne: Presses Polytechniques et Universitaires Romandes).

Rouet, P., 1991, *Les données dans les systèmes d'information géographique*, (Paris: Hermès).

CHAPTER TWENTY

Developing an Environmental Oil Spill Sensitivity Atlas for the West Greenland Coastal Zone

Anders Mosbech, David Boertmann, Louise Grøndahl, Frants von Platen, Søren S. Nielsen, Niels Nielsen, Morten Rasch, and Hans Kapel

20.1 INTRODUCTION

Marine oil spill sensitivity mapping has become widespread. The purpose is to provide oil spill response planners and responders with tools to identify resources at risk, establish protection priorities and identify appropriate response and clean-up strategies. GIS is an important tool in the development of oil spill sensitivity maps and can also be used for presentation. Several different principles for information integration have been used in different countries (Anker-Nilssen, 1994; Dickens et al., 1990; Hall et al., 1997; Nansingh & Jurawan, 1999; Moe et al., 2000). An environmental oil spill sensitivity atlas was produced as part of the preparations for exploratory drilling off the West Coast of Greenland (Mosbech et al., 2000). We have adapted a Canadian sensitivity index system integrating physical, biological and human-use information and combined it with elements from a Norwegian sensitivity mapping system.

The Atlas covers the West coast of Greenland between 62° N and 68° N latitude. Although the area under study stretches only 700 km from north to south it encompasses approximately 18,000 km of coastline. It is the most populated area in Greenland with about 35,000 inhabitants living in 4 towns and 6 settlements. It is extremely important for fisheries and it is ecologically highly important for a number of seabird and marine mammal species.

The Atlas is a multidisciplinary GIS project integrating many kinds of scientific data and local knowledge (traditional ecological knowledge). Although studies on geomorphology and coastal spawning areas were initiated for the project it was a major challenge to compile and get the most out of the existing data from many sources. This paper will outline and discuss how the data was integrated, the principles used to identify and prioritise the sensitive areas, the final Atlas product ,and the dialogue during the community consultation.

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

20.2 METHODS

20.2.1 The Atlas

The following elements are included in the Atlas

- coast types,
- oceanography, ice and climate,
- biological resources (fish, birds etc.),
- fishing and hunting,
- selected areas (e.g. seabird breeding colonies),
- archaeological sites,
- logistics and oil spill response methods.

The Atlas covers the coastline in the scale 1: 250 000 and the offshore area in the scale of 1:3.5 million. In a PDF-version 34 maps cover the coastline and one set of maps shows index values for coastal sensitivity and symbols for the elements of classification (see map and legend example plates 20.1a and 20.1b and colour inserts following page 164). Another set of maps show coast types, logistics, and proposed methods of oil spill response for each area (see legend and map example plates 20.2a and 20.2.b, see colour inserts).

Although a limited number of printed copies of the Atlas have been produced, the main version is published on CD. Since the target group of this atlas is not necessarily familiar with GIS, it was decided to distribute the final product in a form that is simple and quick to use and with an easy-to-use interface. The atlas with all text, maps, tables, and images is distributed on CD and on the Internet as a number of linked PDF-documents, which can be read with the free Acrobat Reader application. This allows a large amount of background information to be included as hypertext and at the same time to have a succinct and operational document. Furthermore, on the CD all geographic data is available for users to pan, zoom and print seamless maps with the free GIS-viewer ArcExplorer.

The Atlas was developed with an ESRI ArcView project and an MS Access database. This database software was chosen because many of the existing data and institutional databases employ MS Access. Spatial information on the model parameters is stored in ArcView whereas all model parameter attribute values are in the Access database. Index calculations are performed with Visual Basic for Applications in the database.

An important component of the Atlas is a sensitivity ranking system, which is used to calculate an index value describing the relative sensitivity of coastal and offshore areas. The sensitivity index value is calculated based on information on resource use (human use), biological occurrences, and physical environment. The sensitivity index system is based on a Canadian system used in Lancaster Sound (Dickins et al., 1990) and modified to the west Greenland fauna and other specific requirements of the Greenland study area. As a supplement to the Canadian ranking system, a number of smaller areas have been selected for priority in case of an oil spill. The selection of these areas is based on the principles from a Norwegian system (Anker-Nilssen, 1994), which gives priority to oil spill-sensitive areas for the purposes of oil spill contingency planning. While the

Canadian index system covers the entire coast in 50-km units, and thus gives a general overview of the sensitivity, the Selected Areas system uses actual borders and thus pinpoints the sensitive areas and leaves the rest unclassified.

20.2.2 The Sensitivity Index Calculation

The shoreline zone in the study area has been divided into 279 shoreline areas, each consisting of approximately 50 km of shoreline or, in the case of archipelagos, groups of islands and skerries having roughly 50 km shoreline. The offshore zone in the study area has been divided into 12 offshore areas (including 1 major fjord). The boundaries of the offshore areas are based on bathymetry and ice conditions during the winter.

The importance of resource use and the abundance of a number of biological occurrences in each of the 279-shoreline and 12 offshore areas were rated on a scale from 0 to 5 by using a number of subindices. Site-specific significant habitats are indicated on each shoreline segment. For example, such sites include important bird colonies and terrestrial haul-out for harbour seals. Photos of the coastal setting for about 50 bird colonies have been included and can be accessed from links.

As a part of the project, classification of the coastline geomorphology has been conducted from aerial photographs, e.g. the occurrence of rocky shores and beaches. An index value (oil residence index) of the self-cleaning ability of the coast after an oil spill has been calculated based on this classification in combination with shoreline exposure to waves and ice. For example, oil on a rocky coast exposed to wave action will be cleaned faster than oil on a beach in a protected lagoon.

An oil spill sensitivity index value has been calculated for each of the 279 shoreline and 12 offshore areas based on:

i) the abundance and sensitivity of selected species (or species groups);
ii) resource use (human use), mainly fishing and hunting;
iii) the potential oil residency on the shoreline (Oil Residency Index) based mainly on wave exposure, substrate and slope of coast;
iv) the presence of towns, settlements and archaeological sites (for shorelines).

The sensitivity index value for each of the 279-shoreline areas and 12 offshore areas is given. All areas are ranked as extreme, high, moderate or low sensitivity areas and a corresponding colour code has been used. The detailed index value calculations for each shoreline and offshore area can be accessed by hyperlinks.

With the settings we used, the average contributions to the final sensitivity values for the shoreline areas are: biological occurrences 49%, resource (human) use 20%, oil residence index 14%, archaeological sites 12%, communities 4% and special status areas (Ramsar sites) 1%. However, this is a simplification since the oil residence index value is a factor in the calculation of the value for biological occurrences, and thus also has an indirect influence on the sensitivity values.

20.2.3 The Selected Areas

To supplement the rather general mapping of shoreline sensitivity using the 50-km long shoreline areas, a number of small sensitive localities have been pinpointed as priority areas. A total of about 80 areas along the coast and within fjords have been selected as priority areas in the case of an oil spill situation. The basis for their selection is, compared to the coastline in general, that they are:

i) of high value either environmentally or for resource use;
ii) sensitive to oil spills; and
iii) of a size and form that may allow effective protection in an oil spill situation with a manageable amount of manpower and equipment.

The last criterion is important because it elucidates the rather limited possibilities to protect the coastline during a large oil spill and shows that tough prioritising is necessary.

20.2.4 Example of Indexing and Data Integration: Seabird Colonies

The major part of the biological information incorporated in the Atlas derives from databases maintained by NERI and the Greenland Institute of Natural Resources. As far as possible automatic selection and integration routines were developed to link these institutional databases and the Atlas database. One of the databases is a file on all seabird breeding colonies known in Greenland, with information on the birds, the sites and all survey results from the sites (Boertmann et al., 1996). A selection from the database was used in the Atlas. This selection was based upon the geographical range between 62° and 68° N and on the most comprehensive surveys, as many colonies have been surveyed several times. However, inferior colonies have been omitted and criteria for inclusion are listed in Table 20.1. Most colonies have a mixed species assemblage and the total number of colonies (with different geographical location) selected is 158.

The seabird colony data is used to produce a relative seabird abundance input value to the Seabird Oil Vulnerability Index scoring system which is integrated into the overall Shoreline Sensitivity Index. The seabird oil vulnerability index component takes into account the sensitivity to oil spills of the bird species both on an individual level and on a population level, as far as possible based on scientifically derived information on the characteristics of each species (Anker-Nilssen, 1987, Mosbech, 2000). These sensitivities are dependent on the behaviour and ecology of the birds, but also the distance to neighbouring colonies, which is a measure of the ability to re-colonise a colony. Moreover they take into account the status of the breeding population within the region, whether they are decreasing, increasing or stable, and finally their international conservation status (Mosbech et al., 1996).

20.2.4.1 Comments on some of the criteria

The breeding population of common eider in West Greenland has decreased seriously for a century, and within the Atlas region large and dense colonies have disappeared. Large breeding populations are mainly found dispersed in extensive

archipelagos. To exclude sites with a few scattered nesting eiders the criterion for inclusion is ≥ 5 birds.

As gulls are only moderately sensitive to oil spills, only the largest colonies are included.

Black-legged kittiwake colonies with less than 50 pairs are excluded as they tend to be less stable over time.

Arctic terns usually breed in dense colonies on low islands. The population in West Greenland is generally decreasing. Small colonies of less than 30 pairs are excluded. Terns are moderately sensitive to oil spills, but colonies situated on low islands are very sensitive to disturbance e.g. from oil spill response activities.

All species of the family auks (alcids) are very sensitive to oil spills. This is caused by their behaviour as well as by their very low population turnover. Therefore, protection of their breeding sites is a high priority. Moreover, the breeding population of Brünnich's guillemot in West Greenland is seriously decreasing due to a very high hunting pressure, and the few breeding sites within the Atlas region are therefore all included.

20.2.4.2 Relative abundance of the species/species groups

Some of the species are pooled into groups, as they have a similar sensitivity to oil spills. The relative abundances applied to species and species groups appear in Table 20.2.

The relative abundance is the input from these site-specific shoreline species elements to the calculation of the sensitivity index for each shoreline area. Similar relative abundance values are calculated for all the biological and human use resources.

20.2.5 Example of Indexing and Data Integration: Archaeological and Historical Sites

Information on the archaeological and historical sites in Greenland has been collected for over 100 years and is contained in a database under development at the Greenland National Museum and Archives. There is information on about 1500 archaeological sites in the mapped area. All known prehistoric and historic sites are included in the present Atlas, but to protect the sites only the most basic information on each site is given e.g. the site type and dating of the site.

There are great differences in what is known about different archaeological sites in the Atlas area due to the source of the information. The major part of the information is derived from secondary sources (i.e. local informants) and may therefore be less accurate. Basic information on each site is often not reported and therefore the degree of sensitivity of each site is based on an estimate.

The sensitivity of each archaeological site is expressed on a scale 1-3;

1. Sites that are unlikely to be endangered.
2. Sites threatened either directly or indirectly. This group includes all sites close to the coastline that are presumed to represent value either as a historical source, as a recreational site or as a special historical highlight.

3. Threatened sites of significant importance that demand special measures be taken in case of a marine oil spill or other activities in connection to mineral or oil investigations and extraction.

Based on this assigned sensitivity each archaeological site is ranked on a scale from 0-5. For each segment (50-km shoreline area) these rankings are added and the results are grouped again on a scale from 0-5 similar to the relative abundance of the biological resources.

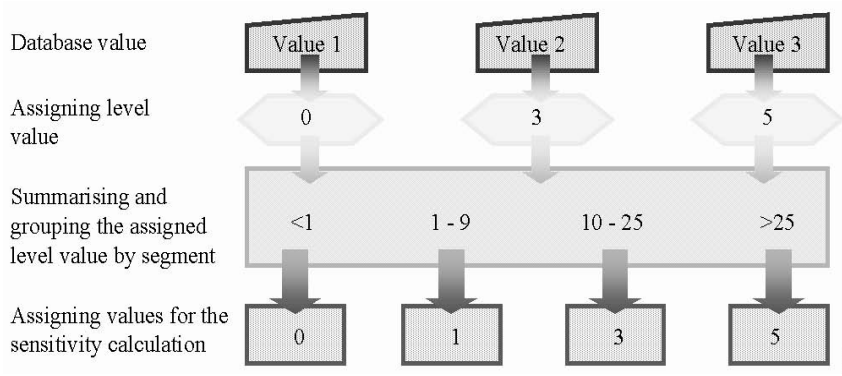


Figure 20.1 The Figure shows how the archaeological sites were given assigned values for the sensitivity calculation.

20.2.6 Integrating Local Knowledge

Local knowledge is included at several levels in the Atlas. Besides specific interview studies much information in the institutional databases is derived from locals helping researchers finding sites such as bird colonies and areas of archaeological significance. A specific interview study has been conducted to map the capelin (*Malotus villosus*) and lumpsucker (*Cyclopterus lumpus*) spawning and fishing areas and arctic char (*Salvelinus alpinus*) rivers (Nielsen et al., 2000). Capelin and lumpsucker spawn in the tidal zone and are therefore very vulnerable to oil spills. The study relied heavily on preliminary maps distributed for comments and the addition of supplementary information. Questionnaires and maps with preliminary information were sent out and returned with additional information, and these results integrated into new versions of the maps. Semi-directed interviews were conducted following the methodology of Huntington (1998). In the semi-directed interviews the maps were used to facilitate and guide the exchange of information among a group of local fishermen.

20.2.7 Community Consultation

A community consultation phase was carried out during the project. A draft version of the Atlas was mailed to local communities and user organizations. Later

meetings were held with all municipal councils and most settlement councils as well as the local hunter and fishermen associations. The purpose of the community consultations was not only to verify the information in the Atlas, but also to allow for comments and discussion of the method we had used and the result of the sensitivity ranking. A computer with the GIS and a printer, which could print overheads, was brought along and used where specific thematic maps were needed.

Maps on overheads explicitly showing the human use ranking were presented and facilitated fruitful discussions from which much supplementary information came forward. Although the sensitivity mapping was positively received the level of abstraction and complexity in the index calculation sometimes lowered the enthusiasm in discussions of the final rankings. As an example an extremely important capelin spawning area could be in an area with low sensitivity ranking, because it takes many sensitive elements to raise the final ranking of the coast to high sensitivity. Here the use of small Selected Areas provides a better tool to accommodate local expectations and wishes.

In the settlements, where Inuit hunters are the majority, the layman – expert divide was very pronounced and personal contacts were of crucial importance for establishing a fruitful dialogue. In some settlements there was concern that the publication of maps with "their" Arctic char rivers would tempt townspeople to come fishing. It was therefore decided to restrict the publication of this information on the Internet until proper regulation to protect fishing rights is in place.

The advent of oil exploration in this area introduces a "low probability – high potential consequence" risk. The oil spill sensitivity Atlas is a valuable tool to minimise the risk. However, it is also important to communicate the realistic limitations of this tool.

20.2.8 Versatile Distribution of the Atlas

The users of the atlas are diverse with different backgrounds and qualifications. The primary target groups are government administrators (both national and municipal) who are involved in planning oil spill response; their counterparts in the oil industry; and oil spill responders in industry, navy and the local fire brigades. Since the target groups of this atlas are not necessarily familiar with GIS it was decided to distribute the final product in a form that is simple and quick to use and with an easy-to-use interface. The atlas with all text, maps, tables and images is distributed on CD and on the Internet as a number of linked PDF-documents, which can be read with the free Acrobat Reader application. Furthermore, all geographic data is available for users to pan, zoom and print seamless maps from the CD with the free GIS-viewer ArcExplorer. There has also been a demand for large plastic-coated map sheets and a limited number of printed copies of the atlas.

The PDF-document consists of 434 pages including 34 map areas covering the entire area at a scale of 1:250 000, suitable for printing on paper. Each map area is covered by two types of maps. One contains the shoreline sensitivity results (plate 20.1) and another the physical environment and logistics (plate 20.2). The PDF-document includes extra information for decision-makers e.g. aerial photos,

photos of specific bird colonies and sensitivity index calculations for each segment, accessible by hyperlinks.

Adobe Acrobat is not a GIS program, but is a good map viewing system. Distributing the Atlas with Acrobat gave us the ability to decide how the final maps were presented for the end user. Another advantage to using Acrobat was the map-producer's ability to decide at what scale the maps should be used. In this way there is no risk of misinterpretation of the spatial data caused by using them at a scale for which they were not designed. The end user must print the maps at the exact scale at which they were meant to be printed – this option is not available with the GIS-viewer ArcExplorer. ArcExplorer gives the end user the ability to view and use the maps and their information without having to learn the basics of a GIS.

20.3 CONCLUSION

The Atlas was produced over a 7-month period and the relatively short campaign worked well. The involved persons could see the end already when they started which helped the facilitation of enthusiasm among relevant participants.

The integration of institutional databases worked well, and the approach makes updates of the sensitivity Atlas and developments/adaptations to other purposes easier. The task initiated valuable contacts and discussions among database managers in the relevant institutions, as well as discussions with a focus on the research opportunities of integrated GIS analysis of georeferenced databases from various fields.

The PDF-format is a good choice for distributing an atlas with a mix of formatted text, tables, maps and images which should be connected with hyperlinks. The quality of the maps is better than what is possible with use of today's Internet mapping solutions. The same documents can be used for printing, CD-distribution and Internet-access. Since Adobe Acrobat is not a GIS, a GIS-viewer was necessary to enable users with no GIS experience to use the feature tables on the CD. ArcExplorer was found to be useful since the software is free of charge and comes with an extensive user guide.

The Selected Area concept helped pinpoint where to focus oil spill response and facilitated communication during community consultations because a single hot-spot could be highlighted based on one single very important item. The need for Selected Areas to supplement the general oil spill sensitivity mapping is related to the size of the coastal units. We chose 50-km units, like in the Lancaster Sound Atlas (Dickens et al., 1990), because of the vast area and the relatively limited amount of information. If the size of the coastal units is reduced to as little as 3 km on average, as is the case for Newfoundland in The Canadian Atlantic Region Sensitivity Mapping Program (Anonymous, 1999), the need for Selected Areas disappears.

The experience from community consultations and collection of local knowledge is that local communities need personal contact on a continuous basis to overcome the layman-expert divide. Continuity and build-up of confidence are important prerequisites for exchange of information and ideas. Some restriction of the publication of information obtained from local sources should be accepted.

20.4 THE STUDY TEAM

The project was carried out by the National Environmental Research Institute, the Geological Survey of Denmark and Greenland, the Greenland Institute of Natural Resources, the University of Copenhagen (Institute of Geography), the Greenland National Museum, The Greenland Secretariat of the Danish National Museum, Danish Meteorological Institute, AXYS Environmental Consulting Ltd. and SL Ross Environmental Research Ltd. The Danish Energy Agency funded the Atlas.

20.5 REFERENCES

Anker-Nilssen, T., 1987, *Metoder til konsekvensanalyser olje/sjøfugl*, (Trondheim: Viltrapport 44, Norsk Institutt For Naturforskning), pp. 114. In Norwegian.

Anker-Nilssen, T., 1994, *Identifikasjon og prioritering av miljøressurser ved akutte oljeutslipp langs norskekysten og på Svalbard*, (Norge: Norsk Institutt for Naturforskning, oppdragsmelding 310). In Norwegian.

Boertmann, D., Mosbech, A., Falk, K., and Kampp K., 1996, *Seabird Colonies in Western Greenland (60^o - 79^o 30´ N. lat.);* NERI Technical Report no. 170, (National Environmental Research Institute), pp. 148.

Dickins, D., Bjerkelund, I., Vonk, P., Potter, S., Finley, K., Stephen, R., Holdsworth, C., Reimer, D., Godon, A., Duval, W., Buist, I., and Sekerak, A., 1990, *Lancaster Sound region. A coastal atlas for environmental protection*, (Vancouver: D.F. Dickins Associates Ltd.)

Halls, J., Michel, J., Zengel, S., Dahlin, J., and Petersen, J., 1997, *Environmental Index Sensitivity Guidelines, version 2;* NOAA Technical Memorandum NOS ORCA 115, (Seattle:NOAA), pp. 79.

Huntington, H., 1998, Observations on the utility of the semidirective interview for documenting traditional ecological knowledge; *Arctic*, **51**, pp. 237-242.

Moe, K.A., Skeie, G.M., Brude, O.W., Lovas, C.M., Nedrebo, M., and Weslawski, J.M., 2000, The Svalbard intertidal zone: A concept for the use of GIS in applied oil sensitivity, vulnerability and impact analyses. *Spill Science & Technology Bulletin* **6** (2), pp. 187-206.

Mosbech, A., 2000, *Predicting Impacts of Oil Spills - Can Ecological Science Cope? A Case Study Concerning Birds in Environmental Impact Assessments*, University of Roskilde, National Environmental Research Institute, Department of Arctic Environment, pp. 129. Available online at URL: http://www.dmu.dk/1_viden/2_Publikationer/3_Ovrige/rapporter/PHD_AndersMosbech.pdf

Mosbech, A., Anthonsen, K.L., Blyth, A., Boertmann, D., Buch, E., Cake, D., Grøndahl, L., Hansen, K.Q., Kapel, H., Nielsen, S., Nielsen, N., Von Platen, F., Potter, S., and Rasch, M., 2000, *Environmental Oil Spill Sensitivity Atlas for the West Greenland Coastal Zone*, (Denmark: The Danish Energy Agency, Ministry of Environment and Energy) pp. 281 + appendix pp. 153. Available on CD-ROM, Print (limited distribution) and on the Internet http://Environmental-Atlas.dmu.dk

Mosbech, A., Dietz, R., Boertmann D., and Johansen, P., 1996, *Oil Exploration in the Fylla Area, An Initial Assessment of Potential Environmental Impacts*,

(Denmark: National Environmental Research Institute), NERI Technical Report no. 156, pp. 92 http://technical-reports.dmu.dk

Nansingh, P. and Jurawan, S., 1999, Environmental sensitivity of a tropical coastline (Trinidad, West Indies) to oil spills. *Spill Science & Technology Bulletin* **5**(2), pp. 161-172.

Nielsen, S.S., Mosbech, A., and Hinkler, J., 2000, *Fiskeriressourcer på det lave vand i Vestgrønland.- En interviewundersøgelse om forekomsten af lodde, stenbider og ørred; Danmarks Miljøundersøgelser*. Arbejdsrapport fra DMU nr. 118.

Table 20.1. Criteria for inclusion of seabird breeding colonies.

Species	Criterion	No. of colonies meeting the criterion	No. of colonies included because other species meet their criterion (mixed colonies)
Single species			
Northern fulmar	all colonies	3	-
Great cormorant	all colonies	34	-
Common eider	colonies with ≥5 individuals	31	42
Iceland gull	colonies with ≥500 individuals	10	32
Glaucous gull	colonies with ≥500 individuals	1	44
Unsp. glaucous/ Iceland gull	colonies with ≥500 individuals	1	3
Black-legged kittiwake	colonies with ≥50 individuals	38	7
Arctic tern	colonies with ≥30 individuals	23	7
Common guillemot	all colonies	5	-
Thick-billed murre	all colonies	5	-
Razorbill	colonies with ≥5 individuals	61	12
Black guillemot	colonies with ≥250 individuals	5	88
Dovekie	all colonies	1	-
Atlantic puffin	all colonies	17	-
Species combinations in mixed colonies (not meeting single species criterion)			
Razorbill and common eider	all colonies	1	-
Common eider and Arctic tern	all colonies	1	-

Table 20.2. Relative abundance of species/species groups in seabird breeding colonies. Add one point to relative abundance if colony/shoreline segment hold three or more alcid species and otherwise only reach a relative abundance of 3 or less. Add one point if colony/shoreline segment hold four or more gull and kittiwake species and otherwise only reach a relative abundance of 2 or less.

Species or species group:	**Alcids** (4 species with different occurrence in the Atlas region)				**Seaducks**	**Gulls**	**Kittiwake**	**Arctic tern**	**Tubenoses**	**Cormorants**	**Relative abundance**
Species in group:	Black guillemot	Razor-bill	Atlantic puffin	Thick-billed murre	Common eider	3 species			Northern Fulmar	Great Cormorant	
Number of individuals in colony:	1-100	1-20	1-5	1-10	1-50	1-200	1-100	1-50	1-200	1-20	1
	101-200	21-50	6-10	11-50	51-100	201-400	101-1000	51-200	201-1000	21-50	2
	201-5000	51-100	11-20	51-100	101-200	401-1000	1001-2000	201-1000	1001-2000	51-100	3
	501-1000	101-200	21-50	101-200	201-500	1001-2000	2001-10,000	1001-2000	2001-10,000	101-200	4
	>1000	>200	>50	>200	>500	>2000	>10,000	>2000	>10,000	>200	5

CHAPTER TWENTY-ONE

Environment Canada's Atlantic Sensitivity Mapping Program

André Laflamme, Stéphane R. Leblanc, and Roger J. Percy

21.1 INTRODUCTION

Canada's Atlantic Region along with other regions across the country have focused on providing consistent and standardized applications related to coastal mapping and data integration/generation during a drill or spill incident. This consistency is crucial if personnel are to be brought in from different regions, as they are immediately familiar with the process and terminology. In an effort to protect the environment and mitigate potential impacts, Environment Canada's Atlantic Sensitivity Mapping Program (ASMP) was designed to provide this level of support to environmental responders. The ASMP has become a very powerful tool providing a consistent terminology through the entire range of pre-spill planning, preparedness, and real-time response activities.

This paper will describe the scope, objectives, and current status of this mapping initiative and highlight recent developments in combining the full range of activities from data generation and decision development to the generation of sensitivity mapping. The desktop mapping application provides an easy-to-use approach to the manipulation, display and output of a wide range of technical and supporting data and information stored in various databases. In the development of its mapping program, Environment Canada relied on crucial partnerships with organizations willing to share data and expertise. Response managers and environmental responders now have access to sensitive resource information that normally would be difficult to collate and present in a map form under the pressures of a spill response.

The objective of developing and maintaining the best possible sensitivity mapping system is to provide planners and managers with the full range of information they require as part of pre-spill activities as well as resource protection recommendations at the time of a spill. The data and information are based on consistent sets of terms and definitions that describe the shore-zone character, the objectives and strategies for a specific response, and the methods by which those

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

objectives may be achieved. These data are linked with other resource information in a GIS based system.

Standard or accepted terms, definitions, and shoreline segmentation procedures are already in place for describing the shore-zone character and shore-zone oiling conditions. In this program, a set of standardized objectives and strategy statements have been developed that can be entered easily into a database; these provide a better level of consistency than do phrases or sentences constructed by different recorders or evaluators. The suggested protection and treatment objectives and strategies are intended for consideration by the spill response management team. The actual type and volume of spilled oil, plus local environmental conditions and local priorities would be brought to bear on the decision process at the time of a spill. The suggested objectives and strategies provide a starting point and a framework for decision makers and planning and operations managers to discuss objectives and priorities. The concept of management by objectives provides a framework for decision-makers to set the goals of an operation at both the regional and a segment-by-segment level (Percy, LeBlanc, Owens, 1997).

The pre-spill database is integrated with the actual Sensitivity Mapping Program which is capable of displaying natural, cultural and man-made features vulnerable to oil spills. The computerized mapping system facilitates quick access and management of multiple data sets. A user-friendly interface allows queries and statistical analysis of data and display of graphical outputs. The system provides a tool for both planning and response; information can be accessed or modified using a laptop computer and real-time spill information or trajectory model outputs can be incorporated.

21.2 PARTNERSHIP

Following the Exxon Valdez spill in Alaska, the government of Canada realized the need for a system to provide access to sensitivity data for planning and response purposes. The Green Plan provided initial funding to develop, create and maintain a sensitivity mapping system to support environmental responders during marine spill incidents. Environment Canada was tasked to lead this project and was assigned the responsibility to gather and manage appropriate data sets from various agencies. Because of its mandate, Environment Canada has environmental emergency officers on duty on a 24/7 basis. Therefore, information must be quickly accessible in order to mitigate a potential impact on marine and coastal resources.

Without ready access to environmental data, the integrity of coastal and marine resources can be compromised during a spill incident if immediate action is not taken to protect them. Partnerships are crucial in order for environmental emergency responders to locate and identify sensitive resources at a spill site, especially within the first few hours/days of an incident. One of Environment Canada's foci is to approach and involve other federal, provincial, municipal agencies, private industry, local communities, etc. to make environmental data accessible. Most of the organizations involved during a spill incident are part of the Regional Environmental Emergencies Team know as REET.

REET has two main operating roles: *planning* and *response*. As part of the planning function, "team" members meet once a year to exchange scientific and technical information on such matters as contingency planning and spill response techniques. During this time, REET members also update and review their respective roles in any emergency response. In its response role, REET operates as a team of experts, advising the On-Scene-Commander, or OSC, in emergency situations. Chaired by Environment Canada, it is composed of scientific advisors, private contractors, community groups, etc.

21.3 GEOGRAPHIC APPLICATION

The coastal area covered by the Atlantic Region Sensitivity Mapping Program encompasses four provinces: these are New Brunswick, Nova Scotia, Prince Edward Island, and Newfoundland/Labrador. Approximately 12,500 unique shoreline segments covering more than 40,000 kilometres of coastline have been identified in Atlantic Canada. Labrador is the only area not presently covered by the shoreline classification; however growing interests and activities in this area will likely require the completion of the pre-spill database in the near future. Portions of the province of Quebec have been included in the mapping system: these are Chaleur Bay (on the north shore of the St. Lawrence River) and the Magdalene Islands, since these areas would likely be impacted by spills in the Atlantic Region. Despite the coastal applications of the mapping system, it also has the flexibility to cover the inland part of the Atlantic Provinces. Environmental data has been collected for the Maine and New Brunswick border. The Atlantic Region Sensitivity Mapping Program in conjunction with the Maine Department of Environmental Protection have agreed to exchange cross border information on coastal areas which can be used for planning and response during marine incidents that could impact both countries. As the information becomes available for inland areas, the mapping system will integrate the information in a format that is compatible with the existing data sets. Most of this information comes from federal, provincial, and municipal government, as well as local knowledge.

21.4 SHORELINE CLASSIFICATION AND PRE-SPILL DATABASE

The objective of the pre-spill database is to collate data and information that would be required and used by the spill response management team in the development of planning, priority and operations decisions. This database plays a fundamental role in the definition of resource protection priorities, and constitutes an introduction to the Shoreline Clean-up and Assessment Techniques (SCAT) process.

The database development procedure involves an initial segmentation of the shoreline followed by the creation of data templates for each segment. This process involves the use of various tools such as low-altitude videotape survey data, aerial photography, and pre-existing mapping materials to define sections of shoreline that have a uniform along-shore character. In Atlantic Canada, each segment has a unique two-letter prefix code followed by a sequential number

(Figure 21.1). The two-letter prefix is unique to one coastal area in Atlantic Canada which makes each code different (e.g.: Halifax Harbour has the following segment codes: HX-01 to HX-75).

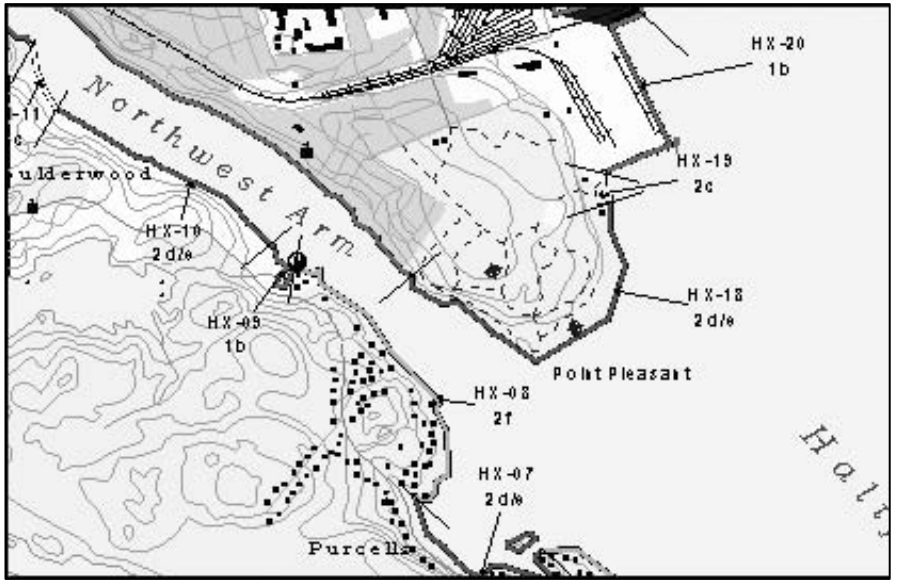


Figure 21.1 An example of the shoreline segmentation of the Northwest Arm area near Halifax, NS.

The description of the shore zone and the development of appropriate response strategies are presented in a systematic format based on four distinct templates: shore zone character, shoreline protection, shoreline treatment, and summary of response and requirements.

These templates contain a total of 143 different attributes which are unique for each shoreline segment. The Shore Zone Character template describes information such as shoreline material/type, nearshore environment, longshore current, oil traps and potential behaviour, resources at risk, etc. (Owens & Dewis, 1995). The shoreline material/type is further subdivided into five distinct categories: lower inter-tidal material, lower inter-tidal form, shoreline type (area located between the high and low tide mark), backshore material, and backshore form (Table 21.1)

The Shoreline Protection and Treatment/Cleanup Templates offer a variety of shoreline data including treatment and protection methods, objectives, strategies, and operational considerations. The last template is known as the Summary of Response Requirements. It is a summary of the protection and treatment templates and includes a response priority code (L = low, M = medium, H = high, VH = very high). The response priority code is defined based on the information available at the time of collection of the pre-spill database. Although it is a starting point in defining priorities, Environment Canada's Sensitivity Mapping Program is now in the process of incorporating other data sets in order to define a response priority

code which will better reflect the actual resource inventory for a specific shoreline segment.

Table 21.1 Shoreline material/type found in Atlantic Canada

Lower ITZ Material	**Lower ITZ Form**	**Backshore Form**
anthropogenic concrete	anthropomorphic breakwater	anthropomorphic breakwater
anthropogenic wood	anthropomorphic pier/jetty	anthropomorphic bridge
anthropogenic riprap	anthropomorphic pilings	anthropomorphic causeway
bedrock resistant	anthropomorphic seawall	anthropomorphic road
bedrock unresistant	anthropomorphic bridge	anthropomorphic dyke
boulder	anthropomorphic wharf	anthropomorphic pier/jetty
cobble	beach	anthropomorphic wharf
mixed coarse with sand	cliff	anthropomorphic railway
marsh grass	dune	anthropomorphic seawall
Mud	platform	barrier beach
pebble	salt marsh	beach
Sand	tidal flat	cliff
	delta	dune
Backshore Material	low-islets	flat
anthropogenic asphalt		peat bog
anthropogenic concrete	**Shoreline Type**	platform
anthropogenic riprap	bedrock	salt marsh
anthropogenic wood	boulder beach	spit
bedrock resistant	man-made solid	wetland bog
bedrock unresistant	mixed sand-gravel beach	delta
mixed coarse with sand	mud tidal flat	
mixed coarse-no sand	pebble-cobble beach	
marsh grass	salt marsh	
peat	sand beach	
sand	sand tidal flat	

These templates use a knowledge-based concept, as data and recommendations are entered, in part, from knowledge and experience rather than from an objective analysis. The templates are described in detail by Owens and Dewis (1995). The shoreline protection and treatment or cleanup techniques that are recommended for each segment are derived from the Environment Canada Field Guide for the Protection and Cleanup of Oiled Shorelines (Owens, 1996).

The description of the physical character of the shore zone for each segment is broken down into the lower intertidal zone, the upper intertidal zone (which corresponds to the nine accepted standard shoreline types (Figure 21.2)) and backshore coastal character.

The shoreline type is a description of that area of the shore zone where oil is most likely to be stranded and the coastal character is described since this is the area in which backshore operations will stage and deploy resources. The description also includes identification of features that are likely to affect the behaviour of persistent oil, such as alongshore traps, potential boulder or riprap

reservoirs, etc. Figure 21.3 gives an example of the Shore Zone Character template for a specific shoreline segment in the Atlantic Region.

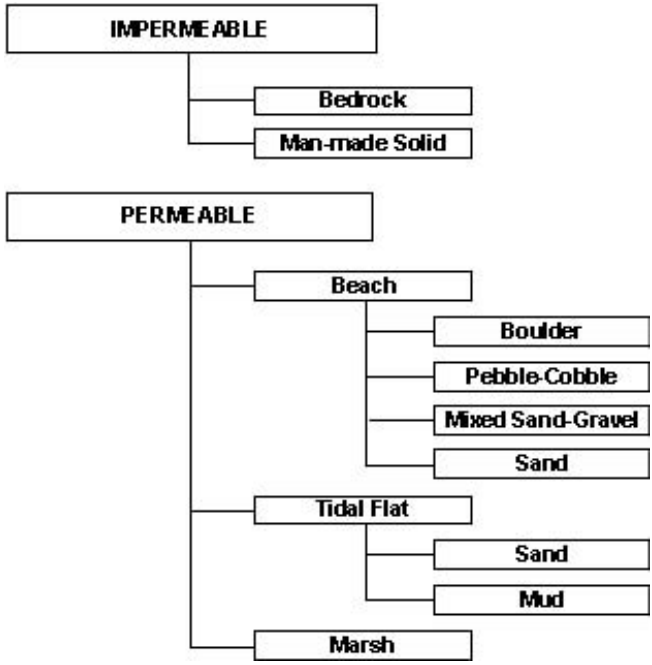
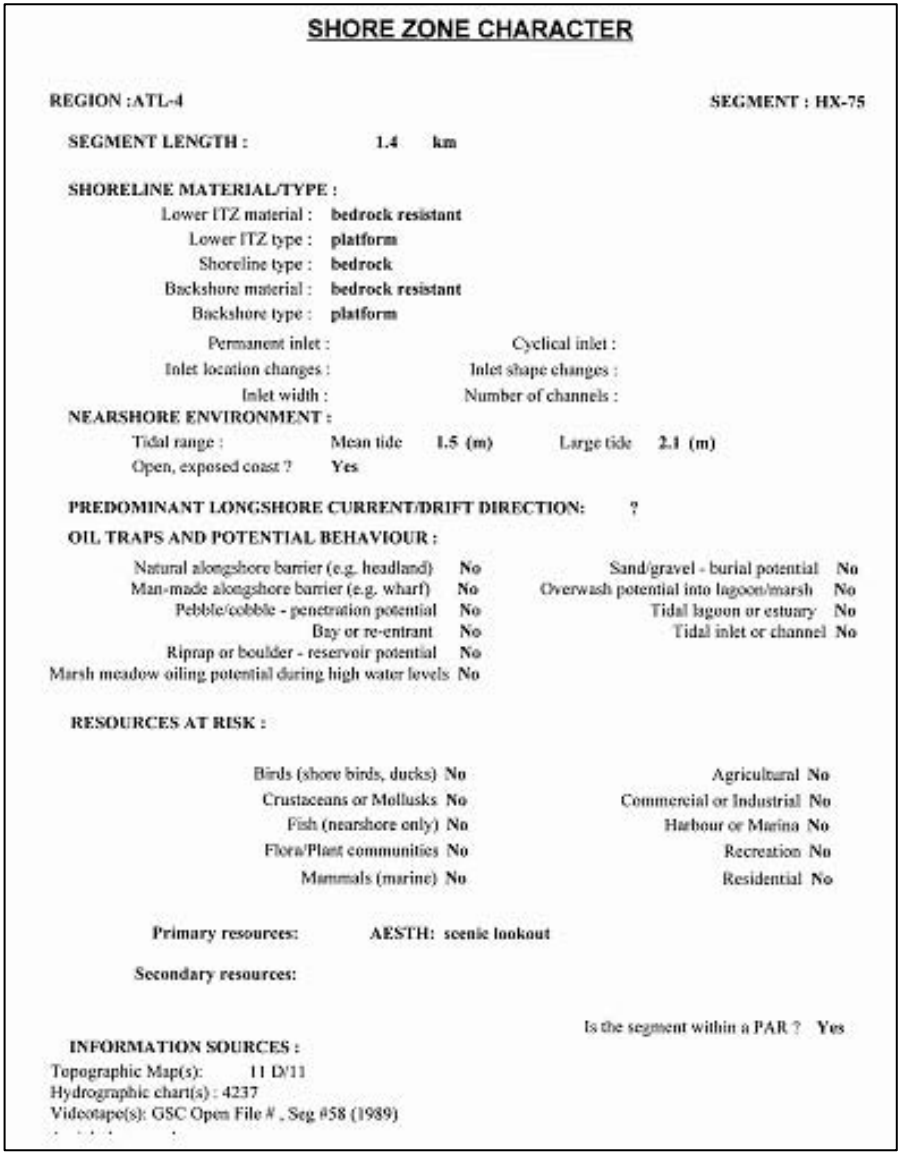


Figure 21.2 Nine standard shoreline types used to describe shore character.

Other areas outside Canada where the same shoreline classification approach has been applied include Hawaii (Honolulu-Waikiki), Russia (Sakhalin Island) and Alaska (Port Valdez). A number of countries around the world have shown interests in the Atlantic Region Sensitivity Mapping Program, including Bangladesh, Brazil, Spain, Israel, Chile, and France.



SHORE ZONE CHARACTER

REGION :ATL-4 **SEGMENT : HX-75**

SEGMENT LENGTH : 1.4 km

SHORELINE MATERIAL/TYPE :

Lower ITZ material : **bedrock resistant**
Lower ITZ type : **platform**
Shoreline type : **bedrock**
Backshore material : **bedrock resistant**
Backshore type : **platform**

Permanent inlet : Cyclical inlet :
Inlet location changes : Inlet shape changes :
Inlet width : Number of channels :

NEARSHORE ENVIRONMENT :

Tidal range : Mean tide 1.5 (m) Large tide 2.1 (m)
Open, exposed coast ? Yes

PREDOMINANT LONGSHORE CURRENT/DRIFT DIRECTION: ?

OIL TRAPS AND POTENTIAL BEHAVIOUR :

Natural alongshore barrier (e.g. headland)	No	Sand/gravel - burial potential	No
Man-made alongshore barrier (e.g. wharf)	No	Overwash potential into lagoon/marsh	No
Pebble/cobble - penetration potential	No	Tidal lagoon or estuary	No
Bay or re-entrant	No	Tidal inlet or channel	No
Riprap or boulder - reservoir potential	No		
Marsh meadow oiling potential during high water levels	No		

RESOURCES AT RISK :

Birds (shore birds, ducks)	No	Agricultural	No
Crustaceans or Mollusks	No	Commercial or Industrial	No
Fish (nearshore only)	No	Harbour or Marina	No
Flora/Plant communities	No	Recreation	No
Mammals (marine)	No	Residential	No

Primary resources: **AESTH: scenic lookout**

Secondary resources:

Is the segment within a PAR ? **Yes**

INFORMATION SOURCES :

Topographic Map(s): 11 D/11
Hydrographic chart(s) : 4237
Videotape(s): GSC Open File # , Seg #58 (1989)

Figure 21.3 The shore zone character template is one of the four available in the pre-spill database. More than 143 attributes are available for each shoreline segment.

21.5 SENSITIVITY MAPPING SYSTEM

All computerized mapping systems require base map layers. Over the past years, Environment Canada purchased National Topographic Data Base (NTDB) digital maps from Natural Resources Canada - Centre for Topographic Information in Sherbrooke (Quebec). Three different scale were purchased: 1:50,000, 1:250,000 and 1:1,000,000. All three scales are used to represent specific information according to whether large- or small-scale maps are required. Once the base maps were visualised to reflect as much as possible the symbols and colours of the hard-copy versions, the digital maps were grouped to form one big region (Figure 21.4).

Figure 21.4 Atlantic Canada Sensitivity Mapping Sub-Regions (not shown are: Labrador).

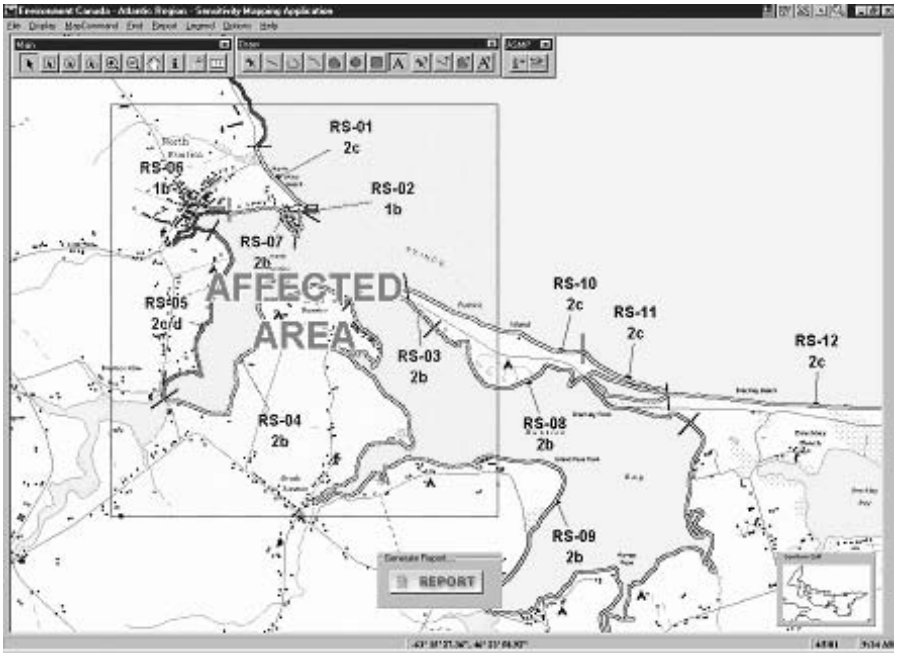
The shoreline classification and the pre-spill database constitute the most important components in the mapping system. The classification describes the physical aspects of the shoreline and provides useful information on protection and clean-up methods. In addition, the mapping system allows for the display of various databases such as birds, fish, shellfish, aquaculture sites, parks, archaeological sites, etc. (Table 21.2). With all this information available, the mapping system can provide a detailed report for any given area.

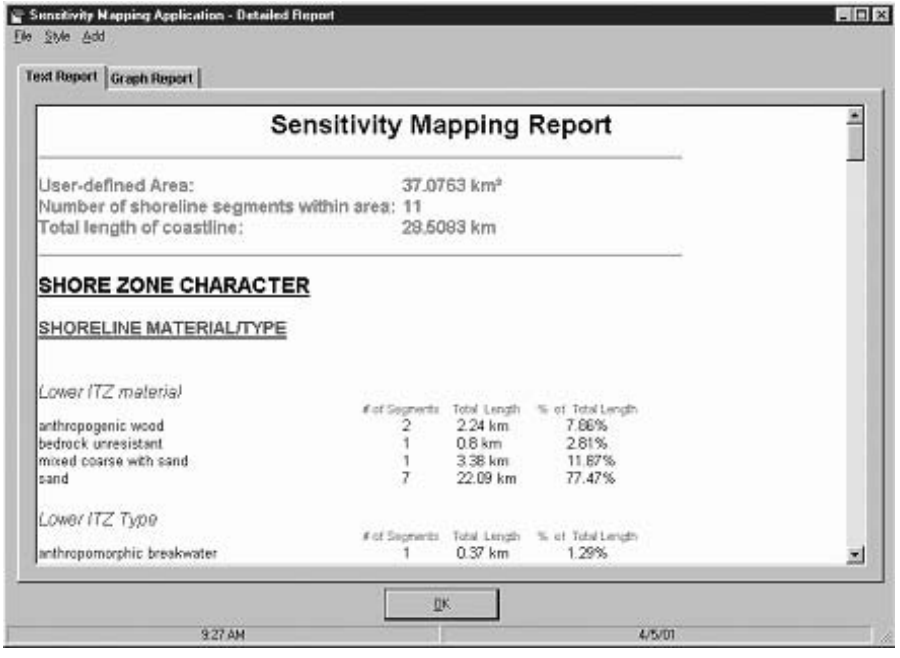
The user has the option of defining a buffer zone which can be used to determine the sensitive resources within an area; to calculate features such as length of shoreline or area affected; or to display data in a graphical form (Figure 21.5). This last example contains shoreline lengths within a polygon for shoreline type, backshore type, biological resources, and human use resources. The information can be displayed in the form of bar graphs or pie charts, and a detailed report on the affected resources may also be generated. The report contains

information on various species or human-use resources, their sensitivity to oil spills and their seasonal vulnerability. A complete database of photographs related to sensitive birds, fish, shellfish, vegetation, and human use structures such as aerial photographs of small craft harbour can also be displayed. In addition to the Natural Resources digital base maps, the system can also display information from digital hydrographic charts or digital elevation models (Figure 21.6).

Table 21.2 Type of information and source (* - Available only during real incidents)

Database	Source
Pre-spill database (MS Access)	Environment Canada
Shoreline Classification	Environment Canada
1:50,000 NTDB maps	Natural Resources Canada
1:250,000 NTDB maps	Natural Resources Canada
Logistical/Operational Data	Response Organization
Fish	Fisheries & Oceans Canada
Shellfish	Fisheries & Oceans Canada
Mammal	Fisheries & Oceans Canada
Amphibian/Reptile	Fisheries & Oceans Canada
Vegetation	Fisheries & Oceans Canada
Small Craft Harbour	Fisheries & Oceans Canada
Whale Sanctuary	Fisheries & Oceans Canada
Fish Weir & Trap	Fisheries & Oceans Canada
Bird	Canadian Wildlife Services
National Wildlife Area	Natural Resources Canada
Wildlife Conservation Area	Natural Resources Canada
Ecological Reserve	Canadian Wildlife Services
Bird Sanctuary	Canadian Wildlife Services
Seabird Colony	Canadian Wildlife Services
Aquaculture Site	Provincial Government
Archaeological Site	Provincial / Federal Government *
Recreational Beach	Provincial Government
Provincial Park	Provincial Government
Municipal Park	Municipal Government
Sewage Outfall	Municipal Government
Water Intake	Municipal Government
Historic Site	Heritage Canada
Sewage Treatment Plant	Municipal Government
Recreational Fishing	Municipal Government
Salmon River	Provincial Government
Fish Processing Plant	Municipal Government
Hydrographic Chart	Canadian Hydrographic Services
Federal Properties	Treasury Boards Canada
National Historic Site	Heritage Canada
Federal Park	Heritage Canada
Aerial Shoreline Video	Geological Survey of Canada
Digital Elevation Model	Natural Resources Canada
Offshore Rigs Location	Environment Canada
Pulp & Paper Mills	Environment Canada


RS-01
2c
RS-06
1b
RS-02
1b
RS-07
2b
AFFECTED
AREA
RS-05
2c/d
RS-10
2c
RS-11
2c
RS-12
2c
RS-03
2b
RS-04
2b
RS-08
2b
RS-09
2b
REPORT


Sensitivity Mapping Application - Detailed Report
File Style Add
Text Report
Graph Report
Sensitivity Mapping Report
User-defined Area: 37.0763 km²
Number of shoreline segments within area: 11
Total length of coastline: 29.5093 km
SHORE ZONE CHARACTER
SHORELINE MATERIAL/TYPE
Lower ITZ material
# of Segments Total Length % of Total Length
anthropogenic wood 2 2.24 km 7.66%
bedrock unresistant 1 0.8 km 2.81%
mixed coarse with sand 1 3.38 km 11.67%
sand 7 22.09 km 77.47%
Lower ITZ Type
# of Segments Total Length % of Total Length
anthropomorphic breakwater 1 0.37 km 1.29%
OK
9:27 AM
4/5/01

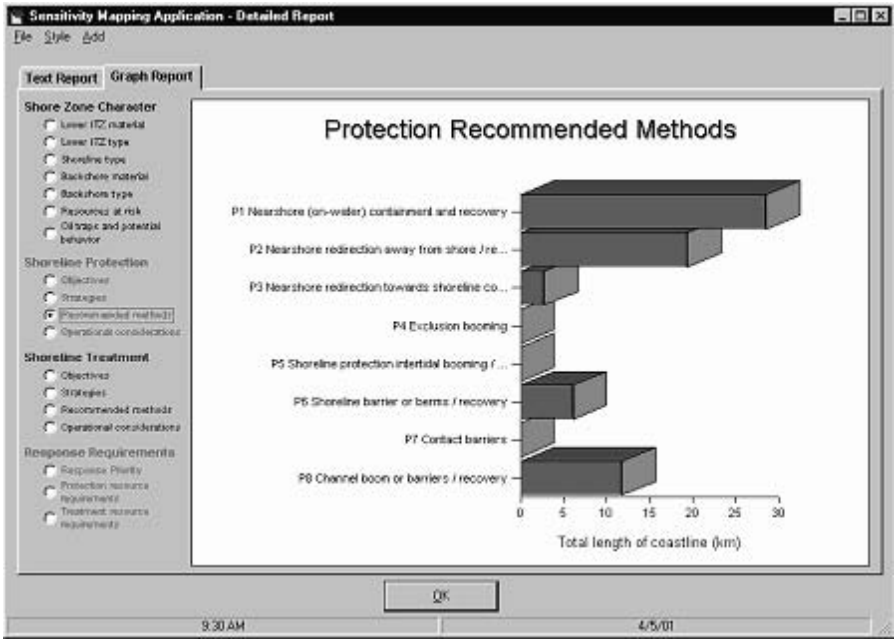


Figure 21.5 a,b,c One of the strengths of the mapping system is the ability to define a buffer zone and extract resource information within that area. The program can generate reports in text and graphical form.

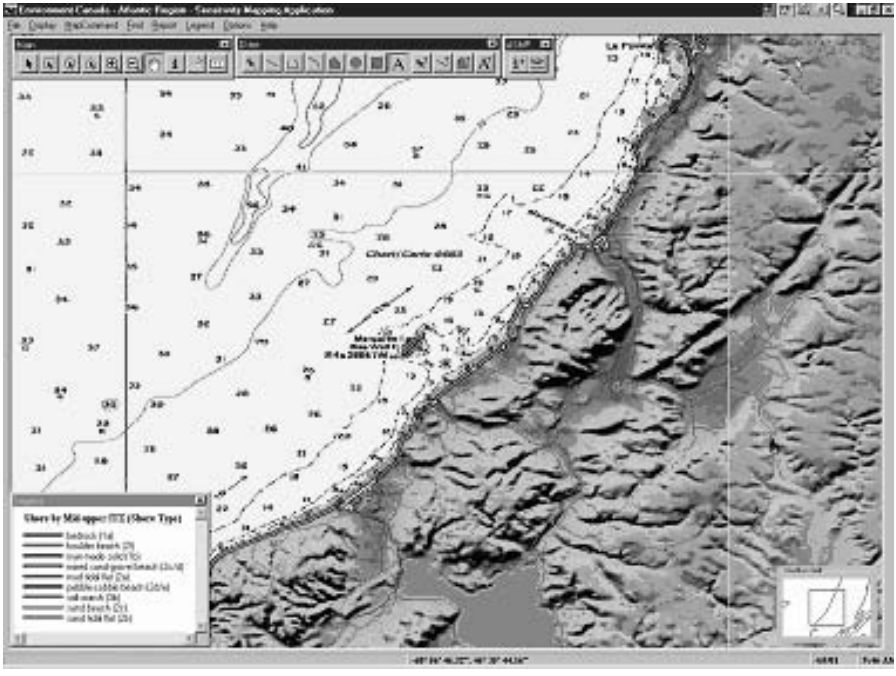


Figure 21.6 Hydrographic chart and digital elevation model used as base layers.

21.6 SUMMARY

The Atlantic Region Sensitivity Mapping Program's main focus is to provide a consistent and standardized application across jurisdictions. This consistency is crucial if personnel are to be brought in from different regions during a drill or spill, as they are instantly familiar with the process and terminology. The same sets of standard terms and conditions are applied to pre-spill database development, sensitivity and pre-spill mapping, SCAT data generation, and the response management decision process. Although the use of standard terms and definitions is almost essential for the description of oiled shorelines, the value of standardization extends into the decision process with regard to shoreline protection and treatment recommendations. This concept is very powerful as it provides a consistent terminology through the entire range of pre-spill planning and response activities.

Sensitivity mapping often has been viewed as an isolated activity in spill planning. Although the identification of resources at risk is a study that can stand alone, the application of that information into the response decision process involves a broader perspective. The integrated approach for spill response in Canada uses sensitivity information as one part of a larger body of relevant information that can and should be used by the decision team in the formulation of an appropriate response. For example, the pre-spill database includes information for each shoreline segment that describes potential oil traps or oil behaviour considerations, possible operational constraints, as well as the resources that are at risk.

The desktop mapping package provides an inexpensive and easy-to-use approach to data management and manipulation. The data can be updated quickly and easily with new information as this becomes available, or during a spill incident, and this new information can be presented immediately in map form.

The Atlantic Region Sensitivity Mapping Program keeps evolving to better serve the needs of environmental responders. All components described in this document will soon be integrated in a web-mapping format which will offer the same features and functionality as the stand alone version. The main advantage of designing a web mapping application resides in its broader accessibility and ease of use. Future developments include the completion of the pre-spill database for Labrador.

21.7 REFERENCES

Owens, E.H., 1996a, *Field Guide for the protection and cleanup of oiled shorelines*, 2nd Edition (Dartmouth: Environment Canada), pp. 201.

Owens, E.H., 1996b, The management of shoreline protection and treatment operation. In *Proceedings 19th Arctic and Marine Oil Spill Programme (AMOP) Technical Seminar* (Edmonton: Environment Canada).

Owens, E.H. and Dewis, W.S., 1995, A pre-spill shoreline protection and shoreline treatment data base for Atlantic Canada. In *Proceedings 18th Arctic and Marine Oil Spill Programme (AMOP) Technical Seminar* (Edmonton: Environment Canada).

Percy, R.J., Leblanc, R.S. and Owens, E.H., 1997, An integrated approach to shoreline mapping for spill response planning in Canada. In *Proceedings 20th Arctic and Marine Oil Spill Programme (AMOP) Technical Seminar* (Edmonton: Environment Canada).

Epilogue: Meeting the Needs of Integrated Coastal Zone Management

Jennifer L. Smith and Darius J. Bartlett

The emergence of Integrated Coastal Zone Management (ICZM) represents a paradigm shift for a range of practitioners who work in the complex, dynamic area where land meets sea. The structure and implementation of geomatics technologies has been strongly affected by this shift, and GIS/RS practitioners have stepped up to meet the information needs of ICZM by creating coastal information systems featuring increased rigour, openness, and useability.

As a result of this forcing, coastal GIS/RS is increasingly differentiating itself from the marine sciences and emerging as a unique discipline. In 1999, Wright wrote: "it may be fair to say that marine applications of GIS have been more in the realm of basic science whereas coastal applications, due in part to the intensity of human activities, have encompassed both basic and applied science, as well as policy and management." This move towards an integrated approach is being realized in parallel with similar shifts in the approach of other actors in the coastal zone, including scientists, managers and planners. The principles of Integrated Coastal Zone Management make explicit what coastal GIS/RS practitioners have known for years: that the coastal zone is a uniquely complex system that requires new and innovative management approaches.

Here we outline the nature of this paradigm shift and the ways in which the authors in this volume have tackled the challenges of applying GIS to the integrated management of the coastal zone.

The principles of Integrated Coastal Zone Management

A thorough discussion of the inception and principles of Integrated Coastal Zone Management can be found in Cicin-Sain (1993) and Cicin-Sain and Knecht (1998). The following draws heavily upon these sources.

ICZM has been described as a process for decision-making: it should be continuous, iterative, and should recognize the contributions of stakeholders and the natural dynamism, both physical and ecological, of the coastal environment.

A primary goal of ICZM is to overcome the compartmentalized approach to managing coastal resources by harmonizing the decisions of diverse jurisdictions and levels of government. In particular ICZM aims to bridge the traditional divide between management of the land and of the water. ICZM, therefore, is also about building institutions that facilitate this integration.

ICZM is founded on principles of sustainable development, recognizing that the coastline is the fount of resources of great value to human communities and

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

that these resources should be managed in ways that conserve their value for future generations. In addition to the extractive resources (such as fishing, hunting, aggregate extraction, and harvesting of algae), ICZM aims to recognize and conserve those resources that provide ecosystem services, such as seagrass beds that provide nursery habitat and mangroves that act as a defense against storm surge and erosion.

The outcome of these principles is a planning methodology that employs an ecosystem approach to management (considering interconnected elements of the ecosystem) and incorporates adaptive management to deal with uncertainty, variability, and change.

The role of GIS/RS in ICZM

The principle function of geomatics technologies in the coastal milieu has traditionally been that of a tool for converting data into information, which is in turn either published as basic science or fed into the balance sheet of the rational decision-making process. With the emergence of ICZM, however, the role of GIS/RS has evolved. In response to the ICZM imperative, coastal geographic information systems are increasingly integrative of scientific and socioeconomic elements, increasingly dynamic, temporal and predictive, and are often the tool that effectively brings stakeholders together.

GIS is often defined as a synergy of hardware, software, data, and people (ESRI 1990). Clearly, the changing role of the technology is accompanied by a changing role for the people who build and operate coastal information systems.

This volume contains a diverse collection of chapters, all of which address the application of GIS/RS and the role of its practitioners in the various facets of an integrated approach to coastal zone management. The authors in this book have begun the process of finding solutions to the challenges presented by ICZM, and are working towards these goals on several fronts.

Contributing to the capacity requirements of ICZM

Cicin-Sain and Knecht (1998) list four areas of critical capacity required for ICZM. In defining "technical capacity" the authors make direct reference to the role of coastal databases and information systems in gathering information on coastal ecosystems and processes, on the patterns of human use that occur in the coastal zone, and on the ongoing effectiveness of management activities. But GIS/RS also has the potential to contribute to the other areas of required capacity outlined by Cicin-Sain and Knecht: authors in this volume show how coastal spatial databases can assist in the formulation, implementation and enforcement elements of the *legal and administrative capacity* required for ICZM. For example, Both Pan and Lindsey *et al.* outline the use of GIS to fulfil legistlative imperatives; the TIS developed by Bourcier has the potential to provoke harmonisation of diverse jurisdictions and regulatory regimes. Some authors have outlined their contribution to the *human resources capacity* of coastal management initiatives, including bringing together multidisciplinary teams and fostering public awareness of and involvement in coastal issues. To this end, Bartlett and Sudarshana examine issues surrounding technology transfer; the standardised approach to risk mapping

developed by Laflamme *et al*. facilitates transfer of workers from one Canadian province to another in response to environmental emergencies; Nwilo describes programs that have facilitated exchanges and training for coastal zone managers in Africa; and many authors in this volume have found avenues for disseminating and interpreting their work to interested stakeholders. And, although GIS are typically expensive to implement, they can sometimes bring economies and recognition to coastal management initiatives, thus enhancing the *financial capacity* of the project in the long-term. Initiatives for advancing geomatics in the coastal zone, like those of Webster *et al*. and Nwilo, have attracted funding with spinoffs for coastal communities.

Integrating data from diverse sources

If the worth of a GIS is determined by how faithfully it represents the real world, creating a servicable GIS is challenging indeed when data of sufficient resolution, coverage and quality are difficult to acquire and unite for reasons of protectionism, conflict and variable format or standards. The ICZM imperative of integration across jurisdictions and institutions makes this issue doubly significant in the coastal zone. Authors in this volume, along with virtually every other coastal GIS practitioner, face this challenge daily and are working to find strategies that yield information systems that meet the needs of ICZM. Bourcier, Laflamme *et al*. and Mosbech *et al*., among many others in this book, have addressed the challenge of collecting and harmonizing data within a distributed, fractured, multi-agency and multi-jurisdictional region (the Seine estuary, the Eastern Canadian coastal zone and the Greenlandic coastal zone, respectively). Gomm considers issues specific to data integration across the land-sea divide, and Gourmelon writes about the process of coordinating a large number of institutions in the creation of a large-scale GIS.

Sometimes the greatest challenge to harmonising datasets is finding them. Longhorn describes the development of Coastal Spatial Data Infrastructures which, at their best, facilitate the formulation of standards, the sharing of data, and, ultimately, a more collaborative, open and connected network of coastal GIS applications.

Even in areas where standards and data-sharing protocols are in force, the increasingly rich and diverse database of many coastal regions leaves to GIS practitioners the task of determining the best available representation of coastal phenomena, as in Li *et al.*, who compared different data sources to arrive at the most accurate depiction of the coastline or Hwang and Ku, who outlined methods for choosing remotely sensed data at a resolution that best meets the needs of a coastal modelling excercise.

The work of all of these authors shows that geography can provide the scale and scope that is needed to unite diverse interest groups. Issues of data integration appear to be casting GIS developers as a uniting force and GIS as a nexus in the project of ICZM.

Adapting tools and technologies to serve the stakeholders of ICZM

We are now faced with the task of expressing the methods and results of what was, until recently, a specialized and technical discipline to a diverse group of scientists, managers, regulators, legislators, resource users and members of the public. These people are increasingly not just data *consumers* but also data *users*, who both demand and deserve to peer inside the black box of GIS technology and GIS-based decision support. GIS practitioners in this volume have responded by creating innovative visualisations, interfaces and forums for participation in the course of applying GIS to the coastal zone.

Involving stakeholders in both the formulation and the output of coastal GIS goes beyond the technical and presents intercultural and interpretive challenges. Bartlett and Sudarshana touch on the pervasiveness of Western metaphors in coastal GIS and the need to consider the cultural background of the stakeholders or collaborators. On Alaska's Pribilof Islands, Lindsay *et al.* collaborated across jurisdictions and cultures to construct a GIS that represents archeological features, important traditional resources, local placenames and the legacy of destructive industries. In another Arctic environment, Mosbech *et al.* gathered traditional knowledge to create an oil spill sensitivity database that considers resources of importance to local people. Both of these projects required that the researchers constructing the GIS consider the needs, concerns, and sensitivities of those people who will ultimately be affected by decisions based upon the outcomes of the exercise. Jude describes an exercise in visualizing the outcome of a coastal development: he applies GIS and RS in a way that makes the effects of change real and explicit to a non-technical audience.

Broad new user groups are now working directly with the data and software that has traditionally been the domain of experts. The Internet has excercised its all-pervasive influence here: Laflamme *et al.*, Green and King and Macharia describe web applications that bring GIS to a broader audience when and where information is needed. Longhorn discusses Coastal Spatial Data Infrastructures, that aid in making data accessible to a greater cross-section of users and in supporting the wider popularization of desktop GIS. Planning for consultation, infrastructures and dissemination is now seen as a critical phase in any coastal GIS initiative, and coastal GIS practitioners must keep the needs of stakeholders and future users in mind.

Improving tools and technologies

Because the land-sea interface is a complex ecotone, the scale of biophysically distinct areas in the coastal zone often necessitates a spatial resolution that has traditionally been difficult or expensive to acquire. GIS/RS practitioners working in the coastal zone were until recently either limited to a few data sources (such as analogue photogrammetry or bathymetric and topographic mapping by traditional survey and sounding means), or required to adapt the scale and resolution of their work to the quality of readily available remotely sensed data, i.e. satellite data at a 30m or greater resolution, often with poor depth penetration, particularly for turbid temperate waters. The challenge of acquiring high-quality data has been somewhat ameliorated by the increasing availability of quick-deploying hyperspectral sensors like CASI (Compact Airborne Spectrographic Imager) and

active laser sensors like LiDAR (Light Detection and Ranging) or active satellite sensors with specialised uses in the marine environment (such as RadarSat). Both Pan and Webster document their experiences and the technical challenges they faced in acquiring and applying data from these sensors, and both authors make explicit the great advances possible with such high (spatial and/or spectral) resolution data. Of particular importance to ICZM is the value of these fast-deploying sensors for monitoring change and perhaps even emergencies in the dynamic and heterogeneous coastal zone. High resolution data is one more piece in the puzzle of building rigorous, accurate and precise models of coastal zone processes.

Adapting the technology to represent complex dynamic systems

Tracking and analyzing spatio-temporal variation is a problem that has received much attention from geographers and GIS practitioners. In the coastal zone, it is compounded by the other challenges described here. Practitioners are faced with the need to deal with (and differentiate between) error and dynamism in an inherently dynamic environment. Bruce provides a detailed outline of the types of error and uncertainty encountered in coastal zone GIS and the various methods that have been developed to help practitioners quantify and work with this uncertainty. Li *et al*. describe methods for defining 'the coastline' in such a way that the variability of this elusive entity is accounted for. Schupp *et al*. have measured historical shoreline change in order to learn about future change. Webster *et al.* describe how dynamic coastal processes at different time scales (extreme tidal range, seasonal variability, and sea level rise) entered into the decisions that were made in the process of acquiring high resolution imagery and elevation data in a coastal area, then uses these data to plan for future changes. If change is the currency of ICZM, then developing data structures and modes of analysis that integrate dynamism into coastal GIS is critical.

Modelling for decision support

Applying GIS/RS in ICZM presents the challenge of using new technology in a real-life context, where models and findings are fed into a legislative and regulatory application and are subjected to not only academic but also governmental, public and vested private sector scrutiny. Coastal GIS practitioners have acknowledged the need to continue to make their models more robust at the same time as they become more complex and comprehensive. Populus *et al*. and Mayerle and Toro, for example, both undertook to model complex hydrodynamic processes with the explicit goal of using these models for decision support. Euan-Avila *et al.* considered land-based activities and their socio-economic drivers in modeling effects on water quality.

Conclusions

Researchers and practitioners are increasingly taking an interdisciplinary, integrative approach to applying GIS/RS in the coastal zone. As the body of work in this area grows, so too does our understanding of coastal processes and the role of humans in perturbing and managing these systems. GIS has much to contribute to Integrated Coastal Management, and GIS practitioners have many more challenges to face in constructing truly integrative coastal spatial information systems.

REFERENCES

Cicin-Sain, B., 1993, Sustainable development and integrated coastal management. *Ocean & Coastal Management*, **21** (1-3), pp.11-43.

Cicin-Sain, B. and Knecht, R. 1998, *Integrated Coastal and Ocean Management: Concepts and Practices*, (Washington and California: Island Press).

ESRI, 1990, *Understanding GIS: The ARC/INFO Method*, (Redlands, CA: Environmental Systems Research Institute).

Wright, D. J., 2000, Down to the Sea in Ships: The Emergence of Marine GIS. In *Marine and Coastal Geographic Information Systems*, edited by D.J. Bartlett and D. J.Wright, (London: Taylor and Francis).

Index

A

Accuracy, 18, 23, 29,51
  classification, 76, 80, 85-91
  as component of SDI, 10
  of data, 129, 195, 220-221, 260
  of DGPS, 168
  of Global Map, 9
  of LiDAR, 43, 199, 203-215
  measures of uncertainty, 54
  of models, 110, 117, 232
  positional, 19-23
  of shorelines, 30-31
  Standards for Nautical Charting, 4
  of thin-layer chromatography, 170
Adobe Acrobat PDF, 276
Aerial photography, 35, 39, 99-101, 167, 199, 220, 243, 247, 283
Aggregate
  data and information, 253, 262 266
  mineral resources, 35, 47, 296
AGI, see Association for Geographic Information (UK)
Agricultural pollution, 1, 4, 126, 138, 182, 190, 229-237, 259
Airborne Laser Hydrography, 53
Alaska, 165-180, 282, 286, 298
Aleut community, 166-179
AML, see Arc Macro Language
Analytical Hierarchy Process (AHP), 232
Angola, 185
Apache Tomcat, 66
Aquifers, 227-237
Archaeological sites, 270-273, 288
Arc Macro Language, 100, 212, 221-224
Association for Geographic Information (UK) GIGateway, 21
Australia, 11, 53, 55-57, 156, 239, 240, 242, 243, 252
Autodesk, 72

B

Bathymetry, 4, 7, 10, 12, 25-26, 29, 30, 32, 47, 64, 114, 116, 126, 144, 168, 183, 210, 214, 247, 271
  interpolated, 128
  of inter-tidal areas, 129
Bay of Fundy, 196-215
Bayesian probability theory, 56
Belize, 230
Benin, 184, 187, 188, 192
Biosphere Reserve, 141-2, 146
British Geological Survey, 18, 99

C

Cameroon, 184-8, 192
Canada, 2, 58, 72-3, 196-215, 281-293

0-41531-972-2/04/$0.00+$1.50
©2004 by CRC Press LLC

Canadian Geospatial Data Infrastructure, see GeoConnections
Canadian National Program of Action (CNPA), 64
Capacity building, 187, 190, 192
CARIS, 65-67
CASI, 36, 38, 44-48, 99, 100-101, 195, 197-199, 240, 244, 298
Centre for Ecology and Hydrology (UK), 101
Clearing houses, 57, 150
Cliff instability, 38, 48
Climate change, 53, 195
Coastal
change detection, 33
defence policies, 96
erosion, see Erosion
realignment, 97-98
Terrain Model, 27, 53
Coastal Zone Management, 5,105, 146, 150, 160, 188, 190, 191, 239, 295
Act (USA), 166
GIS for, 126, 147, 157, 296, 300
in India, 154
in the UK, 95-97
COASTWATCH, 240
Cognition, 57
Common Object Request Broker Architecture (CORBA), 66
Comoros, 191
Compact Airborne Spectrographic Imager, see CASI
Contamination, see pollution
Continuously Operating Reference Stations (CORS), 178
Coral reefs, 54, 230
Côte d'Ivoire, 184-186, 188, 192
Crown Estate (UK), 36
Cultural origins of GIS, 153
Currency,
of data, 18, 20, 23
Currents
Guinea, 183-186, 192-193
longshore, 284
marine, 112, 114, 116, 117, 121, 126-128, 136, 144, 168, 183

D

DBASE, 223
Decision support systems, 72, 109-124, 147, 187, 239, 247, 248
DEMs, see Digital Elevation Model
Department of Fisheries and Oceans (Canada), 65
Differential Global Positioning System (DGPS), see Global Positioning System
Digital Earth, 3
Digital Elevation Model (DEM), 100, 101-2, 167, 168, 196-197, 200, 202-204, 209-210, 211, 213, 233-234, 235, 236, 289
DEMIS, see Dynamic Estuary Management Information System
Digital Line Graph, 28
Digital Shoreline Analysis System, 221, 222
Digital Terrain Model (DTM), 28, 30, 32, 126
Dredging, 36-37, 43, 47, 125, 242
DTM, see Digital Terrain Model
Dykes, 205-206, 210-212, 285
Dynamic Estuary Management Information System, 253

E

ECDIS, see Electronic Chart Display and Information System
Education, 142, 149, 153, 241, 249-250
Electronic Chart Display and Information System (ECDIS), 11
Emergency management personnel, 57, 210, 282
Emergency response, 283

English Channel, 127, 134
English Nature, 239
Enviromapper for Watersheds, 65
Environment Agency (UK), 43, 98, 100-101
  Shoreline Defence Survey, 102
  Shoreline Management System, 100, 103
Environment Canada, 72, 197, 282-292
Environmental Protection Agency
  Nigeria, 184, 188
  USA, 65, 253
Environmental Systems Research Institute
  ArcExplorer, 249, 270
  ArcGIS, 189
  ArcIMS, 72, 250-251, 252
  Arc/Info, 88
  ArcPad, 72, 240
  ArcView, 37, 112, 168, 189, 270
  Shapefile, 20, 65
ENVISAT, 240, 244
ERDAS, 28, 100
ERMapper, 250
Erosion
  aeolian, 167
  coastal, 20, 57, 183, 184, 185, 187, 190, 191, 196, 240, 242, 246, 296
  shoreline, 32, 158, 178, 219, 225-226
  soil, 230, 233
ERS, see European Remote Sensing Satellite
ESRI, see Environmental Systems Research Institute
Estuary, 43, 58, 127, 132-133, 182, 183, 192, 212, 230, 233, 240-254, 258-267, 297
ETeMII, see European Territorial Management Information Infrastructure
Euclidean distances, 80, 156
European Commission, 2, 7, 8
European Habitats Directive, 36, 97
European Institute for Marine Studies, see IUEM
European Remote Sensing Satellite, 244
European Shellfish Hygiene Directive, 130
European Territorial Management Information Infrastructure, 7
European Union, 7, 11, 45
  Birds Directive, 98
  Demonstration Programme on Integrated Coastal Zone Management, 96
  Water Framework Directive, 2, 7-8, 12, 139
EUROSION, 246
eXtensible Markup Language, 11, 21, 72
Exxon Valdez, 282

F

FastMap, 240
Federal Emergency Management Agency (USA), 220
Federal Environmental Protection Agency (NIGERIA), see Environmental Protection Agency
Federal Geographic Data Committee (USA), 2, 3, 21, 168, 208
FGDC, see Federal Geographic Data Committee
Fish, 171, 241, 242, 245, 288
  catch as system of accounting,159
  kill, 229, 243
Fisheries, 1, 58, 65, 135, 177, 182, 184-186, 192, 259, 269
Fishermen, 178, 274
Fishermen's associations, 274
Fishing, 190, 230, 241, 270, 271, 274, 296
Flooding, see Floods
Floods, 138, 187, 259
  defences, 97-100, see also Dykes

as factor in pollution, 126
as hazard, 158
risk, 195-196, 206, 210-214
role of mangroves in attenuation of, 182
Fly-through, 101
Fractal dimension, 78, 89
Fraser River Estuary Management Program (FREMP), 252
Fuzzy sets, 56-57, 232

G

GAP analysis, 144-146
General Packet Radio Service (GPRS), 71
GeoConnections, 5
Geography Markup Language, 21
GESAMP, see Group of Experts on the Scientific Aspects of Marine Pollution
Ghana, 184, 187, 188
GIGateway, 21
Glaciation
Devensian, 39
Global Environment Facility, 183, 184, 190
Global Environmental Trust Fund, 191
Global Land Cover Characteristics Database, 9
Global Map Project, 3, 9
Global Monitoring of Environment and Security (GMES), 240
Global Positioning Systems (GPS), 28, 30, 69, 72, 115, 170, 177, 178, 199, 200-210, 215, 239, 243, 245, 246
differential (DGPS), 42-43, 44, 166, 168, 169, 170, 178-179, 203, 220
Global Spatial Data Infrastructure Association (GSDI), 3, 6, 9
SDI Cookbook, 9
Global System for Mobiles (GSM), 69, 70, 71
Globalisation, 154, 156, 157
GMES, see Global Monitoring of Environment and Security
Governance
marine, 12, 64,73
Graphical User Interface (GUI), 112, 114, 117
Great Lakes Forecasting System, 29
Greenland, 270-277
Group of Experts on the Scientific Aspects of Marine Pollution (GESAMP), 135
GSDI, see Global Spatial Data Infrastructure Association
GSM, see Global System for Mobiles
GTOPO30, 9
GUI, see Graphical User Interface
Guinea Current Large Marine Ecosystem Project, 183-189
Gulf of Mexico, 230

H

Habitat mapping, 38, 44, 47, 48
Handheld Device Markup Language (HDML), 69-70
Handheld Device Transport Protocol (HDTP), 69-70
HandyGIS, 239
Holocene, 36
HTML, see Hypertext Markup Language
Hurricanes, 158, 230
Hydrodynamic models, 110-123,126-139, 146
Hydrographic Office (UK), 17, 22, 25, 247
Hydrography, 1, 7, 9, 12
Airborne Laser, 53
Hypertext Markup Language (HTML), 11, 69-70, 249

I

Ice packs, 168, 173
ICZMap, 18, 19

IDRISI, 232, 237
Ievoli Sun, 134-139
IFREMER, 127, 149
IFSAR, see Interferometric Synthetic Aperture Radar
IKONOS, 27, 28, 30-33, 78, 195, 197, 198, 205, 210, 212, 244
India, 154, 156-7, 159, 161
Inertial Reference System, 42, 201
Information and Communications Technologies, 154
Infrastructure for Spatial Information in Europe (INSPIRE), 2, 7, 8, 10, 12
INSPIRE, see Infrastructure for Spatial Information in Europe
Institut Universitaire Européen du Mer, see IUEM
Integrated Satellite Targeting of Polluters (I-STOP), 73
Interferometric Synthetic Aperture Radar, 167, 197
Intergovernmental Oceanographic Commission (IOC), 3, 9, 12
  International Oceanographic Data and Information Exchange, 3, 9, 11
Intergovernmental Panel on Climate Change (IPCC), 100, 196
International Development Association, 191
International Geographic Data Committee, 27
International Hydrographic Bureau, 11
International Hydrographic Organisation, 11
International Standards Organisation (ISO)
  ISO standard for geospatial metadata, 21, 149
  ISO TC/211, 149
Internet, 6, 11, 57, 64, 65-71, 155, 166, 170, 198, 220, 226, 239, 240, 243, 245, 248-249, 270, 275-276, 298
Internet-based information systems, 250-253
Invasive species, 259
IODE, see International Oceanographic Commission: International Oceanographic Data and Information Exchange
Isarithm method, 78
IUEM, 142-143, 148-149

J

Java, 65-66, 129
Jeffreys-Matusita distance, 79-80, 82, 86, 89-90
Judeo-Christian approach to nature, 156-8, 159

K

Kappa Index, 88
Kenya, 190, 191
Korea, 83

L

La Réunion, 191
Lake Victoria Environmental Management Project, 190
Land cover, 210, 212, 214
  classification of, 83, 86, 87, 91
  mapping of, 191
Landsat
  Landsat-7, 197, 198
  Multispectral Scanner (MSS), 77
  Thematic Mapper ™, 77, 83
Large Marine Ecosystems, 183
LiDAR, 36, 38, 42-44, 47-48, 99, 100, 102, 196-215, 240, 244, 299
Light Detection And Ranging, see LiDAR
Lizardtech
  Content Server, 250
  MrSID, 250

Local knowledge, 247, 269, 274, 277, 283; see also Traditional Ecological Knowledge
Local variance method, 78, 81, 89
Location Based Services, 68-70, 72

M

MaB, see UNESCO Man and Biosphere program
Madagascar, 191
Maine Department of Environmental Protection, 283
MAGICC, 100
Mahalanobis distance, 80-82, 86-87, 89, 90
Managed retreat, 97
Mangroves, 52, 58, 182, 184, 186, 191, 230, 241, 296
MapInfo, 20, 65, 72, 129
MapExtend, 72
MapExtreme, 72
Marinas, 242, 246
Marine Aggregate Resource Management and Planning System (MARMPS), 47-48
Marine Geospatial Data Infrastructure (Canada), 2, 5-6
Marine mammals, 171, 241
Marine Protected Areas, 53, 245
MARS Hydrodynamic Model, 129
  2-D, 127, 128, 131, 136
  3-D, 127
Massachusetts Office of Coastal Zone Management, 220
Mauritius, 191
Mental maps, 157
Metadata, 3, 6, 9, 11, 12, 21, 57, 138, 149-150, 168, 187, 252, 253, 260
  FGDC Shoreline Metadata Profile, 4, 11
  ISO standards, 11, 21, 149
Mexico, 229-237
Microsoft, 68
  Access, 219, 223
  Internet Information Services (IIS), 66
  MapPoint, 72
  Powerpoint, 121
  Visual Studio, 129
  Windows, 72
  Word, 121
Migratory Birds Act (Canada), 73
Mobile devices, 66, 68, 69-70, 72, 73
Modifiable Aerial Unit Problem (MAUP), 78
Model for the Assessment of Greenhouse-gas Induced Climate Change, see MAGICC
Moran's I coefficient, 79, 81-82, 84-85, 90
Mozambique, 190, 191
MrSID, see Lizardtech
Multi-criteria evaluation, 48, 232, 237, 245
Multi-Spectral imagery, 78, 167, see also Landsat MSS

N

National Assembly for Wales, 35, 36, 37, 43, 48, see also Welsh Office
National Imagery and Mapping Agency (USA), 9
National Ocean Service (USA), 220
National Oceanic and Atmospheric Administration (NOAA), 2, 3-5, 27-29, 65, 165-166, 172, 177, 178, 184
  as land management agency, 165-6
  mapping and charting, 28-30, 167-168, 173
  National Marine Fisheries Service (NMFS), 174
  Pribilof Islands Environmental Restoration Project, 166
Natural Environment Research Council (UK), 100

Natural Resources Canada, 200, 288-289
NETCDF, 128, 129, 138
Newlyn Datum, 20, 25
New South Wales Coastal Protection Act (Australia), 57
Niger Delta, 182, 189, 190, 192
  Environmental Survey, 189-190, 192
Nigeria, 184, 185, 186, 187-189, 192
North Sea, 97, 113
NSDI, see Spatial Data Infrastructures
Numerical models, 109-111, 112, 115, 117-122, 126, 146

O

OGC, see Open GIS Consortium
Oil spill, 177, 185, 192, 282, 289, 298
  contingency planning, 147, 270-277, 282-286
OpenGIS Consortium, 11, 21
Open GIS Consortium Europe), 11
Ordnance Survey (Great Britain), 17, 20, 22, 25, 26, 99, 100, 247, 248

P

PC-IDEA, see Permanent Committee for SDI for the Americas
PCGIAP, see Permanent Committee on GIS Infrastructure for Asia and the Pacific
PDA, see Personal Data Assistant
Permanent Committee for SDI for the Americas, 2, 6
Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP), 2, 6, 8
Personal Data Assistant, 72, 239, 243, 245, 247, 250
Pesticides, 230, 231,233
Photogrammetry, see aerial photographs
Place names, 172
Planning
  coastal, 58, 96, 104, 188, 190, 197
  emergency, 177, 270, 238, 282-3
  environmental, 143, 193, 242, 259, 263, 283, 296
  maps for, 170
  participatory, 58, 96, 248-249, 298
  project, 197
  resource management and, 47, 160, 183, 193, 210
  SDI, 2
  town, 210, 212
  transport, 53
PocketGIS, 240, 245, 246, 247
Pollution, 5, 7, 125, 135, 183-185, 187, 190, 191, 229, 265
  agricultural, 125, 229-238, 246-247
  of estuaries, 240, 242, 246-24
  industrial, 184
  noise, 264
  oil, 125, 141, 188
  response, 138, 147

Q

Quality control, 201, 209, 215, 220
Quickbird, 196

R

RADARSAT, 73, 196, 197, 198, 244, 299
Radar, see Radio Detection And Ranging
Radio Detection And Ranging (Radar), 42, see also Synthetic Aperture Radar
Ramsar sites, 271
Rational functions, 28

Reefs
artificial, 245, 246
coral, 34, 229
oyster, 241
Religion, 156-157
Remote sensing, 2, 27, 35-49, 70, 78, 90, 144, 147, 148, 157, 243, 244, 247, 253
Runoff, 131

S

Salt marshes, 78, 83-86, 98, 102, 220, 240, 241, 285
SCAT, see Shoreline Clean-up and Assessment Techniques
Scorecard mapping application, 65
Scottish Natural Heritage, 239
SDI, See Spatial Data Infrastructures
Sea level rise, 53, 100, 182, 196, 212, 299
Sea lions, 165, 170, 173-174, 177
Seabirds, 166, 168, 171, 173, 174, 175, 177, 270, 271, 272-273, 289
Seagrass, 54-55, 56, 296
Seals, 165, 171-174, 177, 271
Search and Rescue, 247-248, see also Emergency response
Secretariat for Eastern African Coastal Area Management (SEACAM), 190-191
Sediment, 35, 48, 53, 131, 133, 144, 147, 183, 229, 247
cells, 95
movement of, 39, 240, 242, 244
Sedimentation
modelling of, 138
rates of, 55-56
Seine Estuary, 258-267
Setback, 97
Seychelles, 190, 191
statement, 190-191
Shellfish, 125, 138, 241, 242, 288, 289
farming, 130-134, 259
fisheries monitoring with GIS, 130-134
Shoreline, 1, 4, 9, 12, 27, 52, 53, 158, 178, 271, 273, 274, 292
change, 53, 64, 170, 220-227, 242, 299
classification, 283-286, 288
Clean-up and Assessment Techniques, 283-286, 292
as component of SDI, 10
data, 3, 11, 99
Defence Survey, 102
management plans, 96-98
Management System, 103
management units, 231
mapping, 27-33, 269-277
Metadata Profile, 4
segmentation, 282, 283
sensitivity, 272, 276, 288
Somalia, 191
Spatial autocorrelation, 79, 81-82, 84-85, 89, 90
Spatial Data Infrastructure, 1-13, 182, 297, 298
SPOT, 77
Stakeholders, 1, 2, 5, 6, 96, 126, 144, 146, 185, 188, 190, 191, 193, 252, 263, 296-299
Standards, 5, 9, 11, 12, 52, 170, 297
cartographic, 53, 220
data, 254
de facto, 20
health, 131
metadata, 11, 149
organisations, 11
spatial data quality, 57,182, 208
technical, 4
STELLA modelling software, 249
Sustainable development, 64, 65, 154, 184, 185, 188, 190, 193, 257, 264, 266, 296

T

Tanzania, 155, 190, 191

Technology transfer, 154-155, 161, 296
Terraffectors, 102-103
TINs, see Triangular Irregular Networks
TNT maps, 233, 237
Togo, 184, 187, 188, 192
Topology, 100, 222
Tourism, 63, 142, 157, 230, 239, 241, 242, 246
Traditional ecological knowledge, 177-178, 270
Training, 104, 149, 179, 180, 187, 249-250, 297
  image processing, 79, 82, 86, 90
Transport Canada, 72
Triangular Irregular Networks (TINs), 101, 202, 209
Triangular prism method, 78
Trimble, 166-167, 168, 201
Tsunami, 158

U

Uganda, 190
Uncertainty, 51-59, 100, 296, 299, see also Error
UNESCO Man and Biosphere program, 142, 143
United Kingdom, 36, 39, 48, 104, 181, 187
  coastal management in, 95-96
United Kingdom Hydrographic Office (UKHO), 17, 22, 25, 247
United Nations Conference on Environment and Development, 190
United Nations Education and Scientific Cultural Organisation (UNESCO), 11, 142, 143, 184, 190
United Nations Industrial Development Organisation (UNIDO), 183, 184, 187
United Nations Regional Cartographic Conferences, 2, 6
United States, 2, 3-5, 27, 65, 146, 169, 174, 181, 242
United States Geological Survey (USGS), 172
Universal Mobile Telephone System, 69, 71
Useability, 296
User interface, 112, 249, 260, see also Graphical User Interface

V

Variogram method, 78
Virtual Reality, 40, 96, 97, 103-105, 245
Virtual Reality GIS (VRGIS), 96-97, 104
Virtual Reality Modelling Language (VRML), 101, 103
Visualisation, 40, 71, 96-105, 115, 126, 146, 244, 249, 253, 265, 298
VMAP, 9
VRML, see Virtual Reality Modelling Language

W

Wales, 35-49, 239
WAP, see Wireless Application Protocol
Water
  level, 27-29, 30, 32, 53, 64, 112, 114, 116, 117, 119, 120, 126, 128, 195, 210-212
  quality, 64, 126, 130, 132, 134, 138, 230, 231, 232, 236, 237, 247, 299
Watersheds, 65, 230, 231, 232, 237
WebGIS, 65
Weedmats, 240, 244, 246-247

Wetlands, 52, 57, 78, 182, 192, 230, 260, 262
optimum resolution for studying, 78-91
WIFS, 232, 233
Wireless Application Protocol (WAP), 69-71, 73, 239, 243
World Bank, 191
World Construction Set, 101-104
World Wide Web, see Internet

X

XML, see eXtensible Markup Language
Xtools, 247

Y

Yucatan Peninsula, 229-237

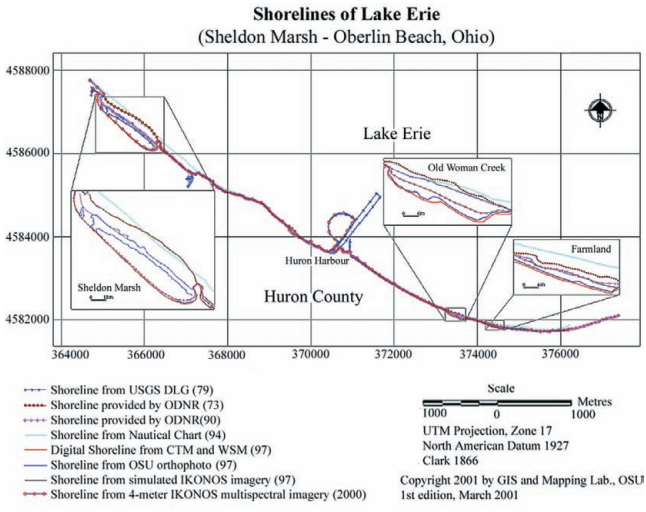


Plate 3.1 Shorelines compared and analysed in Chapter 3, Li et al.

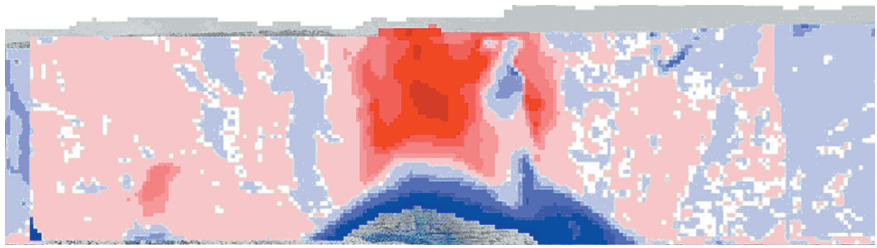


Plate 4.1 Changes between August 2000 and February 2001.

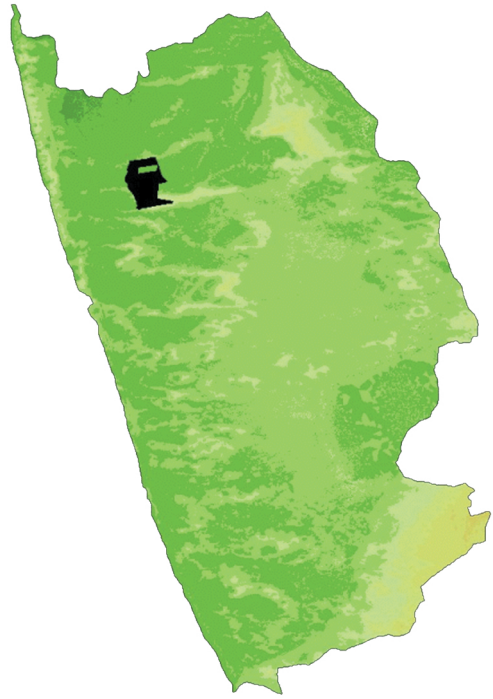


Plate 4.2 The Digital Elevation Model (DEM) for Kenfig NNR. Elevations are represented by different shades of green. Lower to higher ground is graded from darker to lighter shades.

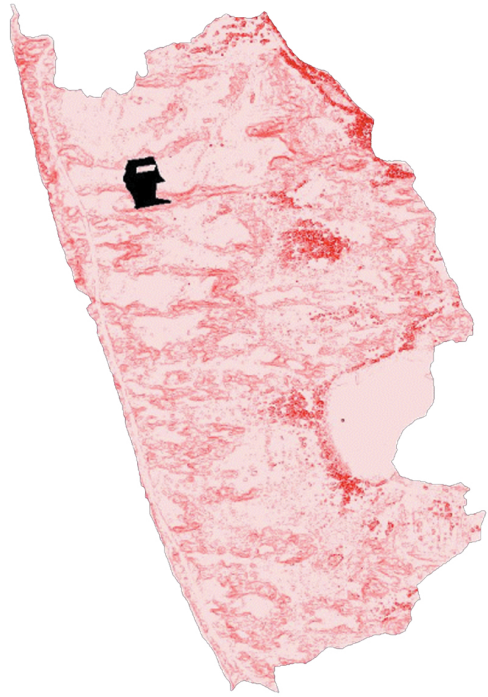


Plate 4.3 Slope data for Kenfig NNR. Slope "steepness" is represented by different shades of red, with darker shades representing steeper slopes. Clusters of red in circular shapes are individual trees.

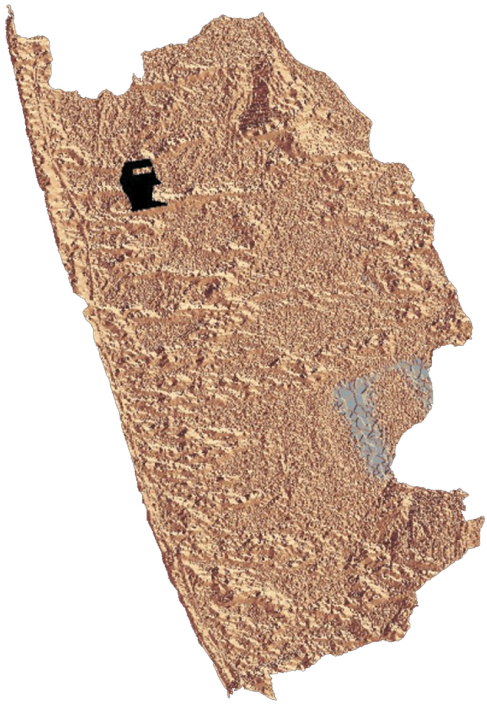


Plate 4.4 Aspect data for Kenfig NNR. Orientation of the slopes is represented by different shades of brown. Lighter shades represent a northerly orientation, and darker shades indicate a southerly direction.

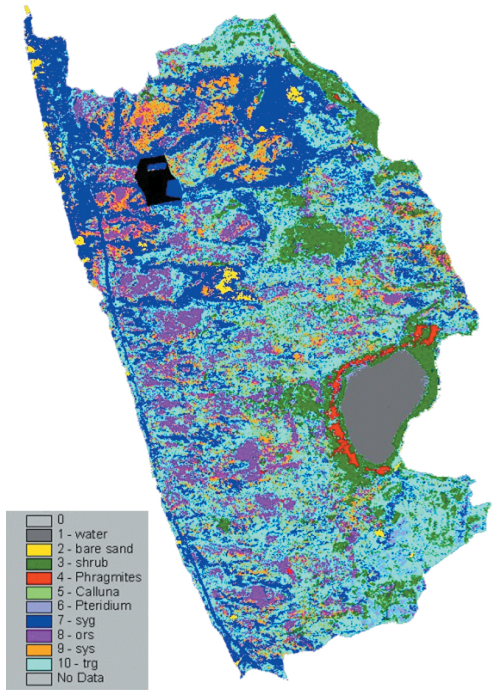


Plate 5.5 The classified habitat map of Kenfig NNR using CASI data.

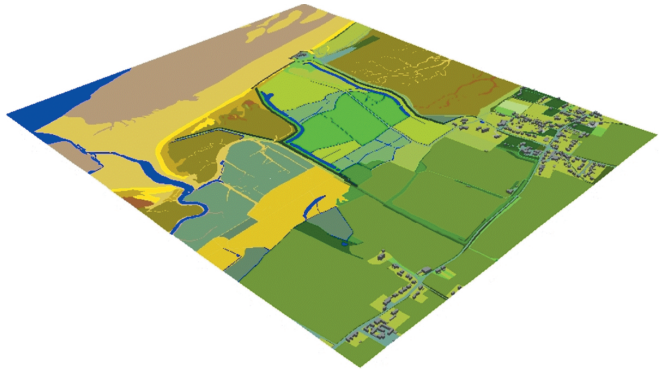


PLATE 8.1 ArcView 3D Analyst view of the site prior to the scheme. © Crown Copyright Ordnance Survey. An EDINA Digimap/JISC supplied service.

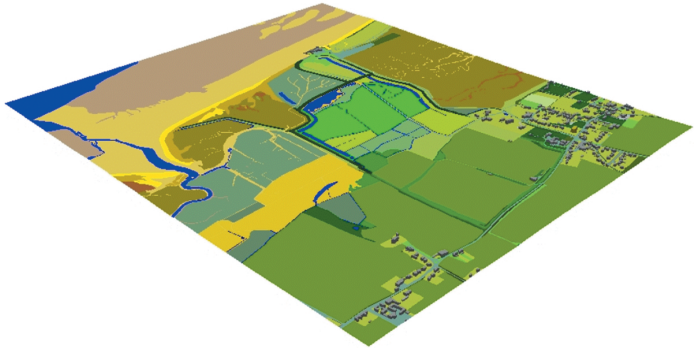


PLATE 8.2 ArcView 3D Analyst view of the site following construction of the scheme. © Crown Copyright Ordnance Survey. An EDINA Digimap/JISC supplied service.

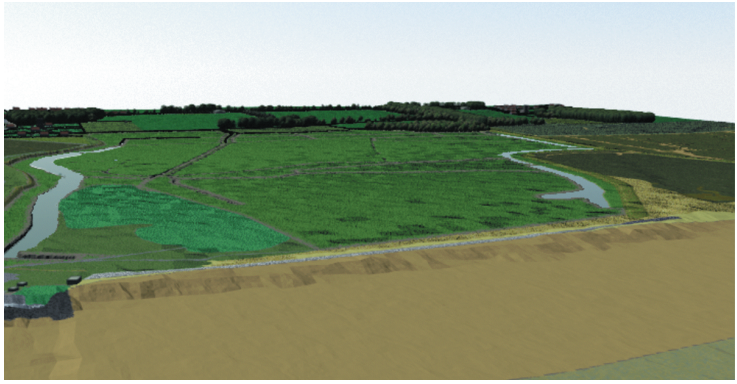


PLATE 8.3 Word Construction Set view looking south over Brancaster West Marshes prior to scheme construction. © Crown Copyright Ordnance Survey. An EDINA Digimap/JISC supplied service.

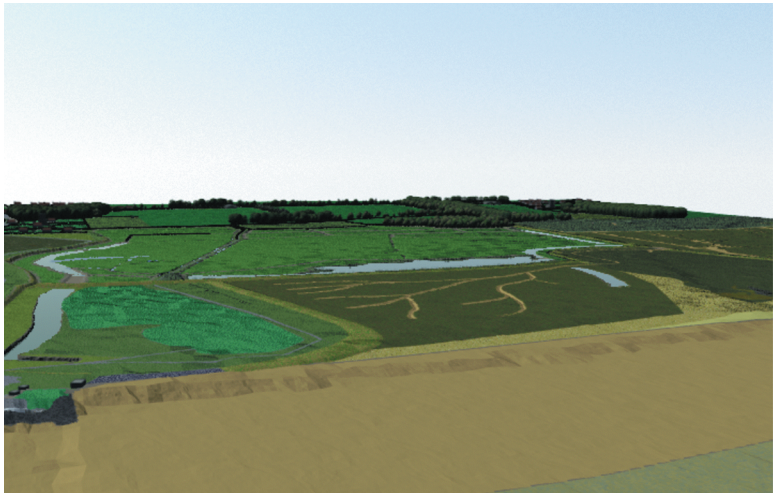


PLATE 8.4 World Construction Set view looking south over Brancaster West Marshes following construction. © Crown Copyright Ordnance Survey. An EDINA Digimap/JISC supplied service.

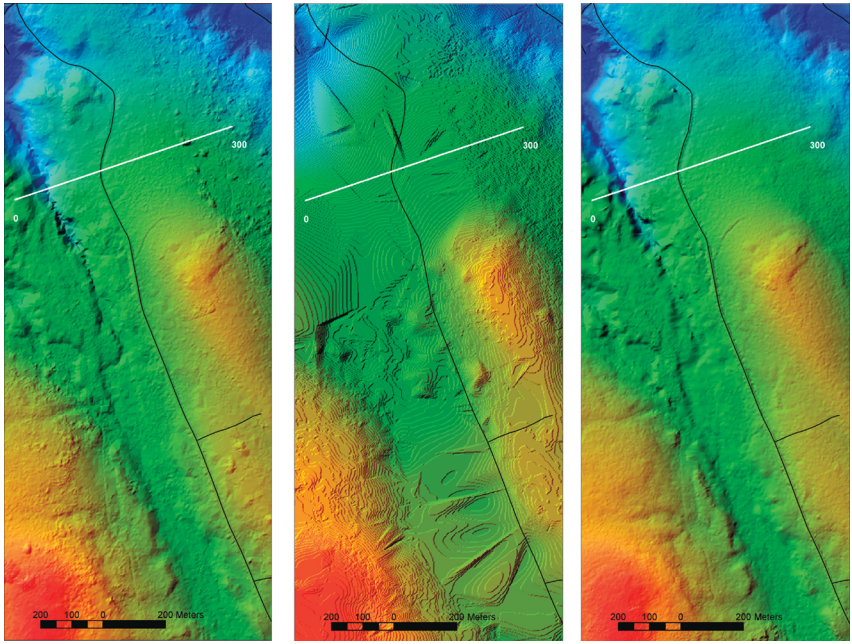


PLATE 15.1 Mosaic of 1 m CASI at low tide (left image), 2 m LIDAR DSM at low tide (centre image), and 3 m CASI at high tide (right image) for Port Lorne along the Bay of Fundy. Overall image is approximately 4 km across.

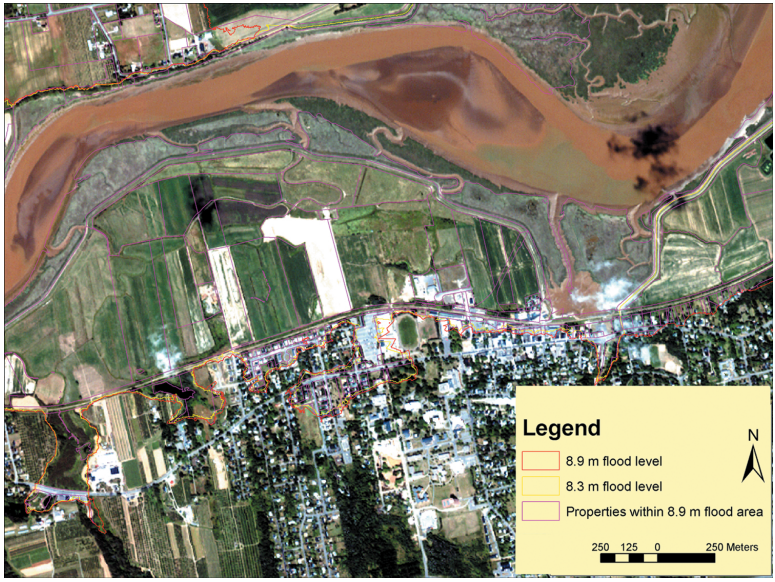


PLATE 15.2 Close-up of flood levels of one-in-twenty year (yellow) and one-in-one-hundred year (red) storm events with property boundaries (purple) that intersect the 100-year flood level displayed over an Iknonos image. Economic impact of such an event can be assessed by using GIS overlay functions to determine the value of infrastructure and properties affected. Includes material copyright Space Imaging, LCC.

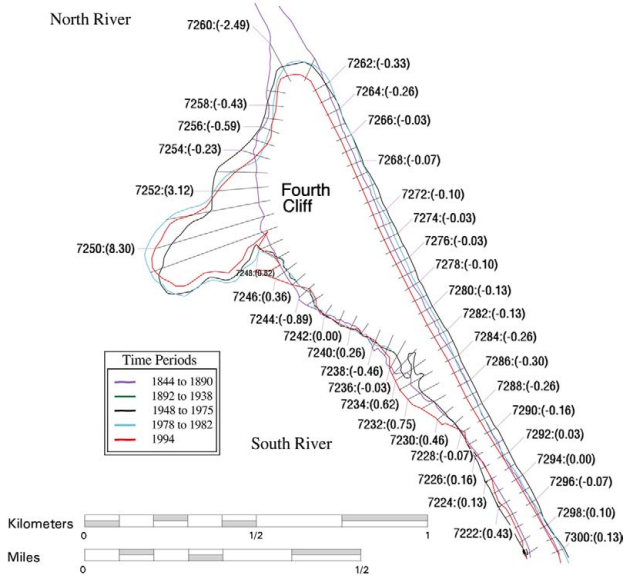


PLATE 16.1 Enlarged section of a shoreline change map for Scituate, Massachusetts, displaying historical shorelines, transects, and linear regression rate. (For clarity at page size, the orthophotograph underlay is not shown.)

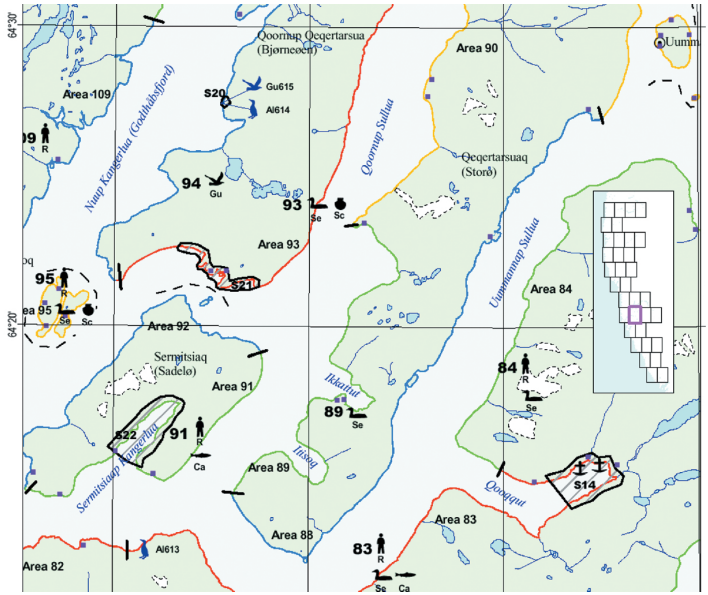


Figure 20.1A An example from the Atlas showing part of a shoreline sensitivity map. The shoreline areas (segments) have four different colours, indicating the level of sensitivity to oil spill. The black symbols indicate the important biological elements and resource use at the shoreline area, the blue symbols shows the site specific bird colonies.

Shoreline Sensitivty Maps

Shoreline Species

- Al Alcids
- Ar Arctic char
- Ca Capelin
- Co Cormorants
- De Deep sea shrimps
- Gu Gulls
- Ha Harbour seals
- Lu Lumpsuckers
- Sb Seaducks breeding
- Sc Scallops
- Se Seaducks
- Sn Snow crabs
- Tu Tubenoses

Site Specific Shoreline Species

- Al Alcids
- Co Cormorants
- Gu Gulls
- Ha Harbour seals
- Sb Seaducks breeding
- Se Seaducks
- Tu Tubenoses

Shoreline Resource Use

- R Resource use (Human use)
- Archaeological site

Shoreline Areas Sensivity Ranking

- Extreme (> 45)
- High (33 - 45)
- Moderate (22 - 33)
- Low (< 22)

Selected Areas

- S Selected area

Figure 20.1B The legend for the shoreline sensitivity map. Symbols are displayed with the species name and the abbreviation for the name.

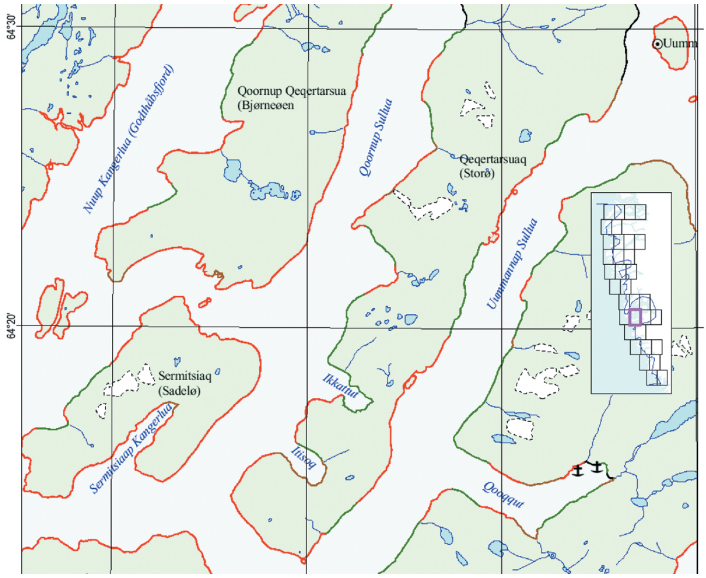


Figure 20.2A An example from the Atlas showing the Physical Environment and Logistics map. The map shows the shoretype classification and where harbours/anchorages are located.

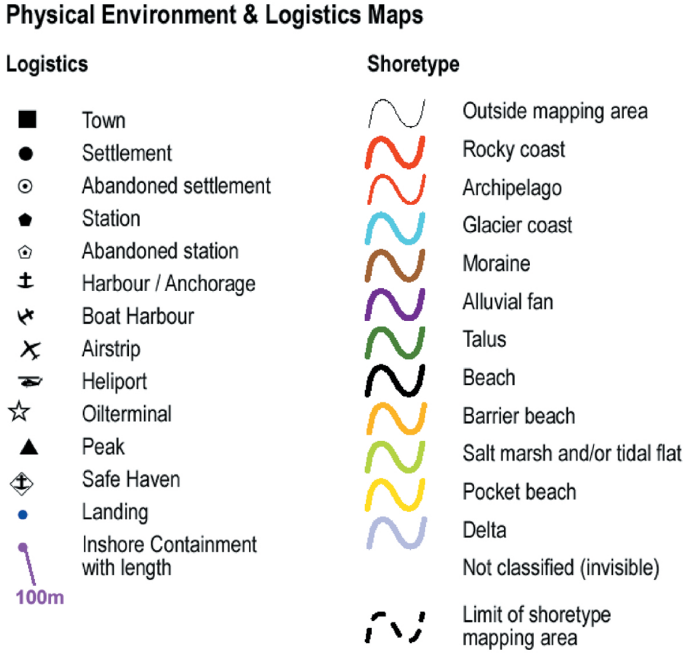


Figure 20.2B The legend for the Physical Environment and Logistics map. The twelve geomorphologic shoretypes all have individual colours.

GIS for Water Resources and Watershed Management

Edited by
John G. Lyon

Taylor & Francis
Taylor & Francis Group
LONDON AND NEW YORK

Also available as a printed book
see title verso for ISBN details

GIS for Water Resources and Watershed Management

GIS for Water Resources and Watershed Management

Edited by
John G. Lyon

Taylor & Francis
Taylor & Francis Group
LONDON AND NEW YORK

First published 2003 by Taylor & Francis
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
By Taylor & Francis Inc,
29 West 35th Street, New York, NY 10001

Taylor & Francis is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2003.

© 2003 Taylor & Francis
except Chapters 1, 3, 5 & 6 © American Water Resources Association; Chapter 13 © Elsevier Science; Chapter 14 © American Society for Photogrammetry and Remote Sensing; Chapters 2, 4, 7, 17 and 18, which originate from a US Government source and are in the public domain.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known, or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Every effort has been made to ensure the advice and information in this book is true and accurate at the time of going to press. However, neither the publisher nor the authors can accept any legal responsibility or liability for any errors or omissions that may be made. In the case of drug administration, any medical procedure or the use of technical equipment mentioned within this book, you are strongly advised to consult the manufacturer's guidelines.

British Library Catloguing in Publication Data
A catalogue record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested.

ISBN 0-203-21791-8 Master e-book ISBN

ISBN 0-203-27349-4 (Adobe eReader Format)
ISBN 0-415-28607-7 (Print edition)

Contents

About the editor ..... **ix**
Contributors ..... **xi**
Acknowledgements ..... **xv**

Introduction
GIS for Water Resources and Watershed Management **1**
John G. Lyon

Chapter 1
Channel Network Delineation and Watershed Segmentation in the TOPAZ Digital Landscape Analysis System **7**
Lawrence W. Martz and Jurgen Garbrecht

Chapter 2
Assessing the Performance of Automated Watershed Segmentation from Digital Elevation Models **17**
Jurgen Garbrecht and Lawrence W. Martz

Chapter 3
Development of a Basin Geomorphic Information System Using a TIN-DEM Data Structure **25**
Yasuto Tachikawa, Michiharu Shiiba, and Takuma Takasao

Chapter 4
Basinsoft, a Computer Program to Quantify Drainage Basin Characteristics **39**
Craig A. Harvey and David A. Eash

Chapter 5
Deriving Stream Channel Morphology Using GIS-Based Watershed Analysis **53**
Scott N. Miller, D. Phillip Guertin, and David C. Goodrich

Chapter 6
GIS Modeling and Visualization of the Water Balance during the 1993 Midwest Flood **61**
Pawel J. Mizgalewicz, W. Scott White, David R. Maidment, and Merrill K. Ridd

Chapter 7
Selection, Development, and Use of GIS Coverages for the Little Washita River Research Watershed **69**
Patrick J. Starks, Jurgen Garbrecht, F.R. Schiebe, J. M. Salisbury, and D. A. Waits

Chapter 8
Regional Characterization of Inland Valley Agroecosystems in West and Central Africa Using High-Resolution Remotely Sensed Data **77**
Prasad S. Thenkabail and Christian Nolte

Chapter 9
Watershed Characterization by GIS for Low Flow Prediction **101**
Glenn S. Warner, Andrés R. García-Martinó, Frederick N. Scatena, and Daniel L. Civco

Chapter 10
Evaluation of the Albemarle-Pamlico Estuarine Study Area Utilizing Population and Land Use Information **109**
Robert E. Holman

Chapter 11
Application of GIS and Remote Sensing for Watershed Assessment **119**
Lloyd P. Queen, Wayne L. Wold, and Kenneth N. Brooks

Chapter 12
Development of a Database for Lake Ecosystem Studies: Linking GIS with RDBMS **129**
Weihe Guan, Leslie J. Turner, and Sergio L. Lostal

Chapter 13
Historical Aerial Photographs and a Geographic Information System (GIS) to Determine Effects of Long-Term Water Level Fluctuations on Wetlands along the St. Marys River, Michigan, USA **145**
Donald C. Williams and John G. Lyon

Chapter 14
Watershed-Based Evaluation of Salmon Habitat **159**
Ross S. Lunetta, Brian L. Cosentino, David R. Montgomery, Eric M. Beamer, and Timothy J. Beechie

Chapter 15
Physical Characterization of the Navarro Watershed for Hydrologic Simulation **179**
Jeffrey D. Colby

Chapter 16
Hydrologic Modeling Using Remotely Sensed Databases **189**
James F. Cruise and Richard L. Miller

Chapter 17
Technological Advances in Automated Land Surface Parameterization from Digital Elevation Models **207**
Jurgen Garbrecht, Lawrence W. Martz, and Patrick J. Starks

Chapter 18
Aerial Photointerpretation of Hazardous Waste Sites: An Overview **219**
Donald Garofalo

Chapter 19
Remote Sensing and GIS for Site-Specific Farming **241**
John G. Lyon, Andrew Ward, Bruce C. Atherton, Gabriel Senay, and Tom Krill

Index **255**

About the Editor

Lyon was interested early on in water resources, watersheds, and wetlands and other moderately disturbed systems as places to evaluate the condition of native vegetation communities. This interest was honed in his youthful wanderings in the mountains and alpine systems of the Pacific Northwest, California and Nevada, and Alaska. Systematic study of the Lower Columbia River and wetlands in undergraduate work at Reed College in his native Oregon, and graduate work at the University of Michigan yielded Bachelor's and Master's theses and a Doctoral Dissertation devoted to Great Lakes wetlands and other natural systems. Eighteen years as a faculty member and ultimately a full Professor of Civil Engineering and Natural Resources at Ohio State University were further devoted to scholarly pursuits of these interests. A body of work included remote sensor and GIS evaluations of water resources and wetlands, development of field methods for wetlands identification, and other efforts on hydrology, soil, agriculture, riverine, and Laurentian Great Lakes systems.

Contributors

Bruce C. Atherton
Department of Food, Biological and Agricultural Engineering, Ohio State University, Columbus, OH 43210, USA

Eric M. Beamer
Skagit System Cooperative, LaConner, WA 98257, USA

Timothy J. Beechie
Skagit System Cooperative, LaConner, WA 98257, USA

Kenneth N. Brooks
Department of Forest Resources, University of Minnesota, 1530 North Cleveland Avenue, St. Paul, MN 55108, USA

Daniel L. Civco
Department of Natural Resources, University of Connecticut, Storrs, CT 06269, USA

Jeffrey D. Colby
Department of Geography, East Carolina University, Greenville, NC 27858-4353, USA

Brian L. Cosentino
Washington State Department of Fish and Wildlife, Olympia, WA 98501-1091, USA

James F. Cruise
Civil and Environmental Engineering Department, University of Alabama, Huntsville, AL 35899, USA

David A. Eash
U.S. Geological Survey, Water Resources Division, P.O. Box 1230, Iowa City, IA 52244, USA

Jurgen Garbrecht
U.S. Department of Agriculture, Agricultural Research Service, Grazinglands Research Laboratory, 7207 W. Cheyenne Street, El Reno, OK 73036, USA

Donald Garofalo
U.S. Environmental Protection Agency, Environmental Photographic Interpretation Center, National Exposure Research Laboratory/ORD, 12201 Sunrise Valley Drive, 555 National Center, Reston, VA 20192, USA

Andrés R. Garcia-Martinó
U.S. Department of Agriculture, U.S. Forest Service, Caribbean National Forest, Puerto Rico

David C. Goodrich
U.S. Department of Agriculture, Agricultural Research Service, Southwest Watershed Research Center, 2000 E. Allen Road, Tucson, AZ 85719, USA

Weihe Guan
Forest Information Technology, Weyerhaeuser Company, MS WWC 2B2, 33405 8th Ave. S., Federal Way, WA 98003, USA

D. Phillip Guertin
Watershed Resources Program, School of Renewable Natural Resources, University of Arizona, Tucson, AZ 85721, USA

Craig A. Harvey
PixSell, Inc., NASA John C. Stennis Space Center, Bldg. 2105, Stennis Space Center, MS 39529, USA

Robert E. Holman
North Carolina Water Resources Research Institute, North Carolina State University, Campus Box 7912, Raleigh, NC 27695-7912, USA

Tom Krill
John Deere Co., Fort Collins, CO 80524, USA

Sergio L. Lostal
South Florida Water Management District, 3301 Gun Club Road, West Palm Beach, FL 33406, USA

Ross S. Lunetta
U.S. Environmental Protection Agency, National Exposure Research Laboratory/ORD, MD-56, Research Triangle Park, NC 27711, USA

John G. Lyon
P.O. Box 71926, Las Vegas, NV 89170-1926, USA

David R. Maidment
Center for Research in Water Resources, University of Texas, Austin, TX 78712, USA

Lawrence W. Martz
Department of Geography, University of Saskatchewan, 9 Campus Drive, Saskatoon, Saskatchewan, S7N 5A5, Canada

Richard L. Miller
NASA Earth System Science Office, John C. Stennis Space Center, Stennis Space Center, MS 39529, USA

Scott N. Miller
U.S. Department of Agriculture, Agricultural Research Service, Southwest Watershed Research Center, 2000 E. Allen Road, Tucson, AZ 85719, USA

Pawel J. Mizgalewicz
Center for Research in Water Resources, University of Texas, Austin, TX 78712, USA

David R. Montgomery
Department of Geological Science and Quaternary Research Center, University of Washington, Seattle, WA 98195, USA

Christian Nolte
International Institute of Tropical Agriculture (IITA), Nigeria

Lloyd P. Queen
School of Forestry, University of Montana, Missoula, MT 59812, USA

Merrill K. Ridd
Department of Geography, University of Utah, Salt Lake City, UT 84112, USA

J. M. Salisbury
Spatial and Environmental Information Clearinghouse, 003 Life Sciences East, Oklahoma State University, Stillwater, OK 74078, USA

Frederick N. Scatena
U.S. Department of Agriculture, U.S. Forest Service, Caribbean National Forest, Puerto Rico

F.R. Schiebe
SST Development Group, Inc., 824 North Country Club Road, Stillwater, OK 74075, USA

Gabriel S. Senay
EROS Data Center, Sioux Falls, SD 57198, USA

Michiharu Shiiba
Disaster Prevention Research Institute, Kyoto University, Uji 611, Japan

Patrick J. Starks
U.S. Department of Agriculture, Agricultural Research Service, Grazinglands Research Laboratory, 7207 W. Cheyenne Street, El Reno, OK 73036, USA

Yasuto Tachikawa
Disaster Prevention Research Institute, Kyoto University, Uji 611, Japan

Takuma Takasao
Department of Civil Engineering, Kyoto University, Kyoto 606, Japan

Leslie J. Turner
People Soft Global Services, 100 Four Falls Corporate Center, Suite 515, West Conshohocken, PA 19428, USA

Prasad Thenkabail
Center for Earth Observation, Department of Geology and Geophysics, Yale University, 210 Whitney Avenue, New Haven, CT 06520, USA

D. A. Waits
SST Development Group, Inc., 824 North Country Club Road, Stillwater, OK 74075, USA

Andrew Ward
Department of Food, Biological and Agricultural Engineering, Ohio State University, Columbus, OH 43210, USA

Glenn S. Warner
Department of Natural Resources, University of Connecticut, Storrs, CT 06269, USA

W. Scott White
Department of Geography, University of Utah, Salt Lake City, UT 84112, USA

Donald C. Williams
U.S. Army Corps of Engineers, Mississippi Valley Division, Vicksburg, MS 39180, USA

Wayne L. Wold
Boise Cascade Corporation, P.O. Box 50, Boise, ID 83728-0001, USA

Acknowledgements

Chapter 1 is reproduced by kind permission of the American Water Resources Association from the *Symposium on Geographic Information Systems in Water Resources.*

Chapters 2, 4, 7, 17 and 18 originate from a US Government source and are in the public domain.

Chapter 3 is reproduced by kind permission of the American Water Resources Association from the Journal of the American Water Resources Association.

Chapter 5 is reproduced by kind permission of the American Water Resources Association from *GIS and Water Resources*.

Chapter 6 is reproduced by kind permission of the American Water Resources Association from the symposium proceedings *GIS and Water Resources*.

Figures 10.1 and 10.2 are in the public domain.

Figures 12.1, 12.3, 12.5 © by South Florida Water Management District

Chapter 13 is reprinted from *Aquatic Botany* **58**, Donald C. Williams and John G. Lyon, "Historical aerial photographs and a geographic information system (GIS) to determine effects of long-term water level fluctuations on wetlands along the St Mary's River, Michigan, USA", 363–378, ©1997 with kind permission of Elsevier Science – NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands.

Chapter 14 is reproduced with permission, the American Society for Photogrammetry and Remote Sensing. Lunetta R., Cosentino B., Montgomery D., Beamer E. and Beechie T. "GIS-Based Evaluation of Salmon Habitat in the Pacific Northwest". *Photogrammetric Engineering and Remote Sensing*, Vol 63 no. 10 (October 1997), 1219–1229.

Figures 15.2 and 15.3 are reproduced by kind permission of the American Water Resources Association from Colby, Jeffrey D. 1996, Physical Characterisation of the Navarro Watershed for Hydrologic Simulation, *GIS and Water Resources* (Proceedings of Symposium), pp 383–392.

Figure 16.3 is reprinted from *Remote Sensing of Environment* **53**, Miller R.L. and J.F. Cruise, "Effects of Suspended Sediments on Coral Growth: evidence from remote sensing and hydrologic modeling", 177–187, ©1995 with permission from Elsevier Science.

Figures 16.4, 16.5, 16.6 and 16.7 are reproduced by kind permission of American Water Resources Association from Water Resources Bulletin (now the Journal of the American Water Resources Association).

INTRODUCTION

GIS for Water Resources and Watershed Management

John Grimson Lyon, Editor

The chapter authors and the Editor are pleased to bring you this volume of outstanding contributions to water resources research. These authors and others have pioneered and developed viable applications of Geographic Information Systems (GIS) and continue to do so. As a result, the current state of water resources and watershed modeling has been advanced by the use of GIS and allied technologies (Maidment and Djokic, 2000). This volume reports their accomplishments and shares their insight with others who may find these contributions valuable to their own work.

A number of GIS applications to water resources and watersheds have been completed over the years, and they nicely illustrate the potential of the technology. Many of these efforts have resulted from the need to address difficult-to-achieve project goals. GIS applications are applied due to the variability of the resources over time and space, and the number of variables that must be evaluated.

Geographic Information Systems are databases that usually have a spatial component to the storage and processing of the data. Hence, they have the potential to both store and create maplike products. They also offer the potential for performing multiple analyses or evaluations of scenarios such as model simulations.

Data are stored in multiple files. Each file contains data in a coordinate system that identifies a position for each data point or entry. Characteristics of the data point are stored as "attributes." A database of individual files is developed and the combined files may contain characteristics or attributes such as stream locations, topography, water or soil chemical sampling, management practices, ownership, biota, point sources, and any other data that can be collected and have meaning for the analysis.

The GIS supplies value by virtue of the detail of the database, and how these data can be used to address the application of interest. Each variable or "layer" in the GIS supports the application with data on the characteristics of water resources or watersheds.

An important capability of GIS is the simulation of physical, chemical, and biological processes using models. GIS can potentially be used with deterministic or complex models based on algorithms simulating processes, or they can be applied with statistical models. The requirement is that the model to be applied has the capability to take spatial and/or multiple file or "layer" data as input to computations.

Geographic Information Systems are also amenable to evaluations as to quality. They may be used to evaluate the quality characteristics of the product, and to determine the inherent accuracy (Lyon, 2001). Many methods can assist the assessment of accuracy, and these methods can be used in experiments.

Advanced applications utilize GIS, remote sensor, and Global Positioning Systems (GPS) tech-

nologies to better address the project goals. These advanced approaches make use of more complex models and GIS technologies, data collection along with GPS (Kennedy, 1996; Van Sickle 2001), and remote sensing data collection and analyses.

There are a number of strengths that GIS technologies bring to water resources research. GIS allows for improved database organization and storage. The objectives of many watershed studies include watershed segmentation, identification of drainage divides and the network of channels, characterization of terrain slope and aspect, catchment configuration, and routing of the flow of water. Obtaining these variables has been difficult to do from paper maps and aerial photographs. These traditional methods are subject to errors related to manual operations. The work using traditional measurement methods has also proved to be time-consuming.

Once the GIS system is developed, it is straightforward to produce needed data, make maps, and provide simple or complex displays of model results.

The GIS also allows for integration of other sources of data. Historical maps and aerial photos can be of assistance. The U.S. Geological Survey (USGS), the U.S. Department of Agriculture, and the National Archives and Records Service hold originals of maps and photos. Copies of these can be obtained, corrected for inconsistencies, and scanned into a computer file for analyses (Garofalo, 2001).

The USGS also creates and sells Digital Elevation Model (DEM) data sets for the U.S. These data are sets of point elevations on a regular sampling grid. These point elevations are organized by USGS 1:24,000 scale or smaller scale quadrangles. The use of DEM products and custom DEMs are the subject of much of the book, and others (Maidment and Djokic, 2000).

The above USGS products are available at low cost. They may also be acquired at no cost from the USGS address on the Internet.

These USGS products do have the potential for being inaccurate. This may be due to systematic or nonsystematic errors, old source quadrangle maps, or other problems. Normally, the user will make any corrections or modifications for the needs of a given project. Corrections and methods to make corrections in DEM data are an important issue in this book.

It is also possible to purchase enhanced USGS products and other products from vendors who develop the improved DEM products. These enhanced or "value-added" products are quite good, and they represent a considerable cost savings as compared to generating the data anew or correcting available USGS versions. Vendors may also have unique products, or products that can be further enhanced to meet user needs. Examples would include customizing an existing data set to provide only information for a given watershed that is found on several map sheets, or developing a fine detailed DEM by use of airborne LIDAR measurements.

Another, potentially valuable data set is U.S. Department of Agriculture Natural Resource Conservation Service (NRCS, formerly Soil Conservation Service) soils data. Several programs have resulted in general soils data stored in digital form, and products are available at different scales and levels of detail. For example, the STATSGO data sets are files of soil survey soil type and boundary data, and the chapters present some of the capabilities of these and similar databases.

A number of projects have used satellite remote sensor measurements. The resulting remote sensor data provide valuable inputs to GIS, due to the difficulty of access to research areas or the size of the study areas. Supplying data on watersheds over sizable areas has been a valuable use of satellite systems, and has been conducted in a variety of ecosystems including Florida, Washington, the Great Lakes, and Africa as discussed in this book (Lyon, 2001).

GIS also allows for advanced analysis and modeling methods to be implemented in support of research efforts. The use of simulations of models provides detail on water movement and transport of materials.

GIS databases and technologies allow for optimization of model results. The running of model

scenarios supplies detailed information for the development of plans, management decisions, and informed leadership. Repetitive processing facilitates predictions and forecasting of events.

Certain characteristics of and applications in water resources research lend themselves to GIS databases and GIS analyses. As we know, water flows downhill and supplies a directional characteristic to the modeling efforts. The flow can be determined by gauging, and the simulations of water discharge and water quality above a gauge can be compared with the reality measured at the gauge.

The history of watershed and water resources research has included many models that utilize spatial data. These "traditional" models incorporate data with a spatial basis or the components of spatially averaged data. This is because applications have been dictated by need, and models exhibit sensitivity to these variables.

The future promises a variety of enhanced applications of GIS and allied technologies for water resources and watershed research. We will see better integration of data within databases, and the advent of more uses of three-dimensional visualization of data. Databases will become available in greater numbers and detail, and will be procured through the Internet and the World Wide Web.

The advantages of using GIS for watershed studies has been recognized. It is apparent that the capabilities of GIS are potentially valuable in a number of efforts. Over time the need for information has resulted in the development of appropriate algorithms to facilitate the utility of GIS and statistical or deterministic models. Much traditional modeling and analyses preceded the advent of GIS and many efforts are ongoing to join the approaches and optimize their capabilities. It is these kinds of efforts that have been brought together here to help share the experience of the authors with the readers.

The chapters presented here demonstrate a number of practical applications of theory and applications of GIS technologies to solving real-world problems. Some outstanding authors have contributed. They wish to showcase their high quality efforts, and share their knowledge and experience with other professionals. The result is an informative and practical book that will be enjoyed for years to come.

The book begins with two contributions from Martz and Garbrecht, who are well-known for their thoughtful work in watershed and water resources research. These papers help to characterize the variable that can be obtained through analysis of Digital Elevation Model (DEM) data.

Here, DEM data were preprocessed to identify drainage networks using the Digital Elevation Drainage Network Model (DEDNM). Application of the model and software allows extraction of topographic and topologic drainage network and watershed properties, using available DEM data from 1:24,000 scale USGS quadrangle maps. Data corrections such as removal of depressions and flat areas are addressed and methods discussed. Watershed characteristics are developed such as channel tracing by flow direction using codes to identify links, nodes and subwatersheds. Other computations to develop variables include calculations of slope and overland flow distances, elevations, and associated flow paths.

The second chapter by these two authors further develops the efforts to calculate watershed characteristics. Included are efforts to correct artifacts in DEM data using DEDNM, and use of other algorithms in the TOpographic PArameteriZation program (TOPAZ). The TOPAZ digital analysis system processes and analyzes DEM data to characterize topography, measure parameters, identify surface drainage, subdivide watersheds, and quantify the drainage network.

An additional contribution of this chapter is an evaluation of an automated approach to development of watershed data including segmentation, channel network identification, and subwatershed definition. This is completed for a small watershed in low relief terrain. The channel network generated by DEDNM is compared with a traditional blue-line method to determine the capability and accuracy of the automated approach.

The chapter by Tachikawa, Shiiba, and Takasao presents a TIN-DEM based topographic model which incorporates the advantages of grid-based methods and contour-based methods for generating watershed characteristics. The results were applied to the question of determining the structure of a distributed rainfall-runoff model, and determinations of source of flow and direction of runoff flow.

The chapter by Harvey and Eash provides a history and description of capabilities of the Basinsoft program. This program processes data through modules, and can quantify 27 different basin characteristics. Watershed characteristics can be generated for a given basin or multiple basins, and the repetitive steps can be automated.

The capabilities of Basinsoft were compared to manual measurement in this chapter. Twelve selected drainage basin characteristics were made from USGS topographic data for drainage areas upstream of 11 streamflow gauging stations in Iowa. Results of Wilcoxon signed ranks test indicated that Basinsoft quantifications were not significantly different from manual measurements for 9 of 10 drainage basin characteristics tested. This work was significant because it demonstrated statistically that GIS- and DEM- derived parameters were similar to those derived from traditional, manual measurement procedures. Such validation of automated procedures is a necessary and important step in GIS research and applications.

The chapter by Miller, Guertin and Goodrich presents a methodology for predicting channel shape from watershed characteristics that could be readily derived from commonly available GIS data. A high resolution database was constructed and tested for the Walnut Gulch Experimental Watershed in Arizona. Channel cross-sectional area and width were found to be significantly related to channel order, upstream watershed size, and maximum contributing flow length within a watershed.

The chapter by Mizgalewicz, White, Maidment and Ridd, describes procedures developed to model the water balance of the large-scale Mississippi River flooding event of 1993, and to utilize hydrologic and meteorologic data sets in a GIS database. The goals were to develop a comprehensive understanding of the climate and hydrologic conditions related to the floods. The chapter also presents ideas on how to use visualization software in the depiction of water storage change over the land surface during the flood period.

Starks, Garbrecht, Schiebe, Salisbury, and Waits address the development of hydrographic data layers from DEM. Included in their chapter is the evaluation of reliability of the soil property data extracted from soils coverage and Soil Survey data, and land use and land cover data derived from satellite sensors. The topographic information was developed from TOPAZ, a set of algorithms and software described earlier. The use of Soil Surveys and attribute data was examined, and data employed in the project were taken from the STATSGO database. The concept of critical source area (CSA) was also described and its utility addressed.

The chapter contribution by Thenkabail and Nolte demonstrates the use of remote sensor data, Global Positioning Systems (GPS), and ground data for input to a GIS. In this chapter, three large study areas in Africa were evaluated, and the capabilities of these technologies to locate resources was demonstrated. Fascinating information was developed on the presence of inland valley bottoms, their cultivation, and the proximity of roads to farm fields in a poorly mapped and developed area. The utility of the derived information was great, and the results can demonstrate how these technologies supply details unavailable from any source, and help to fix the location of resources in space and in time.

The chapter by Warner, García-Martinó, Scatena, and Civco is an interesting example of GIS and watershed modeling for use in low-flow watersheds. Low-flow or dry streams in watersheds are difficult to model and this paper demonstrates a thoughtful way to deal with the problems presented by such a watershed in Puerto Rico.

In his chapter, Holman demonstrates how land cover data from Landsat satellites can assist in modeling land cover and use in a rural and urban study area of large size. Numerous and interesting results were generated by evaluation of the database using GIS.

The chapter by Queen, Wold, and Brooks presents a GIS that was developed to manage and analyze landscape data for nine representative subwatersheds in the Nemadji basin of Minnesota. Output from the GIS was used in regression modeling of the relationships between the frequency of soil mass movements and the landscape characteristics. Significant results included the relationship between slump frequency and total forested area and percent nonforested land.

Guan, Moore, and Lostal illustrate in their chapter how a large data set can be accumulated by submitting individual results of projects to a shared database. They discuss the requirements of managing such a database, and making the data available to many users. They also examine the problems associated with changing hardware and software, and how to maintain a useful database over multiple-year time periods.

The chapter by Williams and Lyon presents the utility of historical aerial photos and GIS analyses to inventory and monitor a dynamic water resource over a more than 50-year period. The results address the hydrologic forcing function of fluctuating Great Lakes water levels and their influence over time on coastal wetlands resources. The results have allowed resource managers to examine the natural variability of the wetlands, and any potential influence of human activities.

How the influence of habitat conditions can be evaluated for a given wildlife species is shown in the chapter by Lunetta, Cosentino, Montgomery, Beamer, and Beechie. The issue of salmonid habitat and human activities is a potent one, yet little quantitative, spatial-based work has been done. The combination of GIS analysis and data from a variety of sources including Landsat satellites allowed this difficult issue to be addressed and quantitative information be brought to the dialogue.

In his chapter, Colby examines the role of GIS in pulling together a number of data sources to address watershed characteristics and runoff yields for a rural watershed in Puerto Rico.

The chapter by Cruise and Miller examines the capabilities of remote sensor data to facilitate hydrologic modeling. The combination of remote sensor data, GIS, and models provides the spatial, spectral, and numerical capacity to address topical interests. Results demonstrate that macroscale hydrologic modeling with conventional remote sensing tools can provide simulations of observed climatological features over fairly long time periods.

The chapter by Garbrecht, Martz, and Starks presents the results of preprocessing DEM data to treat problems associated with depressional and flat areas. The additional image processing technologies better capture the channel networks from overland or hillslope areas, and enable automated management of the water flow routing process in large drainages, and help generate network and subcatchment data necessary for rainfall-runoff modeling.

Garofalo demonstrates the practical value of using historical aerial photographs and photointerpretation techniques for historical analyses. This very informative treatment details how valuable information can be derived on hazardous waste disposal areas and the associated water resource characteristics. Use of these techniques can greatly assist users in developing hard-to-obtain factual information on most sites of interest, and for populating GIS databases.

The chapter by Lyon, Ward, Atherton, Senay, and Krill demonstrates a number of techniques that can be applied to agricultural and water resource applications (Ward and Elliot, 1995). The combination of advanced techniques such as remote sensing, GIS, GPS, variable rate applications, detailed soil chemical sampling, and related approaches has greatly improved decision making in agriculture. The use of site-specific or precision agriculture can be done in a cost-friendly manner, and greatly improve production and protect water quality.

BIBLIOGRAPHY

Maidment, D. and D. Djokic, 2000. Hydrologic and Hydraulic Modeling Support with Geographic Information Systems. ESRI Press, Redlands, CA.

Garofalo, D., 2001. Aerial photointerpretation of hazardous waste sites. In Lyon, J., Ed., 2001.

Kennedy, M., 1996. *The Global Positioning System and GIS*. Ann Arbor Press, Chelsea, MI.

Lyon, J., 2001. *Wetland Landscape Characterization.* Ann Arbor Press, Chelsea, MI.

Lyon, J., Ed., 2003. *GIS for Water Resources and Watershed Management*. Taylor & Francis, London.

Van Sickle, J., 2001. *GPS for Land Surveyors*. Ann Arbor Press, Chelsea, MI.

Ward, A. and W. Elliot, 1995. *Environmental Hydrology*. Lewis Publishers, Boca Raton, FL.

CHAPTER 1

Channel Network Delineation and Watershed Segmentation in the TOPAZ Digital Landscape Analysis System

Lawrence W. Martz and Jurgen Garbrecht

INTRODUCTION

Research over the past decade has demonstrated the feasibility of extracting topographic information of hydrological interest directly from digital elevation models (DEM). Techniques are available for extracting slope properties, catchment areas, drainage divides, channel networks and other data (Jenson and Domingue, 1988; Mark, 1988; Moore et al., 1991; Martz and Garbrecht, 1992). These techniques are faster and provide more precise and reproducible measurements than traditional manual techniques applied to topographic maps (Tribe, 1991). As such, they have the potential to greatly assist in the parameterization of hydrologic surface runoff models, especially for larger watersheds (i.e., >10 km^2) where the manual determination of drainage network and subwatershed properties is a tedious, time-consuming, error-prone, and often highly subjective process. The automated techniques also have the advantage of generating digital data that can be readily imported and analyzed by Geographic Information Systems (GIS).

This chapter presents an overview and discussion of the original DEDNM (Digital Elevation Drainage Network Model) computer program developed to measure drainage network and subcatchment parameters directly from a DEM. A detailed discussion of the original program structure and algorithms is available in Martz and Garbrecht (1992). Improvements and modifications to the program continue to be made, although the essential approach to drainage analysis remains largely unchanged. DEDNM now functions as the core component in a more comprehensive digital landscape analysis system known as TOPAZ (TOpographic PArameteriZation) (Garbrecht and Martz, 1999).

DEDNM OVERVIEW

The main purpose of the program is to provide an automated, rapid, and reproducible evaluation of the topographic properties of large watersheds by processing raster DEMs. The primary application target for the program is the parameter determination for hydrologic surface runoff models. The following general objectives guided the program development:

1. Program input should be DEMs that are both commonly available to hydrologists and represent the land surface in sufficient detail for hydrologic analysis.
2. The program should not be limited by the nature of the topography represented in the DEM and should be applicable to both high and low relief terrain.

3. The program should extract all topographic and topologic drainage network and watershed properties relevant to the parameterization of hydrologic surface runoff models.
4. The program should incorporate established and proven algorithms from earlier research, and supplement these with new algorithms as required.
5. The program output should be structured so that it can be used directly for hydrologic model parameterization and be readily imported by a GIS.

The program first preprocesses the DEM to remove depressions and flat areas without modifying other parts of the DEM. Flow simulation algorithms are then applied to determine the catchment area of each grid cell, and to define the boundaries of the watershed to be analyzed. A continuous, unidirectional channel network is delineated by selecting all grid cells with a catchment area in excess of a user-specified critical source area.

Network evaluation involves tracing the course of the channels in the previously defined watershed, and identifying the junctions in the network. Channel tracing is directed by flow direction codes generated in the flow simulation analysis. The channels are traced initially to order the network according to the Strahler system, and to determine the length of each network link and the spatial coordinates of its upstream and downstream end. These data are used, in turn, to calculate the slope and catchment area properties of each link, and the number and average properties of the channel segments of each Strahler order. The channels are traced again to assign unique identification numbers to each node in the network. These node numbers are used to associate link, node, and subwatershed data, and can also be used to optimize flow routing in hydrologic models. A final trace is used to identify the subwatersheds of each source node and of the left and right banks. The individual subwatershed areas are then evaluated to determine their slope and overland flow travel distances.

The program generates both tabular and raster output (Table 1.1). The tabular output gives the watershed and drainage network properties of the total drainage network and of the individual network links. The raster output provides maps of the watershed and drainage network components which can be imported into a GIS, registered to other data layers, and used as templates to extract

TABLE 1.1. Summary of the Major DEDNM Program Outputs

Raster Output	
• DEM, depressions filled	• Template of main watershed
• DEM, depressions filled, relief on flat areas	• Channel network: links by Strahler order
• Extent of depressions and flat areas	• Channel network: links by node number
• Flow direction at each grid cell	• Subwatersheds: by node number and subarea code
• Drainage area at each grid cell	

Tabular Output	
• Main watershed (by Strahler order)	• Individual network links
1. Number of channel segments	1. Strahler order
2. Average channel length	2. Upstream and downstream node coordinates
3. Average channel slope	3. Upstream node number
4. Average upstream and direct drainage area	4. Channel length
• Individual subwatersheds	5. Upstream and downstream node elevation
1. Channel area	6. Upstream and direct drainage area
2. Left-bank, right-bank and source node direct drainage area	

additional, hydrologically-relevant information (i.e., soil type, land cover, climate, etc.) for individual subwatersheds and network components.

DEDNM is coded in FORTRAN–77. It is modular to allow easy modification and reorganization to meet specific needs. Each major operation is performed by a set of subroutines called from the main program. Program input is a raster DEM in a standardized format, DEM parameters (i.e., number of rows and columns, elevation range, etc.) and user-specified job parameters that control processing and output. DEDNM keeps all data in dynamic memory during execution to minimize time-consuming I/O operations. While memory requirements are minimized by performing all computations on and storing all elevations as integer data, the processing of large DEMs can create significant memory demands. An option to aggregate the DEM to a coarser grid cell size is provided to permit operation in limited memory or to permit more rapid preliminary analyses of large DEMs. DEDNM can be compiled to run on any hardware platform and, provided sufficient memory is available, can be easily configured to process any size of DEM.

DEM PREPROCESSING

The design objectives for DEDNM emphasize the analysis of generally available DEMs which represent landscapes at a resolution that allows the extraction of hydrologic variables. The 7.5-minute DEMs distributed by the United States Geological Survey (USGS) are typical of such data. They provide elevation values in 1 m increments for 30 m grid cells over areas corresponding to USGS 7.5-minute quadrangles. Much of the United States is now covered by such DEMs, and coverage continues to increase. DEMs with the same general structure and order of resolution are provided by other national jurisdictions and can be generated from SPOT data (Quinn et al., 1991; Tribe, 1991).

A fundamental problem in using DEMs of this order of resolution for hydrologic analysis is the presence of sinks in the data. Sinks are grid cells with no neighbors at a lower elevation and, consequently, with no downslope flow path to a neighbor. By this definition, sinks occur on both flat areas and in closed depressions. Sinks are quite common in DEMs with a spatial and vertical resolution similar to that of the USGS 7.5-minute DEMs. They also tend to be more common in low relief terrain than in high relief terrain. While a few of these sinks may represent real landscape features, the majority are spurious features which arise from interpolation errors during DEM generation, truncation of interpolated values on output, and the limited spatial resolution of the DEM grid (Mark, 1988; Fairchild and Leymarie, 1991; Martz and Garbrecht, 1992).

Figure 1.1 shows the spatial distribution of sinks in a DEM of an 84 km^2 low relief watershed used as a test data set in developing DEDNM (Garbrecht and Martz, 1997). The DEM elevations are in 0.9 m (3 ft) increments for 30 m grid cells. Ten percent of the watershed is covered by sinks, of which 25% are in closed depressions and 75% are on flat areas. The sinks are concentrated along valley bottoms where local slopes are gentle. The banded appearance of many of the flat areas suggests that they are artifacts of the limited vertical resolution of the DEM.

The ability to simulate flow across sinks is essential for effective hydrologic analysis of DEMs of this type. DEDNM provides this capability through a two-phase operation which has the same effective result as the flow modification method of Jenson and Domingue (1988). The first phase involves filling all closed depressions to the elevation of their local outlet using the method of Martz and de Jong (1988). The subwatershed of a depression is delineated and the lowest outlet on the edge of this area is identified. The elevation of all cells within the subwatershed and below the elevation of the lowest outlet are raised to the outlet elevation to simulate depression filling (Figure 1.2). Complex, nested, or truncated depressions are readily evaluated by the filling algorithm.

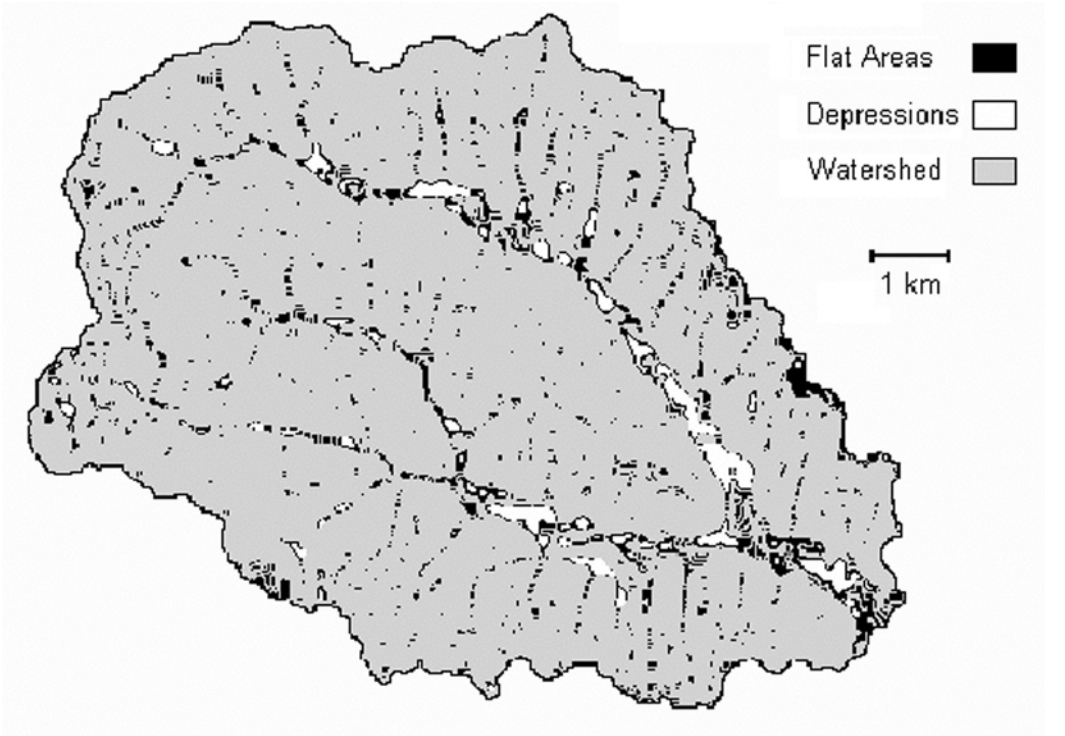


Figure 1.1. Type and spatial extent of sinks in DEM of Bill's Creek watershed in southwestern Oklahoma.

Subtracting elevations in the original DEM from those in the DEM modified by depression-filling gives the location, spatial extent, and depth of depressions.

The second phase imposes relief on all flat areas (both those created by depression-filling and those inherent to the DEM) to permit the unambiguous definition of flow paths across these areas. Two assumptions are implicit to this operation: (1) flat areas are not truly level, but have relief that is simply not detectable at the resolution of the DEM, and (2) this relief directs flow entering or originating on the apparent flat area along a reasonably direct path over the flat area to a point on its perimeter where a downward slope is available. Relief is imposed by adding an elevation increment of 1/1000th of the vertical resolution of the DEM to each cell with no neighbor at a lower elevation (Figure 1.2). This is repeated until no flat areas remain. This approach to the treatment of sinks is superior to the use of prior smoothing to remove sinks because it focuses directly on problem areas without reducing the information content of the DEM elsewhere. It also relies on an assessment of local boundary conditions that bear directly on flow patterns.

The current version of DEDNM incorporated into the TOPAZ digital landscape analysis system introduced two significant improvements for the treatment of depressions and flat areas. These improvements allow the breaching (i.e., lowering) of the outlet of some depressions and apply a two-phase relief imposition algorithm to direct flow across flat areas simultaneously away from higher and toward lower surrounding terrain (Garbrecht and Martz, 1997).

FLOW VECTOR AND DRAINAGE AREA ANALYSIS

Following the preprocessing operation, a flow vector code is assigned to each grid cell using the D8 method (Fairchild and Leymarie, 1991). The flow vector indicates the direction of the

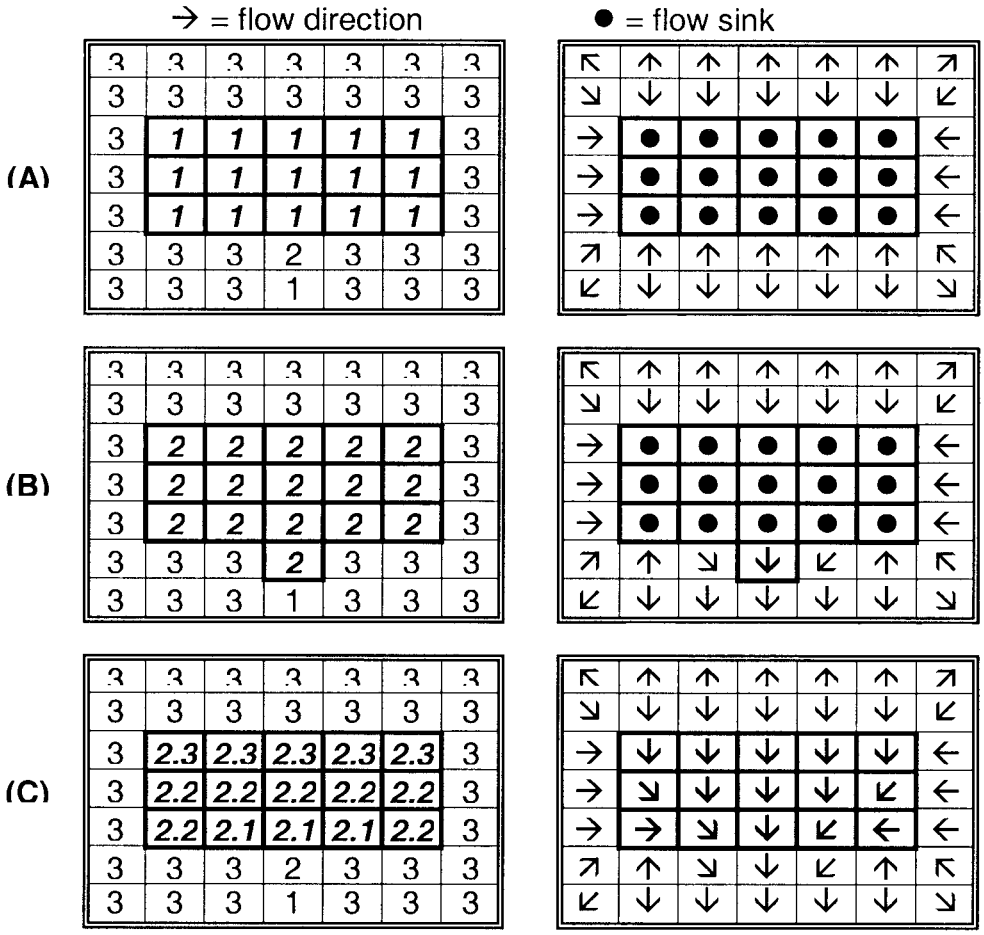


Figure 1.2. DEM preprocessing by DEDNM. Elevations (left column) and associated flow paths (right column) are shown for: (A) original DEM with depression; (B) DEM with flat area created by depression filling; (C) DEM with relief imposed on flat area (for simplicity of illustration, an elevation increment of 1/10th is used).

steepest downward slope to an immediately neighboring cell. Where more than one downward slope maxima exist, the flow vector is arbitrarily assigned to indicate the direction of the maximum first encountered. At cells on the edge of the defined DEM (i.e., cells in the outer rows and columns, or adjacent to a cell with a missing elevation value), the flow vector points away from the defined DEM if no other downward slope to a neighbor is available. All flow originating on or entering a cell is assumed to move in the direction indicated by the flow vector, and no divergent flow out of a cell is accommodated.

The catchment area of each grid cell is determined using the method of Martz and de Jong (1988). The flow vectors are used to follow the path of steepest descent from each cell to the edge of the DEM, and the catchment area of each cell along this path is incremented by one. After a path has been initiated from each cell, the catchment area value accumulated at each cell gives the number of upstream cells which contribute overland flow to that cell (Figure 1.3).

The boundary of the watershed to be analyzed is also determined from the flow vectors. The user specifies the location of the grid cell at watershed outlet, and all grid cells which contribute

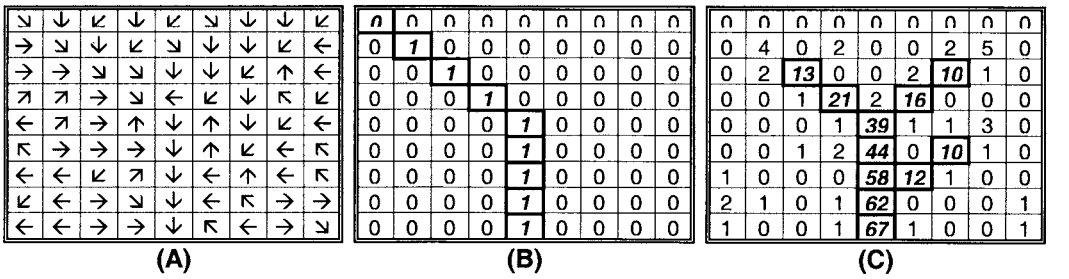


Figure 1.3. Catchment area determination by DEDNM: (A) flow vectors at each DEM cell; (B) catchment area at each DEM cell after the first flow path (highlighted) has been traced from the upper-left cell; (C) catchment area at each DEM cell after flow paths have been traced from all cells. Highlighted cells in (C) are those classified as channels using a critical source area of 10 cells.

overland flow to the outlet cell are identified. This provides a mask of the watershed that is used for subsequent operations.

CHANNEL NETWORK ANALYSIS

The channel network within the watershed is delineated from the catchment area grid. All cells with a catchment area greater than a user-specified critical source area are classified as part of the channel network. This yields a fully connected, unidirectional network. However, the network may contain some very short exterior links which represent valley side indentations, gully outlets, and other features that normally would not be classified as part of the channel network. The network is pruned to remove these spurious exterior links. The pruning involves tracing each exterior link downstream from its source to determine if a user-specified threshold distance (the minimum source channel length) can be traveled before a junction is reached. If so, the link is retained. If not, the link is a candidate for removal. Where two links less than the threshold length join, the shorter link is removed and the remaining link is reevaluated to determine if it now meets the threshold length criterion for retention. A slightly more involved analysis is required at junctions of more than two links.

Once the network has been pruned, the length, starting and ending cell coordinates, and Strahler order of each channel link are determined. Each exterior link is traced downstream from its source until a junction is reached. Each cell along the link upstream of the junction is marked as Strahler order 1, while the cell at the junction is marked as Strahler order 2. On reaching a junction, the link coordinates and link length are stored, and the trace terminates. Once all exterior links have been traced, interior links of successively higher Strahler order are evaluated.

Each junction where a cell was previously marked as being of the current Strahler order (initially order 2) is examined to determine if all links entering the junction have already been assigned a lower order. If this is the case, the junction is the upstream end of a channel segment of the current order, and a trace is initiated. The trace proceeds downstream from the junction, with each cell along the interior link being assigned the current Strahler order until another junction is reached. On reaching a junction, the length and the upstream and downstream coordinates of the link are stored, and the other links entering the junction are examined. If all other links entering the junction were previously assigned a lower order, the order of the downstream link will not increase and the trace continues. If this is not the case, the junction is flagged as being of the next

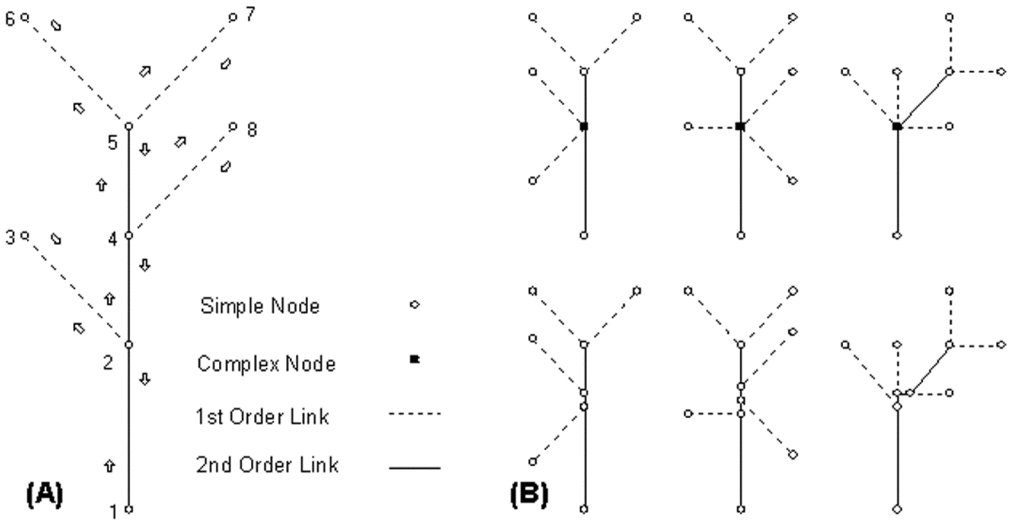


Figure 1.4. Node numbering by DEDNM: (A) the general method of assigning node numbers by tracing the network as shown by arrows; (B) the upper row shows networks with complex nodes (i.e., junctions of more than two channels) in the grid representation of the channel system and the lower row shows corresponding networks in which the complex junctions are effectively replaced by multiple simple nodes in the DEDNM node numbering procedure. The multiple simple nodes are assumed to fall within the grid cell containing the complex node which they replace.

highest order, and the trace terminates. Analysis continues until all links in the network have been ordered.

Following network ordering, the link coordinates are used to find the elevation and the catchment area at the upstream and downstream end of each link. These are used to calculate the longitudinal slope and direct drainage area, respectively, for each link in the network. In addition, the number of channel segments, and the average slope, length, and total and direct drainage area of the segments of each Strahler order are calculated.

A unique identification number is then assigned to each network node using the system described by Garbrecht (1988). Flow vectors are used to simulate a walk along the left bank of the channel beginning and ending at the watershed outlet. Each node is assigned a number the first time it is passed during the walk (Figure 1.4). These numbers can be used by hydrologic models to optimize flow routing, and are used by DEDNM to associate node, link, and subwatershed data. Channel networks often have some junctions where more than two links join. This situation is treated by assuming that more than one node occupies a single grid cell, and assigning one number to each assumed node. Some special analysis is required to ensure that the node number assignment remains consistent with the network ordering applied earlier. The method by which this is done is beyond the scope of this chapter, but some illustrative results are presented in Figure 1.4.

SUBWATERSHED DELINEATION AND ANALYSIS

Subwatershed delineation is accomplished in two stages. In the first stage, flow vectors are used to follow each link from its upstream end to its downstream end, and to identify the cells

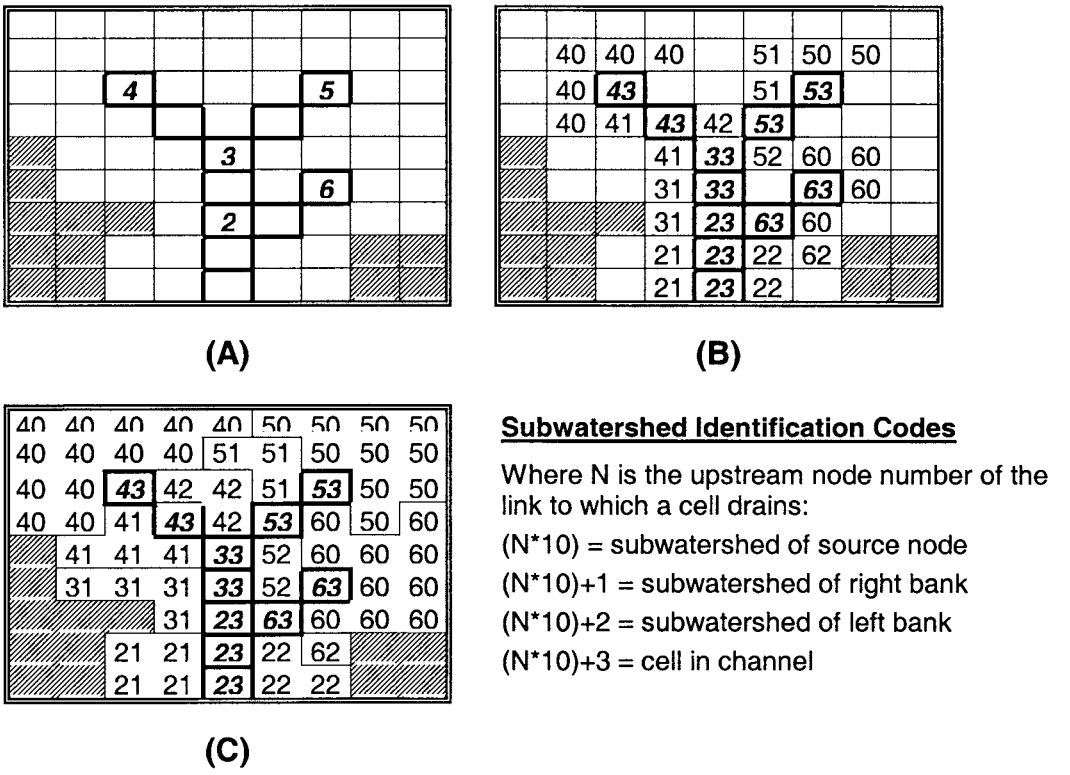


Figure 1.5. Subwatersheds defined by DEDNM from the flow vectors in Figure 1.3: (A) highlighted channel network cells and node numbers of sources and junctions; (B) subwatershed codes assigned to all cells adjacent to the network cells; (C) subwatershed codes assigned to all cells in the watershed. Shaded cells are outside the watershed.

which are immediately adjacent and contribute flow to the link. These cells are classified according to whether they are on the left bank, the right bank or, in the case of exterior links, contribute flow directly to a source node. A subwatershed identification code based on the previously assigned node numbers is applied to each classified cell. In the second stage, the entire grid is scanned and all cells in the main watershed that have not been assigned a subwatershed code are evaluated to determine if they contribute flow to a cell which has been assigned a subwatershed code. If so, it is assigned the code of the cell to which it flows. This is repeated until all cells have been assigned a code. The subwatershed identification process is illustrated in Figure 1.5.

Once the subwatersheds have been identified, they can be evaluated to determine their individual slope, aspect, and overland flow distance properties. These variables were measured directly in the original DEDNM. In the TOPAZ digital landscape analysis system, of which DEDNM is now a part, a separate raster processing program (RASPRO) now performs this and several other functions (Garbrecht and Martz, 1997).

CONCLUSIONS

The computer program DEDNM was developed with the primary objective to assist the rapid parameterization of hydrologic surface runoff models using DEMs similar to those provided by the USGS for 7.5-minute quadrangles. The set of algorithms developed for this purpose provide:

1. Raster maps including the corrected DEM, the flow patterns, the drainage network, and the extent of the main watershed and its subwatersheds.
2. Channel and subwatershed indexing by node identification numbers.
3. Tables of channel and subwatershed attributes, and of drainage network and subwatershed composition.

DEDNM is able to process limited resolution DEM data of low and high relief terrain to provide results that are both conceptually valid and consistent with the topographic characteristics of the landscapes to which it is applied. A more comprehensive evaluation of the capabilities of the program to reproduce the results of a traditional manual analysis of the blue-line network of topographic maps has been undertaken and is reported in the following chapter.

While the essential approach to drainage analysis introduced in DEDNM has remained largely unchanged, the software has been continuously modified and improved. One of the more significant changes has been the incorporation of DEDNM into TOPAZ, a more comprehensive digital landscape analysis system (Garbrecht and Martz, 1997), in which DEDNM provides the core function of network delineation and watershed segmentation from raster digital elevation data.

The TOPAZ digital landscape analysis system processes and evaluates raster DEMs to identify topographic features, measure topographic parameters, define surface drainage, subdivide watersheds along drainage divides, quantify the drainage network and parameterize subcatchments. The core drainage analysis provided by DEDNM allows TOPAZ to maintain consistency between all derived data, the initial input topography, and the physics underlying energy and water flux processes at the landscape surface. The primary objective of TOPAZ is to support hydrologic modeling and analysis, but it can also be used to address a variety of geomorphological, environmental and remote sensing applications.

REFERENCES

Fairchild, J., and P. Leymarie, 1991. Drainage networks from grid digital elevation models. *Water Resources Research,* 27(4):29–61.

Garbrecht, J., 1988. Determination of the execution sequence of channel flow for cascade routing in a drainage network. *Hydrosoft* 1(3):129–138.

Garbrecht, J., and L.W Martz, 1999. *TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation and Subcatchment Parameterization; TOPAZ Overview*. USDA-ARS Publication No. GRL 99–1, 26 pp.

Garbrecht, J., and L.W. Martz, 1997. The assignment of drainage direction over flat surfaces in raster digital elevation models. *Journal of Hydrology,* 193: 204–213.

Jenson, S.K., and J.O. Domingue, 1988. Extracting topographic structure from digital elevation data for geographical information system analysis. *Photogrametric Engineering and Remote Sensing,* 54(11):1593–1600.

Mark, D.M., 1988. Network models in geomorphology. In *Modeling Geomorphological Systems*, M.G. Anderson, Ed., John Wiley and Sons, pp. 73–96.

Martz, L.W., and E. de Jong, 1988. CATCH: A FORTRAN program for measuring catchment area from digital elevation models. *Computers and Geosciences,* 14(5):627–640.

Martz, L.W., and J. Garbrecht, 1992. Numerical definition of drainage networks and subcatchment areas from digital elevation models. *Computers and Geosciences,* 18(6):747–761.

Martz, L.W., and J. Garbrecht, 1998. The treatment of flat areas and closed depressions in automated drainage analysis of raster digital elevation models. *Hydrological Processes,* 12: 843–855.

Moore, I.D., R.B. Grayson, and A.R. Ladson, 1991. Digital terrain modeling: A review of hydrological, geomorphological and biological applications. *Hydrological Processes,* 5(1):3–30.

Quinn, P., K. Beven, P. Chevallier, and O. Planchon, 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Processes,* 5(1):59–79.

Tribe, A., 1991. Automated recognition of valley heads from digital elevation models. *Earth Surface Processes and Landforms,* 16(1):33–49.

CHAPTER 2

Assessing the Performance of Automated Watershed Segmentation from Digital Elevation Models

By Jurgen Garbrecht and Lawrence W. Martz

INTRODUCTION

Watershed segmentation and channel network definition is often required in distributed hydrologic modeling. Manual segmentation from maps is a tedious, time-consuming, and subjective task, particularly for large watersheds. The automated watershed segmentation and extraction of channel network and subwatershed properties from raster elevation data represents a convenient and rapid way to parameterize a watershed. The increasing availability of DEM coverage for many areas of the United States makes this automated watershed segmentation and characterization a promising approach for a wide range of hydrologic investigations. However, assessment of the generated watershed data beyond the usual visual inspection is required to develop confidence in the automated approach.

Early research on automated landscape analysis focused on algorithm development and treatment of unique situations, such as depressions, flat areas, or the connectivity of the network. DEM processing models which identify upward concave areas (Pueker and Douglas, 1975; Jenson, 1985; Band, 1986) often produce discontinuous network segments that must subsequently be connected (O'Callaghan and Mark, 1984) and may require additional adjustments to produce a reasonable pattern (Douglas, 1986). Other models rely on flow routing concepts. In this approach, the steepest downslope direction defines the flow paths (Jenson and Domingue, 1988; Martz and de Jong, 1988; Morris and Heerdegen, 1988). For either approach, problems arise in low relief terrain when the vertical resolution of the DEM is insufficient to identify either upward concave areas or a downslope flow direction. Little work has been done to extend DEM processing methods to low relief terrain, such as found in the central plains of the U.S.

This chapter investigates the performance of automated watershed segmentation, channel network identification, and subwatershed definition in low relief terrain for an 84 km^2 watershed. The landscape analysis computer program chosen from this study is the Digital Elevation Drainage Network Model DEDNM (Martz and Garbrecht, 1992, 2001). Program DEDNM is the main component of a larger landscape analysis program called TOPAZ (TOpographic PArameteriZation) (Garbrecht and Martz, 1999, 2000). The approach used by program DEDNM is similar to that of other DEM processing models that are based on flow routing concepts, but it includes enhancements for processing low relief landscapes where the rate of elevation change may be only a few meters per kilometer over large areas. The program is designed for the resolution of the USGS 7.5′ DEMs (1 meter or 1 foot elevation increments for 30 m grid cells) and targets parameters that are of

hydrologic interest. Even though program DEDNM was designed primarily for hydrologic and water resources investigations, it is equally applicable to address a variety of geomorphological, environmental, and remote sensing applications. The performance of program DEDNM is evaluated by comparing the generated watershed subdivisions, channel network, and other derived parameters to those obtained by traditional evaluation methods using USGS 7.5′ minute topographic maps.

THE APPLICATION WATERSHED

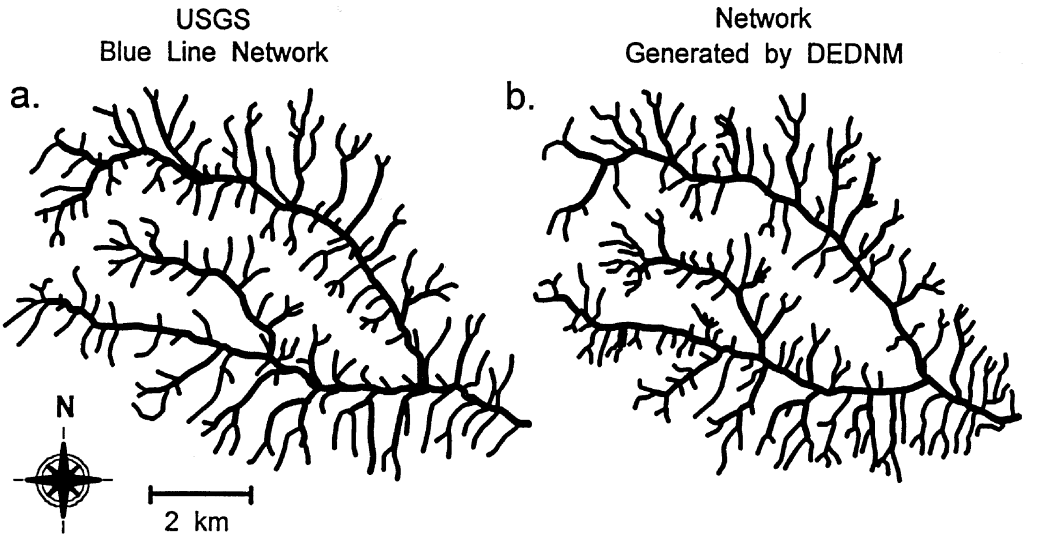
Bills Creek watershed drains an 84 km ^{2} area of the Little Washita River watershed, a USDA-ARS experimental watershed in southwestern Oklahoma. The terrain consists of gently rolling hills, and land use is predominantly rangeland with some cultivated areas. The source and the outlet of Bills Creek are at 420 m and 336 m above MSL, respectively. With a longitudinal distance of 16.5 km, the average main channel slope is about 0.005 m/m. Channel slopes as low as 0.003 m/m and flat flood plain areas extending over several hectares are not uncommon toward the outlet of Bills Creek watershed.

USGS 7.5′ DEM coverages from the geographic area under consideration are not available, and a previously generated DEM of the ARS experimental watershed is used to conduct this study. The DEM was generated in 1987 by the NASA Stennis Laboratory, Slidell, Louisiana, from USGS 7.5′ topographic maps. Elevation data are given for 30 m grid cells in 0.91 m (3-foot) increments. These DEM characteristics are similar to the USGS 7.5′ DEM coverages which have a 30 m grid spacing and a 1 m vertical resolution. A three-dimensional, vertically enhanced representation of Bills Creek watershed is shown in Figure 2.1 (see color section). With a vertical resolution of 0.91 m, parts of the gently sloping floodplains appear as flat areas and include a number of apparent depressions or pits which are artifacts of the DEM. Such spurious depressions are often encountered in low resolution DEMs of terrain with limited relief and usually arise from input errors, interpolation procedures, elevation rounding and limited DEM resolution. These produce under- and overestimation of elevation values of individual or groups of DEM cells which then can result in spurious depressions. The extent of the flat areas and depressions in Bills Creek watershed are illustrated in Figure 2.2 (see color section). They can be seen to be closely associated with channel bottom lands and drainage divides. The low channel slopes, flat areas, and spurious depressions in the DEM of Bills Creek watershed make it an appropriate low relief application for evaluating watershed segmentation and channel network generation capabilities of program DEDNM.

The channel network, defined by the blue-line channels on the USGS 7.5′ topographic maps (Figure 2.3a), is the map-based network against which the generated network is assessed. It should be recognized that a channel network is a dynamic drainage feature that can change in time as a function of climate, land use, and other land surface parameters. Thus, any channel network defined from topography alone is the reflection of past runoff and erosional activities and may not necessarily be representative of current runoff conditions. Other methods such as contour crenulation or slope analysis can also be used to define the channel network, but they also have limitations. The blue-line method (Morisawa, 1957) was selected because it is the simplest and one that most readers are familiar with, and despite its limitations, it is a standard that is readily available for most watersheds in the United States. The map-based network of Bills Creek watershed is of 5th Strahler order (Strahler, 1957) and of magnitude 180 following Shreve's (Shreve, 1967) classification.

BOUNDARY CONDITIONS FOR WATERSHED SEGMENTATION

Program DEDNM requires that two channel network parameters be specified for the automated watershed segmentation: the critical source area and the minimum source channel length. The crit-

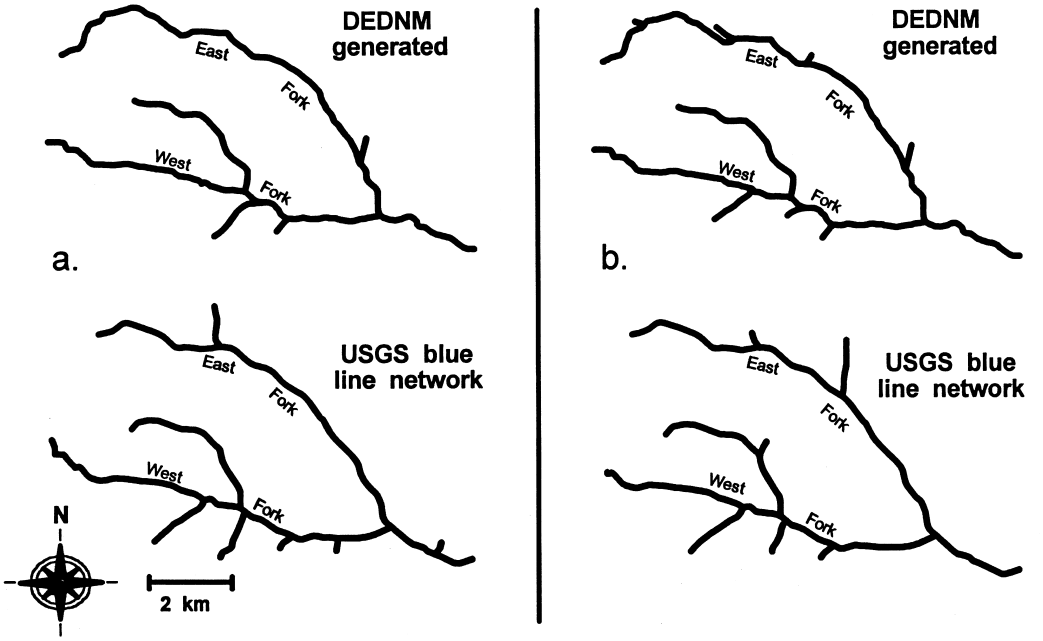


Figures 2.3a and 2.3b. Channel network of Bills Creek watershed: (a) blue-line network from USGS 7.5′ topographic maps; (b) generated network by DEDNM.

ical source area is the drainage area required to support a permanent channel. Its value is related to soil characteristics, vegetation cover, climatic conditions, and terrain slope. It also varies with map scale when maps are used as the basis for network definition (Scheidegger, 1966). For the present application, the critical source area was determined from the USGS 7.5′ topographic maps and was about 9 ha. The minimum channel length for source channels is a parameter that is necessary to control the identification of very short channels that satisfy the critical source area criterion, but have no real significance at the scale of the USGS 7.5′ topographic maps. Therefore, a threshold length below which first order channels are not generated is specified. This threshold length for Bills Creek was determined from the USGS topographic maps to be 130 m. With only these two input channel network parameters, the DEM can be processed by program DEDNM, and a segmented watershed, and channel network and subcatchment parameters can be extracted from the DEM. As part of this processing, the spurious depressions in the DEM are first removed by raising the elevation of the cells within the depressions to the elevation of the lowest outlet cell on the outside edge of the depression. The watershed segmentation and channel network and subcatchment parameter extraction is then performed on this depression-free DEM.

RESULTS

Program DEDNM generates raster maps of the channel network, subwatersheds, and other drainage parameters. The watershed segmentation is represented by the drainage boundaries of the subcatchments and the channel network. As examples of generated raster maps, the raster map of elevation contours, flow vectors, and subwatersheds boundaries, including the generated channel network, are displayed in Figures 2.4, 2.5, and 2.6, respectively (see color section). Program DEDNM also produces attribute tables that can be used for subwatershed parameterization, distributed surface runoff modeling, or other purposes. In the following discussion, network and sub-



Figures 2.7a and 2.7 b. Generated and measured channel networks for (a) channels of third and higher Strahler order; (b) channels of fifth and higher Shreve magnitude.

watershed raster and attribute data are compared to values derived from USGS 7.5′ topographic maps by traditional methods.

Evaluation by Visual Appearance

The channel network generated by program DEDNM (Figure 2.3b) is similar to the blue-line map-based network derived from the USGS 7.5′ topographic maps (Figure 2.3a). Differences are apparent, particularly at the level of source channels. Some source channels are generated at locations where there are none on the USGS maps, although contour crenulation suggests the presence of a channel in most of these cases. The reverse is also true; source channels with small upstream source area are shown on the USGS maps, but are not generated by program DEDNM. Other source channels differ by their length. The primary reasons for these differences are the ambiguity in the definition of a source of a channel, natural variation in channel characteristics within the network, and the use of only two parameters to generate the entire network. The two-parameter approach used by program DEDNM (critical source area and minimum source channel length) can reproduce average channel properties, but cannot account for random or spatial variability of these properties within the watershed. When such spatial variability is important, the user must either apply the program to each homogeneous subarea separately, or introduce a variable channel maintenance constant and minimum source channel length.

The differences between the two networks diminish for higher channel magnitude or Strahler order. This increasing similarity is demonstrated for third and higher Strahler order channels (Figure 2.7a), and for fifth and higher Shreve magnitude channels (Figure 2.7b). In both cases, the main channel, the East and West Fork, and several larger tributaries are well reproduced. The few

remaining differences are channels that exist in both networks, but are not represented because they did not meet the order or magnitude selection criterium used to generate the illustrations.

Comparison of Selected Channel Network Parameters

The selected parameters used to compare the generated and map-based channel networks are listed in Table 2.1 and represent the topologic, geometric, and hydrographic characteristics of the networks. The parameter values are representative of the average network characteristics of the entire network and do not reflect spatial variations within the watershed. As indicated earlier, spatial variability can be accounted for by either evaluating each homogeneous subarea separately, or introducing a variable channel maintenance constant and minimum source channel length.

The watershed Strahler order and the Shreve magnitude are nearly identical for the generated and map-based channel networks. The number of channel links and the link lengths are reproduced with a mean difference of less than 2%.

The channel slope values shown in Table 2.1 are length weighted values; i.e., long channels are assumed to be more representative of average network conditions. The mean slope of the generated channel links is about 23% lower than the slope derived from the map-based network. For interior links the difference is 10%, whereas for exterior links it is 27%, suggesting that exterior links are the primary reason for the discrepancy. Exterior links are the same as first order channels or source channels.

Table 2.1. Selected Network Parameters for Generated and Map-Based Networks and Corresponding Deviations

Parameter	Generated Network	Map-Based Network	Deviation %
Strahler order	5	5	0.0
Shreve magnitude	182	180	1.1
Number of channel links total	363	359	1.1
Exterior links	182	180	1.1
Interior links	181	179	1.1
Channel length [m]			
Total length	188706	187980	0.4
Mean link	520	524	–0.8
Mean exterior link	648	643	0.8
Mean interior link	391	407	–3.9
Channel slope [m/m]			
Mean	0.0170	0.0221	–23.1
Exterior links	0.0212	0.0289	–26.6
Interior links	0.0101	0.0112	–9.8
East Bills Creek	0.0055	0.0051	7.8
West Bills Creek	0.0066	0.0063	4.7
Drainage area [ha]			
Total area	8405	8442	–0.4
Mean links	23.2	23.7	–2.1
Above channel source	9.4	9.0	4.4
Drainage density [1/m]	0.0023	0.0022	4.5

The ambiguous and subjective definition of source channels, natural variations in source channel characteristics, the coarse vertical resolution of the DEM, and the use of only two parameters to generate the entire network are believed to lead to the observed differences. A more representative comparison of channel slope is achieved by computing the slope over long channel stretches, such as for the East and West Fork of Bills Creek. For these two channel stretches, the mean slope from source to watershed outlet is reproduced with less than 10% discrepancy. This illustrates that even though large differences in slope may exist for individual channel links, the average slope over long channel stretches is reasonably approximated by program DEDNM.

The total watershed area and mean direct drainage area feeding channel sources and channel links are reproduced within 4%. Finally, the drainage density of the generated network is within 5% of that of the validation network.

Comparison of Channel Network Composition

The channel network composition is quantified by the bifurcation, length, slope, and upstream area ratios, as defined by Horton (1945) and Schumm (1956). These ratios define the rate of change of a variable with channel order and incorporate a relative measure of the variable magnitude for each order. The channel network composition analysis complements the previous analysis of selected network parameters because it measures change of parameter value with channel order.

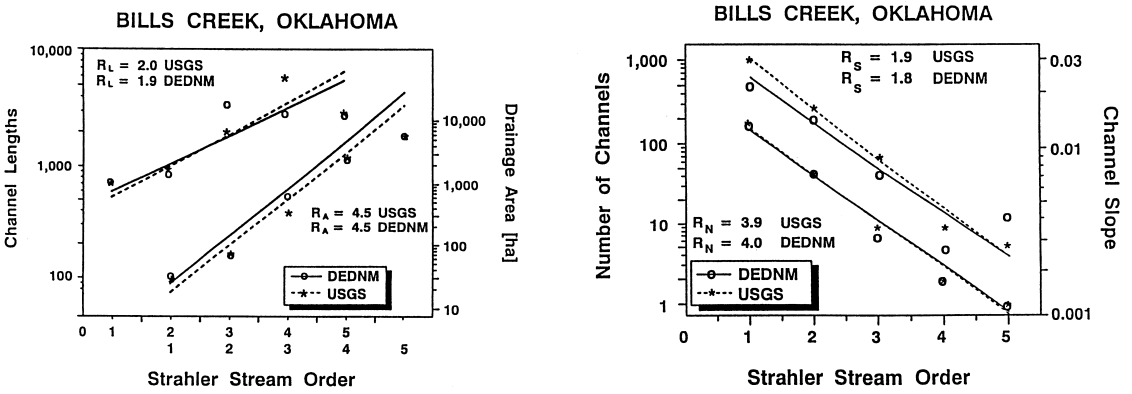
The average value of the four ratios for the generated network is within 4% of that obtained for the map-based network (Table 2.2). The slope and relative position of the regression lines in Figures 2.8a and 2.8b graphically represent the ratios. In all four cases, the regression lines are quite similar in slope and position. This close agreement shows that the general character of a system of channels is reproduced by the automated network extraction. Variations in parameter values for individual Strahler orders are primarily the result of the stochastic nature of the Strahler ordering system. As previously reported by Gregory and Walling (1973), and Scheidegger (1966 and 1970), the stochastic aspect of Strahler's ordering sometimes result in different order values for corresponding channel segments in two very similar networks. This is why direct comparisons of parameter values for individual Strahler order are not necessarily conclusive and are not performed here.

Table 2.2. Ratios of Channel Network Composition

Parameter	Generated Network	Validation Network	Deviation %
Bifurcation ratio	4.00	3.92	2.0
Length ratio	1.87	2.02	–7.4
Slope ratio	1.79	1.90	5.8
Area ratio	4.54	4.54	0.0

DISCUSSION AND CONCLUSIONS

In this chapter, the channel network generated by program DEDNM for an 84 km^2 low relief watershed is compared to the one defined by the blue-line method on the USGS 7.5′ topographic maps. The selected watershed included flat areas and depressions along the valley bottom, and flat areas near drainage divides. The depressions are usually artifacts of the DEM and have been removed by raising the elevation of the cells within the depression to the elevation of the lowest outlet cell on the outside edge of the depression. The watershed segmentation and channel network and subcatchment properties are generated from this depression-free DEM.



Figures 2.8a and 2.8b. Channel network composition: (a) length and area ratios, R_L and R_A; (b) bifurcation and slope ratios, R_N and R_S.

The visual appearances of the generated and map-based channel networks are very similar, and so are the channel network composition parameters which display an average discrepancy of less than 5%. Other selected network parameters that describe the channel network also show good correspondence. For example, the generated watershed Strahler order, the Shreve magnitude, the number of channel links, the channel link length, the drainage area, and the drainage density are within less than 5% of those of the map-based channel network. The average discrepancy for all parameters used in this study is also less than 5%. The largest differences are found for channel slope. The reason for the larger discrepancies in longitudinal channel slope values is primarily the ambiguous definition of first order channels in the map-based channel network, the coarse DEM resolution, and natural variations in channel characteristics within the network which cannot be reproduced by the two-parameter model of program DEDNM. Spatial differences between generated and map-based channel networks occur because channel generation criteria are applied uniformly to the entire network and, therefore, cannot reproduce spatial variability within the channel network. For investigations where spatial variation is important, program DEDNM should be applied to subareas that are homogeneous. In general, the close agreement between the various parameters describing the overall channel network and subwatershed characteristics demonstrates the ability of program DEDNM to overcome the problems associated with ill-defined drainage boundaries and indeterminate flow paths in low relief terrain.

Based on the experience of this application, further improvements have been introduced into program DEDNM. These include the capability to reproduce spatial variability in the generated channel network and subcatchments within the watershed and to treat spurious depressions by a combination of breaching and filling which is more appropriate given that spurious depressions arise from elevation over- and underestimation. These improvements are covered in a separate chapter of this book. Program DEDNM with these improvements and other additions are available in a software package called TOPAZ (TOpographic PArameteriZation) (Garbrecht and Martz, 1997).

REFERENCES

Band, L. E., 1986. Topographic partition of watersheds with digital elevation models. *Water Resources Research,* 22(1):15–24.

Douglas, D. H., 1986. Experiments to locate ridges and channels to create a new type of Digital Elevation Model. *Cartographica* 23(4):29–61.

Garbrecht, J., and L. W. Martz, 1999. TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation and Subcatchment Parameterization; TOPAZ Overview. U.S. Department of Agriculture, Agricultural Research Service, Grazinglands Research Laboratory, El Reno, OK, ARS Publication No. GRL 99–1, 26 pp.

Garbrecht, J., and L. W. Martz, 2000. TOPAZ: An Automated Digital Landscape Analytical Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation and Subcatchment Parameterization: TOPAZ User Manual. U.S. Department of Agriculture, Agricultural Research Service, Grazinglands Research Laboratory, El Reno, OK, ARS Pub. No. GRL 2-00, 144 pp.

Gregory, K. J., and D. E. Walling, 1973. *Drainage Basin Form and Process: A Geomorphologic Approach.* John Wiley and Sons, New York, NY.

Horton, R. E., 1945. Erosional development of streams and their drainage basins; Hydrophysical approach to quantitative morphology. *Bul. Geol. Soc. Am*., 56:275–370.

Jenson, S. K., 1985. Automated Derivation of Hydrologic Basin Characteristics from Digital Elevation Data. *Proc. Auto-Carto* 7, Digital Repres. of Spatial Knowledge, Washington, D.C., pp. 301–310.

Jenson, S. K., and J. Q. Domingue, 1988. Extracting topographic structures from digital elevation data for Geographic Information System analysis. *Photogrammetric Engineering and Remote Sensing,* 54(11):1593–1600.

Martz, L. W., and E. DeJong, 1988. CATCH: A FORTRAN program for measuring catchment area from Digital Elevation Models. *Computers and Geosciences,* 14(5):627–640.

Martz, L. W., and J. Garbrecht, 1992. Numerical definition of drainage network and subcatchment areas from Digital Elevation Models. *Computers and Geosciences,* 18(6):747–761.

Martz, L. W., and J. Garbrecht, 2001. Channel network delineation and watershed segmentation in the TOPAZ digital landscape analysis system. In Lyon, J., Ed., 2001, *GIS for Water Resources and Watershed Management.* Ann Arbor Press, Chelsea, MI.

Morisawa, M., 1957. Accuracy of determination of stream lengths from topographic maps. *Trans. Am. Geo. Un.,* 38(1):86–88.

Morris, D. G., and R. G. Heerdegen, 1988. Automatically derived catchment boundary and channel networks and their hydrological applications. *Geomorphology,* 1(2):131–141.

O'Callaghan, J. F., and D. M. Mark, 1984. The extraction of drainage networks from digital elevation data. *Computer Vision Graphics and Image Processing,* 28:323–344.

Peucker, T. K., and D. H. Douglas, 1975. Detection of surface specific points by local parallel processing of discrete terrain elevation data. *Computer Vision Graphics and Image Processing,* 4(3):375–387.

Scheidegger, A. E., 1966. Effect of map scale on stream orders. *Bul. Int. Assoc. Sci. Hydrol*., XI, 3.

Scheidegger, A. E., 1970. *Theoretical Geomorphology*. Second Revised Edition, Springer Verlag, New York, NY.

Shreve, R. L., 1967. Infinite topologically random channel networks. *J. Geol*., 75:178–186.

Schumm, S. A., 1956. Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Bul. Geol. Soc. Am*., 67:595–646.

Strahler, A. N., 1957. Quantitative analysis of watershed geomorphology. *Trans. Am. Geophys. Union,* 38(6):913–920.

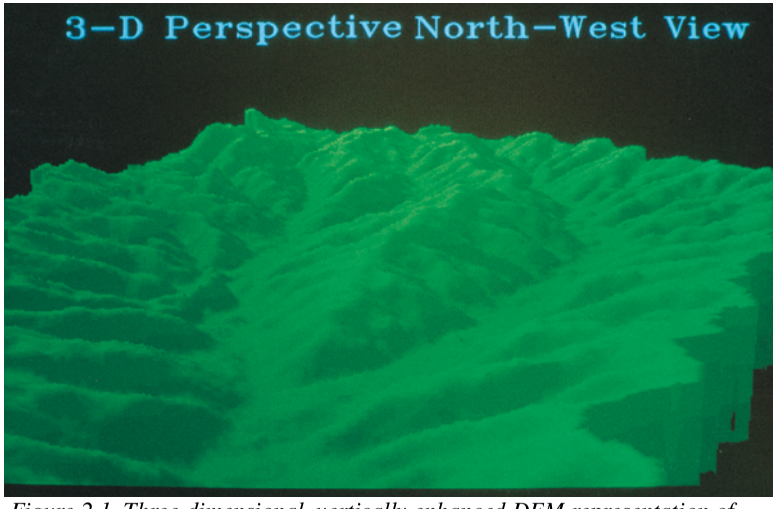


Figure 2.1. Three-dimensional, vertically enhanced DEM representation of Bills Creek watershed.

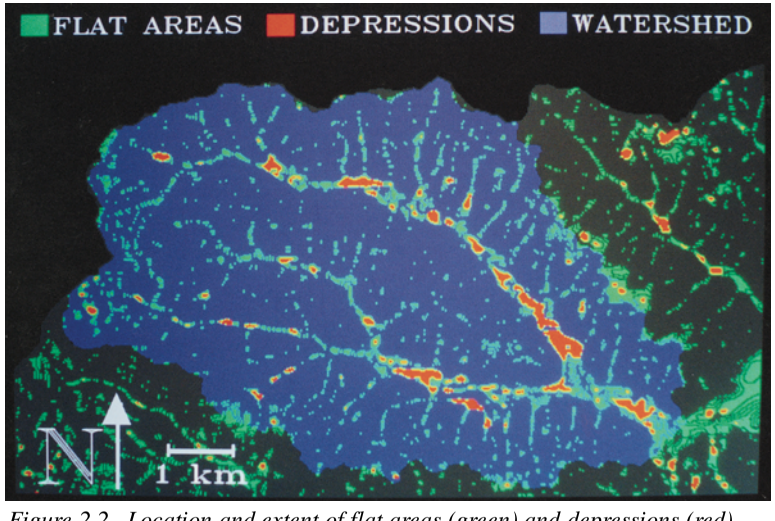


Figure 2.2. Location and extent of flat areas (green) and depressions (red) within the watershed area (blue).

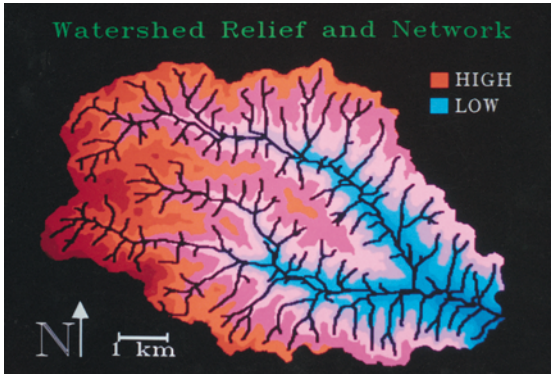


Figure 2.4. Generated elevation contours and channel network.

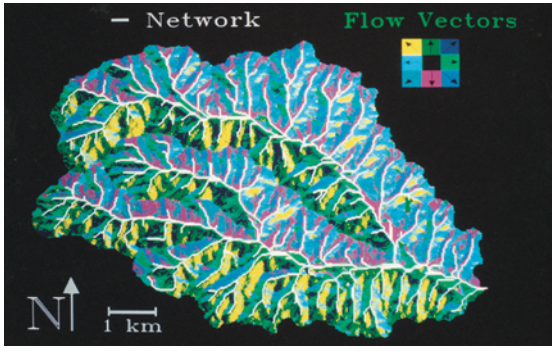


Figure 2.5. Generated flow vectors (or aspect) and channel network.

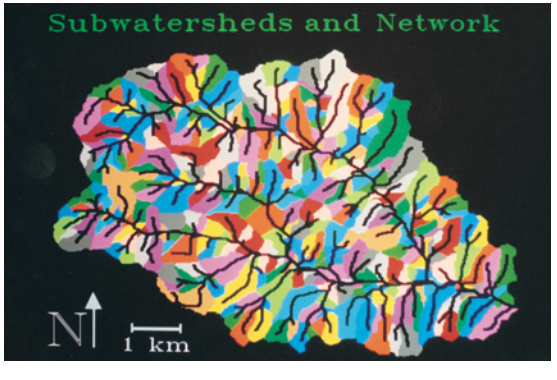


Figure 2.6. Watershed subdivision with corresponding channel network.

CHAPTER 3

Development of a Basin Geomorphic Information System Using a TIN-DEM Data Structure

Yasuto Tachikawa, Michiharu Shiiba, and Takuma Takasao

INTRODUCTION

When developing a distributed rainfall-runoff model using Digital Elevation Models, it is important to consider the method with which a spatial distribution of elevations is represented, because the method of a surface representation determines the structure of a distributed rainfall-runoff model. Three principal methods for structuring a network of elevation data are square-grid networks, contour-based networks, and triangulated irregular networks (Moore et al., 1991).

Using contour-based networks, a watershed basin can be subdivided into irregular polygons bounded by contour lines and adjacent to their orthogonals (flow trajectories) that define the boundaries of drainage areas (O'Loughlin, 1986; Moore et al., 1988). Moore and Foster (1990) modified these methods and provided a structure for modeling overland flow, TAPES. For dynamic hydrologic modeling, contour-based methods have a great advantage in considering the directions of water flow, but they need heavy data storage and much computation time.

Square-grid networks are the most common form of DEMs used for topographic analysis of a river basin (O'Callaghan and Mark, 1984; Band, 1986; Hutchinson, 1989; Tarboton et al., 1989; Takasao and Takara, 1989; Takara and Takasao, 1991), and rainfall runoff modeling (Lu et al., 1989; Wyss et al., 1990; Quinn et al., 1991). Grid-based DEMs have advantages for their ease of computational implementation, efficiency, and availability of topographic databases. However, when considering the directions of water flow, these methods are not appropriate for hydrological applications because those trajectories represent only crudely the movements of water from one grid to one of the eight neighboring grids.

A more applicable approach for hydrological modeling is the Triangulated Irregular Networks (TINs). Palacios and Curvas (1986; 1991) made it possible to delineate river-course and ridge of a watershed basin automatically with these methods and to simulate surface runoff production. Jett et al.(1979), Jones et al.(1990) and Vieux (1991) also used TINs for representation of a watershed basin.

This chapter describes a TIN-based topographic model which incorporates the advantages of grid-based methods and contour-based methods. First, a topographic surface is represented by a TIN-DEM generated by a grid-DEM and a DLG (Digital Line Graph) of river courses. Then, these triangle facets are subdivided by the steepest ascent lines (flow trajectories), so each triangle has only one side through which water flows out. Using these triangles, the discretization of a basin similar to contour-based methods is realized.

TIN-DEMS DATA STRUCTURE

In the TIN-DEMs generating system for representing a natural topography of a basin, three datasets are produced: (1) a triangle network data set, (2) a vertex data set, and (3) a channel network data set. A sample triangle data set and its network are illustrated in Tables 3.1, 3.2, and Figure 3.1. Each of the triangles, squares, and vertices is indexed by a number which is given to specify it.

The vertex data set contains the x, y, and z values of the vertices. The triangle network data set contains the properties of triangles. Each triangle is described by an index of the square in which the triangle is contained; indices of its three vertices; indices of three triangles which are adjacent to the triangle; three 'side-attribute-indices' which specify whether water flows into the side, along the side, or out of the side; three 'side-component-indices' which specify whether the side forms a part of valley, channel, slope, ridge, or boundary of a study area; and unit normal vectors of a triangular facet. The indices of vertices, the side-attribute-indices, and the side-component-indices are ordered in a counterclockwise direction.

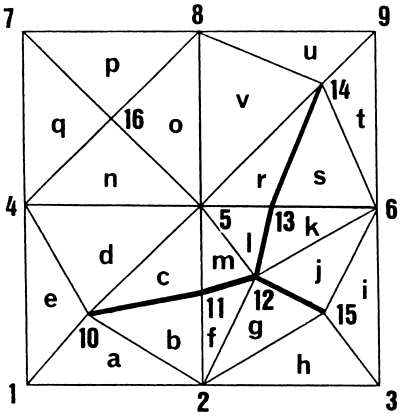


Figure 3.1. Sample triangle network.

Table 3.1. Triangle Network Data Set for Sample Triangle Network Shown in Figure 3.1

Triangle	No. of Squares	No. of Vertices	Adjacent Triangles	Side-Attribute Indices[a]	Side-Component Indices[b]	Unit Normal Vectors
a	1	1 2 10	null b e	3 1 2	0 2 2	–0.71 0.71 0.07
b	1	2 11 10	f c a	2 1 3	2 3 2	–0.71 0.71 0.07
c	1	11 5 10	m d b	3 3 1	2 2 3	–0.89 –0.41 0.09
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

[a]Side-Attribute Index: 1 = out-flow side; 2 = along-flow side; 3 = in-flow side.
[b]Side-Component Index: 0 = boundary of TIN-DEM; 1 = valley segment; 2 = slope segment; 3 = channel segment; 4 = ridge segment.

A side-attribute index of a side is set to be 1, 2, or 3, depending on whether water flows out of a side, along a side, or into a side, and the side is defined as an out-flow-side, an along-flow-side, or an in-flow-side, respectively. Whether water flows out of a side, along a side, or into a side is decided by the cross product of the steepest descent vector of a triangle and a side of the triangle. For example, in Figure 3.2 water on triangle abc flows out to the adjacent triangle through ab. In this case, the z-component of the cross product g x ab is positive, so the side-attribute-index of ab is set to be 1. Similarly, water flows into this triangle from the adjacent triangles through bc and ca. In this case, the z-components of g x bc and g x ca are negative, so side-attribute-indices of bc and ca are set to be 3. If water flows along a side, the cross product is equal to zero, and the side-attribute-index is set to be 2.

Table 3.2. Vertex Data Set

Vertex	x	y	z
1	25.00	100.00	301.25
2	50.00	100.00	287.55
3	75.00	100.00	288.89
.	.	.	.
.	.	.	.
.	.	.	.

A side-component-index of a side is set to be 0, 1, 2, 3, or

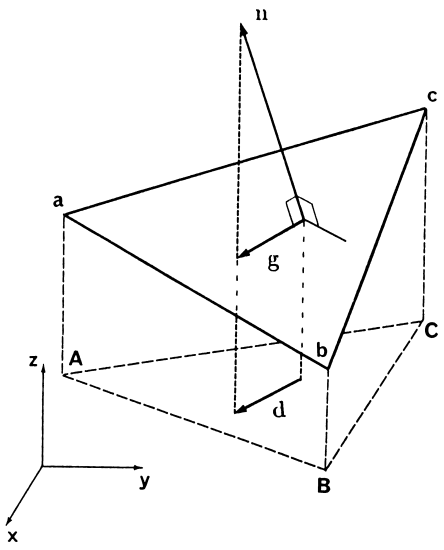


Figure 3.2. Sample triangle facet.

4, depending on whether the side constitutes a part of a boundary of a TIN-DEM, valley, slope, channel, or ridge, respectively. What value a side-component-index is set to be is decided by the side-attribute-indices of the sides which are held in common by the adjacent triangles.

If the common sides of the adjacent triangles are composed of an out-flow-side and an out-flow-side, the sides represent part of a valley. Similarly, if composed of an in-flow-side and an in-flow-side, the sides represent part of a ridge. The relation between a side-attribute-index and a side-component-index is shown in Table 3.3.

A sample channel data set and its network is illustrated in Table 3.4 and Figure 3.3. For a logical representation of a channel network in a computer, a channel network is represented by a set of links which are composed of the sections of a channel network between the terminal point of a channel network and a confluence, a confluence and another confluence, or a confluence and the upstream ends. Each link is also indexed by a number which is given to specify it. The channel network data set is represented by the index of a link, the index of the downstream link, the indices of the upstream links, the indices of vertices which form the link, and the indices of the triangles which are adjacent to the link.

BGIS (BASIN GEOMORPHIC INFORMATION SYSTEMS)

The BGIS consist of interactive software for generating TIN-DEMs data structure and topographic analysis software which contain an automatic delineation of source areas to arbitrary part

Table 3.3. The Relation Between a Side-Attribute-Index and a Side-Component-Index

	Out-Flow-Side	Along-Flow-Side	In-Flow-Side
Out-Flow-Side	Valley segment	Valley segment	Slope segment
Along-Flow-Side	Valley segment	Slope segment	Ridge segment
In-Flow-Side	Slope segment	Ridge segment	Ridge segment

Table 3.4. Channel Network Data Set

No. of Link	No. of Downstream	No. of Upstream	No. of Vertices	Right Triangle	Left Triangle
I	Null	II, III	10, 11	c	b
			11, 12	m	f
II	I	Null	12, 13	l	k
			13, 14	r	s
III	I	Null	12, 15	j	g

of a channel network and mapping of a distribution of elevations, slopes, aspects, flow path lengths, and upslope contributing areas. A schematic outline of the BGIS is shown in Figure 3.4.

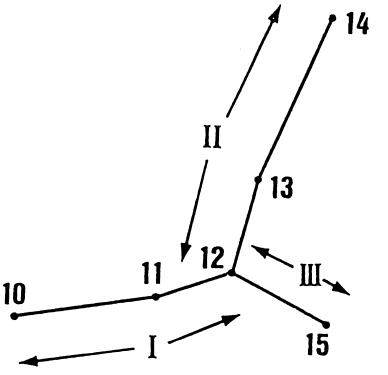


Figure 3.3. Sample channel network.

Source Data Sets

Source data sets are grid DEMs and DLGs of river courses. These data sets, produced by government agencies such as the United States Geological Survey (USGS) or the National Land Agency in Japan, are easily obtained. If source data sets for a particular study area are not available, they can be derived by digitizing contour lines and river courses on a topographic map by using a flatbed digitizer.

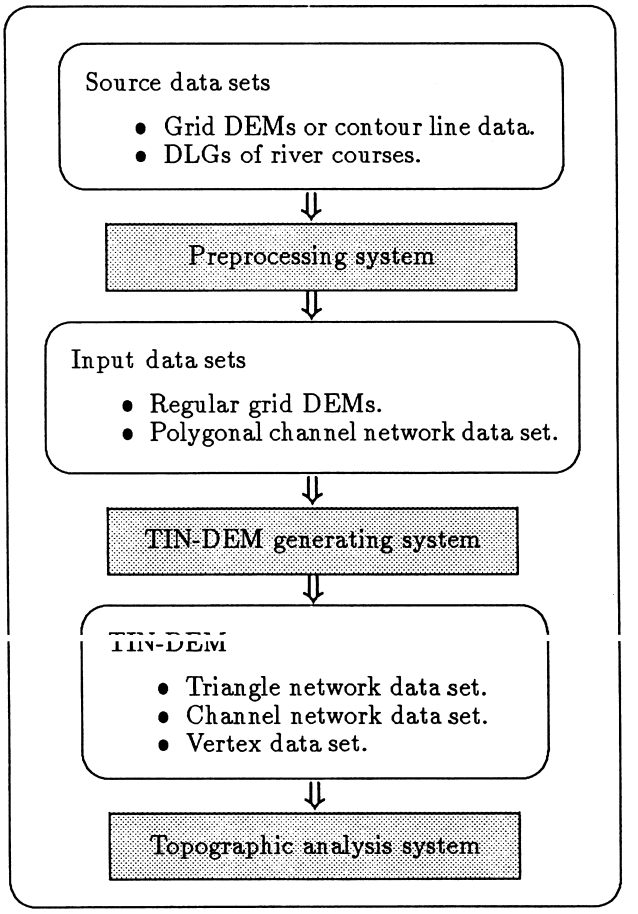


Figure 3.4. Schematic outline of BGIS.

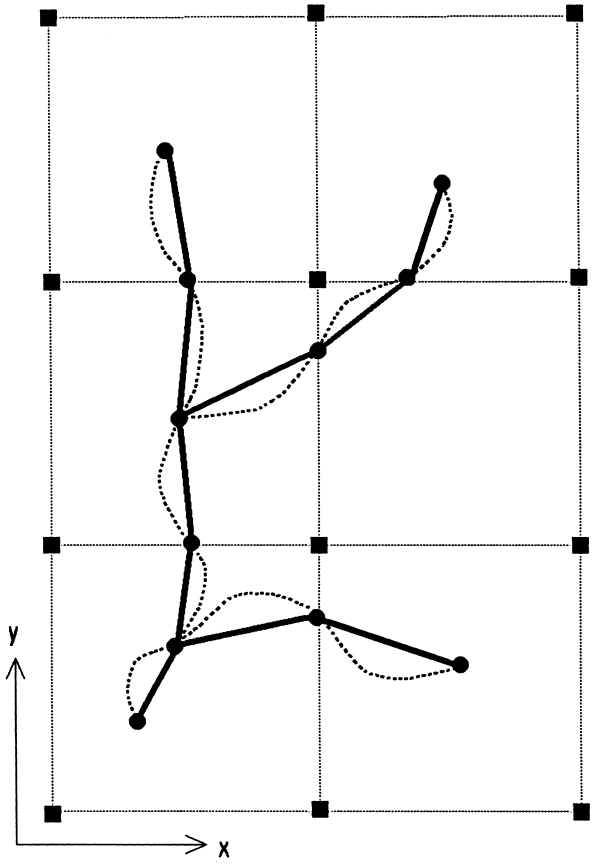


Figure 3.5. Schematic representation for making a polygonal channel network. Dashed lines denote a DLG of river courses. Solid lines denote a polygonal channel network.

Preprocessing System

From these source data sets, (1) a regular grid DEM, and (2) polygonal channel network data for a study area are produced. A regular grid DEM is interpolated from a grid DEM or contour line data. Polygonal channel network data are made up of polygonal lines which are derived by calculating the intersection of a straight line which connects two points on a regular grid DEM and a segment which connects two continuous points on a DLG of river courses (Figure 3.5).

TIN-DEMs Generating System

The two data sets made by the preprocessing systems are input into these systems and the three data sets noted in the TIN-DEMs DATA STRUCTURE section are generated. These systems include the following modules:

(a) a module for generating triangles from a regular grid DEM;
(b) a module for getting rid of pits;
(c) a module for joining discontinuous valley segments to a channel network; and
(d) a module for subdividing triangular facets.

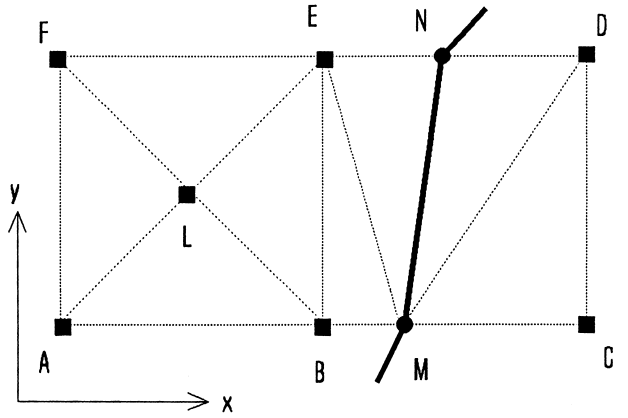


Figure 3.6. Automatic division of squares into triangles.

Module for Generating Triangles from a Regular Grid DEM

A data set which represents a basin with triangular facets based on a regular grid DEM and a polygonal channel data set are generated. For example, in Figure 3.6, the points A, . . . , F represent vertices on a grid DEM, and the segment MN represents a part of a channel network. For the square ABEF which has no channel segment in it, the point L is added in the center of the square, and it is subdivided to four triangles. The elevation of the added point L is interpolated using the elevations of four neighboring points. For the square BCDE which has one channel segment in it, it is subdivided to several triangles under the rule that the channel segment results in a side of a triangle. These two cases are processed automatically. In other cases, for example, a square which has more than one channel segment in it, and a square which has confluence points, upper ends of a channel network or a downstream end of a channel network in it, are subdivided using an interactive software. An operator can add new points if needed, make triangles manually, and see the result of a subdivision on a computer display. Figure 3.7 shows the example of a subdivision. The shaded area has already been subdivided into triangles. Squares which an operator needs to subdivide into triangles are not so many that the interactive handling of these subdivisions is not laborious and not time-consuming. After subdividing all the squares into triangles, side-attribute-indices, side-component-indices, and unit normal vectors of each triangular facet are computed, and the vertex data set, the triangle network data set, the channel network data set are produced.

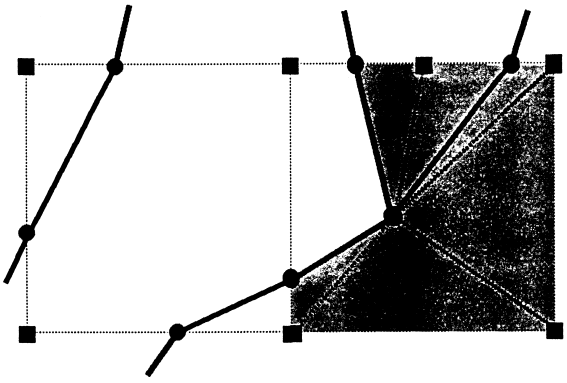


Figure 3.7. Interactive division of squares into triangles.

Module for Getting Rid of Pits

A pit is a vertex whose surrounding vertices have higher elevations. If a natural topography is so complicated to represent it using a grid DEM with a current grid spacing, sometimes false pitting occurs. In this module, a pit is found automatically and solved by adding a new

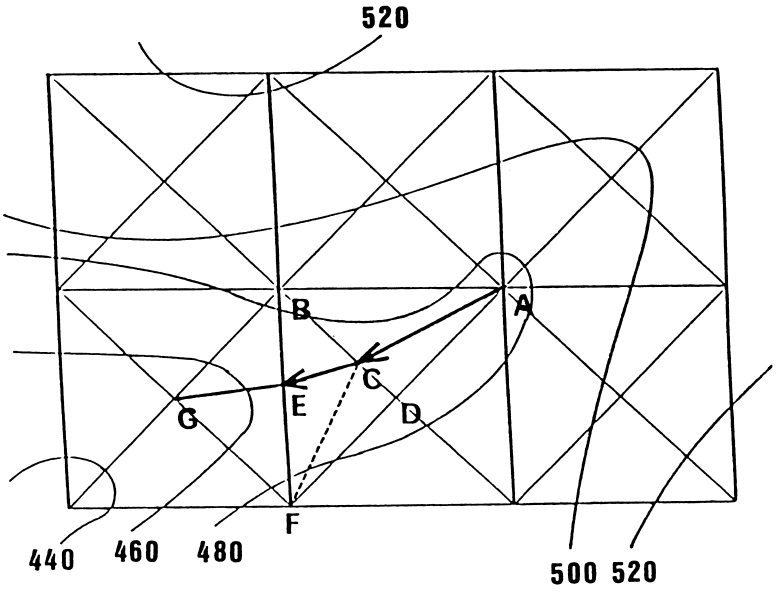


Figure 3.8. Schematic representation for getting rid of pits.

point and subdividing to triangles interactively. An algorithm for getting rid of a pit can be accomplished by following five steps:

Step 1: Find a vertex whose elevation is lower than the surrounding vertices (A, in Figure 3.8).
Step 2: Based on a topographic map, add a new point C, considering the direction of water flow, and give an appropriate elevation to the point referring to a topographic map.
Step 3: Using the point C, subdivide triangle ABD to triangle ABC and triangle ACD, triangle BFD to triangle BFC and triangle FDC.
Step 4: Update the vertex data set and the triangle network data set.
Step 5: If the new point C is a pit, return to step 1 and repeat Step 1–5 until no false pits exist.

Module for Joining Discontinuous Valley Segments to a Channel Network

In a current model of a watershed basin, many valley segments exist. If these valley segments do not join a channel network, the channel segment the triangles contribute to cannot be defined. For example, for triangle bgk and triangle ikg in Figure 3.9, the channel segment they contribute to cannot be defined. To correct this, the channel segment that the valley segments reach to is determined, after which, these valley segments are included in the channel network and the channel network is reconstructed. An algorithm for this procedure is as follows:

Step 1: Find the lowest vertex in the continuous valley segments g in Figure 3.9.
Step 2: Trace the path of steepest descent from the lowest vertex until it reaches either a channel network or the boundary of the DEM.
Step 3: If it reaches to the channel networks, subdivide to triangles along the path of the steepest descent (in Figure 3.9, triangle ceg into triangle chg and triangle heg, triangle cde into triangle cdh and triangle deh).
Step 4: Update the vertex data set and the triangle network data set.

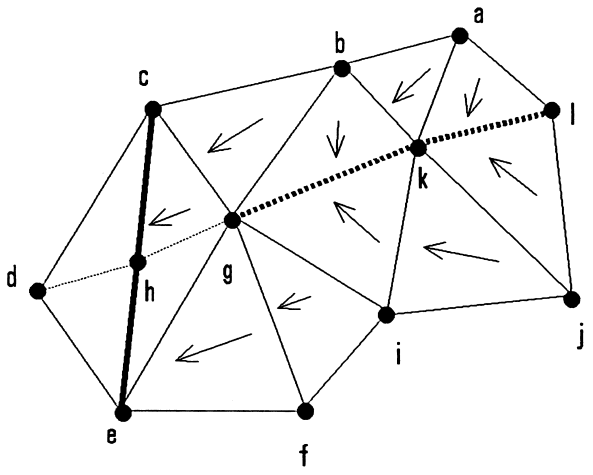


Figure 3.9. Schematic representation of discontinuous valley segment.

Step 5: Reconstruct a channel network and update the channel network data set. The channel networks before and after joining discontinuous valley to channel networks are shown in Figure 3.10.

Module for Subdivision of Triangles

Most of the triangles have two sides through which water flows out. To identify source areas, these triangles must be subdivided so that each triangle has only one out-flow-side contributing to only one adjacent triangle. An algorithm for this procedure is as follows:

Step 1: Trace a path of steepest ascent from a vertex, and find coordinates of an intersection on an opposite side.

Step 2: If the intersection is found on an opposite side (e on the segment bd in Figure 3.11), subdivide the triangle bcd to triangle bce and triangle cde, triangle abd to triangle abe and triangle aed.

Step 3: If the intersection exists on a ridge segment, stop. Otherwise continue until it encounters a ridge segment or a boundary of a TIN-DEM.

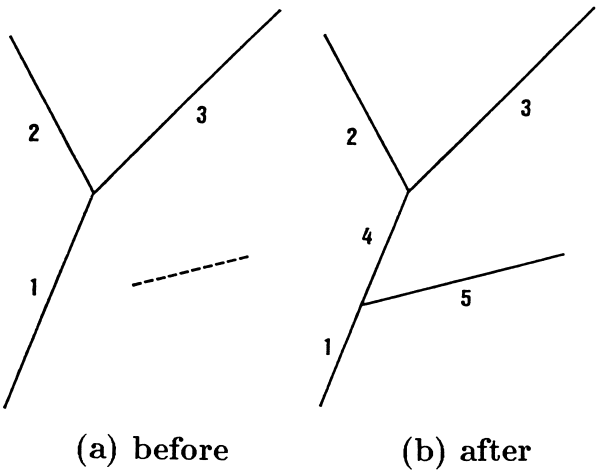


Figure 3.10. Reconstruction of channel network.

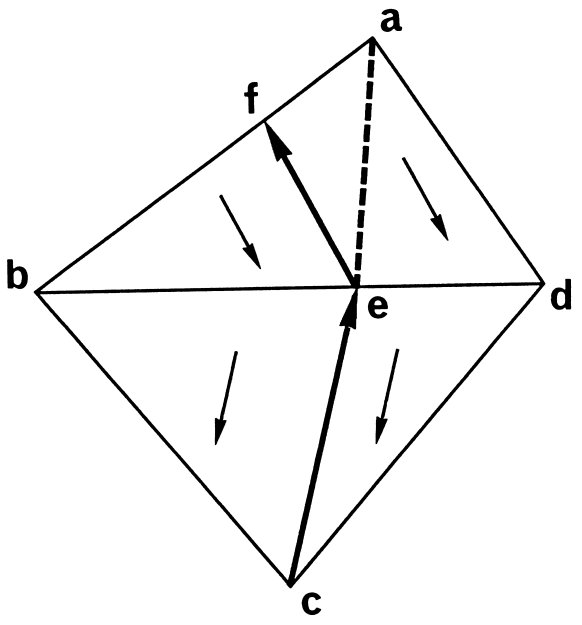


Figure 3.11. Subdivision to triangles which have one side through which water flows out.

This subdivision procedure is accomplished for all the vertices included in TIN-DEMs, but for new vertices added by this subdivision it is not necessary to apply this procedure.

APPLICATIONS

The BGIS was applied to three basins. Figure 3.12 shows a topographic map of the upper part of the Ara experimental basin. From this map, contour lines and a channel network were digitized manually by using a flatbed digitizer. Figure 3.13 shows the directions of water flow, ridges (bold solid lines), valleys (dashed lines), and the channel network (solid lines) for this study area. Figure 3.14 shows a three-dimensional representation of the basin, and the shaded areas represent the watershed basin delineated automatically.

Once the TIN-DEM data structure is generated, it is easy to identify source areas contributing to an arbitrary triangle. Each triangle has only one triangle which water flows into. When triangles which contribute to a particular triangle are found, a triangle from which water flows into it is found and added to a list of source areas recursively, until a triangle which has two ridge sides, or one ridge side and one along side, is added to the list. The channel network data set includes the numbers of the triangles which contact with a channel network, so by beginning this procedure with these triangles, all the triangles included in the watershed are identified.

Figures 3.15 and 3.16 show the Ara experimental basin (0.184 km^2) and the Ina basin (54.0 km^2). The number of vertices and triangles after processed by each module are represented in Table 3.5.

CONCLUSIONS

Geographic information systems in hydrologic modeling, the BGIS (Basin Geomorphic Information Systems) were presented for modeling a river basin using a TIN-DEM data structure. The BGIS are made up of interactive software for generating three data sets, (1) a vertex data set, (2) a triangle network data set, and (3) a channel network data set, and includes topographic analysis

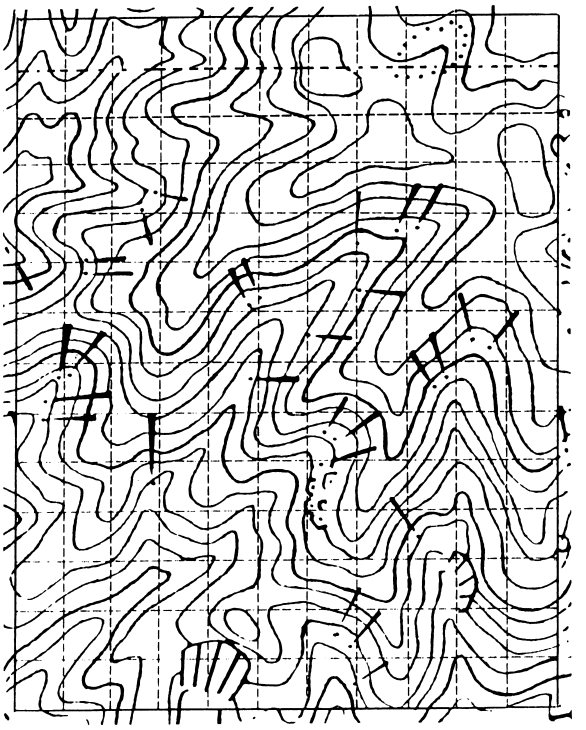


Figure 3.12. Topographic map for the upper part of the Ara experimental basin.

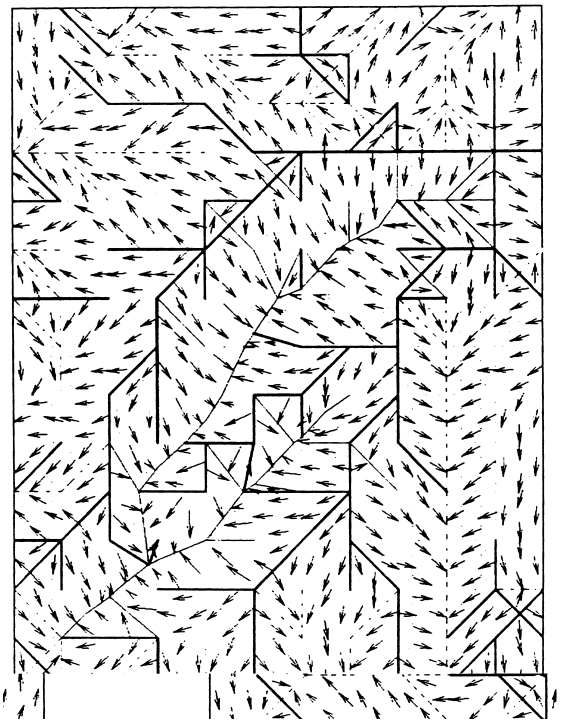


Figure 3.13. Directions of water flow of the upper part of the Ara experimental basin.

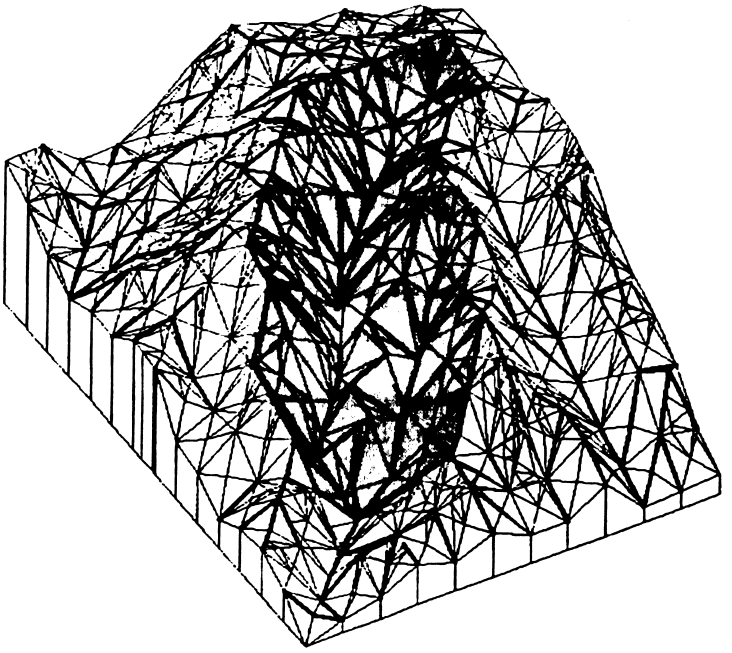


Figure 3.14. TIN-DEM for the upper part of the Ara experimental basin.

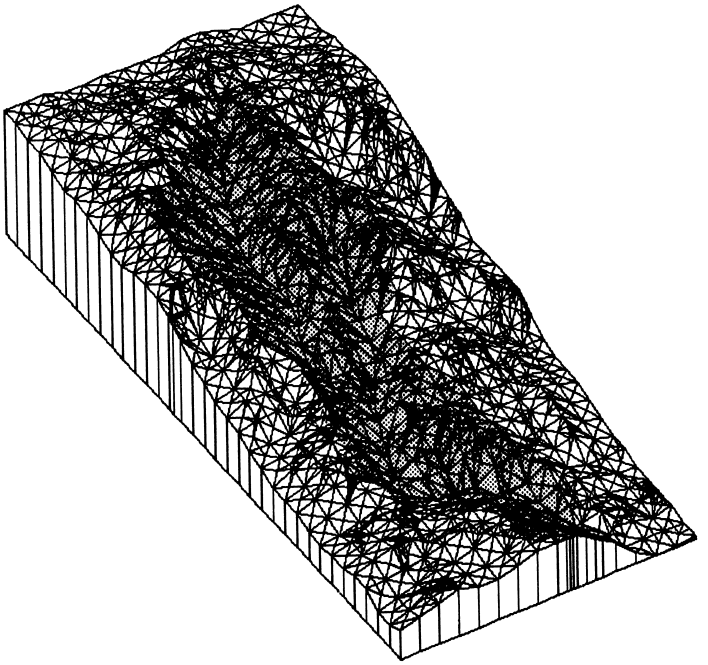


Figure 3.15. TIN-DEM for the Ara experimental basin.

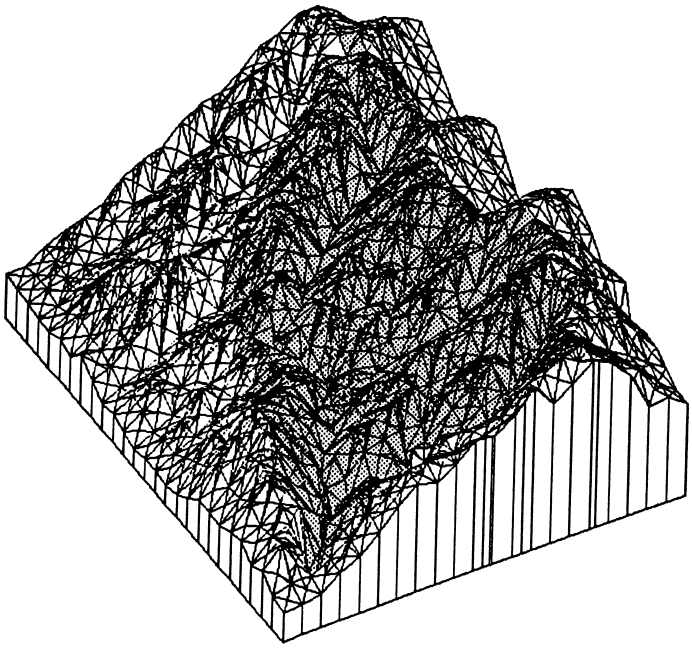


Figure 3.16. TIN-DEM for the Ina basin.

Table 3.5. The Number of Triangles and Vertices after Processed by Each Module

	Ina Basin		Ara Experimental Basin	
Module	No. of Vertices	No. of Triangles	No. of Vertices	No. of Triangles
Module for generating triangles from a regular grid DEM	1123	2146	1058	2016
Module for getting rid of pits	1151	2202	1071	2042
Module for joining discontinuous valley segments to a channel network	1185	2240	1100	2100
Module for subdividing triangular facets	3900	7682	3304	6483
Grid spacing	250 m		25 m	
Number of pits	5		2	

software which makes an automatical delineation of source areas to an arbitrary part of a channel network and mapping of a distribution of elevations, slopes, aspects, flow path lengths, and upslope contributing areas.

This TIN-based topographic model incorporates the advantages of grid-based methods with their ease of computational implementation, efficiency, and availability of topographic databases, and combines the advantages of contour-based methods such as subdivision of a basin considering the direction of water flow. This form of discretization is advantageous for modeling water movement of a natural watershed basin.

REFERENCES

Band, L., 1986. Topographic partition of watersheds with Digital Elevation Models. *Water Resources Research*, 22(1):15–24.

Hutchinson, M. F., 1989. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. *Journal of Hydrology*, 106:211–232.

Jett, S. C., A. A. Weeks, W. M. Grayman, and W. E. Gates, 1979. Geographic Information Systems in hydrologic modeling. In: *Hydrologic Transport Modeling Symposium*, 10–11 December 1979, A.S.A.E, New Orleans, LA, pp. 127–137.

Jones, N. L., S. G. Wright, and D. R. Maidment, 1990. Watershed delineation with triangle-based terrain models, *J. Hydraul. Eng*., 116(10):1232–1251.

Lu, M., T. Koike, and N. Hayakawa, 1989. A rainfall-runoff model using distributed data of radar rain and altitude. *Proceedings of JSCE*, 411(2–12):135–142.

Moore, I. D., E. M. O'Loughlin, and G. J. Burch, 1988. A contour-based topographic model for hydrological and ecological application. *Earth Surface Processes and Landforms*, 13:305–320.

Moore, I. D., and G. R. Foster, 1990. Hydraulics and overland flow. In *Process Studies in Hillslope Hydrology*, Anderson, M. G. and T. P. Burt, Eds., John Wiley and Sons, pp. 215–254.

Moore, I. D., R. B. Grayson, and A. R. Ladson, 1991. Digital terrain modelling : A review of hydrological, geomorphological and biological applications. *Hydrological Process*, 5(1):3–30.

O'Callaghan, J. F., and D. M. Mark, 1984. The extraction of drainage networks from digital elevation data. *Computer Vision Graphics and Image Processing*, 28:323–344.

O'Loughlin, E. M., 1986. Prediction of surface saturation zones in natural catchments by topographic analysis. *Water Resources Research*, 22(5):794–804.

Palacios-Velez, O. L., and B. Cuevas-Renaud, 1986. Automated river-course, ridge and basin delineation from digital elevation data. *Journal of Hydrology*, 86:299–314.

Palacios-Velez, O. L., and B. Cuevas-Renaud, 1991. SHIFT : A distributed runoff model using irregular triangular facets. *Journal of Hydrology*, 134:35–55.

Quinn, P., K. Beven, P. Chevallier, and O. Planchon, 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological Process*, 5:59–79.

Takara, K., and T. Takasao, 1991. Fractal dimension of river basins based on digital terrain maps. *Proceedings of Hydraulic Engineering*, Japan Society of Civil Engineers, 35:135–142.

Takasao, T., and K. Takara, 1989. A basic analysis of geomorphologic features of drainage basins and channel networks based on the digital national land information. *Annals of the Disaster Prevention Research Institute*, Kyoto University, 32(B–2), pp. 435–454.

Tarboton, D. G., R. L. Bras, and I. Rodriquez-Iturbe, 1989. The Analysis of River Basins and Channel Networks Using Digital Terrain Data. Dept. of Civil Engineering, M.I.T, TR No. 326, Cambridge, MA.

Vieux, B. E., 1991. Geographic Information Systems and non-point source water quality and quantity modeling. *Hydrological Process*, 5:101–113.

Wyss, J., E. R. Williams, and R. L. Bras, 1990. Hydrologic modeling of England river basins using radar rainfall data. *Journal of Geophysical Research*, 95(D3):2143–2152.

CHAPTER 4

Basinsoft, a Computer Program to Quantify Drainage Basin Characteristics

Craig A. Harvey and David A. Eash

INTRODUCTION

Surface water runoff is a function of many interrelated factors including climate, soils, land-use, and the physiography of the drainage basin. A practical and effective method to quantify drainage basin characteristics would allow analysis of the interrelations of these factors, leading to an improved understanding of the effects of drainage basin characteristics on surface-water runoff. Historically, the quantification of drainage basin characteristics has been a tedious and time-consuming process. Recent improvements in computer hardware and software technology have enabled the developers of a program called Basinsoft to automate this process. Basinsoft requires minimal preprocessing of data and provides an efficient, automated procedure for quantifying selected morphometric characteristics and the option to area-weight characteristics for a drainage basin. The user of Basinsoft is assumed to have a limited amount of experience in the use of ARC/INFO, a proprietary geographic information system (GIS). (The use of brand names in this chapter is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey [USGS].)

In 1988, the USGS began developing a program called Basinsoft. The initial program quantified 16 selected drainage basin characteristics from three source-data layers that were manually digitized from topographic maps using the versions of ARC/INFO, Fortran programs, and prime system Command Programming Language (CPL) programs available in 1988 (Majure and Soenksen, 1991). By 1991, Basinsoft was enhanced to quantify 27 selected drainage-basin characteristics from three source-data layers automatically generated from digital elevation model (DEM) data using a set of Fortran programs (Majure and Eash, 1991; Jenson and Dominique, 1988). Due to edge-matching problems encountered in 1991 with the preprocessing of the DEM data, the Basinsoft program was subsequently modified to quantify 24 selected drainage-basin characteristics from four source-data layers created from three types of data (topographic maps, digital line graph [DLG] data, and DEM data) (Eash, 1993, 1994). The early versions of Basinsoft relied primarily on Fortran programs and prime system CPL to manage data and calculate statistics, thus making them platform dependent.

In 1994, Basinsoft was redeveloped entirely using Arc Macro Language (AML), a postprocessing language written to run in ARC/INFO, to increase portability among systems using ARC/INFO version 7.0.2 or later (Environmental Systems Research Institute, 1994). The current (1997) version of Basinsoft processes four source-data layers representing selected aspects of a drainage basin to quantify selected morphometric drainage basin characteristics (Harvey and Eash, 1996). The 27 selected basin characteristics quantified by Basinsoft are listed in Table 4.1 as meas-

Table 4.1. Selected Drainage Basin Characteristics Quantified Using Basinsoft

Basin-Area Quantifications

TDA—Total drainage area[a,b], in square miles, an internal measurement maintained by ARC/INFO. TDA is acquired from the drainage-divide data layer (cover_bas) attribute table as the area measurement and it includes noncontributing areas.

NCDA—Noncontributing drainage area[a], in square miles, is the total area that does not contribute to surface-water runoff at the basin outlet. NCDA is obtained by computing summary statistics on the drainage-divide data layer (cover_bas) attribute table based on the item CONTRIB.

Basin-Length Quantifications

BL—Basin length[a], in miles, measured along a line areally centered through the drainage-divide data layer (cover_bas) from basin outlet to where the main channel extended meets the basin divide. This process uses ARC/GRID to calculate the centerline.

BP—Basin perimeter[a], in miles, measured along entire drainage-basin divide. BP is an internal measurement maintained by ARC/INFO and is acquired from the drainage-divide data layer (cover_bas).

Basin-Relief Quantifications

BS—Average basin slope[a,b], in feet per mile, measured by the "contour-band" method, within the contributing drainage area (CDA). Summary statistics are performed on the hypsography data layer (cover_con). The output from the statistics command is used in conjunction with the user-designated elevation-contour interval as input into the formula to calculate BS. BS = (total length of all selected elevation contours) (contour interval) / CDA.

BR—Basin relief[c], in feet, measured as the difference between the elevation of the highest grid cell and the elevation of the grid cell at the basin outlet. BR uses the lattice (grid) data layer (cover_lat) to determine the minimum elevation as the land-surface elevation at the basin outlet. The maximum elevation is determined from the lattice data layer (cover_lat) statistics INFO file item ZMAX.

Basin-Aspect Quantifications

BA—Basin azimuth[a,b], in degrees, compass direction of a line projected from where the main channel, if extended, meets the basin divide downslope to the basin outlet. Measured clockwise from north at 0°.

Basin Computations

CDA—Contributing drainage area[a,b], in square miles, defined as the total area that contributes to surface-water runoff at the basin outlet, CDA = TDA – NCDA.

BW—Effective basin width[a], in miles, BW = CDA / BL.

SF—Shape factor[a], dimensionless, ratio of basin length to effective basin width, SF = BL / BW.

ER—Elongation ratio[a], dimensionless, ratio of (1) the diameter of a circle of area equal to that of the basin to (2) the length of the basin, $ER = [4\ CDA / \pi\ (BL)^2]^{0.5} = 1.13\ (1/SF)^{0.5}$.

RB—Rotundity of basin[a], dimensionless, $RB = [\pi\ (BL)^2] / [4\ CDA] = 0.785\ SF$.

CR—Compactness ratio[a], dimensionless, the ratio of the perimeter of the basin to the circumference of a circle of equal area, $CR = BP/2\ (\pi\ CDA)^{0.5}$.

RR—Relative relief[d], in feet per mile, RR = BR / BP.

Channel- or Stream-Length Quantifications

MCL—Main channel length[a,b], in miles, measured along the main channel from the basin outlet to where the main channel, if extended, meets the basin divide. Summary statistics are computed on the hydrography data layer (cover_str) based on the item CODE.

TSL—Total stream length[c], in miles, computed by summing the length of all stream segments within the CDA using summary statistics on the hydrography data layer (cover_str) based on the item LENGTH.

Table 4.1. Selected Drainage-Basin Characteristics Quantified Using Basinsoft (Continued)

Channel-Relief Measurement

MCS—Main-channel slope[a,b], in feet per mile, an index of the slope of the main channel computed from the difference in streambed elevation at points 10% and 85% of the distances along the main channel from the basin outlet to the basin divide. A route system is developed based on the INFO-item CODE equal to 1 in the hydrography (cover_str) data layer. The 10% and 85% distances from the basin outlet are calculated and nodes are placed at those positions along the route. The nodes are converted to points and elevations are determined for each point from the lattice data layer (cover_lat) and attributed to a temporary data layer for use in the MCS formula. MCS = $(E_{85} - E_{10})$ / (0.75 MCL).

Channel or Stream Computations

MCSR—Main-channel sinuosity ratio[a], dimensionless, MCSR = MCL / BL.

SD—Stream density[a], in miles per square mile, within the CDA, SD = TSL / CDA.

CCM—Constant of channel maintenance[a], in square miles per mile, within the CDA, CCM = CDA / TSL = 1 / SD.

MCSP—Main channel slope proportion[c], dimensionless, MCSP = MCL / $(MCS)^{0.5}$.

RN—Ruggedness number[e], in feet per mile, RN = (TSL) (BR) / CDA = (SD) (BR).

SR—Slope ratio of main-channel slope to basin slope[c], dimensionless, within the CDA, SR = MCS / BS.

Stream-Order Quantifications

FOS—Number of first-order streams within the CDA[f], dimensionless. FOS is computed using Strahler's method of stream ordering and summary statistics on the hydrography data layer (cover_str).

BSO—Basin Stream Order[f], dimensionless, stream order of the main channel at the basin outlet. BSO is computed by intersecting the main channel with the drainage-divide data layer and determining the Strahler-stream order of the stream at the basin outlet.

Stream-Order Computations

DF—Drainage frequency[c], in number of first-order streams per square mile, within the CDA, DF = FOS / CDA.

RSD—Relative stream density[g], dimensionless, within the CDA, RSD = (FOS)(CDA)/$(TSL)^2$ = DF/$(SD)^2$.

Footnotes reference the literary sources for drainage-basin charactertistics:
[a]Modified from Office of Water Data Coordination (1978, pp. 7–9—7–16).
[b]Modified from National Water Data Storage and Retrieval System (Dempster, 1983, pp. A–24–A–26).
[c]Modified from Strahler (1958, pp. 282–283 and 289).
[d]Modified from Melton (1957).
[e]Modified from Robbins (1986, p. 12).
[f]Modified from Strahler (1952, p.1120).
[g]Modified from Melton (1958).

urements or computations. Twelve of the drainage-basin characteristics constitute specific quantifications of area, length, relief, aspect, and stream order; the other 15 characteristics are computations that make use of the various quantifications to calculate other drainage basin characteristics or statistics.

Development of the Basinsoft program is continuing. Information presented in this chapter on Basinsoft is republished in a slightly condensed and modified version (reflects revisions to the current version of Basinsoft as of July 1997) from the report "Description, Instructions, and Verification for Basinsoft, a Computer Program to Quantify Drainage-Basin Characteristics" (Harvey and Eash, 1996). Information regarding the status of this program may be obtained by accessing the PixSell Home Page at http://www.pixsell.com or by sending an email to basinsoft@pixsell.com or by contacting PixSell, Bldg. 2105, John C. Stennis Space Center, Bay St. Louis, MS 39529.

Table 4.2. "Core" AMLs and Associated Functions

AML Name	AML Function[a]
basinsoft.aml	Driver AML, initiates other core AMLs.
basinit.aml	Initializes variables and creates an empty INFO table to hold drainage basin characteristics as they are quantified by Basinsoft.
bam.aml	Quantifies TDA, NCDA, and CDA.
clm.aml	Quantifies MCL and TSL.
azim.aml	Quantifies BP and BA.
orient.aml	Calculates and sets the global variable ".orient".
basinl.aml	Quantifies BL.
dfm.aml	Quantifies FOS, BSO, and DF.
brm.aml	Quantifies BS and BR.
crm.aml	Quantifies MCS.
computations.aml	Quantifies BW, SF, ER, RB, CR, RR, MCSR, SD, CCM, MCSP, RN, SR, and RSD.

[a]See Table 4.1 for description of drainage-basin characteristic quantifications.

BASINSOFT PROGRAM STRUCTURE

The AML programs comprising Basinsoft are initialized so that each program performs specific tasks and when necessary, passes information required by the following program for subsequent processing. By programming in this modular manner, additional or future modules are easily appended to Basinsoft by the user.

The "core" AML programs comprising Basinsoft, listed with their associated functions in Table 4.2, are referred to as modules in this chapter. Basinsoft.aml controls the modules found in the Basinsoft directory (Figure 4.1) and can be modified to control some of the utility programs found in the tools directory (Figure 4.1). Modules are executed each time Basinsoft is initiated. Drainage-basin characteristics are quantified within the modules. Examples of the directory structure required by Basinsoft are listed in Table 4.3 and shown in Figure 4.2.

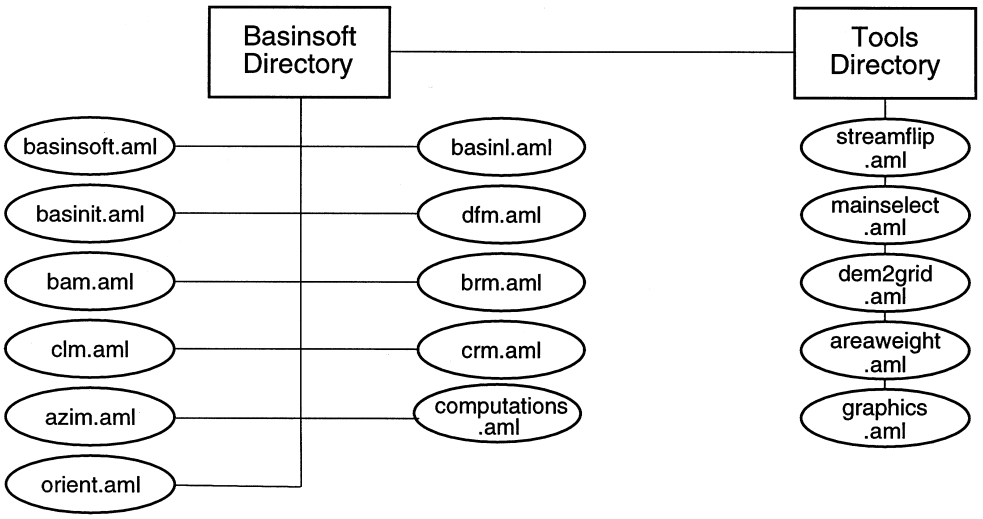


Figure 4.1. Example of Basinsoft program directory structure.

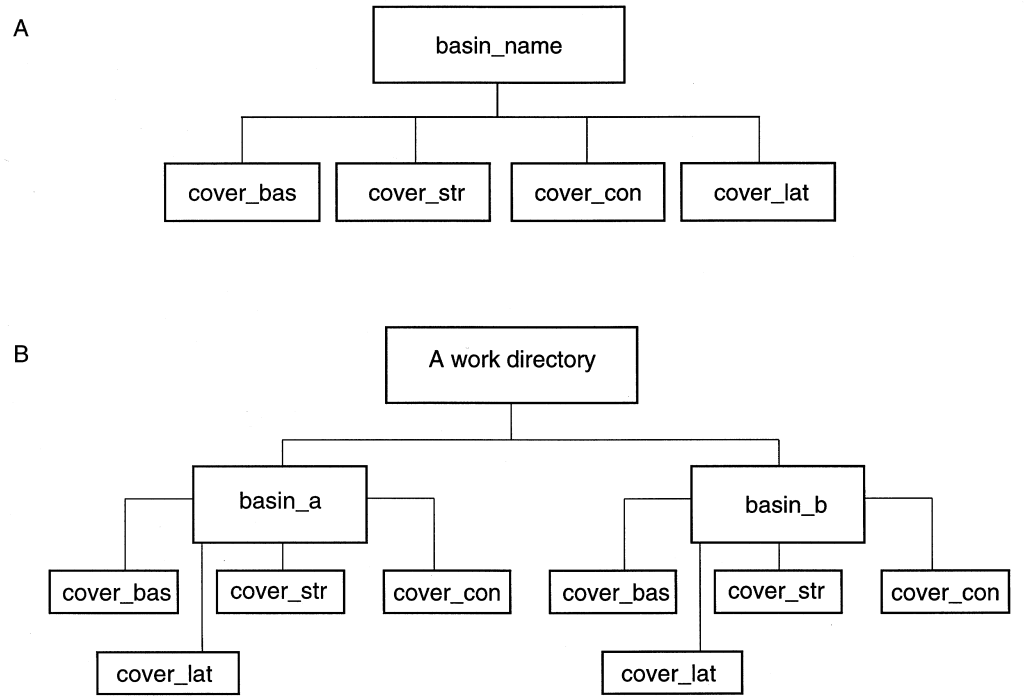


Figure 4.2. Naming conventions and directory structure required for processing multiple drainage basins—(A) generic directory structure of single basin, (B) actual directory structure for multiple basins (e.g., basin_a and basin_b) maintained under a work directory.

COMPUTER RESOURCES REQUIRED TO EXECUTE BASINSOFT

Basinsoft was developed and tested on Data General Avion model 300, 530, and 550 workstations using ARC/INFO version 7.0.2, and was subsequently ported to run on NT with ARC/INFO version 8.1. Basinsoft requires no compilation. Processing time to execute Basinsoft varies between 5 and 20 minutes on a Data General Avion 530 for drainage basins with areas less than 3,000 square miles. Processing time on other systems may vary depending on their hardware capabilities. These processing times do not include time for preprocessing, which will vary according to the type of source data available for generating the four source-data layers required as input to Basinsoft. The time to execute the TOPOGRID command may range from 5 minutes to more

Table 4.3. Naming Conventions and Description of Directory Structure Required by Basinsoft

Directory Name	Description of Directory
basin_name	This is the unique name defining a specific drainage basin (e.g., basin_a). The following data layers are source-data layers maintained under the basin_name directory.
cover_bas	Drainage-divide data layer.
cover_str	Hydrography data layer.
cover_con	Hypsography data layer.
cover_lat	Lattice data layer with areal extent clipped to drainage-basin boundary.

than 1 hour, depending on the size of the drainage basin and hardware capabilities. The basin-length module also is computer-time, memory, and disk-space intensive.

SOURCE-DATA REQUIREMENTS OF BASINSOFT

Basinsoft requires the user to provide four source-data layers. Figure 4.3 (see color section) shows examples of the four source-data layers, which are: (1) a drainage-divide (Figure 4.3A), (2) hydrography (Figure 4.3B), (3) hypsography (Figure 4.3C), and (4) a lattice elevation model (Figure 4.3D). These data layers are not required to be of any specific scale; however, the hydrography and hypsography data layers should be clipped to the areal extent of the drainage-divide data layer. The lattice elevation model is preprocessed into a lattice data layer—cover_lat, with the same areal extent as the drainage-divide data layer.

Several programs were developed to assist the user in preprocessing source-data layers in preparation for use with Basinsoft. These programs, which are found in the tools directory, perform various tasks such as creating a hydrologically accurate lattice elevation model, facilitating the creation of a hypsography data layer, and converting a 1:250,000-scale DEM to a lattice data layer. Several programs were developed to assist the user in assigning attributes to data layers according to a specialized attribute scheme required by Basinsoft.

PREPROCESSING OF SOURCE DATA FOR PROGRAM INPUT

Basinsoft was developed to run with minimal preprocessing; however, there are several steps required to prepare the source-data layers for subsequent processing by Basinsoft. Preprocessing steps include the generation of four source-data layers as described above and the assignment of attributes. The arc segments of the hydrography data layer must be edited using the streamflip.aml to orient the arc segments in a downstream direction. The main channel of the drainage basin must be delineated, attributed, and extended to the drainage-divide. AMLs were developed to assist the user with nearly all aspects of the preprocessing required by Basinsoft.

Preprocessing the Drainage-Divide Data Layer

The drainage-divide data layer (cover_bas) (Figure 4.3A) must be attributed with an INFO-item named CONTRIB. This item should have a value of ‘0’ (the default) for contributing drainage areas (CDA) and a value of ‘1’ for noncontributing drainage areas (NCDA); thus no attribute generation is necessary unless NCDA exist. A second INFO-item, BASINNAME, must be added to the drainage-divide data layer and attributed with an identifying name for the drainage basin.

Preprocessing the Hydrography Data Layer

The topology of the surface-water drainage network or hydrography data layer (cover_str) (Figure 4.3B) must be clean, and the main channel must be delineated and attributed with an INFO-item named CODE. The arc segments in the hydrography data layer (cover_str) must be oriented in a downstream direction. AML programs named streamflip.aml and mainselect.aml were developed to assist in performing these tasks and specialized attributing.

Streamflip.aml

The number of first-order streams (FOS) within a drainage basin is used to quantify three drainage basin characteristics. The number of first-order streams is determined within Basinsoft

using ARC/INFO's STRAHLER program. The STRAHLER program requires that all FROM/TO nodes of arc segments representing the hydrography be oriented in a downstream direction. The streamflip.aml was developed to help the user accomplish this editing. The AML will orient the FROM/TO nodes of each arc segment in a downstream direction.

Mainselect.aml

The main channel length (MCL) is used to quantify four drainage basin characteristics. By definition, MCL is measured along the main channel from the basin outlet to where the main channel, if extrapolated, were to meet the basin divide. The quantification of MCL requires the delineation or attribution of the main channel and, in most instances, the addition of an arc segment from the upstream node of the main channel to the drainage basin divide. An automated procedure named mainselect.aml was developed to assist in extending and defining the main channel. The mainselect.aml requires arc segments to be oriented in a downstream direction. The AML will assign a value of '1' to the INFO-item named CODE for arcs which represent part of the main channel, and a value of '0' (the default) to all other arcs in the hydrography data layer (cov_str).

Preprocessing the Hypsography Data Layer

Elevation contour or hypsography data may be available in the form of DLG data or may need to be generated from DEM data. Other sources of data, such as triangular irregular network (TIN) data or digitized vector data, may be used to generate elevation contour data.

Several steps are necessary if the hypsography data layer (cover_con) (Figure 4.3C) is to be created from DEM data. A utility AML, dem2grid.aml, located in the tools directory, was developed to convert DEM (raster) data to hypsography (vector) data. This task also can be accomplished interactively using standard ARC/INFO commands.

Preprocessing the Lattice Data Layer

If the hypsography data layer is developed using DEM data, and the utility AML dem2grid.aml is used to generate hypsography data, then the lattice elevation model (Figure 4.3D) is a by-product of executing dem2grid.aml. A lattice data layer needs to be generated. The lattice data layer, cover_lat, will have its areal extent defined by the drainage-divide data layer (cover_bas). An optional module, dem2grid.aml was developed to create the lattice data layer and the hypsography data layer (cover_con) from DEM data.

The elevation source data (lattice elevation model) (Figure 4.3D) may be unavailable, or may be available at an inappropriate scale. Under these circumstances it may be necessary to create or enhance the elevation data. One method is to use the ARC/INFO command TOPOGRID to create the lattice data layer. This command uses hypsography and a single-line hydrography network with options to use point elevation data and linear data as break lines in the topography to create a hydrologically accurate lattice data layer.

PROCESSING DATA LAYERS TO QUANTIFY DRAINAGE BASIN CHARACTERISTICS

Basinsoft processing can be performed on either single or multiple basins. The process of quantifying basin characteristics for a number of drainage basins is highly repetitive; therefore, it is possible to automate much of this task once the preprocessing has been completed. Processing

involves invoking the program Basinsoft.aml to quantify 27 selected morphometric drainage basin characteristics.

A single drainage basin can be processed from the ARC prompt. Basinsoft is executed with the appropriate basin name as a command line argument. Upon execution of Basinsoft, the drainage basin is processed and selected output is written into the basin_name directory (Figure 4.2A).

Multiple drainage-basin processing can be performed using the program multi_p.aml, an AML provided in the tools directory. The user modifies the AML to include a list of names of drainage basins to be processed. Strict adherence to the naming convention defined in Table 4.3 is necessary to automate this procedure. Upon execution, the program multi_p.aml loops through the list of basin names provided by the user, and output files are written into the basin_name directory (Figure 4.2B). The directory structure, as shown in Figure 4.2B, is needed to execute the program multi_p.aml.

BASINSOFT OUTPUT

Basinsoft output is generated in two formats. An INFO table named COVER_CHAR and a flat file named CHARS.TXT are generated in the basin_name directory (Table 4.3). The flat file is generated from the INFO table. Both files are output with the following format: characteristic_name = value. Table 4.4 shows an example output in the flat-file format. An optional graphic-output module named graphics.aml is available in the tools directory.

Table 4.4. Example of Flat File (ASCII) Output from Basinsoft

Basin Name	=	beavtr			
TDA	=	11.890	MCL	=	5.689
NCDA	=	0.000	TSL	=	15.086
CDA	=	11.890	MCS	=	23.381
BL	=	5.271	MCSR	=	1.079
BP	=	15.528	SD	=	1.269
BS	=	53.549	CCM	=	0.788
BR	=	151.246	MCSP	=	1.176
BA	=	55.063	RN	=	191.896
BW	=	2.256	SR	=	0.437
SF	=	2.336	FOS	=	6.000
ER	=	0.739	BSO	=	3.0
RB	=	1.834	DF	=	0.505
CR	=	1.270	RSD	=	0.313
RR	=	9.740			

BASINSOFT VERIFICATION

To verify the accuracy of the drainage basin characteristics quantified using Basinsoft, manual measurements of 12 selected drainage basin characteristics were made from USGS topographic maps for drainage areas upstream of 11 streamflow-gauging stations in Iowa. Manual measurements were made at scales identical to the quantifications done by Basinsoft. The results of the comparisons are shown in Table 4.5.

The Wilcoxon signed-ranks test was applied to determine the statistical significance between median manual measurements and median Basinsoft quantifications of 10 of the 12 drainage basin characteristics listed in Table 4.5. No test was performed for noncontributing drainage area

Table 4.5. Comparisons of Manual Measurements and Basinsoft Quantifications of Selected Drainage Basin Characteristics for 11 Streamflow Gauging stations in Iowa[a]

Station Number	Measurement Technique	Drainage Basin Characteristic											
		TDA[b]	NCDA	BL	BP	BS	BR	BA	MCL	TSL	MCS	FOS	BSO
05411600	MAN	177	0	27.0	73.3	166	297	144	36.4	242	5.61	84	4
	BSOFT	178	0	25.6	73.9	73.6	280	143	36.3	236	6.14	84	4
	% DIFF	+0.6	0	–5.2	+0.8	–55.7	–5.7	–0.7	–0.3	–2.5	+9.4	0	0
05414450	MAN	21.6	0	8.81	21.9	426	444	104	11.4	31.9	19.1	10	3
	BSOFT	22.3	0	8.47	21.3	264	439	104	11.1	31.5	23.1	10	3
	% DIFF	+3.2	0	–3.9	–2.7	–38.0	–1.1	0	–2.6	–1.3	+20.9	0	0
05414600	MAN	1.54	0	2.31	5.32	246	280	68	2.63	2.63	101	1	1
	BSOFT	1.53	0	2.10	5.97	208	299	75	2.58	2.58	108	1	1
	% DIFF	–0.6	0	–9.1	+12.2	–15.4	+6.8	+10.3	–1.9	–1.9	+6.9	0	0
05462750	MAN	11.6	0	4.84	15.0	157	160	52	5.74	15.2	28.3	6	3
	BSOFT	11.9	0	5.27	15.5	53.5	151	55	5.69	15.1	23.4	6	3
	% DIFF	+2.6	0	+8.9	+3.3	–65.9	–5.6	+5.8	–0.9	–0.7	–17.3	0	0
05463090	MAN	56.9	0	11.6	33.5	ND	181	91	17.4	73.5	7.27	28	3
	BSOFT	57.0	0	11.4	33.1	52.3	187	89	16.8	73.9	8.67	28	3
	% DIFF	+0.2	0	–1.7	–1.2	ND	+3.3	–2.2	–3.4	+0.5	+19.3	0	0
05470500	MAN	204	0	24.4	69.8	99.6	318	150	37.8	210	7.52	60	4
	BSOFT	208	0	25.3	67.7	49.0	309	151	35.7	192	7.81	51	4
	% DIFF	+2.0	0	+3.7	–3.0	–50.8	–2.8	+0.7	–5.6	–8.6	+3.9	–15.0	0
05481000	MAN	844	0	51.5	139	ND	303	175	88.1	685	2.04	152	5
	BSOFT	852	0	52.3	139	33.4	269	176	88.9	685	1.50	155	5
	% DIFF	+0.9	0	+1.6	0	ND	–11.2	+0.6	+0.9	0	–26.5	+2.0	0
05489490	MAN	22.9	0	10.5	24.8	289	280	70	13.3	28.4	14.8	10	2
	BSOFT	22.2	0	10.6	26.2	165	286	68	13.0	27.6	20.2	10	2
	% DIFF	–3.1	0	+1.0	+5.6	–42.9	+2.1	–2.9	–2.3	–2.8	+36.5	0	0
06609500	MAN	871	0	80.9	206	352	537	201	101	1,230	3.18	477	5
	BSOFT	869	0	81.7	210	197	491	203	99.9	1,270	3.34	487	5
	% DIFF	–0.2	0	+1.0	+1.9	–44.0	–8.6	+1.0	–1.1	+3.3	+5.0	+2.1	0
06807780	MAN	42.7	0	21.1	47.4	346	268	204	22.2	52.7	9.37	18	3
	BSOFT	42.8	0	21.5	48.8	195	283	205	22.2	55.3	10.1	19	3
	% DIFF	+0.2	0	+1.9	+3.0	–43.6	+5.6	+0.5	0	+4.9	+7.8	+5.6	0
06903400	MAN	182	0	21.9	79.0	152	224	57	39.5	228	3.24	80	4
	BSOFT	184	0	21.3	79.6	82.2	200	57	39.6	231	3.37	80	4
	% DIFF	+1.1	0	–2.7	+0.8	–45.9	–10.7	0	+0.3	+1.3	+4.0	0	0
WSRT p-value[c]		0.1192	NT	0.6248	0.2125	0.0092	0.2296	0.3447	0.0908	0.9291	0.1000	0.3742	NT

[a]TDA, total drainage area, in square miles; NCDA, noncontributing drainage area, in square miles; BL, basin length, in miles; BP, basin perimeter, in miles; BS, average basin slope, in feet per mile; BR, basin relief, in feet; BA, basin azimuth, in degrees; MCL, main channel length, in miles; TSL, total stream length, in miles; MCS, main channel slope, in feet per mile; FOS, number of first-order streams; BSO, basin stream order; MAN, manual measurement; BSOFT, Basinsoft quantification;% DIFF, percentage difference between MAN and BSOFT; ND, not determined; WSRT, Wilcoxon signed-ranks test; NT, no test performed because all values for % DIFF = 0.

[b]Manual TDA measurements are streamflow gauging station drainage areas published by the U.S. Geological Survey in annual streamflow reports.

[c]In general, p-values greater than 0.05 indicate that there is not a statistically significant difference between the median of the manual measurements and the median of the Basinsoft quantifications, using a 95% level of significance for a two-tailed Wilcoxon signed-ranks test.

(NCDA) and basin stream order (BSO) because these characteristics either were equal to zero or there was no difference between manual measurements and Basinsoft quantifications for all 11 drainage basins. Results of the statistical comparison tests indicate that Basinsoft quantifications were not significantly different (p-value >0.05) from manual measurements for 9 of the 10 drainage basin characteristics tested (Table 4.5). Basin slope (BS) was the only characteristic tested for which Basinsoft quantifications were significantly different (p-value <0.05) from manual measurements.

The results of a comparison test for average basin slope (BS) using three methods of measuring elevation contour lengths are listed in Table 4.6. The results indicated that Basinsoft quantifies basin slope with acceptable results; however, ARC/INFO was unable to generate appropriate elevation contours from 1:250,000-scale DEM data for comparison with manual measurements of elevation contours from 1:250,000-scale topographic maps. Comparisons for basin slope (Tables 4.5 and 4.6) appeared to indicate that the 1:250,000-scale DEM data were too coarse for ARC/INFO to accurately reproduce elevation contour data with the sinuosity found on the 1:250,000-scale topographic maps. Figure 4.4 shows elevation contours generated from DEM data using ARC/INFO (Figure 4.4A) are much more generalized than elevation contours digitized from topographic maps of the same scale (Figure 4.4B). Thus, the total length of contours generated from DEM data are underrepresented when compared to contours shown on topographic maps (Table 4.6). This contour overgeneralization illustrates how the total length of the elevation contours are underestimated by Basinsoft using the "contour-band" method to quantify basin slope (Table 4.1).

Understanding the reason for differences between the manual measurements and quantifications made by Basinsoft is important in determining the type of comparisons which may be relevant in analysis of this type of data. The data in Table 4.6 indicated that it was not preferable to compare Basinsoft quantifications of contour data generated from DEM data by ARC/INFO with either digitized contours or contours generated from DLG data; however, measurements based on elevation contours digitized from contour maps and those based on elevation contours generated from DLG data were similar.

Basinsoft quantifications of main channel slope (MCS) have the greatest range in percent dif-

Table 4.6. Comparison of Elevation-Contour Length Measurements Used to Quantify Basin Slope

Source Data (All Data Were 1:250,000-scale)	Method of Measurement	Elevation-Contour Length (CL) Mesurement, in Miles	Contributing Drainage Area (CDA), in Square Miles	Contour Interval (CI), in Feet	Average Basin Slope Quantified Using CL * CI/CDA, in Feet per Mile
Topographic map	Manual measurement of elevation contours from topographic map	37.23	11.89	50	156.6
Topographic map (Figure 4.4B)	Basinsoft quantification of elevation contours digitized from topographic map	35.05	11.89	50	147.4
DEM data (Figure 4.4A)	Basinsoft quantification of elevation contours generated using ARC/INFO	12.73	11.89	50	53.5

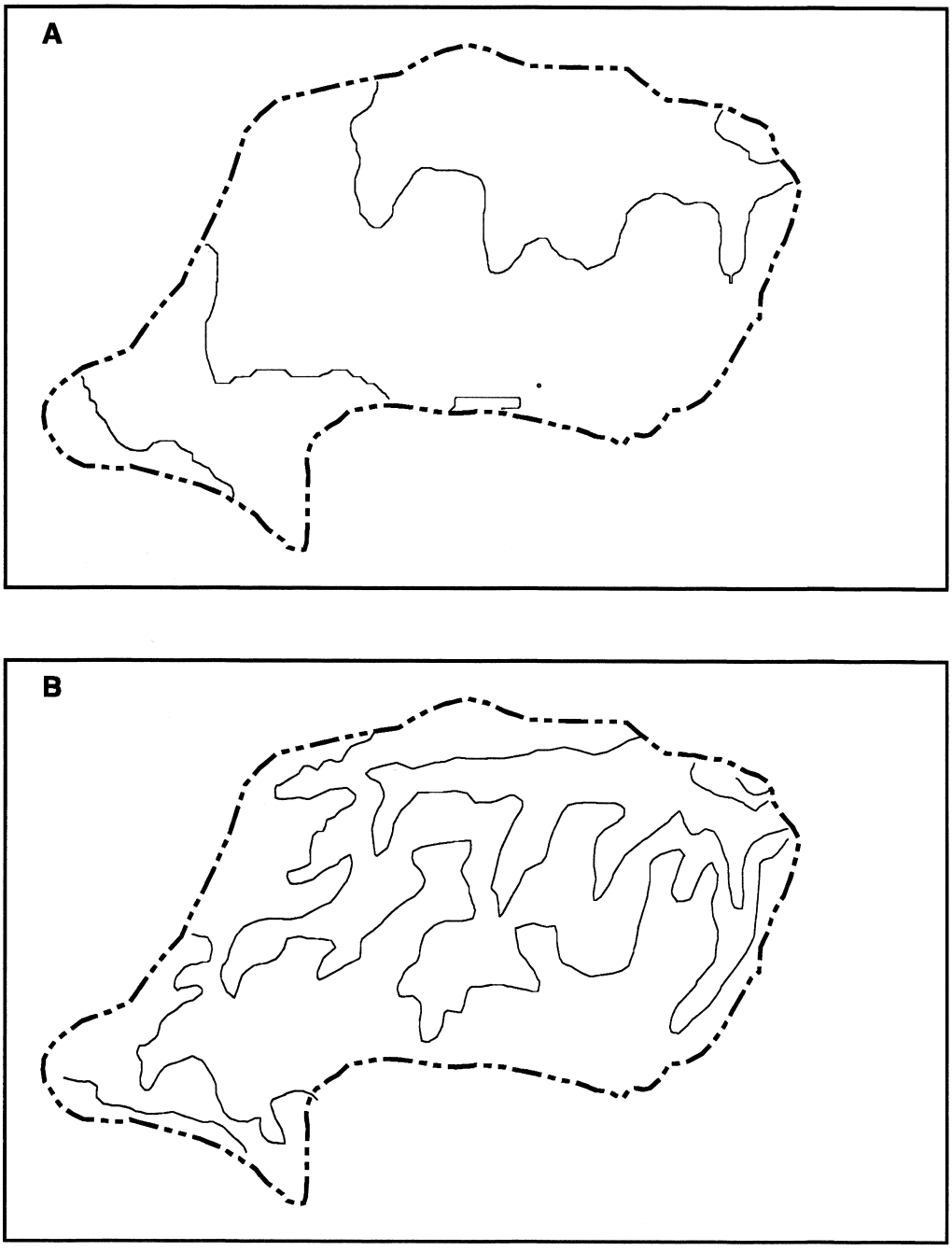


Figure 4.4 A-B. Elevation contours generated (A), with a 50-foot interval from 1:250,000–scale DEM data using ARC/INFO and (B), with a 50–foot interval digitized from a 1:250,000–scale topographic map.

ference between manual measurements and Basinsoft quantifications of the 10 selected drainage basin characteristics listed in Table 4.5. There are two main components of the MCS equation, length and elevation (Table 4.1). Quantification differences for MCS ranged from –26.5% to +36.5% (Table 4.5). However, quantification differences for main channel length (MCL) ranged from –5.6% to +0.9% (Table 4.5). These quantification differences indicated that the variation between manual measurements and Basinsoft quantifications of MCS were due mainly to differences in the determination of the streambed elevation at points 10% and 85% of the distance along the main channel from the basin outlet.

The approximate time required for an experienced ARC/INFO user to process three of the drainage basins listed in Table 4.5 using Basinsoft was 8 hours. The three basins represent large, intermediate, and small drainage areas. Manual measurements made from the same scales of topographic maps as used by Basinsoft for these three drainage basins required approximately 50 hours.

Preliminary comparisons between basin characteristic measurements made from various scales of cartographic data for these three drainage basins appear to indicate that several of the basin characteristics (BS, MCL, TSL, MCS, FOS, BSO) are map-scale dependent (Eash, 1993; Eash, 1994). Map-scale dependency refers to the effect on a measurement when that measurement is made from a cartographic data source of some specific scale as compared to that same measurement made from a different scale. Thus, interbasin comparisons of Basinsoft quantifications may be unreliable if different scales of cartographic data, or different sources of digital data (that is, raster versus vector), are used in generating the four source-data layers for each drainage basin.

Source-data layers obtained from larger-scale cartographic data and processed using Basinsoft may provide the best drainage basin quantifications for a study area. However, the scales and sources of cartographic data available for a study area may be a limiting factor in generating the four source-data layers required by Basinsoft. In general, Basinsoft can process most scales and sources of cartographic data available for a study area once the preprocessing is complete for the four source-data layers (Majure and Soenksen, 1991; Majure and Eash, 1991; Eash, 1993).

OPTIONAL PROGRAMS

Optional programs included in the tools directory (Figure 4.1) for use with Basinsoft are described below. The tools directory contains several variations of some programs, such as dem2grid.aml, which was developed to process one DEM at a time. The dem3grid.aml and dem4grid.aml process three and four DEM files at a time, respectively.

Area Weighting

A program to weight data by area, named areaweight.aml, is located in the tools directory. This program can be used as a module or as a stand-alone program to quantify characteristics from a variety of data, such as climatic data (annual precipitation and the two-year 24-hour precipitation intensity), which are stored in multipolygonal data layers (Eash, 1993). Minimal preprocessing is needed to execute this module. The areaweight.aml requires statewide or large area multipolygonal data layers representing the distribution of a characteristic, such as precipitation values, land-use values, pedologic values, geologic values, etc. The multipolygonal data layer must be larger than and encompass the drainage-divide data layer. Output is written to an auxiliary file specified by the user on the command line.

Dem2grid

The optional dem2grid.aml preprocessing program, located in the tools directory, is designed to convert a DEM file to a lattice data layer and project it into a user-specified projection. This module creates the required lattice data layer. There are several variations of this program in the tools directory. The various forms of dem2grid simultaneously preprocess up to four DEM files at a time.

Graphical Output

The graphical output module, graphics.aml, is located in the tools directory and can be used as a module or as a postprocessing program. Graphical output can be used in the interpretation of output generated by Basinsoft and can be useful in quality-control checking for obvious errors. The graphical output is in the form of an ARC/INFO graphics file which can be readily converted to a postscript file. Figure 4.5 (see color section) shows an example of graphical output. Data contained in the graphical output include: (1) tables of the basin characteristics quantified, (2) variables used by Basinsoft, and (3) a plot of the drainage-divide data layer, elevation contours, main channel, basin-length measurement line, points depicting the 10% and 85% distances along the main channel from the basin outlet, and a point at the basin outlet (outfall).

CONCLUSIONS

A computer program named Basinsoft has been developed by the U.S. Geological Survey to quantify 27 morphometric drainage basin characteristics using a geographic information system. Basinsoft was developed entirely using AML to increase portability among systems. Basinsoft uses ARC/INFO AMLs written for ARC/INFO version 7.0.2 or later. Basinsoft requires four source-data layers representing the drainage-divide, hydrography, hypsography, and a lattice elevation model of a selected drainage basin. Minimal preprocessing is required to prepare the source data used by Basinsoft. Optional programs and modules were developed to enhance the usability and functionality of Basinsoft.

Statistical comparison tests indicate that Basinsoft quantifications are not significantly different (p-value >0.05) from manual topographic-map measurements for nine of ten basin characteristics tested. Results also indicate that elevation contours generated by ARC/INFO from 1:250,000-scale DEM data are substantially overgeneralized when compared to elevation contours shown on 1:250,000-scale topographic maps and that quantification of basin slope thus is underestimated using the DEM data. A comparison test indicated that the Basinsoft module used to quantify basin slope is valid and that differences in quantified basin slope are due to source-data differences.

Basinsoft provides an automated computer procedure for the quantification of drainage basin characteristics and reduces the amount of time required to quantify drainage basin characteristics when compared to manual methods of measurement.

ACKNOWLEDGMENTS

The authors express their gratitude to the Spatial Data Support Unit (SDSU), Office of the Assistant Chief Hydrologist for Technical Support, USGS, for providing support in the development of Basinsoft; especially to Bob Pierce, Chief, SDSU, for providing technical support to the project, and to Mark Negri, Computer Specialist, SDSU, for providing programming expertise during the development of Basinsoft.

REFERENCES

Dempster, G. R., 1983. Instructions for Streamflow/Basin Characteristics File. U.S. Geological Survey National Water Data Storage and Retrieval System (WATSTORE), Vol. 4, Chap. II, sect. A.

Eash, D. A., 1993. Estimating Design-Flood Discharges for Streams in Iowa Using Drainage-Basin and Channel-Geometry Characteristics. U.S. Geological Survey Water-Resources Investigations Report 93–4062.

Eash, D. A., 1994. A Geographic Information System Procedure to Quantify Drainage-Basin Characteristics. *Water Resources Bulletin*, 30(1):1–8.

Environmental Systems Research Institute, Inc. (ESRI), 1994. ARC/INFO Users Guide, Version 7.0. Redlands, CA.

Harvey, C. A., and D. A. Eash, 1996. Description, Instructions, and Verification for Basinsoft, a Computer Program to Quantify Drainage-Basin Characteristics. U.S. Geological Survey Water-Resources Investigations Report 95–4287.

Jenson, S. K., and J. O. Dominique, 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric Engineering and Remote Sensing,* 54, (11):1593–1600.

Majure, J. J., and D.A. Eash, 1991. An automated method to quantify physical basin characteristics. In *U.S. Geological Survey Toxic Substances Hydrology Program*, G. E. Mallard and D. A. Aronson, Eds. Proceedings of the Technical Meeting, Monterey, CA, March 11–15, 1991. U.S. Geological Survey Water-Resources Investigations Report 91–4034, pp. 558–561.

Majure, J. J., and P. J. Soenksen, 1991. Using a geographic information system to determine physical basin characteristics for use in flood-frequency equations. In *U.S. Geological Survey National Computer Technology Meeting*, B. H. Balthrop and J. E. Terry (Eds.) Proceedings, Phoenix, AZ, November 14–18, 1988. U.S. Geological Survey Water-Resources Investigations Report 90–4162, pp. 31–40.

Melton, M. A., 1957. An Analysis of the Relations Among Elements of Climate, Surface Properties, and Geomorphology. Office of Naval Research, Geography Branch, Columbia Univ. Dept. of Geology, New York, Technical Report 11, Project NR 389–042.

Melton, M. A., 1958. Geometric properties of mature drainage systems and their representation in an E_4 phase space. *Journal of Geology*, 66:35–54.

Office of Water Data Coordination, 1978. Physical basin characteristics for hydrologic analysis. In *National Handbook of Recommended Methods for Water-Data Acquisition*. U.S. Geological Survey, Reston, VA., Chap. 7.

Robbins, C. H., 1986. Techniques for Simulating Flood Hydrographs and Estimating Flood Volumes for Ungaged Basins in Central Tennessee. U.S. Geological Survey Water-Resources Investigations Report 86–4192.

Strahler, A. N., 1952. Hypsometric (area-altitude) analysis of erosional topography. *Bulletin of the Geological Society of America*, 63:1117–1142.

Strahler, A. N., 1958. Dimensional analysis applied to fluvially eroded landforms. *Bulletin of the Geological Society of America*, 69:279–300.

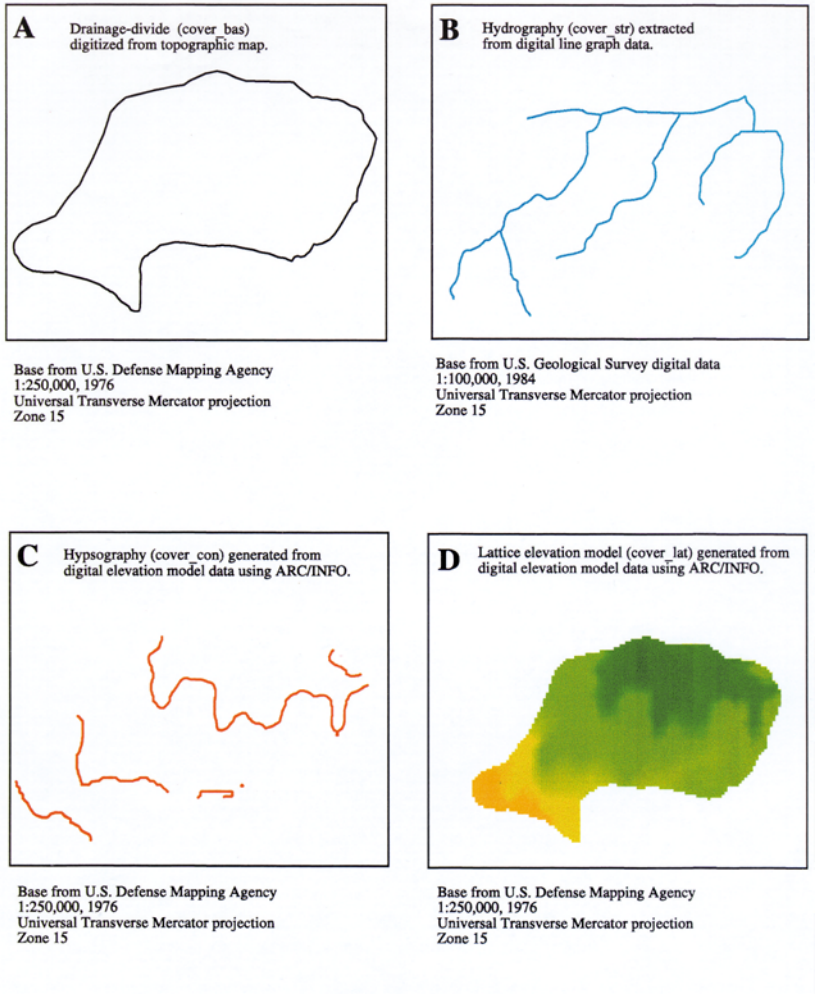


Figure 4.3 A–D. Examples of four source-data layers required by Basinsoft—(A) drainage-divide data layer, (B) hydrography data layer, (C) hypsography data layer, and (D) lattice elevation model data layer.

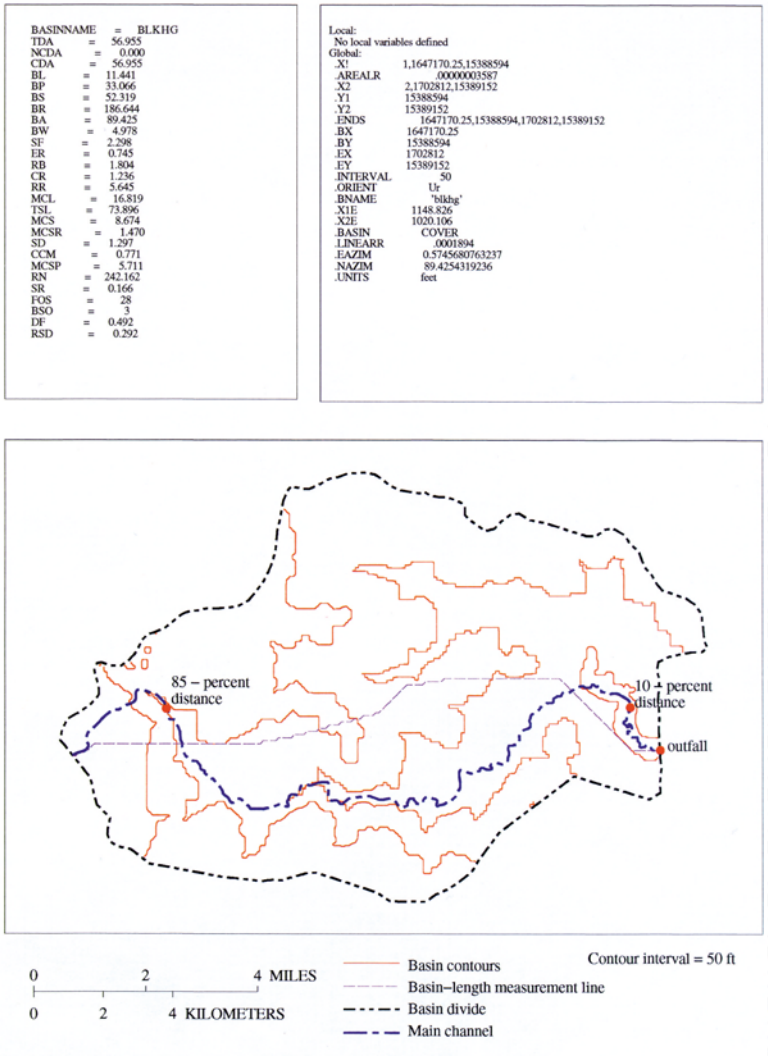


Figure 4.5. Examples of graphical output from Basinsoft.

CHAPTER 5

Deriving Stream Channel Morphology Using GIS-Based Watershed Analysis

Scott N. Miller, D. Phillip Guertin, and David C. Goodrich

INTRODUCTION

Because of the time and degree of technical skill required for the completion of geomorphology studies, individual projects have historically been limited in size and scope. With the advent of geographic information systems (GIS), these technical problems have been assuaged. The GIS capability of storing large and diverse quantities of spatial data allows for the complex analysis of many sites to be carried out quickly, efficiently, and with a high degree of repeatability (Burrough, 1986). However, GIS-based projects often fail to integrate field-collected data with GIS spatial data. This project was designed to relate the GIS characterization of spatially distributed watershed characteristics with field measurements of point-attribute data (channel cross-section surveys). These data sets were related using statistical analysis to derive relationships between watershed characteristics and channel shape.

Watershed characterization based on geometric and physical properties was carried out in a GIS on 222 subwatersheds within the Walnut Gulch Experimental Watershed. At the same time a field measurement program was completed in which channel shape characteristics were measured at the outlet of each subwatershed. Statistical analysis between the two data sets showed a strong relationship between channel shape and watershed characteristics. It was shown that the derivation of hydrologic model parameters may be effectively carried out in a GIS on a large number of data points in a relatively short amount of time.

With its long history of data collection and observational data, Walnut Gulch serves as an excellent location on which to conduct research into geomorphologic and hydrologic processes (Renard et al., 1993). Relatively little work, however, has focused on the characterization of the entire watershed. This lack of data has limited the ability to model landscape processes on a basin scale (Lane et al., 1994). Additionally, most of the research on the relationship between channel and watershed characteristics has been conducted on intermittent and perennial streams. A knowledge gap therefore exists for this type of data on aridland watersheds and the processes acting on ephemeral channel systems such as exist on Walnut Gulch (Osterkamp et al., 1983; Lane et al., 1994). Instead of limiting this work to a small section of the watershed, it was decided to characterize as much of the watershed as possible. Sample sites were located randomly across the entire 148 km^2 watershed within all soil types and many hydrologic conditions. Strahler ordering analysis (Strahler, 1952) and other measures of channel and watershed characteristics were utilized to describe the watershed as quantitatively and thoroughly as possible.

Analyses of basin characteristics have been carried out in a GIS environment for many years (Burrough, 1986; Garbrecht and Martz, 1995), but many of these processes were found to be

incompatible or unworkable for the data collected during this project. Therefore, a suite of GIS analysis tools and ARC/INFO Macro Language (AML) programs was developed to facilitate the GIS investigation (trade names are mentioned solely for the purpose of providing specific information and do not imply recommendation or endorsement by the U.S. Department of Agriculture). Channel shape, required for hydraulic routing, cannot be accurately predicted (or extracted) from DEMs. Therefore a principal goal of this project was to develop a methodology for predicting channel shape from watershed characteristics that could be readily derived from commonly available GIS coverages. During this process, field research was synthesized with GIS applications and photogrammetry to more thoroughly describe the channel and geomorphologic characteristics than had previously been attempted.

SITE DESCRIPTION OF WALNUT GULCH

Located in southeastern Arizona (approx. 110°W, 31°45′N) and comprised of rolling hills and some steep terrain, the elevation of Walnut Gulch Experimental Watershed ranges between 1190 and 2150 m A.M.S.L. Some urbanization exists in and around the town of Tombstone, but cattle grazing and recreational activities are the major land uses. Vegetation within the watershed is representative of the transition zone between the Chihuahua and Sonoran deserts, and is composed primarily of grassland and shrub-steppe rangeland vegetation.

Underlying Walnut Gulch is the geology of a high alluvial fan contributing to the San Pedro River watershed (Renard et al., 1993). Due to the enormous thickness and extent of the alluvial fill, the groundwater reserves are substantial, and can be found at depths ranging from 50 to 145 m (Libby et al., 1970). Some geologic control over the hydrology exists in the western regions of the watershed where metamorphic and orogenic activity has resulted in the fracturing and faulting of the bedrock. In 1994 the USDA Soil Conservation Service completed a detailed soil survey, finding that the watershed is dominated by 30 principal soil types (Breckenfield et al., 1995). Major soil units are Elgin-Stronghold (*Ustollic Paleargid, Ustollic Calciorthid*), Luckyhills-McNeal (*Ustochreptic Calciorthid*), McAllister-Stronghold (*Ustollic Haplargid, Ustollic Calciorthid*), and Tombstone (*Ustollic Calciorthid*).

The climate of Walnut Gulch can be classified as semiarid or steppe. Mean annual temperature in the city of Tombstone is 17.6°C, with a mean annual precipitation of 324 mm. Annual precipitation is highly variable in both timing and total depth. Rain occurs mainly during two seasons: summer rains are the product of monsoonal, highly localized, convective storms; winter rains are generally low-intensity events that cover a larger proportion of the watershed. The majority of runoff occurring on Walnut Gulch is the product of summer storms, and is therefore episodic and of relatively high intensity (Renard et al., 1993).

FIELD DATA COLLECTION

A field measurement program was undertaken wherein 222 channel cross-sections were surveyed for morphometric assessment. To account for basin-scale variability, a large number of randomly selected points were used, and multiple measurements were taken at each site to account for local variability in channel shape. These randomly located sample point locations were prestratified by soil type using a GIS procedure: each major soil type was assigned a weighted proportion of the sample points based on the areal extent of the soil coverage. At each site three cross-section surveys were taken to characterize the channel section just above the outlet of the subwatershed. Width and depth were measured at breakpoints (changes in slope). The three sur-

veys were then combined to determine the average width and depth for the channel segment, and these results were combined to derive average cross-sectional area.

A strict protocol was followed at each sample location in order to ensure proper measurement and consistency between sites. Upon arrival at a site, an inspection of the bank morphology, vegetation, and soil characteristics along the entire reach was completed to ensure that cross-sections were located where they would be most representative of the channel section. A site description was recorded in a logbook for future analysis, and potential problems related to channel complexity and morphology were noted where applicable. Bankfull indicators, including slope breaks, changes in bed or bank materials, a shift in vegetative type, debris lines, and bank staining were noted in order to determine the bankfull depth (Osterkamp et al., 1983; Gordon et al., 1992; Harrelson et al., 1994). Wherever possible, evidence indicative of a constructive, rather than destructive process, was used to determine bankfull depth. In the southwestern United States, channel processes are governed by rapid and violent runoff events, and many of the channels on Walnut Gulch are actively degrading. Channels that were clearly degrading and out of equilibrium were not subjected to channel measurement since an adequate determination of bankfull depth was not possible.

At each of the site locations a minimum of three cross-sections was surveyed. If the channel reach was complex, up to five cross-sections were measured to ensure adequate representation. At each of the cross-sections a light line was pulled level across the channel top at the bankfull depth. The line was leveled and pulled taut to reduce sag. Measurements of channel depth and distance from the left bank (looking upstream) were taken at each break in slope across the cross-section.

Channel width was more easily measured with precision than channel depth. Although determining the stage to which floodwaters rise proved difficult, the possibility for error was greater when measuring depth. This is due to a number of factors. First, depth was only measured at break points, which are to some degree subjective. Second, there was always a slight amount of sag in the line when it was stretched across a channel, lending a source of imprecision to the depth measurements. Third, more random deposition or scour of the stream channels tends to impact local channel depth measurements to a greater degree than width measurements.

GIS DATABASE DEVELOPMENT AND ANALYSIS

Given Walnut Gulch's history as a research site into various aspects of hydrology and natural resource management and its extensive rainfall and runoff database, it was decided that the GIS database would be created at a resolution not ordinarily attempted. Throughout the database development, an answer to a basic question was sought: what are the highest levels of precision and accuracy that could be achieved? There can be a tendency by GIS developers to overestimate the level to which data may be discretized. By attempting to create maps with a higher resolution than is allowable by the data, errors may be introduced, and a false level of analysis can be attempted (Burrough, 1986). Fortunately, data available for Walnut Gulch were of a quality that allowed for a very high level of resolution and positional accuracy.

Of particular relevance was the stream channel coverage. In many GIS studies, the channel network is derived from a DEM in a raster environment and then translated into vector data. Alternatively, channels may be digitized from USGS topographic maps, but channels drawn on these maps are partly based on DEMs. Traditional GIS technique dictates that the majority of channels be digitized as single vectors bisecting the channel position, with a few of the larger drainages characterized with two lines to illustrate relative width. Since a correlation was to be made in this study between channel shape and watershed variables, a channel network map was constructed whereby only the smallest channels were digitized as single vectors. Channels wider than

approximately 1.5 meters were drawn as polygonal features. This highly detailed procedure relied on the 1:5,000 orthophotographs as the base from which the stream positions and characteristics could be extracted. Most of the channels on Walnut Gulch were thus characterized in the GIS database as polygons, with associated width characteristics. In addition, where channel islands and bars were visible on the orthophotographs they were digitized. Thus, the channel network theme layer provides a detailed record of the channel system and its hydrologic characteristics as existed at the time the aerial photographs were taken (April 1988).

An important variable for the understanding of geomorphologic relationships is stream order. In this case the intensive channel network map was a drawback: because most of the channels were digitized as polygonal features it was not possible to automatically order the streams. To take advantage of GIS arc-node topology, the stream channels were vectorized. First, the map was translated into GRASS and rasterized with a one-meter resolution. The GRASS module "r.thin", which draws a parallel bisector through polygons, was executed on the stream map (Geographic Resources Analysis Support System, 1991). Upon completion of the vectorizing process, the maps were appended together and edited to remove spurious vectors created as a by-product of the thinning process.

The vector stream channel map was then re-exported into ARC/INFO, which supports both vector and routing functions. An ordering routine was created that takes advantage of the "from" and "to" node data structure that ARC/INFO imposes on vector maps. All the streams first had to be oriented in the downward direction (i.e., pointing downstream). Once the streams were all pointing in the downstream direction, the ordering program was initiated. By assigning all vectors that had an open-ended "from" node an order value of one, it was possible to stimulate a cascading effect, whereby all vectors were assigned a stream order based on their relationship and connectivity to other channels.

A 10 m resolution DEM provided the basis for the articulation of subwatersheds and the creation of many theme layers important to the statistical analysis of field data. Created from a large number of spot elevation points, contour data, and a thinned version of the channel network using the ARC/INFO tool "topogrid" (Environmental Systems Research Institute, 1994), the DEM was resolved to a 10 by 10 meter gridded surface. Using the "selectpoint" option within the "watershed" command in GRID, subwatersheds were delineated above each of the 222 channels surveyed in the field. From the DEM theme layers for flow direction and flow length were created for each watershed. Watershed characteristics that were derived with the GIS included: watershed area; maximum flow length; average slope; elevation characteristics; and watershed shape variables.

STATISTICAL ANALYSIS

Descriptive statistics implied that stream order was significantly related to channel shape variables. An analysis of variance showed that significant differences exist for channel width, depth, and cross-sectional area between each stream order. Multiple analysis of variance proved average watershed soil clay content to have no influence on channel shape. Relationships between channel shape variables and watershed parameters were investigated using simple linear regression analysis. Having found strong relationships between these variable sets, multiple regression analysis was employed to further refine these relationships.

RESULTS

For the purposes of evaluating the relationship between channel morphology and the contributing area, the relationships describing the channel cross-sectional area were of primary interest, and

Table 5.1. Relationship of Channel Morphology Variables to Stream Order

Order / N	Average Width (cm)	Average Depth (cm)	Average Cross-Sectional Area (m^2)
1 / 58	279.65	26.32	0.802
2 / 65	404.32	34.57	1.47
3 / 40	563.03	40.10	2.54
4 / 26	960.39	54.94	5.63
5 / 20	1967.42	52.58	10.58
6 / 13	3329.99	79.69	26.21

deterministic models were derived using regression analysis. Channel cross-sectional area is a function of both channel width and average depth and thus reflects the total channel response to its hydrologic regime. Channel width can be extracted from a high resolution GIS such as exists for Walnut Gulch. Therefore, given a strong statistical relationship between cross-sectional area and watershed parameters, it would be possible to fully articulate channel geometry (width, depth, cross-sectional area) for all channels throughout Walnut Gulch. This ability to model channel shape accurately when a minimum of field data are available may benefit the application of a host of hydrologic models that incorporate hydraulic channel routing (i.e., the USGS DR3M model—Alley and Smith, 1982; KINEROS—Woolhiser et al., 1990; HEC–1, Army Corps of Engineers—Feldman, 1995).

Horton (1945) investigated the role of stream order on channel shape and hydrologic processes. He found that stream order was highly correlated to many watershed and channel variables, and that stream order could be used as a predictive tool for these variables. Strong relationships between stream order and channel shape were also found to exist on Walnut Gulch (Table 5.1). In this project, statistically significant differences were found to exist between the means of channel width, depth, and cross-sectional area for each step in stream order. Stream order, which is closely related to contributing area, was found to exert a strong effect on channel shape, and was used to stratify the data into subcategories for further analysis.

Channel characteristics were related strongly and in a semilog fashion to stream order. Average channel width, depth, and cross-sectional area were all directly related to order, with a break in the trend occurring between the fourth and fifth order channels, but only for channel width and depth; cross-sectional area maintains a semilog relationship throughout each step in order. The average value for channel depth shows a decrease between channel orders four and five, which is out of trend for every other increase in order (Figure 5.1). However, there is a significant increase in channel width between the fourth and fifth order channels, effectively counteracting the decrease in depth so that the relationship between cross-sectional area and stream order remains consistent

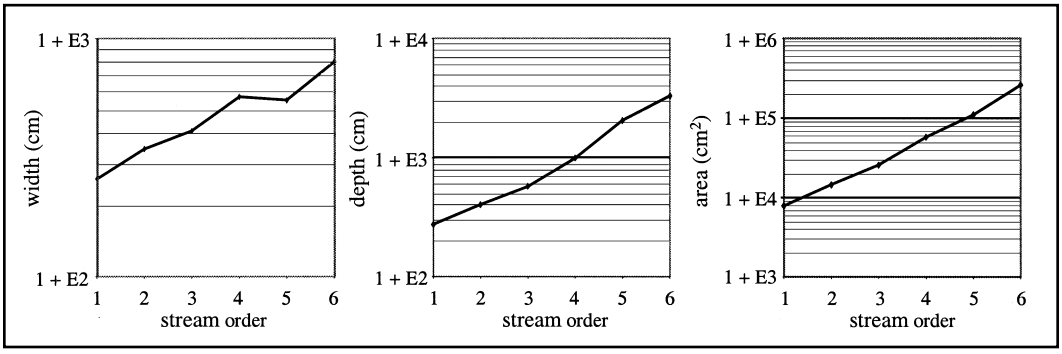


Figure 5.1. Semi-log plots of channel morphologic properties as a function of stream order.

across each order. The overall effect on channel shape is an increase in the channel width:depth ratio, while the relationship of cross-sectional area to order (and, hence, upstream watershed area) remains consistent (Figure 5.1).

Statistical Properties of Channel Shape

Channel width appears to be more sensitive to the influence of watershed parameters than channel depth. Measured values of width have a large spread in their data, while the values for depth show a more central tendency with a lower variation. Without exception, channel width proved to have a higher coefficient of determination than depth (e.g., $R^2 = 0.33$ for depth and 0.72 for width when related on a log-log basis to watershed area) when regression analysis was performed. In fact, depth proved to be resistant to any deterministic model based on the variables used in this study. Some of this resistance to forming a deterministic relationship may be a function of the difficulties associated with precisely measuring depth in the field. Fluvial characteristics are undoubtedly important to this tendency: as flow energy increases in a channel, the channel will adjust its shape to accommodate the increased level of power and erosive energy. This can be accomplished through the widening and/or deepening of the channel. In the loosely consolidated soils of Walnut Gulch, the channels appear to respond to elevated flow energy by increasing their channel width proportionally more than depth.

Responding to the runoff they receive from uplands, stream channels constantly adjust their shape to achieve equilibrium with the flow volume. Changes in channel morphology may result in either degradation or aggradation, with a resultant change in the width:depth ratio, but the net effect is a change in the channel cross-sectional area. As such, the measurement and analysis of channel cross-sectional area is an effective method of illustrating the manner in which channels are responding to watershed characteristics.

A strong relationship exists between channel area and the maximum flow length within a watershed ($R^2 = 0.79$). Table 5.2 shows the results of regression models involving channel area. Long flow lengths within a watershed have been directly related to discharge (Leopold et al., 1964). With higher flows, the channel will become enlarged, either through bed scour or bank erosion, to accommodate the larger flows, resulting in an increased channel cross-sectional area. Following the same reasoning, a strong relationship between channel cross-sectional area and watershed area would also be expected. Data collected in this research support that logical extension. A log-log relationship ($R^2 = 0.68$) exists between channel cross-sectional area and watershed area. A strong relationship ($R^2 = 0.77$) exists between channel cross-sectional area and the watershed area: perimeter ratio, a measurement of the rotundity of a basin, and hence an indicator of basin response. Neither average watershed slope nor the relief ratio correlated strongly with channel cross-sectional area. The log of cumulative drainage length (total length of all channels in a subwatershed) had a moderate relationship to the log of channel cross-sectional area ($R^2 = 0.62$).

In order to improve on the relationships derived using simple linear regression, channel variables were related to watershed characteristics using multiple linear regression. Multiple regres-

Table 5.2. Results of Linear Regression Analysis between Channel Area and Watershed Variables

Variable	Watershed Characteristic	R^2	Coefficient	Constant	Se_{yx}
Channel area	maximum flow length	0.79	0.001	1.83	3.46
Log channel area	log watershed area	0.68	0.49	–2.44	0.34
Channel area	area:perimeter ratio	0.77	0.03	0.17	3.60
Log channel area	log cumulative channel length	0.62	0.51	–1.38	0.40

Table 5.3. Results of Stepwise Backwards Multiple Linear Regression Analysis for Channel Cross-Sectional Area as a Function of Watershed Variables[a]

Case	Regression Model	R^2	Se_{yx}
1	Ca = 0.686(So) + 0.065(Aw) + 0.909(Lm) – 0.006(h)	0.849	3.36
2	Ca = 0.40(So) + 0.009(Aw) + 0.821(Lm) – 0.006(h)	0.851	3.35
3	Ca = 0.72(So) + 0.095(Aw) +0.001(Lm) – 0.007(h) – 0.001(Dl)	0.851	3.34
4	Ca = 0.616(So) + 0.001(Lm) + 0.001(S)	0.849	3.42

[a] where: Ca = channel cross-sectional area (m^2); So = stream order; Aw = subwatershed area (m^2); Lm = maximum flow length (m); h = relief (m); Dl = sum of drainage lengths (m); S = basin slope.

sion analysis of channel cross-sectional area revealed the relatively strong role that channel order played in the determination of channel cross-sectional area. Systematic exploration of the watershed data, using both stepwise forward and backward regression analysis, showed that channel area was heavily dependent on stream order and the area of and maximum flow length within the contributing watershed (Table 5.3). Depending on the subset of parameters investigated, it was possible to extract a significant regression model with a number of different independent variables. To avoid collinearity, multiple pools of data were used during the regression analysis. For example, the relief ratio, a product of the maximum flow length and maximum elevation change, was considered separately from those two variables. The same separation was used for basin shape variables and watershed size. Note that a constant was not used in the analysis, and the equations were driven through the origin.

CONCLUSIONS

Strong statistical relationships were derived between channel variables measured in the field, such as width, depth, and cross-sectional area, and a host of watershed parameters, including channel order, watershed area, shape, drainage properties, and elevation characteristics that were defined using a GIS. Channel cross-sectional area was related in a deterministic manner to multiple watershed variables, yielding models with strong coefficients of determination ($R^2 > 0.84$). Channel shape (and, hence, bankfull stage) may thus be predicted from watershed characteristics readily extracted from common GIS coverages.

Field data were successfully integrated with GIS-derived results. Channel cross-sectional area and other field-measured channel morphometric parameters were found to be strongly related to watershed characteristics extracted from a high-resolution GIS. It is preferable to collect field data when developing parameters for application in hydraulic routing models, but field collection can be costly and time consuming. The channel coverage created for Walnut Gulch contains information on channel width. Using the values for width that can be extracted for the GIS, in conjunction with the developed regression models, values for channel depth and cross-sectional area may be calculated for all channel segments within the watershed. Relationships developed upon verification outside Walnut Gulch have the potential to overcome the inability of DEMs to parameterize channel cross-section properties. In this fashion hydrologic models can be parameterized using a GIS to aid in the understanding of hydrologic processes in the southwestern United States.

ACKNOWLEDGMENTS

Support for this project was provided by the United States Department of Agriculture (USDA) Agricultural Research Service, Southwest Watershed Research Center, Tucson, Arizona. The au-

thors would also like to thank the Advanced Resource Technology Group (ART), University of Arizona, for the use of equipment, space, and technical advice.

REFERENCES

Alley, W. M., and P. E. Smith, 1982. Distributed Routing Rainfall-Runoff Model—Version II. Computer Program Documentation User's Manual. USGS-WRD open-file report 82–344. Gulf Coast Hydroscience Center, NSTL Station, MS.

Breckenfield, D. J., W. A. Svetlik, and C. E. McGuire, 1995. Soil Survey of Walnut Gulch Experimental Watershed. United States Department of Agriculture, Soil Conservation Service.

Burrough, P. A., 1986. *Principles of Geographical Information Systems for Land Use Assessment*. Oxford University Press, New York, NY.

Environmental Systems Research Institute (ESRI), 1994. ARC/INFO ver. 7.0 manual (on-line documentation). Environmental Systems Research Institute Corp. Redlands, CA.

Feldman, A. D., 1995. HEC–1 flood hydrograph package. Chapter 4 of *Computer Models of Watershed Hydrology.* Water Resources Pub., Highlands Ranch, CO, pp. 119–150.

Garbrecht, J., and L. Martz, 1995. TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation, and Subcatchment Parameterization; TOPAZ Overview. USDA-ARS Publication NAWQL 95–1.

Geographic Resources Analysis Support System (GRASS), 1991. GRASS ver. 4.0 Manual. U.S. Army CERL. Champaign, IL.

Gordon, N. D., T. A. McMahon, and B. L. Finlayson, 1992. *Stream Hydrology An Introduction for Ecologists*. John Wiley and Sons, New York, NY.

Harrelson, C. C., C. L. Rawlins, and J. P. Potyondy, 1994. Stream Channel Reference Sites: An Illustrated Guide to Field Technique. USDA Forest Service General Technical Report RM–245.

Horton, R. E., 1945. Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin* 56:275–370.

Lane, L. J., M. H. Nichols, M. Hernandez, C. Manetsch, and W. R. Osterkamp, 1994. Variability in discharge, stream power, and particle-size distributions in ephemeral-stream channel systems. In *Variability in Stream Erosion and Sediment Transport; Proceedings of the Canberra Symposium*, *December 1994.* IAHS Publication 224: 335–342.

Leopold, L. B., M. G. Wolman, and J. P. Miller, 1964. *Fluvial Processes in Geomorphology.* W.H. Freeman and Co., San Francisco, CA.

Libby, F. J., D. E. Wallace, and D. P. Spangler, 1970. Seismic Refraction Studies of the Subsurface Geology of the Walnut Gulch Experimental Watershed, AZ, USDA-ARS 41–164.

Osterkamp, W.R., L.J. Lane, and G.R. Foster, 1983. An Analytical Treatment of Channel-Morphology Relations. USGS Professional Paper 1288.

Renard, K. G., L. J. Lane, J. R. Simanton, W. E. Emmerich, J. J. Stone, M. A. Weltz, D. C. Goodrich, and D. S. Yakowitz, 1993. Agricultural impacts in an arid environment: Walnut Gulch studies. *Hydrologic Science and Technology* 9(1–4): 145–190.

Strahler, A. N., 1952. Dynamic basis of geomorphology. *Geological Society of America Bulletin* 63:923–938.

Woolhiser, D. A., R. E. Smith, and D. C. Goodrich, 1990. KINEROS- A Kinematic Runoff and Erosion Model; Documentation and User Manual. USDA-ARS Pub. ARS–77.

CHAPTER 6

GIS Modeling and Visualization of the Water Balance during the 1993 Midwest Flood

Pawel J. Mizgalewicz, W. Scott White, David R. Maidment, and Merrill K. Ridd

INTRODUCTION

The characterization of the water balance during a major flooding event is useful to scientists, land-use planners, and government agencies who need such information to develop strategies for coping with future floods. Recent technological developments in the form of computer mapping and analysis databases have provided the means to accurately analyze flood-related data sets, such as streamflow and precipitation measurements, over extended periods of time. This study investigates the use of geographic information systems (GIS) to model spatially distributed and time-varying hydrologic and meteorologic data sets, and the use of scientific visualization techniques in the interpretation of the results. The data sets referred to deal specifically with the 1993 Midwest flood which affected a large part of the Upper Mississippi River Basin.

The study area encompasses the main stem of the Mississippi River above Cairo, Illinois, and the main stem of the Missouri River below Gavins Point Dam near Yankton, South Dakota. Also included are the many tributaries of these rivers such as the James, Des Moines, Illinois, and many other rivers and streams. This region of nearly 700,000 km^2 was previously defined by the Scientific Assessment and Strategy Team (SAST) as the area most affected by the spring and summer floods of 1993 (SAST, 1994). States included in this study are Illinois, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota, and Wisconsin (Figure 6.1). Exactly 180 Hydrologic Unit Code (HUC) subbasins (8 digit) cover the study area.

The goal of this project is twofold:

1. Develop, demonstrate, and document procedures used to model the water balance of a large-scale flooding event, such as the 1993 Midwest floods, utilizing hydrologic and meteorologic data sets in a GIS database; and
2. Develop a comprehensive understanding, in terms meaningful to scientists as well as policy and decision makers, of the climatic and hydrologic conditions related to the 1993 floods and the spatiotemporal dynamics of the events. This understanding will result from the use of visualization software in the depiction of water storage change over the land surface during the flooding period.

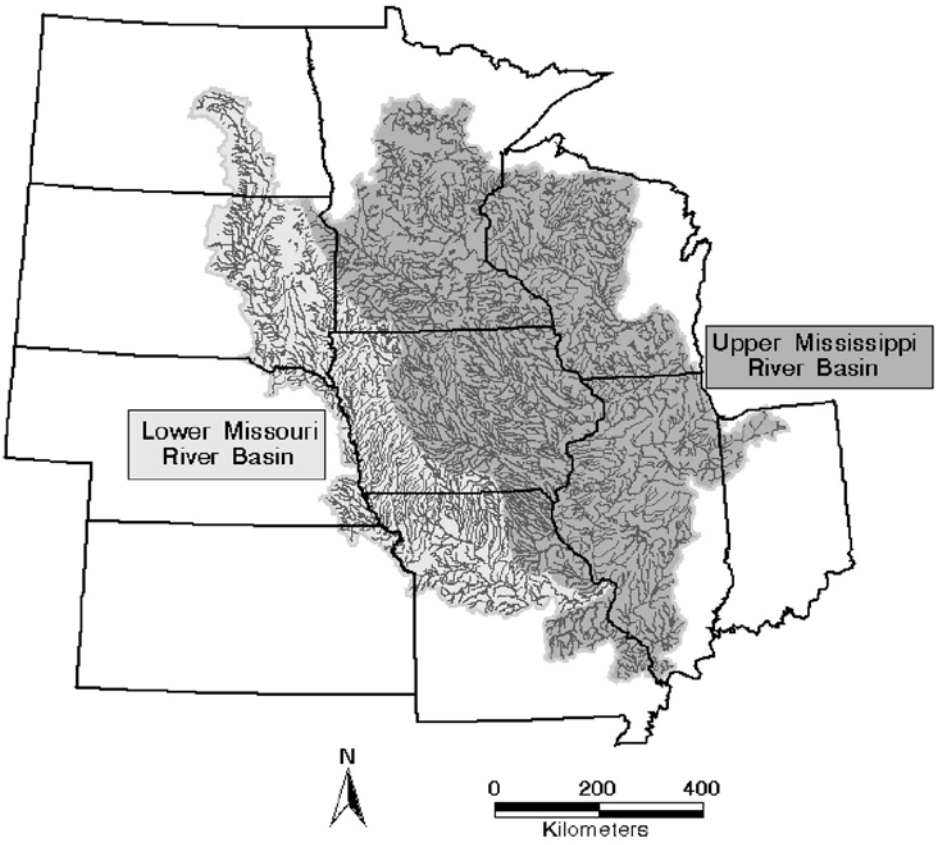


Figure 6.1. Study area—Upper Mississippi and Lower Missouri River Basins.

SURFACE HYDROLOGY CONSIDERATIONS

The water balance of a particular area represents a measure of the inputs to the hydrologic system and the outputs from that system over a specified period of time. In the case of this study, daily water balance maps were desired, therefore a decision was made to partition the study area into incremental drainage units based on the location of stream gauge stations. These drainage units or gauging station zones provided the basis on which areal interpolation of precipitation measurements were made. The outlets of the 180 HUC boundaries rarely corresponded with a stream gauge site, thus necessitating the subdivision of the study area into customized zones. The creation of the gauging station zones is shown in Figure 6.2. There are three possible zones: (1) no inflow; (2) one inflow; or (3) two or more inflows.

The equation for calculating the water balance of each gauging station zone is:

$$dS/dt = I - Q \tag{1}$$

where S is the volume of water stored in each gauging station zone, t is the time index, and I and Q are inflow to the gauging station zone and outflow from the gauging station zone, respectively. The term on the right $(I - Q)$ can be rewritten to account for the area of the gauging station zone:

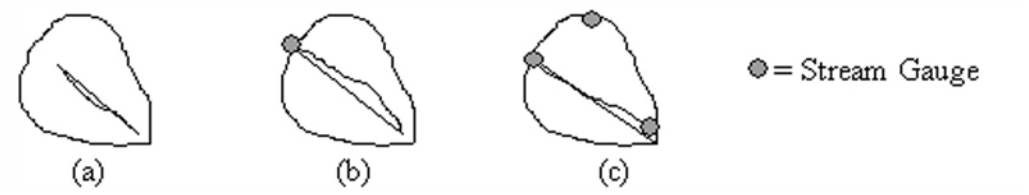


Figure 6.2. Gauging station zone configurations based on (a) no inflow, (b) one inflow, and (c) two or more inflows.

$$I - Q = (P - E)A + SQ_{in} - Q_{out} \tag{2}$$

where P is the rate of precipitation, E is the rate of evaporation, A is the area of the gauging station zone, SQ_{in} is the sum of the inflows recorded at gauges whose flows enter the zone and Q_{out} is the outflow at the downstream end of the zone.

The change in storage (ΔS) can be found by combining Equations 1 and 2 and integrating each term over a time interval of one day (Δt):

$$\Delta S = P\Delta tA - E\Delta tA + SQ_{in}\Delta t - Q_{out}\Delta t \tag{3}$$

In order to express ΔS in terms of average depth over the gauging station zone area, Equation 3 can be rewritten as the following ($\Delta y = \Delta S/A$):

$$\Delta y = P\Delta t - E\Delta t + (1/A)[SQ_{in}\Delta t - Q_{out}\Delta t] \tag{4}$$

where Δy is the change in storage depth per unit area of a gauging station zone. The values of Δy were computed on a daily basis (01/01/93 through 09/30/93) for each zone.

GIS AND WATERSHED DELINEATION

The raster GIS program GRID, a module available in the ARC/INFO® software package, allows for several hydrologic modeling procedures including the determination of flow direction, flow downstream accumulation, and watershed delineation. One of the strengths of this cell-based modeling package is the availability of map algebra functions. With map algebra, the variables in a logical expression actually are map (raster grid) layers. Algebraic manipulation of these grids can be performed at the local or individual cell level, neighborhood level (cells surrounding the cell of interest), zonal level (entire cell groups change in value), or at a global level (entire grid changes in value) (Tomlin, 1990).

Prior to delineating the gauging station zones, several data sets were acquired from the GCIP Reference Data Set (GREDS) CD-ROM, produced by the U.S. Geological Survey. This CD-ROM is a collection of several geographic reference data sets of interest to the global change research community (Rea and Cederstrand, 1994). ARC/INFO export files of 8-digit HUC boundaries and current streamflow gauge sites were obtained from this CD-ROM, as was a 15″ digital elevation model (DEM) of the region (500 meter resolution). RF1 river reach files in ARC/INFO export format for the Upper Mississippi and Missouri Rivers were downloaded via the Internet from the USGS node of the National Geospatial Data Clearinghouse (http://h2O.er.usgs.gov/nsdi/wais/water/rf1.HTML).

The first step in creating the gauging station zones was to edit the RF1 vector coverage to re-

move circular arcs and isolated lakes. Once a definite stream network was available, the vector coverage was converted to a raster grid, and then embedded into the 15″ DEM, which had been corrected for spurious sinks or pits. This embedding procedure "raised" the elevation of the off-stream grid cells surrounding the network in the DEM relative to the stream grid cells. By embedding these stream cells, a more precise network was created on the DEM that exactly matched the original paths of the RF1 vector coverage. This stream "burn in" technique has been shown to be particularly effective in areas of low relief (Maidment, 1996). The resulting grid was then clipped with a polygon coverage of the study boundaries. A 50 km buffer had been applied to this boundary coverage to account for drainage outside of the study area.

The next steps involved the actual delineation of the stream network within their embedded channels on the DEM. In ARC/INFO's version of this process, the flow direction is first determined by examining the neighborhood of eight grid cells that surround the cell of interest. The flow direction function identifies the lowest cell value in the neighborhood, and assigns a flow direction value to the corresponding cell in an output grid, thus creating an implicit stream network between cell centroids. Once this process was complete over the entire study area, a flow accumulation grid was made. The flow accumulation function in ARC/INFO uses the flow direction grid to determine the number of "upstream" cells. A new value is assigned to the cells in an output grid showing the number of cells that contribute or flow to downstream cells. High values indicate confluences of streams, whereas values of zero indicate watershed boundaries (Maidment, 1995). A conditional statement was then set up in ARC/INFO which isolated those cells that met a certain threshold of flow accumulation. Stream links were created, and the stream network was then in place.

After a suitable terrain model had been made, the next step was to precisely locate the USGS gauge stations on the stream grid. The point coverage of station locations was converted to a grid, and then viewed as a background grid against the stream grid. Unfortunately, most of the gauge cells did not lie directly on top of a stream cell, so the stream cell closest to a gauge cell was given a unique value (the USGS-assigned station number) to differentiate it from adjacent nongauge stream cells. Initially, 460 gauge cells were located on the stream grid; however, it was later determined that a number of these stations contained incomplete streamflow records. A total of 50 gauges were removed which did not have complete records. This left 410 gauges, which were unevenly distributed over the stream network. Some streams had many gauges, whereas other stream segments contained one or less. The gauge locations were viewed along with the HUC coverage to provide a better spatial representation of the gauges with respect to watersheds. It was decided that those gauges whose contributing drainage areas (an attribute in the ARC/INFO point coverage of the gauges) were less than 1,000 km^2 would be removed from the collection. Also removed were gauges that were concentrated in a particular watershed. Most HUCs contained between one and three gauge stations, but several contained more than three. Unless the HUC was large in area, those extra gauge stations exceeding three were also removed from the collection. Approximately 260 gauge stations on the stream grid were retained, and these gauge locations were for the most part uniformly distributed throughout the stream network.

The gauging station zones were then determined using the flow direction grid and the edited stream gauge grid (Figure 6.3). The resulting grid contained approximately 260 subbasins defined on the basis of the stream gauge locations. Each gauging station zone was checked against its corresponding HUC to ensure that the two sets of boundaries were mutually compatible.

GIS DATABASE AND MAP CREATION

Streamflow daily values for the 1993 water year were provided by the Water Resources Division of the USGS. After removing the 1992 values and reformatting the data into comma-delim-

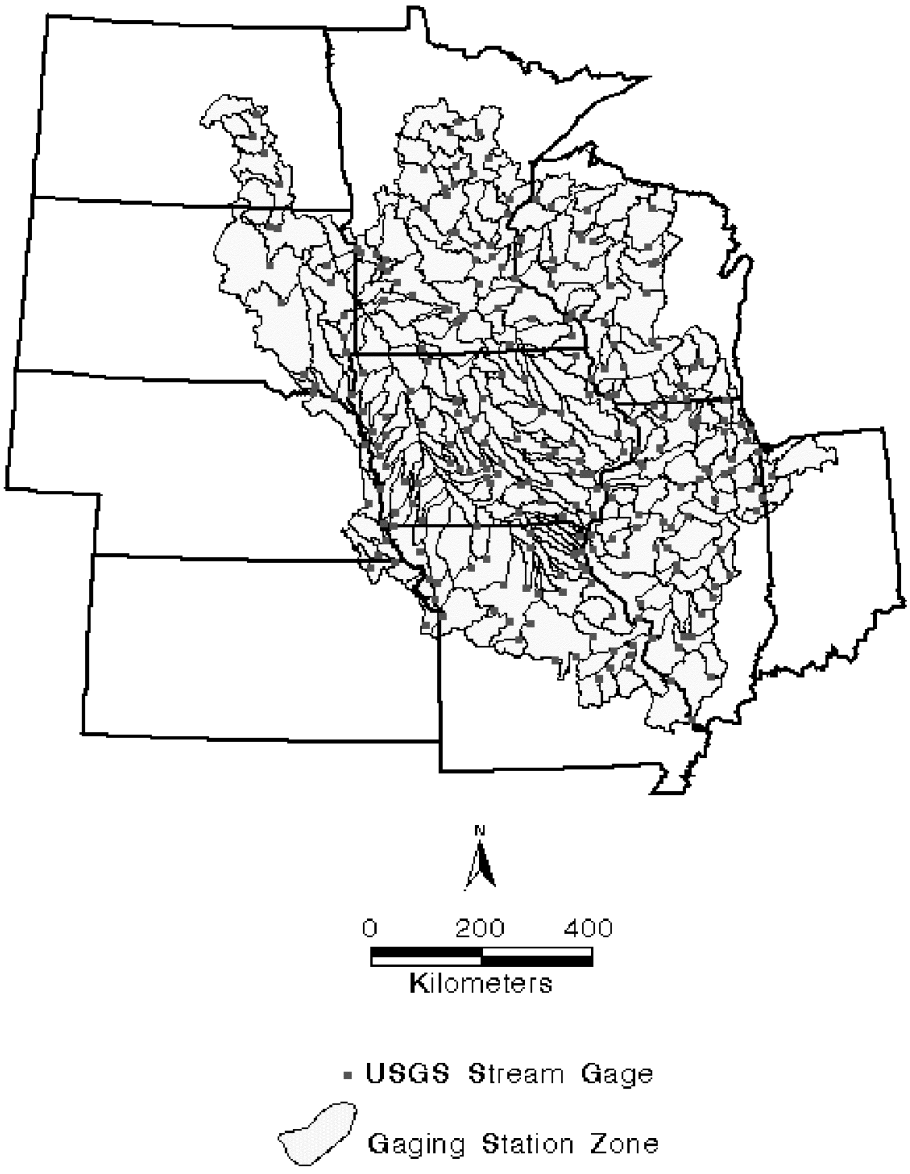


Figure 6.3. Map of gauging station zones defined on the basis of the location of USGS streamflow gauge stations.

ited ASCII text, the values were incorporated into the attribute table of the gauge station point coverage. This point coverage reflected the edits made to the stream gauge grid. Complete daily records for the gauge stations were assigned to the point coverage for a total of nearly 71,000 daily values from January 1 to September 30, 1993 (273 days).

Precipitation values obtained from National Climatic Data Center Summary of the Day files were treated in a manner similar to the streamflow data. A total of 1,078 climate stations were mapped. This number includes stations within the study area, and those within a 50 km buffer of

the study area. Stations in the buffer zone were included so that a more precise interpolation of the precipitation measurements could be achieved for the regions at the study area borders. Over 290,000 precipitation measurements were incorporated into the point attribute table of the stations coverage. Some of the climate stations had missing precipitation measures, so an Arc Macro Language (AML) program was written to eliminate those stations with missing data.

A total of 273 daily precipitation grids with 4 km cell resolution was interpolated from the gauged precipitation data using the inverse distance weighting method. The ARC/INFO GRID function "zonalaverage" was subsequently used to obtain an average value of precipitation depth, as well as potential evapotranspiration (PET), over the gauging station zones. PET monthly values were compiled from the National Weather Service. Since values of PET vary slowly from month to month and the study area was saturated during the 1993 floods, it was felt that evaporation was most likely constant (i.e., one PET value per month) over the month in each gauging station zone.

The resulting grids show unique gauging station zones with average precipitation and evaporation depth values in each zone. Net streamflow grids show daily values averaged over each gauging station zone. These grids were created by subtracting Q_{out} from the sum of Q_{in}, and dividing the answer by the area of the gauging station zone. The final map series involve a combination of all of the preceding maps. Daily storage change (Δy) grids were produced using the map algebra functions in ARC/INFO GRID and Equation 4. Essentially, the equation changed from

$$\Delta y = P\Delta t - E\Delta t + (1/A)[SQ_{in}\Delta t - Q_{out}\Delta t] , \tag{4}$$

to

$$S\ Grid = P\ Grid - E\ Grid + (1/A)[NS\ Grid] , \tag{5}$$

where *S Grid* is a grid of daily storage depth change, *P Grid* is a daily precipitation grid, *E Grid* is a daily evaporation grid, and *NS Grid* is a daily net streamflow grid. Maps of cumulative storage depth were obtained by summing the grids. The resultant 273 grids show water storage depth of each gauging station zone in the study area on a daily basis.

Balancing the water inputs into the basin (precipitation and streamflow), and the outputs from the basin (evapotranspiration and runoff) meant that the database had to be extended to include the whole calendar year of 1993. The number of precipitation, streamflow, evaporation, and subsequent daily water storage grids was increased from 273 to 365 each to take into account the rest of the year (10/01/93 to 12/31/93). It was therefore assumed that the amount of water entering the basin on 01/01/93 was the same as the amount leaving the basin on 12/31/93, and an evaporation multiplication factor was introduced to make the necessary calculations. Based on these calculations, it was found that the water storage in the UMRB study area, when the basin is considered as one modeling unit, rose by approximately 110 mm in about a two-month time period (from the end of May to the beginning of August). However, the range of water storage depths was not geographically consistent throughout the basin. As one might expect, the highest storage depths increases took place in the most downstream units of the UMRB, where calculations show that the depth was nearly 300 mm for the two-month period described above.

VISUALIZATION

The final products of this study consisted of maps that represented the water storage depth which occurred on each subbasin or gauging station zone as a result of the heavy precipitation events between 01/01/93 and 12/31/93. Each day's maps represented a snapshot of how much pre-

cipitation occurred in the study area, and how it affected the land surface. Although the GIS can perform many of the tasks necessary to determine water storage, a final step remains in order to fully understand the temporal aspects of the flooding events. This visualization step involves the use of various computer software packages, coloring and animation procedures, and output types.

GIS software packages have provided the spatial sciences with an excellent set of tools for performing hydrologic analyses. The main problem with the technology today is its relative inability to handle temporal information. This chapter has described a method of incorporating temporal data into the attribute files of GIS layers, but the final product, as described up to this point, remains a series of static portrayals of a dynamic event. To overcome this problem, various visualization methodologies are being investigated as additional components to the GIS-based hydrologic modeling process. Buttenfield and Mackaness (1991) described GIS-related visualization as the interface of three processes: (1) computer analysis (data collection, organization, modeling, and representation); (2) human cognition (perception, pattern identification, and mental imaging); and (3) graphic design principles (construction of visual displays). MacEachren (1994) further described the concept of "geographic visualization" as stressing map use that can be conceptualized as a three-dimensional space. Both descriptions of visualization imply that maps and associated images can now be constructed that incorporate 3-D and 4-D perspectives. This ideally suits geographic visualization techniques as an end product to temporal representation using a GIS.

Currently, the authors are investigating several methods of geographic visualization. Most involve some form of geographic animation as a way of representing the temporal aspects of the flood. Initially, the 273 maps of water storage depth, averaged over each gauging station zone from 01/01/93 to 09/30/93, will be color-coded based on the depth of water storage in each zone. Blue is the most likely color, and will range in shades from light or pale (no water saturation of the land surface) to dark or bright (saturated land surface). These choropleth maps will be chronologically arranged from 01/01/93 to 09/30/93, and then put into an animation software package. Viewing these maps in sequence will provide the user with a dynamic display of water leaving gauging station zones and saturating downstream zones. Figure 6.4 (see color section) shows a sequence of four daily maps (07/29/93–08/01/93) with corresponding water storage in 8-digit HUC boundaries.

Another means of visualizing flow through the basin would be through the use of the stream network. The representation of the network in the GIS does not really account for its changing form and velocity during the flood months of 1993. A series of 273 numbers representing streamflow velocity could be assigned to the cells representing stream gauge locations along the rasterized network (500 m resolution), but the actual channel dimensions remain the same throughout the time period. Visualization procedures can be used in conjunction with the streamflow values to produce a dynamic representation of streamflow velocity through the basins. The flow velocity can be represented by some type of hydrologically significant symbol which appears to speed up or slow down depending upon the streamflow value at the gauge location. This iconic representation of streamflow will increase in size as flow becomes faster downstream. Combining the symbology of flow velocity with the choropleth map described above is also feasible.

The authors are also experimenting with various means of stream channel visualization. As previously mentioned, the streams reside in the GIS as a 500 meter wide grid network. Of course, the streams vary in size in the Upper Mississippi and Missouri River basins from creeks a few meters wide, to the Mississippi River at Cairo, Illinois, which is nearly 1 km wide during normal flow conditions. Vector representation of the stream network would allow some flexibility in the appearance of the channels, and this method is currently being investigated with the GIS software packages ARC/INFO and ArcView. Satellite and radar imagery will also be incorporated into a vi-

sualization package to provide a more realistic view of the flooding events. By using imagery draped over a DEM of the study area, three-dimensional viewpoints can be achieved which provide even greater flexibility in terms of visualization and human cognition of the floods.

SUMMARY AND CONCLUSIONS

In this study, geographic information system software has been utilized as a first step to visualizing the water balance that was in place during the 1993 Midwest floods. A series of daily maps for 1993 are being produced that combine streamflow, precipitation, and evaporation measurements. This series of maps, while informative and meaningful to the group of people who understand GIS, needs to be made available to the public in a different form. The second step will be the development of a dynamic representation of this temporal data through the use of geographic visualization procedures. The final product of this research will be useful to many different people who study such natural hazards as flooding. Methodologies developed in this study can be used to investigate future floods, and will hopefully provide the end users with several different views of the 1993 Midwest flood which will enable them to make wise decisions regarding floodplain management and land use practices.

ACKNOWLEDGMENTS

This research has been supported by a grant from the U.S. Geological Survey to Drs. Maidment and Ridd. Thanks are also due to Dr. William Kirby, of the USGS Water Resources Division in Reston, VA, for supplying the authors with necessary data sets, and to Dr. Barbara Buttenfield for providing some ideas about geographic visualization.

REFERENCES

Buttenfield, B.P., and W.A. Mackaness, 1991. Visualization, in Maguire, D., M. Goodchild, and D.W. Rhind, Eds., *Geographical Information Systems: Principles and Practice.* Longman, London, pp. 427–443.

MacEachren, A.M., 1994. Visualization in modern cartography: Setting the agenda, in MacEachren, A.M., and D.R. Fraser-Taylor, Eds., *Visualization in Modern Cartography*. Pergamon, Oxford, UK pp. 1–12.

Maidment, D.R., 1995. *GIS and Hydrology: A Workshop of the 15th Annual ESRI User Conference*. Environmental Systems Research Institute, Redlands, CA.

Maidment, D.R., 1996. GIS and Hydrologic Modeling—An Assessment of Progress. Paper Presented at the Third International Conference on GIS and Environmental Modeling, Santa Fe, NM. National Center for Geographic Information and Analysis, Santa Barbara, CA, 1 CD-ROM.

Rea, A., and J.R. Cederstrand, 1994. GCIP Reference Data Set (GREDS): U.S. Geological Survey Open-File Report 94–388. USGS, Reston, VA, 1 CD-ROM.

SAST, 1994. Science for Floodplain Management into the 21st Century, Preliminary Report of the Scientific Assessment and Strategy Team Report of the Interagency Floodplain Management Review Committee to the Administration Floodplain Management Task Force. Washington, D.C.

Tomlin, C.D., 1990. *Geographic Information Systems and Cartographic Modeling*. Prentice Hall, Englewood Cliffs, NJ.

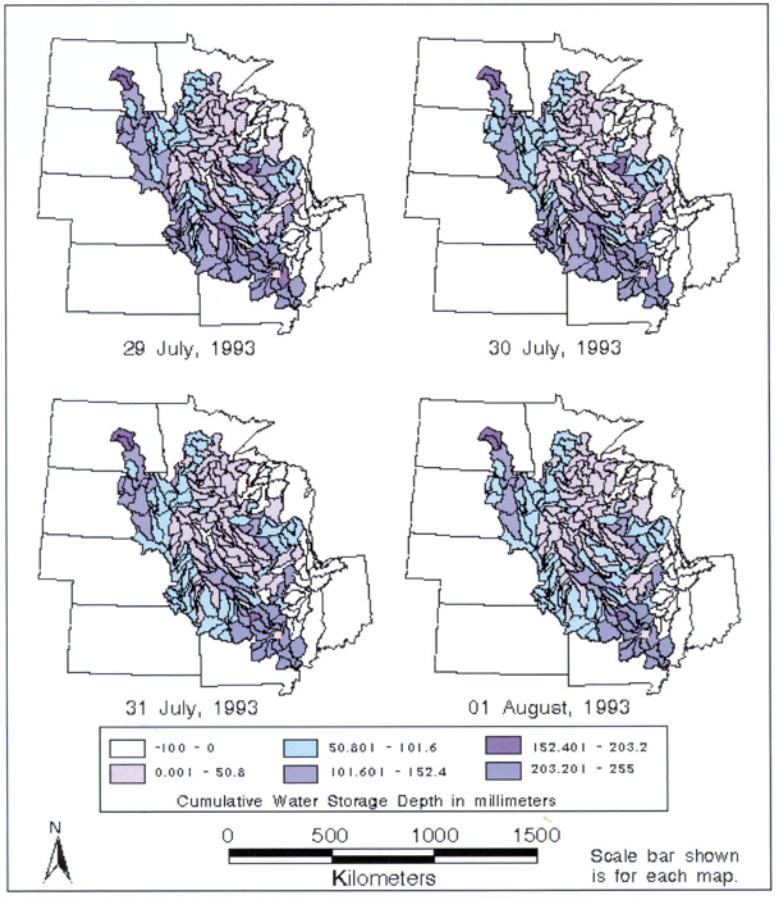


Figure 6.4. Cumulative water storage amounts within USGS HUC boundaries (spatially averaged values from gauging station zones) for the time period 07/29/93 - 08/01/93. Flooding peaked in the St. Louis area on 08/01/93. The city is shown by a small light-colored dot placed at the southern end of the basin.

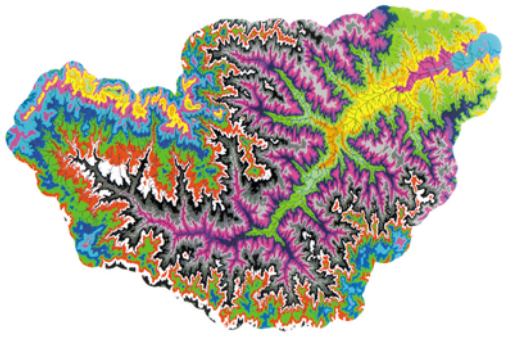


Figure 7.2. Digital elevation model of the Little Washita River Watershed overlain by TOPAZ-derived stream network. Each change in color on the DEM represents an elevation change of 7 m.

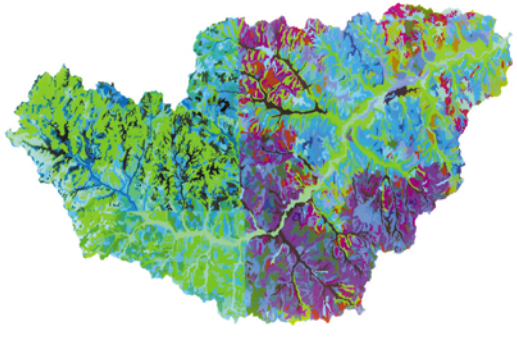


Figure 7.3. Soils coverage for the Little Washita River Watershed. Note that the soil mapping unit codes delineate the county boundaries.

Figure 7.4. April 1994 synoptic temporal land cover classification for the Little Washita River Watershed.

CHAPTER 7

Selection, Development, and Use of GIS Coverages for the Little Washita River Research Watershed

Patrick J. Starks, Jurgen D. Garbrecht, F.R. Schiebe, J.M. Salisbury, and D.A. Waits

INTRODUCTION

The Little Washita River Watershed (LWRW) in south central Oklahoma (Figure 7.1) is the largest and one of the longest-studied research watersheds operated by the United States Department of Agriculture's (USDA) Agricultural Research Service (ARS). The watershed drains an area of 611 km^2 and has been studied since 1961 for rainfall runoff, impact of flood control structures, water quality, sediment transport and best management practices. A series of Geographical Information Systems (GIS) raster coverages have been developed to support present and future ARS research on the LWRW and to complement the historical and current databases. The selection of the

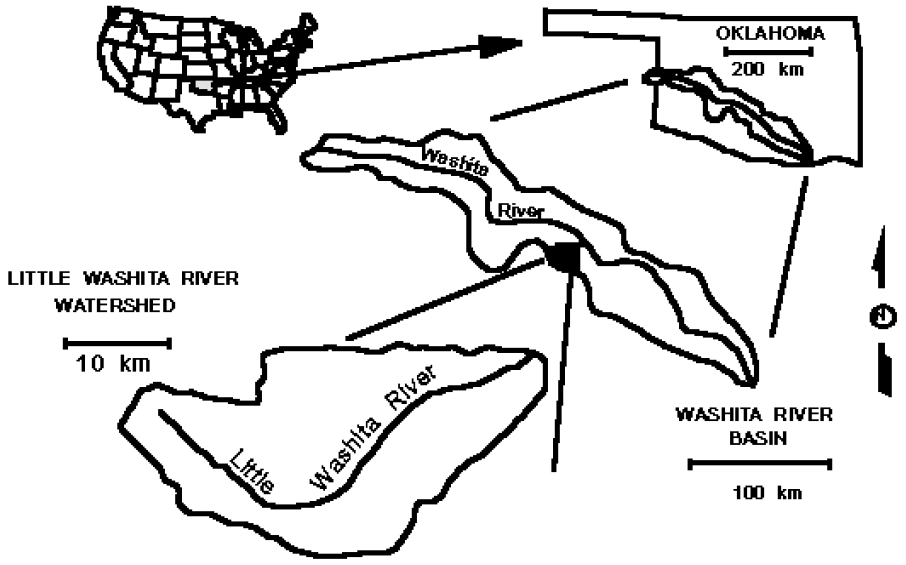


Figure 7.1. Location of the Little Washita River Watershed.

GIS coverages was guided by hydrologic research needs (Goodrich et al., 1994) and an effort to support both distributed and lumped parameter watershed modeling. The coverages are grouped into three categories: topography, soils, and land cover.

The object of this chapter is to present topics relating to the development and use of the three categories of GIS coverages. The selected topics address: (1) the development of hydrographic data layers from the Digital Elevation Model (DEM); (2) reliability of the soil property data extracted from the soils coverage and county soil survey data; and (3) land cover identification from Landsat satellite remotely sensed data.

DIGITAL ELEVATION MODEL AND TOPOGRAPHIC GIS DATA

Automated extraction of topographic parameters from DEMs has established itself in GIS over the past decade. This is attributed to the importance of and need for landscape derived data and to the increasing availability of DEMs and software products that derive topographic data from DEMs. In the field of water resources and hydrology, the main uses of digital landscapes are watershed segmentation, definition of drainage divides and channel networks, determination of catchment geometry, and parameterization of landscape properties such as terrain slope and aspect (Jenson and Domingue, 1988; Mark, 1988; Moore et al., 1991; Martz and Garbrecht, 1992). Such landscape evaluation tasks are generally tedious, time-consuming, error-prone, and often subjective when performed manually from topographic maps, field surveys, or photographic interpretations (Richards, 1981). The automated techniques are faster and provide more precise and reproducible measurements than traditional manual techniques applied to topographic maps (Tribe, 1991). Digital data generated by automated techniques also have the advantage that they can be readily imported and analyzed by GIS.

Most of the topographic coverages presented here were automatically derived from the DEM of the LWRW using software TOPAZ (Garbrecht and Martz, 1999). TOPAZ (TOpographic PArametriZation) is a software package for automated digital landscape analysis. It uses raster DEMs to identify and measure topographic features, define surface drainage, subdivide watersheds along drainage divides, quantify the drainage network, and parameterize subcatchments. TOPAZ is designed primarily for hydrologic and water resources investigations, but is equally applicable to address a variety of geomorphological, environmental and remote sensing applications. TOPAZ is discussed more fully elsewhere in this volume. Topographic GIS coverages of the LWRW, DEM resolution and quality, and degree of watershed segmentation and drainage density are addressed below.

Topographic GIS Coverages

A DEM with a horizontal and vertical resolution of 30×30 meters and 0.3 meter, respectively, is available for the LWRW. The DEM was developed in 1996 by the U.S. Geological Survey, Rolla, MO, from digitized contour data. The quality of the DEM corresponds to a Level 2 DEM; thus it has been processed at production time for consistency and edited to remove identifiable systematic errors. This DEM is the basic GIS coverage representing the landscape topography of the LWRW.

DEM preprocessing is necessary before deriving additional hydrographic and topographic data because DEMs commonly contain localized depressions and flat surfaces, many of which are artifacts of the horizontal and vertical DEM resolution, DEM generation method, and elevation data noise. Depressions and flat surfaces are problematic for drainage identifications. Depressions are sinks at the bottom of which drainage terminates, and flat surfaces have indeterminate drainage.

Therefore, TOPAZ preprocesses the input DEM to rectify these features and allows the unambiguous determination of the drainage over the entire digital landscape. Rectifications are strictly limited to cells of depressions and flat surfaces so as to minimize the impact on the overall information content of the elevation data. Further details on the rectification procedure are given below. With this rectified DEM the hydrographic and topographic coverages listed in Table 7.1 can be derived using software TOPAZ.

Table 7.1. Topographic GIS Coverages and Data Source

GIS Coverage	Source and Development Procedure
Digital Elevation Model	USGS, contour interpolation
Watershed boundary	TOPAZ, downslope flow routing concept
Drainage network	TOPAZ, downslope flow routing concept
Subcatchment drainage boundaries	TOPAZ, downslope flow routing concept
Terrain slope	TOPAZ, surface derivative
Terrain aspect	TOPAZ, surface derivative

The watershed boundary, subcatchment drainage divides, and drainage network computed by TOPAZ are based on the D8 method, the downslope flow routing concept, and the critical source area (CSA) concept. The D8 method (Douglas, 1986; Fairfield and Leymarie, 1991) defines the landscape properties for each individual raster cell as a function of itself and its eight immediately adjacent cells. The downslope flow routing concept defines the drainage on the landscape as the steepest downslope path from the cell of interest to one of its eight adjacent cells (Mark, 1984; O'Callaghan and Mark, 1984; Morris and Heerdegen, 1988). The CSA concept defines the channels draining the landscape as those raster cells that have an upstream drainage area greater than a threshold drainage area, called the critical source area (CSA). The CSA value defines a minimum drainage area above which a permanent channel is maintained (Mark, 1984; Martz and Garbrecht, 1992). The CSA concept controls the watershed segmentation and all resulting spatial and topologic drainage network and subcatchment characteristics.

DEM Resolution and Quality

The spatial resolution and quality of the DEM data are important considerations when deriving hydrographic data from DEMs. The choice of an appropriate resolution must be made under consideration of the landscape characteristics that are to be represented and the use of the derived data products. In the case of the LWRW it has previously been shown (Garbrecht and Martz, 1993) that the network and drainage divides can be adequately derived from a DEM with a 30×30 meter horizontal resolution.

A more difficult question was that of the quality of the elevation data. Accuracy standards, data noise, interpolation errors, and systematic production errors often create spurious depressions, flat areas, and flow blockages which cause problems for the identification of drainage features (Garbrecht and Starks, 1995). These in turn impact the drainage identifications and indirectly the drainage divides and network. The DEM for the LWRW is a Level 2 DEM and has been processed for consistency and systematic errors. However, a Level 2 DEM is not hydrographically corrected and problems such as spurious depression and flat areas remain. These must either be rectified, for example through data preprocessing as described above, or accepted as data limitations. For the LWRW project these problem features in the DEM have been rectified.

Spurious depressions are rectified by a procedure of depression breaching/filling (Garbrecht et

al., 1996), and flat surfaces are rectified by imposing two gradients which are a function of the landscape configuration surrounding the flat surface (Garbrecht et al., 1996). Both these rectifications result in an approximated, yet realistic drainage pattern. It was also common to see discrepancies in the position of derived and actual channels. Such discrepancies were expected because a DEM is only an approximation of the true landscape topography, and the derived drainage divides and network represent the drainage features of the approximated topography.

Degree of Watershed Segmentation and Drainage Network Density

The degree of watershed segmentation and drainage network density are a function of the value given to the critical source area parameter (CSA). The CSA is the minimum drainage area that is required to initiate a first-order channel (Strahler, 1957). Any area smaller than the threshold value does not produce enough runoff to form and maintain a channel. The threshold CSA value depends upon, among other things, terrain slope characteristics, soil properties, land use, and climatic conditions. It can be as small as a fraction of a hectare or as large as tens of hectares, depending on the landscape characteristics under consideration.

Often the CSA value is also used to represent different degrees of watershed segmentation and drainage network densities to address scaling issues. At a small scale, one would choose a small CSA value to represent the smallest channels and hill slopes. As a result, a high degree of partitioning, many subcatchments, and a dense drainage network are obtained. In the case of large-scale applications, only the major streams in the watershed may be needed. A large CSA value would result in the desired low degree of watershed segmentation, few subcatchments, and only large streams. This capability to generate GIS coverages of different degrees of watershed segmentation and drainage network densities is important for landscape modeling.

Figure 7.2 (see color section) is an overlay of the basic LWRW DEM coverage and a TOPAZ-generated stream network from the DEM. The relief of the DEM is approximately 190 m. Each color change on the DEM represents a 7 m change in elevation from the neighboring color. In this example, a CSA of 8 hectares was used to generate the stream network, resulting in 1,218 first order, 264 second order, 54 third order, 12 fourth order, 4 fifth order, and 1 sixth order (Strahler number) channels.

DIGITAL SOILS COVERAGE

The LWRW soil coverages were developed from two data sources provided by the Natural Resources Conservation Service (NRCS): (1) county soil survey maps, and (2) the digital State Soil Geographic Database (STATSGO). The STATSGO database contains the soil attributes for the county soil survey maps. From this information a basic soil coverage and derived soil attribute coverages can be developed.

The basic soil coverage for the LWRW consists of the soil mapping units digitized from the county soil survey maps and rasterized to a 30 m pixel cell size. It was readily apparent that the name given to the soil mapping units is influenced by which county the mapping unit falls within (i.e., the county lines are seen in the soils coverage) [Figure 7.3 (see color section)]. This systematic trend is partly related to the different experience and interpretation of the county's soil scientist in charge of soil classification (Arnold et al., 1994). The trend can also be attributed to different dates at which a county's soils were classified. Soil names of an earlier classification in one county may not correspond to an updated soil classification in another county at a later date. However, it is not the difference in name that is important, but that the soil physical properties are similar for both soil names. Therefore, special care was taken to ensure that the attribute data were

correct for all soils, regardless of soil name, before including them in the GIS data base for the LWRW. The consistency was ensured by involvement of NRCS personnel in the review of the soils attribute data, particularly for the soils along the county lines.

The soil attribute data were transferred from the NRCS STATSGO database into a computer worksheet. The data consist of both generalized attributes for the soil mapping units as well as soil profile data for up to six layers within a mapping unit. The soil layer data include surface and subsurface horizons. Tables 7.2 and 7.3 list examples of attribute data for the soil mapping units and profile layers, respectively. Some attributes are given as maximum and minimum values, and it is up to the investigator's discretion as to which value within this range is appropriate for use in a particular application.

Table 7.2. Examples of Attribute Data for Soil Mapping Units

Slope	Surface Soil Texture
Taxonomic classification	Annual flooding frequency
Flood duration class	Depth to water table
Ponding depth of surface water	Depth to bedrock
Hydrologic soil group	

Table 7.3. Examples of Soil Profile Attribute Data

Depth of Soil Layer	Permeability
Soil texture	Clay content
Water holding capacity	Bulk density

A number of soil attribute coverages can be developed from the basic GIS soil coverage and the attribute data contained in the worksheet by matching the soil mapping units codes in the GIS coverages to soil characteristics in the worksheet attribute file and creating a corresponding soil attribute coverage.

LAND COVER

Land cover is a dynamic entity which varies both spatially and temporally. For example, in agricultural areas it is typical for crop canopies to cover a field for part of the year, while at other times that field is fallowed or bare. Also, crop rotation patterns, crop type, and total acreages planted in crops vary from year to year.

A series of land cover coverages was produced to gain a better understanding of the vegetative dynamics for the LWRW. These coverages were derived from Landsat MSS satellite images for the spring, summer, and fall seasons of every even-numbered year from 1972 through 1994. Each seasonal image was subjected to an unsupervised classification on all four wavebands from the satellite. In the unsupervised approach, a cluster analysis was used to examine the reflectance properties of the land surface and to aggregate related reflectance values into a number of classes. These classes were derived by cluster analysis and represent natural groupings of reflectance values (Eastman, 1992; Lillesand and Kiefer, 1994). Finally, ground truthing was used to assign a land cover category to each class. Land cover categories developed from the seasonal data are water, urban, bare soil, woodlands, native rangeland, tame rangeland (planted), and cultivated lands.

The seasonal remotely sensed data were combined to yield a synoptic temporal view of the agricultural landscape. This was achieved by converting the spring, summer, and fall remotely sensed data into "greenness" values using the greenness vegetation index algorithm (Kauth and Thomas, 1976). The greenness index quantifies vegetation presence and vigor. Colors were assigned to crops in accordance with their "greenness value" and in which season they were actively growing. For example, crops growing only in summer were assigned a color of green, with intensity of that color in proportion to the greenness value. By overlaying these three separate seasonal data sets, a new GIS coverage of the temporal dynamics of the agricultural landscape was produced [e.g., Figure 7.4 (see color section)]. Table 7.4 lists the GIS categories for the coverages of the temporal variability of land cover.

Table 7.4. GIS Categories for Temporal Land Cover Variability

Rangeland:	-native and planted
Wheat :	-spring only
	-fall only
	-spring only
Other crops:	-summer
	-spring and summer
Woodlands	
Water:	-also includes urban and bare soil

CONCLUSIONS

The development and use of three categories of GIS coverages for the LWRW were presented. Selected topics addressed the development of hydrographic coverages from a DEM, reliability of soil property data extracted from county soil survey data, and identification and dynamics of land cover derived from remotely sensed data. Hydrographic data sets were discussed with reference to the quality and spatial resolution of the DEM from which the coverages were derived, and also with reference to the degree of watershed segmentation and drainage network density. Development of the soils coverage revealed inconsistencies in the soil mapping units along county boundary lines within the LWRW. Therefore, NRCS personnel were involved in verifying the consistency of the soil attribute data before it was transferred from the STATSGO data files into the LWRW GIS. Two types of land cover information were derived from remotely sensed data. These show the spatial distribution and temporal dynamics of the land cover. Together these topographic, soil and land cover coverages provide a comprehensive GIS database in support of our research program in water resources and watershed modeling.

REFERENCES

Arnold, J.G., J.D. Garbrecht, and D.C. Goodrich, 1994. Geographic Information Systems and large area hydrology. In Richardson et al., Eds., *Proceedings of the ARS Conference on Hydrology*, Denver, CO, September 13–15, 1993. United States Department of Agriculture, Agricultural Research Service, 1994–1995.

Douglas, D. H., 1986. Experiments to locate ridges and channels to create a new type of digital elevation Model. *Cartographica*, 23(4):29–61.

Eastman, J.R., 1992. IDRISI: User's Guide. Clark University, Graduate School of Geography, Worcester, MA.

Fairfield, J., and P. Leymarie, 1991. Drainage Networks from Grid Digital Elevation Models. Water Resources Research, 27(4):29–61.

Garbrecht, J., and L.W. Martz, 1993. Network and subwatershed parameters extracted from digital elevation models: The Bills Creek experience. *Water Resources Bulletin*, 29(6):909–916.

Garbrecht J., and L. W. Martz, 1999. TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation and Subcatchment Parameterization: TOPAZ Overview. United States Department of Agriculture, Agricultural Research Service, Grazinglands Research Laboratory, El Reno, OK, ARS Pub. No. GRL 99-1, 26 pp.

Garbrecht, J., and P. J. Starks. 1995. Notes on the use of USGS Level 1 7.5 minute DEM coverages for landscape drainage analysis. *Photogrammetric Engineering and Remote Sensing*, 61(5):519–522.

Garbrecht, J., P. J. Starks, and L. W. Martz, 1996. New digital landscape parameterization methodologies. In Hallam et al., Eds., *Proceedings of the AWRA Annual Symposium, GIS and Water Resources*. American Water Resources Association Symposium, Herndon, VA, TPS–96–3.

Goodrich, D.C., P.J. Starks, R.R. Schnabel, and D.D. Bosch, 1994. Effective use of USDA-ARS experimental watersheds. In Richardson et al., Eds., *Proceedings of the ARS Conference on Hydrology,* Denver, CO, September 13–15, 1993. United States Department of Agriculture, Agricultural Research Service, 1994–1995.

Jenson, S. K., and J. O. Domingue, 1988. Extracting topographic structure from digital elevation data for Geographical Information System analysis. *Photogrammetric Engineering and Remote Sensing*, 54(11):1593–1600.

Kauth, R.J., and G.S. Thomas, 1976. The tasseled cap—A graphical description of spectral-temporal development of agricultural crops as seen by Landsat. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*. Laboratory of Applied Remote Sensing, Purdue University, West Lafayette, IN, pp. 41–51.

Lillesand, T.M., and R.W. Kiefer, 1994. *Remote Sensing and Image Interpretation*, 3rd ed., John Wiley and Sons, New York, NY.

Mark, D. M., 1988. Network models in geomorphology. In Anderson, M. G., Ed., *Modeling Geomorphological Systems*. John Wiley and Sons, New York, NY, pp.73–96.

Mark, D. M., 1984. Automatic detection of drainage networks from digital elevation models. *Cartographica*, 21(2/3):168–178.

Martz, L.W., and J. Garbrecht, 1992. Numerical definition of drainage networks and subcatchment Areas from Digital Elevation Models. *Computers and Geosciences* 18(6):747–761.

Moore, I. D., R. B. Grayson, and A. R. Ladson, 1991. Digital terrain modeling: A review of hydrological, geomorphological and biological applications. *Hydrological Processes*, 5(1):3–30.

Morris, D. G., and R. G. Heerdegen, 1988. Automatically derived catchment boundary and channel networks and their hydrological applications. *Geomorphology,* 1(2):131–141.

O'Callaghan, J. K., and D. M. Mark, 1984. The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing,* 28:323–344.

Richards, N. S., 1981. General problems in morphology. In Gordie, A. S., Ed., *Geomorphological Techniques*. Allen and Umnin, London, UK, pp. 26–30.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union*, 38(6):913–920.

Tribe, A., 1991. Automated recognition of valley heads from digital elevation models. *Earth Surface Processes and Landforms*, 16(1):33–49.

CHAPTER 8

Regional Characterization of Inland Valley Agroecosystems in West and Central Africa Using High-Resolution Remotely Sensed Data

Prasad S. Thenkabail and Christian Nolte

BACKGROUND

Inland valleys (IVs) (locally also known as fadamas, bas-fonds, and dambos) have the potential to become agroecosystems with a substantial impact on African food production (Izac et al., 1991). These agroecosystems are favorable for rice cultivation and dry season cropping and have the potential to increase acreage and yields in Africa if careful attention is paid to technical, environmental, and socioeconomic constraints (Juo and Lowe, 1986).

Inventorying, mapping and characterizing inland valleys at regional level (meso/semidetailed level, typically mapped on scales of 1:50,000 through 1:250,000) in sample areas of macrolevel agroecological zones (Figure 8.6, color section) is crucial for the selection of representative research sites, which then allows for the development of appropriate technologies that can be reliably tested and transferred to larger regions (regionalization or technology transfer).

Recognition of the importance of inland valley agroecosystems led the International Institute of Tropical Agriculture (IITA) to adopt a new research agenda, as put forth in Izac et al. (1991). This strategy involved a combination of biophysical and socioeconomic research issues to be addressed at three spatial scales:

a. level I (macro or continental or subcontinental, typically mapped in scales of greater than or equal to 1:5,000,000);
b. level II (meso or regional or semidetailed mapped in scales of 1:100,000 to 1:250,000); and
c. level III (micro or research site/watershed related, mapped in scales of less than or equal to 1:50,000).

This was conceived to facilitate the design of appropriate technology, able to sustain the highly varying resource base and at the same time to be acceptable to smallholder farmers in the diverse socioeconomic and ethnic environment of West and Central Africa.

The first step to characterize parameters critical to land use of IVs constituted the (macro) level I map based on secondary agroecological and soil data using a Geographic Information System (see Thenkabail and Nolte, 1995a for details). The objective was to map on a subcontinental

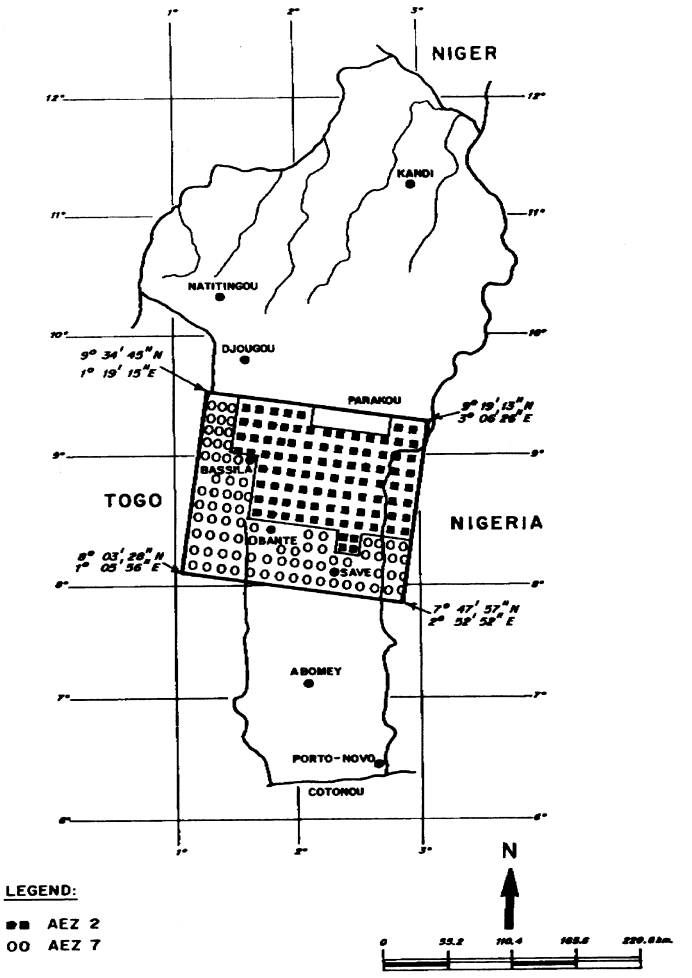


Figure 8.1 Spatial distribution of level 1 agroecological and soil zones (AEZ) in Landsat 192/54. Of a total area about 3.12 Mha, 49% fall into AEZ 2 and 46% into AEZ 7. The remaining 5% area is outside the 18 zones mapped in Figure 8.1.

(macro) scale broad agroecological and soil zones in the mandate area of IITA in West and Central Africa (Figure 8.1b). The level I map was the result of combining two parameters:

a. IITA's agroecological zones; and
b. major soil groupings according to the FAO classification (FAO/UNESCO 1974, 1977).

The five agroecological zones of IITA's map, namely northern Guinea savanna, southern Guinea savanna, derived and coastal savanna, humid forest, and midaltitude savanna, were overlaid with the 23 zones of major soil groupings. This resulted in 18 zones of more than 10 million ha each in West and Central Africa, for a total of 492 million hectares (Figure 8.6, color section).

Regional (or level II) characterization of inland valley agroecosystems were planned within the "windows" of macro (level I) zones. For rapid characterization and mapping at a regional scale of such a large area, which is spread across a subcontinent, it was only feasible using high-resolution

satellite images (Thenkabail and Nolte, 1995a; Thenkabail and Nolte, 1996). The location of the acquired Landsat TM and SPOT HRV satellite images for regional (level II) characterization study in relation to the IITA level I (macroscale) map are displayed in Figure 8.6 in color section.

Four international research centers (IITA, WARDA, CIRAD, Winand Staring Centre and Wageningen Agricultural University) and several national research systems (from Republic of Benin, Burkina Faso, Côte d'Ivoire, Ghana, Mali and Nigeria) acknowledged this holistic research approach to characterization of IVs at a workshop at WARDA headquarters, Bouaké, Côte d'Ivoire, June 1993, since it fit well with the ideas developed by these institutes.

RATIONALE AND JUSTIFICATION

Characterization of the agroecosystems in which IVs occur is available at subcontinental/macrolevel of West and Central Africa (Hekstra et al., 1983; Andriesse and Fresco, 1991; Izac et al., 1991; Windmeijer and Andriesse, 1993). Country/mesolevel studies of inland valley agroecosystems were published for the Republic of Benin by Kilian (1972), for Senegal by Bertrand (1973), for Burkina Faso and Mali by Albergel et al. (1993), and for Côte d'Ivoire by Becker and Diallo (1992). Turner (1985, 1977) gives mesolevel details upon biophysical data of fadamas in Central and Northern Nigeria, which were completed by the studies of Kolawole (1991) on fadama economics and management. Many studies upon dambos and their cultivation pattern have been done in South and Southeastern Africa. For example, Rattray et al. (1953) described vlei cultivation in Rhodesia/Zimbabwe. Other regional studies in this part of sub-Saharan Africa are reviewed by Ingram (1991).

Given the fact that the characteristics of IVs are known to vary dramatically within and across agroecosystems, it is inadequate to characterize only a few IVs as most conventional studies often do. Lack of representativeness of a few study sites in the context of a regional agroecological zone leads to limited extrapolation or regionalization of the results of key sites to other areas within the same agroecological zone.

Lack of an appropriate approach to characterization constitutes a major constraint to research activities in developing technologies that are able to sustain the resource base of large regions and are adoptable by diverse groups of farmers with highly differing socioeconomic and ethnic backgrounds. Therefore, IITA attempted to systematically characterize agroecosystems of IVs in West and Central Africa. This three-tier methodology from macro- to microlevel is intended to lead to the development of technologies at benchmark sites.

OBJECTIVES AND APPROACH

Level-II (regional) characterization work involves establishing detailed characteristics of IV agroecosystems in sample areas of level I. The specific objectives envisaged in level-II characterization fall within the overall objectives of Izac et al. (1991) and were:

1. inventory the area of inland valley systems and their uplands;
2. map the spatial distribution of IVs, and study their spatial variability;
3. study the existing land-use pattern of IVs and uplands, and establish their interactions;
4. explore the potential of IVs for dry-season cropping;
5. determine the existing crop types and cropping pattern in IVs, and on uplands;
6. map the major road systems and significant settlements;
7. establish the cultivation pattern with respect to distance from road network and settlements;
8. study the watershed, and establish morphometric characteristics of IV watersheds;

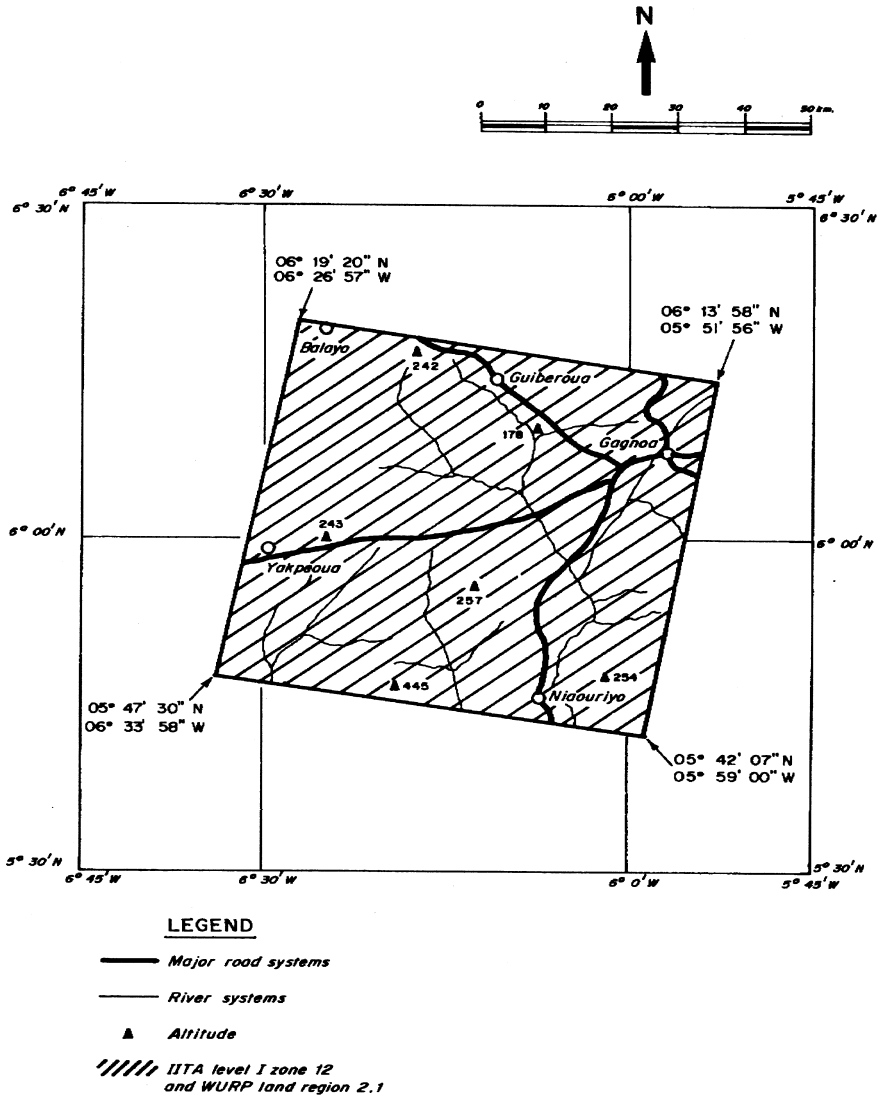


Figure 8.2 Spatial distribution of level 1 agroecological and soil zones in SPOT K:J of 47/338. Of the total area of).49 Mha, the entire 100% are falls in AEZ 16.

9. describe morphological IV characteristics such as shape, size, and slope;
10. make it possible to compare and contrast IVs within and across agroecological and socioeconomic zones in West and Central Africa;
11. rationalize the selection of representative benchmark site/s or benchmark IV watershed/s for the main phase research (i.e., technology development) activity; and
12. reveal the socioeconomic conditions of IV use. (Note that this is not part of the ongoing study).

IITA adopted a specific approach to achieve the above specific level-II objectives through integration of high-resolution satellite data with Global Positioning System (GPS) data, and ground-truth data in a Geographic Information Systems (GIS) framework. A comprehensive methodology

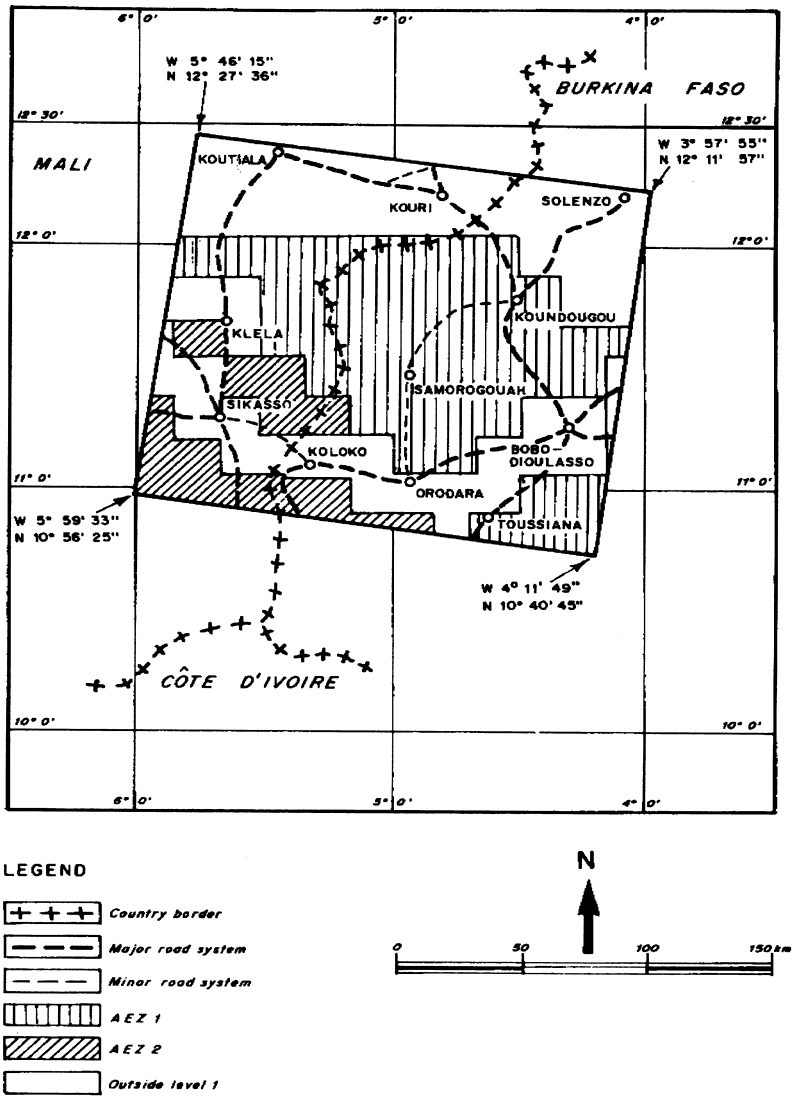


Figure 8.3 Spatial distribution of the level I agroecological and soil zones in Landstat 197/52. Of the total area of about 3.14 Mha., 45% is in AEZ 1 and 12% in AEZ 2. As for the rest of the area, 43% is outside the 18 AEZs mentioned in Figure 8.1.

for rapid characterization and mapping of inland valley systems at regional level using satellite imagery was developed and reported by Thenkabail and Nolte (1996, 1995a).

In this chapter two Landsat TM images (path 192 and row 54 and path 197 and row 52, [(see Figure 8.1and 8.2; and Table 8.1 for details and spatial location)], and one SPOT HRV image (Figure 8.3, 8.6 in the color section and Table 8.1)] were was used for the study. These images were selected, based on a set of criteria that included availability of cloud-free (0% cloud cover) satellite scenes, location of images in different agroecological and socioeconomic zones, most recently available dates of satellite overpasses (images were of early 1990's-see Table 8.1), accessibility of

the region for ground-truthing, and coverage during wet and dry seasons. The images covered four agroecological and soil zones (AEZ) of level I—AEZ 1, 2, 7, and 16.

Inland valleys, their fringes, and uplands distributed in each of the above AEZs were be reported along with their land-uses, cultivation intensities , and their spatial distribution.

The level II work offers :

a. the capability to zoom in on IV agroecosystems from macro- to microlevels ("top-down" approach) leading to the rationalization of benchmark site selection for technology development research activities; and
b. the capability to extrapolate research results to larger regions or the regionalization of research results ("bottom-up" approach) for the transfer of technology.

DEFINITION USED FOR MAPPING INLAND VALLEYS

It is obvious from the various accounts in the literature that there is no single, widely accepted definition of IVs. They (IVs) are one of the many forms of wetlands that are characterized in their bottom by hydromorphic soils. Beyond this, there are significant variations in the definitions and characteristics of IVs even within the same region as perceived by different people (e.g., Savvides, 1981; Raunet, 1982; Hekstra et al., 1983; Acres et al., 1985; Mäckel, 1985; Turner, 1985; Andriesse, 1986; Oosterbaan et al., 1987; Andriesse and Fresco, 1991; Izac et al., 1991; Mokadem, 1992). For example, Mäckel (1985) defines a uniform zonation of dambos in Southeast Africa depending on vegetation, soil type, moisture content, and morphodynamics of the dambo.

Lack of a consistent definition of IVs has been a constraint in comparing or synthesizing different studies. To most researchers, characterization of IVs meant valley bottom characterization. Others included fringes also in their studies. Andriesse and Fresco (1991) proposed a physiohydrographic model of an inland valley.

The working definition adopted in this study will be as follows: Inland valleys (IVs) comprise valley bottoms and valley fringes (Figure 8.5). Valley bottoms are characterized by hydromorphic soils that constitute a relatively flat surface with or without a central stream. Valley fringes refer to areas along the slopes of the valley; rainfall either runs off above the surface of these areas or interflows horizontally on impervious subsurface layers toward the valley bottom and the central stream. Valley fringes, typically, have two distinct characteristic zones (Figure 8.5):

1. the lower part of the valley fringe immediately adjoining the bottoms that may have a high likelihood of a seasonal hydromorphic zone with significant potential for dry-season cropping; and
2. the upper part of the valley fringe with steeper slopes, in zones with less than 1400 mm rainfall (Guinea savanna) characterized predominantly by impervious layers (ironstones or carapaces) from which rainwater quickly runs off to the valley bottom). Soils in these upper portions of valley fringes dry out rapidly once the rains have ceased, and therefore, have no potential for dry-season cropping.

METHODOLOGY

An overview of the methodology has been shown in Figure 8.4 and described in detail in Thenkabail and Nolte (1995) and Thenkabail and Nolte (1996). The methodology permits a rapid characterization of large areas on a regional scale. The methodology includes the description of techniques for:

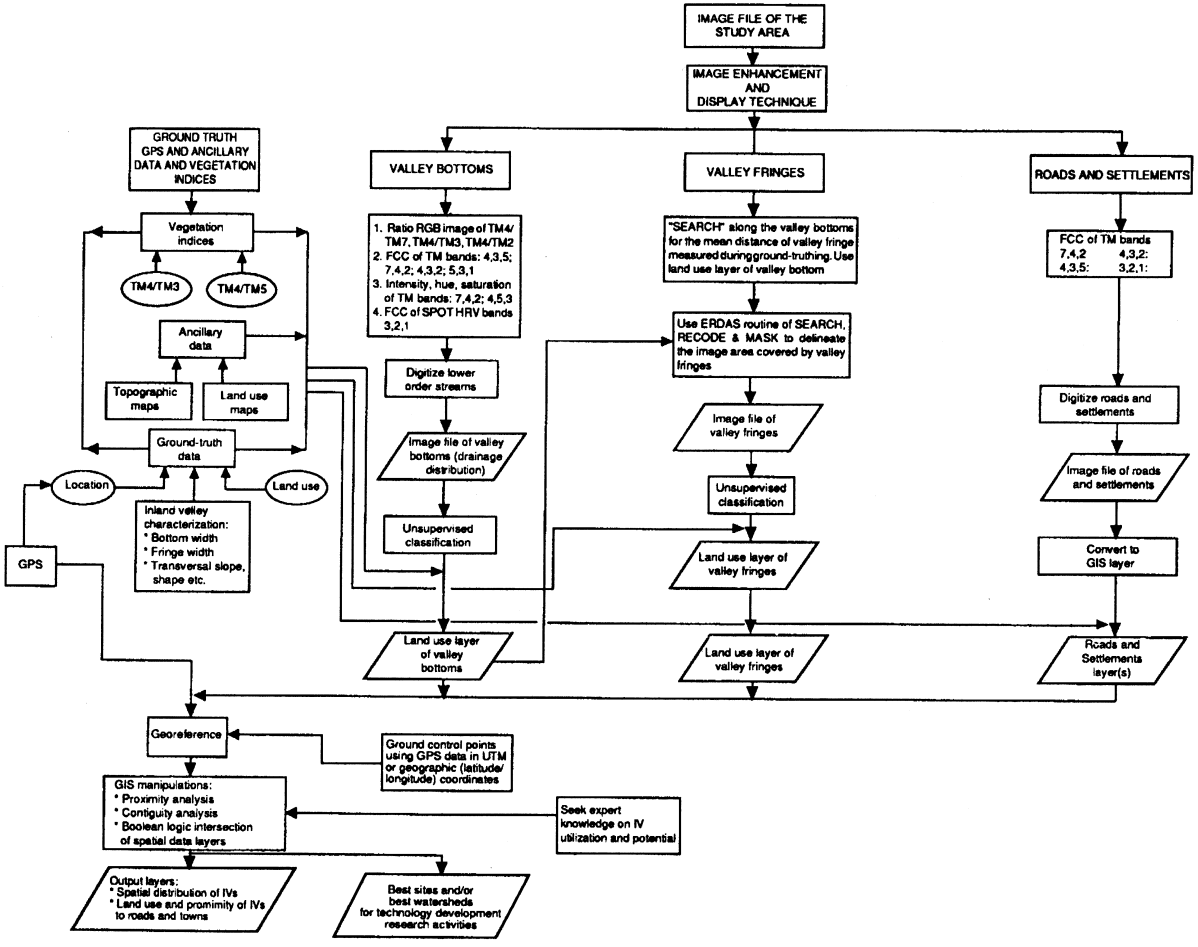


Figure 8.4 Methodology for mapping and characterizing Inland Valley (valley bottoms plus valley fringes) agroecosystems using Remote Sensing, Global Positioning System (GPS), and Ground-Truth Data.

1. valley bottom mapping;
2. valley fringe mapping;
3. mapping roads and settlements;
4. establishing land-use land-cover characteristics across the IV toposequence (uplands, valley fringes, and valley bottoms); and
5. establishing other IV characteristics such as IV stream densities and stream frequencies.

Figures 8.7 to 8.11 in the color section demonstrate the application of the methodology (Figure 8.4) for certain subareas of the study. Using the methodology, valley bottoms were highlighted (Figure 8.7, see color section) and delineated (Figure 8.8, see color section) for a subscene in Landsat TM path and row 192/54 that is part of agroecological and soil zone (AEZ) 7. The fringes adjoining the bottoms were mapped as illustrated in Figure 8.9 in the color section. A similar procedure was used to delineate valley bottoms from a SPOT subscene (Figure 10, see color section) which is in AEZ 16. Figure 8.11 in the color section (in AEZ 2) illustrates the large widths of the valley bottoms that are seasonally flooded and are characteristically dissimilar to valley bottoms in other AEZs. Sixteen land-use classes (Table 8.2) were mapped consistently across each study area based on the percentage distribution of 10 different land-cover types (Table 8.3).

Table 8.1 Parameters Describing the Level I Agroecological and Soil Zones[a]

Level I AEZ[b]	Agroecological zone According to IITA's Definition	LGP[c] (days)	Major FAO Soil Grouping[d]	Area[e] (million ha)
1	**Northern Guinea savanna**	**151–180**	**Luvisols**	**25.2**
2	**Southern Guinea savanna**	**181–210**	**Luvisols**	**18.4**
3	Southern Guinea savanna	181–210	Acrisols	12.4
4	Southern Guinea savanna	181–210	Ferralsols	11.9
5	Southern Guinea savanna	181–210	Lithosols	10.7
6	Derived savanna	211–270	Ferralsols	47.2
7	**Derived savanna**	**211–270**	**Luvisols**	**24.9**
8	Derived savanna	211–270	Nitosols	14.2
9	Derived savanna	211–270	Arenosols	14.0
10	Derived savanna	211–270	Acrisols	11.7
11	Derived savanna	211–270	Lithosols	10.8
12	Humid forest	> 270	Ferralsols	150.1
13	Humid forest	> 270	Nitosols	27.2
14	Humid forest	> 270	Gleysols	19.2
15	Humid forest	> 270	Arenosols	18.9
16	**Humid forest**	**> 270**	**Acrisols**	**18.0**
17	Mid-altitude savanna[b]		Ferralsols	45.4
18	Mid-altitude savanna[g]		Nitosols	12.3

[a]AEZs in bold have been investigated in this study at next level (level II) using high resolution satellite imagery, and the results are reported here
[b]AEZ: level I agroecological and soil zones.
[c]LGP: length of growing period.
[d]Names refer to the FAO soil classification scheme of 1974 (FAO/UNESCO 1974).
[e]The area figures are for West and Central Africa.
[f]Area distribution of LGP in AEZ 17 is: 151–180 days 11%, 181–210 days 9%, 211–270 days 59%, > 270 days 21%.
[g]Area distribution of LGP in AEZ 18 is: 151–180 days 2%, 181–210 days 5%, 211–270 days 53%, > 270 days 40%.

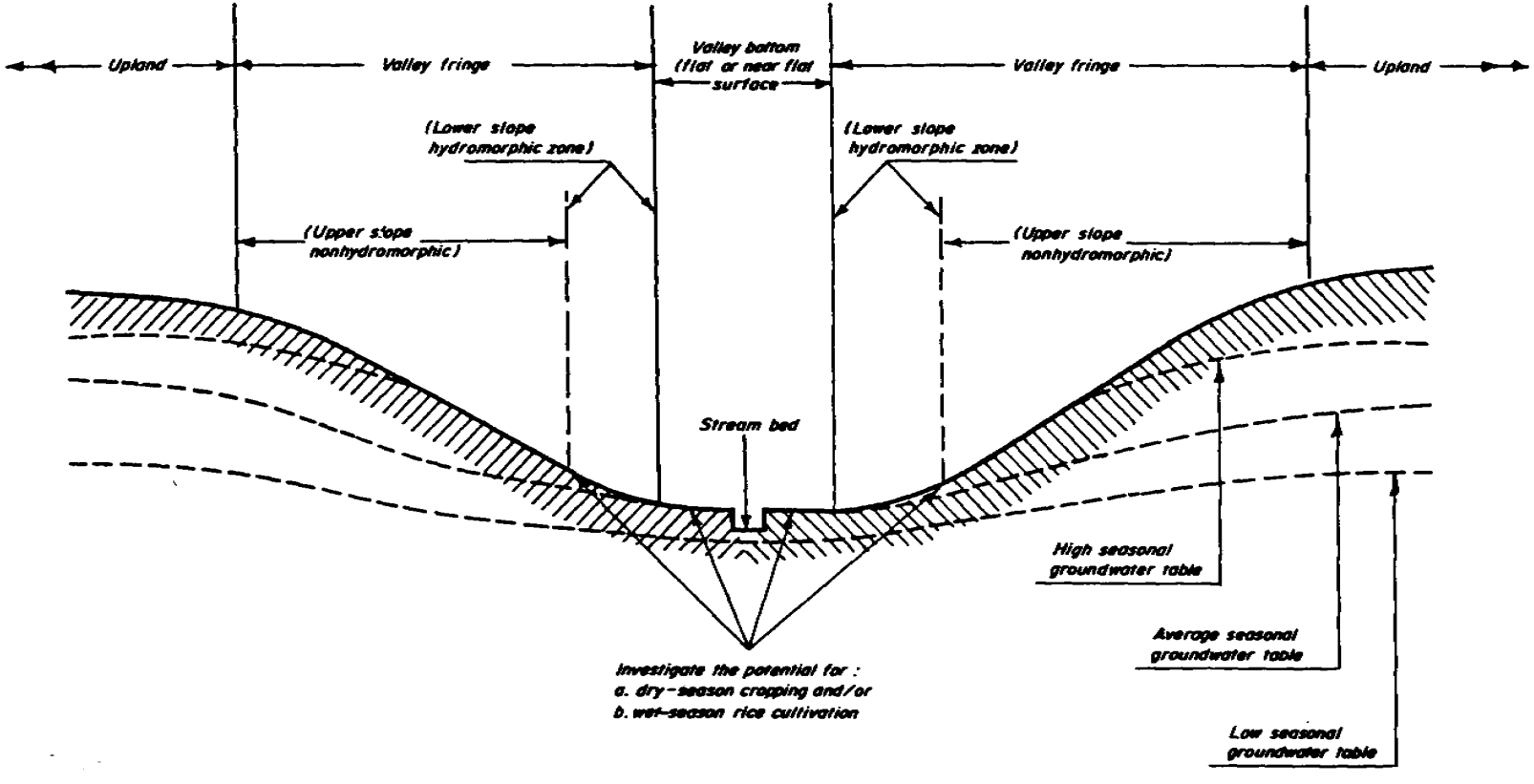


Figure 8.5 Cross-section showing model inland valley as defined in this study.

Table 8.2 Land-Use classes Mapped in Level-II Characterization of IV Agroecosystems of West and Central Africa

Code	Land-Use Class Description	Designated Color Derived Vegetation Indices
	Upland	
1	significant farmlands	gray
2	scattered farmlands	seafoam
3	insignificant farmlands	violet
4	wetland/marshland	mocha
5	dense forest	rose
6	very dense forest	red-orange
	Valley Fringe	
7	significant farmlands	white
8	scattered farmlands	pine-green
9	insignificant farmlands	red
	Valley Bottom	
10	significant farmlands	cyan
11	scattered farmlands	yellow
12	insignificant farmlands	magenta
	Others	
13	water	blue
14	built-up area/settlements	tan
15	roads	navy
16	barren land or desert land	sand

Table 8.3 Land-Cover Types Identified in Level-II Characterization of IV Agroecosystems of West and Central Africa

Code	Land-Cover Type Description	Code	Land-Cover Type Description
1	water	6	barren farms
2	tree	7	barren lands
3	shrub	8	built-up area/settlement
4	grass	9	roads
5	cultivated farms	10	others

STUDY AREAS AND GROUND-TRUTHING

In this chapter results of the three study areas (each satellite image comprising a study area) have been discussed (Table 8.4 and 8.1; Figures 8.1, 8.2, and 8.3; Figure 8.6, see color section). These study areas cover four distinct agroecological and soil zones (AEZ) in:

1. AEZ 2 (49% of the total area) and AEZ 7 (46%) by Landsat path 192 and row 54 (having an total area of 3.12 million hectares) (Figures 8.1, 8.6, see color section, Table 8.1 and Table 8.4);
2. AEZ 16 (100% of the total area) by SPOT HRV K 47 and J 338 (having an total area of 0.39 million hectares); (Figures 8.2, 8.6, see color section, Table 8.1 and Table 8.4) and
3. AEZ 1 (45% of the total area) and AEZ 2 (12%) by Landsat path 197 and row 52 (having an total area of 3.14 million hectares) (Figure 8.3, 8.6, see color section, Table 8.1 and Table 8.4).

Table 8.4 Attributes of Landsat TM and SPOT HRV Data Acquired for the Level II Characterization of Inland Valley Agroecosystems in West and Central Africa [a]

Landsat TM Path	Landsat TM Row		Center Coordinates		Date of
SPOT HRV K	SPOT HRV J	Region / Country	longitude	latitude	Overpass
192	54	Save / Republic of Benin	02° 06′ E	08° 41′ N	12/29/91
47	338	Gagnoa / Côte d'Ivoire	06° 12′ W	06° 00′ N	12/27/90
197	52	Sikasso / Mali	04° 59′ W	11° 34′ N	09/27/91

[a](See their location in Figure 8.1)

Ground-truthing was conducted to correspond with satellite overpass dates during 1993 and 1994 (Table 8.4). The location of each ground-truth site was determined using a global positioning system (GPS) Garmin 100-SRVY. Locations noted were geographic co-ordinates (latitude/longitude) in degree, minutes and seconds, and universal transverse mercator (UTM) coordinates (x,y) in meters. The accuracy of these GPS readings was within ± 30 m. The GPS was also used to collect ground-control points to georeference the satellite image.

Land-use measurements were made in a 90 m by 90 m plot of the valley bottoms, valley fringes, and uplands. Its location was determined from the center with the GPS. The leaf area index of trees, shrubs, and grasses was measured in the same plot. Each land-use class has varying degrees of land cover types (Table 8.3). A total of 10 land-cover types was recorded at each ground-truth site: water, trees, shrubs, grasses, cultivated farms, barren farms, barren lands, built-up areas or settlements, roads, and others. Different combinations of these landcover types constitute specific land-use categories (Table 8.2).

The measurements (in meters) of valley-bottom and valley-fringe width were taken at a transect covering the 90 m by 90 m plot for land-use measurements. Other characteristics recorded at each inland valley sample site included valley bottom width (in meters), valley fringe width, transversal slope (degree), stream order (number), and qualitative observations of the soil moisture status as well as the level of water management system prevailing in valley-bottom fields.

RESULTS AND DISCUSSIONS

The results of each of the three study areas have been discussed seperately in the subsections below. Data from Landsat and SPOT satellites for the study areas described below were integrated with GPS location data and ground-truth data in a GIS framework. The spatial data layers were georeferenced to Universal Transverse Mercator (UTM) coordinates obtained during ground-truthing using a GPS. Digital georeferenced databases generated include spatial distribution of inland valleys and land-use classes of inland valleys and uplands.

Save (Republic of Benin) study area

The study area [Figures 8.1, 8.6 (see color section), Table 8.1, and 8.4] encompassed two level I (macro or subcontinental) agroecological and soil zones: 1.AEZ 2 (southern Guinea savanna with Luvisols) and AEZ 7 (derived savanna with Luvisols) [see Table 8.1, Figures 8.1, and 8.6 (see

color section)]. AEZ 2 is representative of 18.4 million ha and AEZ 7 of 24.9 million ha in West and Central Africa (see spatial distribution of these zones in Figure 8.6 in the color section and Table 8.1). Both these agroecological and soil zones have Luvisols as major soil grouping. The zones differ in their length of growing period, with AEZ 2 having 181–210 days and AEZ 7 having 211–270 days. The study area is part of a region that gets rainfall varying between 975 mm and 1336 mm per year. Of the total study area 49% falls into AEZ 2 and with 46% into AEZ 7, while the remaining 5% is located outside the level I zones.

A total of 9.0% (281,500 ha) of the entire study area (3.12 million ha) is occupied by valley bottoms (Figure 8.12, see color section and Table 8.5). The corresponding numbers for the subzones were 7.9% in AEZ 2 and 10.2% in AEZ 7. Valley fringes accounted for 18.3% in AEZ 2, 21.8% in AEZ 7, and 19.9% in the entire study area, while uplands represented 73.3% in AEZ 2, 67.1% in AEZ 7, and 70.3% in entire study area. These results indicated a greater percentage of inland valleys in the wetter zone (AEZ 7) compared to the drier zone (AEZ 2).

The cultivation intensity differed only marginally between the AEZs (Table 8.5). A total of 7.9% of the valley bottoms in the entire study area were cultivated with 7.9% in AEZ 2 and 7.7% in AEZ 7. A total of 15.9% of the valley fringes was cultivated in the entire study area with 15.1% in AEZ 2 and 16.6% in AEZ 7. Compared to the valley fringes, less of the upland area was used as farmland, 13.2% across the entire study area, 11.2% in AEZ 2, and 14.3% in AEZ 7. These results indicated a low percentage of cultivation across the toposequence. This was attributed to: (a) a low density of the road network and (b) a low number of settlements, hence low population densities. In addition, socioeconomical issues such as the traditional preference of farmers in cultivating uplands, debilitating diseases in lowlands, lack of appropriate technologies (e.g., low-cost water management techniques), lack of knowledge about lowland agriculture, lack of market access, and lack of governmental support for farmers imply less than optimal utilization of inland valleys for agriculture.

TM data facilitated mapping the shape, size, spatial distribution, density, and frequency of inland valley bottoms occurring in the study area. The spatial distribution of these characteristics is illustrated in Figure 8.12 in the color section. They were also found to be very useful in detecting settlements (usually of sizes of >15 ha) and the motorable road network. This was especially useful in mapping the recent spread of settlements and the road systems in the absence of such information on base maps, such as topographic maps (which were mostly available from the early 1960s). Furthermore, the cultivation pattern of inland valleys with respect to their distance from settlements and the road network could be studied.

There was a high spatial correlation in the cultivation pattern of valley bottoms and valley fringes compared with the cultivation pattern of uplands. Significantly cultivated areas are mainly concentrated along major road networks and near settlements. The degree of cultivation decreased significantly as the distance from settlements or the road network increased. Stream drainage densities (km/km^2) were denser in the wetter zone (AEZ 7: 1.07 to 1.20 km/km^2) compared to the drier zone (AEZ 2: 0.9 to 1.0 km/km^2). The same was true for inland valley stream frequencies (number of streams per km^2): 0.95 to 1.10 streams/km^2 in AEZ 7; and 0.74 to 0.79 streams/km^2 in AEZ 2.

About 90% of the inland valleys were classified as U-shaped. Only about 17% of those surveyed during ground-truthing had some improved water management schemes, such as field leveling and bunding. The valley bottom widths of first- to third-order valleys in AEZ 2 and AEZ 7 varied between 40 m and 81 m and hence provided considerable valley-bottom areas (7.7% to 9.0% of the geographic area). The valley bottoms are flat to near-flat. Most valley bottoms are characterized by grassy vegetation that could be easily cleared for cultivation.

Land-use characteristics were mapped separately for valley bottoms, valley fringes, and up-

Table 8.5. Distribution of Valley Bottoms, Fringes, and Uplands, and Their Cultivation Pattern in the Save (Republic of Benin) Study Area[a]

				Valley-Bottom Area		Valley-Fringe Area		Upland Area	
	Subarea	Level I Agroecological and Soil Zones	% of Entire Study Area (% of Full Scene)	as a % of Total Geographic Area (%)	Cultivated as a % of Total Valley-Bottom Area (%)	as a % of Total Geographic Area (%)	Cultimated as a % of Total Valley-Fringe Area (%)	as a % of Total Georgraphic Area (%)	Cultivated as a % of Total Upland Area (%)
A.	Parakou and Tchaourou region of Rep. of Benin	AEZ 2	49	7.9	7.9	18.3	15.1	73.3	11.2
B.	Save/Bante/ Gonka/Bassila region	AEZ 7	46	10.2	7.7	21.8	16.6	67.1	14.3
C.	Entire study area	land region 2.1	100	9.0	7.9	19.9	15.9	70.3	13.2

Notes:
[a]AEZ 2 represents the southern Guinea savanna with Luvisols and AEZ 7 the Derived savanna with Luvisols. AEZ 2 occupies 49% and AEZ 7 46% of the entire study area. The remaining 5% is located outside the level I zones.

Table 8.6 Distribution of Land-Use Classes in the Different Agroecological and Soil Zones of the Save (Republic of Benin) Study Area[a]

		AEZ 2		AEZ 7		Entire Study Area	
	Land-Use Class	Area (ha)	% of Total AEZ 2	Area (ha)	% of Total AEZ 7	Area (ha)	% of Total Area
	Uplands						
1	significant farmlands	57,853	3.8	148,609	10.3	221,399	7.1
2	scattered farmlands	547,694	35.7	492,119	34.2	1,099,828	35.2
3	savanna vegetation[b]	233,120	15.2	155,091	10.8	407,026	13.0
4	wetlands/marshland	129,808	8.5	78,665	5.5	218,185	7.0
5	dense forest	126,096	8.2	63,876	4.4	194,141	6.2
6	very dense forest	30,498	2.0	28,445	2.0	60,052	1.9
	Valley fringes						
7	significant farmlands	12,550	0.8	32,452	2.3	47,861	1.5
8	scattered farmlands	234,627	15.3	250,003	17.4	507,711	16.2
9	insignificant farmlands[c]	33,846	2.2	31,807	2.2	67,597	2.2
	Valley bottom						
10	significant farmlands	2,012	0.1	4,943	0.3	7,537	0.2
11	scattered farmlands	78,174	5.1	82,934	5.8	169,027	5.4
12	insignificant farmlands[d]	41,813	2.7	59,160	4.1	103,895	3.3
	Others						
13	water	402	0.0	629	0.0	1,223	0.0
14	settlements/built-up area	668	0.0	1,279	0.1	4,967	0.2
15	roads	1,856	0.1	1,858	0.1	3,929	0.1
16	barren/desert area	3,938	0.3	8,324	0.6	13,999	0.5
	Total	1,534,955	100.00	1,440,193	100.00	3,128,377	100.00

[a]Level I agroecological and soil zones (see Figure 8.1 and Table 8.1). AEZ 2 represents the southern Guinea savanna with Luvisols and AEZ 7 the Derived savanna with Luvisols. AEZ 2 occupies 49% and [b]AEZ 7 46% of the entire study area. The remaining 5% is located outside the level I zones.
[c]Class 3 occurs only in Guinea savanna zones.
[d]Spectral characteristic of vegetation in class 9 is similar to that of classes 5, 6, and 12; the difference is mainly in the toposequence position,
[e]Mainly riparian vegetation; spectral characteristics of vegetation similar to classes 5, 6, and 9; the difference is mainly in the toposequence position,

lands using the CLUSTR unsupervised classification algorithm of ERDAS (ERDAS, 1992). The areas of each of the 16 classes (Figure 8.15, see color section) in the two agroecological and soil zones and in the entire study area are given in Table 8.6. These land use classes are designed on the lines of Anderson et al. (1976). Each land-use class contains a varying degree of land-cover types (Table 8.3).

Gagnoa (Cote d'Ivoire) Study Area

In this study SPOT–1 HRV data for location K:47, J:338 (of the SPOT grid reference system; see precise coordinates in Figure 8.4) with 27 December 1990 as the date of overpass covering an area of 393,112 ha in the Gagnoa region of southwestern Côte d'Ivoire (Figure 8.4).

Table 8.7. Distribution of Valley Bottoms, Valley Fringes, and Uplands and Their Cultivation Status in the Gagnoa (Côte d'Ivoire) Study Area[a]

	Valley Bottom Area		Valley Fringe Area		Upland Area	
Entire Study Area	as a % of Total Geographic Area (%)	Cultivated as a % of Total Valley-Bottom Area (%)	as a % of Total Geographic Area (%)	Cultivated as a % of Total Valley-Fringe Area (%)	as a % of Total Geographic Area (%)	Cultivated as a % of Total Upland Area (%)
100	18.0	20.6	40.3	16.9	40.1	15.0

[a]The study area falls entirely into agroecological zone 16 of the level I map (Figure 8.1and Table 8.1).

The entire study area [Figure 8.2, and 8.6 (see color section) Table 8.1, and Table 8.4] falls within a single level I (macro) agroecological and soil zone 16 (AEZ 16): the humid forest zone with Acrisols as the major soil grouping according to the FAO soil map of the World (FAO/-UNESCO, 1977) (see Figure 8.2 and Table 8.1). As a result, the characteristics reported for the entire study area will also be true for the macroscale zones AEZ 16 (Figure 8.2).

The study area comprises 18% valley bottoms (Figure 8.13, see color section), 40.3% valley fringes, 40.1% uplands, and 1.6% others (e.g., roads, settlements) (Table 8.7). The area covered by valley bottoms (18%) was relatively high as a result of large bottom widths with mean values of 89 m (first-order valleys), 164 m (second-order), 173 m (third-order), and 244 m (fourth-order). These measurements showed a high variability across the stream orders. First-order valleys had significantly smaller valley bottoms than higher-order valleys. The inland valley stream density in the study area was determined as 0.97 km/km^2 (medium as per Hekstra et al., 1983) and the stream frequencies were determined to be 0.75 streams per km^2 (coarse as per Hekstra et al., 1983). The mean valley fringe widths were 300 m (first-order valleys), 303 m (second-order), 355 m (third-order), and 436 m (fourth-order).

The 16 land-use classes (Figure 8.16, see color section)– six for uplands, three each for valley fringes and valley bottoms, and four others—were mapped for the entire study area (Table 8.8). The total area covered by humid forest vegetation with insignificant farmlands (land-use classes 5, 6, 9, and 12) was 58.3%, whereas the area with humid forest-cropland mosaic (land-use classes 2, 8, and 11) was 23.0% of the total geographic area. The intensity of cultivation was significantly higher for valley bottoms (20.6%) compared with valley fringes (16.9%), and uplands (15%) (Table 8.7). This was mainly a result of the use of valley bottoms for rice cultivation, as revealed by ground-truthing. In the entire study area of the image, the most significant areas of farmlands were in the lower left corner (southwestern portion) of the image near the settlement of Yakpéoua and around predominant settlements, such as Gagnoa and Guibéroua (Figure 8.4). In significant portions of the image there was a strong relationship in the cultivation of uplands and inland valleys.

As expected in a low population density area that belongs to the humid forest zone (Figure 8.6 in color section), the forest vegetation (classes 5, 6, 9, and 12) dominates the image. In the classification system used in this study, the characteristics of areas with insignificant farmlands at valley fringes (class 9) and in valley bottoms (class 12) are similar to those areas with dense and very dense forest on uplands (classes 5 and 6). The only difference is in the toposequential position of the area under consideration. The vegetation of these classes mostly consists of dense and vigorous evergreen trees, and dense and vigorous shrubs. Grasses were less common and farmlands were insignificant in these classes.

Table 8.8 Land-Use Distribution in the Gagnoa (Côte d'Ivoire) Study Area[a]

		Full Study Area	
No.	Land-Use Category	Area (ha)	% of Total Study Area
	Uplands	157,601	40.1
1	significant farmlands	22,589	5.8
2	scattered farmlands	31,992	8.1
3	savanna vegetation[b]	0	0
4	wetlands/marshland	7,024	1.8
5	dense vegetation	54,619	13.9
6	very dense vegetation	41,377	10.5
	Valley fringes	158,606	40.3
7	significant farmlands	26,299	6.7
8	scattered farmlands	39,376	10.0
9	insignificant farmlands[c]	92,931	23.6
	Valley bottom	70,638	18.0
10	significant farmlands	11,490	2.9
11	scattered farmlands	19,058	4.9
12	insignificant farmlands[d]	40,090	10.2
	Others	6,268	1.6
13	water	358	0.1
14	built-up area/settlements	2,703	0.7
15	roads	2,194	0.5
16	barren land or desert lands	1,013	0.3

Notes:
[a]The study area falls entirely into agroecological zone 16 of the level I map (Figure 8.1 and Table 8.1)
[b]Class 3 occurs only in Guinea savanna zones.
[c]Spectral characteristic of vegetation in class 9 is similar to that of classes 5, 6, and 12; the difference is mainly in the toposequence position.
[d]Mainly riparian vegetation; spectral characteristics of vegetation similar to classes 5, 6 and 9; the difference is mainly in the toposequence position.

Areas on uplands, at fringes, and in valley bottoms with forest and savanna vegetation and scattered farmlands (classes 2, 8, and 11) have > 10% and < 30% of cultivation and can be termed as forest-cropland mosaic. This mosaic occurs especially in the middle and western portions of the image, covers about two-thirds of the overall image, and is typically to be found around settlements and major road networks (areas around Gagnoa and Guibéroua in Figure 8.4).

The forest-cropland mosaic land-use classes (2, 8, and 11) contain a significantly higher percentage of farmlands (land-cover types 5 and 6) and a lower percentage of trees and shrubs (land-cover types 2 and 3), but they still have a significant amount of forest vegetation. The overall area under forest-cropland mosaic represents 23.0% of the total geographic area.

Ninety-two percent (92%) of the inland valleys (valley bottoms plus valley fringes) were U-shaped. Ninety percent of these valleys were considered suitable for cultivation, which means that the bottom width was reasonably large enough (> 5–10 m wide and more or less flat). Low-level water management systems, such as leveling and bunding of fields were found in 25% of all the valleys. The characteristics of valley bottoms, such as their widths, spatial distribution (density,

frequency, and location), and land-use characteristics, were mapped for the entire study area as displayed in Figure 8.16, see color section.

The characteristics of inland valleys in the region clearly indicate a vast potential for agricultural development (Table 8.7): over 80% of the valley bottoms and valley fringes are currently unexploited, they offer wide bottom widths, adequate water resource, and are relatively easy to clear compared to uplands due to the presence of significantly fewer trees.

All major and minor roads discernible on SPOT data were mapped, providing a road network density of 0.16 km/km^2 which is considered as a good road infrastructure by Manyong et al (in preparation). Ninety-three settlements of varying sizes were also mapped based on SPOT data, constituting a significant portion of the overall settlements in the study area. The relationship between intensity of cultivation and the distance of farmlands from roads and settlements was established. The intensity of cultivation at all the components of the toposequence (valley bottoms, valley fringes, and uplands) was slightly lower at distance zones beyond 4 km compared with distances within 4 km. This was due to the presence of a large number of well connected small road systems, equally well distributed small settlements, and therefore good access to markets. The spatial distribution of these characteristics are illustrated in Thenkabail and Nolte (in press, 1995c).

Sikasso (Mali), Bobo-Dioulasso (Burkina Faso) Study Area

In this study area Landsat–5 Thematic Mapper (TM) Path: 197, Row: 52, with a total area of 3.13 million hectares, (see the exact coordinates in Figure 8.3) was used. The total study area (3.13 million hectares) [Figure 8.3, 8.6 (see color section) Table 8.4] comprised 8.6% valley bottoms, 20.4% valley fringes, and 70.2% uplands (Table 8.9). Water bodies, roads, and settlements comprised the other 0.8% area (Table 8.10). The drainage density of 0.4 km per km^2, and stream frequency of 0.61 number per km^2 obtained in the study area were classified as low (0.3 to 0.6 km/km^2), and coarse (0.5 to 1.0 number/km^2), respectively, by Hekstra, et al. (1983). In spite of the low and coarse drainage densities, and stream frequencies in the study area, the percentage area of inland valleys (valley bottoms plus valley fringes) were significant mainly as a result of large valley bottom and fringe widths of the inland valley streams (first- to fourth-order streams). The mean bottom widths for the first- to third-order stream were about 90 meters, increased dramatically for the fourth-order to about 400 meters. The mean valley fringe (hydromorphic plus nonhydromorphic) widths for the first three inland valley streams were about 200 meters, and for the fourth-order stream about 920 meters. Hence, even though the stream frequencies and stream densities were coarse and low, respectively, the large sizes of the valley bottoms and valley fringes led to their significant percentages.

Due to significant differences in the geographical areas studied (45% of entire study area for AEZ 1, 12% for AEZ 2; Table 8.9) a direct and realistic comparison of results across zones was not feasible. However, it may be noted that the valleys in AEZ 1 had greater bottom widths than valleys in AEZ 2, resulting in higher percentage area of valley bottoms (Figure 8.14, see color section) in AEZ 1 (9.1 %) compared to AEZ 2 (7.7 %). Due to the same reason, the percentage valley bottom area in land region 2.8 (9.1%) exceeded that of land region 3.3 (8.2%).

The study mapped 16 land use classes [Figure 8.17 (see color section) and Table 8.10] in the total study area of 3.13 million hectares and for the subzones [Figure 8.19 (see color section)]. The spatial distribution of these characteristics are illustrated in Thenkabail and Nolte (1995d).

The grassland dominant savannas dominate the study area. The overall savanna percentage areas were 65.5% for AEZ 1, 70.4% for AEZ 2, 67.1% for land region 3.3, 61.6% for land region 2.8, and 65.4% for the entire study area. The forest classes are predominantly trees along the river bank and were about 4% for all level I zones within the study area. This very low percentage of

Table 8.9. Distribution of Valley Bottoms, Valley Fringes, and Uplands and Their Cultivation Status in the Sikasso (Mali), and Bobo-Dioulasso (Burkina Faso) Study Area

		Valley Bottom Area		Valley Fringe Area		Upland Area	
Study Area	% of Entire Study Area	as a % of Total Geographic Area Scene)	Cultivated as a % of Total Valley-Bottom Area (%)	as a % of Total Geographic Area (%)	Cultimated as a % of Total Valley-Fringe Area (%)	as a % of Total Georgraphic Area (%)	Cultivated as a % of Total Upland Area (%)
AEZ 1[b]	45	9.1	15.9	17.8	16.0	72.2	20.5
AEZ 2	12	7.7	24.0	25.5	21.8	66.5	24.1
Entire study area	100	8.6	18.4	20.4	19.2	70.2	21.9

Notes:
[a]When valley bottoms + valley fringes + uplands are not equal to 100%, the rest of the area falls in water body, roads, and settlements or "round-off" errors.
[b]Level I agroecological and soil zones (see Figure 8.1).

Table 8.10. Land-Use Distribution in the Different Agroecological and Soil Zones (AEZ) Determined in the Sikasso (Mali), and Bobo-Dioulasso (Burkina Faso) Study Area

		AEZ 2		AEZ 7		Entire Study Area	
No.	Land-Use Class	Area (ha)	% of Total AEZ 2	Area (ha)	% of Total AEZ 7	Area (ha)	% of Total Area
	Uplands						
1	significant farmlands	105,468	7.4	32,124	8.3	280,990	9.0
2	scattered farmlands	367,370	25.9	141,550	36.4	901,130	28.7
3	savanna vegetation	406,472	28.7	63,360	16.3	727,987	23.2
4	wetlands/marshland	46,023	3.2	3,433	0.9	85,202	2.7
5	dense vegetation	51,818	3.7	12,078	3.1	79,955	2.6
6	very dense vegetation	10,797	0.8	1,172	0.3	18,175	0.6
	Valley fringes						
7	significant farmlands	31,561	2.2	17,789	4.6	104,879	3.3
8	scattered farmlands	95,177	6.7	52,260	13.4	282,455	9.0
9	insignificant farmlands	126,930	8.9	29,363	7.5	254,088	8.1
	Valley bottoms						
10	significant farmlands	7,717	0.5	4,863	1.2	21729	0.7
11	scattered farmlands	59,062	4.2	16,792	4.3	139,758	4.5
12	insignificant farmlands	62,045	4.4	8,726	2.2	107,519	3.4
	Others						
13	water	9,750	0.7	544	0.1	12,570	0.4
14	built-up area/settlements	299	0.0	82	0.0	5,849	0.2
15	roads	2,170	0.2	498	0.1	6,585	0.2
16	barren/desert area	35,643	2.5	4,600	1.2	106,985	3.4

forest cover was only to be expected in the study area as it falls in the northern guinea savanna and sudan savanna. The study areas in the sudan savanna comprised of maximum barren areas of 6% when compared to other level I zones studied.

The cultivation intensities were nearly the same across the toposequence with 18.4% for valley bottoms, 19.2% for valley fringes, and 21.9% for uplands (Table 8.9). The significant cultivation across the toposequence were mainly attributed to the market-driven conditions. In most cases cultivation intensities were about 3% higher for distance limits within 0–5 km from road network and settlements compared to those areas beyond 5 km.

The valley bottoms in the study area were characterized by flat or near-flat surfaces that have shallow flooding all through the rainy season. Rice cultivation (Figure 8.19, see color section) forms an important component of lowland rainy season cultivation in the entire study area, especially in the valleys surrounding Sikasso, Mali. These broad and flat or near-flat valley bottoms of the study area offer an excellent opportunity for innundated rice cultivation (swamp rice) during rainy season. However, the area of valley bottoms available for cultivation far exceeds their current utilization. A total of 269,006 hectares constitute valley bottoms (Figure 8.14, see color section) in the entire study area, of which only 18.4% (49,497 hectares) was cultivated. Of the cultivated IVs 42% (20,789 hectares) had rice cultivation (Figure 8.19, see color section).

One hundred percent of the inland valleys that were studied were ‘U’ shaped, 74% were fadamas (that is, inland valleys with potential for dry season cropping). At the time of groundtruthing 69% of the inland valleys were wet, 21% were moist, and 10% were dry. A mapping of

biomass levels during summer will show: (a) upland and lowland contrasts in maximum as a result of significant moisture presence in lowlands relative to near dry uplands; (b) biomass variations within valley highlighting moisture availability in the valleys (see an illustration in Figure 8.21, see color section) when upland vegetation is dry. The mean transversal slopes were generally mild with about 1.5 degrees for the first- to third-order inland valley streams, and about 0.5 degrees for the fourth-order inland valley streams.

The study showed a strong relationship between upland cultivation and inland valley cultivation (Table 8.9) proving one of the hypotheses of this study.

SUMMARY

This study used high-resolution satellite images from Landsat TM and SPOT HRV in three separate study areas [Figure 8.1, 8.2, 8.3 and 8.6 (see color section), and Table 8.4) covering four agroecological and soil zones (AEZs; as shown in Figure 8.6 in the color section) of West and Central Africa to characterize and map inland valley agroecosystems. The characteristics studied include: inventorying the areas of inland valleys (valley bottoms plus valley fringes), determining the cultivation intensities of inland valleys and their uplands, and studying the land-use characteristics of inland valley bottoms and fringes, and their uplands.

The paper briefly outlined the holistic three-tier "top-down" approach developed by IITA to rapidly characterize and map inland valley agroecosystems of West and Central Africa at macro (level I), regional (level II), and micro (level III) scales. The focus of the paper was in applying the methodology for regional (level II) characterization and mapping of inland valleys in four major level I agroecological and soil zones (AEZ) using high-resolution satellite imagery from Landsat and SPOT systems (see Figures 8.7–8.11 in the color section for illustration of the methodology along with Figure 8.4). The three study areas were: (1). Save, Republic of Benin (Figure 8.1) covered by a Landsat TM image; (2). Gagnoa, Côte d'Ivoire (Figure 8.2) covered by a SPOT HRV; and (3). Sikasso, Mali/Bobo-Dioulasso, Burkina Faso (Figure 8.3) covered by a Landsat TM image. The three images cover four different agroecological and soil zones [AEZ 1, 2, 7, and 16; see Figures 8.1, 8.2, 8.3, and 8.6 (see color section), and Table 8.4] with a total area of about 6.6 million hectares.

The study made it possible to estimate inland valley (valley bottoms plus valley fringes) areas and their degree of cultivation (Table 8.5, 8.7, and 8.9). Inland valley bottoms (Figures 8.12, 8.13, and 8.14 in the color section) were highest in the humid forest zone (AEZ 16) with 18% (Figure 8.13 in the color section with Figure 2, and Table 7) of the total geographic area of study compared to 9.1% (Figure 8.14 in the color section with Figure 8.3, and Table 8.9) in northern Guinea savanna (AEZ 1), 7.7% (Figure 8.14 in the color section with Figure 8.3, Table 8.9) to 7.9% (Figure 8.12 in the color section with Figure 8.1, and Table 8.5) in southern Guinea savanna (AEZ 2), and 10.2% (Figure 8.12 in the color section with Figure 8.1, and Table 8.5) in the derived savanna (AEZ 7).

Currently, only 8% to 20% inland valley bottoms and 15% to 22% inland valley fringes are exploited for agriculture in AEZ 1, 2, 7, and 16 (Tables 8.5, 8.7, and 8.9). The vast percentage of unexploited inland valleys are rich in soil fertility, have potential for rice cultivation during main cropping season, and significant percentage of valleys have sufficient moisture to sustain upland crops during dry season.

Sixteen similar land-use classes were mapped cutting across the toposequence (uplands, valley fringes, and valley bottoms) for each study area (Tables 8.6, 8.8, and 8.10; and Figures 8.15, 8.16 and 8.17 in the color section). The toposequence oriented land-use mapping has been a unique feature of this study. It is indeed clear from this study that the inland valley bottoms that constituted

7.7% to 18% of geographic area have great economic significance with rich potential for agricultural and ecological exploitation [see an example in Figure 8.19 (see color section) for rice cultivation] and preservation and hence deserve to be mapped as separate land units. The land-use classification resulted in estimates of areas covered by savannas, forests, farmlands, wetlands, and other land-uses (water, settlements, road networks, and barren areas) apart from estimating percentage areas cultivated and the the percentage areas still available for exploitation.

The area of inland valleys was significantly higher in wetter agroecological zones compared to drier zones. For example, valley bottoms in humid forest (AEZ 16) were 18% (Figure 8.13, color section; Table 8.7) and in derived savanna (AEZ 7) were 10.2% (Figure 8.12, color section; Table 8.5) compared to much lower areas in the northern Guinea savanna (AEZ 1) with 9.1% (Figure 8.14, color section; Table 8.9) and southern Guinea savanna (AEZ 2) with 7.7% (Figure 8.14, color section; Table 8.9) and 7.9% (Figure 8.12, color section and Table 8.5). In each study area there was a strong relationship between inland valley and upland cultivation (Tables 5, 7, and 9).

There was no clear relationship between the degree of cultivation and the agroecological zones. One of the hypotheses (Izac et al., 1991) was that drier zones will have a greater degree of valley bottom utilization than the wetter zones as moisture availability in the valley bottoms is more crucial in drier zones than wetter zones. This hypothesis is not a practical reality as other factors play an important role in inland valley utilization. Availability of reliable rainwater supply during wet season, relatively better access to technologies, higher population densities, and higher densities of road networks and socioeconomic factors facilitated utilization of 20.6% (Table 8.7) of valley bottoms, mostly for rice cultivation, in Gagnoa study area. In contrast, the percentage utilization of valley bottoms were, generally, much lesser in drier agroecological zones of Save (Table 8.5) and Sikasso (Table 8.9) due to relative absence of above-mentioned factors.

Digital georeferenced databases generated from the study consisted of the spatial distribution of inland valleys, land-use characteristics of inland valleys and their uplands, and major settlements and road-networks. Other spatial data layers like GPS data, ground-truth data, and expert knowledge databases were integrated with remote sensing derived digital databases into a Geographic Information Systems (GIS) framework. Spatial modeling of these data layers resulted in deriving potential benchmark research areas for technology development research activities (Figure 8.20, color section).

REFERENCES

Acres, B.D., R.B. Blair, R.M. King, Lawton, A.J.B. Mitchell, and L.J. Rackham, 1985. African dambos: Their distribution, characteristics and use. *Zeitschrift für Geomorphologie*, Supplementband 52, Berlin-Stuttgart, pp. 63–86.

Albergel, J., J.M. Lamachère, B. Lidon, A.I. Mokadem, and W. van Driel, (Eds.), 1993. Mise en valeur des bas-fonds au Sahel. Typologie, Fonctionnement hydrologique, Potentialités agricoles. Rapport final d'un projet CORAF-R3S, Ouagadougou, Burkina Faso, CIEH.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. A Land-Use and Land-Cover Classification System for Use with Remote Sensor Data. Geological Survey Professional Paper 964, U.S. Government Printing Office, Washington, D.C.

Andriesse, W. 1986. Area and distribution. In A.S.R. Juo and J.A. Lowe. *Proceedings of an International Conference on Wetlands Utilization for Rice Production in Sub-Saharan Africa*, 4–8 November 1985, IITA, Ibadan, Nigeria, pp. 15–30.

Andriesse, W., and L. Fresco, 1991. A Characterization of rice-growing environments in West Africa. *Agriculture, Ecosystems, and Environment,* 33:377–395.

Becker, L., and R. Diallo, 1992. Characterization and Classification of Rice Agroecosystems in Côte d'Ivoire. WARDA/ADRAO, Bouaké, Côte d'Ivoire.

Bertrand, R., 1973. Contribution à l'étude hydrologique, pédologique et agronomique des sols gris sableux hydromorphes de Casamance (Sénégal). *L'Agronomie Tropical,* 28(12):1145–1192.

ERDAS, 1992. Field guide. ERDAS Inc., 2801 Buford Highway, Suite 300, Atlanta, GA.

FAO/UNESCO, 1974. Soil Map of the World, 1:5,000,000 Vol. VI. Africa. UNESCO, Paris.

FAO/UNESCO, 1977. Soil Map of the World. Revised Legend. FAO World Soil Resources Rep. 60, Rome.

Hekstra, P., W. Andriesse, G. Bus, and C.A. de Vries, 1983. Wetland Utilization Research Project, West Africa. Phase I. The Inventory: Vol. I, Main Report; Vol. II, The Physical Aspects; Vol. III, The Agronomic, Economic and Sociological Aspects; Vol. IV, The Maps. WURP-Report, ILRI/STIBOKA, Wageningen, the Netherlands.

Ingram, J., 1991. Wetlands in Drylands: The Agroecology of Savanna Systems in Africa. Part 2: Soil and Water Processes. International Institute for Environment and Development (IIED), Drylands Programme, London, UK.

Izac, A. -M. N., M.J. Swift, and W. Andriesse, 1991. A Strategy for Inland Valley Agroecosystem Research in West and Central Africa. RCMP Research Monograph No.5, Resource and Crop Management Program, IITA, Ibadan, Nigeria.

Izac, A.-M., S.S. Jagtap, A.I. Mokadem, and P.S. Thenkabail (in preparation). Agroecosystems characterization in International Agricultural Research: The case of inland valley systems in West and Central Africa. *Agriculture, Ecosystems, and Environment.*

Juo, A.S.R., and Lowe, J.A., 1986, The Wetlands and Rice in Sub-Saharan African. The International Institute of Tropical Agriculture, Ibadan, Nigeria.

Kilian, J. 1972. Contribution à l'Etude des Aptitudes des Sols à la Riziculture des Bas-Fonds sans Aménagements dans le Nord Dahomey (Bas-Fonds de Sirarou et de Bagou*). L'Agronomie Tropical* 27 (3): 321–357.

Kolawole, A., 1991. Wetlands in Drylands: The Agroecology of Savanna Systems in Africa. Part 3a: Economics and Management of Fadama in Northern Nigeria. International Institute for Environment and Development (IIED), Drylands Programme, London, UK.

Mäckel, R., 1985. Dambos and related landforms in Africa: An example for the ecological approach to tropical geomorphology. *Zeitschrift für Geomorphologie*, Supplementband 52, Berlin-Stuttgart, 1–24.

Mokadem, A.I., 1992. Apports de la Télédétection à l'étude des bas-fonds. Cas d'étude au Sierra Leone. Faculté des Sciences Agronomiques, Dép. du Génie Rural, U.E.R. Hydraulique Agricole, Laboratoire de Télédétection et d'Agrohydrologie, Gembloux, Belgique.

Oosterbaan, R.J., H.A. Gunnewig, and A. Huizing, 1987. In Water Control for Rice Cultivation in Small Valleys of West Africa, pp. 30–49. Annual Report 1986, ILRI, Wageningen, the Netherlands.

Rattray, J.M., R.M.M. Cormack, and R.R. Staples, 1953. The Vlei areas of Southern Rhodesia and their uses. *Rhodesian Agricultural Journal,* 50: 465–483.

Raunet, M., 1982. Les Bas-Fonds en Afrique et à Madagascar. Formation, Caractères Morphopédologiques, Hydrologie, Aptitudes Agricoles. IRAT, Service de pédologie, Montpellier, France.

Savvides, L., 1981. Guidelines to Improve Valley Swamps (Fadamas). Bida Agric. Dev. Project, Nigeria.

Thenkabail, P.S., and C. Nolte, 1996. Capabilities of Landsat–5 Thematic Mapper ™ data in regional mapping and characterization of inland valley agroecosystems in West Africa. *International Journal of Remote Sensing*, 17(8): 1505–1538.

Thenkabail, P.S. and C. Nolte, 1995a. Mapping and Characterizing Inland Valley Agroecosystems

of West and Central Africa: A Methodology Integrating Remote Sensing, Global Positioning System, and Ground-Truth Data in a Geographic Information Systems Framework. RCMD Monograph No.16, IITA, Ibadan, Nigeria.

Thenkabail, P.S. and C. Nolte, 1995b. Level II Characterization of Inland Valley Agroecosystems in Save, Parakou, Bassila, Bante, and Tchaourou Regions of the Republic of Benin through Integration of Remote Sensing, Geographic Information Systems and Global Positioning Systems. Inland Valley Characterization Report No.1. Resource and Crop Management Division, IITA, Ibadan, Nigeria.

Thenkabail, P.S. and C. Nolte, 1995c. Regional Characterization of Inland Valley Agroecosystems in Gagnoa (Côte 'Ivoire) through Integration of Remote Sensing, Geographic Information Systems and Global Positioning Systems. Inland Valley Characterization Report No.2. Resource and Crop Management Division, IITA, Ibadan, Nigeria.

Thenkabail, P.S. and C. Nolte, 1995d. Regional Characterization of Inland Valley Agroecosystems in Sikasso (Mali) and Bobo-Dioulasso (Burkina Faso) through Integration of Remote Sensing, Geographic Information Systems and global positioning systems. Inland Valley Characterization Report No.3. Resource and Crop Management Division, IITA, Ibadan, Nigeria.

Turner, B., 1977. The Fadama Lands of Central Northern Nigeria: Their Classification, Spatial Variation, Present and Potential Use. Ph.D. Dissertation, University of London, UK.

Turner, B., 1985. The classification and distribution of fadamas in central Northern Nigeria. Zeitschrift für Geomorphologie, Supplementband 52, Berlin-Stuttgart, 87–114.

Windmeijer, P.N., and W. Andriesse, Eds. 1993. Inland Valleys in West Africa. An Agro-Ecological Characterization of Rice-Growing Environments. International Institute for Land Reclamation and Improvement (ILRI) Publ. 52, ILRI, Wageningen, the Netherlands.

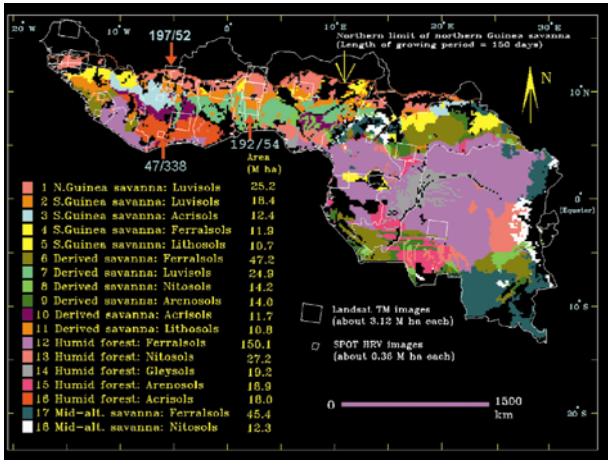


Figure 8.6. Landsat TM and SPOT HRV acquisitions for Level II Characterization of inland valley agroecosystems located in the level I map of agroecological and soil zones of West and Central Africa. Of these images, four agroecological and soil zones covered by Two Landsat TM (path and rows: 192/54; 197/52) and one SPOT HRV image (K and J: 47/338) have been used for illustration in this study.

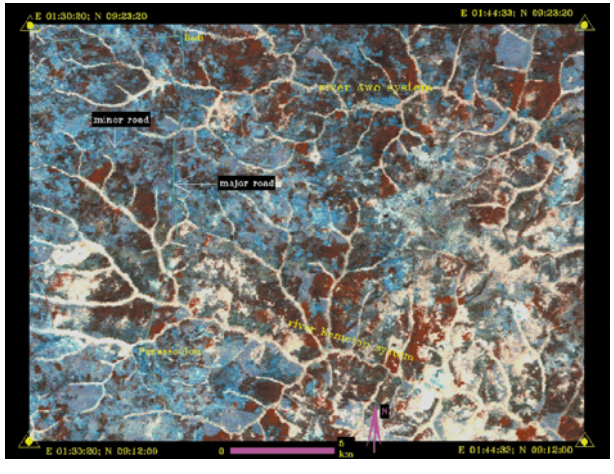


Figure 8.7. Ratio RGB image of Landsat TM bands: TM4/TM7 (red), TM4/TM3 (green), TM4/TM2 (blue) distinguishing the inland valley bottoms from fringes and uplands for a sub-area of Landsat Path:192, Row: 054 in Save, Republic of Benin, IN level I agroecological and soil zone (AEZ) 7.

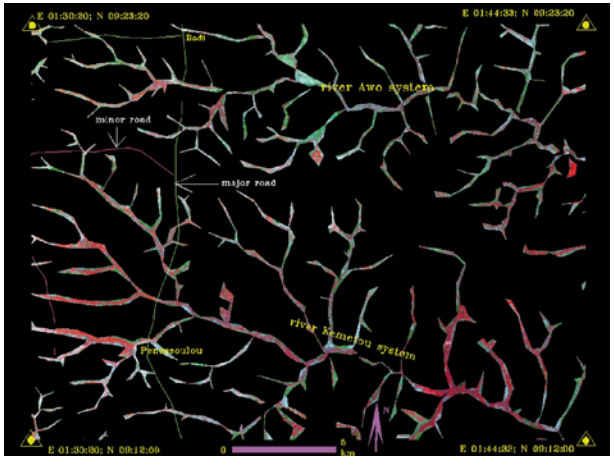


Figure 8.8. Inland valley bottoms delineated (for area shown in Figure 8.7) and displayed as a false color composite of TM bands: TM4 (red), TM3 (green), TM5 (blue). Color key: red (densest vegetation), shades of green and blue (very little and/or dry vegetation). The NDVI values were less than 0.10 for these areas.

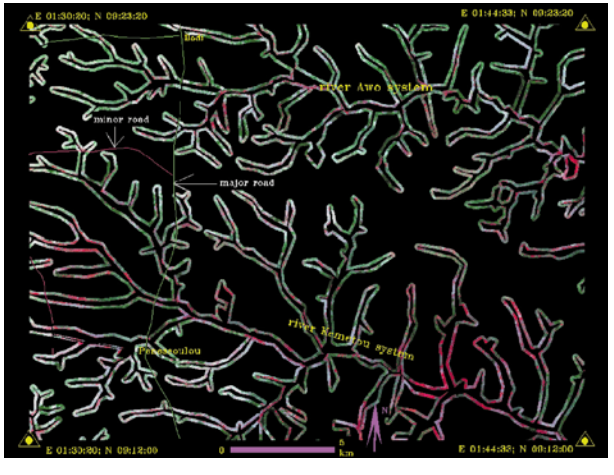


Figure 8.9. Inland valley fringes for the same sub-area as shown in Figures 8.7, and 8.8 displaying the false color image of TM4 (red), TM3 (green), TM5 (blue). Color key: red (densest vegetation), shades of green (lowest density and vigor of vegetation), and shades of blue and gray (intermediate vegetation).

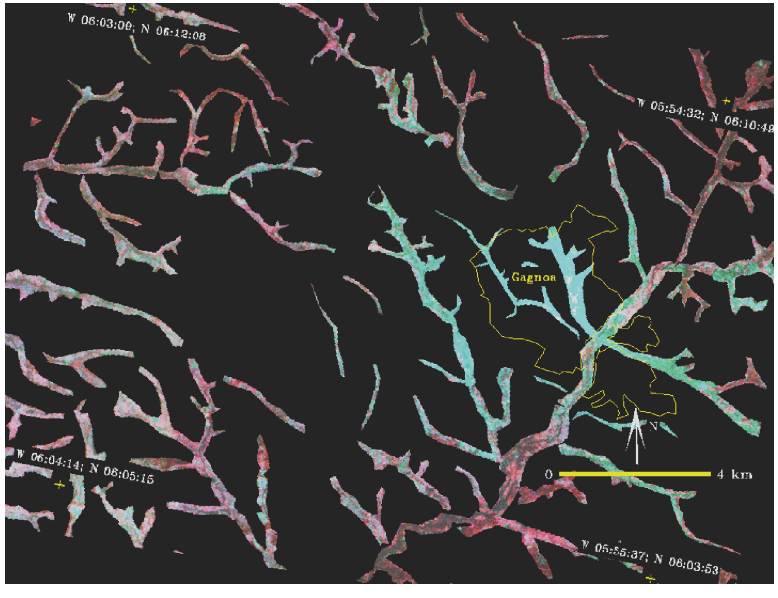


Figure 8.10. Delineated inland valley bottoms around Gagnoa, Cote d'Ivoire, displayed as a false color composite of SPOT HRV bands: band 3 (red), band 2 (green), and band 1 (blue). In AEZ 16.

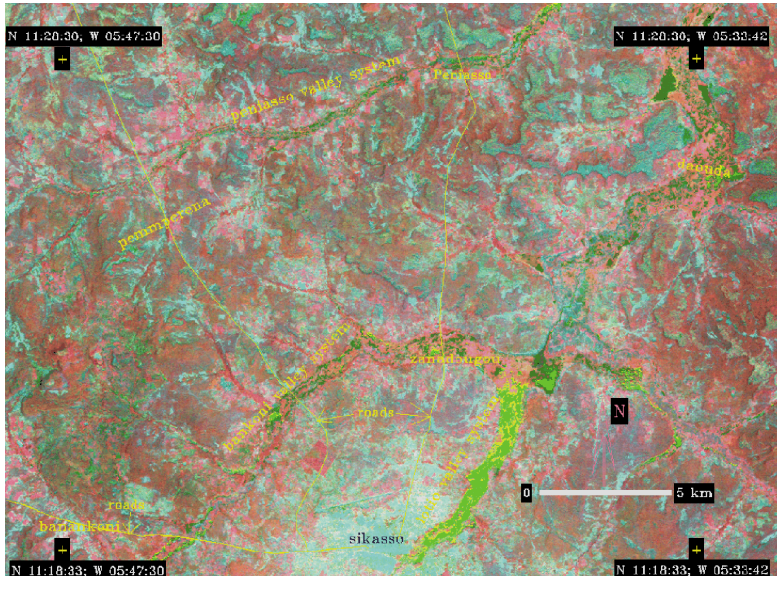
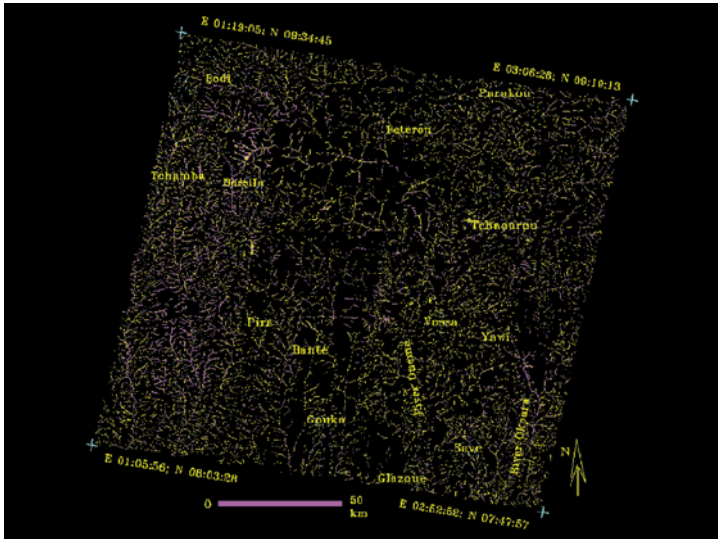
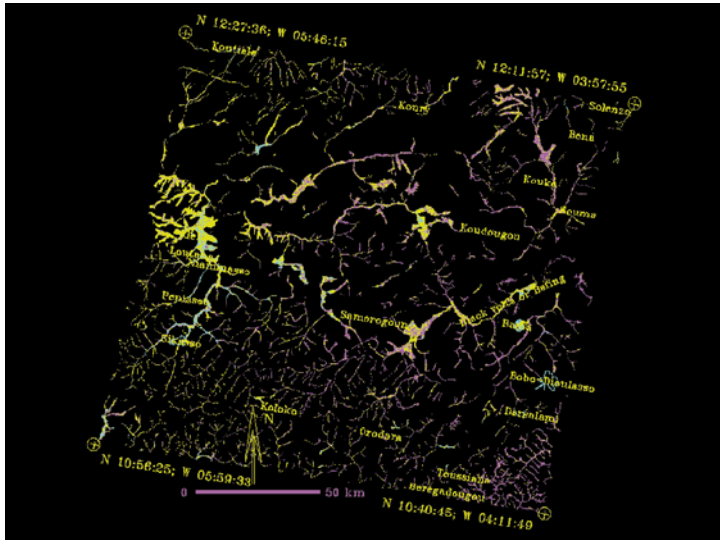


Figure 8.11. Broad inland valley bottoms (often 250 to 800 meter wide) near Sikasso, Mali. These valley bottoms are heavily flooded during rainy seasons and are either used for rice cultivation or have potential for rice and other wetland cropping. Characteristic valley bottoms and their uplands in AEZ 2.





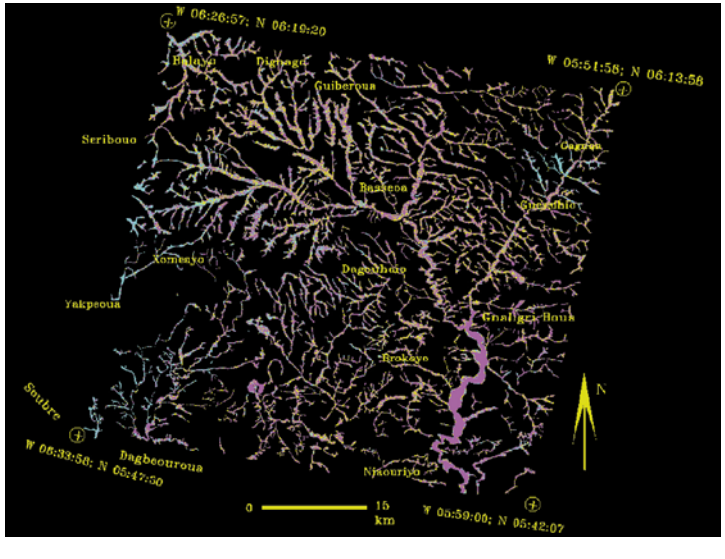
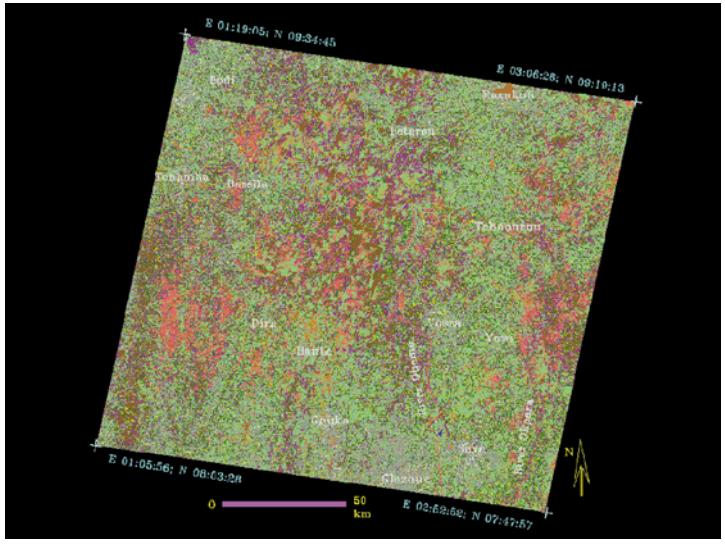
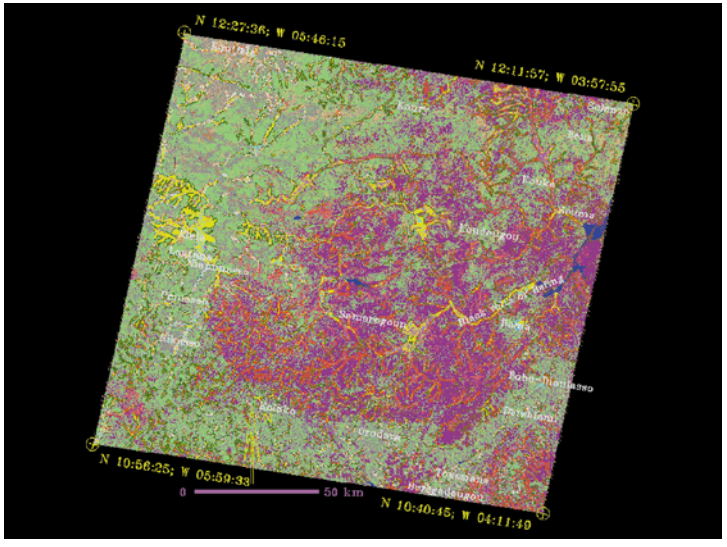


Figure 8.12. (top, left) Inland valley bottoms delineated using the espoused methodology for Landsat path:192 and row 54. There were 9 percent (281,500 ha.) valley bottoms (as shown in this plate) in total area of 3.14 Mha. The valley bottom area as a percent of total area was: 7.9 percent for AEZ 2 and 10.2 percent for AEZ 7. (see Figure 8.6 for spatial distribution of AEZ 2 and AEX 7 within this image; Color legend in Figure 8.18).

Figure 8.13. (left) Inland valley bottoms delineated using the espoused methodology for SPOT HRV K:47 and J:338. There were 18 percent (70,000 ha.) valley bottoms (as shown in this plate) in total area of 0.39 Mha. The valley bottom area as a percent of total area was: 18 percent for AEZ 16. (see Figure 8.6 for spatial distribution of AEZ 16; Color legend in Figure 8.18).

Figure 8.14. (above) Inland valley bottoms delineated using the espoused methodology for Landsat path:197 and row 52. There was 8.57 percent (269,000 ha.) valley bottoms (as shown in this plate) in total area of 3.14 Mha. The valley bottom area as a percent of total area was: 7.7 for AEZ 2 and 9.1 percent for AEZ 1. (see Figure 8.6 for spatial distribution of AEZ 2 and AEZ 7 within this image; Color legend in Figure 8.18).





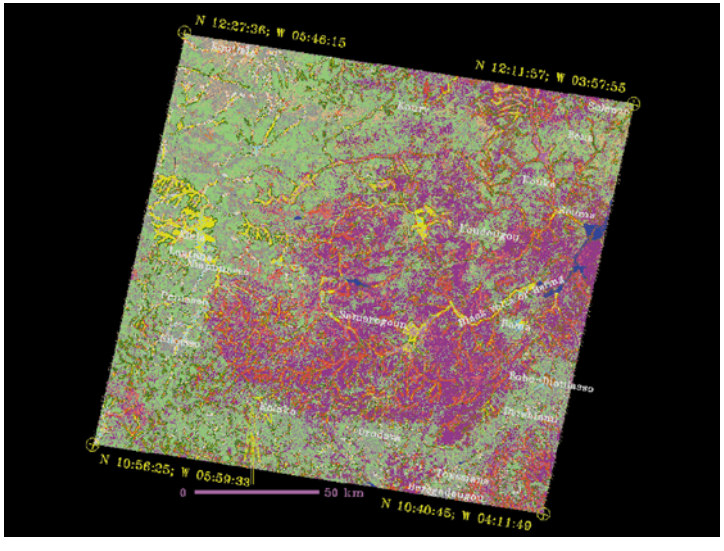


Figure 8.15. (top, left) Land-use classes mapped for the different components of the toposequence (valley bottoms, valley fringes, and uplands) for the entire study area of 3.12 million ha. Covered by Landsat path:192 Row:54 in Save, republic of Benin (see legend in Figure 8.18 for color key).

Figure 8.16. (left) Land-use classes mapped for the different components of the toposequence (valley bottoms, valley fringes, and uplands) for the entire study area of 0.39 million ha. Covered by SPOT HRV K:47 and J:338 in Gagnoa, Cote d'Ivoire (see legend in Figure 8.18 for color key).

Figure 8.17. (above) Land-use classes mapped for the different components of the toposequence (valley bottoms, valley fringes, and uplands) for the entire study area of 3.12 million ha. Covered by Landsat path:197 Row:52 in Sikasso, Mali and Bobo-Dioulasso, Burkina Faso (see legend in Figure 8.18 for color key).

Figure 8.18. Legend of color keys for Figures 8.12–8.17.

Figure 8.20. Location of potential benchmark research sites for technology development research activities in the study area covered by Landsat TM path: 197 and Row:52 (Sikasso, Mali; Bobo-Dioulasso, Burkina Faso). These locations were selected by integrating remote sensing, GIS, ground-truth, and expert knowledge data in a GIS model.

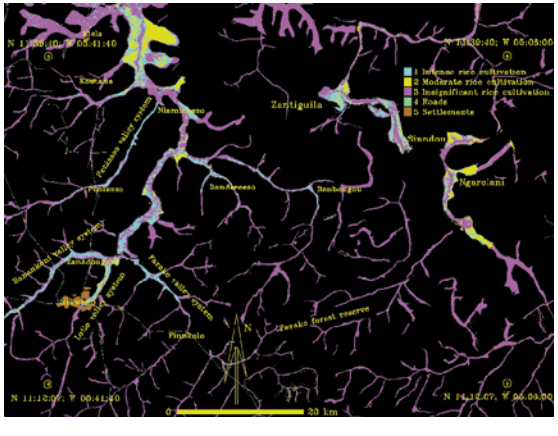


Figure 8.19. Rice cultivation in inland valleys bottoms surrounding Sikasso Mali (AEZ 2). The areas in magenta have either insignificant or no rice cultivation, but have very high potential for exploitation for rice cultivation. The flat and large valley bottoms, which are significantly flooded during rainy season and fertile soils, make these valleys ideal for rice cultivation. Special varieties that can withstand flooding are required.

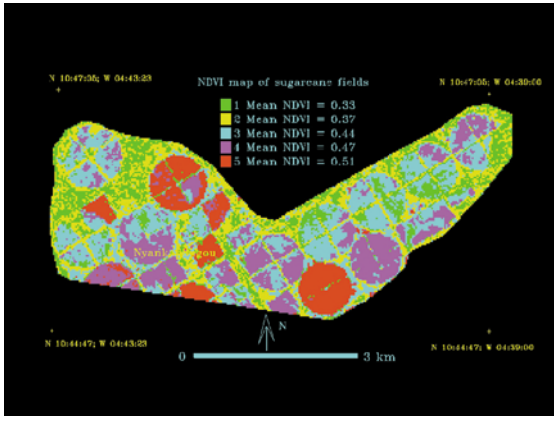


Figure 8.21. Study of biomass levels in the valley bottoms in summer offer a powerful means of assessing potential of valleys for dry season cultivation and in general assessment of moisture levels. The normalized difference vegetation indices (NDVI) has been widely accepted as powerful means of biomass assessment. The above NDVI based biomass levels are for Nyankadougou, South-West of Bobo-Dioulasso, Burkina Faso.

CHAPTER 9

Watershed Characterization by GIS for Low Flow Prediction

Glenn S. Warner, Andrés R. García-Martinó, Frederick N. Scatena, and Daniel L. Civco

INTRODUCTION

A GIS was used to determine watershed characteristics for development of multiple regression equations for prediction of low flow in streams in the eastern part of Puerto Rico. A preliminary selection of 45 gauged watersheds was based on a general Life Zone Map, and vector coverage of precipitation and temperature regions typical of the Caribbean National Forest. The number of watersheds was reduced to 19 by overlaying land use coverage in a GIS and selecting those watersheds having permanent forest cover and available data including digital elevation models (DEMs). A total of 53 initial watershed parameters (grouped into geology, soils, geomorphology, stream network, relief, and climate) were determined by GIS for each of the 19 watersheds. The number of parameters was reduced to 13 through a two-step process using criteria that included correlation among the parameters, correlation with low flows, colinearity, significance of coefficients, and logic. Multiple regression analyses were performed to predict the 7 day–10 yr, 30 day–10 yr, and 7 day–2 yr low flows using combinations of the 13 parameters. Selection of the final regression model for each low flow was based on five statistical tests. Four parameters (drainage density, weighted mean slope, an aspect parameter, and ratio of tributaries to main channel length) were found to be the best predictors, resulting in R^2 from 0.96 to 0.97 for the three low flow variables. A non-GIS model was also developed with easier to measure parameters, but resulted in a higher SE than the GIS-based models.

BACKGROUND

Due to the complexity involved in the generation of low flow in streams, few if any physically based models have been developed for prediction of low flows. Instead, measured low flows are usually regressed against physical parameters that characterize the watersheds involved. Parameters used in low flow studies have varied considerably from one region or study to another, but typically include parameters that reflect basin, climatic, and cultural characteristics (Singh and Stall, 1974) that may play a role in the generation of low flow. Examples of parameters include watershed area, surficial geologic materials, watershed shape, and stream length.

A number of researchers have used GIS for watershed characterization for purposes such as modeling of water quality (Cahill et al., 1993), runoff (Luker et al., 1993), and nonpoint pollution (Robinson and Ragan, 1993). However, GIS application in low flow research is very limited. Low flow studies typically include a large number of watersheds with great differences in drainage areas, topography, geology, and other watershed characteristics. GIS provides not only a format

for more efficient determination of these parameters, but also opportunities for investigation of parameters that are not easily measured by manual means and that therefore have not been used in most low flow studies. The development of automated computer programs for watershed characterization (Martz and Garbrecht, 1993; Eash, 1994) could lead to routine use of GIS in determination of parameter values for multiple regression equations.

The need for low flow prediction and the availability of stream gauge data and GIS coverages for eastern Puerto Rico provides an opportunity to study low flow in streams draining the Caribbean National Forest, also known as the Luquillo Experimental Forest (LEF). The LEF is located in the eastern end of Puerto Rico and covers an area of 11,300 ha. Elevations range from 73 m to 1,075 m above sea level. The LEF is almost completely forested and receives an average of 3864 mm of rainfall annually (García-Martinó, 1996). Water demand by municipalities in the area is putting increased pressure for water withdrawal from streams draining the mountains. The lack of information on low flow for ungauged streams provides a need for better methods of predicting low flows in these streams.

DATA AND METHODS

The GIS database included land use, soils, geology, hydrography (all in vector format), and Digital Elevation Models (DEM). All GIS coverages were obtained as $7.5' \times 7.5'$ quadrangles and georeferenced to the Universal Transverse Mercator (UTM) coordinate system. Soil and geology coverages were digitized from 1:20,000 scale maps from the USDA Soil Surveys (published in the 1970s) and the U.S. Geological Survey (USGS), respectively. Unconsolidated material formations as small as 7,000 m^2 are depicted in the geology coverages. The land use data (produced by the Puerto Rico Dept. of Transportation for 1977) were divided using first degree units based on Anderson et al. (1976) and highly subdivided to specific categories. The stream network was obtained from the USGS as Digital Line Graphs (DLG), showing rivers and major tributaries. The stream network shown in the DLGs was virtually the same as that shown on USGS topographic maps.

We obtained DEMs from the USGS in two different formats: 30 m by 30 m DEMs (USGS, 1987) for the eastern region of the island, specifically for El Yunque, Fajardo, Humacao, Naguabo, Juncos, and Fajardo quadrangles; and a DEM produced by the Defense Mapping Agency (DMA) for the rest of the island. The DEM from the DMA has a resolution of 3×3 arc seconds, which in Puerto Rico is equivalent to an approximate 90 m x 90 m resolution. All DEMs are in UTM coordinates with elevations in meters.

A typical low flow study involves watershed selection, frequency analysis, watersheds characterization, and multiple regression analysis. Since we were interested in the humid montane regions of Puerto Rico, we limited our investigations to watersheds either within the LEF or with rainfall and land use characteristics similar to those of the LEF. The frequency analysis for low flow of the gauged streams and the regression analysis followed standard procedures. The watershed characterization, however, was not limited to commonly used parameters that could be measured by hand, but initially included a large number of potentially important factors that either alone or in combination with other factors might be significant hydrologically. The number and complexity of the parameters considered were only feasible when using GIS, since the effort to evaluate each parameter for each watershed by hand would have been overwhelming.

To select as many representative watersheds as possible, we followed a scheme that first digitized the Life Zones Map for Puerto Rico produced by Ewel and Whitmore (1973) using the Holdridge classification system. We next performed a broad, general selection of USGS stream gauges by a visual inspection of the Life Zones Map and major drainage networks of the island.

We looked first for stations draining the LEF, and second, for stations with temperature and precipitation characteristics similar to those in the LEF. The first GIS analysis consisted of overlaying the watersheds of 45 stations (continuous and partial) with Life Zone and land use coverages. The watershed boundary for each of the 45 stations was drawn on USGS topographic maps with a 1:20,000 scale, and digitized.

We reduced the number of stations to 19 (in three different regions of the island) based on land use and life zone classification, followed by GIS application of the geology, soils, stream network and DEM coverages. We initially chose a large number of parameters based on their potential to influence storm runoff and groundwater dynamics. Many parameters such as soil water-holding capacity, liquid limit, texture, and depth to bedrock did not vary significantly across watersheds and were dropped. A total of 53 parameters (Table 9.1) covering six groups were determined by GIS to vary across the 19 watersheds and were chosen for further analysis. We found that GIS has several advantages over traditional methods: (1) a large number of parameters can be evaluated for numerous watersheds, (2) parameters not easily determined by hand such as percent of area having slopes or aspects less than various thresholds can be calculated, and (3) area-weighted parameters such as the weighted mean elevation within the watershed can be determined easily.

The data for most of the partial stations consist of six or seven low flow measurements taken over a four year period. Some low flow values were obtained from Santiago (1992). Measured low flows for other partial and three continuous stations with short records were correlated with concurrent discharges from nearby watersheds with similar hydrologic characteristics as recommended by Riggs (1972). Daily low flow measurements were randomly selected for each of these stations and correlated with concurrent low flows from nearby continuous stations. All correlations were significant ($\alpha = 0.05$).

We developed two types of regression models: GIS-based models (for 7d–10yr, 30d–10yr, and 7d–2yr low flows) that considered all 13 parameters and a non-GIS model (7d–10yr only) that only considered parameters that could be easily measured without a GIS. A model can be statistically strong, but if the parameters are difficult to measure, then the model may be useless. Although GIS technology is dramatically closing the gap between easy and hard to measure parameters, GIS is not always available. We selected the best models of each type using a number of statistics given by Helsel and Hirsch (1992). Parameters considered in each model and details of the statistical tests are given in García-Martinó (1996).

RESULTS AND DISCUSSION

Four parameters consistently produced the best overall results for the GIS-based models: DD is total length of perennial channels per unit area (m/m^2), SLP is weighted mean slope in the watershed (°), PA90 is percent of total area of watershed with aspects between 0–90°, and CHATRI is ratio of total tributary length to length of main channel. Although watershed area does not appear directly in the equations, it is highly correlated with other parameters. The GIS models (one for each low flow) are:

$$\text{Log7d–10yr} = 1.78 - 1331.61(\text{DD}) - 0.0316(\text{SLP}) + 0.0146(\text{PA90}) + 0.3126(\text{CHATRI}) \qquad R_a^2 = 0.97 \tag{1}$$

$$\text{Log30d–10yr} = 1.60 - 1098.26(\text{DD}) - 0.0301(\text{SLP}) + 0.0160(\text{PA90}) + 0.2957(\text{CHATRI}) \qquad R_a^2 = 0.96 \tag{2}$$

Table 9.1. Definitions of Watershed Parameters Used in Correlation and Regression Analyses

CLIMATE BASED PARAMETERS:

MMYR = mean annual rainfall (mm/yr)
10YRCDNR = number of consecutive days with no rain with recurrence interval of 10 years
M#DNR = average number of days per year with no rain
MMCDNR = maximum consecutive days with no rain

GEOLOGY BASED PARAMETERS:

UNCON and QA = area and % of area of watershed classified as unconsolidated material, respectively
LSEU = length of perennial streams in direct contact with unconsolidated material
INTRU = % of area classified as intrusive igneous rock with low water bearing capacities
STRAT = % of area classified as alluvial deposits and sedimentary formations (sandstone)

SOIL BASED PARAMETERS:

PLASTI = area-weighted mean plasticity index for watershed
PERM = area-weighted mean permeability

STREAM NETWORK BASED PARAMETERS:

PERC = total length of perennial channels (m)
CHAN = total length of main channel (m)
TRIB = total length of tributaries to main channel (m)
DD = total length of perennial channels per unit area (m/m^2)
SD = number of perennial channels per unit area ($\#/m^2$)
TRIPER = ratio of total length of tributaries to total length of perennial channels
CHATRI = ratio of total length of tributaries to total length of main channel
CHANSLP = slope of main channel between 15% and 85% points of main channel (°)
CHSLP = (CHANSLP/CHAN)/100
SINUO = total length of main channel divided by length of watershed

WATERSHED MORPHOLOGY BASED PARAMETERS:

AREA = watershed area (ha)
SLP = area-weighted slope of watershed (°)
PERIM = length of perimeter of watershed boundary (m)
COMP = ratio of PERIM to circumference of a circle with area = AREA
FORM = AREA/(watershed $length^2$)
ELONG = diameter of circle with area = AREA divided by watershed length
AREA4 and A4, AREA7 and A7, AREA10 and 10 = areas and % of AREA with slopes < 4, 7 and 10°
MSLP100, ASLP100, MSLP200, ASLP200 = maximum and mean slope within 100 or 200 m buffers of perennial streams (°)
ASPECT = area-weighted aspect (°)
AREA90 = total area of watershed with aspect between 0 and 90(°)
PA90 = % of AREA with aspects between 0 and 90°
ASP100 and ASP200 = mean area-weighted aspect inside 100 and 200 m buffers along perennial streams

WATERSHED RELIEF BASED PARAMETERS:

EL = area-weighted mean elevation
ELRANGE = maximum minus minimum elevation in watershed (m)
RELIEF = ELRANGE divided by length of watershed
RERA = RELIEF*ELRANGE
SLORAT = CHANSLP/SLP
HYP1 = % change between 0.25 and 0.75 points of hypsometric curve of watershed
HYP2 = HYP1/AREA
HYP3 = HYP1/(% change between 0.50 and 0.75 points of hypsometric curve)

$$\text{Log7d–2yr} = 1.70 - 1157.49(\text{DD}) - 0.0248(\text{SLP}) + 0.0121(\text{PA90}) + 0.3017(\text{CHATRI})$$
$$R_a^2 = 0.97 \quad (3)$$

where R_a^2 is the adjusted R^2 (Velleman, 1992).The best non-GIS model consisted of four parameters that could be relatively easily measured by manual methods for prediction of the 7d–10yr flows:

$$\text{Log7d–10yr} = 0.1827 - 0.0824(\text{CHANSLP}) + 2.24(\text{FORM}) + 0.0012(\text{ELRANGE}) - 882.51(\text{DD})$$
$$R_a^2 = 0.94 \quad (4)$$

where: CHANSLP is slope between 15% and 85% points of main channel, FORM is area divided by square of watershed length, ELRANGE is maximum minus minimum elevation in the watershed, and DD is same as for GIS-based models above.

The error in a model is the product of chance variation and variation from unaccounted differences in watershed characteristics. The objective is to decrease the variation due to the latter. The variation in Y due to each variable (R^2) is very similar in each GIS model (Table 9.2). The strongest variable is DD, contributing over 53% of the variation in Y in each model. The strongest parameter in the non-GIS model was also DD (Table 9.2), which evidently explains a high degree of the subsurface and groundwater dynamics.

The best method of determining the accuracy of the non-GIS model compared to the GIS model is the standard error of the estimate (SE). The SEs varied from 11.5% to 13.8% for the GIS models compared to 19.5% for the non-GIS model (Table 9.2). Values for SE over 100% are common (Riggs, 1973). One of the most accurate low flow models was developed for West Virginia by Chang and Boyer (1977) and had an SE of 31%. Since SE tends to be underestimated for small samples, an adjusted SE (SE_a) as suggested by Ezekiel and Fox (1959) was calculated by the following equation:

$$SE_a = [N/(N - M) * SS_e]^{0.5} \quad (5)$$

where N is the sample size, M is the number of coefficients in the model including the intercept, and SS_e is the sum of squares of the error. The SE_a's for the GIS models shown in Table 9.2 are good, even when compared with the nonadjusted SE more commonly found in low flow research. Although the variation in Y covered by the non-GIS model is relatively high ($R^2 = 0.94$), the SE_a of 82.8% is well above the SE_a of the GIS models. It is still a good model when compared to most low flow research in the U.S. A direct comparison of the GIS model vs. non-GIS model can be made for the 7d-10yr which shows an increase in SE_a from 58.7% (GIS) to 82.8% (non-GIS) that can be attributed to the ability to include additional and more complex parameters by GIS.

The future availability of greater resolution DEMs for areas outside the LEF will significantly improve the analysis. Greater resolutions will permit more detailed and precise watershed characterization, especially when dealing with slopes and aspect. Also, it will be possible to extend the stream network to streams not included in a USGS topographic map. GIS coverages are available for the whole island. These or similar models can be developed and/or adapted to the unique conditions for each region of the island in order to characterize the watersheds and develop better water management plans. On the other hand, the accuracy of the selected models is relatively high, and its application to humid montane regions in Puerto Rico and adjacent areas should provide opportunities for more intensive analysis and greater confidence for low flow predictions.

Table 9.2. Parameters and Comparative Statistics for Selected GIS and Non-GIS Models

GIS Models	Model C-3 (7 day–10 yr)			Model D-4 (30 day–10 yr)			Model E-3 (7 day–2 yr)		
Parameters	R_a^2	SE[1]	SE_a[2]	R_a^2	SE	SE_a	R_a^2	SE	SE_a
DD	0.59	45.68	193.81	0.53	44.67	189.51	0.54	43.19	183.25
DD + CHATRI	0.91	21.84	92.68	0.88	23.41	99.32	0.92	18.11	76.85
DD + CHATRI + PA90	0.96	15.71	66.65	0.95	15.59	66.15	0.96	12.83	54.43
DD + CHATRI + PA90 + SLP	0.97	13.84	58.72	0.96	13.94	59.15	0.97	11.47	48.64

Non-GIS Models	Model P-1 (7 day-10 yr)		
Parameters	R_a^2	SE[1]	SE_a
DD	0.59	45.68	193.81
DD + ELRANGE	0.74	37.55	159.33
DD + ELRANGE + FORM	0.88	26.84	113.88
DD + ELRANGE + FORM + CHANSLP	0.94	19.52	82.81

[1]Standard error.
[2]Adjusted standard error.

SUMMARY AND CONCLUSIONS

The application of GIS for watershed characterization for low flow prediction provides the opportunity to evaluate large numbers of parameters and to evaluate parameters which are difficult to determine by hand methods. Unique and highly qualitative parameters could be readily measured using GIS. The GIS-based models produced better estimates of low flows than use of a non-GIS model developed for gauged watersheds in humid montane regions of Puerto Rico. The adjusted standard error of the estimate (SE_a) of the GIS-based model for the 7d–10yr flow was 58.7% compared to 82.8% for the non-GIS model. These SE_a's are less than the error of most low flow research performed in the U.S. The most significant parameters in the GIS model for each of the three low flows were drainage density (DD), the ratio of the length of the tributaries to the length of the main channel (CHATRI), the percent facing northeast (PA90), and the mean slope in the watershed (SLP).

ACKNOWLEDGMENT

Funding for this study was provided by the USDA Forest Service through the Luquillo Experimental Forest, Rio Piedras, Puerto Rico. Published as Scientific Contribution No. 1700 of the Storrs Agricultural Experiment Station.

REFERENCES

Anderson, J.A., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. A Land Use and Land Cover Classification System for Use with Remote Sensing Data. U.S. Geol. Survey Prof. Paper 954.

Cahill, T.H., J. McGuire, and C. Smith, 1993. Hydrologic and water quality modeling with geographic information systems. In *Geographic Information Systems and Water Resources*, J.M. Harlin and K.J. Lanfear, Ed., American Water Resources Association, Mobile, AL, pp. 313–317.

Chang, M., and D.G. Boyer, 1977. Estimates of low flows using watershed and climatic parameters. *Water Resources Research,* 13(6): 997–1001.

Eash, D.A., 1994. A geographic information system procedure to quantify drainage-basin characteristics. *Water Resources Bulletin,* 30(1):1–17.

Ewel, J.J., and J.L. Whitmore, 1973. The Ecological Life Zones of Puerto Rico and the U.S. Virgin Islands. USDA Forest Services. Resource Paper. ITF–18. Institute. of Tropical Forestry, Rio Piedras, PR.

Ezekiel, M., and K.A. Fox, 1959. *Methods of Correlation and Regression Analysis*. John Wiley and Sons, New York, NY.

García-Martinó, A.R., 1996. "Use of GIS for Low Flow Prediction in Humid Montane Regions in Eastern Puerto Rico." Unpublished M.S. Thesis. University of Connecticut, Storrs.

HEC, 1992. HEC-FFA: Flood Frequency Analysis User's Guide. U.S. Corps of Engineers, CPD–13.

Helsel D.R., and R.M. Hirsch, 1992. *Statistical Methods in Water Resources.* Elsevier Science Publishing, Co., New York, NY.

Luker, S., S.A. Samson, and W.W. Schroeder, 1993. Development of a GIS based hydrologic model for predicting direct runoff volumes. In *Geographic Information Systems and Water Resources*, J.M. Harlin and K.J. Lanfear, Eds., American Water Resources Association, Mobile, AL, pp. 303–312.

Martz, L.W., and J. Garbrecht, 1993. Automated extraction of drainage network and watershed data from digital elevation models. *Water Resources Bulletin,* 29(6):901–916.

Riggs, H.C., 1973. Regional Analyses of Streamflow Characteristics. U.S. Geological Survey Techniques of Water-Resources Investigations Bk. 4, Chap. B3.

Riggs, H.C., 1972. Low Flow Investigations. U.S. Geological Survey Techniques of Water-Resources Investigations Bk. 4, Chap. B1.

Riggs, H.C., 1967. Some Statistical Tools in Hydrology. U.S. Geological Survey Techniques of Water-Resources Investigations Bk. 4, Chap. A1.

Robinson, K.J., and R.M. Ragan, 1993. Geographic information system based nonpoint pollution modeling. In *Geographic Information Systems and Water Resources*, J.M. Harlin and K.J. Lanfear, Eds., American Water Resources Association., Mobile, AL, pp. 53–60.

Santiago-Rivera, L., 1992. Low-Flow Characteristics at Selected Sites on Streams in Eastern Puerto Rico. U.S. Geological Survey Water-Resources Investigations Report 92–4063.

Singh, K.P., and J.B. Stall, 1974. Hydrology of 7-day 10-year low flows. *Journal of the Hydraulics Division, ASCE*, 100:1753–1771.

U.S. Geological Survey, 1987. Digital Elevation Models Data Users Guide. National Mapping Program Tech. Intruc. Data Users Guide 5.

Velleman, P.F. 1992. Data Desk^{4} ® Handbook and Statistics Guide. Data Description, Inc., Ithaca, NY.

Wadsworth, F.H. 1951. Forest management in the Luquillo Mountains. I. The setting. *Caribbean Forestry,* 12:93–114.

Water Resources Council, 1977. Guidelines for Determining Flood Flow Frequency. Bulletin #17A, Hydrology Committee.

CHAPTER 10

Evaluation of the Albemarle-Pamlico Estuarine Study Area Utilizing Population and Land Use Information

Robert E. Holman

BACKGROUND

The Albemarle-Pamlico Estuarine Study (APES) has been funding many information acquisition projects over the last five years in the areas of resource critical areas, water quality, fisheries, and human environment (Steel and Scully, 1991). Most of these projects have transferred their data over to the APES's Geographic Information System (GIS) which was created through a subcontract with the North Carolina Center for Geographic Information System (CGIA). GIS has the ability to bring together (enter, display, edit, and manipulate) data based information with digital mapping (locational attributes). At the time of this study, the Center (CGIA) had or was creating all of the needed databases. CGIA was able to combine the data layers in various ways to analyze the relationship among different layers in a visual as well as a statistical manner.

The study area encompasses approximately 23,250 square miles and includes all or portions of 37 counties in eastern North Carolina and 19 counties in Southeastern Virginia. There are six counties along the coastline, 9 counties along the sounds, and 41 cities/counties that lie in the upper drainage basin (Figure 10.1). This study also incorporates all or portions of 6 major river basins including the Chowan, Pasquotank, Lower Roanoke, Tar-Pamlico, Neuse, and White Oak (Figure 10.2). Each basin is divided into subbasins as follows: Chowan, 13; Pasquotank, 8; Lower Roanoke, 3; Tar-Pamlico, 8; Neuse, 14; and White Oak, 5.

METHOD

The analytical method was broken into three phases. Phase One was the creation of county land use maps from the existing Landsat classification scheme. These map products were sent to U.S. Fish and Wildlife and county officials to determine the accuracy of the defined land use classes. The land use maps were also used by the author during flights over the coastal and metropolitan areas to further clarify classification errors. Phase Two was to correct some of the errors in the existing classification. This was carried out by digitizing the corrections to the map products that were returned from Fish and Wildlife and county officials. The map information was also supplemented with other sources of information such as U.S. Fish and Wildlife Service—National Wetland Inventory, U.S. Forest Service—Forest Inventory and Analysis, U.S. Bureau of Census—Census of Agriculture, and U.S. Soil Conservation Service—National Resources Inventory and Hydric Soils in North Carolina counties. Phase Three was identifying correlation between

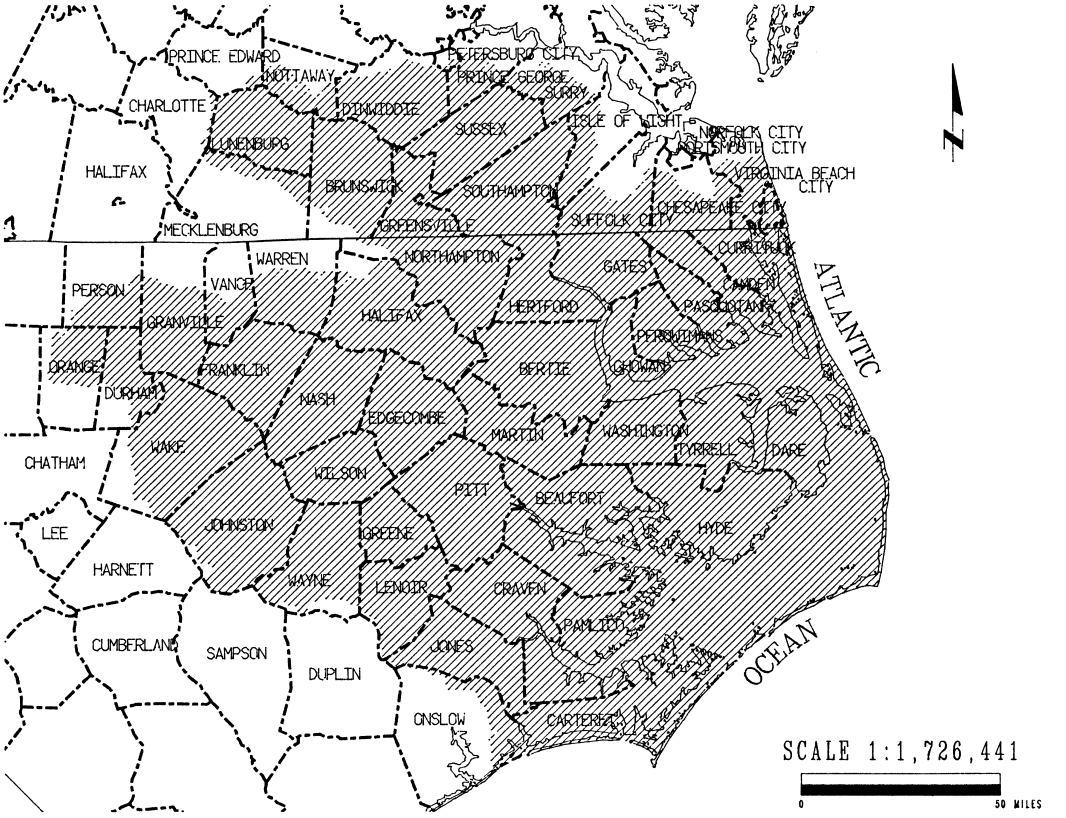


Figure 10.1. Map of study area.

different data sets such as county census population and county acreage of developed land. If a strong correlation was found, then a simple linear regression model was applied in order to predict the relationship between the two parameters. These models were used to correct some of the errors in specific land use categories.

Map Development

There were three tasks in the development of land use and population estimates for the entire APES area: (1) defining the actual drainage area; (2) having all the basin and subbasin boundaries digitized in order to determine the land use and population; and (3) correcting for errors associated with the different land uses.

First, the study area was defined as the entire drainage area of the Albemarle and Pamlico Sound system including Core and Bogue Sounds. The upper Roanoke Basin and a portion of the White Oak Basin were not included because: (1) the upper Roanoke River Basin covers approximately 8,370 square miles in Virginia/North Carolina and stretches over two-thirds the length of North Carolina, and would add one-third more area to the study area; and (2) a decision was made early in APES to have Carteret County as the furthest area south. However, due to the watershed approach in defining the study area in this project, all the subbasins in the White Oak Basin were included except the furthest one southwest that starts at Camp Lejeune. There was no compatible land use data available for this subbasin.

Figure 10.2. The six APES basins and their subbasins.

Second, all North Carolina basins and subbasins were digitized by the Research Triangle Institute and compared closely with the U.S. Geological Survey subbasins for North Carolina. Virginia subbasin information was supplied by Information Support Systems Laboratory within Virginia Polytechnic Institute and State University and was based on Soil Conservation Service (SCS) information. Due to the large number of subbasins identified by SCS in the Virginia portion of the Chowan and Pasquotank Basins, subbasins were combined to create areas of the same size range as subbasins identified in North Carolina. All subbasins were digitized from U.S. Geological Survey's 1:24,000 scale topographic maps. Specific subbasins were identified by a six number code that was broken into two-digit sets. The first two digits identified the regional basin; the second two digits identified the basin; and the third two digits identified the subbasin (Figure 10.2). Codes used in this report were the same ones adopted by the North Carolina Division of Environmental Management.

The third task was to identify the accuracy of the land use data and to develop methods to correct for the errors. Khorram and others (1992) found that with the Landsat data, the urban or built-up land use category was only 46% accurate, and the accuracy of forested wetlands was unknown. In addition, the classification of mixed pixels in the existing land use data set had to be resolved. Mixed pixels are defined as areas that could not be classified because the resolution or pixels were a mixture of many categories. The land use classification from 1987–1988 developed by Khorram will be referred to as the "Landsat" classification in this study.

Accuracy and Errors

The first task was to define which level of land use to utilize. Landsat land use classification defined 18 separate classes that can be generally broken into similar U.S. Geological Survey level I and level II groupings (Anderson and others, 1976). The land use classification was based on Landsat data which Khorram interpreted mostly as land cover (the actual extent of vegetative and other cover) and some land use (interpretation of activities taking place on the land). Interpretation of land use is much more subjective than land cover and is dependent on the knowledge of the individual interpreter.

A level II map with 18 individual classes was provided to officials of two Fish and Wildlife Refuges within the APES area for their evaluation as to land cover accuracy. The Great Dismal Swamp National Wildlife Refuge staff reviewed the Landsat land use map of the refuge. This refuge, located on the border between North Carolina and Virginia just south of Portsmouth, Virginia covers approximately 110,000 acres and is predominantly forested wetland. The staff felt there was good separation among development, agriculture, water, and forest; however, the different forest cover types had serious reliability problems. A major problem was the misclassification of wetter deciduous stands like cypress/gum and maple/gum as pine/hardwood forest. A second land cover map was sent to Mattamuskeet and Swan Quarter National Wildlife Refuges personnel for their review. These two refuges are located entirely in Hyde County, North Carolina, and Swan Quarter is adjacent to the Pamlico Sound. These refuges together cover approximately 65,700 acres and are predominantly water, wetland, and forest. The staff found quite a few areas that were referred to as mixed pixels that were actually open water and irregularly flooded brackish marshes. White Cedar stands were actually marsh impoundment areas around Lake Mattamuskeet, and pine forest was actually mixed pine/hardwood or hardwood/cypress/pine forest. In general, both refuges indicated accuracy problems with the different forest and the mixed pixel classifications.

After indicating they could not evaluate all the classifications in level II land cover maps, officials of counties in the Currituck Sound Basin south of Virginia Beach and adjacent to the Atlantic Ocean were sent, for comment, land cover maps which included the following attributes: USGS level I with 6 categories shown in color; U.S. Census TIGER files that displayed the road network, map scale of 1:100,000; and modified LUDA land use data for the urban or built-up category.

Land Use Data Analysis (LUDA) was an early GIS effort started by U.S. Geological Survey in 1975 to define the land use for the entire United States. All the photographs were manually photointerpreted. The map series consisted of 1:250,000 scale maps of North Carolina defining 37 uses based on the level II classification system. Source imagery was 1:56,000 color infrared photography and 1:80,000 black and white photography dating back to 1970. Resolution was 10 acres for the urban or built-up categories and 40 acres for the remaining classifications (Kleckner, 1981). For comparison, the Khorram classification was based on 1987–1988 Landsat satellite imagery that was semiautomatically interpreted. The county map series was at a scale of 1:100,000 with a final resolution of 1 acre.

In 1991 and 1992, the author flew along the Outer Banks and inland around the estuarine portion of the study area and over portions of Wake, Durham, and Orange Counties to verify the problems with the categories of urban or built-up, wetland, and mixed pixels. The urban or built-up class was underestimated on the 1987–1988 land use maps mainly due to forest crown cover that obscured the true land use on the ground. High spectral reflectance of bare agricultural fields was also a problem because these fields were being classified as developed areas. The problem with fields being identified as developed areas was especially evident in the Landsat scene furthest west that included the Raleigh metropolitan area. This category was a very small percentage (3.3 to

7.9%) of the overall land use of the study area but is critical in that urban or built-up land use can cause the greatest impact on natural resources of the APES area.

Problems associated with the wetland class were found to be interference from forest crown cover. Open marsh and pocosin wetlands were usually accurately defined by the 1987–1988 land use maps but closed forest canopy prevented standing water below the forest to be seen. Therefore these true wetland types were usually defined as forest.

The category of mixed pixels is a grouping the classification scheme could not identify. Flights over the coastal and metropolitan areas verified that in most cases they were a mixture of standing water and wetland vegetation. The only exception to this observation was in Pasquotank County where poorly drained agricultural land was defined as mixed pixels or wetland on the county land use map.

These land use classification problems and others were identified in a workshop the author attended to verify remotely sensed land cover data for the Coastwatch Change Analysis Program of the National Oceanic and Atmospheric Administration (Burgess and others, 1992). The problems fell into four categories: classification error, cover versus land use, categorical resolution, and change detection. Classification errors included "salt and pepper" effect of individual pixels, shadows and bare ground as urban areas, and problems with the degree of wetness during image acquisition. Cover versus land use had the inherent problem with distinguishing land uses and required ancillary data. Categorical resolution was related to the spatial resolution and improper classification. The change detection problem involved the ability to detect a change but not always the nature of the change. From the author's own observations and the results of this workshop, methods were developed to overcome some of the problems associated with remotely sensed land cover data.

Land use information used in this study was analyzed according to the Khorram and others (1992) classification system but condensed from 18 to 7 categories. Certain corrections were incorporated into some classes depending on the observed and documented error associated with each class. The LUDA data set was used to determine "developed land" because the information appeared to be closer to the actual extent and location than the original 1987–1988 Landsat data set. Corrected Landsat built-up areas on the maps returned by county officials were found to have a high degree of correlation (R^2 of 0.9) with the LUDA "developed" category. A linear regression model was used to predict built-up land from the LUDA data. U.S. Fish and Wildlife's National Wetland Inventory (NWI) data were used as a reliable source of wetland acreage for the coastal plains of North Carolina (Wilen, 1990 and Burgess and others, 1992). Wetland acres for twelve of the coastal counties was provided by Kevin Morehead of the Savannah River Ecology Laboratory. The same procedure used to correct built-up areas was used to reconcile the Landsat wetland category with the NWI acreage. A high correlation resulted with a R-squared of 0.9 and a simple linear regression model was used to predict wetland acres from the NWI values. The mixed pixel unidentified category was determined by the U.S. Fish and Wildlife personnel and two overflights of the APES area to be predominantly wetland in nature. The mixed pixel figures were incorporated into the wetland classification.

RESULTS

Land Use

The entire APES area land use classification is based on a modified U.S. Geological Survey's level I classification scheme. One fact to keep in mind is that the water class is not a true *land* use but is a very important classification. There were seven classes with the following percentages: "urban" 4.8%, "agriculture" 28.2%, "forest" 28.4%, "water" 14.6%, "wetland" 20.5%, "shrub

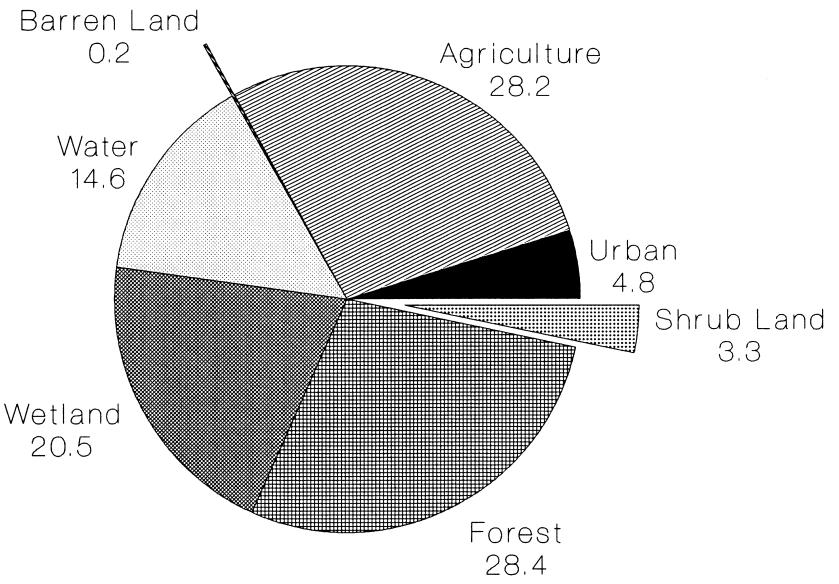


Figure 10.3. APES 1990 land use/land cover.

land" 3.3%, and "barren land" 0.2% (Figure 10.3). In general, the study area is rural in nature with less than 5% of the total area developed. More than 55% of the total APES acreage came from the categories of agriculture and forest.

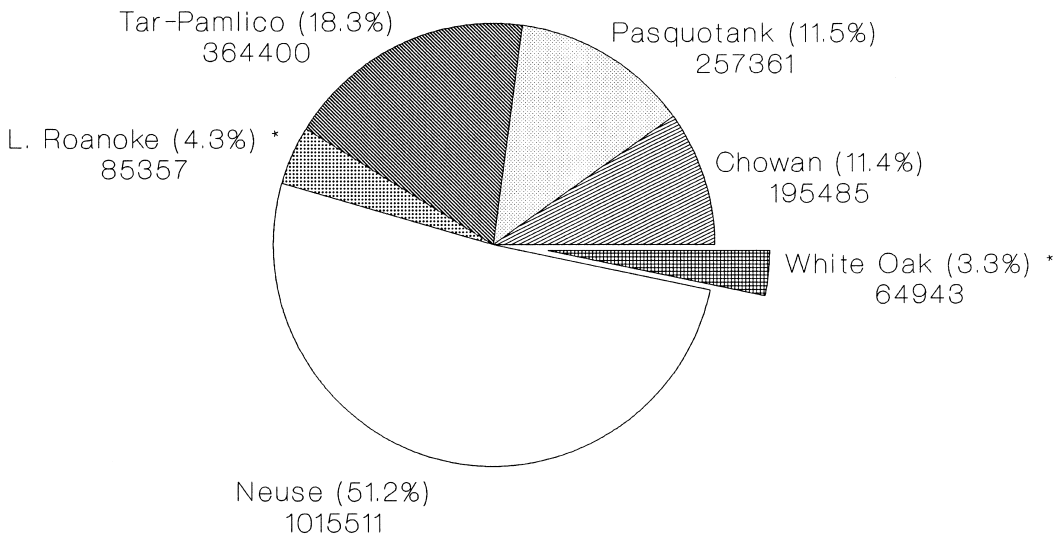
Population

The population of the study area was almost 2 million people (Figure 10.4). About 51.2% reside in the Neuse Basin, which occupies only 26.8% of the land area. Population density for the basins ranged from 163.1 persons/square mile in the Neuse to 39.9 persons/square mile in the Chowan. Since the population density of all but two basins fell below the U.S average of 69 persons/square mile, most of the study area can be characterized as nonmetropolitan (U.S. Bureau of Census 1960, 1970, 1980, 1990).

Population Versus Developed Land

When the subbasins with the highest number of persons/square mile are compared to the subbasins that have the greatest amount of developed land there appears to be a great deal of agreement. If a strong correlation exists between these two parameters then a powerful planning tool can be created to predict the amount of developed land from the existing or projected population for a certain area.

The APES area has only two comprehensive land use databases and one does not correlate well with the category of urban or built-up land. Therefore, another source of long-term land use



*** Only a portion of this basin**

Figure 10.4. APES basin populations in 1990.

data is needed to determine if the relation between population and developed land is statistically sound. Land use data from the State of Maryland has been gathered since 1973 and has been taken as frequently as every five years during the past 15 year period (Maryland State Planning Office, 1991). The acres of total developed land were compared to the closest population census data for each county in Maryland. Three periods (1973, 1981, and 1990) had correlation with a R^2 value of greater than or equal to 0.9. A simple regression model was developed for each correlation and the results were very similar for all three periods. A population of 200,000 people was equated to between 39,000 and 43,000 acres of developed land. Since this relationship held for the Maryland data set, could the same relationship be established with the limited land use data sets in North Carolina? The earlier LUDA data set appeared to correlate well with developed land, but how could the existing Landsat accuracy for developed land be improved? Landsat land use maps of 21 counties in the APES area were sent out to county planners or other county officials for their review. Each county official was to shade in the extent of development that took place in his county during 1990 and change any land use that was not properly classified. The returned maps were digitized and new acreage for developed land was obtained for each county. Both the LUDA and the corrected Landsat land use maps were compared to the population census data in the same manner as the Maryland information. Both correlation had a R^2 value of greater than or equal to 0.8. Again, a simple linear regression model was developed for each correlation and the results were very similar for both 1970 and 1990 (Figures 10.5 and 10.6). A population of 200,000 people was equated to between 45,000 and 60,000 acres of developed land.

A statistical relationship between population census and developed land for the same time frame has been established for land area in Maryland and the APES area. The relationship is not the same for both areas and probably will vary from region to region. Based on this relationship

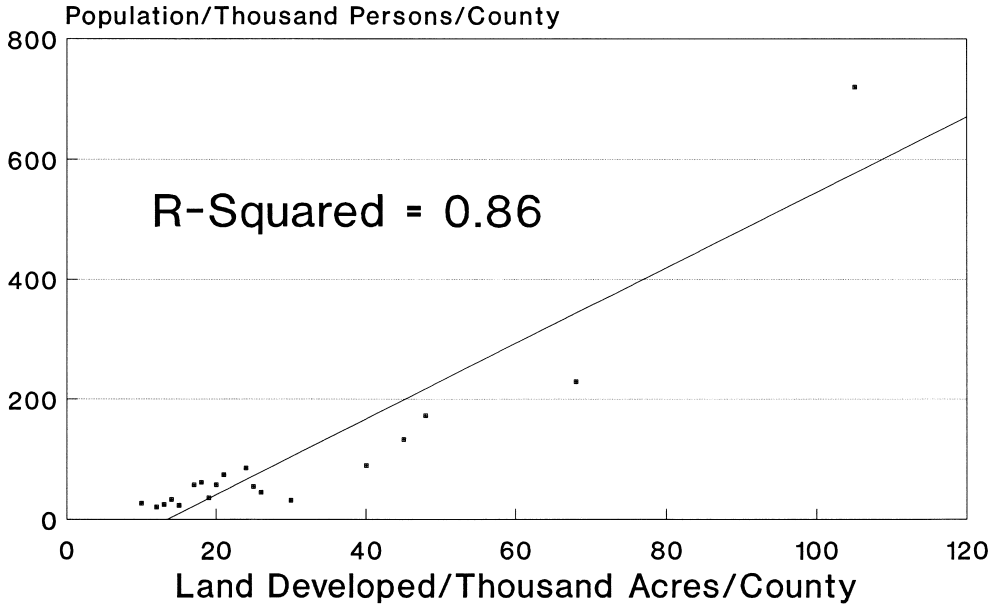


Figure 10.5. 1970 population vs 1972 land use.

the number of acres of developed land for the years 1970, 1980, and 1990 has been estimated for each of the six basins in the APES area. The resulting pattern is similar to that of population, with the Neuse Basin having the largest amount of developed land over the last 30 years. The 1990 figures show the Neuse Basin with approximately 306,000 acres of developed land and the

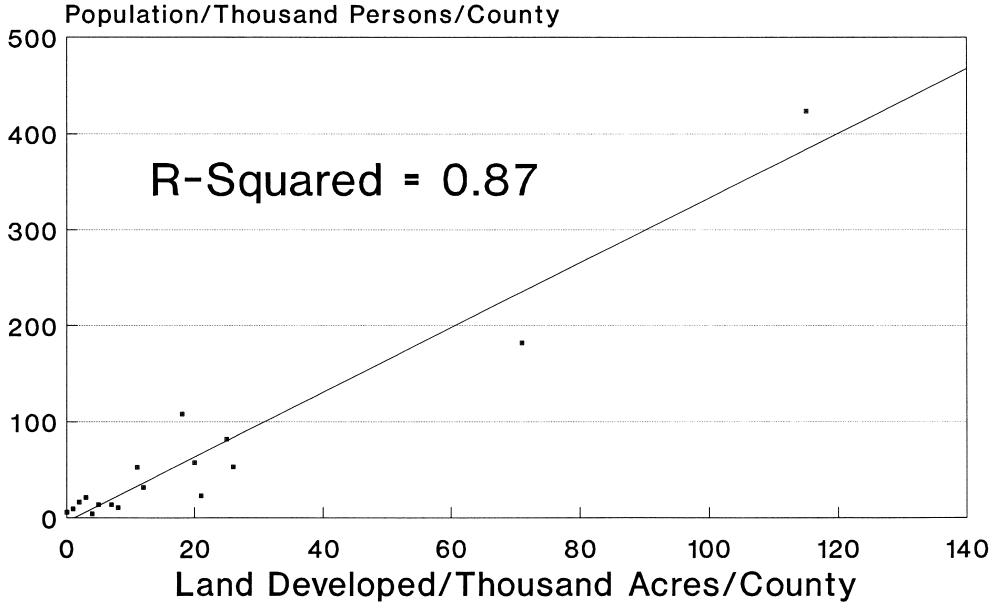


Figure 10.6. 1990 population vs 1990 land use.

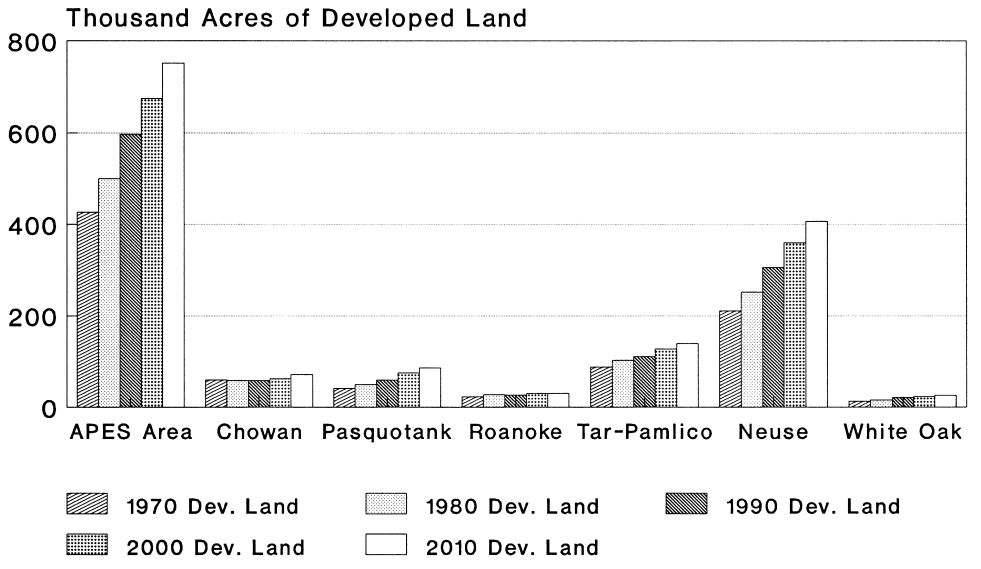


Figure 10.7. Developed land in basins.

White Oak Basin with the least at 65,000 acres of developed land. Developed land for the entire study area for 1990 was approximately 597,000 acres based on this predictive method. Projections for the year 2000 and 2010 (NCDC, 1991; VEC, 1991) have also been estimated and compared to the year 1990. Developed land for the entire study area for the year 2010 is approximately 752,000 acres based on this predictive method. The Neuse River Basin continues to have the most developed land with 407,000 acres and the White-Oak Basin has the least with 26,000 acres (Figure 10.7). The Pasquotank Basin appears to have outpaced the Chowan Basin in the amount of developed land and contains the third largest acreage behind the Neuse and Tar-Pamlico.

CONCLUSIONS

This study found that land use data from Landsat was not adequate by itself to properly identify seven land use classifications. The greatest errors appeared to be associated with the classes of built-up and wetlands. This was due to the forest crown cover obscuring the true land use of rural residential development on the ground and data sources combined with existing data sets in the form of a linear regression appeared to compensate for these two major errors.

A high correlation between population and developed land was found on a county level from Maryland, Virginia, and North Carolina data. Based on this correlation a linear regression model was developed to predict the number of acres of developed land based on the projected population for a particular county. The relation between population and developed land will not be the same for each region but this method can be a powerful tool in predicting where and how development will take place in a particular county.

REFERENCES

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S. Department of the Interior, U.S. Geological Survey Professional Paper 964. U.S. Geological Survey. Washington, D.C.

Burgess, W., E. Christoffers, J. Dobson, R. Ferguson, A. Frisch, P. Lade, and J. Thomas, 1992. Results of a Field Reconnaissance of Remotely Sensed Land Cover Data. Maryland Department of Natural Resources. Annapolis, MD.

Khorram, S., H. Cheshire, K. Sideralas, and Z. Nagy, 1992. Mapping and GIS Development of Land Use and Land Cover Categories for the Albemarle-Pamilco Drainage Basin. Albemarle-Pamlico Estuarine Study. Project No. 91-08. North Carolina Department of Environment, Health, and Natural Resources. Raleigh, NC.

Kleckner, R., 1981. A National Program of Land Use and Land Cover Mapping and Data Compilation. In *Planning Future Land Use*, American Society of Agronomy, Special Publication No. 42.

Maryland State Planning Office, 1991. Maryland's Land 1973–1990: A Changing Resource. Baltimore, MD.

North Carolina Data Center (NCDC), 1991. Population Projections for the Years 2000 and 2010. Raleigh, NC.

Steel, J., and M. Scully, 1991. Projects Funded by the Albemarle-Pamlico Estuarine Study. Albemarle-Pamlico Estuarine Study. Project No. 91-00. North Carolina Department of Environment, Health, and Natural Resources. Raleigh, NC.

Virginia Employment Commission (VEC), 1991. County Population Projections for the Years 2000 and 2010. Richmond, VA.

Wilen, B.O., 1990. The U.S. Fish and Wildlife Service's National Wetland Inventory. In Federal Coastal Wetland Mapping Programs: A Report by the National Ocean Pollution Policy Board's Habitat Loss and Modification Working Group. Kiraly, S.A. and F.A. Cross, Eds., U.S. Department of the Interior. Washington, D.C.

CHAPTER 11

Application of GIS and Remote Sensing for Watershed Assessment

Lloyd P. Queen, Wayne L. Wold, and Kenneth N. Brooks

INTRODUCTION

Located at the western terminus of Lake Superior, the Nemadji River Watershed covers 670 square kilometers in Minnesota (Figure 11.1). Approximately 40% of the area of the Nemadji Basin occurs in what is termed the Red Clay Area, a band of montmorillonite clays, up to 60 meters in depth. The clays, which are a result of offshore sediment deposits from glacial Lake Duluth, are highly erodible and prone to mass wasting. The high levels of sediment transported by the Nemadji River system adversely affect the designated trout streams in the basin, and deposit an estimated 525,000 metric tons of sediment annually into the Duluth/Superior harbor.

The erosion and sedimentation problems associated with the Nemadji Watershed led to the creation of the multiagency research and demonstration "Red Clay Project." This project concluded that erosion in the red clay area and subsequent transport and sedimentation is a natural process that has been intensified by forest harvesting, road building, and land-use conversions from forest to agriculture and open land (Andrews et al., 1980). Forest removal reduced evapotranspiration rates, resulting in a reduced resistance in soils to shear forces because of increased soil moisture content. In addition, the weaker root systems of replacement vegetation compared to the original white pine forests contribute to soil mass movement. The loss of large woody debris in stream channels also is thought to have reduced stream channel stability. In 1991, the conclusions of the Red Clay Project were reassessed, focusing attention on watershed factors that may be of greatest importance in correcting or mitigating the problem of soil mass wasting in the Nemadji basin. This chapter discusses the results of this reassessment.

SETTING

Soil mass wasting is the process of downslope movement of soil that consists of shear stress and displacement along surfaces that are either visible or that can reasonably be inferred (Huang,1983). Movement is thought to be the result of three contributing factors: (1) the high transportability rate potential of the soils; (2) the high soil moisture content; and (3) the absence of strong root systems to hold the soil in place. The Nemadji River Watershed exhibits excessive amounts of soil mass wasting (hereafter referred to as slumping). Most evident in the watershed are the chiefly rotational slumps that occur on hillslopes in the stream valleys and along the streams themselves. These slumps are suspected as a significant source of sediment to both the tributaries and the main stem of the Nemadji (Andrews et al.,1980).

The purpose of this research was to relate the frequency of slumping to watershed characteris-

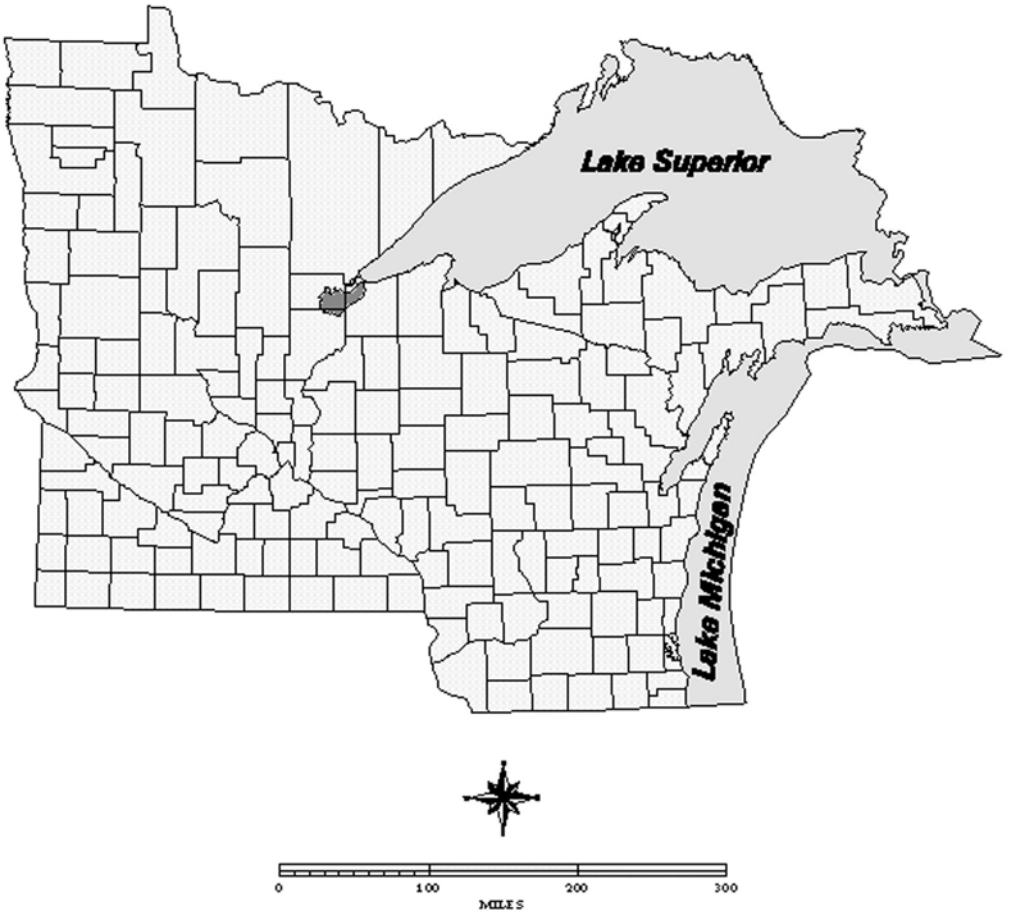


Figure 11.1. Nemadji Basin location map.

tics that potentially have the greatest effect on soil mass wasting. These watershed characteristics were considered independent predictor variables of the dependent variable, frequency of slumping. Secondarily, the research was to determine data requirements for efficient and cost-effective development of a Geographic Information System (GIS) that could be applied for future watershed planning and management.

Based on previous work, this research analyzed the following nine watershed variables thought to affect soil mass wasting: stream channel gradient, time of concentration, stream length, degree of slope, total length of roads in the subwatershed, watershed area, percent coniferous cover, percent deciduous cover, and total forested area. Each of these variables was examined with frequency of slump sites in each of nine subwatersheds (Figure 11.2).

Fortunately, the relationships between forest cover and water yield have been the subject of considerable research. Results from research conducted in humid-temperate regions indicate that, in general, water yield is considerably greater under hardwoods than under conifers due to decreased interception losses and the longer dormant period of deciduous species (Swank and Douglas, 1974).

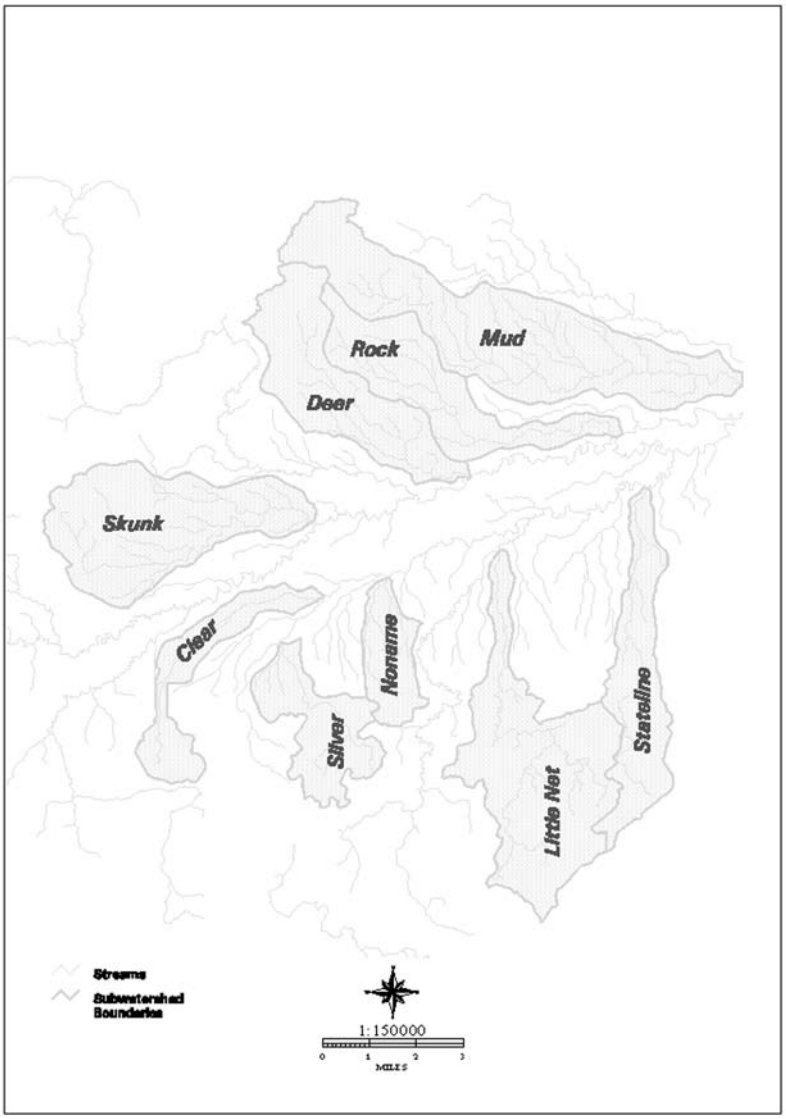


Figure 11.2. Map showing subwatersheds within the Nemadji Basin.

Increased streamflow provides additional energy for streambank degradation. Logging or thinning forests, converting from deep rooted to shallow rooted vegetation, or changing vegetative cover from one with a high interception capacity to one of lower capacity, all have been shown to increase water yield (Brooks et al., 1991).

Studies in Minnesota by Verry (1986) indicate that annual water yield increases and that average annual peak flows can double after cutting upland forests. As the amount of nonforested area in a basin increases, one would expect water yield to increase and with it, higher streamflows with higher velocities and energy. Such conditions could promote the undercutting of slopes and consequently, streambank slumpage. Streambank undercutting, pervasive in the highly erosive red clays, leaves the banks susceptible to slides and slumping. Such effects could be greater following the removal of conifers than hardwoods.

In addition to the role of forest cover type on stream flow, the root strength of different species also can influence soil mass wasting (Sidle, 1985; Abe and Ziemer, 1991). The loss of tree roots diminishes resistance to slumping. On steep slopes bordering streams this combination induces soil creep and slumping. According to the Red Clay Project, high (prefrost) soil moisture levels can be positively correlated with the rate of spring soil mass soil wasting along stream banks (Andrews et al.,1980). Conifers, which transpire later into the fall and retain their interception capacity year-round, should reduce soil moisture during this critical period.

The dependent variable, slump frequency, was determined by tabulating areas where the mineral soil was exposed due to slumping; these sites were identified and mapped from high-resolution low-altitude aerial photographs. Frequency of slumps served as the primary indicator of the relative rate of erosion in each subbasin. In this application, slump frequency was evaluated for both entire subwatersheds and for a series of five discrete buffer zones surrounding each of the streams in those subwatersheds.

APPROACH

Pronounced differences in stream sediment concentrations and related turbidity exist among streams in the Nemadji Basin. To encompass the range of sediment-turbidity conditions, three subwatersheds were selected from each of three turbidity classes, low, moderate, and highly turbid. Classifications based on turbidity were verified using suspended sediment samples and discharge data collected during April and October 1993 and April 1994 (Table 11.1). Suspended sediment, turbidity, and streamflow discharge were measured for both upstream and downstream locations from each of the nine streams (USGS, 1977) at the locations shown in Figure 11.3 (see color section).

Each of the nine predictor variables and maps of slump sites were compiled for each of the nine subbasins in a vector-based GIS. A vector model was chosen for data development and analysis because these systems provide an excellent platform for development and mapping of point (e.g., slump sites), line (e.g., streams), and polygon (e.g., forest stands) data as well as for spatial data analysis. Additionally, local and regional management organizations had previously adopted vector GIS for in-house use. Design criteria for the Nemadji GIS data are map themes at a scale of 1:24,000, with Universal Transverse Mercator (UTM) Zone 15 coordinates applied to all spatial entities, built using conic projections. Base maps for the GIS were derived from existing USGS 7.5 minute quadrangle maps. Quad maps also served as source data for topographic (slope) data, stream locations, and stream gradient estimates. The generation of data themes, the crux of most GIS efforts, is described in the next two sections.

The three primary hydrologic themes are stream gradient, time of concentration, and watershed area. All blue-line (perennial and intermittent) streams were digitized from the USGS 1:24,000 quadrangle maps (Figure 11.4, see color section). The change in elevation from the headwaters to the mouth was taken directly from the quadrangle maps and divided by the length of the stream to determine gradient. Time of concentration was estimated using the Kirpich formula as reported by Gray (1970).

The primary terrestrial variables are slope, stream length, road length, and land cover; and the dependent variable, slump sites. Slope maps were interpolated from a series of contour lines digitized as points from the USGS quads, and a triangulated irregular network (TIN) algorithm was used to create an interpolated slope coverage for all basins. Mean slope was then calculated for each basin (Figure 11.4). The roads theme was acquired in digital vector format from the Minnesota Department of Transportation.

Stereo pairs of conventional color photographs, acquired in May 1992, were used to identify

Table 11.1. Summary of Basin Characteristics

High Turbidity / Moderate Turbidity / Low Turbidity												
Variable	Gradient (m/km)	Discharge (L/s)[a]	S. Sediment (ppm)[a]	TOC (hr)	Stream Length (km)	No. Slumps	Mean Slope (%)	Road Length (km)	Total Area (ha)	% Conifer	% Deciduous	% Forested
Mud	5.3	17–382	6–35	5.37	2.22	45	52	49.2	3889	21	53	74
Rock	7.0	6–49	33–127	3.57	13.8	39	57	22.0	1830	22	33	55
Deer	7.9	54–106	43–105	3.25	11.1	45	50	13.5	2068	18	50	68
Skunk	5.7	26–371	6–30	3.95	14.3	27	66	33.3	2680	23	47	70
Noname	13.2	4–86	10–73	1.53	5.5	25	41	14.2	750	29	52	81
Clear	5.9	19–47	5–24	2.98	10.3	9	18	11.6	1031	29	44	73
Silver	9.1	98–193	1–17	2.13	6.9	2	36	11.7	1234	24	58	82
Little	7.8	66–357	1–11	4.71	19.5	13	52	18.7	3224	18	77	95
Net	12.3	36–95	4–12	3.38	14.3	25	49	11.3	1631	27	49	76
Stateline												

[a]Ranges are taken from three sampling dates (4/93, 10/93, and 4/94) and are reported for the downstream sampling point only.

and map slump sites. Photos were acquired from a helicopter platform using a 35 mm format camera. Negatives were printed at a nominal scale of 1:5,000, and interpreted using a 3-power mirror stereoscope. Land cover was mapped from stereo-pairs of 1:24,000 NHAP photographs into four land cover classes: (1) agricultural/open land; (2) coniferous forest; (3) deciduous forest; and (4) open water. NHAP prints were manually interpreted under magnification using a 0.5 hectare minimum mapping unit. Digital files of land cover were plotted and visually field-checked for accuracy of the cover typing (Figure 11.5, color section).

Slump sites were located on the 1:5,000 scale photos using a magnifying stereoscope. After intensive field checks (using a 75% sample) for accuracy and completeness, sites were entered into the GIS. Large-scale photos were required for accurate interpretation and each slump site could only be represented as a point location. However, due to differences in scale between the photos and the base map, it was difficult to locate the slumps precisely on the base map. Manual editing was employed to reposition the sites in some cases; as a result slump sites were estimated to be located to within 60 meters of their true location at a scale of 1:24,000 (Figure 11.6, color section).

The end result of data compilation was a 1:24,000 scale GIS for all of the nine subwatersheds. Values of land cover class area, ratio measures of percent cover, and lengths and values for the other variables as calculated within the GIS are reported in Table 11.1. The final data preprocessing step was to create data on buffer strips to examine the role of proximity of forest cover to stream channel as a covariate for the independent variables. To assess this effect, five buffer zones were defined along each stream channel, representing buffer strips of 10, 20, 40, 80, and 300 meters on each side of the stream.

Finally, a series of linear regressions were performed based on the frequency counts of soil mass movement sites in each basin and the measurements of the independent variables from the GIS.

RESULTS

The pattern of discharge follows the expected annual trend of highest flows occurring in April, with decreasing flows from August to October. The highest discharge gauged was just over 382 liters per second, measured in April on Skunk Creek. Although the low sample size prevents derivation of sediment rating curves and the use of statistical tests, there is no apparent difference in the observed discharges between the three classes that were grouped on the basis of stream turbidity. Overall, the more highly turbid streams (e.g., Mud, Rock, and Deer Creeks) not only had higher suspended sediment values, but they also exhibited more load variability over the sampling period.

Stream gradients are not higher in the more turbid streams (Table 11.1), and mean slope values are scarcely higher (but not statistically significant) for the more turbid streams. Total subwatershed areas vary from 3,889 hectares for Mud Creek to 750 hectares for Noname Creek; stream length and time of concentration are proportional to the total subwatershed areas for all of the basins. Mud, Rock, and Skunk Creeks have the highest values for total lengths of roads, and all are classified as having high to moderate turbidity. However, it is possible that road building may have its greatest effect at specific road cuts and bridge locations where slope stability was reduced.

Stream gradient was not significantly correlated to the number of slump sites and was relatively constant along the profiles. In a similar fashion, stream length did not exhibit a geographic pattern of variability, nor was it correlated to slump frequency. Time of concentration, conceptually, is an indication of the time it takes for water to travel from the hydraulically most distant point of a watershed to the mouth of the primary stream. The less the amount of time required, the faster a

given stream will flow and thus the greater the energy potential for streamflow through the stream channel. No significant linear relationship was found between time of concentration and the number of observed slumps. However, there is little difference in time of concentration values among the different subwatersheds (Table 11.1).

The relative proportions of land-cover types varies considerably for the subwatersheds. The northern-most subwatersheds (Mud, Rock, and Deer Creeks) generally have lower forest cover area and higher suspended sediment concentrations (Table 11.1) than those subwatersheds to the south. Deciduous forest was the dominant forest type in all of the subwatersheds, with a maximum in the Little Net subwatershed of 77% and a minimum of 33% in Rock Creek basin. There was no strong geographic pattern of differences in the distribution of coniferous forest from one subwatershed to another.

As the width of the buffer zones increases, generally we see a reduction in the proportion of conifer cover to deciduous cover as a percentage of the area of the buffer. Across all nine subwatersheds it appeared that conifer cover types tended to occur most frequently in near-stream locations. When 300 meter widths were considered, the buffer strip usually extended to the ridgetop. This in part explains the increasing proportion of nonforested (agricultural and open) land at the 300 meter width.

The regression analysis of watershed characteristics and slump frequency yielded only one significant linear relationship. When the regressions were repeated for all of the buffer strip widths for all independent variables, again only one relationship was significant. In both cases forest cover percent was the significant variable. Note that the highest turbidity streams tend to have the most slump sites and a smaller relative proportion of forested land. Yet, low turbidity and high turbidity streams occur in watersheds with about the same percentage of conifer cover. The need to consider interactions of other variables and the distribution of the conifer cover in the basin is clear.

The percent slope of hillsides is generally assumed to be a significant covariate when examining causes of soil mass wasting. As percent slope increases, the effects of other contributing factors are amplified. These data show that as the mean slope of a subwatershed increases, there is a tendency for the number of slumps to increase as well. Although mean basin slope was not a significant predictor variable, it was noted that slope effects may be site specific, and when averaged over a large area, this effect was not apparent. Percent slope at each slump site, or the configuration of the hillslope associated with each slump site, could provide a different result.

The relationship between the total length of roads within each subwatershed and the number of slumps within that subwatershed was not significant. The larger the watershed, the greater the road length contained within it. Thus, the length of roads within a subwatershed was correlated with the size of the subwatersheds and not with the number of slumps. Results show the lack of a significant linear relationship of area to the dependent variables. As with slope percent, we suspect that roads need to be investigated on a site specific basis. The effect of roads most likely enhances slump occurrence at or very near the road cut itself. Because they were not mapped in this study, the impact of logging roads and skid trails remains untested.

On a subwatershed level, the occurrence of slumps generally did not diminish with increasing percent of conifer cover. There is a significant inverse relationship between the amount of total forest cover in a subwatershed and the number of slumps observed. As the amount of total forested area in a subwatershed increases the number of observed slumps decreases. However, when the buffer zones were analyzed this relationship did not persist. In fact, the only significant relationship found when repeating all regressions for all variables and all five buffer zone widths was the percent of coniferous cover within 80 meters of the stream channel. This might suggest that the percentage of the total watershed area in open/agricultural land affects the number of slumps, but

the effect of location, or distribution, of the nonforested area in proximity to stream channels could not be tested at the scale and resolution of this study.

Study Implications

The GIS created for the basin contains maps of slump locations, land cover conditions, slope, and other characteristics for portions of the Nemadji River Watershed. In addition to serving as a data management tool, the GIS was used as the basis for data development and analysis. As a result of building GIS capability, several relationships between physical watershed characteristics could be examined systematically and efficiently.

Land cover was disproportionately distributed between the four land cover classes. Deciduous forest cover was more abundant in the upper reaches of the subwatersheds than in the lower reaches. In all subwatersheds, there is more deciduous cover than coniferous cover. Progressing downstream, the amount of conifer cover increased, especially within the stream valleys themselves. Much of the conifer cover was balsam fir (*Abies balsamea*) and spruce (*Picea* sp.) rather than the larger and deeper rooted white pine (*Pinus strobus*) that was the dominant conifer at the turn of the century. Conversely, the amount of open/agricultural area decreased from the upper to lower reaches of the stream. When large open areas occurred from the ridgetop leading down to a stream, there was generally a higher frequency of slumps in association with the open area.

Slope was identified as a significant covariate with slump locations from the Red Clay Project and other sources. Although our results did not show a significant linear correlation between percent slope and slope failure in the subwatersheds, there were plausible explanations. The mean slope and the number of slumps tended to increase closer to the mouth of the stream. Coincidentally, the percentage of conifer cover increased in lower areas as well. While the reason for the increase in conifer cover downstream was not known, it was possible that it would be more difficult to harvest on steep slopes.

Furthermore, the type of coniferous species present, which was not quantified from the aerial photographs, can influence slope stability through differences in root strength and rates of transpiration (for example, balsam fir is shallow-rooted and would not have the same slope stability effects of white pine or red pine). These two effects, among others, greatly complicate the analysis of cover type by slope class.

CUMULATIVE EFFECTS

The occurrence of soil mass wasting in the Nemadji Basin most likely was an expression of the cumulative effects of land use activities that have taken place over the past century. The removal of white pine forests at the turn of the century, the accompanying expansion of agricultural development, road construction, and the conversion from white pine to aspen-dominated forests all affected processes that would be expected to reduce slope and channel stability. The GIS approach allowed the consideration of individual variables as they might affect slumpage; however, the cumulative effects of all these activities over the watershed was not easily quantified and may be a situation of the total being "greater than the sum of the individual parts."

The removal of white pine and the subsequent replacement by either farmland or aspen forests would be expected to reduce slope strength because of loss of root strength on slopes. Overall reductions in evapotranspiration, leading to wetter soils, would increase soil stress. The loss of large woody debris in the channels would be expected to reduce channel stability, which coupled with increased water yield and stormflow volumes and peaks, would promote greater channel scour and streambank erosion. Road construction in the basin adds to these problems. Such processes occur-

ring at the toe of the slope in these deeply incised channels would enhance soil slumps. After nearly 100 years, this watershed system still expresses, through excessive channel erosion, slumpage, and the resulting sediment loading, the cumulative effects of land use change. If we intend to develop models that help explain the cause-and-effect relationships of these systems, a better overall assessment of these cumulative effects is needed.

Because the percent of forested area was inversely related to a greater frequency of soil slumps, management should logically be aimed at increasing and maintaining forest cover in the watershed. Although these results could not identify preferred species for controlling slumps, the literature suggests that certain species such as white pine would be expected to provide more slope stability than other species such as aspen or birch. A more detailed look at species suitable for slope stability is warranted.

The results from this study underscore the need for future research into hillslope hydrology and fluvial processes in the Nemadji Watershed. This study points to certain factors that appear to be related to soil mass wasting and suspended sediment relationships in the Nemadji. Yet no definitive results exist to specifically address the actual relationships of water flow and subsequent soil mass wasting processes. More extensive streamflow data are needed in concert with hillslope process studies. Rainfall and snowmelt events over a range of magnitudes need to be studied so that we can better understand temporal relationships. In fact, the role that large rainfall events play in erosion is as yet unresearched in the Nemadji. Given these results and the availability of the GIS database, a process has been developed that enables managers to continue to examine watershed features and to monitor and track cumulative effects across this critical watershed.

REFERENCES

Abe, K., and R.R. Ziemer, 1991, Effect of tree roots on shallow-seated land slides. In *Proceedings, IUFRO Technical Session on Geomorphic Hazards in Managed Forests.* USDA Forest Service Gen. Tech. Rep. PNW–130, pp. 11–20.

Andrews, S.C., R.G. Christensen, and C.D. Wilson, 1980. Impact of Nonpoint Pollution Control on Western Lake Superior. Red Clay Project Final Report. U.S. Environmental Protection Agency Report, EPA 905/9-79-002-B.

Brooks, K., P. Folliott, H. Gregersen, and J. Thames, 1991. *Hydrology and the Management of Watersheds*. Iowa State University Press, Ames, IA.

Gray, D.M., Ed., 1970. Handbook on the Principles of Hydrology. Reprint. National Research Council of Canada. Port Washington: Water Information Center, Inc.

Huang, Y., 1983. *Stability Analysis of Earth Slopes*. Van Nostrand Reinhold Company, Inc., New York, NY.

Sidle, R.C., 1985. Factors influencing the stability of slopes. In *Proceedings, Workshop on Slope Stability: Problems and Solutions in Forest Management*. USDA Forest Service Gen. Tech. Rep. PNW-180, pp. 17–25.

Swank, W.T., and J.E. Douglas, 1974. Streamflow greatly reduced by converting deciduous hardwood stands to pine. *Science*, 185(4154):857–859.

United States Geological Survey, 1977. Sediment. In *The Handbook of Recommended Methods for Water Data Acquisition.* United States Geological Survey, Reston, VA, Chap. 3.

Verry, E.S. 1986. Forestry harvesting and water: The lake states experience. *Water Resources Bulletin*, 22(6):1039–1047.

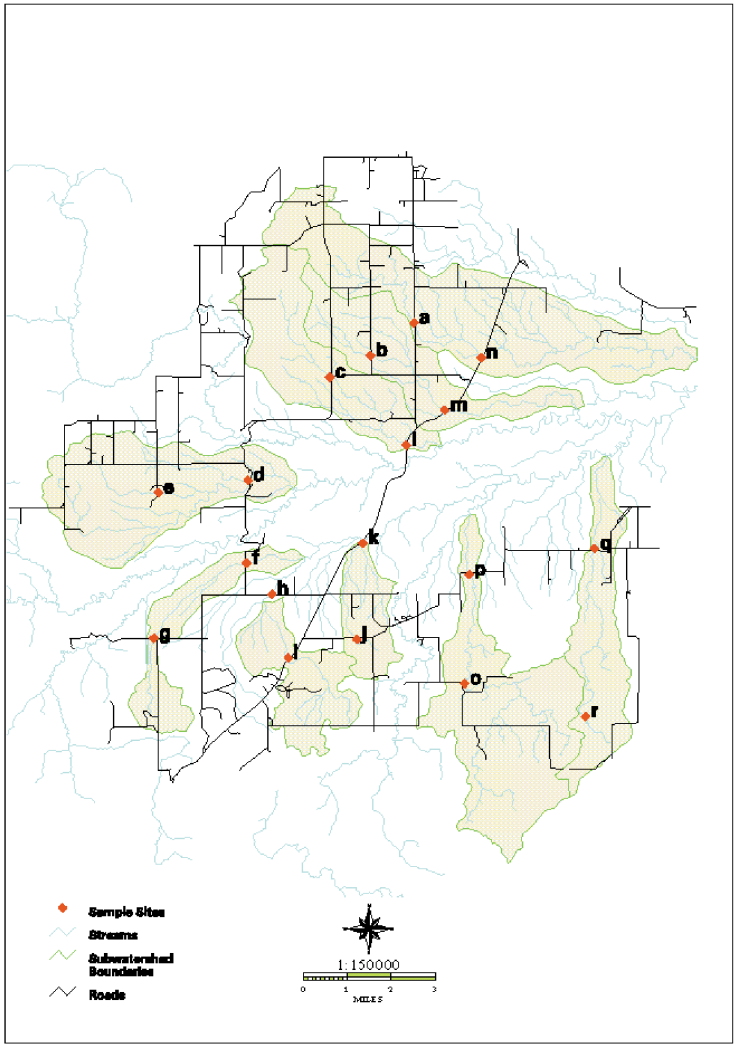


Figure 11.3. Sediment and streamflow sample site map.

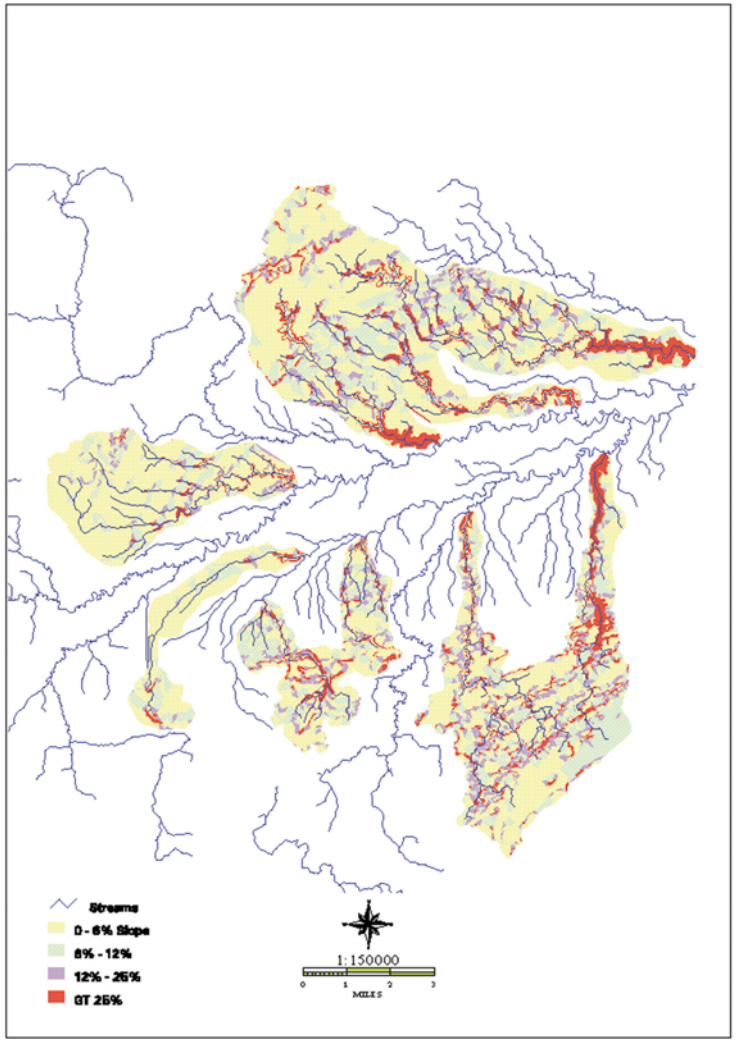


Figure 11.4. Slope maps of subwatersheds.

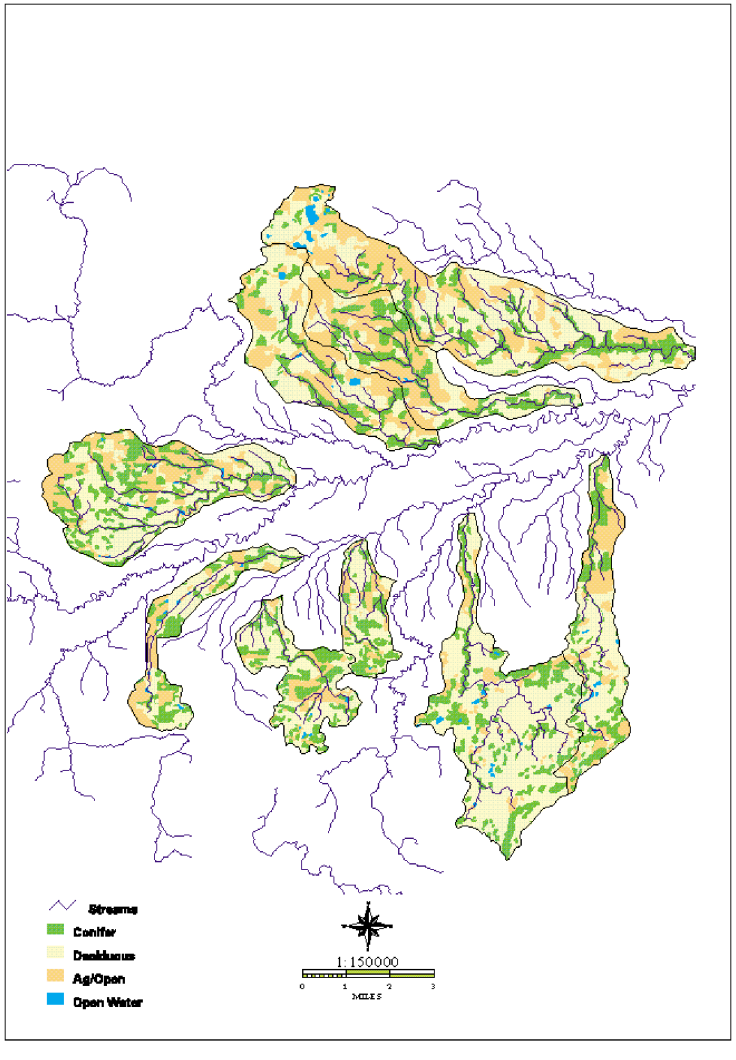


Figure 11.5. Land cover map for the subwatersheds derived from aerial photography.

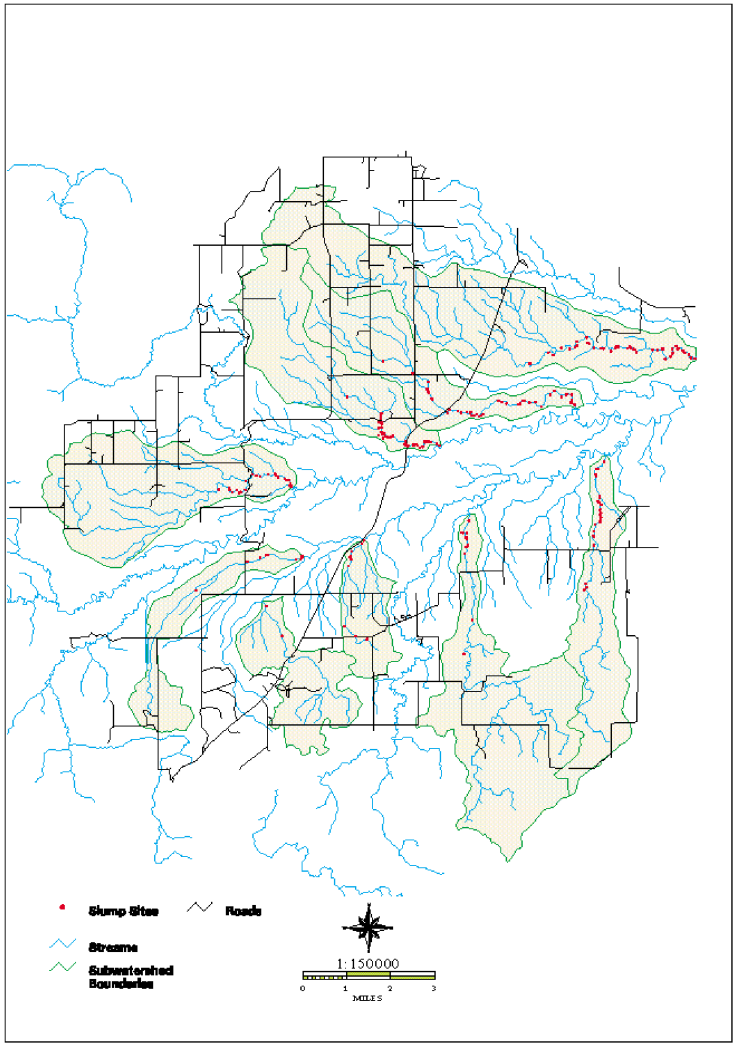


Figure 11.6. Location of mass wasting sites (slumps).

CHAPTER 12

Development of a Database for Lake Ecosystem Studies: Linking GIS with RDBMS

Weihe Guan, Leslie J. Turner, and Sergio L. Lostal

INTRODUCTION

A comprehensive data management system is essential for any ecosystem study. In the Okeechobee Systems Research Division, Ecosystem Restoration Department, South Florida Water Management District (SFWMD), large amounts of spatial data have been accumulated by years of Lake Okeechobee ecosystem studies. This chapter introduces the development of a database in support of the division's lake ecosystem studies. This database links a geographic information system (GIS) with a relational database management system (RDBMS) to best facilitate data use.

Developing and maintaining an integrated GIS-RDBMS database involves the following steps: (1) the establishment of a data server system conveniently accessible to all users; (2) the selection of appropriate RDBMS/GIS software for both attribute data and geographic data; (3) the development of a general data format and database structure; (4) the formalization of data management procedures, including input, update, conversion, QA/QC, and backup; and (5) the implementation of data query and retrieval utilities for end users to search, display, print, or plot information. This chapter documents the major steps in developing a Lake Okeechobee GIS-RDBMS database. The process focuses on the Lake Okeechobee ecosystem studies. Hardware and software availability impacted significantly the entire development approach. The chapter introduces a concept for modeling fuzzy geographic features, which addresses special needs in ecosystem studies. Also introduced in the chapter is the approach for establishing a virtual data server system through a computer network.

BACKGROUND

Lake Okeechobee (Figure 12.1) is the central feature of the Kissimmee River / Lake Okeechobee / Everglades hydrologic ecosystem in south Florida. The lake is a large (approximately 700 square miles), shallow (average depth about 10 feet), subtropical lake which provides water, flood protection, and recreational benefits for a population exceeding 3.5 million people. The lake is also an important biological habitat for economically important fish and wildlife, including several threatened and endangered species (Aumen, 1995).

Various factors have contributed to the deterioration of the Lake Okeechobee ecosystem. Among these factors is excessive nutrient loading from agricultural activities in the watershed, which has caused increased blue-green algal blooms. These blooms, characterized by surface

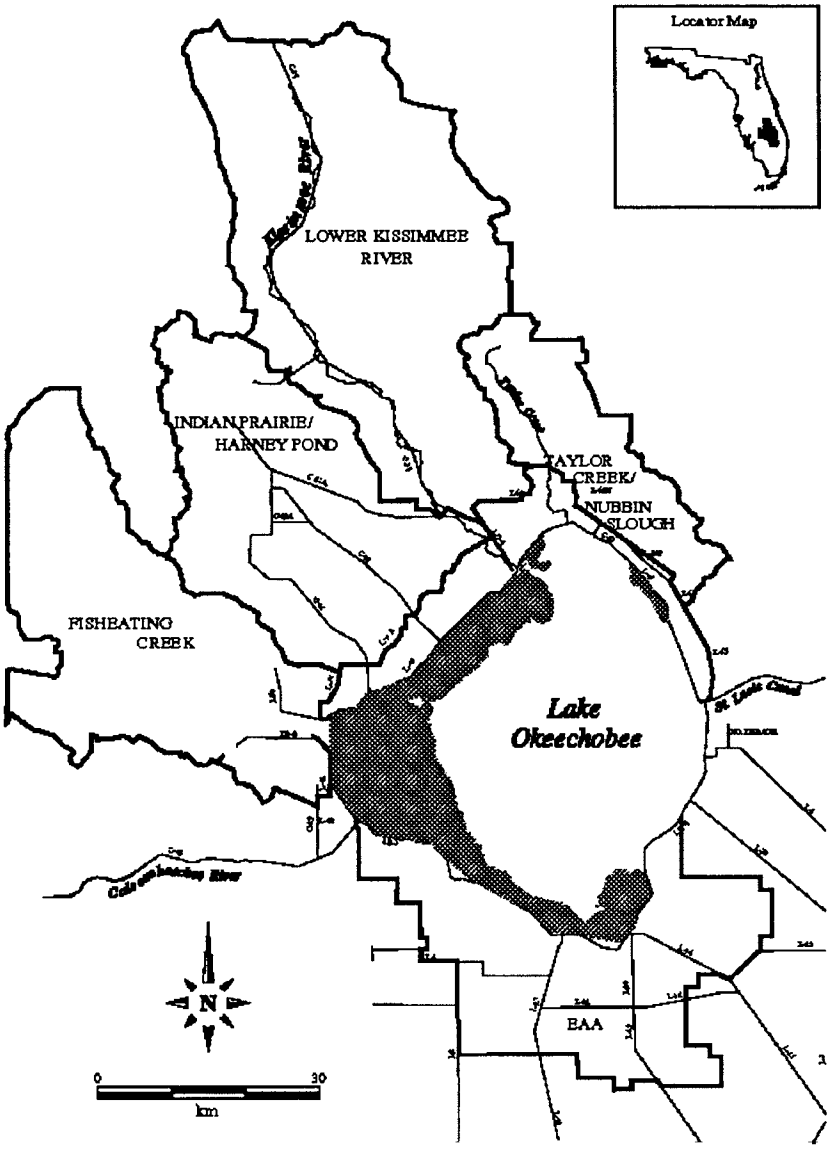


Figure 12.1. Lake Okeechobee and associated ecosystem studies.

scums and unpleasant tastes and odors, raised concerns about declining water quality (Aumen, 1995). Several interdisciplinary, multiyear research efforts were initiated in the late 1980s in response to the algal blooms, including a lake ecosystem study.

The Lake Okeechobee Ecosystem Study (LOES), conducted by the University of Florida under a contract with the South Florida Water Management District (SFWMD), was unique in that it looked beyond excessive nutrient loading to other components of the ecosystem. Research topics include water level effects, water quality, fish and wading bird populations, and wildlife habitat. The study's objective was to provide an ecological baseline against which future ecosystem trends can be compared, and to assess the general health of the ecosystem. The database structure that will be described in this chapter is based on the results of this LOES.

Data Related to Lake Okeechobee Ecosystem Study

The Lake Okeechobee Ecosystem Study addressed the following issues (University of Florida, 1991): (1) data synthesis, modeling, and database management; (2) water chemistry and physical parameters; (3) community and ecosystem ecology of emergent macrophytes; (4) phytoplankton, bacteria, epiphytes, submerged plants, macroinvertebrates, and zooplankton; (5) distribution and abundance patterns and the reproductivity and foraging ecology of wading birds; and (6) larval and juvenile fish. The project involved extensive field data collection and analysis. Data related to the study were archived on floppy disks using Lotus 1–2–3 spreadsheets, WordPerfect documents, ASCII text files, and ERDAS raster files. Data were categorized as follows: (1) plankton—bacteria, bioassays, nitrogen fixation, phytoplankton, and zooplankton; (2) plants—emergent, nutrients, seeds, soils, and submergent; (3) water quality—chlorophyll, nutrients, physical chemistry, and suspended solids; (4) wildlife—birds and fish; (5) hydrology—Lake Okeechobee hydrological data; (6) spatial data—GIS coverages, images, and locational files; and (7) documentation—various text files for clarification, identification, and explanation.

Importance of GIS

Most of the data collected in LOES have locational records. Location is recorded either by x-y coordinates or by verbal descriptions. The geographic information system (GIS) was identified as an important tool for the lake ecosystem study. As stated in a LOES annual report (University of Florida, 1991), the objective of LOES was to use an ecosystems approach to develop a set of tools (models and GIS databases) that integrate the data from various tasks in the project and other projects to provide predictive capabilities with which the SFWMD can evaluate the consequences of various water management options on the marsh littoral zone of Lake Okeechobee and its interaction with the pelagic zone of the lake. Critical concerns include impacts on the fish and wildlife resources, the role of the marsh in nutrient dynamics and exchange with the lake, and estimates of the total flux of nutrients to and from the lake under different water regimes. The following areas of focus were suggested by the LOES review panel (University of Florida, 1991): (1) impact of lake stage on biotic communities; (2) impact of nutrient concentrations on biotic communities and trophic relationships; (3) direct and indirect effects of plant community structure on critical habitat and energy flow to wading birds and fishes; (4) role of littoral zone in the lake's ecology; (5) effects of water pumping from canals on lake communities and productivity; and (6) role of exotics in lake ecology. These focus areas outlined user database requirements.

The LOES review panel (University of Florida, 1991) suggested that a spatially based predictive model utilizing a GIS approach be developed, capable of predicting the responses of fish and wildlife resources to management options such as lake stage manipulation, nutrient loading increases or decreases, or the long-term effects of maintaining the present regimes of stage, nutrients and flows. The model was envisioned to be sufficient to provide indications of the magnitude of the changes in the system and the spatial location of such changes utilizing the GIS information layers and model parameters derived from the various tasks of LOES and other projects.

To effectively use information collected by LOES and other studies, a functional database linking GIS with a relational database management system (RDBMS) was needed. An RDBMS can efficiently manage the large amount of information while a GIS can present data spatially. The effort of developing this database included establishing a data server system, designing an integrated database structure, developing a database management procedure, and implementing a data query and retrieval user interface.

DATA SERVER SYSTEM

A data server system is a computing platform that hosts databases. It may include computers, operating systems, networks, database management systems, geographic information systems, and other hardware and software necessary to input, store, manage, query, and output data (Cowen et al., 1995). Ideally, a data server system should be selected according to the conceptual design of the database to be developed. In reality, many constraints, especially financial, limit available options for the selection of a data server system.

In this study, the computing environment consists of ORACLE as the RDBMS software on a VAX minicomputer, and ARC/INFO and Arcview as the GIS software on SUN SPARC workstations. Some workstations use SUN OS and others use Solaris as the operating system. All computers were networked. Desktop workstations (SPARC 2, 10, 20, or Ultra) and personal computers were available to end users of the GIS-RDBMS database.

The GIS database was a special component of a relational database hosting the information collected by LOES. The GIS database had two components: ARC/INFO coverages for geographic features and unique feature IDs, and ORACLE tables for attribute items with the feature ID as a unique key. The connection between the two was built on the Database Integrator in ARC/INFO and ArcView.

Due to software and storage space limitations, the database cannot be loaded onto a single computer. Several networked computers were needed. Tabular data are stored in ORACLE tables on the VAX, and spatial data were stored in ARC/INFO coverage format on several workstations. One workstation, a SUN SPARC 20, was designated as the "virtual" data server. The graphic user interface for data query and retrieval was installed on this workstation, with the directory structure for all ARC/INFO coverages. Symbolic links in subdirectories point to the actual locations of the coverages on other workstations. When users query on spatial features from a workstation, an ARC-ORACLE interface links to the ORACLE database on VAX and returns appropriate tables and records (Figure 12.2).

This design provides users with a seemingly holistic database on the virtual data server. Users need not know the real storage locations of the database components, nor do they need to interact with any computer platform other than the virtual data server. On the other hand, the group of networked computers collectively provides the required storage and computing capacity necessary for the database, which does not exist in any single computer. When the database grows, more computers may be brought into the group with minimum impact on existing server components.

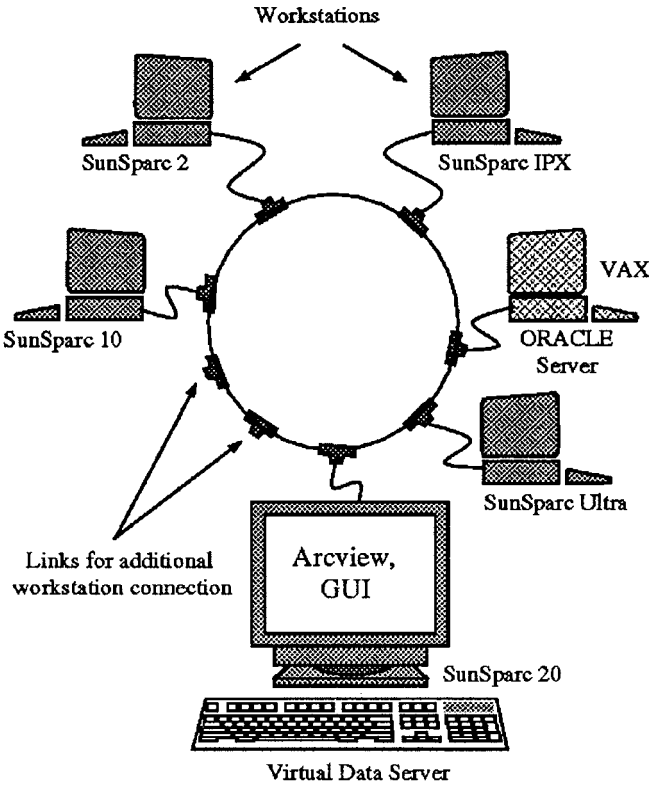


Figure 12.2. The networked data server system.

The disadvantage of this data server structure is its reliance on the network and each computer in the structure. If one data-hosting computer goes off the network, the database will not function properly. Moreover, the database manager must have access to all database computers for maintenance purposes. Because most of the involved workstations are routinely used by SFWMD staff as desktop computers, special effort is needed to coordinate their use for the database.

THE DATABASE MODEL

A model can be thought of as a real-world abstraction where only essential details are kept. Database models are created to understand the data organization before the database and support programs are built. The LOES database model was created using the Object-Oriented Modeling Technique (OMT) (Rumbaugh et al., 1991). Subsequently, the LOES database model was mapped as a relational database and built on a VAX minicomputer using an ORACLE RDBMS.

Originally, LOES data were organized in spreadsheets with different record formats. After data collection was completed, scientists developed a database model from which data could be extracted in convenient formats. The resulting database model has five modules (Lostal, 1996): (1) Locational, (2) Ecological Variable, (3) Biological Species, (4) Time Series, and (5) Organizational.

Locational, Ecological Variable, Biological Species, and Organizational modules have separate, well-defined purposes. The Locational module deals with the measurement site or station. The Ecological Variable module handles information about the type of data (i.e., sample method, parameter, and unit) measured. The Biological Species module describes scientific and common names of measured species. The Organizational module addresses who (i.e., agency, observer) produced the data and why (i.e., project, study, task) the data were produced.

Time Series, the last module, depends on the other modules. Time series are determined by stations, ecological variables, and biological species; they contain summary information and a data point set. Summary information includes items such as period of record, number of observations, average or sum, standard deviation, and maximum and minimum values. The data point set is formed by all data points observed. Each data point encompasses a time series identifier, time stamp, measured or described ecological value, quality indicator information, and organizational code.

The database's architecture allows spatial queries about ecological data. For example, whenever a user selects a location in a map coverage, the GIS client program sends the request to a utility program that chooses the station closest to that location. The program queries the database using the selected station and the database returns a list with all the time series summaries stored for that station. The user reviews the list, selects an appropriate time series, and submits the request. The program queries the database, receives the information, and displays the data point set. As the example shows, the ORACLE database and the GIS client interface initially through locational information.

The Locational Module

The LOES database was modeled using object-oriented methods. As its name suggests, the object-oriented approach organizes real-world concepts as objects. Objects are things or abstractions with boundaries and meaning in the real world. Objects with similar properties are grouped into classes. A database model shows the classes that compose a database, and how these classes associate with each other. The Locational module deals with the classes that represent the measure-

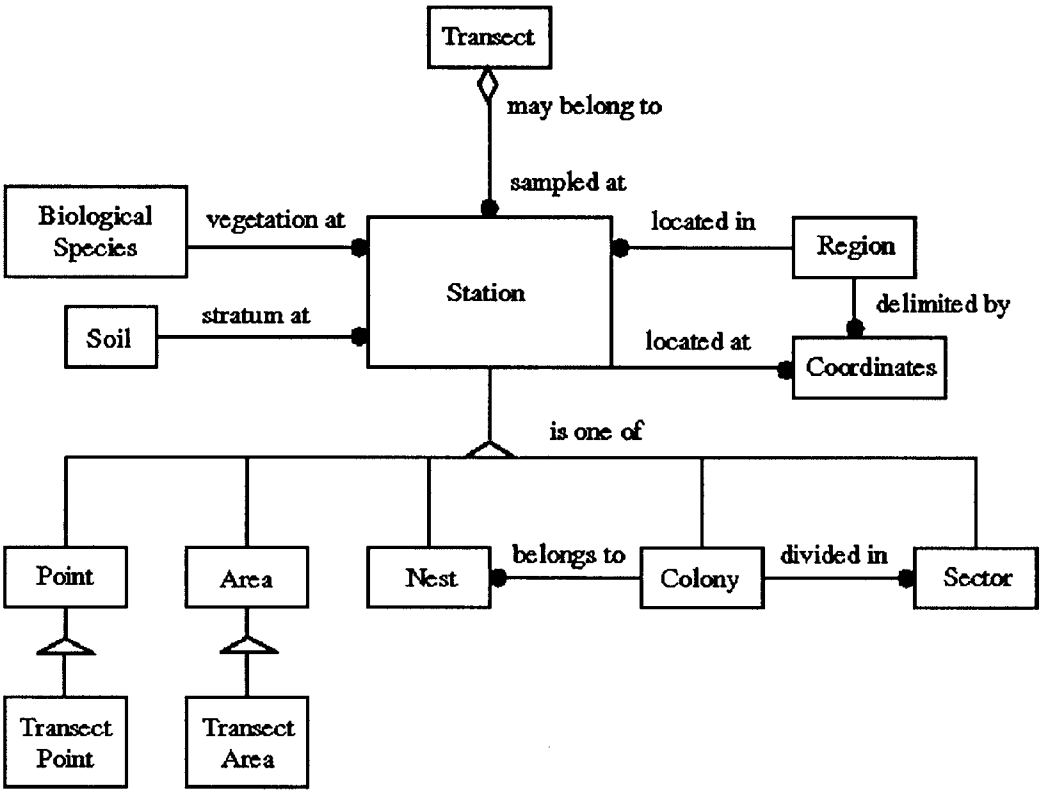


Figure 12.3. Locational module object model.

ment site and their associations. The GIS client is interested particularly in this module, because any information displayed on a map coverage is attached to a measurement site.

Figure 12.3 shows the database model for the Locational module. The measurement site, depicted by the Station class (classes are represented as rectangles in the diagram) is the fundamental component of this module. All classes included in this module associate with the Station class (associations are represented by lines in the diagram). Some classes are specific location types, other classes associate spatially with the Station class, and the rest provide complementary information to the Station class.

Ecological data were collected at a variety of places such as water quality monitoring points, fishing sites, and bird colonies. The Station class was an abstraction that represents any location type. This special association, called a generalization (represented by a triangle in the diagram), was used to classify related classes. As the figure shows, the Station class was a generalization of five location types used to collect data: (1) Point, (2) Area, (3) Bird Colony, (4) Colony Sector, and (5) Bird Nest. Points are sites identified by one pair of x-y coordinates. Areas refer to locations specified by more than one pair of x-y coordinates (two points define a rectangle). Other station types describe bird information. Bird colonies were sites where birds live and procreate. Both colony and point positions were given by one pair of x-y coordinates. However, colonies were classified separately from points because they also associate with other station types such as nests and colony sectors. The use of labels helps to interpret associations. For example, the association between the Nest class and the Colony class indicates that many nests (the filled circle means many) belong to a colony.

Spatially, the Station class associates with regions and transects. Regions are large areas where

stations were located. The diagram indicates that many stations were located in a region. Transects are station aggregates sampled along a transect path. This is a "whole-part" association or aggregation represented as a diamond in the diagram. The transect is the total assembly and the stations are its elements. Of the five station types, only point and area stations may belong to transects. This aspect is shown on the figure by representing Transect Point and Transect Area as separated classes.

The Coordinates classes were related to both Region and Station classes. The Region-Coordinates association shows that a region was delimited (specified) by many coordinates. The association between the Station class and the Coordinates class was more general. The figure indicates solely that a station is located at many coordinates. This association considers all location types explained above. First, a station refers to a location type indicated by one set of x-y coordinates. Second, the station was an area determined by more than one set of x-y coordinates. Third, a station position may be reported without using coordinates values; instead, a descriptive method was used (See Soft Points and Soft Polygons section).

The last two classes in the figure, Biological Species and Soil, serve as "catalogs" for the Station class. For example, the Biological Species class contains all information about vegetation including scientific classification and common name. However, instead of duplicating all the information, the Station class contains solely abbreviated descriptions. When a full description is required, the Biological Species and the Soil classes supply the information that is not stored in the Station class.

Soft Points and Soft Polygons

Due to the biological and ecological nature of the data sets, the LOES database developers needed to process spatial data with uncertain locational coordinates. To address this issue, we introduced the concept of soft points and soft polygons. A soft point is a point without a definite x-y location, and a soft polygon is a polygon without a definite boundary. In ecosystem studies, researchers often have to deal with soft points and polygons when historical and field survey data are involved. Before GIS was implemented, many field observations were made with a verbal description of location, not explicit x-y coordinates. Even with well-established GIS concepts, some ecological features were difficult to describe at a definite x-y location or within a distinct boundary. For example, a given fish species was observed in a certain area of a water body at a certain time. That "certain area" of the water body does not have a clear boundary. Such observations may apply to animals on land, plankton in water, or a floating plant mass in a wetland.

In the soft feature model, the geographic location of a soft point was described by its probability distribution in space. A soft point may appear at any known location with a certain probability. That known location was usually contained in a polygon. When the probability equals one at a known point, the soft point becomes a hard point. Where the probability equals zero, the soft point never appears. The line between none-zero and zero probability areas was the boundary of the probability distribution zone. One soft point may have multiple probability distribution polygons (Figure 12.4).

The geographic location of a soft polygon was more difficult to describe than that of a soft point. Three parameters were required to define a polygon: size, shape, and the location of the gravity center. When any of these parameters is uncertain, the polygon becomes a soft polygon. In theory, this leads to seven types of soft polygons (Table 12.1). In this study, two types of soft polygons are discussed (Types 1 and 5 in Table 12.1), which are most common to ecological studies.

For soft polygons with uncertain size, shape and location, the probability distribution patterns were similar to those of soft points. The probability polygons were stored as an ARC/INFO cover-

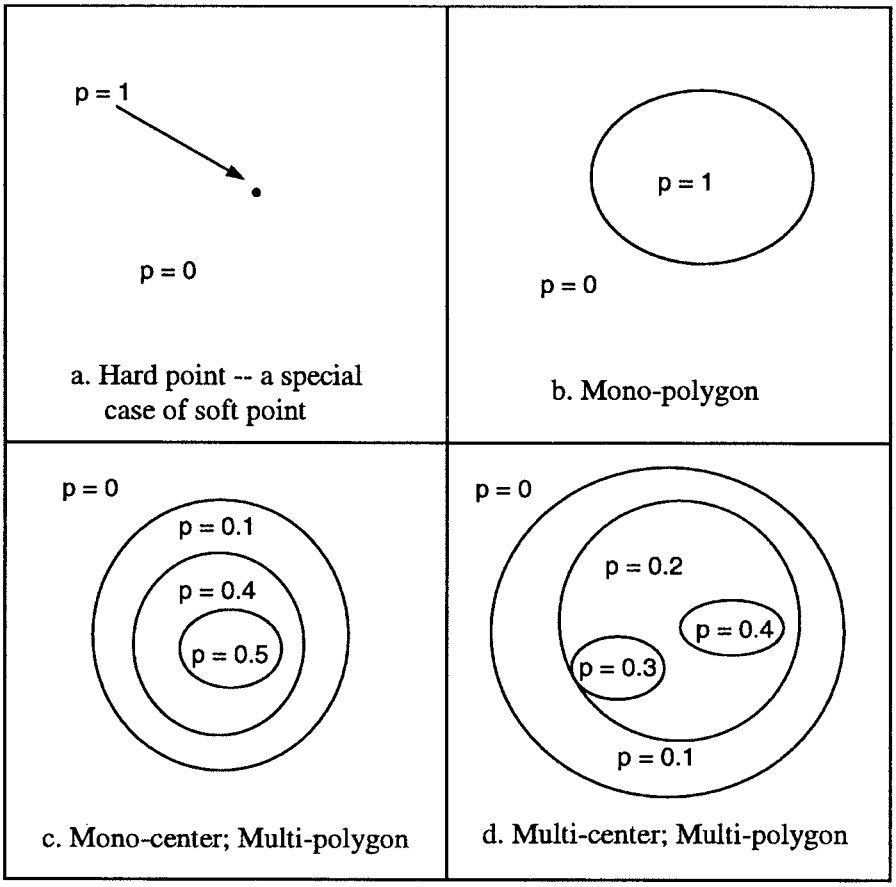


Figure 12.4. Probability distribution patterns of soft points.

age, each with an attribute value indicating the probability of any point in the probability polygon belonging to the soft polygon.

For soft polygons with known size and shape, the only uncertainty was location. The polygon's gravity center may be derived from its size and shape. The probability distribution of the center determines the probability distribution of the polygon. The soft point model discussed above also applies to the center of this soft polygon type. The size and shape of the polygon can be preserved

Table 12.1. Types of Soft Polygons, sort by size, then by shape, then by center location

Type	Size	Shape	Center Location	Note
1	uncertain	uncertain	uncertain	common
2	certain	uncertain	uncertain	uncommon
3	uncertain	certain	uncertain	rare
4	uncertain	uncertain	certain	uncommon
5	certain	certain	uncertain	common
6	certain	uncertain	certain	uncommon
7	uncertain	certain	certain	rare
8	certain	certain	certain	hard polygon, a special case of soft polygon

in a separate coverage or incorporated into the polygon coverage for probability distribution through an appropriate algorithm.

DATABASE IMPLEMENTATION AND MANAGEMENT

Initial database implementation included the following: (1) specifying a directory structure; (2) determining a directory and file-naming convention; (3) specifying read/write permissions and work group arrangements; (4) selecting precision and projection systems for coverages; (5) setting up a standard template for attribute files; and (6) establishing metadata standards.

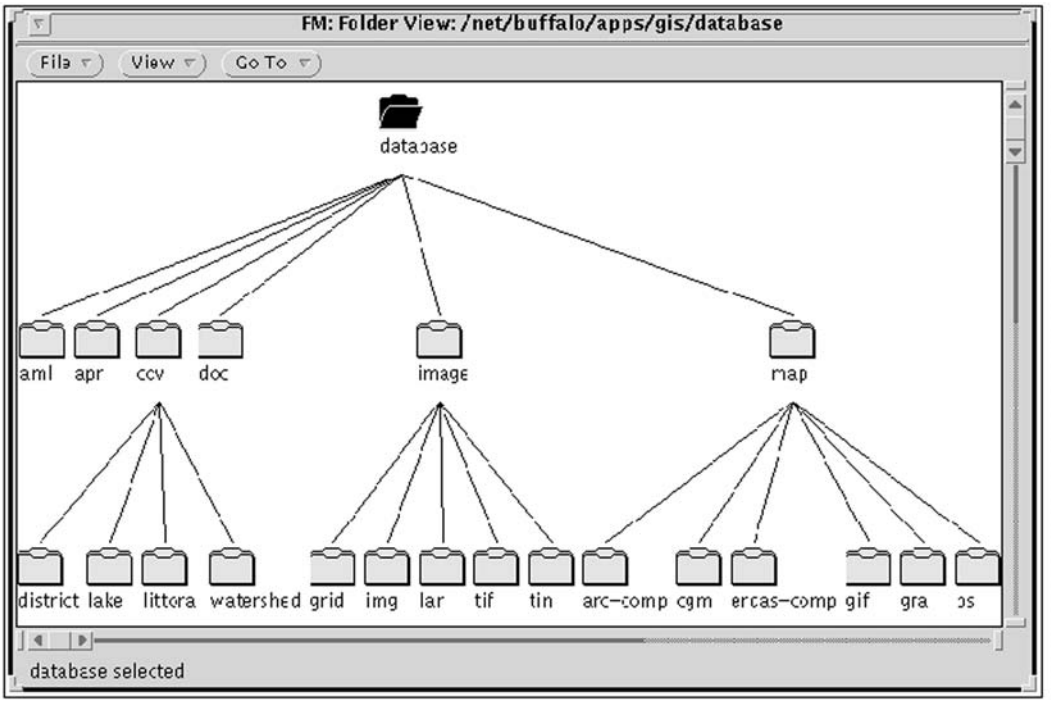


Figure 12.5. The directory structure and naming convention for LOES GIS Database.

The GIS database directory structure was set up on the virtual server. It includes subdirectories for ARC/INFO coverages, ArcView project files, images, map files, programming codes, and documentation files. A naming convention was developed to indicate the subject and format of each subdirectory and file (Figure 12.5).

Access permissions were assigned according to whether an individual was a database developer or user. Developers include database managers, system administrators, and data editors. Database managers have full read/write access to the database and are responsible for managing both the GIS database and the data server system. The system administrators, who also have full read/write access, solve system problems, maintain system security, and perform periodic backups by saving data on external media. Data editors have partial write access to input, update, and/or convert data for the database as well as utilizing metadata standards to document information about the database. End users are an integral part of GIS database management as they provide valuable feedback for database improvement. All end users have read-only access to the database.

Single precision (seven significant digits) and double precision (15 significant digits) are the

only alternatives for storing coverage coordinates in ARC/INFO. Double precision coverages, which provide a more precise geographic location, require more storage space. Due to the large study area (hundreds of square miles) and the variation of parameters within a short distance (inches or feet) in the Lake Okeechobee ecosystem, double precision was used for most coverages.

The projection system was based on the current GIS standard of the SFWMD, which is State Plane zone 3601 (Florida East Zone). The datum for the database was NAD27 (1927 North American Datum). When the SFWMD GIS database migrates to NAD83 (1983 North American Datum), all coverages and images in the Lake Okeechobee database will be transformed accordingly. The transformation will, most likely, use the ARC/INFO PROJECT function.

Minimal attribute data were stored with the coverages in the GIS database. Most ecological data reside in ORACLE tables and are linked to geographic features in the coverages by a unique ID. This ID often is the only external (user specified) attribute in the coverages. The user-specified attribute item (UNIQUE-ID) is indexed to decrease process time across the Arc-Oracle link.

In order to standardize metadata format for the GIS databases, METAMENU, a menu-driven metadata editing user interface was developed. METAMENU provides a convenient tool for users to document GIS data following a predefined standard. It automatically extracts any existing metadata information from GIS files, provides multiple choices whenever applicable, and prompts users to enter required information. The interface also checks for completion of user input, reorganizes entries into a standard metadata format, and saves information at a logical location using a standard naming convention. With this interface, users may define one metadata format for a group of similar coverages, or copy the metadata of one coverage into another and selectively edit some items to document differences.

One major difference between METAMENU and the DOCUMENT command in Arc/Info ver.7 is the metadata file format. DOCUMENT saves metadata in INFO, while the METAMENU interface saves metadata as an ASCII text file. ASCII files can be viewed without an Arc/Info license. Moreover, METAMENU generates metadata more specific to the LOES database users' needs, while DOCUMENT is more general and was designed for a broad range of ARC/INFO users. An example of metadata generated using METAMENU is in Appendix 12.1.

Database Management

The management of a GIS database, similar to that of other databases, involves data development, QA/QC, and backup. Data development includes data input, update, conversion, and construction of metadata. QA/QC, which stands for quality assurance and quality control, is a process following data development. Backup safeguards data from damage.

The unique aspect of GIS data management is the coordination between geographic and attribute data. Geographic data require special techniques and procedures for input, update, conversion, QA/QC, and backup; corresponding attribute data can be managed as regular tabular data. The geographic features and tabular attributes are linked together through a unique ID, which requires special attention when either side of the database is modified.

Data development, QA/QC, and backup of GIS data become more difficult to manage in a multiuser environment. A GIS data management guideline was a necessary tool which assists both database managers and users in systematically handling such transactions. The Lake Okeechobee database management utilized the GIS Data Management Guidelines. These guidelines detail the responsibilities of the database managers, the system administrators, the data editors, and the end users. They also specify the procedures for data QA/QC (SFWMD, 1997).

After data development was completed, data were verified by manual QA/QC procedures. Criteria for assessing data quality were specified for both geographic features and their attributes. Ac-

cording to Montgomery and Schuch (1993), graphic and attribute data "have similar data quality components, such as completeness, correctness, timeliness, and coverage, but the graphic features also are subject to cartographic quality considerations . . . (such as) relative accuracy, absolute accuracy, and graphic quality." The QA/QC procedures for Lake Okeechobee database incorporated these considerations in the data verification process.

An integral part of database management includes database backup. Backup saves database contents on external media as a safeguard against corruption. Three kinds of backup procedures were outlined in the management guidelines for the Lake Okeechobee database: (1) routine backups, which are consistently done at a designated time, regardless of changes in data; (2) nonemergency backups, which are done after changes in data; and (3) emergency backups, which take place preceding the possibility of a natural disaster (e.g., hurricane), power failure, maintenance on critical hard/software, and as otherwise deemed necessary.

DATA QUERY AND RETRIEVAL

The GIS database is accessible to all SFWMD staff for data query and retrieval. A generic user interface is being developed to support users with minimum GIS training to search, display, print or plot data from the database. Thus, the interface must be tailored to user needs.

The first step in interface development was to interview end users and identify their needs for data display, query, and retrieval (Montgomery and Schuch, 1993). Considering the internal structure of the database, user demand, and facilities available, ArcView was selected as the interface platform. Avenue, the object-oriented scripting language for ArcView, was used to communicate with the ORACLE database, customize the display environment, structure query statements, and standardize output formats. A prototype of the interface has been developed, presenting the interface look-and-feel. Development was based on user comments. The interface will be modified based on further input from users.

SUMMARY

This chapter documents the major steps in developing a Lake Okeechobee GIS-RDBMS database. The entire process involved the following components: (1) data server setup; (2) data model design; (3) database implementation; (4) data management standardization; and (5) user interface development. Database development focused on the Lake Okeechobee Ecosystem Study. The approaches adopted in database development were largely determined by hardware and software availability as well as special characteristics of ecological data. Introduced in the chapter is a method for modeling soft geographic features. The chapter also describes the approach for establishing a virtual data server through the network.

Database development is a constant trade-off between "what it should be" and "what it could be." Instead of a "perfect database," the final product of this project is a "functional database," developed under various constraints. Database optimization is a long-term task. By having a functional database first, and improving it constantly, the database will incrementally approach an ideal design that best serves its users.

ACKNOWLEDGMENT

This chapter incorporated review comments from Todd Tisdale, Al Steinman, Chris Carlson, Benjamin Lewis, and Nancy Lin. The metadata editing user interface used a set of AML scripts from a command-drive metadata input program originally developed by Timothy Liebermann. The

data query and retrieval graphic user interface was developed on an ArcView 2.0 data viewing project initially built by Chris Carlson. The authors wish to thank the above-mentioned colleagues—all are staff of the South Florida Water Management District—for their valuable contributions to this study.

REFERENCES

Aumen, N.G., 1995. The history of human impacts, lake management, and limnological research on Lake Okeechobee, Florida (USA). In *Advances in Limnology, Ecological Studies of the Littoral and Pelagic Systems of Lake Okeechobee, Florida*. N.G. Aumen and R.G. Wetzel, Eds., Schweizerbart, Stuttgart, Germany, Vol.45, pp.1–16.

Cowen, D.J., J.R. Jenson, P.J. Bresnahan, G.B. Ehler, D. Graves, X. Huang, C. Wiesner, and H.E. Mackey Jr., 1995. The design and implementation of an integrated geographic information system for environmental applications. *Photogrammetric Engineering and Remote Sensing*, 61(11):1393–1404.

Lostal, S.L., 1996. Software Development for Ecological Data System. Florida Atlantic University, Boca Raton, FL.

Montgomery, G.E., and H.C. Schuch, 1993. *GIS Data Conversion Handbook*, GIS World, Inc., Fort Collins, CO.

Rumbaugh, J. et al. 1991. *Object-Oriented Modeling and Design*, Prentice Hall, Englewood Cliffs, NJ.

SFWMD, 1997. The GIS Data Management Guidelines for Lake Okeechobee Database, unpublished.

University of Florida, 1991. Ecological Studies of the Littoral and Pelagic System of Lake Okeechobee. Annual report prepared for South Florida Water Management District, West Palm Beach, FL.

Appendix 12.1. An Example of Metadata Generated Using METAMENU

H
H SOUTH FLORIDA WATER MANAGEMENT DISTRICT
H 3301 Gun Club Road
H PO Box 24680
H West Palm Beach, FL, USA 33418–4680
H (800) 432–2045
H (561) 686–8800
H
H GEOGRAPHIC INFORMATION SYSTEMS METADATA
H
H
H DOCUMENTATION HISTORY
H ArcMeta Rev. 2.0, lmoore, 10/30/97.11:22:54.Thu.
H Metafile Name: lo_dairy97.met
H Cover: lo_dairy97, Workspace: /home/kos/lmoore/gis/dairy_covs
H Type: cover, Action: template
D
D ARC/INFO DESCRIPTION of cover geodataset LO_DAIRY97
D (Created by lmoore on 10/30/97.11:22:57.Thu)
D Workspace: /home/kos/lmoore/gis/dairy_covs
D Geodataset: lo_dairy97
D
D COORDINATE INFORMATION:
D Precision: DOUBLE
D Minimum X and Y values: 472335.2808281, 1028100.9375
D Maximum X and Y values: 620659.125, 1137946.719614
D Latitude/Longitude Minimum Bound: 27.1413, –81.1082
D Latitude/Longitude Maximum Bound: 27.4848, –80.605
D
D PROJECTION INFORMATION:
D Projection Name: STATEPLANE
D Units: FEET
D Zone: 3601
D Datum: NAD27
D Spheroid: CLARKE1866
D False Easting: 0
D False Northing: 0
D
D COVERAGE INFORMATION:
D Coverage Pathname: /HOME/KOS/LMOORE/GIS/DAIRY_COVS/LO_DAIRY97
D Number of Polygons: 3036
D Number of Arcs: 9877
D Number of Segments: 59870
D Number of Annotations: 0
D Number of Tics: 22
D Fuzzy Tolerance: 0.5
D Dangle Distance: 1
D
D POLYGON INFORMATION:
D Number of Polygons: 3036
D Bytes in PAT: 66
D Polygons Indexed: .FALSE.

Appendix 12.1. (Continued)

```
N
N     NARRATIVE DESCRIPTION of cover geodataset LO_DAIRY97
N          (Created by lmoore on 10/30/97.11:23:25.Thu)
N
N     GENERAL INFORMATION
N          Short Description: A polygon coverage of active dairies in the LO watershed and their associ-
           ated landuses.
N          Geographic Extent: Okeechobee county
N          Accuracy: same as /net/b50home1/proj/erd/gis/kos2/dairies/covs/dairy91
N          Abstract 1: Updated with current information from dairy landowners and Okeechobee Service
           Center personnel.
N          Abstract 2:
N          Abstract 3:
N          Abstract 4:
N          Intended Use 1: LOADSS model runs
N          Intended Use 2:
N          Intended Use 3:
N          Intended Use 4:
N          Limitations 1: same as /net/b50home1/proj/erd/gis/kos2/dairies/covs/dairy91
N          Limitations 2:
N          Limitations 3:
N          Limitations 4:
N
N     DISTRIBUTION INFORMATION
N          Availability: internal
N          Distribution Format: coverage through network
N          Distribution Size (Mb): 1157
N
N     CONTACT INFORMATION
N          Contact Person: Leslie J. Turner or Weihe Guan
N          Contact E-Mail Address: wguan@sfwmd.gov
N          Contact Phone Number: (561)687–6610 or (561)687–6687
N          Contact Address: OSRD, ERD, SFWMD
N
N     SOURCE INFORMATION
N          Source Description: same as /net/b50home1/proj/erd/gis/kos2/dairies/covs/dairy91
N          Source Media:
N          Source Scale:
N          Source Date: 1993
N          Source Contact Person: Joyce Zhang or Weihe Guan
N          Source E-Mail Address: jzhang @sfwmd.gov or wguan@sfwmd.gov
N          Source Phone Number: (561)687–6341 or (561)687–6687
N          Source Address: OSRD, ERD, SFWMD
N
N     ADDITIONAL INFORMATION
N          Related Information 1: None
N          Related Information 2:
N          Related Information 3:
N          Related Information 4:
N          Attribute Description 1:
N          Attribute Description 2:
N          Attribute Description 3:
N          Attribute Description 4:
```

Appendix 12.1. (Continued)

```
N           Processing History 1:
N           Processing History 2:
N           Processing History 3:
N           Processing History 4:
N           Revisions 1:
N           Revisions 2:
N           Revisions 3:
N           Revisions 4:
N           Misc. Notes 1:
N           Misc. Notes 2:
N           Misc. Notes 3:
N           Misc. Notes 4:
I
I     INFO ITEM DEFINITIONS for poly LO_DAIRY97
I     (Entered by lmoore on 10/30/97.11:23:31.Thu)
I
I     ITEM NAME: AREA
I           Item INFO Definition: 8,18,F,5
I
I     ITEM NAME: PERIMETER
I           Item INFO Definition: 8,18,F,5
I
I     ITEM NAME: LO_DAIRY97#
I           Item INFO Definition: 4,5,B,0
I
I     ITEM NAME: LO_DAIRY97-ID
I           Item INFO Definition: 4,5,B,0
I           Short Description:
I           Valid Codes:
I           Data Source:
I           Accuracy:
I           Related INFO Table:
I           Narrative 1:
I           Narrative 2:
I
I     ITEM NAME: TAG
I           Item INFO Definition: 4,4,C,0
I           Short Description: Landuse codes.
I           Valid Codes: IMPA, BARN, OTFL, HIA, MHP, DOL, OTP, SSA, SPFL, WSP, WET, CSF
I           Data Source: same as /net/b50home1/proj/erd/gis/kos2/dairies/covs/dairy91
I           Accuracy:
I           Related INFO Table:
I           Narrative 1:
I           Narrative 2:
I
I     ITEM NAME: DNAME
I           Item INFO Definition: 12,12,C,0
I           Short Description: Dairy names.
I           Valid Codes:
I           Data Source: same as /net/b50home1/proj/erd/gis/kos2/dairies/covs/dairy91
I           Accuracy: same as /net/b50home1/proj/erd/gis/kos2/dairies/covs/dairy91
I           Related INFO Table:
I           Narrative 1:
```

Appendix 12.1. (Continued)

```
|          Narrative 2:
|
|     ITEM NAME: LANDUSE
|          Item INFO Definition: 25,25,C,0
|          Short Description:
|          Valid Codes:
|          Data Source:
|          Accuracy: same as /net/b50home1/proj/erd/gis/kos2/dairies/covs/dairy91
|          Related INFO Table:
|          Narrative 1:
|          Narrative 2:
```

CHAPTER 13

Historical Aerial Photographs and a Geographic Information System (GIS) to Determine Effects of Long-Term Water Level Fluctuations on Wetlands along the St. Marys River, Michigan, USA

Donald C. Williams and John G. Lyon

INTRODUCTION

Water levels of the Laurentian Great Lakes and their connecting channels rise and fall in predictable annual patterns related to the seasonal patterns of precipitation, runoff, and evaporation. Variations of winter low levels to summer high levels average from 0.37 meters on Lake Superior to 0.6 meters on Lake Ontario (U.S. Army Corps of Engineers, 1985). Lake levels also rise and decline in unpredictable multiyear patterns following long-term fluctuations of climatic conditions in and around the Great Lakes basin. Historically, the ranges of Great Lake annual water level averages have been from ± 0.6 to 0.9 meters from their respective historic means (U.S. Army Corps of Engineers, 1985).

The Great Lakes long-term water level fluctuations have recurred about every two to three decades. These fluctuations have broad influences on the areas of the wetlands situated along the lake and connecting channel shores (Jaworski et al., 1979; Harris et al., 1981; Enslin and McIntosh, 1982; Lyon and Drobney, 1984; Busch and Lewis, 1984; Quinlan, 1985; Greene 1987; Bukata et al., 1988; Williams, 1995). The subject of this chapter is the use of a Geographic Information System (GIS) to study the influences of long-term water level fluctuations on the wetland areas of the St. Marys River. The St. Marys River is a connecting channel between Lakes Superior and Huron that borders on the U.S. and Canada. The river water levels are largely controlled by the levels of Lake Huron (Williams, 1995), such that wetland areas along the St. Marys undergo fluctuations corresponding to those of Great Lakes wetlands. A GIS was used to measure area changes in five different St. Marys River coastal wetland classes. These area change data were used to characterize wetland class responses to water level changes, to estimate response times, and to examine how wetland class transfers were affected by water levels.

Geographic Information Systems (GIS) offer excellent capabilities as tools for analysis of wetland changes caused by long-term lake level fluctuations. When accurate historical information on wetland areas and wetland characteristics is available in digital form, it would appear that GIS change detection algorithms could be employed to rapidly provide information on changes in wet-

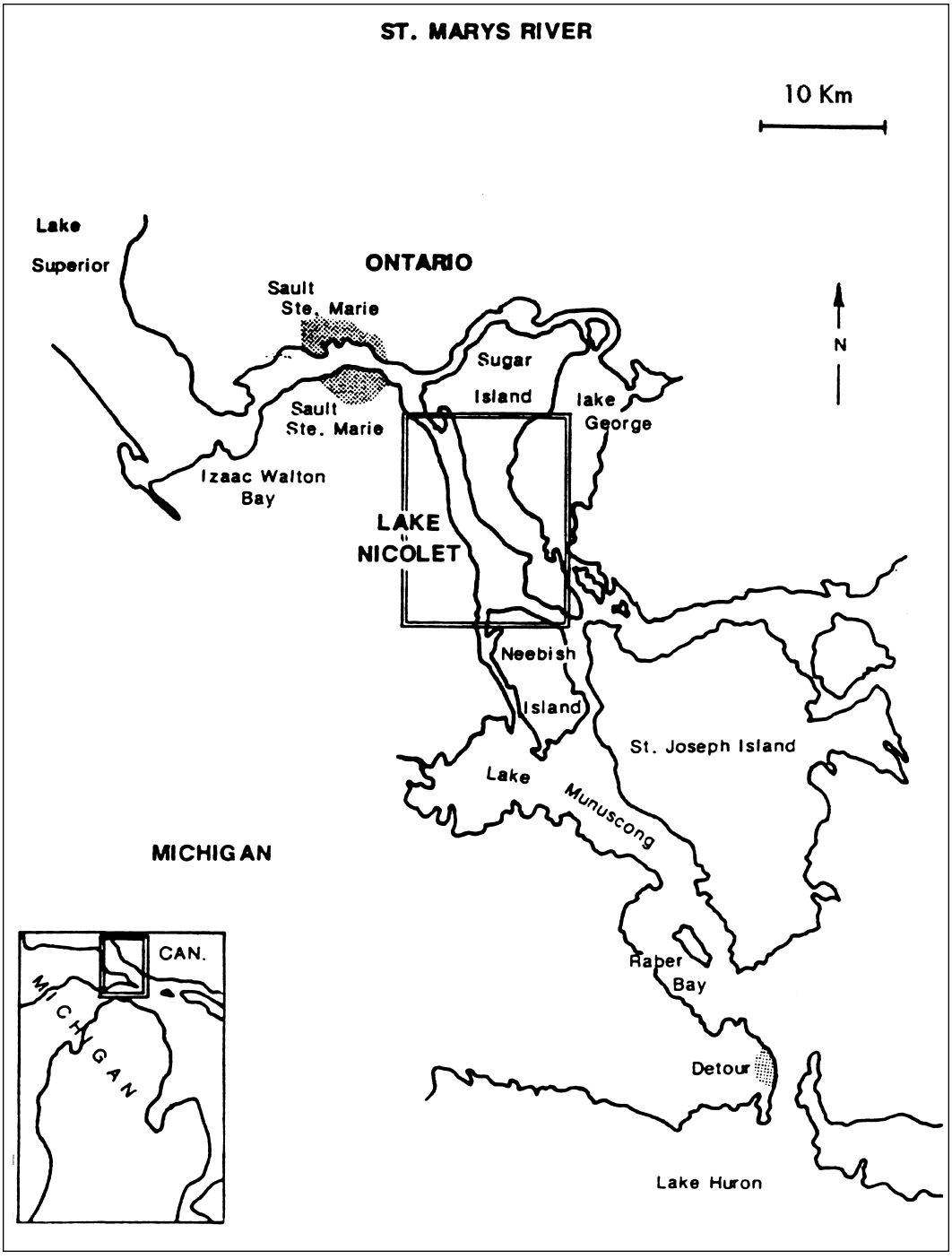


Figure 13.1. Location of St. Marys River study area.

land area and changes in wetland vegetation class associated with water level fluctuations. The following analysis demonstrates the application of such methods to analyze wetland changes due to long-term Great Lakes water level fluctuations. This analysis was conducted on a U.S. Fish and Wildlife Service (USFWS) National Wetland Inventory digital data set on the wetlands of the St. Marys River. This analysis used ERDAS GIS software. Other GIS systems containing similar analytic software would provide similar results.

The St. Marys River connects Whitefish Bay, Lake Superior, with Detour Passage in northern Lake Huron. The wetland study area was directly south of Sault Ste. Marie, Michigan, from Little Rapids Cut at the northern end of Sugar Island, south along both sides of Lake Nicolet, and along the east side of Neebish Island into northern Lake Munuscong (Figure 13.1). The study area is covered by five United States Geological Survey (USGS) 1:24,000 (7 1/2 minute) quadrangle maps.

The wetland types found in the study area are characterized as Unconsolidated Bottom, Emergent Wetland, Unconsolidated Shore, Scrub-Shrub Wetland, and Forested Wetland, using the USFWS Classification System (Cowardin et al., 1979). These wetlands classes are commonly described in nontechnical terms as river bottom, marsh, beach, scrub-shrub, and swamp.

METHODS

Wetland classes and areas in the St. Marys River were determined by interpretation of aerial photos from the summer seasons of 1939, 1953, 1964, 1978, 1982, 1984, and 1985 (Table 13.1). The film types included black and white, black and white infrared, color, and color infrared. The photo scales ranged from 1:12,000 to 1:58,000 (Table 13.1).

The aerial photos were acquired at seven different St. Marys River water levels (Table 13.1). These levels represented a broad sample of the wide range of water levels to which Great Lakes coastal wetlands are subjected. The August 1953 level, 177.52 m, was within 20 cm of the all-time monthly high of 177.64 m August ’52 at the U.S. Slip gauge. The July 1964 level, 176.09, was within 25 cm of the all-time low of 175.86 m (Dec. ’63). These aerial photos recorded the St. Marys wetlands at water levels that varied over a range of 1.43 meters (176.09 to 177.52 m, above the International Great Lakes Datum 1955). This range allowed an analysis of the influence of water levels on the wetlands over a large part of their historic range.

Wetland areas were quantified for each year of aerial photographs by the USFWS National Wetland Inventory contractor (St. Petersburg, FL) experienced with photo interpretation and identification of wetlands. The study area was also visited by the USFWS contractor. Initial interpretations were completed with an analog stereo plotter using the 1984 images. The plotter established

Table 13.1. Dates, Scales, Emulsions and Monthly Average Water Level Elevations of Aerial Photographs

Month	Year	Scale	Emulsion	Monthly Level (m)	Annual Level (m)
July	1939	1:20,000	black & white	177.01	176.62
August	1953	1:16,000	black & white Infrared	177.52	177.19
July	1964	1:16,000	black & white	176.09	176.16
June	1978	1:12,000	black & white	176.74	176.84
October	1982	1:58,500	color infrared	176.82	176.67
September	1984	1:12,000	color infrared	177.18	177.00
October	1985	1:24,000	natural color	177.43	177.20

geometric control and was used to correct any spatial inaccuracies found in the aerial photos. The interpreter viewed stereo pairs of photos, and outlined the wetland boundaries. The other dates of photography were compared to the 1984 maps and boundary adjustments made with a Zoom Transfer Scope. The boundaries were plotted on USGS 7½ minute maps of the study area.

Wetlands up to 800 meters inland from the river shore were included in the analysis to incorporate wetlands that could be influenced by changing water levels. This determination was based on USGS map elevations and mapped marsh areas. The photo interpreter assigned each wetland area to a type or class following the USFWS Classification System (Cowardin et al., 1979; U.S. Fish and Wildlife Service, 1987). Data were summarized in wetland maps and tables of wetland areas reported by the USFWS Wetland Analytical Mapping System (WAMS). The USFWS used its GIS system to create GIS files of the wetland polygons in ELAS format, which could be read by ERDAS software (U.S. Fish and Wildlife Service, 1988; Williams and Lyon, 1991).

The National Wetland Inventory provided maps and WAMS-derived tabular summaries. These were appropriate for completing some analyses. Further GIS analysis using the digital files, however, offered several advantages. First, the numerous wetland classes identified by National Wetland Inventory could readily be aggregated to simplify the analyses and to employ wetland classes used in comparable wetland studies on the Great Lakes. This was effective because of the large number of wetlands found on five USGS quadrangles and the aggregation was accomplished rapidly using the ERDAS software program RECODE. Other data sources on the St. Marys wetlands indicated that the Unconsolidated Bottom class included larger areas with submergent vegetation cover than the USFWS definition of Unconsolidated Bottom (less that 30%). Much of the bottom of the St. Marys is vegetated (Liston et al., 1986), hence the Unconsolidated Bottom class included Aquatic Bed Wetland (Cowardin et al., 1979).

Because the tabular National Wetland Inventory summaries were completed on individual quadrangles, without a GIS it was necessary to sum changes in five separate quadrangles to determine change across the entire study area. Consequently, the GIS was used to combine the quadrangle areas into a single digital file to reduce the number of analyses. Six digital files (one quadrangle was divided into two digital files) from each year were combined into one mosaic. This provided seven mosaics of the study area corresponding to the seven years of historical aerial photographic coverage: 1939, 1953, 1964, 1978, 1982, 1984, and 1985. The mosaics were created using the ERDAS GIS software program SUBSET.

There were also slight discrepancies in the total study area measured in both the tabular data and the individual digital files. Small areas were included in some years that were not in others. The GIS was used to establish an exact common area for each year so that year to year changes in wetland class and area could be determined more accurately. A GIS modeling package (GISMO) was used to create a mask, a sum of areas that were not common to each of the seven files, that was applied to each mosaic so the study area was precisely the same for each of the seven years. Data on the areal extent of each wetland class for each year were then obtained from the corrected mosaics (ERDAS BSTATS). This corrected discrepancies in total area between the seven data sets of up to 1%. Discrepancies in the ranges of the five wetland classes were reduced by as much as 3.8%. Rather than National Wetland Inventory tabular data, these quantities were then used in further analysis (Table 13.2).

The GIS analysis also allowed for the spatial location of changes in wetlands. Analysis of tabular or other data did not allow such identification. Locating exactly where a change occurred was important to identifying other possible causative factors for area changes such as vessel traffic in adjacent commercial navigation channels.

Locational information can also be used with GIS cross-tabulation algorithms to examine the nature of wetland class changes. A transition matrix for the St. Marys River study is shown as

Table 13.2. Quantities of Wetland Classes Calculated by GIS from Mosaics (Hectares). Note: Mask included Upland

Year	Unconsolidated Bottom	Emergent Wetland	Unconsolidated Shore	Scrub Shrub	Forested Wetland
1939	5097.4	1248.7	348.7	352.5	178.7
1953	5586.7	924.7	3.1	321.1	272.3
1964	5100.5	1326.5	99.2	384.3	250.3
1978	5611.2	899.8	9.2	341.4	281.2
1982	5554.5	950.5	7.9	333.1	283.2
1984	5601.1	909.3	1.6	339.9	281.7
1985	5600.8	913.5	1.5	341.7	282.2

Table 13.4. The matrix shows the net area changes that took place between wetland classes from one wetland state to the next. For example, the upper left cell in Table 13.4 indicates that 174.6 ha of Unconsolidated Bottom were gained in 1953 that had been Emergent Wetland in 1939. This matrix was used to examine some of the wetland class dynamics caused by fluctuations in the levels of the St. Marys River.

Class by Class Analysis of Area Changes Due to Long-Term Water Level Changes

The class areas calculated from the six-quadrangle mosaics of the study areas were used for analysis of the effects of water level fluctuations on the wetland class areas. Regressions of class areas on water level were run for the wetland classes Unconsolidated Bottom, Emergent Wetland, Unconsolidated Shore, Scrub-Shrub Wetland, and Forested Wetland. Initially, monthly mean water level from the U.S. Slip gauge was used as the independent variable. Because wetlands do not respond immediately to water level change, and area changes due to water level fluctuations may take months or years to be completed, another set of regressions was run using mean annual water level. Variations in the independent variable, average annual water level, for 1925 through 1985 from the U.S. Slip gauge below the locks at Sault Ste. Marie are shown in Figure 13.2.

Accounting for Lag Times in Wetland Area Changes

Many observations on Great Lakes coastal wetlands (e.g., Jaworski et al., 1979; Busch and Lewis, 1984; Painter et al., 1988) have shown that wetland changes due to the rise or fall of water level may take several years to be completed. To gain some insight into how rapidly wetland areas might have adjusted to water levels, a series of regressions was run using moving average water levels instead of monthly or annual average levels. These regressions were weighted average water levels consisting of the annual water levels of the years of the aerial photographs and annual levels from years prior to the photographs. The levels of the year the wetlands were photographed was given the greatest weight and the preceding years less and less weight. For a two year average the weights were 0.2, 0.1; for a three year average 0.3, 0.2, 0.1, etc.

The standard linear regression equation used for annual mean water levels was:

$$Y = a + b\ (X_{ave}) \tag{1}$$

where Y = wetland area and X_{ave} = average annual water level.

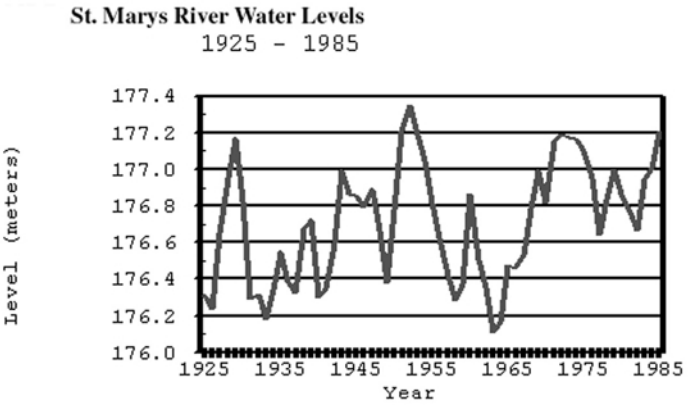


Figure 13.2. Mean annual water levels from the U.S. Slip gauge near the study area (meters above ILGD '55).

For weighted average water levels the regression equation was

$$A = a + b\ (c_1{*}X_t{+}c_2{*}X_t{-}1{+}...{+}c_n{*}X_{t\text{-}n}) \quad (2)$$

where c_1, c_2...c_n were weighting factors for the average annual water levels of the year of photography, the year before the photography, etc., and the weighting factors were 0.2 and 0.1 for a two year regression, 0.3, 0.2, and 0.1 for a three year regression, etc.

The coefficient of determination (R^2) of the regression equation was maximized by adding yearly increments into the linear combination of levels. An optimal combination of water levels was selected for each wetland class based on the maximized R^2. This optimum fit provided an estimate of the number of years that might have been relevant to the adjustment of each wetland class in the St. Marys River to the water level changes, and therefore the importance of the water levels in the years preceding the measured growing season.

RESULTS

Emergent Wetlands

For Emergent Wetland within the study area there was a 426 ha difference between the year with the largest area, 1964, and the year with the least area, 1978. This represented a 32% change, hence Emergent Wetland area varied by about one-third (Table 13.3). The linear regression of Emergent Wetland area on monthly average water level was not significant. The regression of area on average annual level was significant (* $P < 0.05$). This provided strong evidence of a relationship between water level and area of Emergent Wetland areas between 1939 and 1985. As water levels rose, Emergent Wetland area decreased, and as water levels fell, Emergent Wetland area increased. A number of other studies on Great Lakes emergent wetlands have shown this result (Jaworski et al., 1979; Harris et al., 1981; Enslin and McIntosh, 1982; Lyon and Drobney, 1984; Busch and Lewis, 1984; Greene, 1987; Bukata et al., 1988; Payne et al., 1985; Quinlan, 1985). The fact that the regression of area on monthly average level was not significant and the regression of area on annual average level was significant (** $P < 0.05$, $R^2 = 0.64$) suggested that the effect of water level on Emergent Wetland area took a growing season or longer to complete.

Table 13.3. Direction, Rate, and Magnitude of Wetland Class Changes

Wetland Class	Regression Results			Area Change	
	±	P <	Yr. Lag	Hectares	Percent
Emergent	–	0.001	14	426	32
Unconsolidated Bottom	+	0.001	15	514	9
Unconsolidated Shore	–	0.05	22	347	99
Scrub-Shrub	–	0.01	5	63	16
Forested	+	0.01	22	104	37

There is a tendency for the water levels to be higher in the years of photography taken later in the growing season (Table 13.1). This cannot account for the differences in wetland area, however, because the dominant St. Marys emergent plants reach their maximum heights by July, and plants of maximum height are still present in September and generally October (Liston et al., 1986). *Scirpus acutus* Bigelow shoots persist through the winter and are clearly visible protruding through the ice.

Regressions were run for weighted annual water level averages incorporating water levels from 2 to 30 years previous to the years of the area measurement. The regression fit (R^2) increased for each additional year accumulated through the fourteenth year. The analysis thus suggested that water levels in any given year influenced Emergent Wetland area for up to 14 years, or conversely, that Emergent Wetland area was influenced by water levels of multiple years before its occurrence. At 14 years, the regression accounted for 99% of the variability ($R^2 = 0.991$), and was statistically significant (*** $P < 0.001$).

The question of response time for changes in Emergent Wetland has had less attention. Several investigators have postulated that it takes around five years for Emergent Wetlands to adjust to water level changes. Busch and Lewis (1984) used a five-year weighted moving average to explain changes in a marsh in eastern Lake Ontario marsh, and Painter et al. (1988) used a similar method to explain changes in Cootes Paradise on the western end of the same lake. The increase in R^2s for regression using moving averages up to 14 years suggest that the Emergent Wetlands in the St. Marys River might take even longer to adjust to water levels. There is some rationale for supporting the hypothesis that these wetlands may on average take five years or more to complete adjustments to major water level fluctuations. It was clear from the seven sets of aerial photographs that distribution patterns of *Scirpus acutus* Bigelow were very stable. Some areas appeared not to have changed significantly from 1939 to 1985. Although they are not woody plants, many of the emergents in the St. Marys (e.g., *Eleocharis smallii* Britton, *Sparganium eurycarpum* Engelm., *S. chlorocarpum* Rydb., and *Scirpus americanus* Pers.) are long-lived perennials. Storage effects from low levels can occur as gains in colonized areas and increases in robustness and storage products in rhizomes. Substrate (Lyon et al. 1986) and topographic changes brought about by the water level changes may also take some time to complete.

The regressions took into account both gains and losses, and it is common sense to assume that losses of Emergent Wetland could occur quite rapidly in response to high water levels, certainly less than 14 years. Nonetheless, observations of cattail (*Typha* sp.) responses to water level increases at Horicon Marsh in the state of Wisconsin, USA by Mathiak (1971) indicated that maximum die-off occurred in the fourth and fifth years after exposure to high water levels. Given that some perennial marsh species are resistant to flooding, and that the resistance is likely to vary from species to species, it is probable that responses to increased levels will vary from marsh to

marsh depending upon species composition. There appears to be some evidence for this variation (Williams, 1995).

The transition matrix shows some of the interclass transfers that took place between Emergent Wetland and other classes as levels changed. There were large changes from Unconsolidated Bottom to Emergent Wetland associated with the drops in water level from 1953 to 1964 and 1978 to 1982. There were also large transfers from Emergent Wetland to Unconsolidated Bottom during the level increase from 1939 to 1953 and 1982 to 1984. There were fairly large losses of Emergent Wetland in the 1939 to 1953 transitions to Scrub-Shrub, Forested Wetland, and Upland. These would not be expected from water level changes and were probably due to succession that occurred in spite of the level increases.

Unconsolidated Bottom

There was a 514 hectare change in area of Unconsolidated Bottom between the year with the largest area, 1978, and the year with the smallest area, 1964. This amounted to a 9% change in area (Table 13.3). Unconsolidated Bottom area responded in a manner opposite to that of Emergent Wetland. Unconsolidated Bottom expanded during high water levels and contracted during low water levels; as water levels increased, Unconsolidated Bottom area increased. The regression of Unconsolidated Bottom area on monthly water levels was not statistically significant. The regression of area on annual average level was significant (** $P < 0.05$, $R^2 = 0.55$) and suggested that the effect of water level on Unconsolidated Bottom area also took a growing season or more to complete.

A series of regressions was run for weighted average water levels for 2 to 30 years previous to the years of area measurement. The regression R^2 increased for each additional year up to 15 years. The analysis thus suggested that it takes several years before the effects of water levels on Unconsolidated Bottom are complete. At 15 years, the regression (Table 13.3) accounted for 99% of the variability ($R^2 = 0.995$), and was statistically significant (*** $P < 0.001$). The similarity in lag time (15 vs. 14 years) and the opposite response suggested a reciprocal connection between Unconsolidated Bottom and Emergent Wetlands. As water levels increased, many of the losses in Emergent Wetland area were probably attributable to gains in Unconsolidated Bottom and as water levels decreased, the opposite exchanges probably occurred. This was supported by the information in the transition matrix in Table 13.4 discussed above.

Unconsolidated Shore

Unconsolidated Shore is by definition wetland with less than 30% coverage by vegetation and a hydrologic regime varying from intermittently flooded to intermittently exposed. The area of Unconsolidated Shore changed by 347 ha between its minimum and maximum. It varied by 99% of its maximum value (Table 13.3). There was a general decreasing trend until it nearly disappeared in 1984 and 1985. The data suggest that there was gradual colonization of areas that were nearly bare in 1939, possibly because of some disturbance. Regressions of area of Unconsolidated Shore on both monthly and mean annual water level were not significant.

As the regressions on weighted annual mean levels included more and more years prior to the dates of the photographs, the regressions became significant. The results suggested that water levels as early as 22 years prior to the date of the wetland area measurements may have had an influence on the amount of Unconsolidated Shore wetland. The R^2 of the regression reached a maximum of 0.75 with a weighted average that included 22 years of average annual levels prior to

Table 13.4. Transition Matrix of Wetland Classes Transfers Developed from GIS Cross Tabulation[a]

	1939>1953	1953>1964	1964>1978	1978>1982	1982>1984	1984>1985
			Changes in Unconsolidated Bottom			
EM	174.6	–355.6	393.1	–60.5	39.6	–1.8
US	287.8	–77.7	85.5	2.3	4.9	–0.3
SS	18.5	–28.3	24.9	3.3	1.8	–0.0
FO	0.4	2.1	–0.9	1.0	0.1	–0.1
UP	7.7	–26.5	7.9	–2.8	0.2	2.0
			Changes in Emergent Wetland			
UB	–174.6	355.6	–393.1	60.5	–39.6	1.8
US	24.5	–10.9	2.3	0.1	0.5	–0.0
SS	–64.4	–2.4	–13.9	–6.7	–3.5	0.1
FO	–48.9	29.1	–10.5	2.1	0.5	0.3
UP	–60.4	30.2	–11.2	–5.4	0.9	2.0
			Changes in Unconsolidated Shore			
UB	–287.8	77.7	–85.5	–2.3	–4.9	0.3
EM	–24.5	10.9	–2.3	–0.1	–0.5	0.0
SS	–11.7	1.5	1.1	0.1	0.5	–0.8
FO	–1.2	0.1	–0.0	–0.0	–0.1	0.0
UP	–20.3	5.9	–3.2	0.9	–1.2	0.4
			Changes in Scrub-Shrub Wetland			
UB	–18.5	28.3	–24.9	–3.3	–1.8	0.0
EM	64.4	2.4	13.9	6.7	3.5	–0.1
US	11.7	–1.5	–1.1	–0.1	–0.5	0.8
FO	–59.3	3.8	–11.2	–4.7	3.8	0.1
UP	–29.8	30.2	–19.6	–6.9	1.7	1.0
			Changes in Forested Wetland			
UB	–0.4	–2.1	0.9	–1.0	–0.1	0.1
EM	48.9	–29.1	10.5	–2.1	–0.5	–0.3
US	1.2	–0.1	0.0	0.0	0.1	0.0
SS	59.3	–3.8	11.2	4.7	–3.8	–0.1
UP	–15.3	13.1	8.3	0.3	3.0	0.7

[a] Changes in Hectares

the area measurements (* $P < 0.05$). Like Emergent Wetland and Scrub-Shrub wetland, Unconsolidated Shore decreased with increasing water levels and increased with decreasing water levels.

The transition matrix (Table 13.4) shows that 288 ha of the dramatic loss in area in 1939 went to Unconsolidated Bottom in 1953. This occurred simply because the water levels went up; no vegetation change was necessary. There was also a large loss to Unconsolidated Bottom with the level increase from 1964 to 1978. Table 13.4 also shows significant losses of Unconsolidated Shore to Emergent, Scrub-Shrub, and Upland in the 1939 to 1953 transition. These suggested changes due to succession rather than water level.

Because the largest interchanges were between Unconsolidated Shore and Unconsolidated Bottom, which occurred instantaneously with water level changes, the long lag period shown in the regression analysis does not appear reasonable. The regression correctly shows a reciprocal relationship between area and water level, but succession evidently played a significant role and complicated the water level—area relationship.

Shrub-Scrub Wetland

Scrub-Shrub Wetland area varied by 63 ha, about 16% of its maximum area (Table 13.3). The changes in Scrub-Shrub Wetland were clearly influenced by water level as the regressions of both monthly and annual average water levels on area of Scrub-Shrub were significant (* $P < 0.05$). Scrub-Shrub wetland area, like Emergent and Unconsolidated Shore areas, decreased during higher water levels and increased during lower water levels.

Changes in Scrub-Shrub Wetland area appeared to develop rapidly in response to water level changes. The R^2 reached a maximum with the regression that included only five years of weighted average water levels ($R^2 = 0.86$). The relationship was statistically significant (** $P < 0.01$; Table 13.3). This suggested that it took several years for the St. Marys River Scrub-Shrub Wetland vegetation to complete its expansions and contractions in response to water level changes, possibly not as long as Unconsolidated Bottom or Emergent Wetland.

The transition matrix shows that there were losses to Unconsolidated Bottom during the 1939 to 1953, and 1964 to 1978 water level increases. There were also changes attributable to succession: the 59.3 ha loss to Forested Wetland, the 29.8 ha loss to Upland, and the 64.4 ha gain from Emergent Wetland. The same pattern of losses to Forested Wetland and Upland and gains from Emergent Wetland is found in the 1964 to 1978 transition. Successional processes apparently have a significant effect in the Scrub-Shrub changes.

Forested Wetland

Forested Wetland varied by 104 ha from maximum to minimum area, a change of 37%. The regressions of area on monthly and annual average water levels were not significant.

The R^2s of the regressions using Forested Wetland area on weighted average water levels increased (to 0.64) as yearly average levels were added through 22 years prior to the date of the area measurements. With this lag, the regression was significant (** $P < 0.01$). This suggested that water levels in any given year appeared to have had an effect on area of Forested Wetlands for as many as 22 years after they occurred. The regression was positive, i.e., Forested Wetlands appeared to increase during periods of high water levels.

The latter outcome was not expected because woody vegetation is generally thought to be less tolerant to flooding than herbaceous vegetation (Keddy and Reznicek, 1985). Wetland trees are more tolerant than Upland trees, however. An increase in water levels can indeed result in an increase in the quantity of Forested Wetlands. While standing water or saturated soil will often kill upland trees, many of the facultative wetland plants will, according to their classification, grow in the presence or absence of standing water or saturated soils. Higher water over time will influence the presence of facultative wetland trees, and allow their increased growth as compared to facultative upland trees. The species found near the St. Marys River are typically facultative wetland species. Moreover there is evidence for this from the 8.3 ha change from Upland to Forested Wetland in the 1964 to 1978 transition. It is notable that Forested Wetland areas remained consistent despite higher water levels in 1984 and 1985.

The gains during low to high transitions can also in part be explained by succession. Table 13.4 indicates that the largest gain in the 1939 to 1953 transition is from Scrub-Shrub (59.3 ha), followed by Emergent Wetland (48.9). The same pattern of gains is found in the 1964 to 1978 “high to low” transition; gains are mostly from Scrub-Shrub (11.2 ha) and Emergent Wetland (10.5 ha).

The Forested Wetland changes from 1953 to 1964 were puzzling, however. Indeed, 13.1 ha of 1953 Forested Wetland area came from Upland, and 29.1 ha were lost to Emergent Wetland, the opposite of succession. An explanation may be found in the timing of the large water level

increase in that it was affecting the 1953 wetlands. Figure 13.2 shows that St. Marys River levels were in the third year of a high level spike that lasted four years. That spike could have caused some areas of Upland to change to Forested Wetland and also caused the ultimate loss of some Forested Wetland area to Emergent Wetland and Scrub-Shrub, but not so rapidly as to be visible in the 1953 photographs. The loss may have been either not evident , i.e., the trees were dead but still standing, or the loss may have occurred in the next few years. The successional changes did not show up until after 1953, however.

The regression data for Forested Wetland is at best ambivalent. The transition matrix indicates that Forested Wetland increases occur during transitions from low to high levels as successional processes, i.e., conversion of Emergent Wetland and Scrub-Shrub to Forested Wetland. These are not driven by high water levels, hence the association of larger Forested Wetland areas with higher levels shown in the regression is probably an artifact of the timing of successional processes. Large areas of Emergent Wetland and Scrub-Shrub Wetland were evidently caused by the long period of low water in the 1930s (Figure 13.2). These in part "succeeded" to Forested Wetland despite the high levels in the early 1950s.

DISCUSSION

Because a GIS is a convenient tool for manipulation of mapped data, it simplified development of data sets that could be used in data analysis. Of particular interest in this context was the use of a transition matrix derived from a GIS cross tabulation. The matrix showed interclass transfers from wetland state to wetland state. In this study, the transitions matrix was particularly useful in helping to identify successional effects and distinguish them from water level effects.

If the R-Squares of the regressions were taken as a measure of how much variability in the data is due to the regression, the R-Squares then provided a measure of how much of the variability in each wetland class area was due to changes in water levels. This suggested that virtually most of the variability ($R^2 = 0.99$) in the areas of both Unconsolidated Bottom and Emergent Wetland was due to water level fluctuations. A significant part of the variability in Scrub-Shrub Wetland was also due to level fluctuations in the St. Marys ($R^2 = 0.86$). By statistical measures, there were apparently other important factors affecting Unconsolidated Shore ($R^2 = 0.74$) and Forested Wetland ($R^2 = 0.78$).

The transition matrix was useful in identifying strong evidence of successional trends in these wetland classes by allowing quantification of interclass transfers. The matrix also provided a basis for suggesting that the association found in regression between high water levels and increased Forested Wetland was probably anomalous.

There are several factors that could not be accounted for in this study that may have affected conclusions concerning the wetland class area changes. Accuracy of photo interpretation by the USFWS contractor may have affected data accuracy. There may also have been human influences on area changes. Some of these shores may at one time have been farmed; Forested Wetlands may have regrown in areas previously cut down for agriculture. Filling in some areas was visible from changes in the photographs between 1953 and 1978, causing losses of wetland not due to level fluctuations.

CONCLUSION

Use of GIS and Historical Aerial Photography

Using a GIS for a wetland change detection study proved to have several advantages over manual methods for use of historical photographs, such as mapping and measuring areas with a plane

planimeter and use of overlays. Some of the data in this study could have been derived from the tabular data supplied by the National Wetland Inventory. The tabular data was in effect a National Wetland Inventory GIS product however, and would have been produced from the authors' GIS, had the mapping and digitization not been completed by the National Wetland Inventory.

GIS use allowed creation of mosaics of all study USGS quadrangles, greatly reducing the number of necessary analytic operations. It provided a means of quickly establishing an accurate common area of study for each year of wetland mapping. It provided a means of rapidly aggregating wetland types to convenient wetland classes. It provided an efficient method of measuring historical wetland class changes based on archived aerial photographs. The change measurements could then be used in conjunction with the historical water level data to understand how water level changes affected wetland class changes between 1939 and 1985.

Finally, GIS use allowed rapid development of locational information. Digital maps of areas in which Emergent Wetlands were lost provided information on potential losses besides water levels, such as the proximity of commercial navigation channels. Cross tabulation provided the capability to develop a transition matrix that allowed quantification of class to class changes between wetland states in 1939, 1953, 1964, 1978, 1984, and 1985. These data were particularly useful in confirming expected change patterns and in explaining changes that were not intuitive. Hence this study indicates that GIS and historical aerial photography can be used effectively in tandem to study how water level changes affect wetland class areas. These properties of Great Lakes wetlands are important to know, not only because they provide a better understanding of wetland processes, but also in view of the prospect of Great Lakes water level regulation changes, or the adoption of more extensive regulation of the levels of the Great Lakes and connecting channels.

REFERENCES

Bukata, R.P., J.E. Bruton, J.H. Jerome, and W.S. Haras, 1988. A Mathematical Description of the Effects of Prolonged Water Level Fluctuations on the Areal Extent of Marshlands. Scientific Series No. 166, Inland Waters Directorate, National Water Research Institute, Canada Center for Inland Waters, Burlington, Ontario.

Busch, W.D., and C.N. Lewis, 1984. Responses of Wetland Vegetation to Water Level Variations in Lake Ontario. *Proceedings of the Third Annual Conference on Lake and Reservoir Management,* North American Lake Management Society, pp. 519–523. U.S. Environmental Protection Agency, Washington, D.C.

Cowardin, L.M., V. Carter, F.C. Golet, and E.T. LaRoe, 1979. Classification of Wetlands and Deepwater Habitats of the United States. U.S. Department of the Interior, Fish and Wildlife Service, Report No. FWS/OBS–79/31. Washington, D.C.

Enslin, B., and D. McIntosh, 1982. Changes in Aquatic Vegetation in Qanicassee, Nayanquing Point and Wildfowl Bay. Center for Remote Sensing, Michigan State University. Prepared for the East Central Michigan Planning and Development Region.

Greene, R.G. 1987. Effects of Lake Erie Water Levels on Wetlands Areas Measured from Aerial Photographs: Pointe Mouillee, Michigan. M.S. Thesis. Department of Civil Engineering, Ohio State University, Columbus, OH.

Harris, H., G. Fewless, M. Milligan, and W. Johnson, 1981. Recovery processes and habitat quality in a freshwater coastal marsh following a natural disturbance. In *Selected Proc. Midwest Conf. on Wetland Values and Man*, pp. 363–379. Freshwater Society, St. Paul MN.

Jaworski, E., C.N. Raphael, P.J. Mansfield, and B.B. Williamson, 1979. Impact of Great Lakes Water Level Fluctuations on Coastal Wetlands. Office of Research Development, Institute for Water Research, Michigan State University, East Lansing, MI.

Keddy, P.A., and A.A. Reznicek, 1985. Great Lakes vegetation dynamics: The role of fluctuations water level and buried seeds. *J. of Great Lakes Res.* 12:25–36.

Liston, C.R., and C.D. McNabb, (principal investigators) with D. Brazo, J. Bohr, J. Craig, W. Duffy, G. Fleischer, G. Knoecklein, F. Koehler, R. Ligman, R. O'Neal, M. Siami, and P. Roettger, 1986. Limnological and Fisheries Studies of the St. Marys River, Michigan, in Relation to Proposed Extension of the Navigation Season, 1982 and 1983. Submitted to the Detroit District, Corps of Engineers, Detroit, Michigan and U.S. Fish and Wildlife Service, Twin Cities, MN. Issued as U.S. Fish and Wildlife Service, Office of Biological Services Report OBS/85(2).

Lyon, J.G., and R.D. Drobney, 1984. Lake level effect as measured from aerial photos. *ASCE Journal of Surveying Engineering*, 110: 103–111.

Lyon, J.G., R.D. Drobney, and C.E. Olson, Jr., 1986. Effects of Lake Michigan water levels on wetland soil chemistry and distribution of plants in the Straits of Mackinac. *J. Great Lakes Res*., 12: 175–183.

Mathiak, H.A., 1971. Observations on Changes in the Status of Cattails at Horicon Marsh, Wisconsin. Research Report 66, Department of Natural Resources, Madison, WI.

Painter, D.S., K.J. McCabe, and W.L. Simser, 1988. Past and present limnological conditions in Cootes Paradise affecting aquatic vegetation. National Water Research Institute, Environment Canada, Burlington, Ontario. NWRI Contribution No. 88–47, 30 pp. plus figures.

Payne, F.C., J.L. Schuette, J.E. Schaeffer, J.B. Lisiecki, D.P. Regalbuto, and P.S. Rogers, 1985. Evaluation of Marsh Losses: Maisou Island Complex. Prepared for Wildlife Division, Michigan Department of Natural Resources.

Quinlan, C.M., 1985. The Effects of Lake Level Fluctuations on Three Lake Ontario Shoreline Marshes. M.A. Thesis, University of Waterloo, Ontario, Canada.

U.S. Army Corps of Engineers, Detroit District, 1985. Great Lakes Water Level Facts.

U.S. Fish and Wildlife Service, 1987. Photointerpretation Conventions for the National Wetlands Inventory. May 1, 1987.

U.S. Fish and Wildlife Service, 1988. St. Marys River, Michigan, Geographic Information Systems Applications. Final Report, January 19, 1988.

Williams, D.C., 1995. "Dynamics of Area Changes in Great Lakes Coastal Wetlands Influenced by Long Term Fluctuations in Water Levels. Ph.D. Dissertation, School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI.

Williams, D.C., and J.G. Lyon, 1991. Use of a geographic information system data base to measure and evaluate wetland changes in the St. Marys River, Michigan. *Hydrobiologia,* 219: 83–95.

CHAPTER 14

Watershed-Based Evaluation of Salmon Habitat

Ross S. Lunetta, Brian L. Cosentino, David R. Montgomery,
Eric M. Beamer, and Timothy J. Beechie

INTRODUCTION

Dramatic declines in Pacific Northwest (PNW) salmon stocks have been associated with land-use induced freshwater habitat losses (Nehlsen et al., 1991). Substantial resources are being directed toward the restoration of stream habitats in efforts to maintain and/or restore wild salmon stocks in the PNW, yet there is no common scientific framework for guiding the prioritization of where and how salmon habitat preservation and restoration activities should occur. GIS-based analysis can provide a systematic tool for targeting restoration opportunities by rapidly characterizing potential salmon habitat over large geographic regions and by providing baseline data for development of habitat restoration strategies. When integrated, data on stream channels, riparian habitat, and watershed characteristics provide a powerful tool for the development of watershed restoration and management strategies (Delong and Brusven, 1991).

Previous efforts to prioritize salmon habitat preservation and restoration opportunities on state and federal lands in Oregon (Bradbury et al., 1995) and Washington (Oman and Palensky, 1995) have met with some success, but problems associated with data availability over large geographic regions have limited applications. The objectives of this study were to: (a) develop a rapid, cost-effective, and objective analytical tool to support prioritization of specific subbasins and watersheds for salmon habitat preservation and restoration opportunities; (b) investigate the correspondence between forest seral stage and large woody debris (LWD) recruitment and associated pool-riffle stream bed morphologies; (c) illustrate the creation of integrated baseline data to support watershed analyses and the development of preservation and restoration strategies; and (d) explore the use of such data to facilitate the communication of scientific information to decision-makers and the public.

APPROACH

Classification schemes impose order on a system for some particular purpose. Stream channel classifications, for example, provide a means to evaluate and assess the current condition and potential response of channel systems to disturbance (natural and anthropogenic). Identification of functionally distinct channel types can also target fieldwork on stream reaches of particular interest and provide a reference frame for communication between multidisciplinary groups evaluating habitat conditions. No channel classification is ideal for all purposes, and the approach adopted should reflect the goals to which a classification will be applied. Our project needed to identify the likely location and quality of salmon habitat from existing regional data. Numerous channel classification

systems rely on the integration of physical variables such as channel slope, channel morphology, and channel pattern (Paustian et al., 1992; Montgomery and Buffington, 1993; Rosgen, 1994).

Several existing channel classifications can be applied to PNW streams. Paustian et al. (1992), for example, broadly classify stream channels according to fluvial process groups (i.e., estuarine, palustrine, alluvial fan, etc.). Rosgen (1994) combined channel slope, cross-section morphology, and plan view morphologic attributes to classify stream reaches into general categories. Rosgen (1994) then included channel entrenchment, width to depth ratio, sinuosity, slope, and bed material to further refine stream type categories. Of these attributes only channel slope is readily determined from typical digital data available over broad regions. Moreover, neither approach allows modification of channel type due to the influence of large woody debris contributed from streamside forests, which can be a primary influence on the morphology of stream channels in the Pacific Northwest (Swanson and Lienkaemper, 1978; Keller and Swanson, 1979; Montgomery et al., 1995; Abbe and Montgomery, 1996).

For this study we selected the classification system of Montgomery and Buffington (1993) which broadly stratifies channel morphology and allows for adjustment of channel type due to morphologic influences of LWD, and can be applied over large areas on the basis of correlations with reach average slope. At the reach level of classification, channel morphology is controlled by hydraulic discharge, sediment supply, and external influences such as LWD. The classification identifies distinct alluvial bed morphologies (e.g., pool-riffle, plane-bed, step-pool), and the influence of large woody debris on "forcing" stream morphology is designated by modifiers added to a particular reach label (e.g., forced pool-riffle). These specific channel morphologies can be generalized into source, transport, and response reaches (Montgomery and Buffington, 1993). In mountain drainage basins, source reaches tend to be debris-flow-prone colluvial channels that function as headwater sources of sediment to downstream reaches. Transport reaches tend to be step-pool and cascade morphology reaches that rapidly convey increased sediment loads to lower-gradient downstream channels. Response reaches are pool-riffle and plane-bed channels that can exhibit dramatic morphologic response to increased sediment loads. Channel reach slope (S) generally correlates with reach morphology, particularly at the coarse level distinctions of source ($S \geq 0.20$), transport ($0.04 \leq S < 0.20$), and response reaches ($S < 0.04$).

Among response reaches, several morphologic types may occur. Channels with slopes between 0.001 and 0.01 typically exhibit pool-riffle morphology regardless of LWD loading levels, whereas channels with slopes between 0.01 and 0.02 are LWD dependent: at low LWD loading, these reaches typically have either a pool-riffle or plane-bed (i.e., riffle-dominated) morphology, whereas at higher LWD loading, LWD pieces and LWD jams force the formation of pools, hence the name forced pool-riffle channel. Channels in the 0.02 to 0.04 slope range typically exhibit plane-bed or forced pool-riffle morphologies, depending upon LWD loading (Montgomery and Buffington, 1997). Channels with slopes above 0.04 typically exhibit step-pool or cascade morphologies. Of these channel types, salmonid species appear to strongly prefer pool-riffle and forced pool-riffle channels.

Identification of response reaches provides a simple method for identification of potential salmon habitat, as the zone of anadromous fish use typically is restricted to these low-gradient reaches in Pacific Northwest watersheds (Montgomery, 1994). Also, the age class of streamside forests can indicate the potential for a source of abundant large woody debris to stream channels. Channel slope can be readily determined from digital elevation models. We coupled this coarse slope-based classification of channel types with remote sensing data on the associated seral-stage of the streamside forest to generate a regional GIS-driven classification of potential salmonid habitat locations and quality. This approach, however, simply provides an indication of the likely channel type and a general sense of the likely woody debris loading. Channel slopes determined from digi-

tal elevation models (DEMs) and wood loadings derived from forest seral stage correlations can be misleading due to: (a) poor topographic representation in the DEM at the scale of channels; (b) the natural overlap and range in slope for different channel types; (c) variations in channel type due to local controls; and (d) differences between in-channel LWD loading and riparian forest conditions due to removal of LWD from channels flowing through mid- to late-seral stage forests or clear-cutting of streamside forests without removal of in-channel LWD. In spite of these caveats, the simple classification based on general channel type and riparian forest seral stage provides a direct screening tool for identifying likely sites of low- and high-quality salmon habitat.

We infer that pool abundance correlates with overall habitat quality, and that LWD loading is an important factor in determining habitat quality in response reaches. Assuming that riparian forest conditions correlate with increased LWD loading, it follows that the condition of the adjacent riparian forest would correlate with channel type over the slope range of approximately 0.01 to 0.04, with older forests having higher potential for LWD loading and a higher probability of being forced pool-riffle reaches (high quality habitat). Conversely, reaches with young forests or no forest along the channel have a higher probability of being a plane-bed reach (poor quality habitat).

Application of our results to prioritize specific subbasins and watersheds for salmon habitat protection and restoration efforts is based on three major assumptions: (1) salmon stocks are adapted to local environmental conditions; (2) preservation of "natural" conditions will benefit multiple salmonid species; and (3) a general categorization of channels adequately describes key habitat elements for multiple salmonid species. The first two assumptions are more completely explained by Peterson et al. (1992) and Beechie et al. (1996). The third assumption is supported by limited data showing that several species and life history stages select the same two channel types over others (E. Beamer, unpublished data). These preferences appear to be related to factors such as pool area and depth, cover complexity, and the quality of spawning gravels.

METHODS

Watershed screening was performed at both the subbasin (>450 km^2) and watershed (<260 km^2) scales to identify probable high quality and degraded locations in western Washington State. For the purposes of this study, subbasins correspond to Washington Department of Natural Resources (WDNR) Water Resource Inventory Areas (WRIAs) and watersheds correspond to WDNR Watershed Administrative Units (WAUs). The multiple analytical scales provide comparative evaluations of potential salmon habitat across large geographic regions (e.g., western Washington State) or for evaluations across watersheds within an individual subbasin. Data outputs were summarized by WAU, which typically comprise 120 to 260 km^2. WAUs are subunits within larger subbasins (WRIA) that range from 450 to 6,500 km^2 in western Washington State. The GIS-based predictions of potential habitat locations serve to extend the spatial extent of field observations across the entire study area.

Data Sources

Ideally, all spatial data sources should be derived and used at a scale commensurate with the ecological processes of interest. For this project the appropriate source data scale for watershed analysis and management activities across the western Washington project area is 1:24,000 and larger, to accurately resolve the location of salmonid stream habitat. However, large-scale digital data sets such as vegetation cover and land ownership were not available over the project area. At the expense of spatial resolution, some coarser resolution data sets were used (Table 14.1).

Two sources of digital hydrographic data were available: (1) the U.S. Environmental Protection

Table 14.1. Study Data Sets and Corresponding Scale, Formats, and Source Description

Data	Scale	Format	Description
DEM	1:24,000	raster	7.5-minute; 30-meter cell; Levels 1 & 2.
Hydrography	1:24,000	vector	Compiled from USGS 7.5-minute quads & aerial photography.
Transportation	1:24,000	vector	Compiled from USGS 7.5-minute quads & aerial photography.
WAU Boundaries	1:100,000	vector	Variable accuracy due to multiple regional mapping efforts.
WRIA Boundaries	1:24,000 1:62,500	vector	Boundaries developed by state natural resource agencies in cooperation with the USGS.
Forest Vegetation Seral Stage	~1:100,000	vector/ raster	Landsat Thematic Mapper (TM)-derived forest cover.
Land Use/Land Cover	1:250,000	raster	Non-forested lands (ag./urban/etc.) from USGS Land Use/Land Cover.
Validation Data	1:24,000 1:12,000	Hard copy map/pt. data and tabular data.	Field observations.
Land Ownership	1:100,000	vector	Public land ownership.
Landsat TM	~1:100,000	raster	Terrain-corrected imagery ± 15 meters.

Agency's (EPA's) 1:100,000 scale "river reach files"; and (2) 1:24,000 scale hydrography provided by the WDNR. The river reach files have the advantage of unique identifiers for all stream reach locations. Nonetheless, superior mapping resolution associated with the 1:24,000 scale hydrography data was considered a better match for the needs of this study because most field observations and stream habitat measurements were recorded at scales of 1:24,000 and larger, and absolute stream orientation was critical for subsequent spatial data analyses across multiple thematic data layers.

The only available data source for hydrologic unit delineations at both the subbasin and watershed scales for western Washington State was the WRIA and WAU coverages. WRIA boundaries were compiled from 1:24,000 to 1:62,500 scale maps and WAUs were generally compiled at 1:100,000 scale (WDNR, 1988, 1993). Incongruities between the WAU boundary delineations and the larger scale hydrography were common. For example, along wide river main stems, WAU boundaries were not always in agreement with river main stems, especially as river shape became more sinuous in the 1:24,000 scale hydrography data.

Total road length and road density were important attributes used in the assessment of potential habitat quality at the WAU scale of analysis. Transportation data were available for the project area at 1:100,000 and 1:24,000 scales. The 1:24,000 scale data provided the most complete depiction of primary, secondary, and logging roads.

Available source scales for digital elevation data of western Washington were the 1-degree or three arc-second (~85 meter) data and the 7.5 minute (30 meter) DEM data constructed from 1:24,000 scale maps. Given the need to assess channel slope as accurately as possible for this project, the larger scale data provided the best estimate of stream slope over relatively short stream reaches. The slopes of stream arcs were measured over an arc distance of 150 meters. For this purpose, the 30-meter cell size of the 1:24,000 scale elevation models provided superior topographic resolution.

Table 14.2. Study Land Cover Categories Derived from Landsat 5 Thematic Mapper (TM) Data (PMR, 1993; WDNR, 1994)

Class 1
Late Seral Stage
Coniferous crown cover greater than 70%.
More than 10% crown cover in trees greater than or equal to 21 inches diameter breast height (dbh).
Class 2
Mid-Seral Stage
Coniferous crown cover greater than 70%.
Less than 10% crown cover in trees greater than or equal to 21 inches dbh.
Class 3
Early Seral Stage
Coniferous crown cover greater than or equal to 10% and less than 70%.
Less than 75% of total crown cover in hardwood tree/shrub cover.
Class 4
Other Lands in Forested Areas
Less than 10% coniferous crown cover (can contain hardwood tree/shrub cover; cleared forest land, etc.).
Class 5
Surface Water
Lakes, large rivers, and other water bodies.
Class 15
Nonforest Lands
Urban, agriculture, rangeland, barren, glaciers.

Note:
(1) Forest cover derived from Landsat Thematic Mapper™ satellite imagery.
(2) Class 5 derived from Landsat™ and 1:250,000-scale USGS Land Use/Land Cover data.
(3) Class 15 derived from 1:250,000-scale USGS Land Use/Land Cover data.

Forest cover data were originally derived from 1988 Landsat 5 TM data (PMR, 1993) and updated with 1991 and 1993 TM data using image differencing followed by level slicing to identify new clear-cuts (Collins, 1996). The nominal data resolution of 30 meters was interpolated to 25 meters during the terrain correction process. Standard digital image interpretation techniques were then applied to generate the forest cover data (PMR, 1993). Forest cover was broadly categorized into four classes based on forest type and age class (Table 14.2). The overall thematic accuracy of the 1988 TM-based land cover categorization was 92% (PMR, 1993).

The nonforest land cover and most surface water features were derived from 1:250,000 scale U.S. Geological Survey land cover/use data. The data were overlaid on the forest cover classification to discriminate nonforest lands, such as agriculture and urban areas, from forest lands (PMR, 1993). Thus, the final land cover layer contained a mixture of source scales ranging from approximately 1:100,000 to 1:250,000.

Field data used to validate stream channel type prediction were provided as part of an ongoing salmon habitat inventory and management effort. Inventory efforts focused primarily on streams with relatively low channel slopes (<4.0%) and were compiled on 1:12,000 scale orthophotos. Channel slope data were collected using either transit, hand level, or clinometer measurements.

Data Quality and Error Propagation

Errors associated with remote sensing and GIS data acquisition, processing, analysis, data conversion(s), and final data presentation can significantly impact the confidence associated with resultant products and thus influence their utility in the decision-making process. It was not feasible within the scope of this project to explicitly measure discrete error sources and calculate an error propagation budget. Thus, channel type prediction was the only accuracy assessment performed. Channel type accuracy reflects key input data limitations and processing errors. The final GIS data products may give the appearance of uniform thematic accuracy; however, there may be significant variability across specific geographic locations based on the least accurate input data source (Lunetta et al., 1991).

Spatial Data Preprocessing

Data Format Conversions

Data conversion from vector to raster can cause undesirable shifts of objects in the output raster data as well as changes in area and shape (Congalton and Schallert, 1992). This error source was minimized as much as possible by maintaining data in their native format and thus performing limited data conversions. Raster to vector conversions were not performed as part of the project; however, the forest cover data, originally processed from Landsat TM digital imagery were converted to a vector representation prior to processing (PMR, 1993). Also, all single line hydrographic arcs underwent vector to raster conversion to optimize stream buffer calculations (see Stream Buffer Vegetation Tabulation).

Data Generalization

With the exception of the land cover layer, most data sources were not generalized. The Nonforest (class 15) and Surface Water (class 5) classes listed in Table 14.2 were originally compiled under USGS mapping guidelines. The land use and land cover data were interpreted from aerial photography at a scale of 1:60,000 or larger and compiled on 1:250,000 scale topographic maps (USDI, 1993). The guidelines specify a 4.0 hectare minimum mapping unit for urban/built-up lands, surface water, and some agricultural areas. The minimum mapping unit for cropland, pasture, and barren lands is 16.2 hectares. As noted above, the nonforest and surface water data were overlaid on the seral stage coverage to create a combined land cover layer. This layer was subsequently converted to vector format.

Prior to conversion of the land cover data to vector format, a filtering procedure was performed on the raster data coverage to merge polygons smaller than nine pixels into adjacent polygons using a simple majority rule decision criteria. Subsequently, vector polygons smaller than the minimum mapping unit size of 2.0 or 4.0 hectares (depending on adjacent land cover type) were removed (PMR, 1993). Thus, stream bank forest land cover was not accurately represented for patch sizes of less than 4.0 hectares.

Geometric Rectification

GIS processing of multiple data layers requires that all layers reside in a common map projection. A common projection was determined prior to processing based upon possible error sources and processing efficiency. Because project outputs were assumed to be most sensitive to elevation errors, the DEM data were maintained in their native Universal Transverse Mercator (UTM) projection. Changing their projection would have introduced additional error and increased data pro-

cessing time due to interpolation of cell values as both the DEM and vector land cover data existed in UTM space. Therefore all input data not in UTM space were projected to the UTM space of the DEMs and land cover data.

DEM Processing

ARC/GRID analytical routines were used to mosaic 7.5-minute DEMs for each WRIA. Once a grid was created for a WRIA, it was next processed to remove "sinks" (sinks are cells with an undefined flow direction). The processed DEM was then used in stream channel slope calculations. Additional DEM processing was performed to create a slope grid for each WRIA for use in summary statistics compilation. The slope grid was then recoded into three landscape slope classes: (a) Class 1, 0–29%; (b) Class 2, 30–65%; and (c) Class 3, greater than 65%.

Preparation of Hydrography

The 1:24,000 scale hydrography data, originally tiled by township, were appended and clipped to the respective WRIA boundary. Stream direction was set to point upstream. Stream percent slope was then computed from DEM values for each arc. Start and end elevation and slope were written to the hydrography coverage arc attribute table.

Forest Seral Stage

This coverage was checked for positional errors and logical consistency by overlaying it with the ancillary geocoded TM data to serve as a base map. The absolute positional accuracy of the TM base map was plus or minus 15 meters (Table 14.1). If positional errors were found, then a simple x,y shift was performed to improve geometric fidelity. Thematic inconsistencies between the vegetation layer and the TM data were not reviewed. Ideally, obvious errors, such as urban encroachment on forest lands, would have been corrected through editing procedures using the TM data as a validation data source. However, resource limitations precluded the inclusion of such editing.

Spatial Data Analysis

Each WRIA was processed individually using identical protocols. The first step was to compile all data inputs for processing, followed by creation of summary data statistics, hard copy maps, and graphics. Summary statistics and data graphics were generated for both the WRIA and WAU hydrographic units. Additionally, validation procedures were performed using data from nine WAUs (Bacon Creek, Illabot, Jackman, Nookachamps, Finney, Hansen Creek, Gilligan, Mt. Baker, and Alder) located within the Upper and Lower Skagit River WRIAs. The categorization of stream channel types was accomplished using an automated procedure to calculate slope for individual stream reaches.

The sampling procedure was initiated at the low elevation end of the arc, and measured upstream the specified sampling distance of 150 meters. If a slope less than 4% was found over the sample distance, then the arc ID number and UTM coordinate at the end of the sample distance were stored in a file. Upon locating a slope less than four percent, the procedure then moved upstream along the arc another sample distance and measures the slope. The process was repeated if a slope less than 4% was found, otherwise the remainder of the current arc was abandoned and the next one is sampled. Stream segments listed in the output file were then split at the specified UTM coordinates using an automated editing procedure. After the editing procedure, the updated slope and elevation values were written to the edited hydrography coverage.

Stream Buffer Vegetation Tabulation

A 30-meter raster stream buffer was generated along both sides of single-arc streams in the hydrographic data layer. The actual cell size implemented to model stream buffers was six meters; thus, the width of each buffer was modeled with five six-meter cells. Each raster buffer was indexed to the vector hydrography line coverage and the percentage of each land cover category (Table 14.2) was written to its respective arc in the arc attribute table. Raster procedures were incorporated to speed up the buffer processing time through the use of rapid cross tabulation procedures between the buffer areas and the raster vegetation layer. For wider streams and rivers which are depicted with double arcs, buffers were extended from each bank, and vector processing procedures were used to summarize vegetation within each stream buffer. Statistics were then generated from the buffer summary tables for each arc.

Summary Reports

WRIA summary reports were organized by WAU and list attributes for streams, vegetation, roads, slope, and land ownership (Table 14.3). Map and bar chart plot files produced from WRIA coverages and summary table attributes were used to plot graphical aids for watershed assessment teams. WRIA maps can be generated on a large format plotter to depict the following themes: (a) response, transport, and source channel types; (b) vegetation classes; (c) transportation networks; (d) slopes; (e) land ownership; and (f) WRIA and WAU boundaries.

Validation of Stream Channel Type Predictions

Validation was performed by comparing field observations with GIS-generated channel type predictions (Lunetta et al., 2001). Field assessment data were provided to the project for the nine Lower and Upper Skagit River WAUs previously listed. The length of the sample reaches generally ranged from 100 to 300 meters, and the midpoint of each reach was delineated on a topographic map. The comparison was accomplished through creation of an error matrix for each WAU (Story and Congalton, 1986). A Kappa coefficient was calculated using discrete multivariate statistical techniques as a measure of the overall agreement between the stream channel type predictions and field observations (indicated as the major diagonal) versus agreement that is contributed by chance (Congalton et al., 1983). The Kappa coefficient was calculated based on the formula given by Hudson and Ramm (1987).

RESULTS

Of the 164,083 km of stream reaches analyzed, 23.2% (38,002 km) were categorized as response reaches (≤4.0% slope), of which, 8.7% (3,302 km) were associated with late seral and 20.7%(7,867 km) with mid-seral stage forest stream vegetation.

Table 14.3. WRIA Summary Report Attributes, Extent, and Description

Attribute	Extent	Description
Streams	WRIA/WAU	Total kilometers and stream density. Stream density and percent by predicted channel reach type.
Seral Stage	WRIA/WAU/Stream buffer	Hectares and percents.
Slope	WRIA/WAU	Hectares and percent of landscape slope in three classes.
Roads	WRIA/WAU	Total kilometers and road density.

Hydrographic Data Scale

Figure 14.1 (a,b, see color section), clearly illustrates the deficiencies of the 1:100,000 scale hydrography stream network (compared to the 1:24,000 scale product) for depicting the actual stream channel network. In the Finney Creek WAU, a total of 490.1 km of stream length are contained in the 1:24,000 scale hydrography compared to 94.8 km in the 1:100,000 scale product. More importantly, the results of the response reach analyses indicate a significant underestimate of response reaches associated with 1:100,000 scale coverage compared to the 1:24,000 scale (43.0 km and 64.9 km, respectively). The smaller scale EPA hydrographic data in addition to lacking resolution in the number of streams, was also deficient in absolute stream orientation detail.

Stream Slope Sampling

Results of the stream slope sampling procedure are presented in Figure 14.2. The optimal sample length corresponds to the maximum stream arc sampling distance that provides the maximum response reach length. Seven sample distances (100, 125, 150, 175, 200, 225, and 300 meters) were evaluated for each of three WRIAs (Figure 14.2) which represented a broad range of physiographic conditions present throughout western Washington State (Lower Skagit, Willapa River, and Lyre-Hoko). The objective was to determine the maximum effective distance to minimize computational requirements, where 100 meters is the minimum feasible sampling length. Sample length must be sufficiently long to capture the inherent variation of the DEM. Short sample distances are ineffective because the elevation change over the sample length is often very low or zero, and exceedingly long sample lengths tend to mask slope changes.

The stream slope sampling procedure enhanced the detection of response reaches located between stream confluences and the base of steep mountain slopes and identified additional response

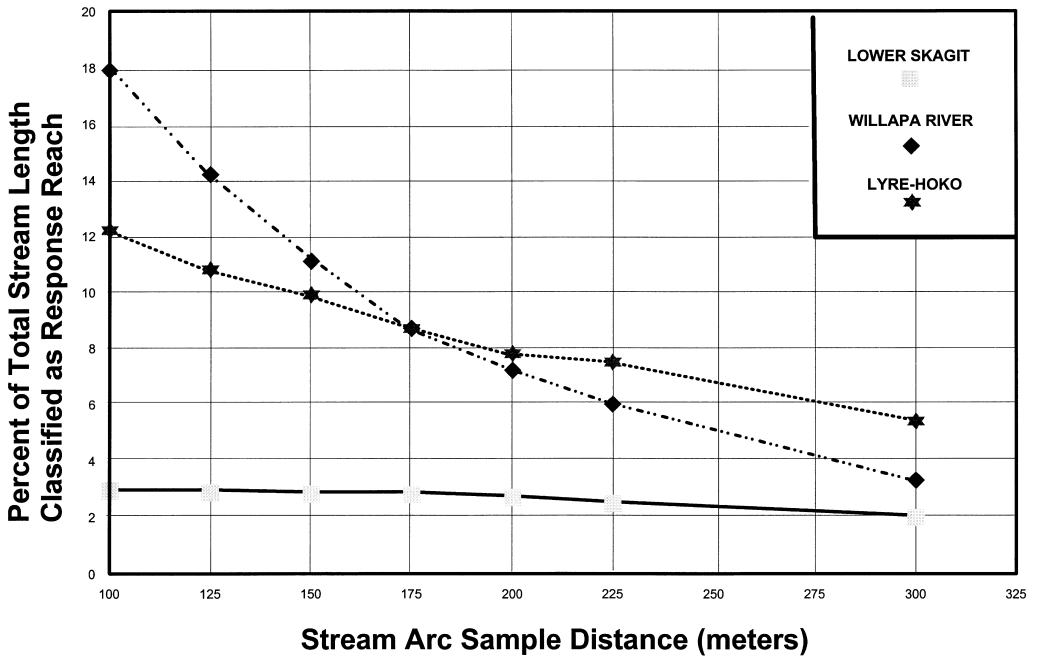


Figure 14.2. Plot of length of predicted response reaches versus sample arc distance as calculated for the Lower Skagit, Willapa River, and Lyre-Hoko WRIAs.

reaches within relatively long stream lengths in moderate terrain. Evaluation of the sample distances indicates that a distance of 100 meters, the shortest distance tested, generated the greatest response length for all WRIAs. As shown for the Willapa WAU (Figure 14.2), the total length of predicted response reaches tends to decrease rapidly as sample length increases from 100 meters to 175 meters, then declines more gradually to 300 meters. The Lyre-Hoko's decline was more gradual than the Willapa, whereas the Lower Skagit was only slightly sensitive to sampling distance. Variations in sampling distance appear to have the greatest effect in locations with moderate to steep terrain. For example, 81% of the landscape of the Lower Skagit WRIA had a slope less than 30 percent; the percentage of area within the Willapa and Lyre-Hoka WRIAs with a landscape slope less than 30% was 75 and 57%, respectively. It appears that hydrologic units with moderate to steep terrain experience the greatest relative increase in response length with decreased sampling distance. Although a l00-meter sample distance maximizes the length of response reaches, a sample distance of 150 meters was applied to minimize errors of commission, and simultaneously reduce processing time and data volume.

Stream Bed Morphology

Field observation data were collected from a total of 120 response reach stream segments in both the Lower and Upper Skagit WRIAs to examine the association between stream buffer zone vegetative land cover and stream bed morphology (Table 14.4). Results indicate that late seral stage forests are associated with forced pool-riffle stream bed morphology. However, the small number of samples (n=8), precludes the drawing of any final conclusions. Response reach buffer zones containing any type of forested land cover had a 77% correspondence to forced pool-riffle stream bed morphology. Nonforested buffer zones were associated with forced pool-riffle morphologies in 35% of the field observations.

Habitat Evaluation

Results applicable to the evaluation of salmon habitat in western Washington State are illustrated in Figure 14.3 (a,b, see color section). The summary bar chart generated for each WRIA and WAU provides a means of comparing potential salmon habitat conditions across WRIAs and to support intra-WRIA assessments. The summary table data for an entire WRIA and individual WAUs include the following information categories: (a) vegetation percent by class; (b) vegetation percent by class within response channel buffers; (c) response, transport, and source channel density; (d) road density; and (e) landscape slope. Summary graphics include drainage density by

Table 14.4. Correspondence between Response Reach Land Cover Categories versus Stream Bed Morphology

Response Reach Land Cover Categories[a]	Percent Reaches Classified as Forced Pool-Riffle
Late Seral Stage	100% (n = 8)
Mid-Seral Stage	78% (n = 18)
Early Seral Stage Other Forest	74% (n = 68)
Non-Forest	35% (n = 26)

[a] observations made along 30m buffers along each bank

channel type and forest seral stage coverage expressed as a percent of total watershed and percent area within buffers around response reaches. These data can facilitate the rapid inference of general streamside conditions and potential for LWD recruitment. In addition, road density and slope data provide some insight to the potential for sedimentation impacts within a given hydrologic unit.

In western Washington more than one-fourth of WRIAs have no late seral stage forest bordering response reaches, and 73% of WRIAs have late seral stage forests along 10% or less of the total response reach length (Figures 14.4 and 14.5). These areas tend to be associated with urban, agricultural or commercial forest land use. Only three WRIAs have more than 20% of their response reach length bordered by late seral stage forests. And these lie partially within national parks or wilderness areas. Overall, only 8.7% of response reaches flow through late seral stage forests. This provides the first quantitative regional characterization of the extent of habitat modifications that accompanied urbanization, agricultural development, and industrial forestry.

Within the Upper Skagit River basin, approximately one-tenth of WAUs had late seral stage forests bordering 10% or less of the total response reach length (Figures 14.6 and 14.7). However, 43% of WAUs had late seral stage forests along 50% or greater of the total response reach length. Of the 20 WAUs identified with late seral stage forests along 25% or greater of total response reach length (highest quality WAUs), eight (40%) were above major dams (Figure 14.7). As in the province scale assessment, land uses in the WAUs with low percent late seral stage tend to be dominated by agricultural and urban development, although some of these WAUs were predominantly industrial forests. WAUs with high percent late seral stage tended to be largely within the boundaries of national parks, national recreation or wilderness areas.

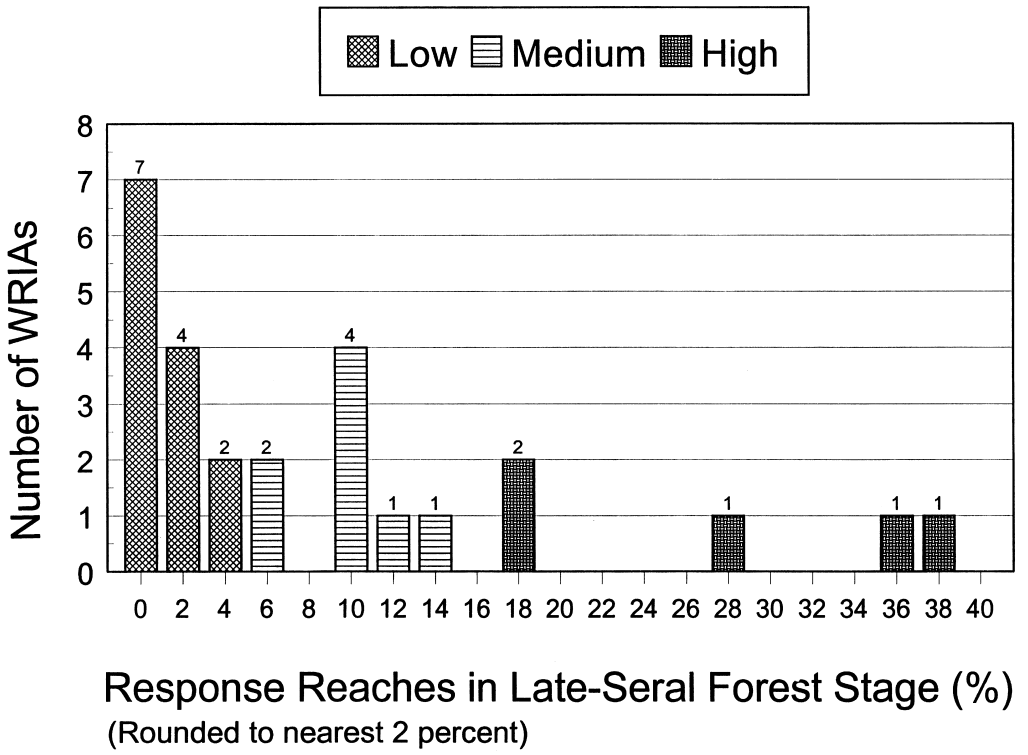


Figure 14.4. Frequency distribution of percent WRIA response reaches in late seral forest stages for western Washington State.

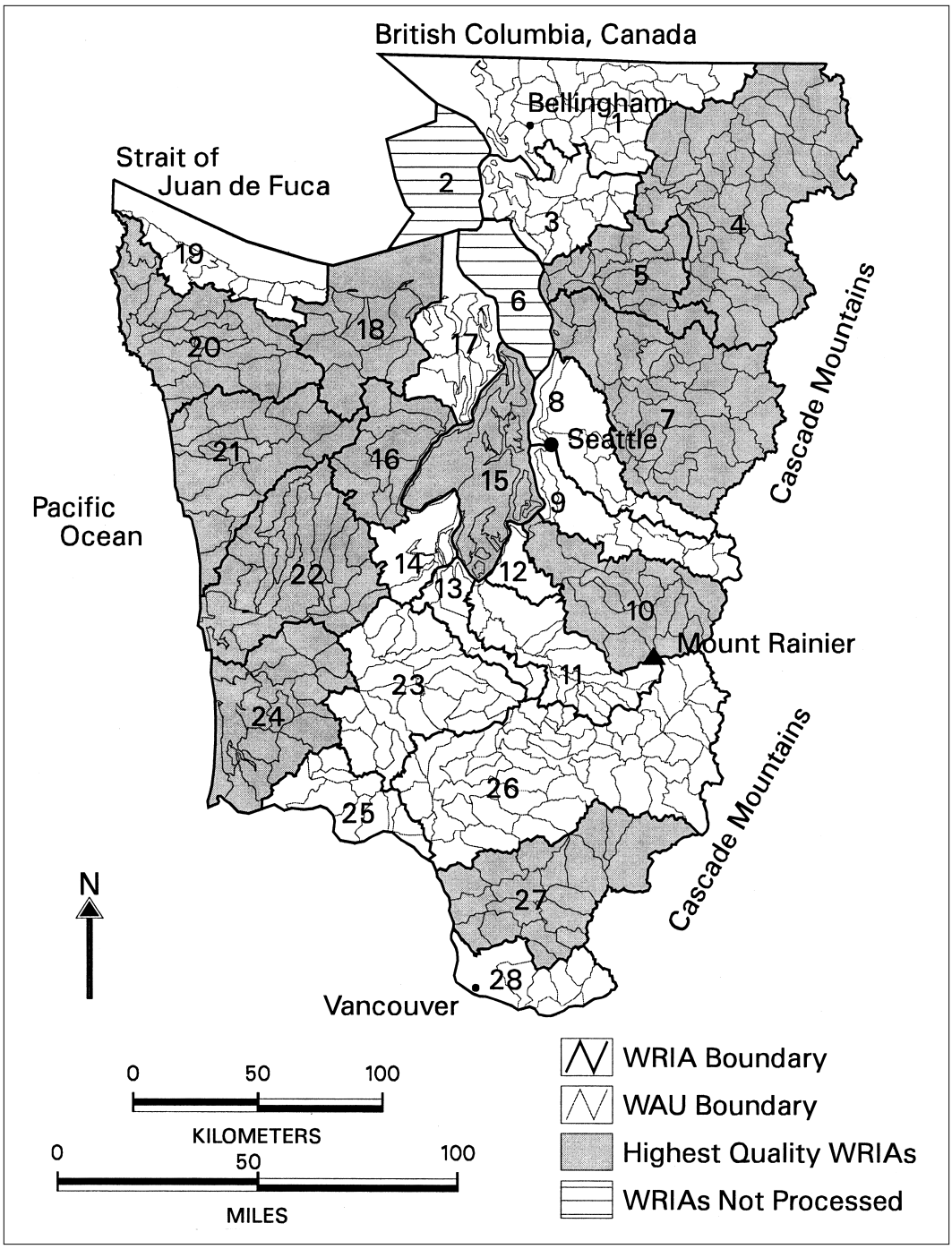


Figure 14.5. Identification and location of the highest quality WRIAs in western Washington State. Note that WRIAs 2 and 6 were not processed because they contain only islands.

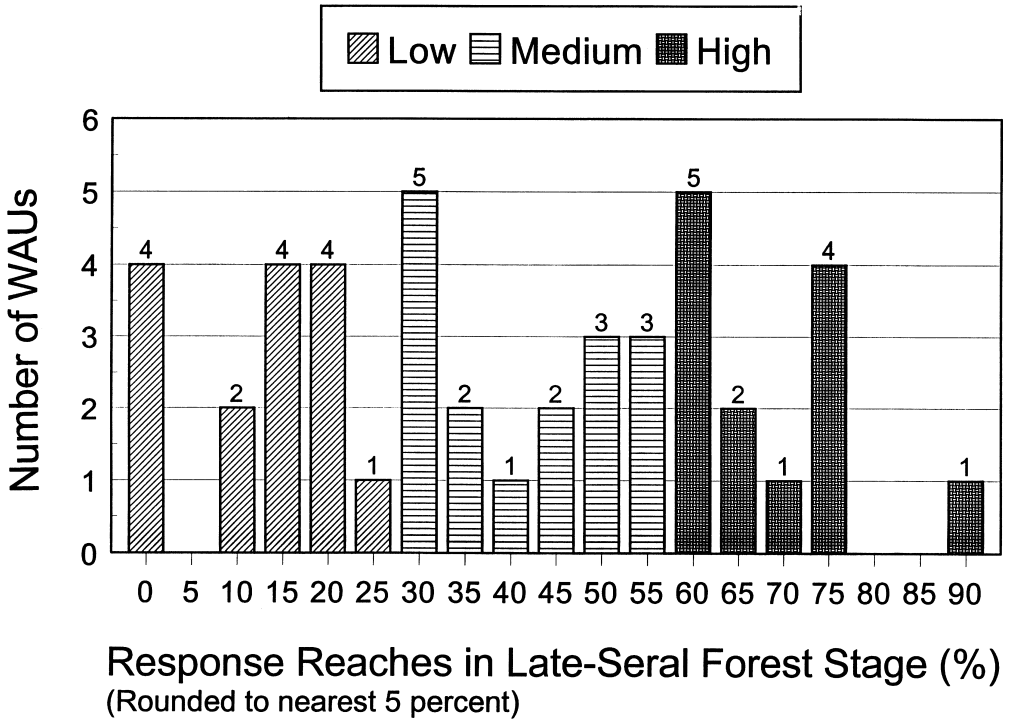


Figure 14.6. Frequency distribution of percent Upper Skagit WAU response reaches in late seral forest stage.

Accuracy Assessment

Although validation was limited to nine WAUs, the basic relationships between physical processes and stream habitat are thought to be consistent across the study area, and the validation for those nine watersheds should be representative for western Washington State. The results of the validation are presented in an error matrix (Table 14.5). The identification of response reaches was 96% accurate, and the overall accuracy of all channel type predictions was 79%, Kappa statistic = 0.64 (n=158). Errors of omission and commission associated with predicted response reaches were 24.0% and 4.0%, respectively. As mentioned above, the use of the 150-meter arc sampling distance tended to minimize commission errors while increasing errors of omission between response and transport channel types. In theory these omission errors could be reduced by using a 100 meter arc sampling distance, but commission errors would likely increase. However, the ultimate limiting factor is the resolution and quality of the DEM data.

DISCUSSION

The intent of this effort was to produce a regionally consistent information base that federal agencies could use for planning or prioritizing salmonid habitat restoration opportunities in the PNW. Our analyses were based on simple concepts that are consistent with our understanding of habitat-forming processes in western Washington State. These are: (a) channel slope largely determines the range of potential channel morphologies; (b) large woody debris abundance modifies within channel type morphology; and (c) salmonid habitat utilization increases with increased

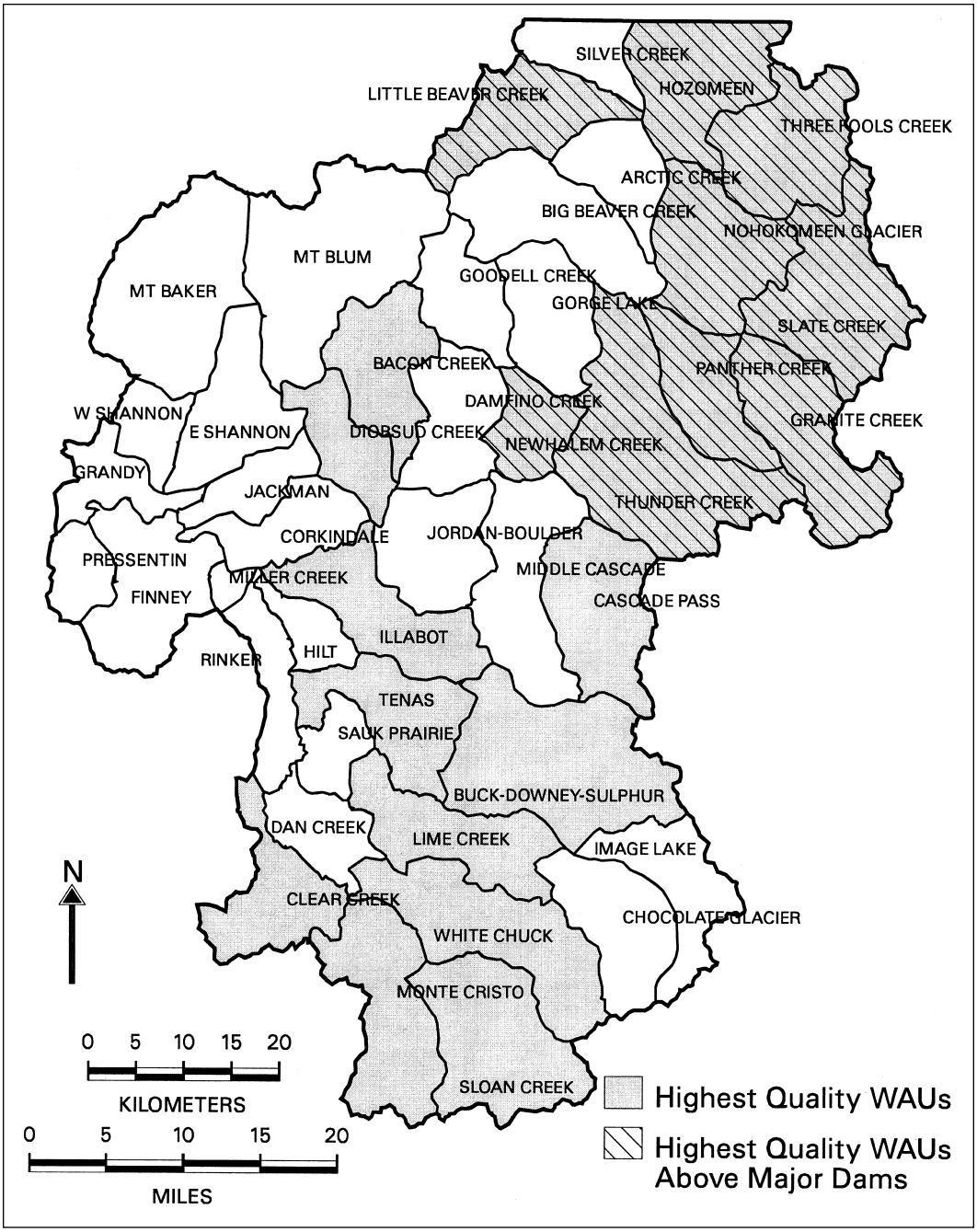
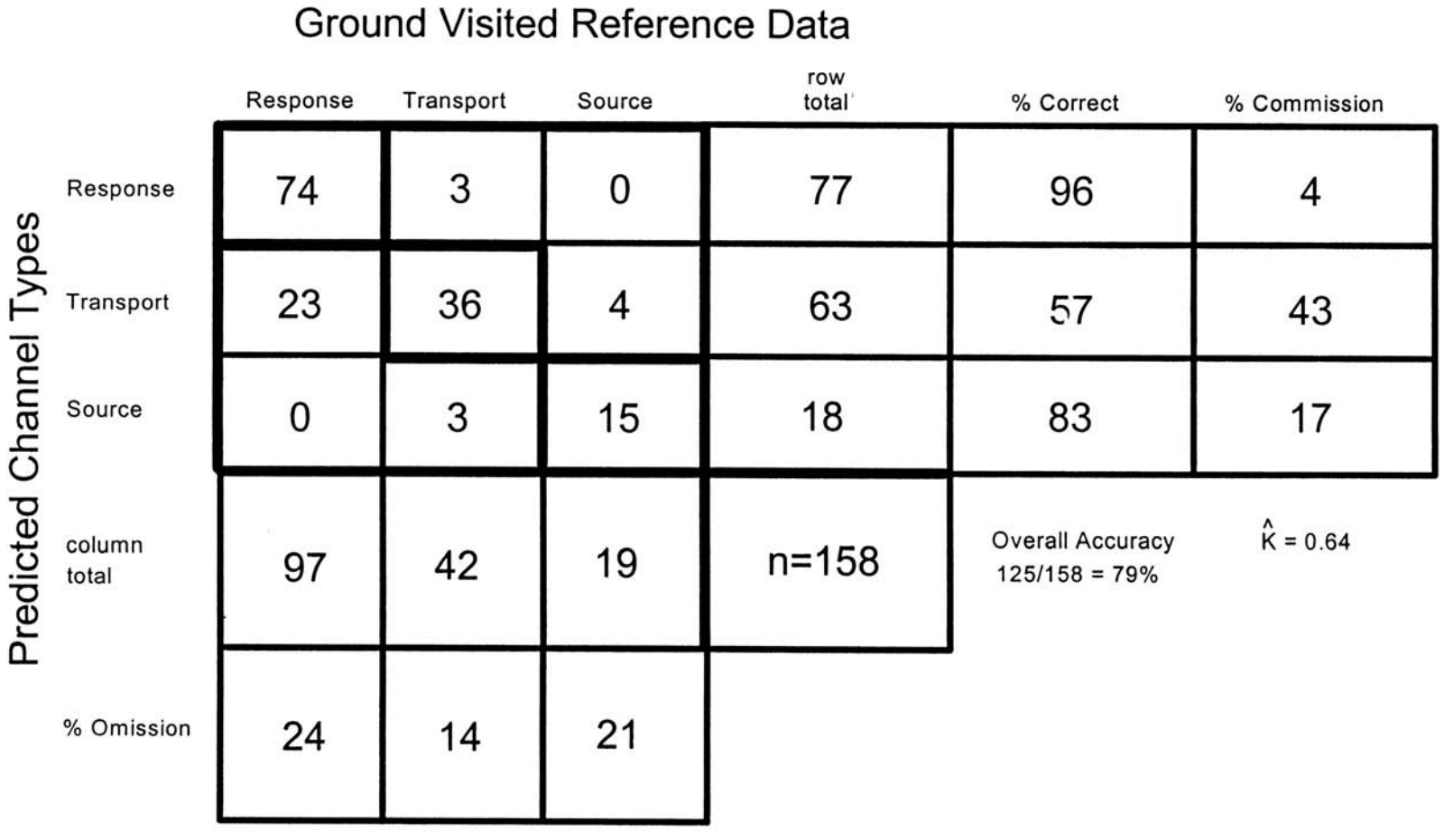


Figure 14.7. Identification and location of the highest quality WAUs in the Upper Skagit WRIA.

LWD abundance in the response reach channel type. We also presumed that large conifer riparian forests tend to be associated with greater LWD abundance than open or early seral stage riparian areas. Hence, the fundamentally important outputs of our analyses are the extent and location of response reaches (slope <0.04) and the condition of riparian forests along response reaches. The extent and location of response reaches identifies areas that may provide suitable habitat for

Table 14.5. Error Matrix Comparing Ground Visited Reference Data to the Predicted Stream Reach Data



salmonids, and riparian forest conditions indicate the likelihood that those reaches have the forced pool-riffle morphology that salmonids favor.

Our accuracy assessment generally supports the assumptions listed above. However, users of such data should be aware that, while the model typically underrepresents the extent of response reaches, areas identified as response reaches are likely to be correct. Field efforts designed to more accurately identify locations of potential salmonid habitat should therefore focus on areas identified as transport reaches. Field data suggest that those response reaches incorrectly identified as transport reaches are often located where tributary channels enter the valleys of larger channels.

The analyses were less accurate at predicting channel morphology within response reaches, although results generally support the hypothesis that increased forest age is associated with increased LWD abundance. Also, we found little difference in the proportion of forced pool-riffle channels between early and mid-forest seral stages. Histograms of GIS-generated data provide a broad-brush description of channel and riparian conditions at scales that are useful to managers with statewide or regional jurisdiction (Figure 14.4). These data provide a crude but comprehensive characterization of landscape and stream channel attributes that influence the abundance and condition of salmonid habitats.

A qualitative comparison between the preceding results and a field-based assessment of habitat losses in the Skagit River basin reveals that our GIS-based predictions are generally consistent with field data collected independently of this study. Based on the results of our analysis, we predict that the greatest habitat losses have occurred in the Skagit river floodplain and delta where little late seral stage forest remains. Beechie et al. (1994) found that by far the greatest proportion (73%) of coho salmon-rearing habitat losses were associated with diking, ditching, and dredging in the floodplain, and that these losses were associated primarily with urban and agricultural land uses. Hence, our GIS-based results at least grossly predict the same result as a field-based assessment.

Beechie et al. (1994) further noted that industrial forestry had less impact on coho-rearing habitat losses at the river basin scale, but was nevertheless strongly associated with habitat losses in tributary streams (channel widths <10 meters). Thus, forestry was associated with less severe impacts to coho salmon-rearing habitat than were urban and agricultural uses. Our results are also consistent with this relative ranking of severity of impact by type of land use. We show no late seral stage forest in WAUs where nonforest land uses dominated response reach zones, suggesting that the most severe impacts to habitat would be located in those WAUs. By contrast, we found a broader range of percent late seral stage in WAUs where forestry borders the majority of response reaches, indicating that impacts to rearing habitat should be less severe in those WAUs.

Although not all response reaches were bordered by late seral stage forest prior to European settlement, our results suggest a dramatic change in riparian conditions during the last 100 to 200 years. Prior to European settlement, forest fires, floods, and channel migration were dominant influences on stand ages and types near streams (e.g., Agee, 1988). Certainly these processes would create in a patchwork of stands along channel networks, resulting in a range of forest types and seral stages along response reaches. Our data for WAUs contained partially or fully within national parks and wilderness areas give some indication of this patchwork (Figure 14.8). The median percentage of response reaches in late seral stage WAUs located substantially in park and wilderness areas was 54%. This compares to 22% for commercial forestry, and <10% for urban-agriculture land uses. We caution, however, that the percentages shown in Figure 14.8A should not be construed as representative of "natural" conditions because many WAUs contain significant amounts of development.

In addition to a relative ranking, the data distributions can provide useful information for the development of preservation and/or restoration prescriptions. For example, some WRIAs have a

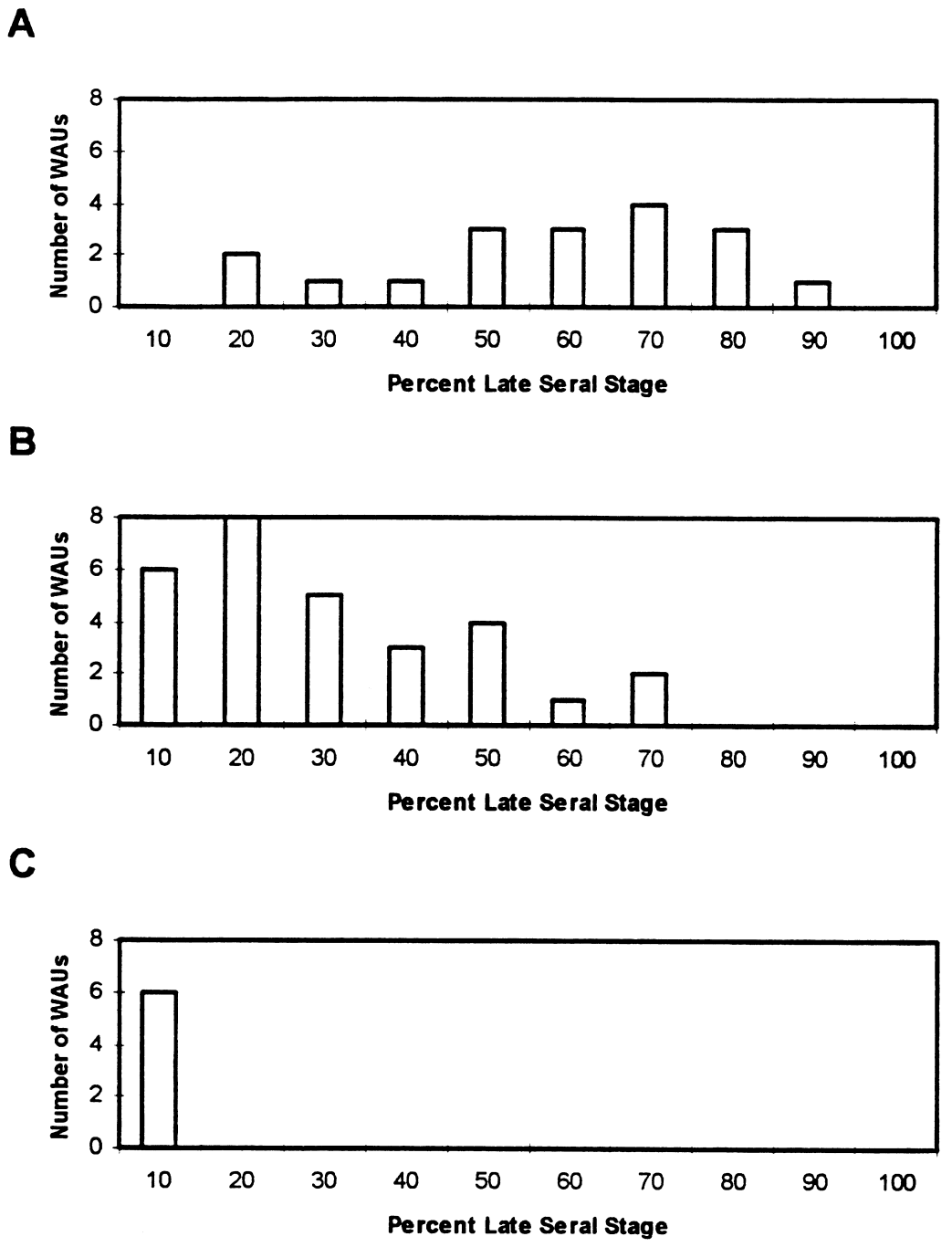


Figure 14.8. Frequency distribution of percent late seral stage along response reaches in WAUs dominated by (a) park and wilderness, (b) commercial forestry, and (c) urban-agriculture land uses.

relatively low percentage of response reaches in the late-seral forests, but a high percentage in mid-seral forests. A rational restoration consideration for these WRIAs may be the preservation of existing mid-seral forests in WAUs with a high density of response reaches. However, use of these analytical tools for identifying tasks or priorities for salmon habitat preservation and restoration can only be accomplished through a process that includes involvement of experts with knowledge of in situ habitat conditions. With the proper expertise and selected ancillary data (e.g., physical barriers to fish migration), map products identifying specific attributes of WRIAs and WAUs could provide a valuable data source to help prioritize the expenditure of preservation and restoration resources.

CONCLUSIONS

Our efforts demonstrate that remote sensing data and GIS methods can be applied to assess landscape attributes that influence the condition of salmon habitat at subbasin to watershed scales. GIS-based analytical products can be used to predict the locations of response reaches likely to provide salmon habitat. By using GIS buffering procedures along response reaches, the likelihood of finding a forced pool-riffle morphology based on the adjacent stream bank vegetation associations can be estimated. Both types of predictions have quantifiable error rates. These products could be used to target reaches where predictions are poor (e.g., the 23% of reaches predicted to be transport reaches that were response reaches), thereby increasing the efficiency of field efforts. Furthermore, such products can rapidly identify the quantity, extent, and condition of habitats at a scale useful for prioritizing regional protection or restoration efforts. We believe that such a wide-area, uniform database (uniform map themes and uniform coordinate system) can complement existing watershed screening protocols and help accomplish prioritization more rapidly and with greater reliability and objectivity.

SUMMARY

Categorization of 164,083 kilometers of stream length has provided the first quantitative measure of the extent and location of potential salmon stream habitat throughout western Washington State. Reach slope and forest seral stage provided a coarse indicator of channel condition across the region. Reach-average slopes calculated for individual stream reaches using 30-meter digital elevation model (DEM) data correctly identified low-gradient (<4.0% slope) response reaches that typically provide habitat for anadromous salmon with an accuracy of 96% (omission and commission error rates of 24.0 and 4.0%, respectively). Almost one-quarter (23.2%) of all stream length categorized consisted of response reaches, of which, only 8.7% were associated with late seral and 20.7% with mid-seral forest stages. Approximately 70% of the total stream length potentially providing anadromous salmon habitat is associated with nonforested and early-seral stage forests. GIS-based analytical techniques provided a rapid, objective, and cost-effective tool to assist in prioritizing locations of salmon habitat preservation and restoration efforts in the Pacific Northwest.

ACKNOWLEDGMENTS

The authors would like to acknowledge Bradford L. Johnson for graphics support. The U.S. Environmental Protection Agency (EPA) partially funded and collaborated in the research described here. It has been subject to the agency's programmatic review and has been approved for publication. Additional funding was provided by the Skagit System Cooperative, the U.S. Department of

Agriculture (USDA) Forest Service through Cooperative Agreement PNW–93–0441, and the USDA Cooperative State Research Service under Agreement No. 94–37101–0321. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.
Reproduced with permission, the American Society for Photogrammetry and Remote Sensing. Lunetta R., Cosentino B., Montgomery D., Beamer E. and Beechie T. "GIS-Based Evaluation of Salmon Habitat in the Pacific Northwest". *Photogrammetric Engineering and Remote Sensing*, Vol 63 no. 10 (October 1997), 1219–1229.

REFERENCES

Abbe, T.B., and D.R. Montgomery, 1996. Interaction of large woody debris, channel hydraulics and habitat formation in large rivers. *Regulated Rivers: Research and Management*, 12: 201–221.

Agee, J.K., 1988. Succession dynamics in forest riparian zones. In *Streamside Management: Riparian Wildlife and Forestry Interactions*. K.J. Raedeke, Ed., University of Washington, Institute of Forest Resources, Contribution No. 59.

Beechie, T. J., E. Beamer, B. Collins, and L. E. Benda, 1996. Restoration of habitat-forming processes in Pacific Northwest watersheds: A locally adaptable approach to salmonid habitat restoration. In *Role of Restoration in Ecological Management*, D.L. Peterson and C.V. Klimas, Eds. Society for Ecol. Restoration, Madison, WI, pp. 48–67.

Beechie, T., E. Beamer, and L. Wasserman, 1994. Estimating coho salmon rearing habitat and smolt production losses in a large river basin, and implications for restoration. *North American Journal of Fisheries Management*, 14: 797–811.

Bradbury, B. et al., 1996. Handbook for prioritizing native salmon and watershed protection and restoration. In review, 1996.

Collins, D., 1996. *The Rate of Timber Harvest in Washington State:* 1988–1991. Washington State Department of Natural Resources Publication, Forest Practices Division, Report No.1.

Congalton, R.G., R.G. Oderwald, and R.A. Mead, 1983. Assessing Landsat classification accuracy using discrete multivariate statistical techniques. *Photogrammetric Engineering and Remote Sensing*, 49(12): 1670–1678.

Congalton, R.G., and D.M. Schallert, 1992. Exploring the Effects of Vector to Raster and Raster to Vector Conversion. U.S. Environmental Protection Agency Internal Report, EPA/600/166.

Delong, M.D., and M.A. Brusven, 1991. Classification and spatial mapping of riparian habitat with applications toward management of streams impacted by nonpoint source pollution. *Environmental Management*, 15(4): 565–571.

Hudson, W.D., and C.W. Ramm, 1987. Correct formulation of the kappa coefficient of agreement. *Photogrammetric Engineering and Remote Sensing*, 53(4): 421–422.

Keller, E.A., and F.J. Swanson, 1979. Effects of large organic material on channel form and fluvial processes. *Earth Surface Processes*, 4: 361–380.

Lunetta, R.S., R.G. Congalton, L.K. Fenstermaker, J.R. Jensen, K.C. McGwire, and L.R. Tinney, 1991. Remote sensing and geographic information system data integration: Error sources and research issues. *Photogrammetric Engineering and Remote Sensing*, 57(6): 677–678.

Lunetta, R.S., J. Iiames, J. Knight, R.G. Congalton, and T.H. Mace, 2001. An assessment of reference data variability using a “virtual field reference database.” *Photogrammetric Engineering and Remote Sensing,* 67(6): 707–715.

Montgomery, D.R., 1994. Geomorphological influences on salmon spawning distributions. Abstracts with Programs, *Proceedings of the Annual Geological Society of America Conference*, pp. A–439.

Montgomery, D.R., and J.M. Buffington, 1993. Channel Classification, Prediction of Channel Response, and Assessment of Channel Condition. Washington Department of Natural Resources Report, TFW-SH10–93–002.

Montgomery, D.R., and J.M. Buffington. Channel reach morphology in mountain drainage basins, 1997. *Geologicical Society of America Bulletin*, 109: 596–611.

Montgomery, D.R., J.M. Buffington, R.D. Smith, K.M. Schmidt, and G. Press, 1995. Pool spacing in forest channels. *Water Resources Research*, 31: 1097–1105.

Nehlsen, W., J.E. Williams, and J.A. Lichatowich, 1991. Pacific salmon at the crossroads: Stocks at risk from California, Oregon, Idaho, and Washington. *Fisheries*, 16(2): 4–21.

Oman, L., and L. Palensky, 1995. Preliminary Priority Watersheds for Restoration and Conservation of Fish and Wildlife. Washington Department of Fish and Wildlife Report, Olympia, WA.

Pacific Meridian Resources (PMR), 1993. Washington State Forest Cover and Classification and Cumulative Effects Screening for Wildlife and Hydrology. Final report submitted to Washington Department of Natural Resources, Olympia, WA.

Paustian, S.J., et al., 1992. A Channel Type User's Guide for the Tongass National Forest, Southeast Alaska. USDA Forest Service, Alaska Region 10, Technical Paper 26.

Peterson, N.P., A. Hendry, and T.P. Quinn, 1992. Assessment of Cumulative Effects on Salmonid Habitat: Some Suggested Parameters and Target Conditions. Washington Department of Natural Resources, Report No. TFW- F3–92–001.

Rosgen, D.L., 1994. A classification of natural rivers, *Catena*, 22(3): 169–199.

Story, M., and R.G. Congalton, 1986. Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52(3): 397–399.

Swanson, F.J., and G.W. Lienkaemper, 1978. Physical Consequences of Large Organic Debris in Pacific Northwest Streams. U.S. Department of Agriculture Forest Service, Pacific Northwest Forest and Experiment Station, General Technical Report PNW–69, Portland, OR.

United States Department of Interior (USDI), 1993. Land Use and Land Cover from 1:250,000 and 1:100,000 Scale Maps. Geological Survey, Data Users Guide No. 4.

Washington Department of Natural Resources (WDNR), 1994. Data documentation for statewide classified canopy coverages.

Washington Department of Natural Resources (WDNR), 1993. Detailed Layer Description for Water Resource Inventory Areas. Forest Practices Division, Olympia, WA.

Washington Department of Natural Resources (WDNR), 1988. Detailed Layer Description for Watershed Administrative Units. Forest Practices Division, Olympia, WA.

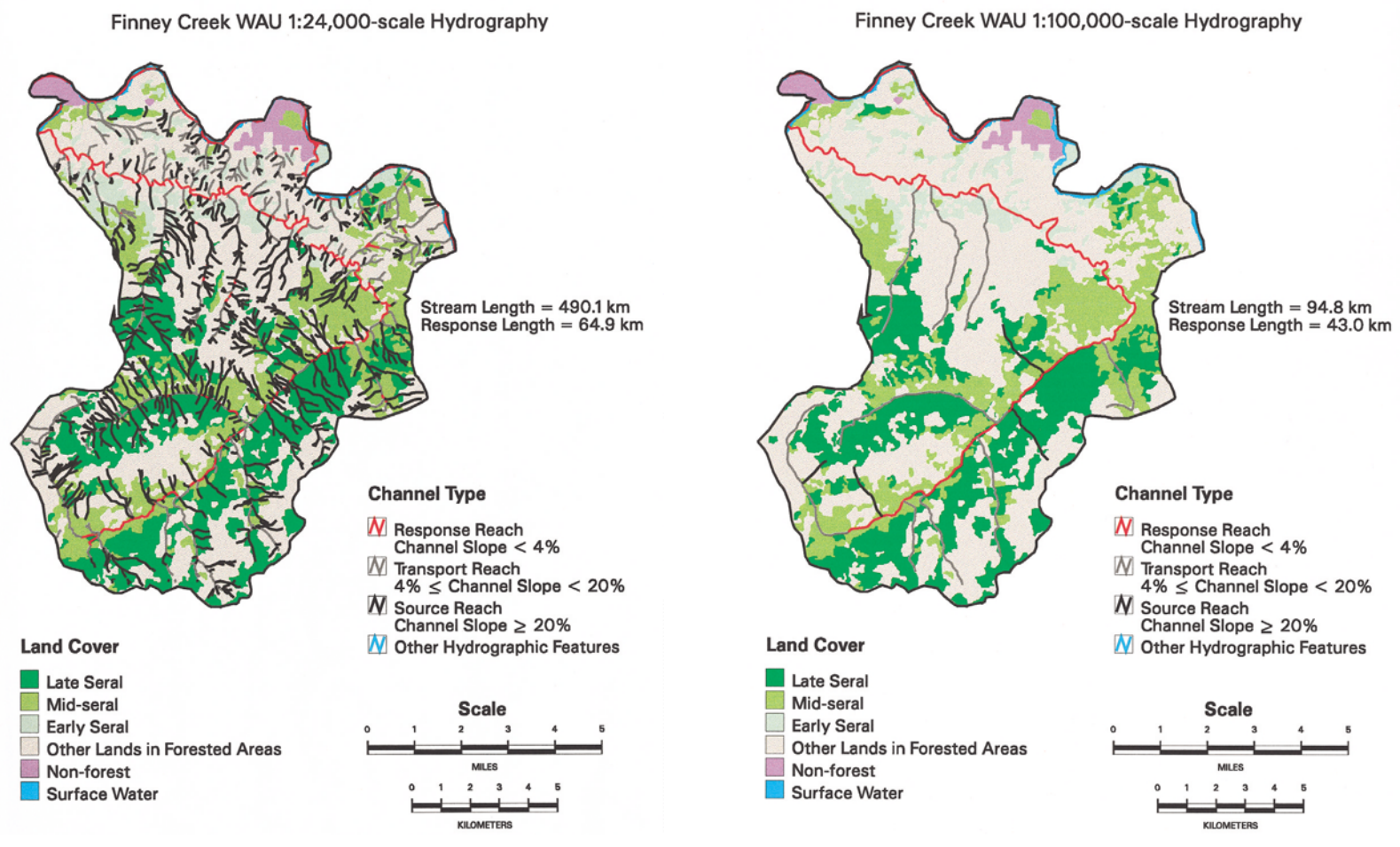


Figure 14.1 (a,b). Comparison of 1:24,000 (a) versus 1:100,000 (b) scale hydrography data. Note the greater stream density and 50% increase in response reach length associated with the 1:24,000 scale data.

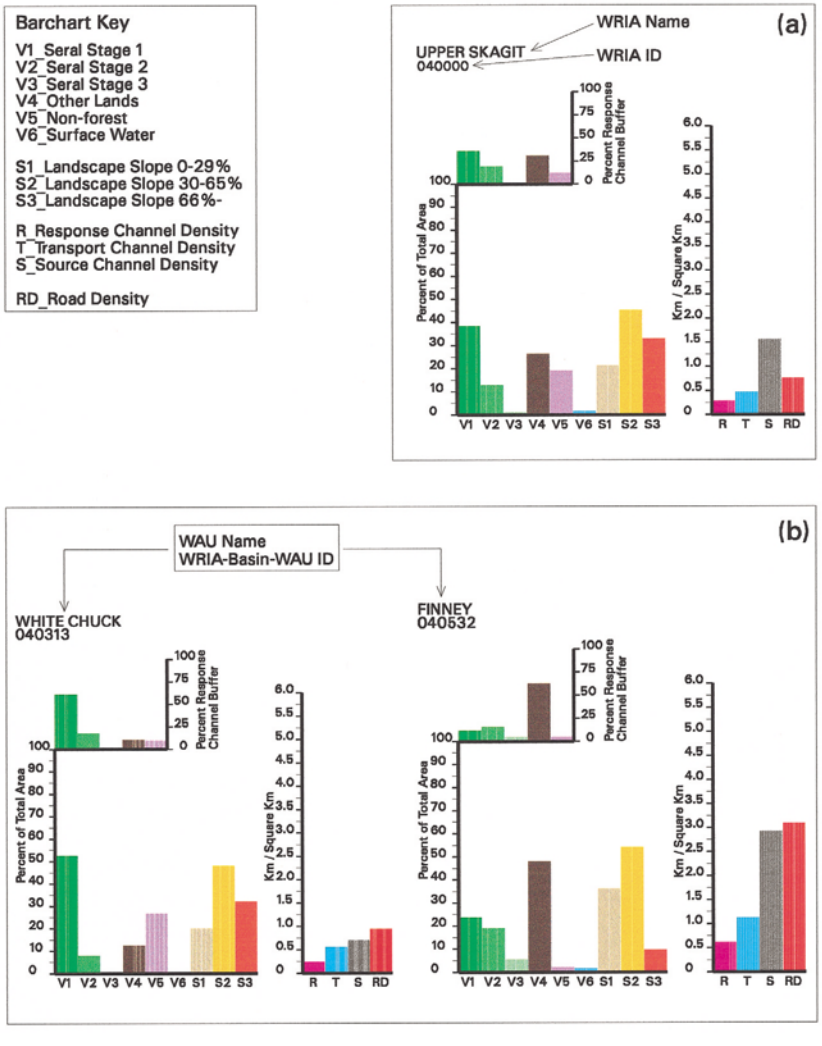


Figure 14.3 (a,b). Example WRIA bar chart (a) data presentation and WAU bar charts (b) standard outputs produced for all of western Washington State.

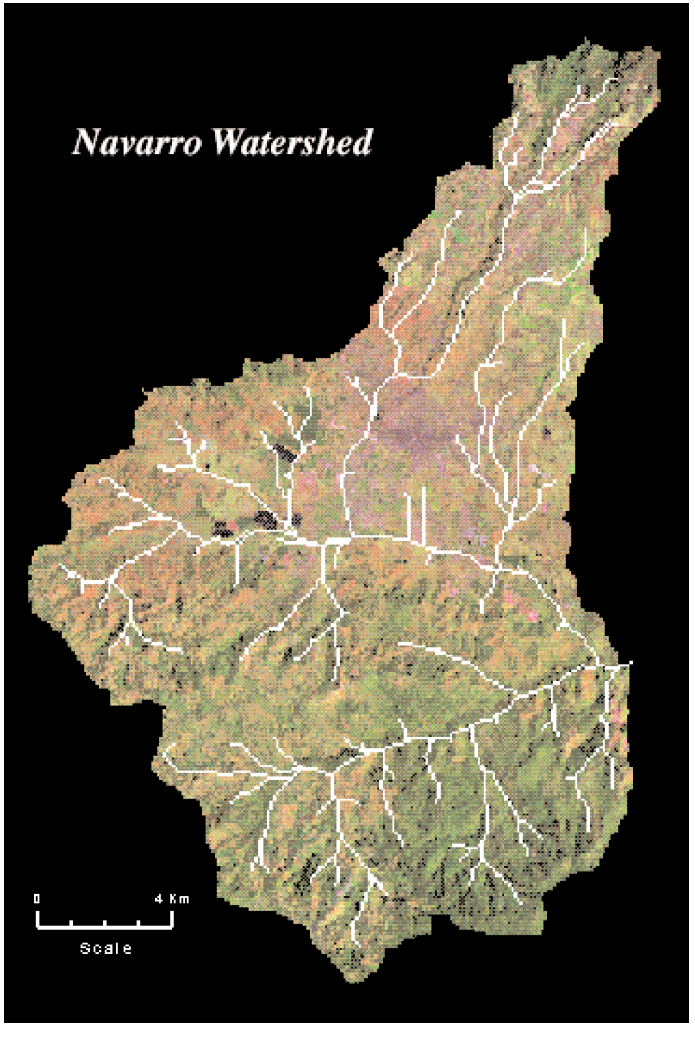


Figure 15.3. Landsat TM image, bands 5,4,3, (with drainage pattern overlay).

CHAPTER 15

Physical Characterization of the Navarro Watershed for Hydrologic Simulation

Jeffrey D. Colby

INTRODUCTION

The Navarro watershed is an important natural resource in Costa Rica (Figure 15.1). The watershed is located at the headwaters of the Reventazón River basin, which may have the greatest hydropower potential of any drainage unit in the country (Quesada, 1979). High sediment loads and streamflow variability affect the proximate Cachí Reservoir (Quesada, 1979; Jansson and Rodriguez, 1992) and additional planned hydroelectric projects. Due to rich volcanic soils, the watershed is an important producer of agricultural products for the country (Cortés and Oconitrillo, 1987). In addition, groundwater is an important water supply source within the watershed.

The city of Cartago is located in the watershed and is part of the Gran Area Metropolitana (GAM). The GAM is the primary population center of the country, and increasing population growth raises important environmental issues (Monzón, 1993). For example, the effects of urban fringe growth on water resources is a concern of the Programa Nacional de Desarrollo Urbano Sostenible (PRODUS).

The 279 km^2 Navarro watershed is one of two located at the headwaters of the Reventazón River basin. The watershed exhibits diverse topography, and particularly dramatic relief (Figure 15.2). The elevation ranges from approximately 3300 m near the summit of the Irazú volcano to 1029 m at the La Troya streamflow gauge.

Management of surface and subsurface water in the Navarro watershed is important for hydroelectric power production and water supply maintenance. Simulating the hydrologic response of the watershed using spatially distributed characteristics could provide a valuable tool for hydrologic research and watershed management. This chapter will describe the physical characterization of the Navarro watershed using geographic information technologies and data to enhance distributed hydrologic modeling efforts. In particular, the processing of digital elevation data will be described.

The hydrologic model used to simulate runoff in the Navarro watershed was the Precipitation Runoff Modeling System (PRMS). The PRMS is a modular modeling system designed to evaluate impacts of climate and land use on surface runoff, sediment yields, and general basin hydrology (Leavesley et al., 1983). Parameterization of PRMS was undertaken based on modeling response units (MRU). MRUs are units of a watershed partitioned on the basis of characteristics such as vegetation type, precipitation distribution, slope, aspect, and soil type. The PRMS was operated within the Modular Modeling System (MMS). The MMS is an integrated computer software sys-

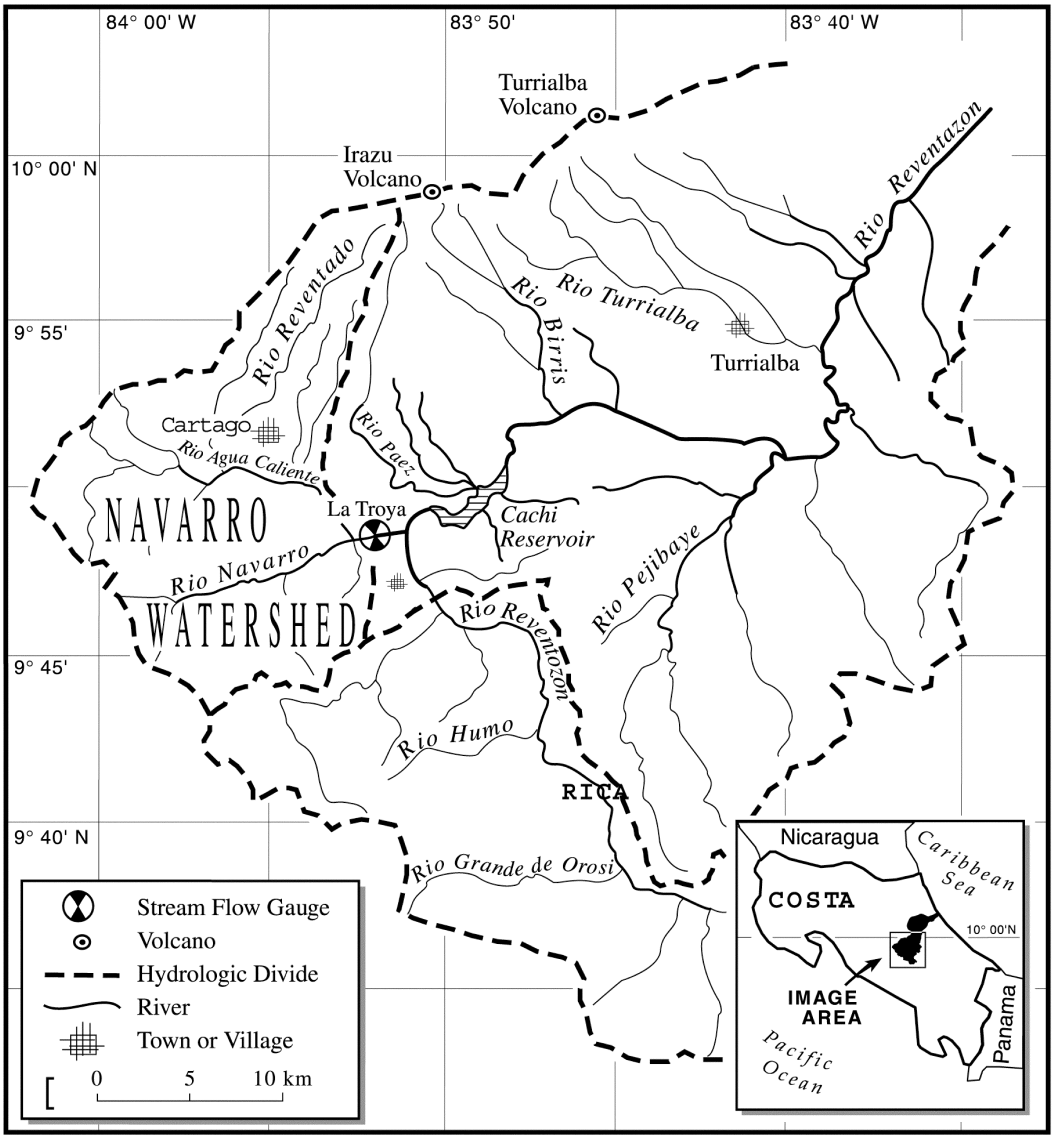


Figure 15.1. The upper watersheds of the Reventazón River Basin (Reprinted by permission, International Journal of Remote Sensing).

tem which provides an operational framework for development of algorithms and their application toward modeling physical processes (Leavesley et al., 1996).

PHYSICAL CHARACTERIZATION

The factors that influenced the selection of data for this study were requirements to complete the study and availability. In comparison to other Central American countries the amount and quality of environmental data available in Costa Rica is relatively high. However, initial assessment of data resources revealed significant gaps in some areas (e.g., soils data) and discrepancies in scales between others.

The availability of data for this watershed may be seen as deficient compared to that required

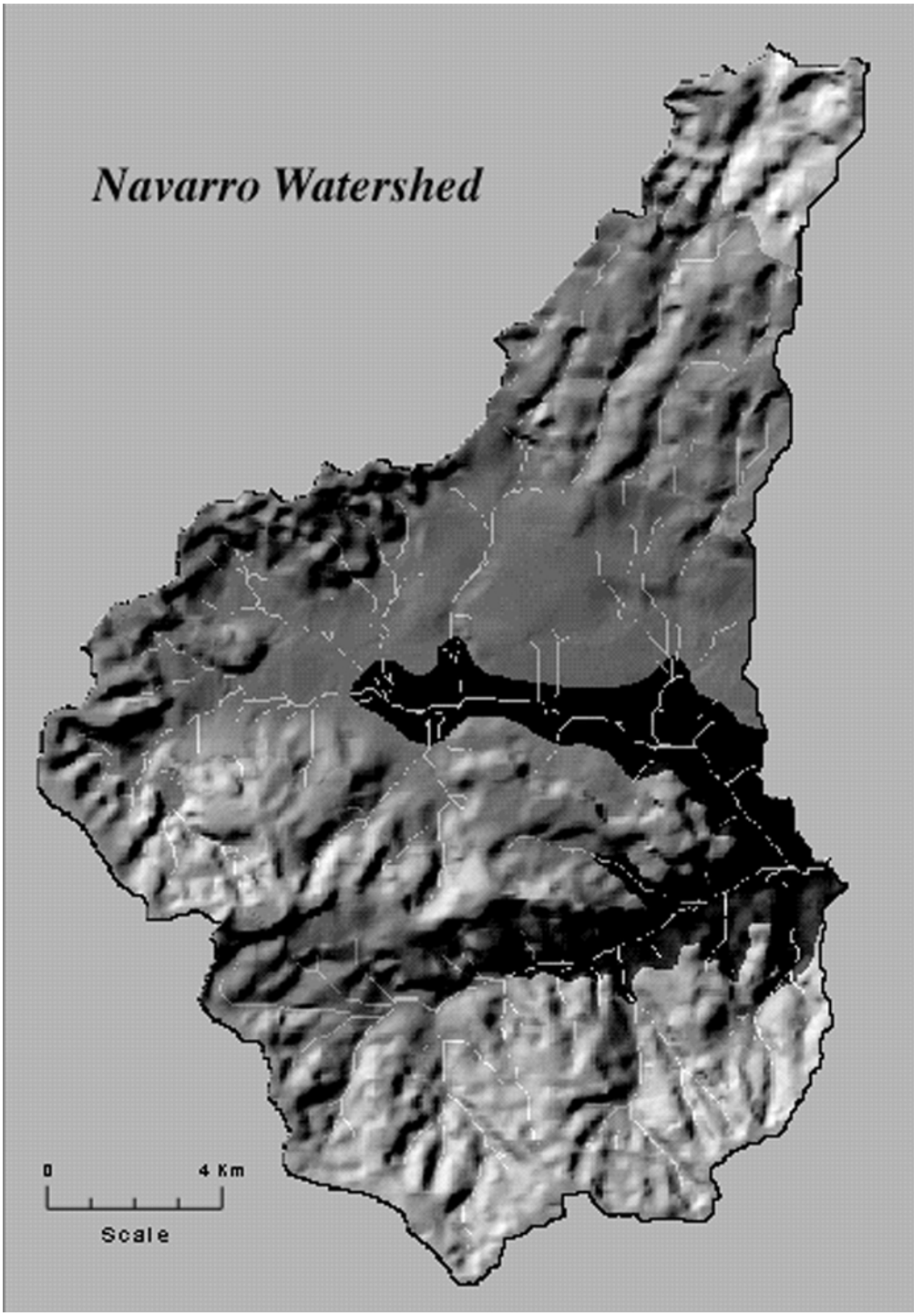


Figure 15.2. Shaded relief image of the Navarro Watershed.

for an experimental watershed in which it is desired to apply a fully distributed hydrologic model. However, this state of data availability lends itself to testing a distributed hydrologic modeling approach utilizing MRUs. In the MRU delineation process used in this study thematic data were aggregated, and homogenous units created from which more generalized parameter values were extracted for modeling purposes. Three types of data were utilized to develop GIS layers for the delineation of MRUs: elevation, digital satellite data, and precipitation.

Digital Terrain Elevation Data Processing

The primary elevation data used was the U.S. National Imagery and Mapping Agency (NIMA) Digital Elevation Terrain Data (DTED). Other elevation data considered bore constraints which were difficult to overcome. For example, topographic maps at 1:50,000 scale which covered the watershed were obtained, but digitizing the entire watershed was judged prohibitive for this study. In addition, complete coverage of the watershed by same scale stereo aerial photographs was not available for construction of a digital elevation model.

The DTED data were distributed in 1° by 1° cells, with a 16-bit range of elevation values. Two cells of DTED Level 1 data were utilized in this study. The boundaries of these cells extended from 9° N to 10° N and from 83° W to 85° W. According to the NIMA, the data for the cells were digitized or scanned from cartographic sources at a scale of 1:250,000, in the mid- to late 1970's. The DTED data were provided in a 3 arc second format. The resolution of this data between 9° N and 10° N latitude was 92.161 m^2. Elevation was represented by a regular grid of post points spaced at 100 m and interpolated with an inverse distance weighted routine using eight neighbors.

Preliminary Steps

A projection file was created in ARC/INFO to enable rectification of the original points from a geographic grid to a Lambert Conformal Conic projection. This projection was chosen because at low latitudes its conformality property possesses true shape of small areas, and also the ancillary maps available for the study area were represented in this projection. To enable integration with additional data sets the DTED file was resampled to 90 m using a bilinear interpolation routine.

Drainage Network and Watershed Delineation

A multiple step process was enacted in ARC/INFO GRID to define the drainage pattern and boundary of the watershed. Essentially four steps were carried out to delineate the drainage network using the elevation data: removing sinks in the DEM; assigning flow direction per cell; assigning flow accumulation values per cell; and determining the threshold flow accumulation value that best represented the drainage pattern (Jenson and Dominique, 1988).

In order to delineate the boundaries of the watershed, the drainage pattern was first displayed and the location of the streamflow gauge estimated. The streamflow gauge location on the derived drainage pattern was accurate in an east-west direction but approximately 550 m south of the actual location, according to a 1:50,000 scale topographic map. The topography in the area where the La Troya streamflow gauge was located opens to relatively level terrain, which may have contributed to the difficulty in defining the location of the gauge. The area draining to the streamflow gauge was identified using the Watershed function in GRID and the flow direction map. The resulting map defined the boundary of the watershed and provided a mask for delineating the boundary on additional GIS data layers.

Physical Characteristics Derived from Elevation Data

Additional analysis steps were carried out to derive physical characteristics of the watershed from the elevation data, including area, elevation parameters, perimeter of the watershed, and drainage density. These parameters were compared with parameters produced in previous studies of the watershed.

In a technical report generated by Solís et al. (1991) and a thesis by Baltodano and Hidalgo (1992), similar physical characteristics were extracted from 1:50,000 scale topographic maps for the Navarro watershed. In an earlier technical report, Elizondo (1979) extracted watershed characteristics from the 1:200,000 scale topographic map, San José CR2-CM–5, published by the Instituto Geographico Nacional (IGN). The report utilized the Puente Negro streamflow gauge which, due to flood damage, was later replaced downstream by the La Troya gauge.

In this study the area of the watershed was calculated using the GRID function Zonalarea. The perimeter was calculated using Zonalperimeter. The elevation parameters, such as maximum, minimum, and mean elevation values, were extracted in ERDAS IMAGINE. These parameters as well as drainage density were compared to estimates from Solis et al. (1991), Elizondo (1979), and estimates from the primary utility company in the country, the Instituto Costaricense de Electricidad (ICE) (Table 15.1).

Table 15.1. Physical Characteristics Derived from Elevation Data

Parameters Station	Colby L.Troya	Solis et al. L.Troya	Elizondo P.Negro	ICE L.Troya	ICE P. Negro
Area (km^2)	278.57	282	273.3	273.6	273.3
Elevation max.	3427	3200	3300	—	—
Elevation min.	1057	1020	—	1028.6	1048.62
Elevation mean	1691	1725	1620	—	—
Perimeter (km)	108.54	87	78	—	—
Drainage density	0.66	0.88	0.46	—	—

Drainage density, or the length of streams per unit area, was calculated using (Black, 1991):

$$D_d = L/A \quad (1)$$

where: D_d = drainage density
L = length of streams (km)
A = area of the watershed (km^2)

Slope values for the watershed were computed in ERDAS IMAGINE and the GIS IDRISI. The following percentage distributions were calculated for the watershed:

(1). 0–4 degrees, 29.5%
(2). 5–9 degrees, 22.8%
(3). 10–15 degrees, 24.9%
(4). 16–19 degrees, 12%
(5). 20–28 degrees, 9.8%
(6). 29–48 degrees, 1%

Elevation Data Quality Assessment

An advantage of using DTED data in this study was the ease with which parameters such as physical characteristics were derived. Use of previously derived digital elevation data provided

flexibility and analysis capabilities in a timely fashion. The trade-off was the lack of quality control in data development.

The drainage pattern derived from the DTED data provided a generally representative depiction of the watershed. However, errors were encountered. Accuracy was assessed by comparisons to drainage patterns from a 1986 Landsat Thematic Mapper (TM) image, 1:10,000 scale land use maps, 1:50,000 scale topographic maps, and aerial photographs. An overlay of the drainage pattern on a false color composite TM image (Figure 15.3, see color section) revealed nonsystematic errors in the correspondence of the location of the drainage pattern, especially along the slopes of the Irazú volcano in the northern section of the watershed. In other areas the drainage patterns matched.

These errors were due in part to the quality of the DTED data and the level of sophistication of GIS software tools used for drainage pattern extraction. The poor representation of the rivers in the northern section of the watershed raised the question of whether the cartographic source for the elevation values was produced before the 1963–1965 eruption of the Irazú volcano.

The most recent eruption of the Irazú volcano began in March of 1963 and continued to spew large volumes of lithic ash through March of 1965. Accumulation of ash on the slopes of the volcano altered the hydrologic regime of the rivers. A hard impervious crust formed on the mantle of ash which reduced infiltration, increased runoff, and resulted in increased slope erosion, frequent flash floods and deadly debris flows. Emergency measures were undertaken such as channel improvements, the construction of levees to protect Cartago, and watershed rehabilitation by terracing, drainage diversion, contour trenching and artificial revegetation (Waldron, 1967).

The area, mean elevation, and drainage density values derived for the Navarro watershed fell between the values produced from analog sources (Table 15.1). The maximum and minimum elevation and perimeter values calculated from the DTED data were somewhat higher than that produced from the 1:50,000 and 1:200,000 scale topographic maps. The absolute vertical accuracy of the DTED data may have affected the differences in elevation values, and the perimeter differences were likely due to grid cells representing the watershed outline, rather than a vector line. Overall, the quality of the elevation data and the derived drainage pattern, though not optimal, supported its application in a distributed hydrologic modeling exercise utilizing MRUs.

Land Cover Classification

The digital Landsat TM satellite image was classified to provide land cover characterization. Variable illumination angles and reflection geometry due to different slope and aspect orientations limit the effectiveness of Landsat TM classification efforts in mountainous terrain. A method for reducing the resulting anisotropic reflectance effects, which was developed in a temperate region (Smith et al., 1980; Colby, 1991; Hodgson and Shelley, 1994) was tested in the neotropical environment of central Costa Rica (Colby and Keating, 1998).

Aerial photographs and land use maps provided reference information for classification of the TM image. Black and white aerial photographs for the north-central section of the watershed (1989) at a scale of 1:20,000 and for the southern section of the watershed (1992) at a scale of 1:60,000 were obtained from the IGN. Also, 1:10,000 scale land use maps (field checked in 1989) were obtained for the central part of the watershed.

Improved classification accuracy was obtained using topographic normalization routines (Colby and Keating, 1998). The resolution and quality of the elevation data was believed to have reduced topographic normalization effectiveness and classification accuracy.

Precipitation Distribution

The Navarro watershed is located between diverse precipitation regimes. One of the driest areas in the country extends from Cartago through the Valle Central (Janzen, 1983). The highest rainfall rates in the Upper Reventazón River basin, 7500 mm a year, are found approximately 25 km southeast of Cartago (Jansson and Rodriguez, 1992). The mean annual precipitation recorded at stations used in this study varied from 1329 mm at the centrally located Comandancia station to 3264 mm at the Belen station located in the southeast near the southern border of the watershed.

To develop an image of the spatial distribution of precipitation across the watershed, a number of trend surface analyses were undertaken using the GIS IDRISI. Second and third order trend surfaces were derived using mean annual rainfall figures calculated from complete years of data for the 20 year period between 1967 through 1986 (IMN, 1988). Trend surfaces were created using data from 7 to 22 stations located in and near the watershed.

Generally, third order surfaces had the highest goodness of fit (R^2) values, but did not represent the spatial distribution of precipitation in the southwest section of the watershed accurately due to a scarcity of precipitation stations in the area. The final image chosen was a second order trend surface created using 14 stations. The R^2 value for this surface was 91.35%.

APPLICATIONS

Once digital spatial characteristics of the watershed had been generated MRUs were delineated using thematic GIS data layers of watershed subbasins, a distance buffer from the stream, precipitation distribution, and land cover categories. The subbasins were delineated using the same techniques as described above to delineate the watershed; however, research and management criteria were taken into consideration in determining outflow points. A distance buffer from the stream was created based roughly on the variable source area concept (Troendle, 1985). The precipitation distribution layer consisted of a three category aggregation of the second order trend surface described above. The land cover layer included the following categories: bare areas, grass, shrubs, trees, and impervious areas. MRUs were delineated using the four thematic layers and a GIS-based methodology (Colby, 1995).

Following delineation of the MRUs, hydrologic simulation of the watershed was undertaken for August through January during the 1987–1988 Costa Rican water year. Parameters for PRMS were extracted using the MRU boundaries, the MRU thematic layers, and additional data layers such as elevation, slope, and aspect. The accuracy of the simulations was determined to be sufficient to proceed with a scaling analysis using multiple resolutions of land cover data.

In the scaling analysis a series of MRU patterns aggregated at 90 m^2 intervals, from 90 m^2 to 1260 m^2, provided areal dimensions to parameterize PRMS. Simulation of hydrologic runoff was undertaken at each resolution. The fractal dimension D values of the MRU patterns were also calculated at each resolution. A strong correspondence was found between the range of resolutions at which accurate hydrologic simulations were achieved and the range of self-similarity of the MRU patterns, as measured by their fractal dimension (Colby, 2001).

DISCUSSION AND CONCLUSION

The data available for this study were sufficient to carry out the intended applications; however, shortcomings did exist. An obvious gap was the lack of soils data available at a useful resolution. The resolution and quality of the elevation data affected the derivation of several watershed characteristics and anisotropic reflectance correction efforts. The status of data available for this study

may be representative of many watersheds in tropical countries in which watershed modeling and management are desired. For example, in these countries topographic maps at a scale larger than 1:200,000 or 1:250,000 may not be available. Physicial characterization of this watershed provided the opportunity to evaluate the effectiveness of utilizing a distributed hydrologic modeling approach based on MRUs. The capability to aggregate thematic data and form hydrologic units provided flexibility to work with less than optimal data. The MRU modeling approach delivered effective watershed characterization and hydrologic simulation accuracy which enabled the desired research to be accomplished. Utilization of geographic information systems provided essential capabilities for improving the processing, management, and analysis of available data.

ACKNOWLEDGMENTS

The author would like to acknowledge the financial support of the Organization of American States, along with the assistance of the U.S. Geological Survey, the U.S. Defense Mapping Agency, and numerous organizations and individuals in Costa Rica for their support in this research effort.

REFERENCES

Baltodano P., J.A., and H. Hidalgo L., 1992. *Definición de Niveles de Inundación en el Río Reventado*. Proyecto de graduacion, Escuela de Ingeniercia Civil, Universidad de Costa Rica, San José, Costa Rica.

Black, P.E., 1991. *Watershed Hydrology*. Prentice-Hall Inc., Englewood Cliffs, NJ.

Colby, J.D., 1991. Topographic normalization in rugged terrain. *Photogrammetric Engineering and Remote Sensing*, 57(5):531–537.

Colby, J.D., 1995. Resolution, Fractal Characterization and the Simulated Hydrologic Response of a Costa Rican Watershed. Ph.D. Dissertation, Geography, University of Colorado, Boulder, CO.

Colby, J.D., 2001. Simulation of Costa Rican Watershed: Resolution Effects and Fractals. *Journal of Water Resources Planning and Management.* 127(4).

Colby, J.D., and Keating, P.L., 1998. Land cover classification using Landsat TM imagery in the tropical highlands: The influence of anisotropic reflectance. *International Journal of Remote Sensing*, 19(8), 1479–1500.

Cortés G., V.M., and G. Oconitrillo C., 1987. Erosion de Suelos Horticulas en el Area de Cot y Tierra Blanca de Cartago. Tesis de Grado para Optar al Grado de Licenciado en Geografia. Departmento de Geografia, Universidad de Costa Rica, San José, Costa Rica.

Elizondo M., J.A., 1979. Estudio Hidrogeológico Preliminar de la Cuenca del Rio Navarro Provincia de Cartago. Servicio Nacional de Aguas Subterraneas Riego y Avenamiento, Departmento de Ingeniería e Hidrología, San José, Costa Rica.

Hodgson, M.E., and B.M. Shelley, 1994. Removing the topographic effect in remotely sensed imagery. *ERDAS Monitor*. 6(1)4–6.

Instituto Meteorológico Nacional (IMN), 1988. Catastro de las Series de Precipitaciones medidas en Costa Rica. San José, Costa Rica.

Jansson, M.B., and A. Rodriguez, Eds., 1992. Sedimentological Studies in the Cachí Reservoir, Costa Rica. Department of Physical Geography, Uppsala University, UNGI Rapport no. 81 Uppsala, Sweden.

Janzen., D.H., 1983. *Costa Rican Natural History*. The University of Chicago Press, Chicago, IL.

Jenson, S., and J. Dominique, 1988. Extracting topographic structure from digital data for geo-

graphic information system analysis. *Photogrammetric Engineering and Remote Sensing*, 54(11):1593–1600.

Leavesley, G.H., R.W. Lichty, B.M. Troutman, and L.G. Saindon, 1983. Precipitation-runoff modeling system — User's manual: U.S. Geological Survey. Water Resources Investigations Report, 83–4238.

Leavesley, G.H., P.J, Restrepo, L.G. Stannard, L.A. Frankoski, and A.M. Sautins, 1996. MMS: A Modeling Framework for Multidisciplinary Research and Operational Applications. In *GIS and Environmental Modeling: Progress and Research Issues,* M.F. Goodchild, L.T. Steyart, and B.O. Parks, Eds., GIS World Books, pp. 155–158.

Monzón P., J.P., 1993. Patrones de Crecimiento del Gran Area Metropolitano. Tesis de Grado para Optar al Grado del Licenciado en Ingenería Civil, Escuela de Ingenería Civil, Universidad de Costa Rica San José, Costa Rica.

Quesada M., C.A., 1979. Effect of Reservoir Sedimentation and Streamflow Modification on Firm Power Generation. Ph.D. Dissertation, Department of Civil Engineering, Colorado State University, Fort Collins, CO.

Smith, J.A., T.L. Lin, and K.J. Ranson, 1980. The Lambertian assumption and Landsat data. *Photogrammetric Engineering and Remote Sensing*, 46(9):1183–1189.

Solis B. H., M. W., Murillo and R. Oreamuno V., 1991. Estudio Hidrologico e Hidráulico para el Control de Inundaciones en la Cuenca del Rio Purires: Valle de Guarco. Servicio Nacional de Aguas Subterraneas Riego y Avenamiento Dirección de Ingeníera, Centro Agronomico Tropical de Investigación y Enseñanza: Proyecto Renarm-Cuencas, Costa Rica.

Troendle, C.A., 1985. Variable source area models. In *Hydrological Forecasting*, M. Anderson, and T. Burt, Eds., John Wiley and Sons, New York, NY, pp. 347–403.

Waldron, H.H., 1967. Debris Flow and Erosion Control Problems Caused by the Ash Eruptions of Irazú Volcano, Costa Rica. Geological Survey Bulletin 1241–1. United States Government Printing Office, Washington, D.C.

CHAPTER 16

Hydrologic Modeling Using Remotely Sensed Databases

James F. Cruise and Richard L. Miller

INTRODUCTION

Recent developments in the evolution of hydrologic modeling using remotely sensed data sources are discussed. The latest remote sensing instruments as well as traditional devices are addressed in terms of spectral characteristics and spatial resolutions. Products of interest to hydrologic modelers are described, as are techniques for their derivation. Hydrologic modeling techniques that make use of these data sources are examined, particularly at the macroscale level. Scientific development in the areas of global change analysis and atmospheric forecasting establish the context of research in hydrologic modeling using spatial databases. Of primary importance are issues involving data scale and resolution requirements for hydrologic modeling. The evolution of this research is discussed and the authors present a study that makes use of recently developed modeling techniques along with remote sensing data products. The results demonstrate that macroscale hydrologic modeling in conjunction with conventional remote sensing data sources can lead to accurate simulations of observed climatologies over fairly long time periods.

Background

Spatial databases are frequently a critical component of hydrologic models. The requirements for modeling over larger spatial domains imposed by the global change and atmospheric forecasting communities have dramatically altered some conceptions relative to modeling strategies and data requirements. The formulation of coupled atmospheric/hydrologic modeling systems to operate in the macroscale spatial domain has placed a premium on data acquisition and processing techniques. The development of comprehensive databases, and hence effective models, is therefore a major focus of most modeling efforts. Remote sensing can provide unique data for constructing spatial databases. These data may include analog aerial photography or digital images representing the emission or reflectance of radiant energy from ground surfaces. Remotely sensed data are available from a variety of instruments mounted on satellite and aircraft platforms. Following acquisition, these data must be processed to correct for the influence of the intervening atmosphere, remapped to a geographic or spatial grid, and analyzed to derive information required specifically for hydrologic modeling.

The purpose of this chapter is to present some of the latest developments in the use of remotely sensed data and geographic information system tools in hydrologic modeling. Sources of remotely sensed data, capabilities of important sensors, acquisition techniques, and hydrologic modeling strategies to make use of these data will be discussed. Data requirements relative to temporal and

spatial scale of modeling domains will be examined, primarily within the context of global change hydrology. Finally, a recent study completed by the authors that made use of many of the data sources and modeling techniques discussed in the chapter will be presented.

DATA REQUIREMENTS FOR HYDROLOGIC MODELING

Hydrologic models require different types of data depending on the processes modeled and the relevant time and space scales of these processes. Models are generally classified temporally as either single event or continuous. Single event models are designed to simulate a single climatic event and thus do not attempt to model hydrologic conditions between events. Continuous models attempt to simulate all hydrologic conditions on a watershed at some specified time step (e.g., hourly or daily) for a long period of time (months or years). In general, continuous simulation models require more extensive climatic data and, possibly, soils and vegetation data sets than do single event models.

Different modeling strategies describe the spatial variation of watershed characteristics at various levels of detail and complexity. Models are classified spatially as either lumped or distributed. The least complex are the lumped parameter models where the spatial variability of select watershed characteristics are described by a few simple parameters (e.g., *lumped* together). Typical parameters include watershed lag (time from centroid of precipitation excess to centroid of runoff hydrograph), time of concentration (travel time from most remote point in the watershed to the basin outfall), and storage factor.

The unit hydrograph method is an outstanding example of a lumped parameter hydrologic model. The more spatially complex distributed models attempt to describe more fully the spatial variation in topography, soils, surface characteristics and meteorology and to explicitly include these in the model. When using this method, the watershed surface is discretized into a spatial grid and the characteristics of the watershed are described within each grid cell. Between the extremes of the lumped and distributed models, several semidistributed techniques have recently been developed which combine the advantages of the distributed models with the simplicity of the lumped parameter method. Of these, the hydrologic response unit technique is quite popular and will be described in further detail in the applications section of this chapter.

All hydrologic models require physical data such as topography and soils characteristics, land-use data such as land cover and vegetation characteristics, and climatic data such as precipitation and temperature. In addition, data quantifying some aspect of streamflow within the watershed are generally required for model calibration and verification.

In constructing databases for hydrologic modeling, consideration must be given to the data requirements for model calibration and the spatial and temporal scale of the model. Products of interest to hydrologic modelers that can be derived from spatial databases include surface representation (slope, aspect, plain geometry), soils characterization (permeability, conductivity, storativity), antecedent moisture conditions, land cover and condition, and vegetation characteristics and biomass. Input parameters are either included directly in the database (i.e., hydraulic conductivity) or calculated from the raw data layers (i.e., slope, land cover). Traditionally, distributed models have employed databases at spatial resolutions on the order of meters. The models may execute on a time scale of minutes or hours, but normally the spatial database is updated on time scales of months or seasons to take into account land use and vegetation changes.

Unfortunately, it is not feasible to construct databases at optimal spatial resolutions for hydrologic modeling on anything but the smallest basin or subbasin scale due to the lack of suitable data and computational resources for the model. Zhang and Montgomery (1994) have suggested a grid resolution of 10 m as appropriate for hydrologic modeling. However, at this resolution the

data are usually not available over large areas. Hydrologic studies are normally conducted using raster (gridded) data sets at spatial resolutions of 30 m or less if possible. More recently, semi-distributed macroscale models constructed at coarser resolutions (1–4 km) have been developed, particularly to function with atmospheric circulation models (Kite et al., 1994; Nijssen et al., 1997).

A number of researchers have conducted investigations of the effect of scale and data resolution on the accuracy of hydrologic models. Wolock and Price (1994) used a topographically driven hydrologic model (TOPMODEL) to investigate the effects of map scale and resolution on hydrologic modeling. They compared results using the 1:24,000 scale USGS maps at grid resolutions of 30 m and 90 m and the 1:250,000 scale, 90 m resolution maps. In general, they found that runoff estimates tended to increase (as did the ratio of surface runoff to total runoff) with increasing map scale and resolution. They attributed this observation to the effect of increasing grid size on the topographical index in the model. However, they did not conclude that the finer scale and resolution maps were necessarily better for hydrologic modeling purposes. Instead, they pointed out that the surface of the saturated soil water zone (which controls runoff generation) may be smoother than fine-scale surface topography, and so would correspond better to coarser resolution data. Likewise, Zhang and Montgomery (1994) found that runoff tended to increase with increasing grid cell sizes in a study based upon topographical information obtained from a 1:4,800 scale base map. In this study grid cell resolutions varied from 2 to 90 m. The authors found that slope representations of an area decreased with increasing grid size, while contributing drainage area increased with grid size. On a similar point, Wood et al. (1988) reported that runoff from areas of variable topography and soils characteristics demonstrated minimum variance at a resolution of about 1 km^2, thus indicating that hydrologic modeling with coarse raster data sets is reasonable. The macroscale model results reported by Kite et al. (1994) and Nijssen et al. (1997) also appeared to show that coarse resolution modeling holds considerable promise.

Representation of basin topography and soil characteristics can be accomplished using either vector- or raster-based data sets. Watershed surface topography is represented by use of a Digital Terrain Model (DTM). The DTM represents the topographic characteristics of the basin through either a grid cell (raster) or data string (vector) formulation (Sole and Valanzano, 1996). The raster-based digital elevation model (DEM) data sets available on-line provide elevation data at grid sizes ranging from 90 to 200 m. Vector-based digital line graph (DLG) data are essentially digitized versions of contour strings from standard topographic maps. The contour intervals range from 5–10 ft (1.5 to 3.0 m) to 50 m.

Surface features may be represented within the GIS by using the raw topographic data in raster or vector format, or by refinements such as the Triangular Irregular Network (TIN). In this method, an area is represented by irregular triangles with the elevation specified at each triangle vertex. The elevation is assumed to vary linearly over the triangle (Singh and Fiorentino, 1996). This formulation allows for the specification of irregular features which might not be possible with gridded data and is commonly employed in hydrologic modeling (Silfer et al., 1987; Maidment et al., 1989; Goodrich et al., 1991; Greene and Cruise, 1996).

Land cover data for hydrologic modeling purposes are usually obtained from interpretation of aerial photographs or multispectral data acquired by satellite or aircraft mounted sensors. High resolution black and white or infrared aerial photographs are available for most of the United States through the U.S. Geological Survey. The photographs are available at scales of 1:30,000 or less for most places. The Survey periodically determines land cover for the U.S. by interpretation of these photographs. Multispectral data sources and products derived from them are discussed later in this chapter.

ON-LINE DATA SOURCES

Several large databases that are available on the internet may be valuable for hydrologic modeling, particular at the macroscale level. In general, these databases are maintained at low spatial resolutions. A valuable source of hydrologic, topographic and land cover data for the United States is the EROS Data Center (EDC) of the U.S. Geological Survey. Topographic data are archived at the EDC in both Digital Elevation Model (DEM) and Digital Line Graph (DLG) format, depending on the location. For instance, the largest scale topographic maps that are commercially available for the U.S. are the 7.5′, 1:24000 scale maps (30m resolution) produced by the U. S. Geological Survey. However, these maps have only been digitized for select areas of the country. Instead, the 15′, 1:250,000 scale digital elevation model (DEM) maps are universally available at the EDC for topography, while the 1:100,000 scale digital line graph maps provide hydrography and transportation data. The grid cell size of the raster-based 15′ DEM data varies from 90 m to 200 m. This type of coarse resolution data may be appropriate for macroscale hydrologic modeling. The EDC is also the repository of the Land Use and Land Cover (LULC) database that is composed of 200 m raster images derived from aerial photographs that cover the entire conterminous United States.

The availability of fine resolution digital soils data for the U. S. is more problematic than are the topogaphic and hydrographic maps. The highest resolution soils data commonly employed in the U.S. are the county soil surveys conducted by the National Resources Conservation Service (NRCS). These surveys are mapped onto aerial photographs at a scale of 1:12,000 or 1:63,360 with individual soil series as the mapping units. NRCS provides soils data in both raster and vector formats at various scales and spatial resolutions. The Soil Survey Geographic Data Base (SSURGO) consists of the digitized county soils maps in vector (DLG) format. These files are released as they are completed and verified and are currently available only in scattered areas of the country.

The State Soils Geographic Database (STATSGO) is a generalization of the county surveys in which the individual mapping units have been aggregated based upon statistical analysis of the soil series. STATSGO is a vector database in which the smallest mapping unit is 1546 acres. The associated data tables include most of the soil matrix and parameter information contained in hard copy surveys. The Penn State University Environmental Sciences Systems Center has developed the Conterminous United States Soils (CONUS) database. CONUS is a raster database at 1 km resolution which represents the STATSGO soils polygons in a standardized 11 layer matrix format.

For land-use determination, various satellite mounted sensors are available that acquire multispectral data at spatial resolutions from 10 m to 1.1 km and temporal resolutions from 6 hours to 16 days. These data sets are used to obtain images of an area of the earth's surface that can be classified in terms of land cover or vegetation characteristics. Classified images are routinely used to obtain land cover information needed in hydrologic modeling such as forest canopy cover and/or biomass, agricultural crop coverage and characteristics (cultivated or noncultivated), urban land use acreages, etc. (Tan and Shih, 1988; Wilkening, 1989; Kite and Kouwen, 1992; Miller and Cruise, 1995). Satellite data are archived in a number of institutions in the United States. The Distributed Active Archive Centers (DAAC) associated with the NASA Earth Observing System Data and Information System (EOSDIS) network are a primary source of satellite data. The EDC serves as the DAAC for the global 1.1 km resolution images from the Advanced Very High Resolution Radiometer (AVHRR) instrument. These images have been archived since 1974. The Marshall Space Flight Center (MSFC) is the primary archive for data associated with the hydrologic cycle.

REMOTE SENSING

A useful operational definition of remote sensing is *to observe an object at some distance without contacting the object*. The large observational scale and unobtrusive nature of remote sensing make this technology a valuable tool for the study of multiscale features of the earth's surface. Although remote sensing cannot provide direct measurements of most hydrologic processes, remote sensing and related technologies such as image processing and Geographic Information Systems (GIS) can provide critical data and products for parameterizing many hydrologic models.

Remotely sensed data are obtained using several different data acquisition technologies. In general, images are formed based on the reflected or emitted energy from a defined spatial element (e.g., pixel, or ground resolution) over select interval(s) of the electromagnetic spectrum. A remote sensing instrument measures the energy at the detectors from either passive (ambient) or active (instrument-generated) energy sources. Although nonimaging instruments such as satellite altimeters do exist, this discussion will focus on the principles and data products derived from imaging instruments, particularly digital multispectral (*few* spectral bands) and hyperspectral (*many* spectral bands) sensors which provide useful data for application in hydrologic modeling. For example, standard CIR (color infrared) aerial photographs, or images, are formed by exposing film to reflected visible and infrared energy. This is an analog process. Important information on landscape patterns, land-use characteristics, and ground cover vegetation can be extracted from such aerial photographs using basic photogrammetric techniques (see for example, Paine, 1981; Graham, and Read, 1986). However, the processing and analysis of photographs using these techniques are time-consuming and labor-intensive. In contrast, the recent rapid development of digital multi- to hyperspectral instruments and low-cost computer systems has established digital remote sensing as an easy-to-use, affordable, and efficient tool for monitoring earth system processes (Miller, 1993; Miller and DeCampo, 1994; Miller et al., 1995).

Multispectral classification is a primary tool for translating remotely sensed data for use in many basic and applied research disciplines. This method is used here to provide a comprehensive example of the processes involved in applying remote sensing to hydrologic models. Multispectral classification is based on a spatial analysis and grouping of surfaces with a similar or correlated reflectance spectrum (i.e., spectral signature). Figure 16.1 shows representative spectra for several surface types along with the spectral bandpass (detection intervals) for the Landsat Thematic Mapper (TM), a satellite-based sensor. A ground point or image pixel is assigned to an image *class* if the pixel's spectral signature is statistically similar to the class mean spectrum. Statistical criteria are derived using multivariate classification methods such as the maximum likelihood classifier, principal component analysis, generalized eigenvalue analysis, and artificial neural networks (Jensen, 1986; Lillesand and Kiefer, 1993; Smith et al., 1994; Venkateswarlu and Singh, 1995; Roger, 1996; Atkinson and Tatnal, 1997; Murai and Omatu, 1997; Paola and Schowengerdt, 1997) These methods yield classified images of land-cover/land-use, soil type and moisture, landscape or drainage patterns, and hydrologic units (Bober et al., 1996; Cialella et al.,1997; Homer et al., 1997; Su et al., 1997; Van Deventer et al., 1997). The digital format of these products allows for a greatly simplified numerical analysis and data extraction process through the use of readily available image processing software. Classification results can then be directly integrated into a GIS, a hydrologic model, or used to derive model coefficients.

The utility of a classified image in hydrologic modeling depends on the accuracy of the classification scheme employed. Classification accuracy is related to the variability in the reflectance spectra for each derived class, band position and width, radiometric sensitivity and noise (e.g., spectral characteristics) of the instrument, and numerical precision of the classification algorithm.

Both the spectral shape and reflectance of a class can vary significantly in time and space as a

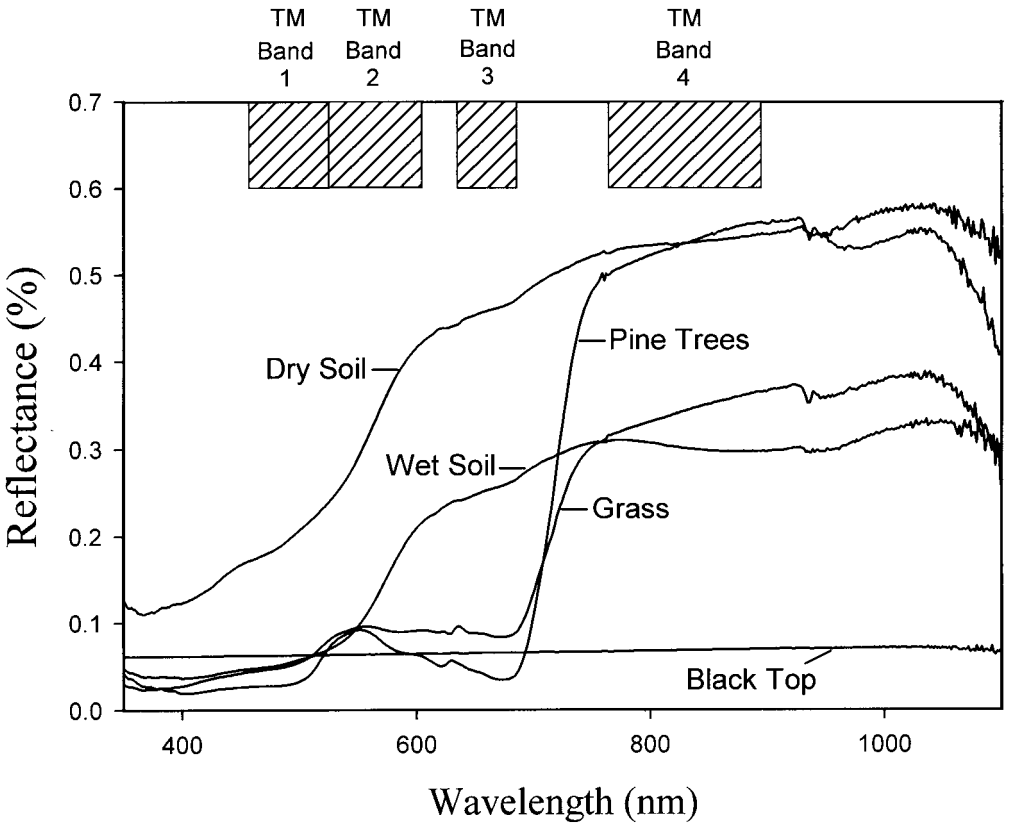


Figure 16.1. Reflectance characteristics of various surfaces observed by Landsat Thematic Mapper.

complex function of many processes. For example, the percent reflectance of vegetation at several wavelengths will increase with decreasing chlorophyll pigment concentration in response to physiological stress (Carter, 1993; Carter and Miller, 1994; Carter et al., 1996). The reflectance of soils in the visible spectrum will change as a function of soil moisture. Surface soil moisture can be measured using passive reflectance and microwave instruments through detecting changes in the soils physical and dialectric properties (Engman, 1990; Jackson et al., 1996; O'Neill et al., 1996). Synthetic Aperture Radar (SAR) systems are also useful in measuring soil moisture (Poncet et al., 1994; Sabburg, 1994; Boisvert et al., 1997; Wegmuller, 1997). The classification of pervious or impervious urban surfaces also changes with surface moisture and temperature. Hence, the ability to detect and resolve changes resulting from intraclass variability from a transition to a different hydrologic unit is an important aspect of current remote sensing research.

Recent advances in remote sensing technology and algorithm development are reducing misclassifications and improving the definition of the hydrologic response of various hydrologic units. Multispectral instruments are generally designed for a specific application. The spectral bands available may be less than optimal for other applications, resulting in large classification errors. An increase in the number of an instrument's spectral bands, as well as narrower band widths, will provide more accurate estimates of the shape (e.g., inflection points) of the surface reflectance spectrum for each class. Narrower bands also help isolate changes in reflectance to specific physical, chemical, and biological processes. Hyperspectral instruments (e.g., AVIRIS, CASI,

Table 16.1. Characteristics of Remote Sensing Instruments

Sensor	Number of Bands	Spectra Range (μm)	Band Width (μm)	Major Products
Advanced Visible/Infrared Imaging Spectrometer (AVIRIS)	224	0.40–2.50	~0.01	1247
Airborne Terrestrial Applications Scanner (ATLAS)	15	0.45–1.75 2.08–4.20 8.20–12.2	~0.1 0.27–0.85 0.40–1.0	
Calibrated Airborne Multispectral Scanner (CAMS)	9	0.45–12.50 2.08–2.35 10.40–12.50	0.06–0.2 (Vis./Near IR) 0.28 (Mid. IR) 2.1 (Thermal IR)	123467
Compact Airborne Spectrograpic Imager (CASI)	288	0.40–1.00	~ 0.002 (adjustable)	12347
European Remote-Sensing Satellite, Microwave Sounder (ERS-1 & ERS-2)	2	23.8 and 36.5 GHz	400 MHz (each band)	4,5
European Remote-Sensing Satellite, Active Microwave Instrument Synthetic Aperture Radar (ERS–1 & ERS–2, AMI-SAR)	C-Band	C-Band (5.3 GHz)	15.55 MHz	45
Hyperspectral Digital Imagery Collection Experiment (HYDICE)	210	0.40–2.50	~ 0.01	1247
Japanese Earth Resources Satellite, Synthetic Aperture Radar (JERS-1 SAR)	L-Band	1.2575 GHz	15 MHz	4
Multispectral Scanner (MSS)	4	0.50–1.10	0.10–0.31	1247
Système Probatoire d'Observation de la Terre (SPOT)	3 1 (Pan)	0.50–0.89 0.50–0.73	~ 0.1 0.22	127
Thematic Mapper (TM)	7	0.45–1.75 2.08–2.35 10.40–12.50	0.06–0.2 (Vis./Near IR) 0.28 (Mid. IR) 2.1 (Thermal IR)	12347
Thermal Infrared Multispectral Scanner (TIMS)	6	8.20–12.20	0.04–0.1	36

1. Land-cover / land-use.
2. Hydrologic units.
3. Surface temperature (thermal analysis).
4. Soils (soil moisture).
5. Precipitation
6. Evapotranspiration.
7. Runoff analysis

HYDICE) with hundreds of bands are effectively imaging spectrometers and may be used for a broad range of applications. The user simply chooses the bands most appropriate for a particular application based on a priori information or comprehensive processing to determine which may reveal information most relevant to a given study. Table 16.1 lists the characteristics of current and planned remote sensing instruments and examples of derived products for hydrologic modeling.

Algorithms based on a combination of bands reduce the effects of nonclass processes. For example, data from two bands, the red and near infrared, are used to compute the Normalized Difference Vegetation Index (NDVI). The NDVI is a simple, yet robust, algorithm to estimate vegetation abundance or *greenness* (Jensen, 1986; Goetz, 1997; Rasmussen, 1997; Ricotta et al., 1997). Similarly, hyperspectral sensors provide the potential to develop highly refined algorithms for classifying ground surfaces and hydrologic processes. Additional research in the spectral response of class components (e.g., plants, soils) to a range of processes will significantly improve

classifications by providing more reliable training fields in supervised (interactive) classification algorithms using hyperspectral data.

Remote sensing technology is well established for measuring hydrologic variables or processes related to transport, snow and ice reservoirs, evaporation, and precipitation. The movement of water through rivers, estuaries, and coastal waters can be monitored using reflectance and thermal imagery. For example, a time series of red and near infrared reflectance images may indicate water movement using suspended particulates as a tracer (Miller and Cruise, 1994; Miller et al., 1994, Schultz, 1996). Multiple images of thermal data can indicate water mass structure, mixing, and flow patterns between areas of different temperatures. Thermal imagery is also used to measure evapotranspiration of forested landscapes (Luvall and Hobo, 1991; Courault et al., 1996) and evaporation in urban environments (Lo et al., 1997). The integration of data from multiple sources and remote sensing can help develop comprehensive models for analyzing the hydrology of a watershed, associated coastal zone, or components of the hydrologic cycle.

CASE STUDY: RIO GUANAJIBO, PUERTO RICO

In this section, we describe a recent study completed by the authors that made use of many of the data sources and hydrologic methods described in the previous sections. A joint project between the NASA Science and Technology Laboratory at Stennis Space Center, MS and the University of Puerto Rico was initiated in 1987 to examine the effects of changing landscapes on the water quality of local rivers and Mayaguez Bay, Puerto Rico (Otero et al., 1992). A particular focus of the study was to examine the correlation of anthropogenic activities such as agricultural practices, deforestation, and urbanization to increased sediment and nutrient-enhanced productivity in the bay. As part of this effort, a mathematical model was used to simulate runoff and sediment yield from the contiguous drainage basins over various periods of time (Cruise and Miller, 1993, 1994; Miller and Cruise, 1995; Mashriqui and Cruise, 1997). Data used during these studies were obtained from several different sources (both in situ and remote sensing) and in varying formats (digital and analog). Remotely sensed data were used in conjunction with soils and topograpic data to provide the basis for the simulation of runoff and sediment yield from the area.

Mayaguez Bay encompasses an area of approximately 100 km^2 on the west coast of Puerto Rico (Figure 16.2). As shown on the figure, three relatively small watersheds drain into the bay; the Anasco (360 km^2), the Guanajibo (311 km^2), and the Yaguez (17.4 km^2). Land-use activities within these basins include various agricultural enterprises (sugar, coffee, and dairy) as well as commercial and industrial developments. Sugar harvesting is confined to the floodplains along the main stems and major tributaries while coffee plantations are located in the upland areas. In addition, large portions of the Anasco and Guanajibo basins consist of undisturbed forested areas. The present study focused on the Guanajibo basin which is composed of agricultural and forested areas with only a small amount of urbanization.

Three streamflow gauging stations are maintained by the U.S. Geological Survey within the Guanajibo basin. Two stations are located on the main stem and one is on Rio Rosario which is the major tributary to the main channel. The Rio Rosario gauge is a continuous stage recorder and water quality station and encompasses a drainage area of 47.4 km^2. This subbasin was the initial focus of the study because in addition to discrete water quality samples, continuous suspended sediment records were available for this gauge. The simulation model was first constructed and verified for this small basin (Cruise and Miller, 1993; Mashriqui and Cruise, 1997) and then extended to the main Guanajibo watershed. Initial Guanajibo simulations covered the period 1988–1990 (Cruise and Miller, 1994), while subsequent simulations were run for the period

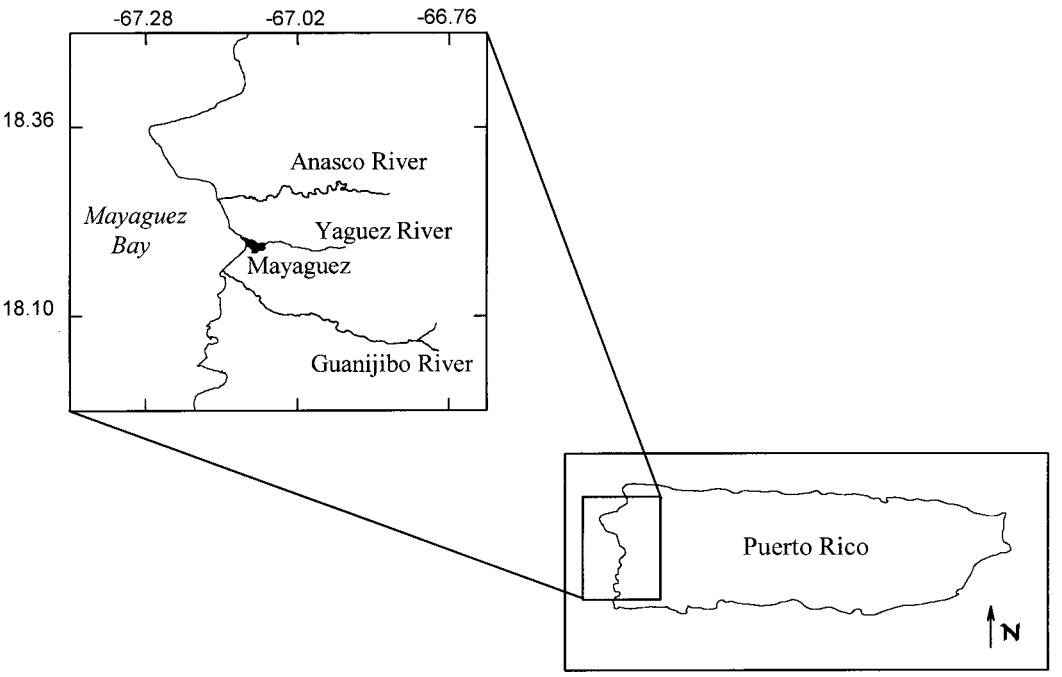


Figure 16.2. Location map: Western Puerto Rico and Mayaguez Bay.

1973–1982 (Miller and Cruise, 1995). Different sources of remotely sensed data were used to derive land cover information for the basin corresponding to these periods of time.

The northern portion of the Guanajibo watershed is composed of steep mountainous ridges (including the Rosario) which comprise about 58% of the total area. Terrain slopes within these regions vary from 32% to 56%. The less significant mountains on the southern rim (average slope = 25%) comprise about 14% of the area. The remaining 28% is comprised of the very mild floodplain area where sugar harvesting is the primary activity. Soil associations in the mountainous areas are generally clays or silty clays while silty clay loams dominate the floodplain region (Gierbolini, 1975).

Digital Data Acquisition and Processing

Data requirements for the hydrologic model included climate records, topography, soil characteristics, land use activities, and ground cover. Regional topography and hydrography (streams) were obtained from DLG files of 1:20,000 scale quadrangle maps available from the U.S. Geological Survey. Contour intervals were 10 m in areas of high relief with intermediate lines of 5 m in the coastal floodplain regions. Soil associations were hand-digitized from published soil surveys available from NRCS.

Estimates of land use and land cover were obtained for the various simulation periods using three sources of remotely sensed data: aerial photography (1975), Calibrated Airborne Multispectral Scanner data (CAMS, 1990), and a Landsat TM image (1985). The photography consisted of black and white images at a scale of 1:30,000 (U.S. Geological Survey). The photos were acquired using an antivignetting filter to minimize tonal variation within a frame. Individual photos were digitized to 256 gray scale levels using a Microtek (PC-based) desktop scanner at a 100 DPI reso-

lution to yield a pixel ground resolution of 7.5 m. Only the portion of each photo with a consistent illumination was scanned. Twenty-six photos were required to provide complete coverage of the Guanajibo watershed. The digitized images were digitally mosaicked using the ELAS (Beverly and Penton, 1989) and Figment (Miller, 1993) image processing software. The mosaicked image was then resampled to a resolution of 30 m to correspond to the CAMS and TM data. Statistical analyses were performed on pixels within areas of known ground cover in order to determine threshold reflectance values (gray scale levels) for separation of classes. The 90th percentile point in each distribution was selected as the threshold reflectance value for that particular class (i.e., the reflectance for which 90% of the pixels were less than that value). Next, a masking routine was applied to the mosaicked image using the threshold gray scale values to separate the image into four relevant classes (e.g., highly reflectant, clear or bare ground, agriculture, and forest) for hydrologic analysis. The resulting image was georeferenced using Universal Transverse Mercator (UTM) map coordinates for the location of gauging stations (highway bridges) identified on the image. The digitized boundaries of the study areas were overlain onto the image to extract statistics and data required for the hydologic model.

Verification of the classification procedure was based on the original aerial photographs. Sample sites of approximately 50 x 50 pixels were selected at 20 random locations within the study area. These sites were then cross-referenced with the photographs. Results indicated that in 95% of the selected sites at least 80% of the pixels were correctly classified. In 50% of the cases, at least 90% of the pixels were correctly classified. Based on these results the errors in the classification procedure applied to the scanned photographs were deemed to be within acceptable limits. The classified image is shown in Figure 16.3a (see color section).

An unsupervised classification was performed on the 1985 Landsat TM data using the ERDAS software package (Figure 16.3b, see color section). A maximum likelihood classification was performed using TM channels 2, 3, and 4 with a maximum of 10 classes specified. Classes representative of hydrologically similar land cover were combined to yield class types consistent with the classification of the digitized photography. Because the TM scene had a number of clouds over the study area, these clouds and their shadows were left in separate classes and were reconciled later.

A land-cover / land-use classification of the CAMS data was created using ELAS (Figure 16.4, see color section). The CAMS data were preprocessed to remove the effects of the atmosphere using Figment. A single flight line of CAMS 30 m data was georeferenced to a UTM projection for classification. ELAS module SRCH provided unsupervised training statistics for 14 classes based on CAMS channels 2 (0.52 to 0.60 mm), 4 (0.63 to 0.69 mm), and 6 (0.76 to 0.90 mm). Using these statistics, the ELAS module CLMAXL produced an image based on a maximum likelihood classifier. The final classified image was then cross-referenced with the CAMS photography to assign specific land-use / land-cover definitions to each class. All digital data were converted to an ELAS image format for final analysis and data extraction for the hydrologic model.

Hydrologic and Sediment Yield Modeling

The hydrologic and sediment yield modeling for the Rosario and Guanajibo watersheds was accomplished using the "grouped response units" technique developed for macroscale modeling by Tao and Kouwen (1989), Kite and Kouwen (1992), and Kite et al., (1994). This method is a semi-distributed technique that makes maximum use of remotely sensed data and GIS capabilities (Cruise and Miller, 1993, 1994; Mashriqui and Cruise, 1997). Database environments utilized in this study included ELAS, ARC/INFO, and Map II (Panzer et al., 1992). The spatial database was used to represent the study area by identification of preliminary computational zones for modeling

purposes based upon relative homogeneity of soils and topographic characteristics. Important soils characteristics include depths of layers corresponding to particular matrix definitions, porosity, and hydraulic conductivity. Topographic parameters include mean slope, stream density, and flow plain geometry. These characteristics were derived from the database and used to define preliminary computational zones (CZ) within the study area. Frequency histograms of hydrologically relevant properties were used to examine spatial homogeneity for CZ definition (Mashriqui and Cruise, 1997). The land cover classes were then overlain onto the computational zones and final hydrologic response units (HRU) were defined as areas of identical land cover. HRU are areas of homogeneous hydrologic characteristics such that the response to climatic forcing would be uniform (Vieux, 1988; See et al., 1992). The assumption is made that areas of equal land cover within a computational zone can be aggregated into HRU regardless of their location within the zone (Kite and Kouwen, 1992; Kite et al., 1994; Mashriqui and Cruise, 1997). Figure 16.5 (see color section) shows the computational zones for the Rosario subbasin overlain onto the land classes derived from the 30 m 1990 CAMS image. The 1985 TM image and the 1975 aerial photo image were employed in a similar manner in the Guanajibo simulations (Miller and Cruise, 1995).

A water budget model was executed once for each land class present in each computational zone (Cruise and Miller, 1993, 1994; Mashriqui and Cruise, 1997). The results from all the HRU were then summed to determine the total response from the zone. The responses from all the zones were summed in turn on a weekly basis to estimate the total basin response at the outfall point. The U.S. Agricultural Research Service model CREAMS (*C*hemical, *R*unoff, *E*rosion from *A*gricultural *M*anagement *S*ystems, Knisel, 1980) was employed for water balance accounting and sediment yield modeling using the methodology outlined above. A particular point of interest is that the CREAMS model uses the Soil Conservation Service curve number method (USDA, 1986) to separate surface runoff from infiltrated water. The curve number is significantly related to the land cover of a particular area. Research has shown that remotely sensed data, particularly TM images, can be employed with great reliability to estimate runoff curve numbers (Ragan and Jackson, 1980; Slack and Welch, 1980; Rango et al., 1983) as was done in this study. Observed runoff and sediment loading from the Rio Rosario gauge was used to calibrate and verify the initial model while runoff data were used to verify the extended Guanajibo model.

RESULTS

The results of the runoff and sediment yield simulations for the initial Rosario simulations are shown in Figures 16.6a, b and 16.6c, d, respectively (Cruise and Miller, 1993). The model was initially calibrated for the period 1986–1989 and then verified compared to observations for 1990–1991. The results of the two sets of runoff simulations for the entire Guanajibo watershed are shown in Figures 16.7 (1988–1990) and 16.8 (1973–1982), respectively (Cruise and Miller, 1994). These results demonstrate that the modeling strategy was fairly consistent in simulating runoff and sediment yield from these watersheds. As with any real-world modeling effort, some inconsistencies in model veracity are inevitable. These irregularities were due primarily to the sparse rain gauge network in the region which caused several storms to be underestimated or missed altogether in the modeling. However, taking these irregularities into account, the model appeared to respond properly to changing soil moisture regimes, thus indicating that the water budget mechanism was adequately calibrated.

The results appear to verify the modeling strategy which was largely based upon the land cover characteristics of the spatial database. The classified spectral images revealed the pattern of land use and land cover throughout the basin and aided in the subdivision of the Rosario and Guanajibo watersheds into computational zones and response units for the model. The results appear to ver-

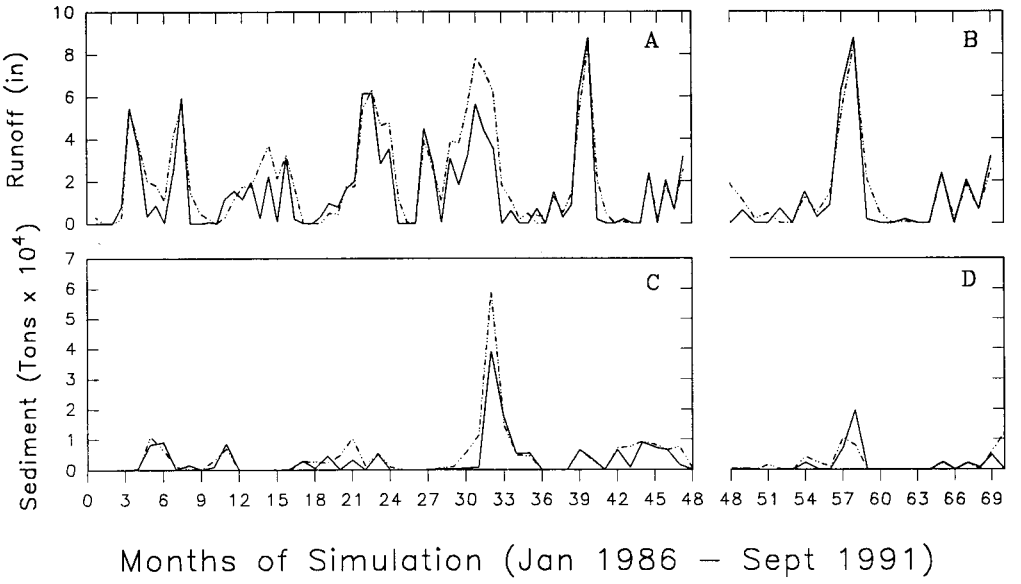


Figure 16.6A, B. Runoff simulation results versus observed data for Rio Rosario subbasin; 1986–1991; A) Calibration runs; B) Verification runs (Cruise and Miller, 1993).
Figure 16C, D. Sediment yield simulation results versus observed data for Rio Rosario subbasin; 1986–1991; C) Calibration runs; D) Verification runs (Cruise and Miller, 1993).

ify that hydrologic variables such as SCS curve numbers and vegetation characteristics can be reliably estimated from remotely sensed images. More significantly, the results support the contention that remotely sensed data from different sources can be integrated into a consistent data base for hydrologic modeling. The land-cover data were obtained from different sources corresponding to the periods of time when particular sources were available. Of particular note are the results of the 1973–1982 simulations (Figure 16.8) that were based upon the land cover obtained from the classified aerial photographs and the 1985 TM image. The simulations for this period were as accurate (r = 0.89) as were the more recent simulations (r = 0.87) for the Guanajibo based on the CAMS images (Figure 16.7).

CONCLUSIONS

Remotely sensed data integrated into a GIS environment has been shown to be a valuable asset in hydrologic modeling. In fact, recent developments have shown that remotely sensed data and GIS tools are virtually indispensable resources for modeling over large spatial domains. The development of macroscale hydrologic models, either to serve as land surface components of atmospheric circulation models, or to function in conjunction with such models, is largely predicated upon remote sensing/GIS principles. Remotely sensed data makes it possible to update hydrologic parameters at relevant temporal scales (monthly, seasonally, annually) over very large spatial domains. The increased speed, efficiency, and cost-effectiveness of GIS and image processing hardware and software have brought these capabilities down to the desktop level. In addition, the products derived from remotely sensed data, as well as the raw images themselves (many of which are available on-line at no charge) are available to virtually every hydrologic modeler. A large variety of remote sensing instruments are currently available with even more planned for launch in

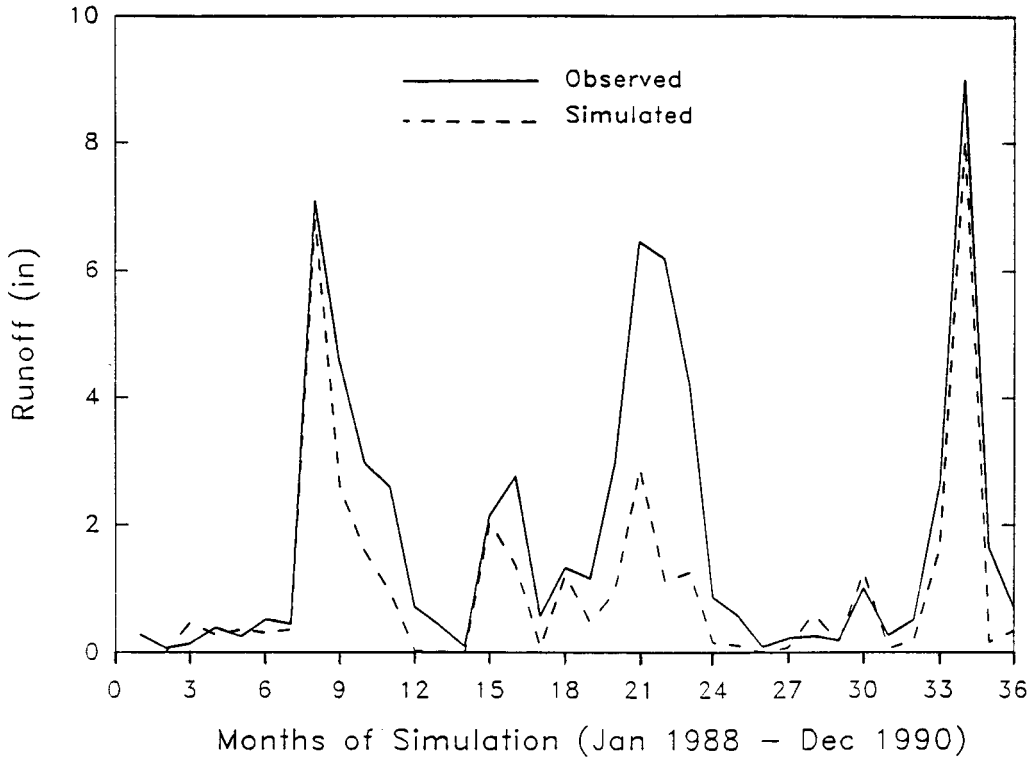


Figure 16.7. Runoff simulation results versus observed data for Rio Guanajibo, Puerto Rico; 1988–90 (Cruise and Miller, 1994).

the near future. These new hyperspectral instruments will greatly increase the use of remotely sensed data in hydrology by providing the basis for more accurate estimation of such products as soil moisture, snow cover, precipitation distribution, land cover, and vegetation biomass. Advances in hydrologic modeling methods to make use of these data at varied spatial and temporal resolutions will continue to be a major focus of research in the hydrologic sciences.

ACKNOWLEDGMENTS

A large portion of the research on hydrologic processes in Puerto Rico was conducted while the senior author was supported by a joint NASA/National Research Council Associateship at Stennis Space Center (SSC), MS. Dr. Armond Joyce, Education Affairs officer at SSC, provided assistance during this period. Additional financial support for this work was provided by NASA under grant NAGW–849 from the office of University Affairs, NASA Headquarters. Mr. Hassan Mashriqui of Louisiana State University, Baton Rouge, LA aided in classification of the Landsat TM data. We thank Jim Anderson, Greg Booth, Greg Carter, and Marco Giardino for their comments on an earlier version of this chapter.

REFERENCES

Atkinson, P.M., and A.R.L. Tatnal, 1997. Introduction: Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4): 699–710.

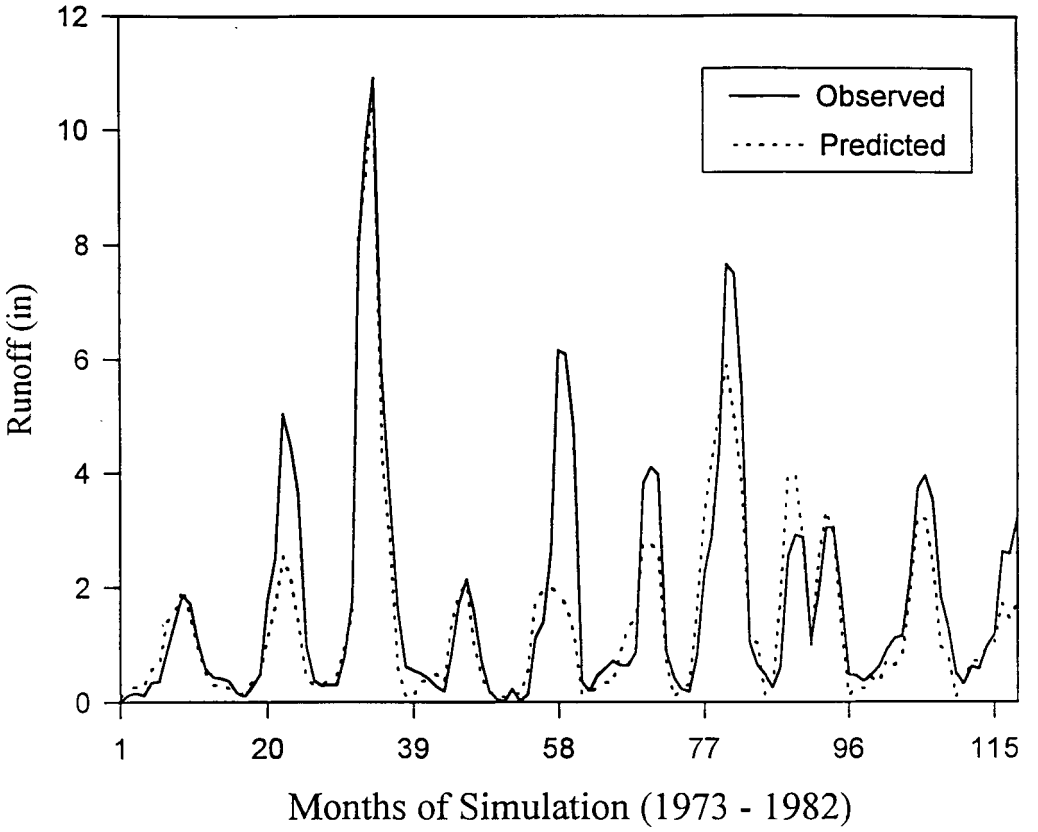


Figure 16.8. Runoff simulation results versus observed data for Rio Guanajibo, Puerto Rico; 1973–82 (Miller and Cruise, 1995).

Beverly, A.M., and P.G. Penton, 1989. ELAS Users Reference, Volume II, Unpublished Report, Science and Technology Laboratory, John C. Stennis Space Center, MS.

Bober, M.L., D. Wood, and R.A. McBride. 1996. Use of digital image analysis and GIS to assess regional soil compaction risk. *Photogrammetric Engineering and Remote Sensing*, 62(12): 1397–1404.

Boisvert, J.B., Q.H.J. Gwyn, A. Chanzy, D.J. Mayor, B. Brisco, and R.J. Brown. 1997. Effects of soil moisture gradients on modelling radar backscattering from bare fields. *International Journal of Remote Sensing*, 18(1): 153–170.

Carter, G.A., W.G. Cibula, and R. L. Miller. 1996. Narrow-band reflectance imagery compared with thermal imagery for early detection of plant stress. *Journal of Plant Physiology*, 148: 515–522.

Carter, G.A., and R.L. Miller. 1994. Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing Environment*, 50: 295–302.

Carter, G.A., 1993. Responses of leaf spectral reflectance to plant stress. *American Journal of Botany*, 80(3): 239–243.

Cialella, A.T., R. Dubayah, W. Lawrence, and E. Levine, 1997. Predicting soil drainage class using remotely sensed and digital elevation data. *Photogrammetric Engineering and Remote Sensing*, 63(2):171–178.

Courault, D., B. Aloui, J.P. Lagouarde, P. Clastre, H. Nicolas, and C. Walter, 1996. Airborne thermal data for evaluating the spatial distribution of actual evapotranspiration over a watershed in oceanic climatic conditions—application of semi-empirical models. *International Journal of Remote Sensing,* 17(12): 2281–2302.

Cruise, J.F., and R.L., Miller. 1993. Hydrologic modeling with remotely sensed databases. *Water Resources Bulletin*, 29(6): 997–1002.

Cruise, J.F., and R.L. Miller, 1994. Hydrologic modeling of land processes in Puerto Rico using remotely sensed data. *Water Resources Bulletin*, 30(3): 419–428.

Engman, E.T., 1990. Progress in microwave remote sensing of soil moisture. *Canadian Journal of Remote Sensing*, 16: 6–14.

Gierbolini, R.E., 1975. Soil Survey of Mayaguez Area of Western Puerto Rico. United States Department of Agriculture, Conservation Research Report No. 26.

Goetz, S.J., 1997. Multi-sensor analysis of NDVI, surface temperature and biophysical variables at a mixed grassland site. *International Journal of Remote Sensing,* 18(1): 57–70.

Goodrich, D.G., D.A. Woolhiser, and T.O. Keeper, 1991. Kinematic routing using finite elements on a triangular irregular network. *Water Resources Research*, 27(6): 995–1003.

Graham, R., and R. Read, 1986. *Manual of Aerial Photography*. Focal Press, Boston, MA.

Greene, R.G, and J.F. Cruise, 1996. Development of a geographic information system for urban watershed analysis. *Photogrammetric Engineering and Remote Sensing*, 62(7): 863–870.

Homer, C.G., R.D. Ramsey, T.C. Edwards, Jr., and A. Falconer, 1997. Landscape cover-type modeling using a multi-scene thematic mapper mosaic. *Photogrammetric Engineering and Remote Sensing,* 63(1): 59–68.

Jackson, T.J., J. Schmugge, and E.T. Engman. 1996. Remote sensing applications to hydrology: soil moisture. *Hydrological Sciences Journal,* 41(4): 637–647.

Jensen, J.R., 1986. *Introductory Digital Image Processing: A Remote Sensing Perspective.* Prentice-Hall, Englewood Cliffs, NJ.

Kite, G.W., and N. Kouwen, 1992. Watershed modeling using land classifications. *Water Resources Research*, 28(12): 3193–3200.

Kite, G.W., A. Dalton, and K. Dion, 1994. Simulation of streamflow in a macroscale watershed using general circulation model data. *Water Resources Research*, 30(5): 1547–1559.

Knisel, W.G., Ed., 1980. CREAMS, a Field Scale Model for Chemical, Runoff and Erosion from Agricultural Management Systems. Conservation Research Report No. 26, United States Department of Agriculture, Washington, D.C.

Lillesand, T.M., and R.W. Kiefer, 1993. *Remote Sensing and Image Interpretation*. 3rd ed., John Wiley and Sons, New York, N.Y.

Lo, C.P., D.A. Quattrochi, and J.C. Luvall. 1997. Application of high-resolution thermal infrared remote sensing and GIS to assess the urban heat island effect. *International Journal of Remote Sensing,* 18(2): 287–304.

Luvall, J.C., and H. R. Holbo, 1991. Thermal remote sensing methods in landscape ecology. In *Quanitiative Methods in Landscape Ecology*, M.G. Turner and R.H. Gardner, Eds., Springer, Verlag, New York, NY, pp. 127–152.

Maidment, D.R., D. Djokic, and K.G. Lawrence, 1989. Hydrologic modelling on a triangulated irregular network. EOS, 70(43):1091.

Mashriqui, H.S., and J.F. Cruise, 1997. Sediment yield modeling by grouped response units. *Journal of Water Resources Planning and Management*, ASCE, 123(2): 95–104.

Miller, R.L., J. DeCampo, T. Leming, and K. Hughes, 1995. Affordable solutions for analyzing coastal processes using remotely sensed data, *Proceedings of the Third Thematic Conference, Remote Sensing for Marine and Coastal Environments*, Seattle, WA, pp. 308–314.

Miller, R.L., J.C. Cruise, E. Otero, and J.M. Lopez. 1994. Monitoring suspended particulate matter in Puerto Rico: Field measurements and remote sensing. *Water Resources Bulletin,* 30(2): 271–282.

Miller, R.L., and J.D. DeCampo, 1994. C COAST: A PC-based program for the analysis of coastal processes using NOAA CoastWatch data. *Photogrammetric Engineering and Remote Sensing*, 60(2):155–160.

Miller, R.L. 1993. High resolution image processing on low cost microcomputers. *International Journal of Remote Sensing*, 14:655–667.

Miller, R.L., and J.F. Cruise, 1995. Effects of suspended sediments on coral growth: evidence from remote sensing and hydrologic modeling. *Remote Sensing of Environment*, 53: 177–187.

Murai, H., and S. Omatu, 1997. Remote sensing image analysis using a neural network and knowledge-based processing. *International Journal of Remote Sensing*, 18:655–667.

Nijssen, B, D.P. Lettenmaier, X. Liang, S.W. Wetzel, and E.F. Wood, 1997. Streamflow simulation for continental-scale river basins. *Water Resources Research*, 33(4): 711–724.

O'Neill, P., N. Chauhan, and T. Jackson. 1996. Use of active and passive microwave remote sensing for soils moisture estimation through corn. *International Journal of Remote Sensing,* 17: 1851–1865.

Otero, E., R.L. Mille, and J.M. Lopez. 1992. Remote sensing of chlorophyll and sediments in coastal waters of Puerto Rico. *Proceedings, First Thematic Conference on Remote Sensing for Marine and Coastal Environments*, New Orleans, LA.

Paine, D.P., 1981. *Aerial Photography and Image Interpretation of Resource Management*. John Wiley and Sons, New York, NY.

Panzer, M., K.C. Kirby, and N. Thies, 1992. *Map II: Reference Manual*. John Wiley and Sons, New York, NY.

Paola, J.D., and R.A. Schowengerdt, 1997. The effect of neural-network structure on a multispectral land-use / land-cover classification. *Photogrammetric Engineering and Remote Sensing*, 63(5):535–544.

Poncet, F.V., C. Prietzsch, and M. Tapkenhinrichs, 1994. Regionalization of Soil Physical Parameters Using ERS–1 PRI SAR Data. IGARSS'94.

Ragan, R.M., and T.J. Jackson, 1980. Runoff synthesis using Landsat and SCS model. *Journal of the Hydraulics Division, ASCE*, 106(HY5):667–678.

Rango, A., A. Feldman, T.S. George, III, and R.M. Ragan, 1983. Effective use of Landsat data in hydrologic models. *Water Resources Bulletin*, 19:165–174.

Rasmussen, M.S., 1997. Operational yield forcast using AVHRR NVDI data: Reduction of environmental inter-annual variability. *International Journal of Remote Sensing*, 18(5):1051–1058.

Ricotta, C., G.C. Avena, and F. Ferri, 1997. Analysis of human impact on a forested landscape of central Italy with a simplified NDVI texture descriptor. *International Journal of Remote Sensing*, 17(14): 2869–2874.

Roger, R.E., 1996. Principal components transform with simple, automatic noise adjustment. *International Journal of Remote Sensing*, 17:2719–2728.

Sabburg, J.M., 1994. Evaluation of an Australian ERS–1 SAR Scene Pertaining to Soil Moisture Measurement. IGARSS'94.

Schultz, H.,1996. Shape reconstruction from multiple images of the ocean surface. *Photogrammetric Engineering and Remote Sensing*, 62(1):93–101.

See, R.B., D.L. Naftz, and C.L. Qualls, 1992. GIS-assisted regression analysis to identify sources of selenium in streams. *Water Resources Bulletin,* 28(2):315–330.

Silfer, A.T., G.J. Kinn, and J.M. Hassett, 1987. A geographic information system utilizing the triangulated irregular network as a basis for hydrologic modeling. *Auto Carto 8, Proceedings of*

the 8th International Symposium on Computer Assisted Cartography, N.R. Chrisman, Ed., American Society of Photogrammetry and Remote Sensing/American Congress of Surveying and Mapping, pp. 129–136.

Singh, V.P., and M. Fiorentino, 1996. Hydrologic modeling with GIS. *Geographical Information Systems in Hydrology*. V.P. Singh and M. Fiorentino, Eds., Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 1–13.

Slack, R.B., and R. Welch, 1980. Soil Conservation Service runoff curve number estimates from Landsat data. *Water Resources Bulletin*, 16(5): 887–893.

Smith, C., N. Pyden, and P. Cole, 1994. *ERDAS Field Guide*, 3rd ed., ERDAS, Inc., Atlanta, GA.

Sole, A., and A. Valanzano, 1996. Digital terrain modelling. *Geographical Information Systems in Hydrology.* V.P. Singh and M. Fiorentino, Eds., Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 175–194.

Su, Z., P.A. Troch, and F.E. de Troch, 1997. Remote sensing of bare surface soil moisture using EMAC/ESAR data. *International Journal of Remote Sensing*, 18: 2105–2104.

Tan, C.H., and S.F. Shih, 1988. A geographic information system for study of agricultural land use changes in St. Lucie county, Florida. *Soil Crop Science Society of Florida, Proceedings*, 47:102–105.

Tao, T., and N. Kouwen, 1989. Remote sensing and fully distributed modeling for flood forecasting. *Journal of Water Resources Planning and Management*, *ASCE*, 115(6): 809–823.

United States Department of Agriculture, 1986. Urban Hydrology for Small Watersheds, Technical Release 55. United States Department of Agriculture, Soil Conservation Service, Washington, D.C.

Van Deventer, A.P., A.D. Ward, P.H. Gowda, and J.G. Lyon, 1997. Using Thematic Mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing*, 63(1):87–92.

Venkateswarlu, N. B., and R.P. Singh, 1995. A fast maximum likelihood classifier. *International Journal of Remote Sensing*, 16:313–320.

Vieux, B.E., 1988. Finite Element Analysis of Hydrologic Response Areas Using Geographic Information Systems. Ph.D. Dissertation, Michigan State University, East Lansing, MI.

Wegmuller, U, 1997. Soil moisture monitoring with ERS SAR interferometry. *Proceedings ERS Symposium, Florence'97*.

Wilkening III, H.A., 1989. Landsat data processing and GIS for regional water resources management in northeast Florida. *Proceedings, GIS/LIS 1989, American Congress of Surveying and Mapping*, pp. 110–119.

Wolock, D.M., and C.V. Price, 1994. Effects of digital elevation model map scale and data resolution on a topography-based watershed model. *Water Resources Research*, 30(11): 3041–3052.

Wood, E.F., M. Sivapalan, K.J. Beven, and L. Band, 1988. Effects of spatial variability and scale with implications to hydrologic modeling. *Journal of Hydrology*, 102: 29–47.

Zhang, W., and D.R. Montgomery, 1994. Digital elevation model grid size, landscape representation, and hydrologic simulations. *Water Resources Research*, 30(4):1019–1028.

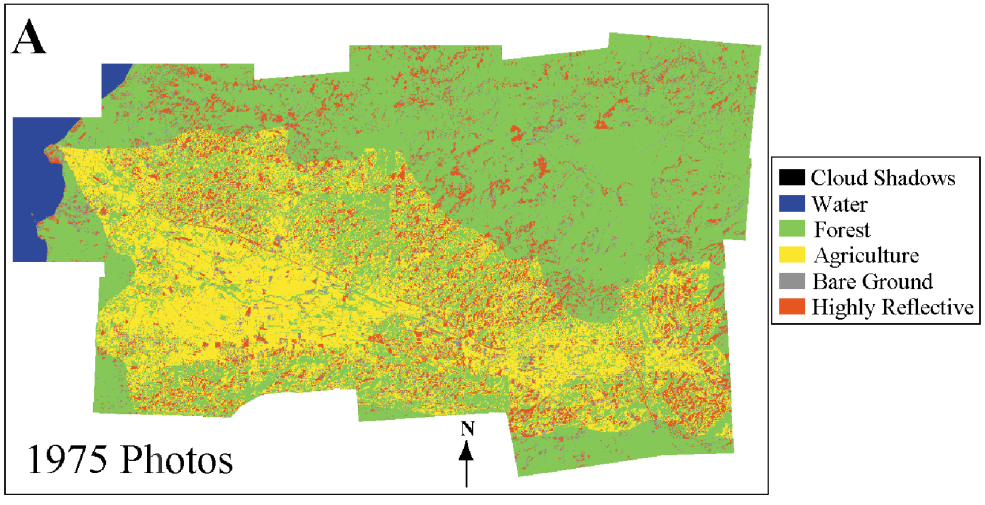


Figure 16.3A. Classified land-cover image of western Puerto Rico derived from 1975 black and white aerial photographs (Miller and Cruise, 1995).

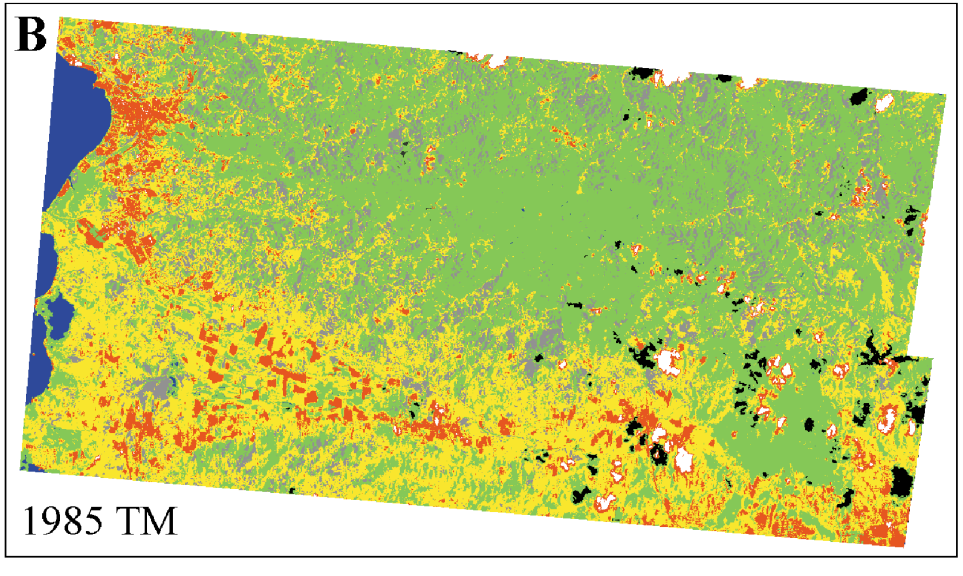


Figure 16.3B. Classified land-cover image of western Puerto Rico derived from 1985 Landsat Thematic Mapper image (Miller and Cruise, 1995).

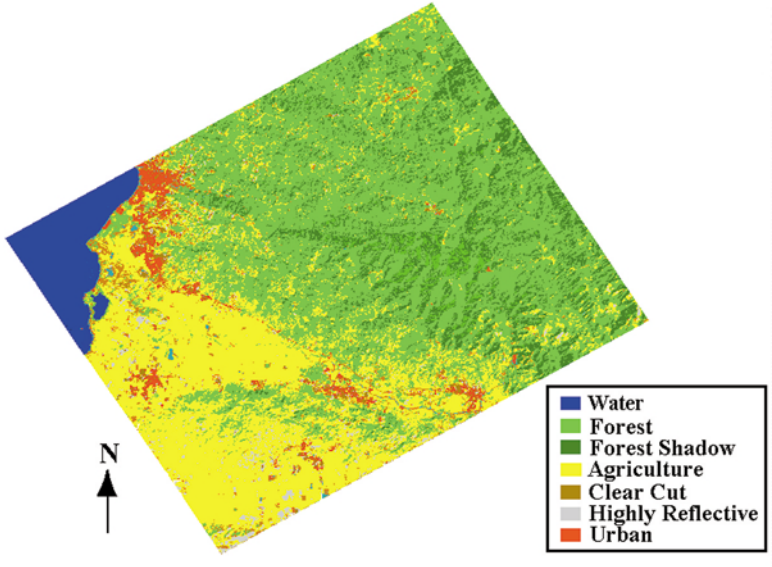


Figure 16.4. Classified land-cover image of western Puerto Rico derived from 1990 CAMS data (Cruise and Miller, 1993).

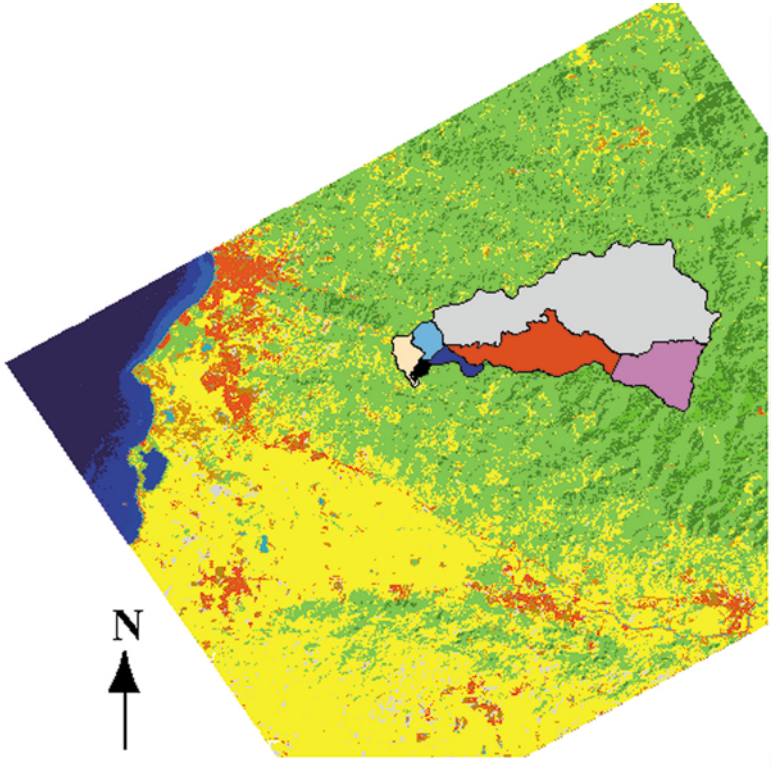


Figure 16.5. Computational zones of Rio Rosario subbasin overlain on classified 1990 CAMS image (Cruise and Miller, 1993).

CHAPTER 17

Technological Advances in Automated Land Surface Parameterization from Digital Elevation Models

Jurgen Garbrecht, Lawrence W. Martz, and Patrick J. Starks

INTRODUCTION

Topography plays an important role in the distribution and flux of water and energy within natural land surfaces. Classical examples include surface runoff, evaporation, infiltration, and heat exchange which are hydrologic processes that take place at the ground-atmosphere interface. The quantitative assessment of the processes depend on the topographic configuration of the land surface, which is one of several controlling boundary conditions. Many topographic parameters can be computed directly from a Digital Elevation Model (DEM) (Band, 1986; Jenson and Domingue, 1988; Mark, 1988; Martz and Garbrecht, 1992, 1993; Moore et al., 1991; Tarboton et al., 1991; Wolock and McCabe, 1995). This automated extraction of topographic parameters from DEMs is recognized as a viable alternative to traditional surveys and manual evaluation of topographic maps, particularly as the quality and coverages of DEM data increase. Manual evaluation of topography is general tedious, time-consuming, error-prone, and often subjective (Richards, 1981).

This chapter presents four advances in computerized methods to extract topographic parameters from DEMs. The first two advances address the treatment of depressions and flat surfaces in the DEM. Most existing methods for handling depressions and flat areas in DEM processing for drainage analysis are based on some common and fundamental assumptions about the nature of these features. These assumptions are largely implicit to the methods and are usually not recognized explicitly. These are: (1) that closed depressions and flat areas are spurious features that arise from data errors and limitations of DEM resolution; (2) that flow directions across flat areas are determined solely by adjacent cells of lower elevation; and (3) that closed depressions are caused exclusively by the underestimation of DEM elevations. While the first of these seems reasonable, the others are not. It is possible to make more reasonable assumptions about the controls on flow direction and the cause of closed depressions and to incorporate these assumptions into new algorithms for handling difficult topographic situations encountered in raster DEM processing for drainage analysis. The two new algorithms presented here are based on a deductive but qualitative assessment of the most probable nature of depressions and flat areas in raster DEM.

The last two advances address the identification of the topology of the channel network from raster images, and the parameterization of irregular overland or hillslope areas. The topology of the channel network is captured in terms of network node indexing and channel ordering by the Strahler method (Strahler, 1957). Channel ordering and node indexing is fundamental to the automation of flow routing management in distributed surface hydrology models and morphometric

evaluation of channel network structure. The node index numbers can also serve to link network nodes and channel data stored in tabular format. The parameterization of subcatchments quantifies the length, width, and slope of rectangular planes representing irregularly shapes overland and hillslope contributing area. This rectangular subcatchment conceptualization is often used in distributed modeling of hillslope runoff and erosion processes. The presented subcatchment parameterization algorithms are an important contribution to traditional watershed modeling because the algorithms automate a task that is subjective and requires experience in interpretation and conceptualization of irregular hillslope features.

In the following section the fundamental principles underlying the algorithms and the essential components of the algorithms are presented. Related discussions in the broader context of water resources can be found in Garbrect and Martz (1999a). Issues relating to technical details and the implementation of these algorithms into digital land surface processing exceed the framework of this chapter. The presented technological advances are incorporated in the topographic parameterization model TOPAZ (TOpographic PArameteriZation) which automatically segments and parameterizes watersheds from DEMs for water resources, hydraulic, and hydrologic applications (Garbrecht and Martz, 1999b, 2000).

TREATMENT OF SPURIOUS DEPRESSIONS IN DEMS

Depressions are groups of raster cells completely surrounded by other cells of a higher elevation. They are usually artifacts that arise from data inaccuracies, interpolation procedures, and limited horizontal and vertical resolution of the DEM (Mark, 1983, 1988; Tribe, 1992; Zhang and Montgomery, 1994). They represent a major difficulty for DEM processing procedures that are based on the downslope flow routing concept (Martz and Garbrecht, 1992) because the existence of a downslope flow path at every cell is assumed. In the case of a depression there is, by definition, no outflow, and procedures based on an assumed downslope flow path are bound to fail.

The traditional solution to this problem is to remove all depressions in the DEM by raising the elevations within the depression to the elevation of its lowest outlet. This procedure is called "filling" of the depression. Two assumptions are implicit to this approach: (1) depressions are spurious features that arise from interpolation errors or insufficient precision in elevation values; and, (2) all depressions are due to the underestimation of elevation and should be filled. The practice of eliminating depressions solely by filling is likely to introduce systematic bias into the modified DEM. In reality, elevation errors in the DEM are as likely to result from elevation overestimation as from underestimation, and some depressions arise from the obstruction of flow paths by over-

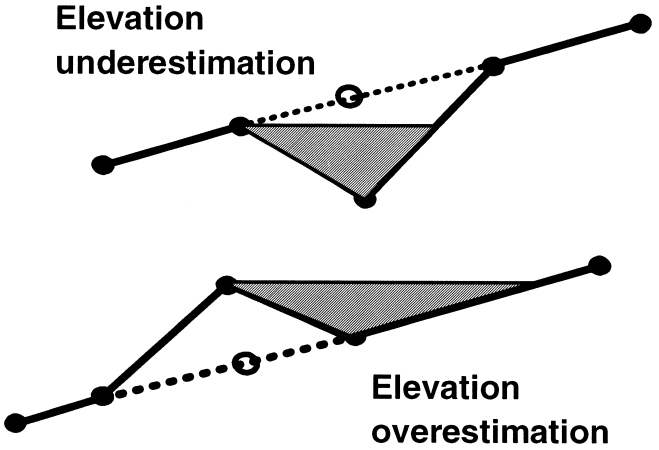


Figure 17.1. Two dimensional schematic profiles illustrating depressions arising from elevation underestimation and elevation overestimation (figure from Martz and Garbrecht, 1997).

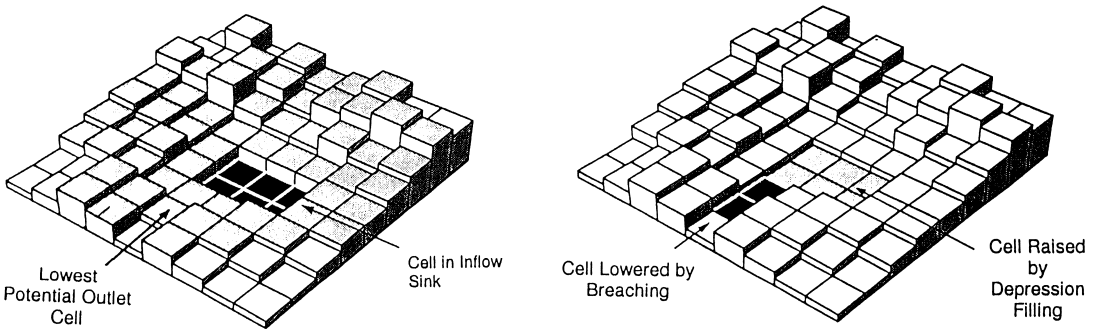


Figure 17.2. Breaching and filling of a spurious depression in a DEM: (2a) depression and outlet identification; (2b) outlet breaching and final filling of depression (figure from Garbrecht, Starks, and Martz, 1996).

estimated elevations (Figure 17.1). In such cases, breaching the obstruction is more appropriate than filling the depression created behind the obstruction. This breaching approach reduces or eliminates the filling of depressions created by narrow obstructions. It is particularly effective in DEMs of low relief landscapes in which obstruction of flow paths are more prevalent.

A three-step algorithm is used for breaching narrow obstructions along flow paths. First, the spatial extent of each depression and its contributing area is delineated. Second, potential outlets or overflow points on the edge of the depression area are defined. Potential outlets are those raster cells within the depression area which are: (1) adjacent to a cell outside the depression area, and (2) at a higher elevation than a cell outside the depression area. The lowest of these potential outlets is selected as the depression outlet. Third, the selected outlet is evaluated for possible lowering to simulated breaching (Figure 17.2a). The number of cells at the outlet that may be lowered by breaching is termed the breaching length. To restrict breaching to relatively narrow obstructions, the breaching length is arbitrarily set to one or a maximum of two cells. If the breaching length is longer than two cells the flow obstruction is likely to be a true topographic feature, and outlet breaching is not permitted. In the presence of a one- or two-cell breaching length, the elevation of the outlet cell(s) are lowered to the lesser elevation of the cells outside or inside of the depression at the breaching site (note: both cells outside and inside the depression at the outlet are of lower elevation per definition of the breaching length). This third step changes the elevation of the outlet and effectively breaches the obstruction responsible for the depression (Figure 17.2b). If more than one potential breaching site exists, the one with the greatest breaching depth (primary criterion), and the shortest breaching length (secondary criterion) is selected. Finally, regardless of whether a breach is performed or not, the elevations of the remaining cells inside the depression and at a lower elevation than the outlet are changed to the elevation of the outlet (Figure 17.2b). This produces a continuous flat surface at the location of the depression. A more detailed coverage of depression breaching/filling can be found in Martz and Garbrecht (1999).

TREATMENT OF FLAT SURFACES IN DEMS

Truly flat land surfaces seldom occur in nature. However in DEMs, areas of limited relief can translate into perfectly flat surfaces. Perfectly flat surfaces in DEMs can be attributed to the following three causes: (1) too low a vertical and/or horizontal DEM resolution to represent the landscape, particularly affecting low relief landscapes; (2) filling of depressions; and (3) landscape that is truly flat, which seldomly occurs. Whatever their origin, flat surfaces are problematic because

flow direction on a perfectly flat surface is indeterminate (Speight, 1974; Tribe, 1992). This problem arises in automated drainage analysis with both the widely used D–8 flow routing approach (Fairchild and Leymarie, 1991) and for various multiple-direction and aspect-driven approaches (Costa-Cabral and Burges, 1994).

Traditionally, flow direction over flat surfaces in DEMs is defined using a variety of methods ranging from landscape smoothing to arbitrary flow direction assignment. For a short review of existing methods the reader is referred to Tribe (1992). Flow direction assignment over flat surfaces is particularly difficult within the framework of the D–8 method (Fairchild and Leymarie, 1991) because landscape properties are defined by the DEM cell at the point of interest and its immediate surrounding eight adjacent cells. Since all DEM cells on a flat surface have the same elevation value, a unique flow direction cannot be assigned. In the following, a generic numerical scheme is presented that allows for the identification of flow direction over flat surfaces.

This numerical scheme is based on the recognition that natural landscapes generally drain toward lower terrain while simultaneously draining away from higher terrain. This effect is incorporated into DEMs by incrementing elevations on flat surfaces to produce two gradients: one forces flow away from higher terrain; the second draws flow toward lower terrain. The selected elevation increment is arbitrarily small (say, 1 mm). Such small elevation increments are sufficient to identify flow direction over the flat surface, yet from a practical point of view they do not significantly alter the elevation of the digital land surface.

The gradient toward lower terrain is imposed by incrementing the elevation of all cells in the flat surface that are not adjacent to a cell with a lower elevation (outlet) or an existing downslope gradient. This incrementation is applied successively and repeatedly to all cells that after each incrementation pass still remain with no downslope gradient. In this way, a flow gradient toward lower terrain is constructed as a backward growth from the outlet(s) into the flat surface while at the same time satisfying all boundary conditions imposed by the higher and lower terrain surrounding the flat surface.

The gradient away from higher terrain is imposed by first incrementing the elevation of all cells in the flat surface that are adjacent to higher terrain and have no adjacent cell at a lower elevation. The imposed increment introduces a downslope gradient away from higher terrain for all cells immediately adjacent to higher terrain. In subsequent passes the incrementation is applied to all cells that have been incremented in previous passes, and also those cells that are in the flat surface and adjacent to an incremented cell, but not adjacent to a cell of lower elevation. The result of this incrementation is a gradient away from higher terrain which is grown from the edges of the higher terrain into the flat surface.

In a final step, the cumulative gradients applied in the previous two steps are linearly added for each cell to determine the total incrementation. Adding the total incrementation to the initial elevation of each cell results in a surface that is no longer flat and includes a gradient away from higher terrain, and a gradient toward lower terrain. The net effect of the elevation incrementation is the modification of elevations on the flat surface which will produce, by means of subsequent DEM processing, a flow direction pattern that is consistent with the topography surrounding the flat surface and that displays flow convergence properties. A more detailed description of drainage identification over flat surfaces can be found in Garbrecht and Martz (1997a).

The effect of the algorithm is illustrated using a saddle topography between two mountains (Figure 17.3). The saddle consists of a flat surface between higher terrain to the right and left (hatched) and three locations of lower terrain (circles). Additional complications are introduced by a wedge of higher terrain protruding into the flat surface from the bottom, and a rectangular indentation of the flat surface into higher terrain at the top right corner of the figure. The arrows show the computed drainage over the flat surface. The arrow size is proportional to the upstream

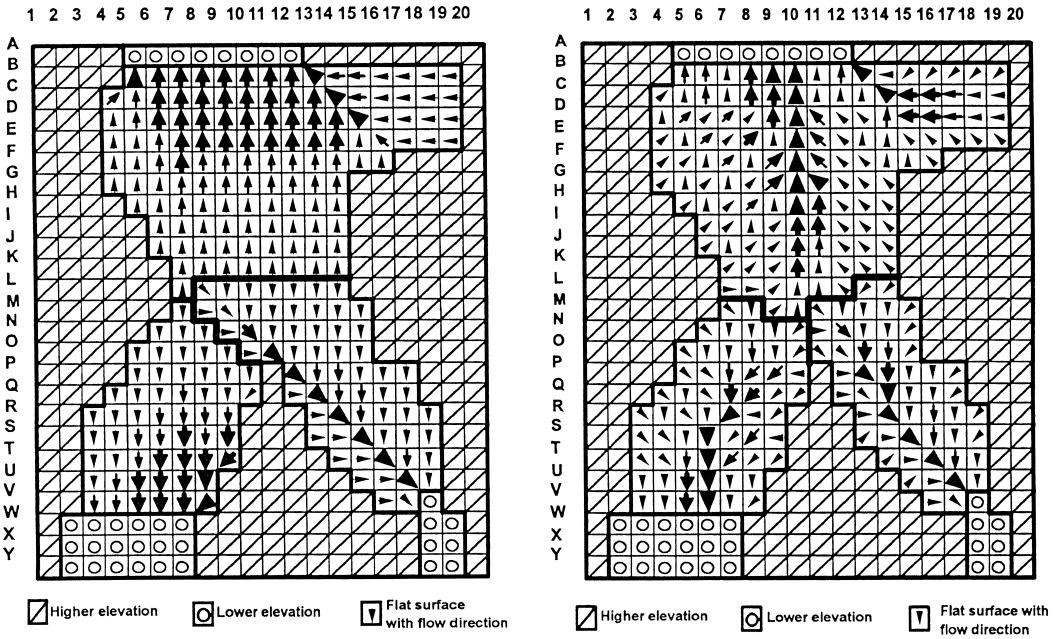


Figure 17.3. Drainage pattern over a saddle topography: (a) traditional approach; (b) new approach (figure from Garbrecht and Martz, 1997a).

drainage area of the flat surface. Figure 17.3a shows the drainage pattern produced by the model of Martz and De Jong (1988) and Jenson and Domingue (1988). This drainage pattern suffers from the "parallel flow problem" and a lack of flow convergence (Fairchild and Leymarie, 1991). Figure 17.3b shows the corresponding drainage pattern produced by the presented procedure. The drainage displays flow convergence properties and is much more consistent with the topography of the overall saddle configuration.

A second example illustrates a curved valley with a flat floor flanked by higher terrain (Figure 17.4). In addition, a small hill in the valley center creates an obstruction to drainage. The arrows show the path of the main drainage line around the inside corner of the valley bend. Drainage from behind and below the small hill converge rapidly and join the main drainage line. Any tributary from the higher valley sides would enter the flat surface, follow the indicated arrows, and join the main drainage line within a short distance. The flow convergence and drainage pattern in Figure 17.4 is reasonable, given that the initial valley floor was flat and contained no topographic information to guide the drainage identification.

NETWORK AND SUBCATCHMENT INDEXING

Once the channel network and direct contributing areas are automatically defined from DEMs, they are usually displayed as a raster image which consists of strings and groups of raster cells with numeric codes or colors that distinguish the network and subcatchments. For these images to be useful in watershed management and runoff modeling, individual channel links and contributing areas must be explicitly identified and associated with topological information for upstream and downstream connections. Such identification is often possible in vector GIS, but usually not in raster GIS. An algorithm that can analyze images of raster channel networks, index network nodes,

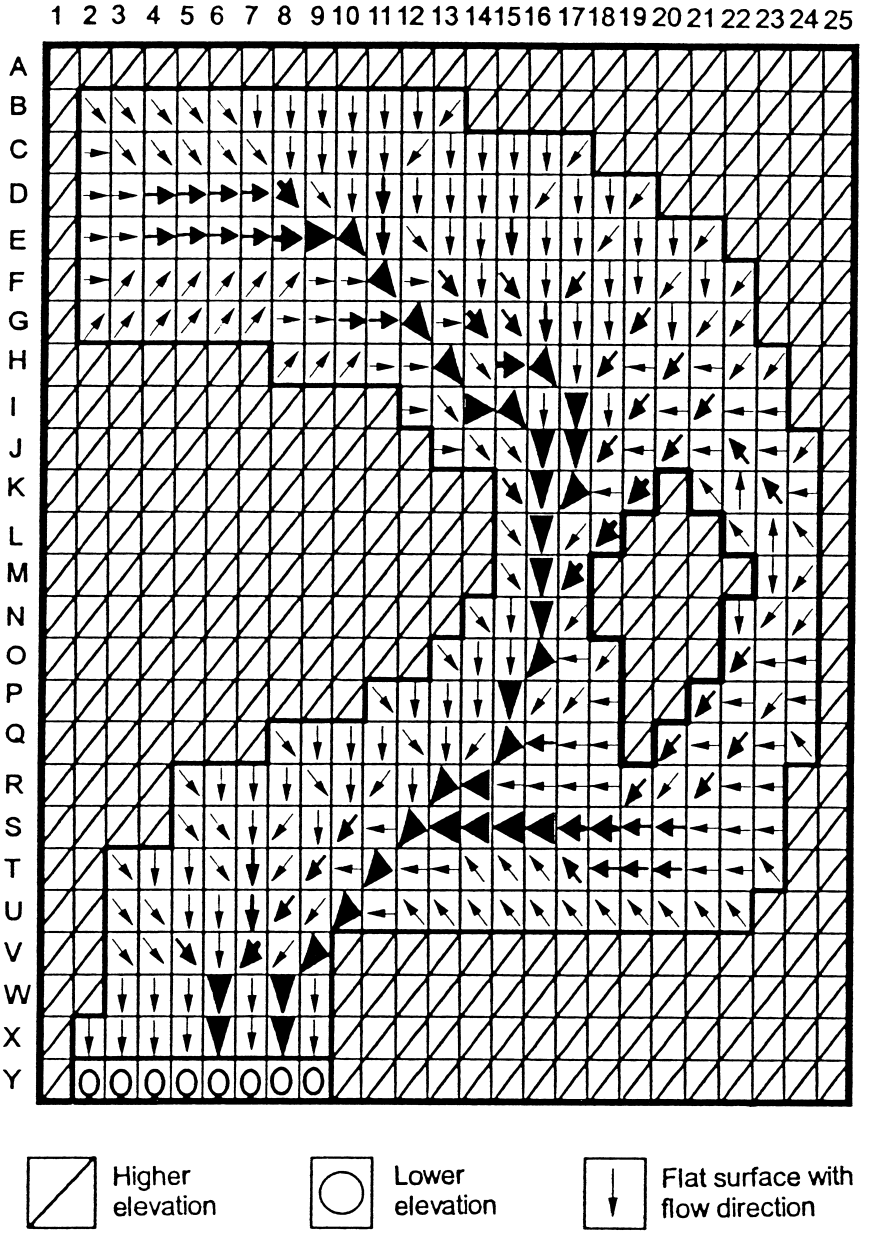


Figure 17.4. Drainage pattern on a flat valley floor (figure from Garbrecht and Martz, 1997a).

and order the channels by the Strahler method is presented. This algorithm provides a direct link between GIS images and hydrologic models, and leads to automated processing of segmented watersheds by distributed hydrologic models.

It is assumed that an image of an unidirectional, fully-connected network has already been defined. Four steps are required to fully identify the topology of the network and subcatchments. In the first step, flow direction information at each raster cell is used to move cell by cell along the channels of the network from upstream to downstream to determine the Strahler order (Strahler,

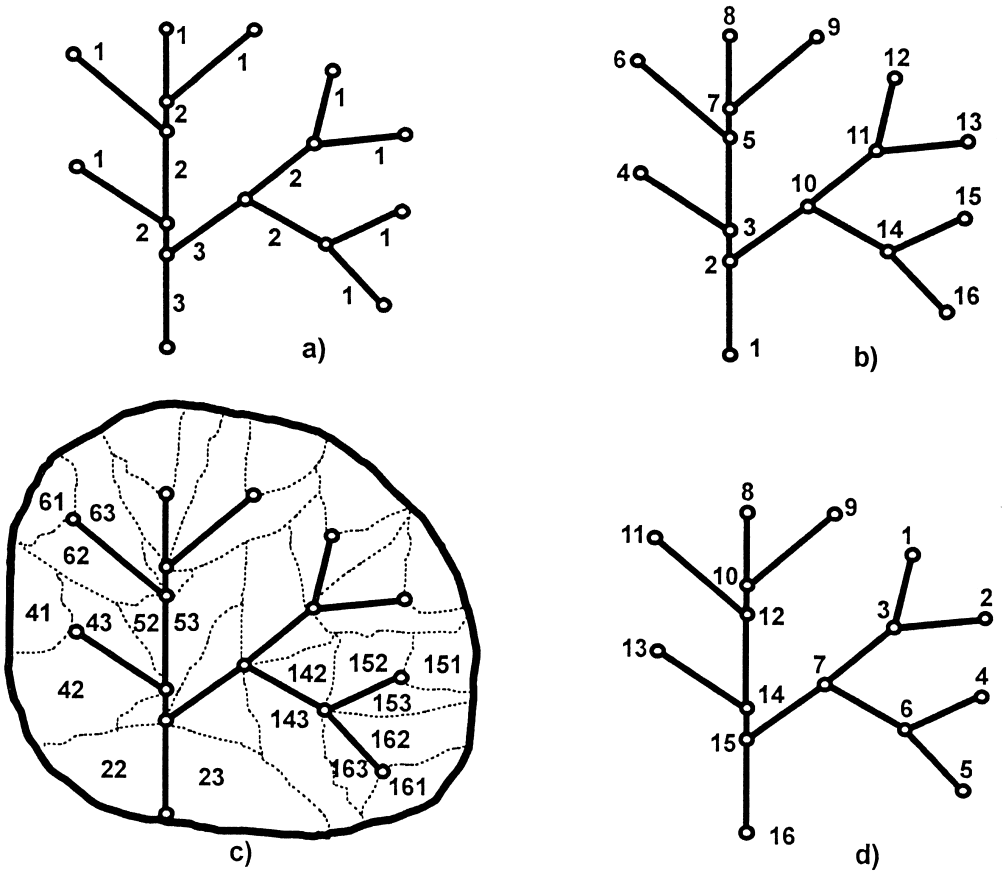


Figure 17.5. Drainage network and node indexing: (a) Strahler orders; (b) node indexing; (c) subcatchment indexing for selected node numbers 2, 4, 5, 6, 14, 15 and 16; (d) sequence in which the channels are to be processed for flow routing (figure from Garbrecht and Martz, 1997b).

1957) of each channel link and to identify the location of all source and junction nodes (Figure 17.5a). Flow direction is determined as the steepest downslope flow path from the current cell to one of the eight neighboring cells. In the second step, the node location and flow direction information are used to simulate a walk along the left bank of the channel network beginning and ending at the watershed outlet (Croley, 1980). The walk begins at the watershed outlet which is assigned node number 1. Then the walk traces the channel that ends at the outlet to the next node upstream. At that node the left channel is followed further upstream. As each node is passed during this walk, it is assigned the next sequential node number (Figure 17.5b). When a source node is encountered the walk moves in the downstream direction until a partially evaluated node is encountered, at which time the node-by-node walk is resumed again in an upstream direction along the unevaluated channel branch.

In the third step, the subcatchment areas for each channel link are identified. These consist of direct contributing areas into the left-bank, right-bank and, in the case of exterior links, into the source of the channel link. Subcatchments are assigned an identification code based on the previously assigned node numbers. For all subcatchments, a base identification number is assigned which is the node number (NN) at the upstream end of the link to which the subcatchment drains multiplied by 10 (NN*10). For source node subcatchments a value of 1 is added to (NN*10), for

left-bank subcatchments a value of 2 is added, and for right-bank subcatchments a value of 3 is added. Channel cells (which are not considered to be part of the right-bank, left-bank, or source node subcatchments) are assigned an identification code of (NN*10) plus 4 (Figure 17.5c).

These identification codes provide a basis on which network links, nodes, and subcatchments can be associated with one another. Most importantly, in a fourth step the node numbering scheme can be used to determine the optimal routing sequence to be used in modeling streamflow through large and complex networks (Garbrecht, 1988) (Figure 17.5d). The algorithm makes possible the automated quantification of network structure from raster network images and greatly enhances the direct linkage of GIS-generated channel networks and hydrologic and hydraulic models. Further details on this algorithm can be found in Garbrecht and Martz (1997b).

OVERLAND AREA PARAMETERIZATION

Automated identification of landscape parameters for individual overland areas within a subdivided watershed is the next step in automated DEM processing for hydrologic/hydraulic model application (Goodrich and Woolhiser, 1991). Overland areas are defined as undissected hillslopes of irregular shape that drain directly into a channel link. For hydrologic modeling these overland areas are often approximated by a rectangular plane of given width (W), length (L) and slope (S) (Wooding, 1965; Smith at al., 1995; Feldman, 1995). Such a Wooding representation of a subcatchment consists of two rectangular planes joined to form a V-shaped valley along which the stream flows (Figure 17.6). The geometric dimensions W, L, and S of the plane are essential for the determination of the magnitude, shape, and timing of the overland runoff hydrograph, and are used in models such as KINEROS (Smith et al., 1995) and the kinematic option of HEC–1 (Feldman, 1995). Three numerical expressions are presented that identify the plane parameters to reproduce the hydraulic runoff characteristics of the original overland areas. These expressions are implemented using flow path, accumulated area and elevation data derived from the DEM raster. This raster data can be obtained by suitable DEM processing software such as the digital landscape analysis tool TOPAZ (Garbrecht and Martz, 1999b, 2000).

The definition and assumption necessary for the first algorithm are: (1) a flow path is the route traveled by the water from an upstream source to the channel at the downstream edge of the overland area; and, (2) flow paths with large discharge contribute proportionally more to the runoff hydrograph characteristics than flow paths with small discharge. Based on this definition and assumption the model for hydraulically representative plane length (L) can be formulated as a discharge-weighted mean length of all flow paths within the irregular overland area. Furthermore, the discharge on hillslopes is often proportional to upstream area, and the expression for plane length can be formulated as follows:

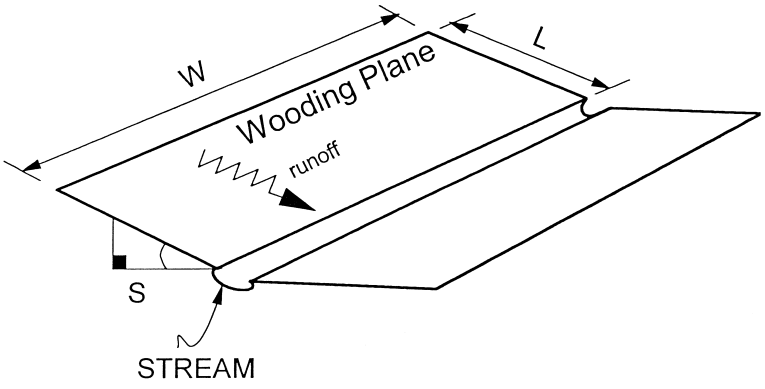


Figure 17.6. Wooding catchment representation (figure from Garbrecht, Martz, and Goodrich, 1996).

$$L = \frac{\sum_{i=1}^{n} l_i * a_i}{\sum_{i=1}^{n} a_i} \tag{1}$$

where l is flow path length, a is upstream drainage area, and subscript *i* is flow path counter.

The second numerical expression defines the plane width and is derived from area and water conservation considerations. To ensure water continuity and area conservation, plane width is computed as the overland area (A) divided by the plane length (L) computed by Eq. 1.

Finally, the expression for plane slope is a weighted average of all flow path slopes. Flow path slope is defined as the change in elevation from the top to the bottom of the flow path divided by flow path length. All flow paths in the irregular overland areas are considered. The expression for the plane slope is given by:

$$S = \frac{\sum_{i=1}^{n} s_i * k_i}{\sum_{i=1}^{n} k_i} \tag{2}$$

where s is flow path slope and k is a weighting factor. The weighting is defined in one of three ways: (1) upstream drainage area weighted with $k_i=a_i$; (2) flow path length weighted with $k_i=l_i$; or, (3) upstream drainage area times flow path length weighted with $k_i=a_i*l_i$.

In the first case, length weighting is used because slope estimates of long flow paths are generally more accurate since the elevation and length values have been obtained over a larger number of discrete raster units, effectively reducing the raster resolution noise. Length weighting also favors flow paths with larger drainage areas (i.e., larger discharge) because drainage area and flow path length are often related. In the second case, drainage area weighting is used to emphasize flow paths with larger drainage areas which contribute proportionally more to the runoff hydrograph characteristics than those with smaller drainage areas. The third case is a combination of the first and second case with the product of length and drainage area as the weighting factor. Until further research establishes the most appropriate method for the estimation of plane slope, the results of each of the three methods should be considered in the determination of the final plane slope.

A more detailed discussion of the overland area parameterization can be found in Garbrecht et al. (1999).

CONCLUSIONS

Advances in the treatment of depressions and flat surfaces, the identification of network topology, and the parameterization of overland areas are presented. The proposed depression removal by a combination of breaching and filling, as well as the gradient imposition of flat surfaces, produce a more realistic and consistent drainage pattern than traditional depression filling and local drainage searches over flat areas. The presented improvements are particularly important for drainage and erosion investigations.

The identification of channel network and subcatchment topology from raster images provides an important linkage between DEM-derived drainage features and automated watershed management and hydrologic modeling. Finally, the length and slope of the rectangular overland area ap-

proximation are computed as drainage area weighted averages over the length and slope of all flow paths in the irregular overland areas. The strength of the procedure lies in its consideration of the geometric properties of each contributing flow path rather than relying on a lumped approach, and in the consideration of qualitative cause-effect relations between flow path and runoff characteristics to emphasize flow paths that are more important to the runoff process.

REFERENCES

Band, L. E., 1986. Topographic partition of watersheds with digital elevation models. *Water Resources Research,* 22(1):15–24.

Costa-Cabral, M. C., and S. J. Burges, 1994. Digital Elevation Model Network (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resources Research*, 30(6):1681–1692.

Croley, T. E., 1980. A micro-hydrology computation ordering algorithm. *Journal of Hydrology*, 48:211–236.

Fairchild, J., and P. Leymarie, 1991. Drainage networks from grid digital elevation models. *Water Resources Research,* 27(4):29–61.

Feldman, A. D., 1995. HEC–1 flood hydrograph package, Chapter 4 of *Computer Models of Watershed Hydrology*, Singh, V. J., Ed., Water Resources Publication, Highlands Ranch, CO, pp. 119–150.

Garbrecht, J., 1988. Determination of the execution sequence of channel flow for cascade routing in a drainage network. *Hydrosoft,* 1(3):129–138.

Garbrecht, J., L. W. Martz, and D. C. Goodrich, 1996. Subcatchment parameterization for runoff modeling using digital elevation models. In *Proceedings of the American Society of Civil Engineers Hydraulics Conference,* North American Water and Environment Congress '96, ASCE, Anaheim, CA, 24–28 June, 1996.

Garbrecht, J., P. J. Starks, and L. W. Martz, 1996. New digital landscape parameterization methodologies. In *Proceedings of the 32nd Annual Conference and Symposium on GIS and Water Resources, American Water Resources Association*, Herndon, VA, TPS–96–3, September 22–26, 1996, Fort Lauderdale, FL, pp. 357–365

Garbrecht J., and L. W. Martz, 1997a. The assignment of drainage direction over flat surfaces in raster digital elevation models. *Journal of Hydrology*, 193:204–213.

Garbrecht J., and L. W. Martz, 1997b. Automated channel ordering and node indexing for raster channel networks. *Computers and Geosciences*. 23:961–966.

Garbrecht J., and L. W. Martz, 1999a. TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation and Subcatchment Parameterization, Overview, ARS-GRL 99–1, U.S. Department of Agriculture, Agricultural Research Service, Grazinglands Research Laboratory, El Reno, OK, 26 pp.

Garbrecht, J., and L.W. Martz, 1999b. Digital elevation model issues in water resources modeling. In *Hydrologic and Hydraulic Modeling Support in GIS,* D. Maidment and D. Djokic, Eds., ESRI Press, Redlands, CA, 216 pp.

Garbrecht, J. and L. W. Martz, 2000. TOPAZ: An Automated Digital Landscape Analysis Tool for Topographic Evaluation, Drainage Identification, Watershed Segmentation and Subcatchment Parameterization, TOPAZ User Manual, ARS-GRL, 2-00, U.S. Department of Agriculture, Agricultural Research Service, Grazingland Research Laboratory, El Reno, OK.

Garbrecht, J., D.C. Goodrich, and L. W. Martz, 1999. Method to quantify distributed subcatchement properties from digital elevation models. In *Proceedings of the 19th Annual AGU Hydrology Days,* Atherton, CA, pp. 149–160.

Goodrich, D. C., and D. A. Woolhiser, 1991. Catchment hydrology. *Reviews of Geophysics*, Supplement, April 1991, pp. 202–209, U.S. National Report to International Union of Geodesy and Geophysics. 1987–1990.

Jenson, S. K., and J. O. Domingue, 1988. Extracting topographic structure from digital elevation data for geographical information system analysis. *Photogrammetric Engineering and Remote Sensing,* 54(11):1593–1600.

Mark, D. M. 1984. Automatic detection of drainage networks from digital elevation models. *Cartographica*, 21 (2/3):168–178.

Mark, D. M., 1988. Network models in geomorphology. In Anderson, M. G., Ed., *Modeling Geomorphological Systems*. John Wiley and Sons, Chichester, UK pp. 73–96.

Martz, L. W., and E. De Jong, 1988. CATCH: A FORTRAN program for measuring catchment area from digital elevation models. *Computers and Geosciences*, 14(5):627–640.

Martz, L.W., and J. Garbrecht, 1992. Numerical definition of drainage networks and subcatchment areas from digital elevation models. *Computers and Geosciences* 18(6):747–761.

Martz, L. W., and J. Garbrecht, 1993. Automated extraction of drainage network and watershed data from digital elevation models. *Water Resources Bulletin*, 29(6):901–908.

Martz, L. W., and J. Garbrecht, 1998. The treatment of flat areas and depressions in automated drainage analysis of raster digital elevation models. *Hydrologic Processes* 12:843–855.

Martz, L. W., and J. Garbrecht, 1999. An outlet breaching algorithm for the treatment of closed depressions in a raster DEM. *Computers and Geosciences,* 25:835–844.

Moore, I. D., R. B. Grayson, and A. R. Ladson, 1991. Digital terrain modeling: A review of hydrological, geomorphological and biological applications. *Hydrological Processes*, 5(1):3–30.

Richards, N. S. 1981. General problems in morphology. In Gordie, A. S., Ed., *Geomorphological Techniques.* Allen and Umnin, London, pp. 26–30.

Smith, R. E., D. C. Goodrich, D. A. Woolhiser, and C. L. Unkrich, 1995. KINEROS—A kinematic runoff and erosion model. Chapter 20 of *Computer Models of Watershed Hydrology,* V. J. Singh, Ed., Water Resources Publication, Highlands Ranch, CO, pp. 697–732.

Speight, J. G., 1974. A Parameteric Approach to Landform Regions. Special Publication of the Institute of British Geographers, 7:213:230.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. *Trans. AGU*, 38(6): 913–920.

Tarboton, D. G., R. L. Bras, and I. Rodriguez-Iturbe, 1991. On the extraction of channel networks from digital elevation data. *Water Resources Research*, 5:81–100.

Tribe, A., 1992. Automated recognition of valley lines and drainage networks from grid digital elevation models: A review and a new method. *Journal of Hydrology*, 139:263–293.

Wolock, D. M., and G. J. McCabe, 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research*, 31(5):1315–1324.

Wooding, R. A., 1965. A hydraulic model for the catchment-stream problem. *Journal of Hydrology*, 3:254–267.

Zhang, W., and D. Montgomery. 1994. Digital elevation model grid size, landscape representation and hydrologic simulations. *Water Resources Research*, 30(4):1019–1028.

CHAPTER 18

Aerial Photointerpretation of Hazardous Waste Sites: An Overview

Donald Garofalo

INTRODUCTION

Much interest has been expressed recently by scientists interested in the capabilities and limitations of various remote sensors for characterizing hazardous waste and waste disposal sites. Special attention has been focused on the ability of these sensors to locate and characterize waste and waste site activities at abandoned facilities. As such, sophisticated new radar systems, thermal sensors, and other multispectral and hyperspectral scanners are being assessed as potential candidates for addressing the waste discovery and characterization problem.

Airborne radars can penetrate sand in hyperarid environments, between 1–6 meters in depth (Sabins, 1987); penetration is minimal in humid soil areas. Thermal sensors can display temperature variations at the surface, sometimes suggestive of subsurface conditions and multispectral and hyperspectral scanners can detect subtle changes in the surface environment which may be suggestive of buried features. The historical aerial photograph, however, frequently has resolutions for detecting barrel size or smaller features, and the ability to show site conditions long ago. Thus, the historical aerial photograph, unlike many newer and more sophisticated sensors, is an inexpensive source of invaluable information which can be used to locate a waste disposal site and measure accurately the size and dimensions of currently buried or overgrown features and to generally track the history of waste disposal site activity from beginning to end.

THE HISTORICAL AERIAL PHOTOGRAPH

The historical aerial photograph is the sensor of choice of the Environmental Protection Agency's Environmental Photographic Interpretation Center (EPIC), which for the past 25 plus years has been applying this tool for locating potential waste disposal sites and characterizing these sites and associated waste disposal practices. The historical aerial photograph is an extremely powerful remote sensing tool. It is the only remotely sensed data to have recorded events at sites frequently as far back in time as the 1930s. This is of immense value to the Superfund program which is charged with looking for and evaluating abandoned hazardous waste disposal sites, sites which today may display no evidence at the surface of their former use (Slonecker et al., 1999). The tool is of substantial value in litigation as evidence of the past waste disposal practices of PRPs (Principal Responsible Parties), and has been highly successful at assisting EPA and Department of Justice lawyers in winning their cases. In addition, the ability to view the aerial photo stereoscopically (in 3-D) and to measure, using photogrammetry, the heights, depths, volumes, and other dimensions of features and materials currently present, long removed from, or buried at

a site has also contributed invaluable evidence in courtroom situations, and has helped to reconstruct pictorially for jurors and judges alike and in an easily understandable format, the activity at a site over time. Finally, the availability of historical aerial photographs from both government and private sources, and the extent of aerial photo coverage of the United States, through various federal agency mapping programs such the Agricultural Stabilization and Conservation Service (now Farm Services Agency), and the U.S. Geological Survey, to name only two, have literally ensured that a hazardous waste disposal site located within the conterminous US has been overflown and site conditions documented more than once during the site's disposal history.

This chapter focuses on the value of the historical aerial photograph and the various image analysis and mapping functions which use aerial photos for analyzing hazardous waste disposal sites.

THE BASICS OF AERIAL PHOTOINTERPRETATION FOR WASTE SITE CHARACTERIZATION

Backlighting and Variable Magnification

Using backlighted tables with adjustable illumination, and high-power magnification zoom stereoscopes, aerial photographs are interpreted for site size, drainage patterns, type of fill materials, leachate, burial sites, lagoons, impoundments and their contents, and general condition of the site. Locations and descriptions of tanks, drums, open storage areas, evidence of vegetation stress, on-site obstacles, structures, equipment, access routes, and other details may also be obtained through photo analysis. Historical analysis provides the information necessary to obtain a chronological understanding of a site's development and activities. This information is particularly important for describing and illustrating past activities and conditions at abandoned hazardous waste disposal sites which fall under the jurisdiction of the Comprehensive Environmental Response, Compensation, and Liability Act (aka Superfund) program.

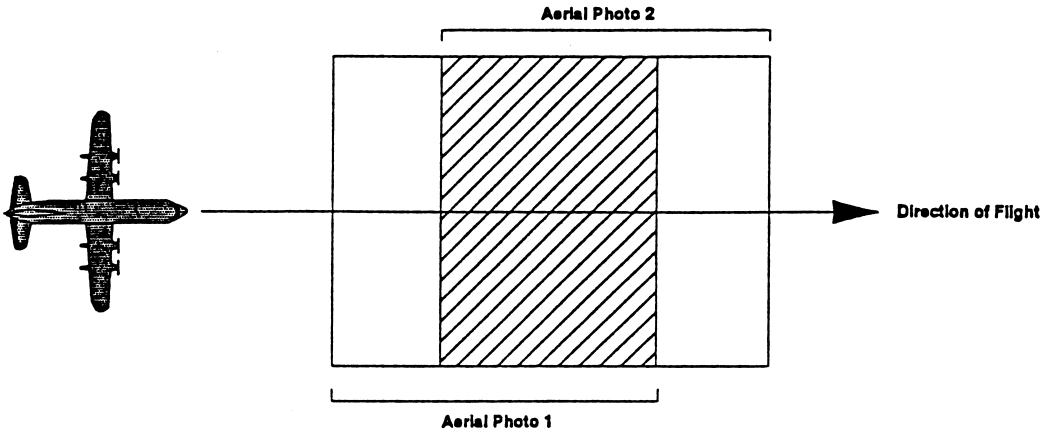
Film Transparencies

The use of aerial photo transparencies on backlighted variable illumination tables maximizes the available information content of analyzed aerial photographs. Aerial photo transparencies are first generation copies of originally exposed film. Each additional step of film processing, such as producing additional film copies or photographic prints degrades from the original product and reduces the amount of information contained in the original photo. High-powered magnifying scopes are used to identify subtle, but often significant features on aerial photos which can easily be overlooked if not viewed with the benefit of backlighting and variable magnification.

Stereoscopy

The importance of stereoscopy in the photointerpretation process cannot be ignored. Stereoscopy allows the photo analyst to see features on an aerial photograph in three dimensions (3-D). Through stereoscopic parallax (the apparent displacement of the position of a feature in an image caused by a change in the position of observation) a stereoscope may be used to view overlapping aerial photographs to provide a three-dimensional effect of features on the ground (Figure 18.1). When coupled with various measuring devices such as stereo comparators or other digital photogrammetric devices highly accurate measurements can be made of the dimensions (height, widths, lengths, depth) of features seen on the aerial photograph. Volumes of materials and volu-

Acquisition of Stereoscopic Aerial Photographs and Stereoscopic Parallax *



Area of 60% Forward Lap
Area of Forward Lap represents the same ground area photographed along a flightline by two overlapping aerial photos.

* Stereoscopic parallax is the displacement of the position of a feature in an image caused by a shift in the observation position. A stereo pair are two overlapping photos containing stereoscopic parallax, and which may be viewed using a stereoscope to provide a three-dimensional effect. The same effect can be obtained with a hand-held camera by taking a picture of the same object from two separate locations and then viewing the pair of pictures together with a stereoscope (Sabins, 1987).

Using special photogrammetric equipment accurate measurements can be made on stereo photo pairs. Contour maps which show differences in terrain elevation are also made using this technique.

Dimensions of features such as their length, width, and height, as well as volumetric measurements such as amounts of mounded materials, or depths and sizes of excavations can be calculated using photogrammetric tecniques.

Figure 18.1. Acquisition of stereoscopic aerial photographs and stereoscopic parallax. The area of 60% forward overlap is shown as the cross-hatched area, and it represents the same ground area photographed along a flightline by two overlapping aerial photos.

metric capacities of excavations can be calculated. This same technique is used to make topographic maps which show land surface elevations using contour lines.

The Signature Concept

A basic part of photointerpretation is the extraction of useful information from an image. Although this may be performed by human (manually) or machine (electronically by computer) depending on the form of the original data (e.g., as analog, hard copy photos versus digital images, respectively) the photo analysis performed by EPIC in support of EPA's hazardous waste program is conducted almost exclusively by manual, not machine analysis. Using manual methods, some types of information displayed on aerial photos are obvious to anyone used to reading a map. An example includes large bodies of water. However, the vast bulk of information is not evident to the untrained or inexperienced viewer.

The training and experience that makes feature identification possible on aerial photos is based on learning to recognize combinations of imagery characteristics called *signatures*. A *signature* is a combination of visible characteristics (such as color, tone, shadow, texture, size, shape, pattern, and association) which permit a specific object or condition to be recognized on an aerial photograph. The relative potential of a given signature to enable possible, probable, or positive identification of an object or condition is the *degree of certainty*. *Possible* is a term signifying a degree of certainty of signature identification when only a few characteristics are discernible on a photo or these characteristics are not unique to a signature. *Probable*, on the other hand, is a term signifying a degree of certainty of signature identification when most characteristics, or strong or unique characteristics of a signature are discernible but fall short of positive identification.

When interpreting an aerial photograph, the analyst is generally searching for the signature of one or more objects or conditions by viewing aerial photo stereopairs through stereoscopes. An analyst relies either on experience or "ground truth" information (corroborative information obtained through other data sources or on-site field visits to the site) in identifying signatures. When working with historical aerial photographs of areas that have changed or are inaccessible, experience becomes critically important. For some forms of photo-interpretation, e.g., to map wetlands vegetation or forest types, signatures representative of various vegetation types are confirmed by on-site visits and then extrapolated over a much larger area to map vegetation types displaying the same signatures, but not visited on the ground.

The *signature* concept is not an all or none concept since signatures can vary in degree of certainty. Because of this it is essential to clearly distinguish between positive identifications and calls of lesser certainty. Because hazardous waste site characterization by aerial photo interpretation may be used to support civil or criminal litigation, it is important that the degrees of certainty be clearly stated by the photo interpreter.

The characteristic signature of a given object or condition can vary with the type of film or imagery, scale, resolution, and other factors. Therefore, aerial photos vary in suitability depending on the object of the analysis. For example, 55-gallon drums can be positively identified on average quality, 1:6,000 scale aerial photographs. On the other hand, typical 1:20,000 scale aerial photos do not normally allow for positive identification of drums, but may allow for *possible* or *probable* identifications to be made. Even smaller scale (higher altitude) imagery is so inappropriate for drum identification that even *possible* identifications cannot be made. Variability in photo type also affects the interpretability and signatures of objects. Natural color, color infrared, and black and white aerial photos, for example, have no sensitivity for heat detection, while thermal scanner imagery can reveal temperature differences of objects. An experienced photo interpreter is able to use a variety of remote sensing tools and is fully knowledgeable of the capabilities *and* limitations

Table 18.1. Hazardous Waste Disposal Site Features/Activities Routinely Extracted by Image Analysis from Historical Aerial Photographs

Access road	Extraction area	Pressure tank
Berm	Feature boundary	Probably underground drainage
Building	Feature outline	Railroad
Channelized drainage	Fence	Refuse
Chemical storage	Fenced site boundary	Revegetated
Cleared area	Fill	Revetment
Container	Flow direction	Site boundary
Crates/boxes	Graded area	Sludge
Culvert/bridge	Ground scar	Solid waste
Cylindrical object	Historical boundary	Stacked objects
Dark-toned	Horizontal tank	Stain
Debris	Impoundment	Standing liquid
Dike	Indeterminate drainage	Structure
Disposal area	Lagoon	Study area
Disturbed ground	Landfill	Surface runoff
Drainage	Leachate	Suspected drainage
natural	Light-toned	Tank farm
channelized	Liquid	Tank trailer
suspected or historical	Material	Trench
indeterminate flow	Medium-toned	Unfenced site boundary
natural	Mounded material (extensive)	Vegetated
channelized	Mounded material (small)	Vegetation stress
tidally influenced	Objects	Vehicle
natural	Open storage	Vehicle access
channelized	Outfall	Vertical tank
Drums	Pipeline	Waste disposal area
Edge of slope	Pit	Wastewater treatment plant
Excavation	Pond	Wetland
Excavation, pit (extensive)	Possible drum area	

of these tools for specific applications. Table 18.1 is a listing of the kinds of features, natural resources, and site activities which are routinely identified on aerial photographs by skilled image analysts for characterizing waste disposal sites.

OVERVIEW

Hazardous Waste Site Analyses

Hazardous waste disposal site characterization using historical and current aerial photographs comprises a major part of EPIC's workload. Utilizing the vast archives of aerial photographs of the country maintained by government and private sources, dating back to the 1930s, EPIC's analysts reconstruct the waste handling and disposal history of a site in order to support site cleanup and regulatory or enforcement efforts. Aerial photographs have proved to be powerful tools in court in the form of evidence and to support expert witness testimony and for facilitating the recovery of millions of dollars in site cleanup costs and penalties from responsible parties (Garofalo and Wobber, 1974; Erb et al., 1981; Evans and Mata, 1984; Stohr et al., 1987; Mata and Christie, 1991).

The information interpreted from an aerial photograph is annotated onto a clear film overlay which identifies and delineates the location of significant ground features and activities. Accom-

Figure 18.2. Sample historical site analysis prepared by EPIC. October 19, 1958. No significant activity was evident on the 1946 or 1947 photographs (not shown here). Significant features observed on the 1951 photographs are annotated and discussed in conjunction with the following analysis.

The landfill study area consists of two landfills, the Western Landfill and the Eastern Landfill. Both landfills are south of Wheatfield Road. No significant activity is visible within the portion of the study area north of Wheatfield Road throughout the analysis, and this area will not be discussed further.

Each landfill is divided into three portions for the purpose of discussion. The northern, central, and southern portions will not be annotated further. Fill areas are annotated in both landfills throughout the analysis, but are not discussed individually. Disposal activities noted within fill areas are annotated and discussed.

Drainage at and around the landfill is shown in the Wetland and Drainage analysis (Figure 18.10). Minimal changes were evident in the overall drainage routes throughout the years of analysis. Minor variations in drainage resulted from the gradual development of the landfill. These transient drainage routes are annotated for each year of analysis, but are not discussed unless a significant change is visible.

Western Landfill. No significant activity is evident.

Eastern Landfill. Northern portion. In 1951, possible refuse (R) was noted in a possible fill area (FA) outside the eastern site boundary.

In 1958 the east side has been partially cleared (not annotated). Material (M) is piled in three areas within the clearing. A possible excavation (EX) with dark-tone (DK) liquid (LQ) is noted on the south edge of the clearing. Possible refuse remains evident in the possible fill area outside the eastern site boundary.

Central and southern portions. No significant activity is evident.

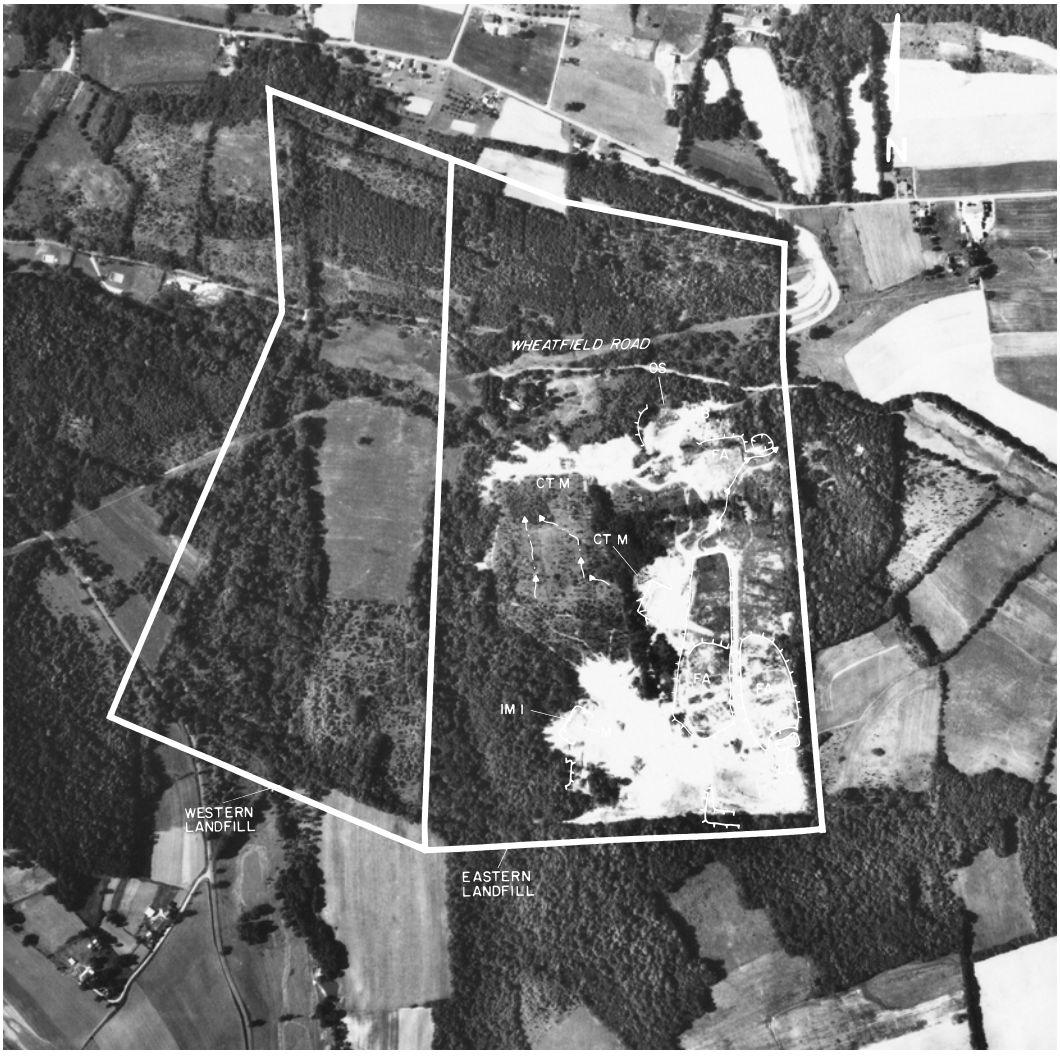


Figure 18.3. June 17, 1964. Western Landfill. No significant activity is evident.

Eastern Landfill. Extensive clearing (not annotated) is underway in the eastern landfill, and brush piles (not annotated) are scattered throughout the clearing.

Northern portion. The material and possible excavation with liquid seen in 1958 are no longer visible. Piles of coarse-textured (CT) material, an open storage (OS) area, a building (B) with a parking area, and an empty excavation, possible a catchment basin, are noted in the northern portion of the landfill. A variety of objects and equipment are visible in the open storage area. The open storage area, building and parking area remain active and expand throughout the analysis, and will be annotated but not discussed in detail.

The possible refuse seen outside the eastern site boundary in 1958 is no longer visible.

Central Portion. Piles of coarse-textured material and a rectangular bermed area are visible. The bermed area appears empty, and its use is not evident.

Southern Portion. An impoundment (IM1) and an excavation are visible. IM1, a leachate collection impoundment, is dry inside and contains a small amount of material. The excavation contains murky liquid.

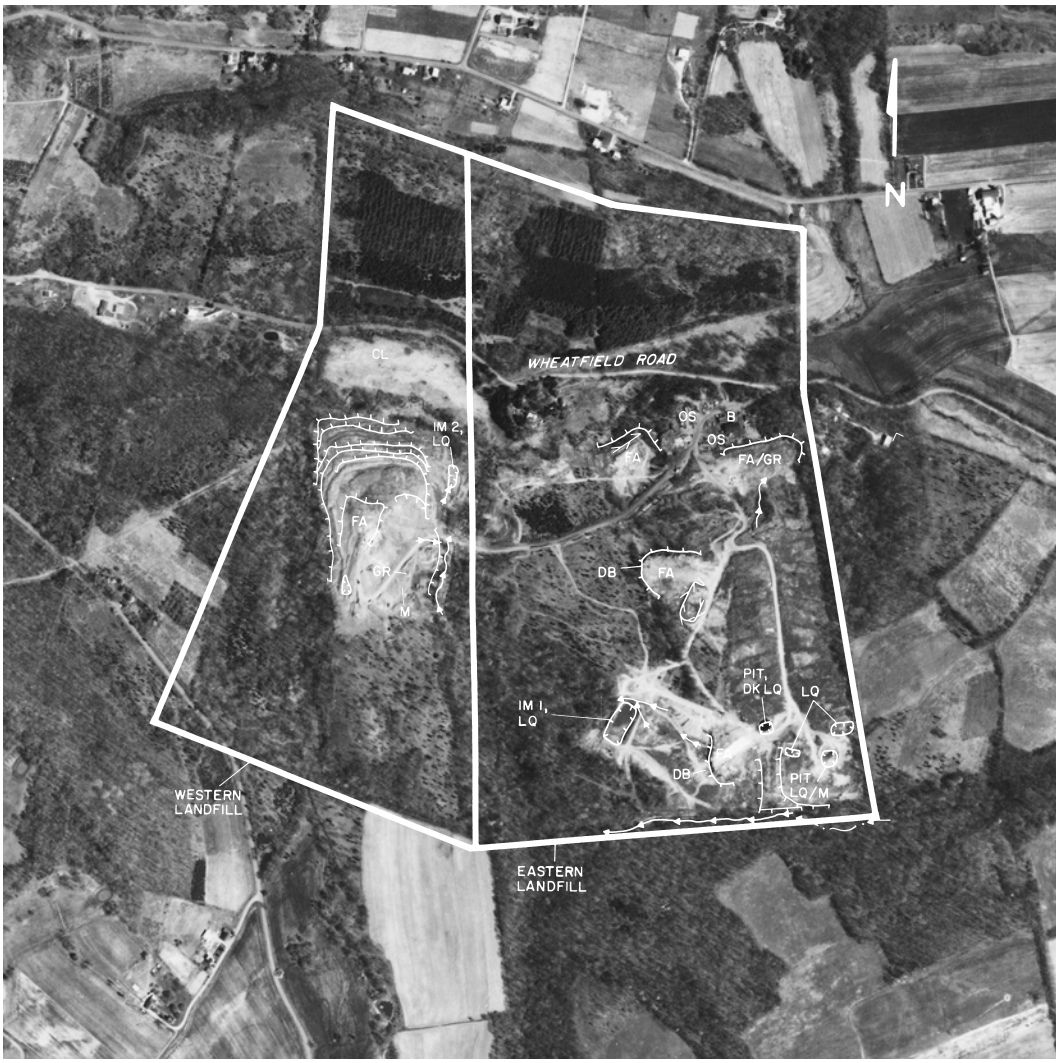


Figure 18.4. April 17, 1968. Western Landfill. North portion. A wetland has been cleared (CL); however, there is no evidence of fill activity in this location.

Central Portion. Extensive filling has occurred since 1964, an impoundment (IM2), an excavation, grading (GR), and material are evident. A channel leads from the east edge of the landfill into IM2. Murky liquid is visible in the channel and in IM2. The excavation is on the southwest side and appears to be empty.

Southern portion. No significant activity is evident.

Eastern Landfill. Northern portion. The coarse-textured material noted in 1964 is no longer present. A berm and the grading of a fill area are evident in the northern portion. The graded fill area includes the former excavation which was noted in 1964.

Central portion. The bermed area and debris (DB) are visible. The bermed area remains empty, and the debris is along the edge of a fill area.

Southern portion. IM1, debris, two pits, and two excavations are visible. IM1 contains liquid, the debris is along the edge and sides of a fill area, the northern pit is full of dark-toned liquid, and the southern pit contains a small amount of dark-toned liquid/material. Both excavations contain murky liquid.

A drainage channel is visible at the south edge of the site. It receives runoff from slopes south and east of the landfill as well as from on-site. The channel leads west to the natural drainage, which flows north (Figure 18.10).

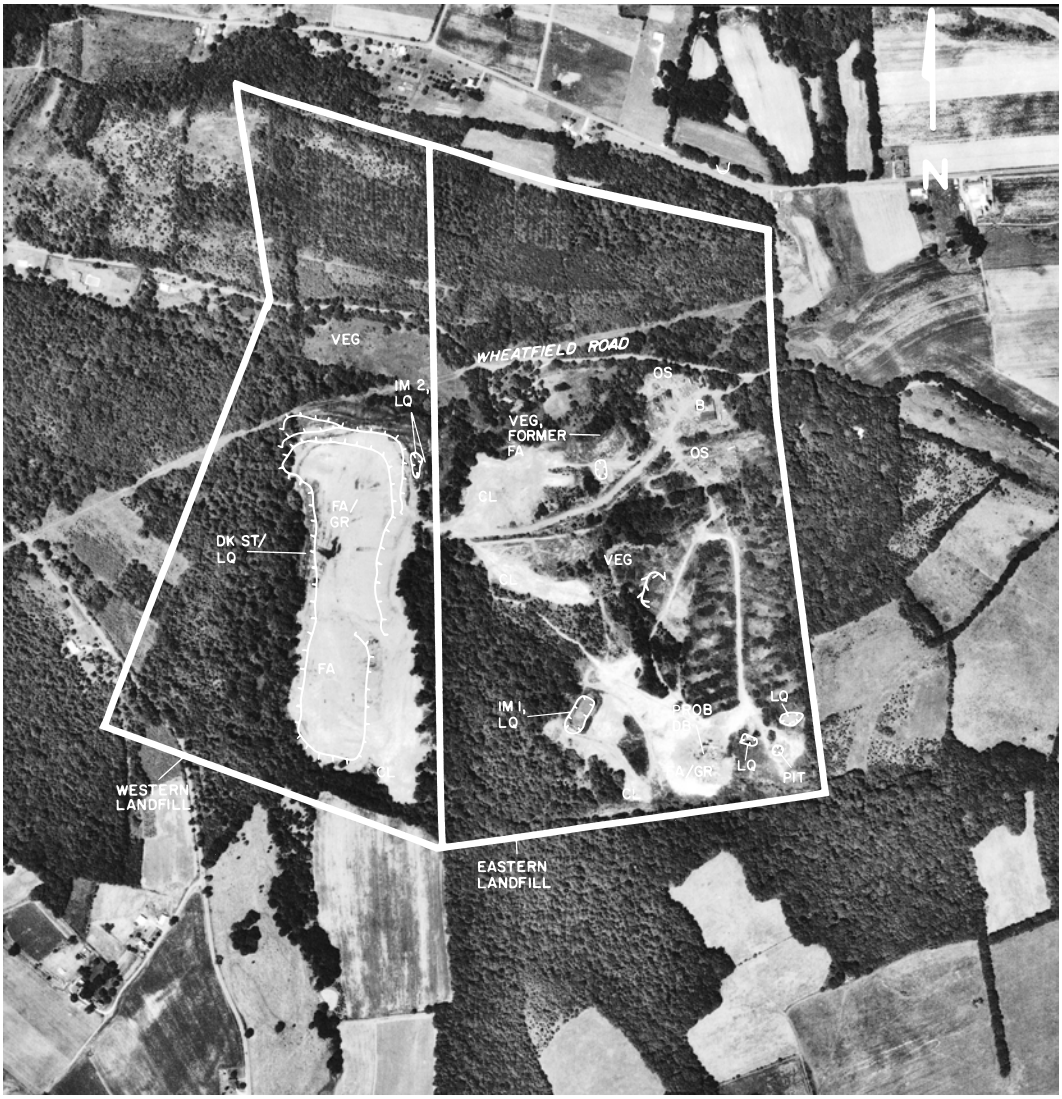


Figure 18.5. July 18, 1971. Western Landfill. North portion. Vegetation (VEG) is growing where the clearing of a wetland was noted in 1968.

Central Portion. Filling has continued since 1968, and the upper surface of the landfill has been graded. IM2 contains dark-toned liquid. Dark-toned liquid is adjacent to the east side of IM2, indicating the possibility of a breach or overflow in the impoundment. The western excavation is no longer visible. A dark-toned stain (ST) extends from the top of the landfill downslope to the base, as if caused by a liquid.

Southern portion. Extensive filling has occurred since 1968. A clearing is noted along the southeast edge.

Eastern Landfill. Northern portion. The fill area with a berm noted in 1968 is vegetation. A shallow excavation is noted south of the former fill area. The graded fill area noted in 1968 and part of the parking area around the building are in use for open storage. A large clearing is seen on the west side of the landfill.

Central Portion. Vegetation is growing where a fill area with debris was noted in 1968. The bermed area appears inactive and is partially overgrown with vegetation. A clearing is seen on the west side of the landfill.

Southern portion. The northern pit seen in 1968 is no longer present; the southern pit is smaller and appears empty. IM1 contains liquid. The debris seen along the edge of a fill area in 1968 is no longer visible, and the top of the fill area has been graded. A small amount of probable debris is evident on top of the fill area. The two excavations remain and contain murky liquid.

Tree canopy obscures the drainage channel seen along the south edge of the landfill in 1968.

panying text provides a full site description. Figures 18.2 through 18.11 illustrate a standard historical site analysis of a hazardous waste disposal site, and show the kinds of information that can be extracted from an analysis of these photos, and changes at the site over time. The historical analysis clearly shows and documents the location of major fill areas which were evident during the life of the landfill. Disposal-related activities identified within the fill areas include debris, material, refuse, dark-toned liquid in erosion rills, excavations, and pits containing liquid. Between 1964 and 1984 a leachate collection impoundment was present on the site and a second leachate collection impoundment was present from 1968 through 1984. An upgraded leachate collection impoundment was present from 1980 through 1992.

Inventories of Potential Hazardous Waste Sites

Inventories of potential hazardous waste sites covering large areas and decades in time are a very cost-effective way to discover sites for future investigation. The aerial photographs are systematically searched for specific features or to identify types of sites. Type might include landfills, open dumps, scrap salvage yards, chemical handling and storage facilities, impoundments, or abandoned industrial sites. Identified sites are located on overlays to topographic maps accompanied by data sheets describing site conditions. The site conditions are presented chronologically with the period of site activity shown on the map overlay. This approach is helpful to determine the origin of a progressive problem and to identify a hazardous site that is currently hidden by new development.

Emergency Response

EPIC reacts through quick response capabilities to emergency situations such as hazardous material releases and natural disasters like hurricanes (Hugo) and earthquakes (San Francisco/Oakland, CA). Aerial photographs are flown, processed and analyzed to provide immediate information to on-site personnel regarding circumstances not easily or safely observed from the ground. Typical products for an emergency response include an immediate telephone report to on-site personnel followed by photographs or positive film transparencies with interpretation results annotated on overlays, annotated topographic maps, and a short letter report describing analysis results.

Wetlands

Wetlands analyses are performed by EPIC in support of various sections of the Clean Water Act concerning enforcement, permitting, and advance identification. Analysis of historical aerial photographs is often the only means of establishing the prior existence of wetlands on lands that have been dredged or filled, and for calculating wetlands loss acreage necessary for mitigation settlements. Aerial photographs also provide information concerning vegetative type, periodicity of flooding, tidal influences, and affected drainage patterns.

EPIC image analysts perform various types of wetlands mapping depending on the needs of the requesting EPA headquarters or regional program office. A *wetlands/upland* boundary delineation is the simplest form of mapping performed (Figure 18.10). The purpose for this delineation is to identify the location of wetlands and to separate wetlands from nonwetlands areas. A *detailed wetlands analysis* is performed to identify and classify various wetland types, often using the wetlands and deepwater habitat classification developed for the Fish and Wildlife Service by Cowardin et al. (1979). Using *jurisdictional wetlands delineation* procedures and associated fieldwork, EPIC maps wetlands in support of Section 404 of the Clean Water Act which protects wet-

lands from unpermitted dredge and fill activities. Also, in support of the *Advance Identification* process of Section 404, wetlands maps are prepared as a cost-effective way to identify wetlands in advance of permit application and evaluation. EPIC also performs image interpretation for wetlands delineation in support of EPA enforcement cases. Historical aerial photographs are used along with field checking to map wetlands losses and change due to filling and/or dredging activities. Photogrammetry is used to make accurate quantitative measurements of wetlands losses, extent of filling, and overall area changes. Geographic Information Systems (GIS) technology is applied to produce maps and graphics displays which are suitable for courtroom presentations and which clearly illustrate these changes.

Photogeology

EPIC employs aerial photointerpretation to study the geology of an area from an analysis of landforms, drainage, tones/textures, and vegetation distribution. EPIC conducts two types of photogeologic analysis and mapping: fracture trace and lithologic.

Fracture trace analysis involves the use of aerial photographs and other types of remote sensing imagery to identify linear features on the earth's surface that are naturally occurring and are surface manifestations of subsurface fracture zones in the bedrock (Figure 18.9A). When viewed in cross-section, fracture traces are seen to be vertical or near vertical breaks in the bedrock. Fractures are of particular environmental concern because contaminants are likely to move more easily through zones of fractured bedrock than through the surrounding more consolidated bedrock material. Thus, fracture traces can be used to identify possible migration routes of pollutants and are often used in the placement of monitoring/remedial wells around hazardous waste sites (Stohr et al., 1987; Scheinfeld et al., 1988; Mata and Christie, 1991).

EPIC conducts *lithologic mapping* (mapping of distinct rock types or units) from aerial photographs in order to produce a more accurate geologic map in areas where geologic mapping is incomplete due to limited fieldwork, small map scale, or other factors (Figure 18.9B). This procedure is usually performed by consulting available geologic maps of the area.

Photogrammetric Mapping

In support of EPA's mission, EPIC produces highly accurate topographic and planimetric maps, generally at a large scale, which conform with National Map Accuracy Standards and EPA Photogrammetric Mapping Specifications. Map scales, contour intervals, and planimetric details can be varied to suit specific requirements.

EPIC uses analytical stereoplotters, digital video plotters, or other digital photogrammetric methods to measure the area and volume of hazardous wastes; determine the height and placement of containment berms, dikes, and impoundments (Figure 18.8); and determine the depth of waste pits. Changes in size, shape, and other physical characteristics of a waste site are documented through sequential photogrammetric mapping (Slima, 1980).

Photogrammetric techniques are also used by EPIC to establish precise location and orientation data to support geophysical monitoring or for monitoring well placement.

Geographic Information Systems (GIS)

EPIC applies GIS technology to support a variety of EPA Headquarters or Regional Program Office Needs. For example, a National Priorities List (NPL) hazardous waste site investigation was performed using information from diverse sources (including numerous years of historical

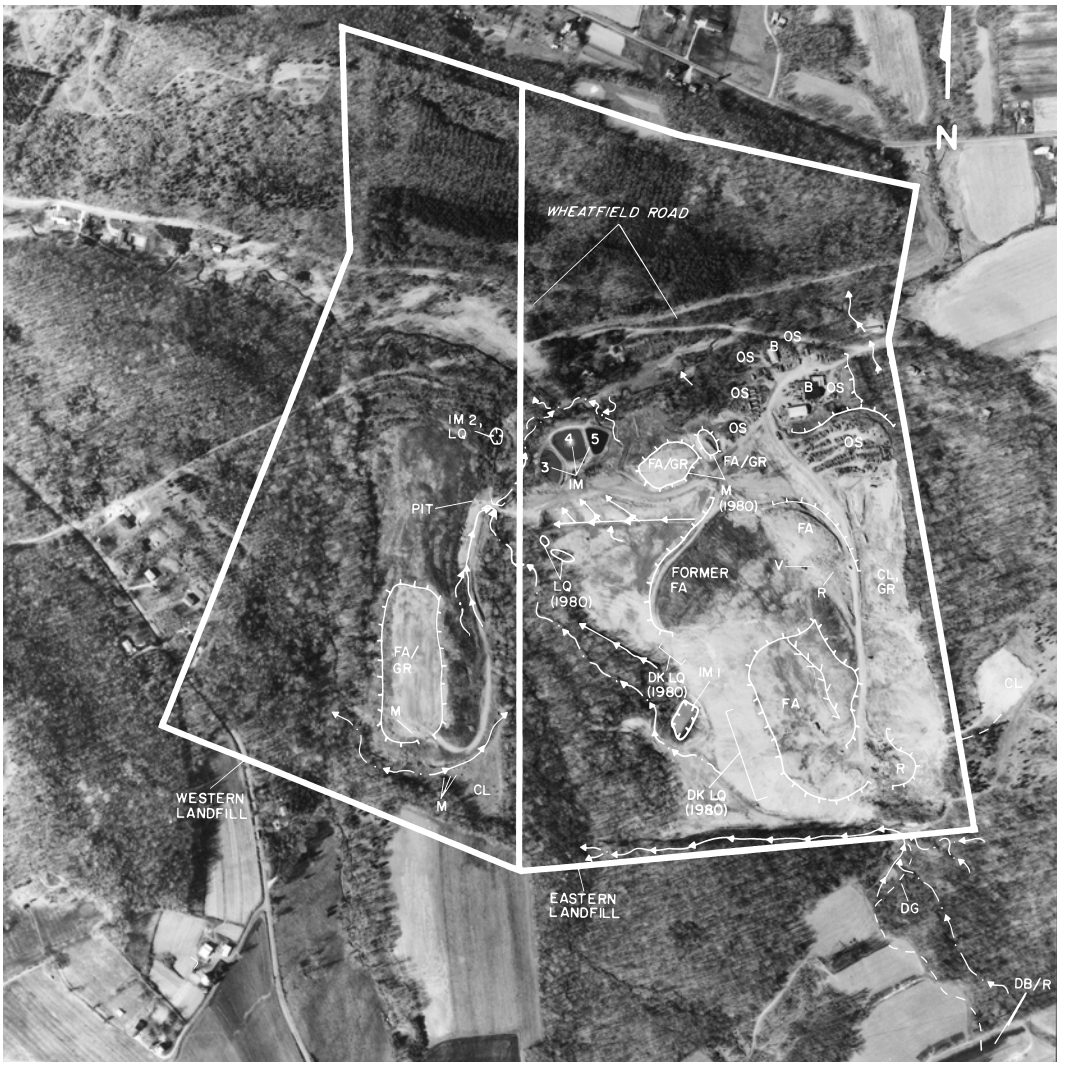


Figure 18.6. April 7, 1981 Significant features seen on the 1980 photograph are annotated and discussed with the following analysis on page 231.

Figure 18.6, continued. Western Landfill. Central portion. In 1980 IM2 contained a small amount of liquid. In 1981 there is an increase in the amount of liquid in IM2. The road (not annotated) on the east side of the landfill has been widened. A shallow pit, at the bottom of which appears to be a drain, possibly associated with the leachate collection system, is visible on the northwest side of the widened road.

Southern portion. In 1980 a clearing was visible on the west side. In 1981 the clearing (not annotated) is expanded and graded. A slight increase in the elevation indicates fill activity since 1980. Debris (not annotated) is evident along the edges of the fill area. Several piles of earthen-toned material are visible on the south side of the fill area, and in the clearing along the southeast edge.

Eastern Landfill. Northern portion. In 1980 the leachate collection system was visible in the large clearing where grading was seen in 1975. The leachate system consisted of IM3, IM4 and IM5, all of which contained liquid. One large and one smaller pile of material were seen in a former fill area which was used for open storage in 1975. This may have been excess soil dug out while excavating the leachate collection impoundments. An increase in open storage was noted around the buildings.

In 1981 IM3, IM4, and IM5 contain liquid. Both piles of material are graded. There has been a further increase in open storage since 1980.

Central portion. In 1980 extensive filling was evident. Dark-toned liquid was pooled in two areas on the west side. Partial clearing occurred along the east side. In 1981 filling is in progress in a new area. A vehicle (V) and refuse are visible in the new fill area, indicating current activity. The east side has been completely cleared and grading is in progress.

Southern portion. In 1980 extensive filling had occurred since 1975 and the east edge was partially cleared. Dark-toned liquid was visible in numerous erosion rills on the west side of the fill area. IM1 contained liquid, and its edge was more well-defined than in previous years, suggesting that the impoundment was dredged and redug. An access road led to a clearing east of the study area. Disturbed ground (DG) was noted along a second access road which led southwest of the study area, turned southeast, and continued to a wooded area where debris/refuse was evident. The same access road also continued southwest, and debris/refuse was evident at the edge of a second wooded area (not shown). The two wooded areas south of the study area are also accessible from roads outside of the landfill.

In 1981 further filling has taken place. The east edge is completely cleared and grading is in progress. The dark-toned liquid seen in 1980 is no longer visible. There is a significant increase in the grade east of IM1, which contains liquid. Refuse is abundant in the southeast corner.

The access road and clearing remain east of the study area. Disturbed ground remains evident along the access road seen south of the study area, and debris/refuse remains visible in both wooded areas.

The drainage channel previously seen along the south edge of the landfill is visible and contains liquid.

Figure 18.7. April 17, 1988. Significant activity observed on the 1983, 1984, and 1987 photographs is discussed, but not annotated in conjunction with the following analysis on page 233.

Figure 18.7, continued. Western Landfill. Central portion. In 1983 IM2 contained liquid, and the pit seen northwest of the widened road in 1981 was no longer visible. In 1984 IM2 contained liquid, and a small clearing was noted at the end of an access road. In 1987 IM2 was no longer visible, trees had been cut around the area where it was previously located, and revegetation was evident. The clearing seen at the end of an access road in 1984 was no longer visible. In 1988 no significant activity is noted.

Southern portion. In 1983 the clearing seen in 1981 was vegetated except for a small portion on the south side. The debris and material were no longer visible, and the clearing seen at the southeast edge in 1981 was vegetated. In 1984 the southern clearing remained and no further significant activity was evident. In 1987, an access road led west from the southern clearing to a ground scar at the creek. Clearings were seen on the southwest and southeast edges; debris was noted in the southwest clearing. In 1988 linear objects (LO), probably uprooted trees, are piled in the southern clearing. A series of objects, possibly straw bales for erosion control (not annotated), are noted between the southern clearing and a long clearing. This is the same location where an access road and ground scar were noted in 1981. Refuse is seen in the southeast clearing, and linear objects, probably uprooted trees, and refuse are seen in the southwest clearing.

Eastern Landfill. Northern portion. In 1983 IM3, IM4, and IM5 contained dark-toned liquid. Disturbed ground/light-toned material was noted southwest of the office buildings. In 1984 IM3, IM4, and IM5 contained dark-toned liquid. IM6 had been constructed and was empty. The disturbed ground/light-toned material was no longer visible. Clearing and fill activity were evident southeast of IM3, IM4, and IM6. In 1987 IM5 was full of dark-toned liquid, IM4 was partially filled with dark-toned liquid, IM3 appeared empty, and IM6 was surrounded by vegetation and contained dark-tone liquid. Extensive filling and grading occurred in the area southeast of the impoundments. In 1988 IM5 and IM6 contain dark-toned liquid and IM3 and IM4 contain medium-toned liquid. The vegetation has been cleared from around IM6 and the edges appear higher. No further filling or grading is evident in the area southeast of the impoundments.

Central portion. No significant change was noted in 1983 or 1984. In 1987 extensive filling and grading had occurred. In 1988 a series of excavations, possibly catchment ponds, containing murky liquid and linear objects, probably uprooted trees, are noted on the west side.

Southern portion. In 1983 IM1 remained and contained liquid. The access road and clearing remained active east of the study area. In 1984 IM1 remained and contained liquid, and trees were cleared west of IM1. The access road and clearing remained active east of the study area. In 1987 IM1 was no longer present. Disturbed ground was evident in the area where it was previously located. The access road and clearing east of the study area were vegetating. In 1988 no significant change is evident. Debris/refuse remains visible in both wooded areas south of the study area. Between 1984 and 1987 the drainage channel along the south edge of the site was rerouted northward through a culvert as shown in the Wetlands and Drainage Analysis (Figure 18.10). It continues west to the natural drainage, which flows north.

Figure 18.9a. Introduction. This fracture trace analysis presents findings based on the study of aerial photographs and geologic literature (see References) for the area on and around the landfill.

Fracture traces and other geologic lineaments are considered to be the surface expressions of vertical to near vertical zones of fracture concentration in bedrock. Fracture trace analysis is the technique of using aerial imagery for locating fracture traces or geologic lineaments on the earth's surface based on photogeologic signatures such as soil-tonal variations and vegetational and topographic alignments.

The landfill and surrounding area are astride the boundary separating the Great Valley section of the Valley and Ridge Physiographic Province. The bedrock geology of the landfill and surrounding area consists of Paleozoic and Triassic age sedimentary rocks, Triassic age igneous rock, and assorted contact metamorphosed rock.

Findings. Of particular note are the upper Cambrian carbonates (Millback Formation) which straddle the northern section of the site (Figure 18.9B). The Millback carbonates, reported to be highly permeable (well yields in the 225 gpm range), are surrounded by rock (diabase, meta sandstones and shales) having significantly lesser yields. Because of this, contaminants from the site may have difficulty moving out of the Millback aquifer, possibly concentrating in the Millback.

Efforts to characterize the groundwater regime around the landfill should in part focus on delineating the areal extent of the Millback (mapped by MacLachlan et al. as inferred). Although a photogeologic analysis of presite imagery (1946, 1947, and 1951) basically corroborates the MacLachlan et al. map, a more detailed field analysis is warranted.

Groundwater flow through the bedrock in this area is predominantly through secondary openings such as frac-

tures, and in the Millback perhaps through solutionized conduits. Identification of fracture traces is important when modeling aquifer contamination because the fracture traces may act as conduits for concentrated groundwater flow. Fracture trace analysis can be used to map possible pathways along which contaminants may migrate away from sites via subsurface routes. Once these possible pathways have been mapped, a monitoring well network can be designed to take into account the fracture pattern identified during the photogeologic analysis.

Figure 18.9b, above. A total of 25 fracture traces was identified using 1946, 1947, and 1951 imagery. They are presented on a photographic enlargement of the 1980 aerial photograph (Figure 18.9a). Of note are fracture traces labeled A, B, and C. A and B bisect the landfill area and therefore have a high likelihood of carrying contaminants. Fracture trace C, oriented north-south, is located along the expected direction of regional groundwater flow away from the site (west-north-west), and could deflect the groundwater flow away from this expected route.

Secondary openings in the Millback carbonates may be solutionized, resulting in an enhanced groundwater flow. Depending on the degree of solutionization, any cones of depression developed around wellheads can greatly influence groundwater flow direction. Caution should be exercised if fracture traces are to be used in the location of monitoring wells. Differences of a few feet can determine whether or not a well is located on a fracture trace; therefore, it is important that a geology field team and a photogeology analyst with stereo aerial photographic transparencies and related interpretation equipment be on site for precise placement of the monitoring wells or well network.

Recommendations. Based on the findings of this photogeologic analysis, future groundwater characterization efforts at the landfill should be directed toward: (1) defining the areal extent of the Millback Formation, (2) placing monitoring wells on fracture traces A, B, and C, and (3) determining the degree of solutionization in the Millback Formation.

Figure 18.10. Wetlands and drainage analysis. Wetlands (W), open water (OW), uplands (U), and drainage are annotated for the area at and around the landfill using photographs from April 17, 1988. The general direction of drainage flow is north and west through a series of unnamed streams which eventually lead into the north-flowing Main Creek (not shown). The on-site drainage flows through a series of wetlands that contain emergent and forested vegetation.

aerial photographs, geological data, digital line graph data, soil data, property ownership, monitoring well data, etc.). These data were combined to produce topical maps and analyses for use in the Remedial Investigation/Feasibility Study decision-making process under Superfund. Another project in support of the Environmental Monitoring and Assessment Program (EMAP) involved the ecological characterization of a pilot site using photo-derived information on land use, vegetation,

Figure 18.8. Landfill Area Measurements from Photogrammetry of Leachate Collection Impoundments.

YEAR	IM1	IM2	IM3	IM4	IM5	IM6
1971						
Hectares	0.143	0.073	not	not	not	not
Acres	0.353	0.181	present	present	present	present
1981		no				
Hectares	0.108	significant	0.095	0.145	0.077	not
Acres	0.267	change	0.235	0.359	0.189	present
1988						
Hectares	not	not	0.116	0.166	0.097	0.172
Acres	present	present	0.287	0.409	0.240	0.424
Increase (+) or Reduction (–)		no				N/A
Hectares	–0.035	significant	+0.021	+0.021	+0.020	(one year
Acres	–0.083	change	+0.052	+0.050	+0.051	measured)

Area measurements were obtained for impoundments 1–6 using photographs representing the year of peak landfill activity (1971), an intermediate year (1981), and the year of final closure (1988). The impoundments were digitized from photographic prints using an XY digitizing table. The digitizing table coordinates were transformed into the UTM coordinated system to facilitate area measurements, using a GIS.

IM1 was visible from 1964 through 1984, and decreased in size by 0.035 hectare (0.083 acre) between 1971 and 1981. IM2 was visible from 1968 through 1984, and was 0.073 hectare (0.181 acre) in 1971. Changes in the size of IM2 were minimal, and therefore have not been measured. IM3, IM4, and IM5 were visible from 1980 through 1992, and were each enlarged by approximately 0.021 hectare (0.05 acre) between 1981 and 1988. IM6 was visible from 1984 through 1992, and was 0.172 hectare (0.424 acre) in 1988.

wetlands, and landforms to produce maps and overlays for landscape characterization and trend analysis (Norton and Slonecker, 1990).

Global Positioning Systems

Global positioning systems (GPS) technology is used by EPIC primarily to produce accurate latitude/longitude coordinates for sites under investigation, and as a means to evaluate and quantify the spatial accuracy of digital map data, and to create accurate cartographic products. In 1990 EPA adopted its Locational Data Policy (LDP) for the purpose of ensuring the collection of accurate, fully documented locational coordinates as part of all agency-sponsored data collection activities. An accuracy goal of 25 meters was established and GPS is considered the best method for achieving this goal. GPS is used to provide accurate latitude/longitude ground control points for producing accurate maps from aerial photographs and to precisely locate hazardous waste disposal sites (Wells et al., 1986; Slonecker and Groskinsky, 1993).

Miscellaneous Analyses

Additional image analyses performed by EPIC in support of EPA needs include interpretation of thermal infrared imagery for detection of illegal river discharges and landfill and mine fires;

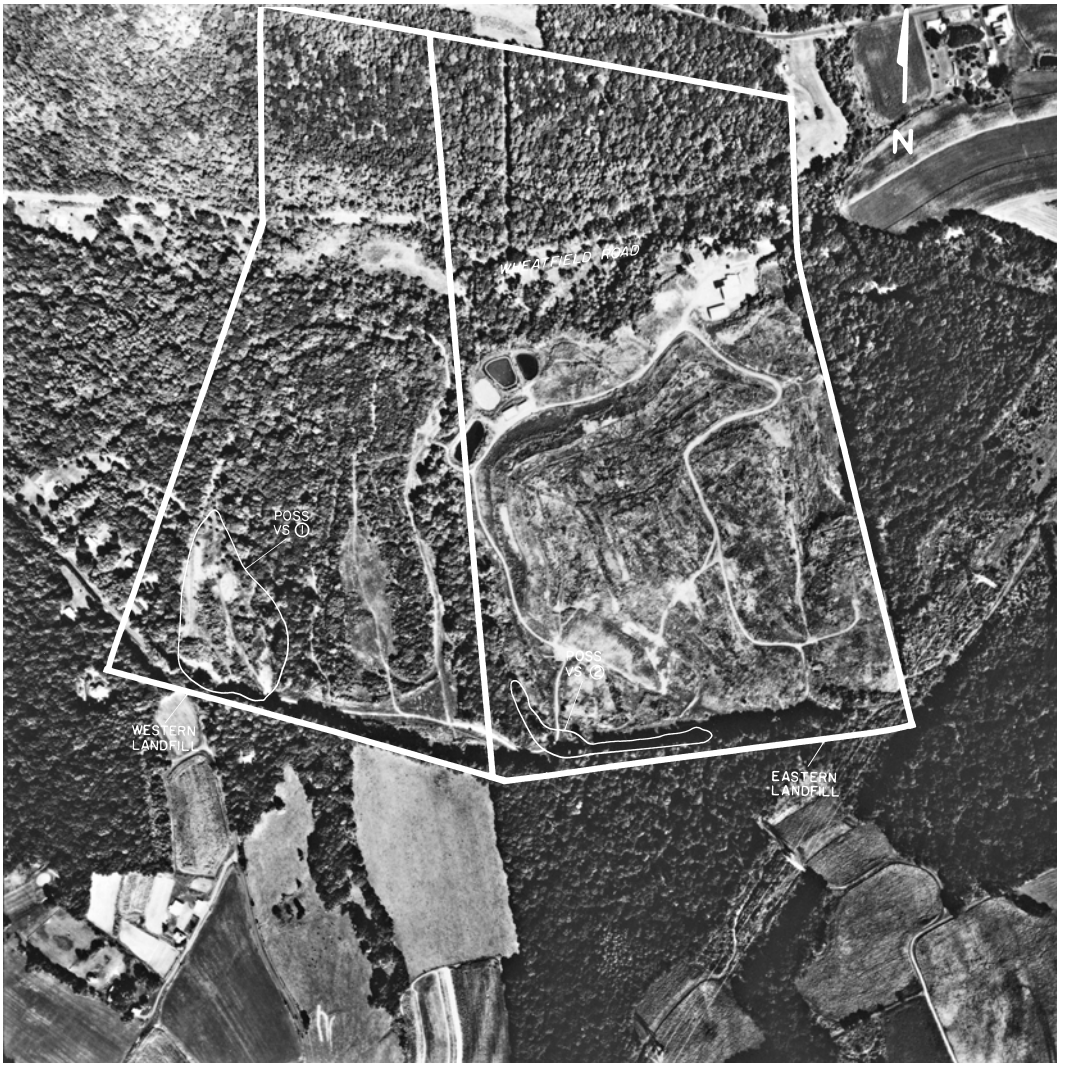


Figure 18.11. Vegetation stress analysis, September 24, 1992. A color infrared overflight was acquired for the landfill during full leaf-on conditions to facilitate the vegetation stress analysis. The photo shown here is a black and white copy of the color infrared image which normally shows vegetation as various shades of red.. Vegetation stress (V) can be caused by many interrelated factors such as insect infestation, disease, drought, water inundation due to natural and/or man-made changes in hydrology, and soil compaction, as well as contaminants. Often vegetation can be weakened by one factor and consequently become susceptible to one or more of the others. Due to the abundant rainfall and below-normal temperatures throughout the eastern United States during the summer of 1992, it is unlikely that vegetation stress has been caused by drought conditions.

Since individual dead trees are likely to be naturally occurring snags, they have not been identified. In some cases dead or stressed vegetation may have been overcome with climbing vines, resulting in a healthy vegetation signature.

A standard or "normal" signature was developed for the vegetation at the landfill by studying the areas at and around the site and establishing a generalized spectral reflectance pattern. Possible vegetation stress was identified based on deviations from the "normal" vegetation signatures and is discussed for the western and eastern landfills.

and aerial photointerpretation for detection of abandoned oil, gas, and water wells; mapping of submerged aquatic vegetation; and land use and drainage mapping.

Aerial Photo Acquisition

EPIC acquires historical photographs, dating to the late 1930s, from a wide range of federal, state, and local government agencies and private aerial survey companies. EPIC sets the specifications for overflight of new aerial photography, to meet the Agency's varying needs. New aerial photography is acquired through a network of private aerial survey companies across the country. These companies are also available on short notice for emergency response efforts.

SUMMARY

EPIC is a facility of the Landscape Ecology Branch, Environmental Sciences Division, National Exposure Research Laboratory, Office of Research and Development, and has been performing aerial photointerpretation and mapping in support of EPA headquarters and program offices since the early 1970s. Over the years the technology of remote sensing and mapping has advanced dramatically. EPIC has maintained pace with this technology and today provides the agency with the benefits derived from its use. An evolution has also been occurring with regard to environmental concerns. The health and well-being of our nation's ecological resources, and concerns about global climate changes have added a new dimension to the way we view the environment, and require methods for efficiently collecting data inexpensively and over large areas. Advances in remote sensing and mapping technology provide important tools to help us learn more about these regional and global problems. EPA's EPIC is poised and prepared to take on these new challenges through the next century.

Notice: The U.S. Environmental Protection Agency, through its Office of Research and Development (ORD) funds and performs the work described here. This document has not been sub-

Western Landfill. The overall vegetation patterns appear healthy, and a few small patches of irregular vegetation growth (not annotated) are evident. 1. Possible vegetation is noted in the most recently active clearing, which is located in the southeast corner of the landfill. Irregular patches of vegetation are growing within barren areas, and vegetation along the adjacent natural drainage route (Figure 18.10) appears stressed. The healthiest vegetation within the clearing is growing on top of the remaining probable uprooted trees noted in 1988.

Eastern Landfill. In general, the vegetation is growing irregularly, with healthy patches interspersed with possible stressed patches (not annotated), and numerous areas which are void of vegetation. 2. Possible vegetation stress is noted along the drainage route at the south edge of the landfill, and continuing west to the natural drainage (Figure 18.10). The lowlands along drainage routes are of particular importance because they offer a natural pathway for the migration of contaminants. It was not possible to determine deviations in the vegetation signature of the on-site wetlands due to the small size and number of these areas (Figure 18.10). A generalized spectral reflectance pattern can only be established when a number of similar areas within the site vicinity can be compared to each other.

The most effective way to verify possible vegetation stress and to investigate causes of vegetation stress is with a combination of aerial photograph analysis and a field check of the study areas during the active growing season.

jected to the Agency's review; therefore, it does not necessarily reflect the views of the EPA. The U.S. government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this article.

REFERENCES

Cowardin, L., V. Carter, F. Golet, and E. LaRoe, 1979. Classification of Wetlands and Deepwater Habitats of the United States. U.S. Department of Interior, Fish and Wildlife Service, FWS/OBS–79/31.

Erb, T., W. Philipson, W. Tang, and T. Liang, 1981. Analysis of landfills with historical airphotos. *Photogrammetric Engineering and Remote Sensing*, 47:1363–1369.

Evans, B., and L. Mata, 1984. Aerial photographic analysis of hazardous waste disposal sites. In *Hazardous Wastes and Environmental Emergencies*. HMCRI, Houston, TX.

Garofalo, D., and F. Wobber, 1974. Solid waste and remote sensing. *Photogrammetric Engineering and Remote Sensing*, 40:45–49.

Mata, L., and G. Christie, 1991. The use of remote sensing techniques in environmental assessments at CERCLA and RCRA facilities. In *Proceedings of the Eighth Thematic Conference on Geological Remote Sensing*, Vol. 2, Denver, CO, pp.931–944.

Norton, D., and E. Slonecker, 1990. The environmental monitoring and assessment program's landscape characterization database: New opportunities in spatial analysis. In *Proceedings of GIS/LIS'90,* Anaheim, CA.

Sabins, F., 1987. *Remote Sensing Principles and Interpretation*. W. Freeman Co., New York, NY.

Scheinfeld, R., L. Mata, P. Landry, and A. Logue, 1988. *Use of Fracture Trace Analysis in Thickly Mantled Karst Terraces and Their Solutions.* National Water Well Associate, Dublin, OH, pp.183–210.

Slima, C., 1980. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA.

Slonecker, T., and B. Groskinsky, 1993. Environmental applications of GPS technology. In *Proceedings of the Federation Internationale Des Geometres*, New Orleans, LA.

Slonecker, T., M. Lacerte, and D. Garofalo, 1999. The value of historical imagery. *EOM,* 8:39–41.

Stohr, C., W. Su, P. DuMontelle, and R. Griffin, 1987. Remote sensing investigations at a hazardous waste landfill. *Photogrammetric Engineering and Remote Sensing*, 53:1555–1563.

Wells, D., N. Beck, D. Dlikaraoglou, A. Kleusberg, E. Krakiwsky, G. Lachapelle, R. Langely, M. Nakiboglu, K. Schwarz, J. Tranquilla and P. Vanicek, 1986. *Guide to GPS Positioning*. Canadian GPS Associates, Fredericton, NB, Canada.

CHAPTER 19

Remote Sensing and GIS for Site-Specific Farming

John G. Lyon, Andrew Ward, Bruce C. Atherton,
Gabriel S. Senay, and Tom Krill

INTRODUCTION

There are a number of capabilities that have been postulated to be useful in agriculture. These capabilities have been realized through numerous efforts over the years (e.g., Moran et al., 1997). The focus now is to demonstrate the capabilities of available technologies for operational and practical applications.

The hope is that research can demonstrate that available remote sensor, GPS, and GIS technologies can supply good information for management of crops, soils, and waters on a within-field basis, and do so at reasonable cost. The goals are to potentially improve yields while maintaining or improving soil tilth and water quality (Ward and Elliot, 1995; Blackmer and Schepers, 1996; Gowda et al., 1999).

Precision agriculture approaches can be implemented using a suite of technologies. To address within-field management concerns it is necessary to navigate over short distances and small elevations. The advent of Differential Global Positioning Systems (Van Sickle, 1996) now provides for the required precision and accuracy. The need for mapping the spatial data collected through a variety of means is met by position information from GPS and mapping capabilities of Geographic Information Systems (GIS)(Lyon and McCarthy, 1995). GIS supplies the maps of soil grid sampling, management applications, and on-the-go yield, and can do so for each season and each year of evaluation with high accuracy (Bolstad and Smith, 1995). To supply detail as to crop, soil, and hydrological conditions over the growing season and from year to year, remote sensor data from aircraft or spacecraft are employed (Eidenshink and Haas, 1992; Thenkabail et al., 1992, 1995; Ward and Elliot, 1995; Yang et al., 1998). Use of different parts of the electromagnetic spectrum can help to separate types of crops and weeds, general soil and soil moisture characteristics, and the potential influences of hydrology on soils and crops (Huete and Tucker, 1991; Lyon et al., 1998).

Our work over the last 12 years and the work of others helps to illustrate the experimental and operational capabilities of remote sensor, GPS, and GIS technologies for agriculture. Here, the background to these methods, their results, and options for the future are described and discussed using our experience from three experiments on commercial farms in Ohio, and the experience of other researchers.

BACKGROUND

In this chapter we focus on the variables which are most useful in site-specific farming and can be best related to remotely sensed data. These techniques and methods can be incorporated with GIS and watershed models to answer important water resource questions. In general, the following variables can be measured in fields and related to remotely sensed measures of light: wet and dry biomass, crop residue, leaf area index (LAI), plant population density, areal cover of crop canopy, concentration or quantity of chlorophyll, leaf tissue nutrient content, crop height, plant moisture content, yield after harvesting, on-the-go yield, soil moisture content, soil texture, soil fertility, soil nitrogen, soil phosphorus, soil potassium, soil water release characteristics, the presence and location of weeds, wet and droughty areas, and plant stresses associated with diseases, insects, or other factors.

Crop Characteristics

The wet, green biomass of healthy plants can be measured in a number of ways remotely using the visible, near, and middle infrared regions of the electromagnetic spectrum (Singh, 1989; Thenkabail et al., 1994; Senay et al., 2000a, 2000b). A number of people make use of the fact that green plants absorb most of the red light from sunshine for photosynthesis, and reflect very little red light to the sensor above the crop (Price, 1992). In addition, green plants absorb little near infrared light, but reflect great quantities of it to the sensor above. This differential light reflectance can be used as a tool to identify green plants from the background materials including soils, crop residue, and water.

Conversely, plants that do not perform optimally or are "stressed" will often have less chlorophyll and be chloretic or yellow (Blackmer and Schepers, 1996). This decrease in chlorophyll can be detected by a decrease in red light absorbance and infrared light reflectance. This differential light reflectance can be used as a tool to identify green plants from the background materials including soils, crop residue, and water (Figure 19.1).

Residue

An important consideration to many farm management activities is the maintenance of crop residue cover during the nongrowing season. The reasons for this practice include decreasing soil erosion from water and wind, reduction of plowing costs, decreasing water quality problems downstream, and others. The implementation of crop residue cover or conservation tillage practices has been of great interest, as have methods for measuring the extent of this practice (Messer et al., 1991; Jakubauskas et al., 1992). We have conducted work on measuring crop residue to track the implementation of conservation tillage, and to better understand residue influences on spectral responses during the growing season. This work began in the early 1980s and continues today (Thenkabail et al., 1992, 1994; Van Deventer et al. 1997; Gowda et al., 1999).

Crop residue or senescent plants reflect different amounts of light as compared to green plants or soil. It is possible to identify crop residue because of these differences, that include the absence of chlorophyll, presence of nonchlorophyll plant pigments, difference in water holding capabilities as compared to live or green plants, different leaf or stalk structure as compared to live plants, and different water holding, pigment, and structural characteristics as compared to soils. These differences manifest themselves as a distinct spectral signature or differential light reflectance that can be measured in the visible, near, and middle infrared regions of the spectrum (Ward and Elliot, 1995).

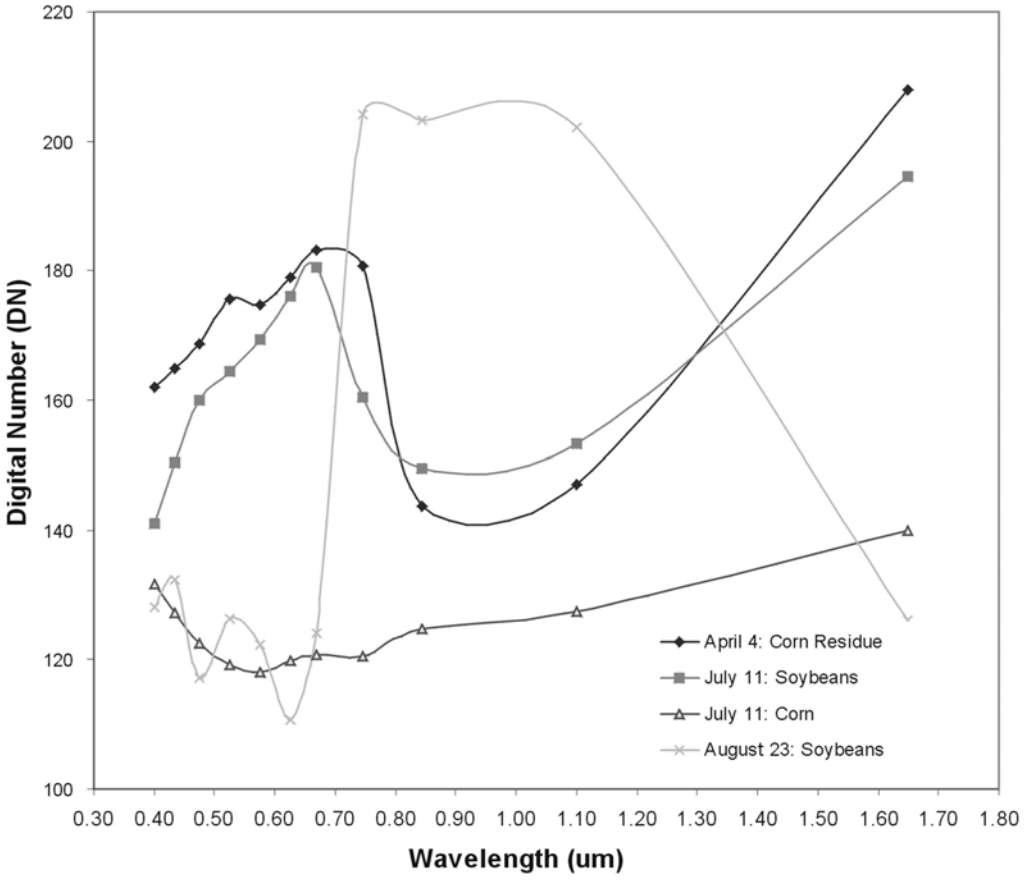


Figure 19.1. Spectral responses of corn and soybeans during vegetative growth stages, soybeans at maturity and corn residue.

Soils

The reflectance characteristics of soils are related to the parent materials, the texture, the moisture content, organic matter, and to a certain extent slope and elevation. Soils are generally lighter or brighter toned than plants and water. The exceptions would include soils of high organic matter or low bulk density or high moisture content. Bright tones of soils can be distinguished from plant materials in the visible and infrared portions of the spectrum. Work on using remote sensing to study a wide range of soil properties has been conducted by many researchers (Hatfield and Pinter, 1993; Moran et al., 1997)

An important consideration is that a dense crop canopy will obscure the bright tone of soils. This supplies a ready method to make estimates of canopy closure or leaf area index due to the darkness of the plant biomass obscuring the soil background. Conversely, measures of the presence of bare soil can help assess the absence of crop or crop canopy that has not developed (Lyon et al., 1986).

Water in a Crop Environment

Water characteristics of plants and soils can be identified and differentiated from other materials due to the fact that water or wet soils or wet plants have a lower reflectance of light to the sensor above (Lyon and McCarthy, 1995; Ward and Elliot, 1995). In general, the presence of water in any concentration decreases the light reflectance of materials. This is well known for the case of soils, where the presence of either water at the surface, water saturation, or standing water will greatly reduce the generally relatively bright reflectance of dry or relatively dry soils (Lyon, 1993).

A major use of remote sensor data is in the evaluation of crop moisture conditions. In particular, the middle infrared (approximately 1.5 and 2.2 μm) and to a certain extent the near infrared (approximately 0.7 to 1.1 μm) supply good detail as to relative plant moisture conditions of the leaves and stalk.

METHODS

Several experiments were developed over time to test the capabilities of remote sensor data for precision agriculture. The results of those experiments are presented here and in other publications. In general, the experiments included: evaluations of satellite and ground measurements of crop, soil, and hydrology variables for more than 50 commercial farms in Seneca County of north central Ohio (Thenkabail et al., 1992, 1994; Van Deventer et al., 1997); evaluations of crop, soil, hydrology, and weather characteristics using fine spatial resolution (1 m picture elements or pixels) and moderate spectral resolution (12 spectral bandwidths) for a commercial farm in Pike County of south central Ohio used in the Midwest Systems Evaluation Area (MSEA) studies (Senay et al., 1998, 2000a, and 2000b); and evaluations of low-cost aerial photographs taken from low altitudes of four commercial farms and one experimental farm station in Van Wert County of west central Ohio, and the MSEA site. Details on the methods to collect and analyze field data for these experiments are reported in several publications (Thenkabail et al., 1992, 1994; Nokes et al., 1997; Senay et al., 1998).

Landsat Thematic Mapper Data

During a three-year period, ground and satellite data were collected and analyzed for commercial farms in Seneca County (Thenkabail et al., 1992, 1994; Van Deventer et al., 1997). The goal was to evaluate the utility of Landsat satellite Thematic Mapper (TM) data for crop, soil, residue, and hydrology characteristics. This large experiment involved many field data collections during the growing season as detailed elsewhere. For remote sensing purposes, the research team visited the county on a biweekly basis to collect data during the Landsat satellite overpasses. The result was a very rich data set to evaluate field and satellite measured characteristics of crops, soils, and management practices.

Landsat 5 TM data and now Landsat 7 ETM+ data provides 30 m by 30 m resolution data every 16 days or so, and some type of Landsat data has been available for 25 years. It provides information in the visible, infrared, and thermal wavelengths.

Airborne Multispectral Scanner Data Collection

Several overflights of the MSEA site in Pike County were conducted in 1994. The platform was an Aero Commander twin engine aircraft with a camera bay holding the Daedalus Multispectral Scanner (MSS) instrument (model 1260) and an aerial camera by Wild (RC–8). The overflights were made on April 4, July 11, August 15, August 23, and September 15, and corresponded to

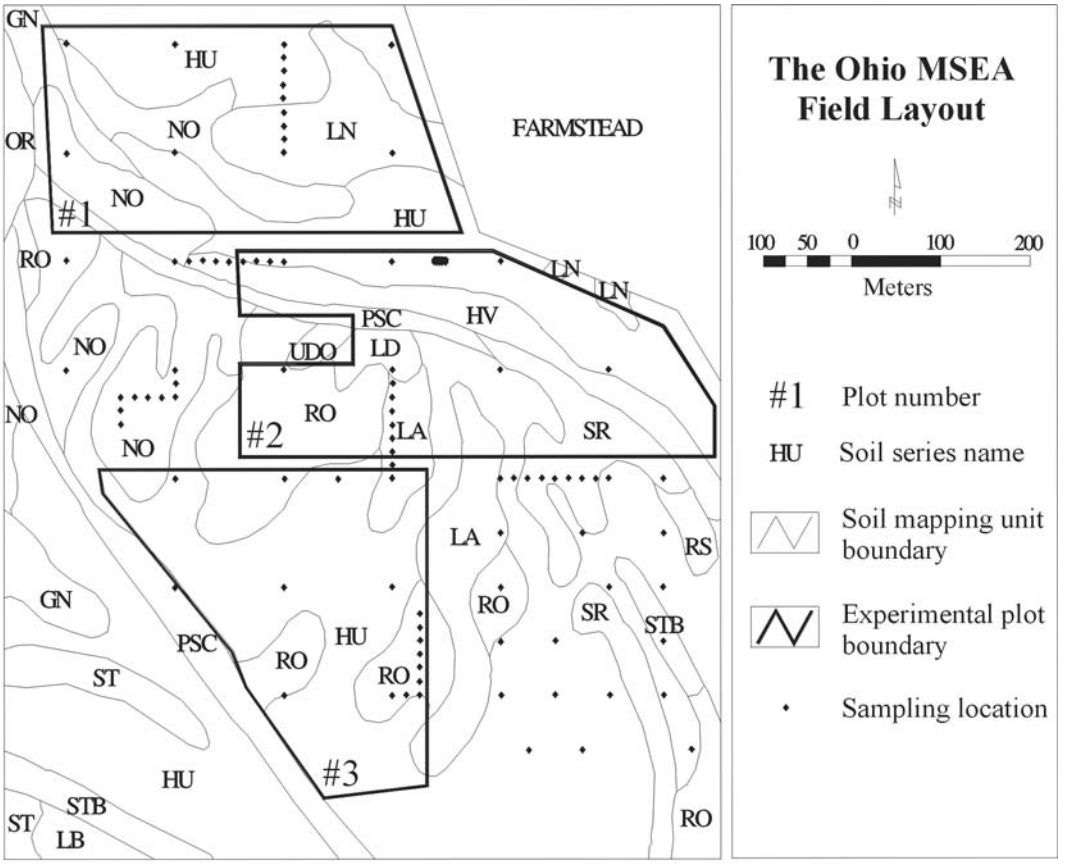


Figure 19.2. Ohio MSEA soil map (Wu et al., 1996). Source: US Department of Agriculture, Natural Resources Conservation Service, Columbus, Ohio.

early spring and preplanting, early planting and germination, mature crop, and senescent soybean, and early senescent corn crop growth periods, respectively. Ground data were collected close in time to the overflights, by a crew that measured the MSEA on a regular basis in support of a number of studies, and by personnel concerned specifically with the overflights.

Other remote sensor data were collected during 1994 and have been analyzed. These data included airborne radar data in X, C, and L-bands from a Lockheed P–3 Orion aircraft and sensor supplied by the Navy, and the hyperspectral sensor AVIRIS flown by a NASA U–2 aircraft.

The MSEA effort is ongoing, and is significant because of the extensive data collection effort in the field. In addition, products such as a Digital Elevation Model (DEM) were developed from a Differential Global Positioning System (GPS) experiment, on-the-go yield measurements were made from a combine with GPS capability, and a very extensive soil type map was developed for the site (as shown in Figure 19.2, Wu et al., 1996), and much of these data were processed for analysis using Geographic Information System (GIS) technologies (Senay et al., 1998).

Low Altitude, Small Film Format Data

The Van Wert County area was flown approximately six times before and during the growing season of 1997. A single engine, six place aircraft was supplied by the local airport flight service for these flights of approximately 45 minutes duration. One or two photographers collected photo-

graphs from approximately 1,000 feet of altitude above ground level (AGL). Each used a single lens reflex type 35 mm camera of general manufacture. The optimal lens was a zoom telephoto with a focal length of approximately 100–180 mm or similar sort of lens. The film types included: black and white film for interpretation and publication, color slide film for interpretation and presentation to audiences, color infrared film for interpretation, and color print film for interpretation and for quick processing and printing to support immediate analyses (Lyon, 1993). The color infrared film was taken at ASA 100 to 125, using a yellow (minus blue light) filter and processed by Rocky Mountain Films in Aurora, CO. The color print and slide films were approximately 100 ASA and were taken with a polarizing filter to reduce atmospheric haze.

In 1998, a study which focused on evaluating crop responses to spatial variability was conducted at the Ohio MSEA site. As part of this study digital infrared data was obtained in May, August, and September. This information has a resolution of about 0.3 m by 0.3 m and was obtained from commercial services which use cameras mounted on aircraft flying at low altitude. We also used an AgLeader yield monitor to obtain on-the-go yield information. Scouting was performed every 1–4 weeks and extensive crop and soil data were obtained on several occasions at 27 to 36 sites for each of the three farming systems which were located on 9 ha fields.

Field Data Collection

The methods to collect and analyze field data are described in a number of publications cited here. In general, the following variables can be measured in the field and related to remotely sensed measures of light. They include: wet and/or dry biomass, crop residue by field scouting method or by removal and weighing, leaf area index (LAI), number of plant stocks in rows, areal cover of crop canopy, concentration or quantity of chlorophyll, crop height, plant moisture content, soil moisture content, soil color or reflectance, yield after harvesting, on-the-go yield, soil organic matter content, soil texture, soil fertility, soil nitrogen, soil phosphorus, soil potassium, presence and location of weeds, presence or location of wet areas during the growing season, presence of sandy or rocky or salty or other low quality soils, and related measures.

RESULTS

Crop Characteristics

All three experiments have demonstrated the capabilities of remote sensor measurements to supply information about crops. Whether the sensor was deployed in space, or in an expensive aircraft or in a relatively low-cost aircraft the analyses yielded information.

Remote sensing methods were useful in measuring the presence of wet biomass and dry biomass. The results include very good relationships between wet biomass and satellite and aircraft measured light (Price, 1992, 1993; Thenkabail et al., 1992, 1994; Van Deventer et al., 1997; Seney et al., 2000a, 2000b), and with dry biomass (Thenkabail et al., 1992).

A particularly valuable analysis approach demonstrated by our work and the work of a number of researchers was the use of ratio indices of remote sensed variables as compared to ground measured variables (Eidenshenk and Haas, 1992; Qi et al., 1993; Hatfield and Pinter, 1993; Lyon et al., 1998; Lunetta and Elvidge, 1998; Lyon, 2000, 2001). The most famous and widely used examples are those that employ the division of the red portion of the spectrum with that of the near infrared (Thenkabail et al., 1992; Price, 1994; Lyon et al., 1998). The normalized difference vegetation index (NDVI) has yielded very good results in our work with wet biomass (Thenkabail et al., 1992, 1994) and other ground measures (Van Deventer et al., 1997).

Other ratio indices have been tested by us (Thenkabail et al., 1992, 1994; Van Deventer et al., 1997; Lyon et al., 1998; Senay et al., 2000a, 2000b) and by others. Using combinations of middle infrared (TM5), red (TM3), and near infrared (TM4) provided information on crop moisture conditions or "stress" [e.g., STVI, Landsat (TM5 x TM3)/TM4,Thenkabail et al. 1992]. However, the NDVI is the most commonly used index and has been found to be the optimal example in several studies (Thenkabail et al., 1992; Lyon et al., 1998).

Yield and Crop Stresses

The low-altitude and low relative cost method of flying 35 mm photographs has demonstrated very good detection of spring, perennial weed growths. The optimal method is to use color print film and a 35 mm camera with a local aircraft service (Figure 19.3), and results are available within 24 hours or less of the flight. This method allows for a timely flight when weed conditions are obvious as compared to newly emerged or newly germinated crop, and provides adequate time to develop a plan and execute spraying of weeds. A similar approach can be used with infrared film but processing usually takes several days or longer (Figure 19.4a). Alternatively, digital cameras can be used and the data downloaded directly to a personal computer for near real-time analyses.

Remote sensing has the potential to be used to determine yield variability. Our earlier studies with Landsat TM data often showed low correlation between multispectral data and yields (Thenkabail et al., 1992, 1994). However, in the recent experiment at the MSEA site there was a strong linear correlation between yields and the near-infrared bands (Senay et al., 1998). It is possible that this improved ability to use spectral data to determine crop yields is associated with the higher resolution (1 m compared to 30 m) and with the availability of on-the-go yield data for use in developing relationships between yields and the spectral data.

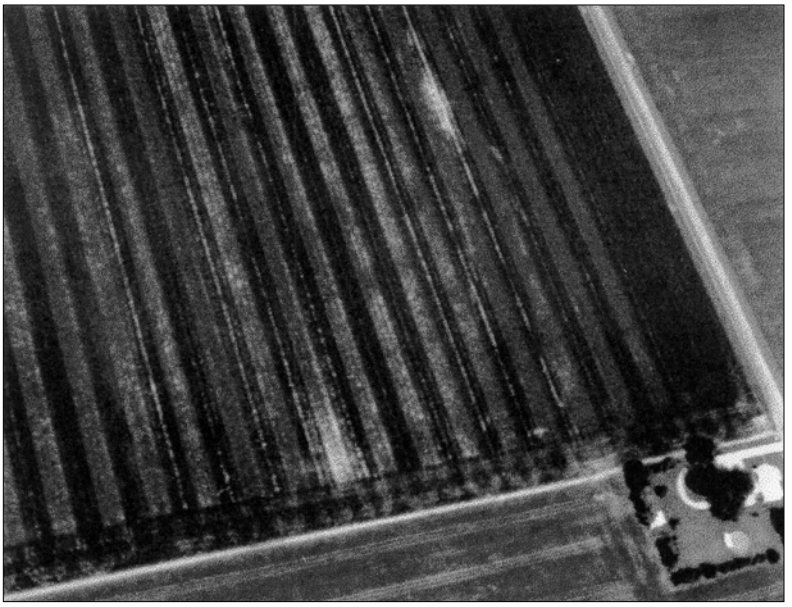
Whether these types of relationships can be used to accurately determine yields in fields where on-the-go yield data are not available remains to be seen. However, it appears that even low cost infrared data obtained from a camera mounted in an aircraft will provide useful information on relative yield and biomass differences within a field. Figure 19.4b shows reduced crops in areas with high weed pressures (Canada Thistle and Giant Ragweed) that are presented in Figure 19.4a.

Figures 19.5a,b,c and d, show yield data and infrared images of a cornfield at the MSEA site during different growth stages. A well is located in the largest bare area, which is connected to an access road, and a power pylon is located in the other bare area (lower corner). This set of figures illustrates both the usefulness and the difficulties in using remotely sensed data. The August image suggests that there is little spatial variability in the crop characteristics. This image was taken when mature plants had a high moisture content and were fairly green. In contrast, the September image depicts senescent plants and considerable spatial variability is evident. Unlike the information presented in Figures 19.4a and 19.4b the variable crop responses in this field are primarily due to soil differences. In particular the soils in the southern (lower) part of the field are coarser, have lower soil water retention characteristics, and are less fertile than soils in other parts of the field. In Figure 19.5b, the small green patches in the middle of the field and near the bottom right-hand corner are areas with low plant densities. They correspond well to the lowest yielding (dark green) patches in Figure 19.5d.

Residue

We have used satellite Landsat TM data and high resolution multispectral data to study residue. On a field basis, in the northern Ohio study, we were able to differentiate between contrasting soil plains, and conventional versus conservation tillage (Van Deventer et al., 1997). However, we re-

a.



b.

Figure 19.3. Low altitude aerial photographs of farm fields in Van Wert County showing weeds in a field, and poor management resulting in low seeding, germination, or fertilizer applications in some rows.

cently tested the same linear logistic regression model on the Maquoketa River Watershed in Iowa and there was poor agreement between remote sensed based estimates and those from field surveys (P. Gowda, personal communication).

Soil Characteristics

A valuable result of the use of remote sensor data is in the evaluation of soil moisture conditions. In particular, the middle infrared (approximately 1.5 and 2.2 μm) and to a certain extent the

near infrared (approximately 0.7 to 1.1 μm) supply good detail as to relative soil moisture conditions (Huete and Tucker, 1991; Van Deventer et al., 1997; Seney et al., 2000a, 2000b).

Soil fertility is an important issue in precision agriculture. Variables related to soil fertility are often not measured directly with a remote sensor, but are estimated from remote sensor variables using a related or surrogate variable (Lyon and McCarthy, 1995). Relative soil moisture and soil organic matter of bare soils can be measured directly. Relative soil phosphorus is measured in the field and compared to a surrogate variable measured remotely such as crop "greeness" or biomass, or leaf area index (Blackmer and Schepers, 1996).

Greater spectral resolution capabilities have allowed more detailed work on soil characteristics using indices similar to those previously published. Slightly modified indices that use the increased capabilities of Multispectral Scanner (MSS) aircraft data have proved useful. For example, the Simple Normalized Difference (SND) of the middle infrared (1.55–1.75 μm) divided by the near infrared (NIR, 0.92–1.10) which is similar to the Simple Vegetation Index (SVI) (Senay et al., 1998), and the Normalized Difference (ND) index using the equation (MIR-NIR)/(NIR + MIR) which is similar to NDVI in formulation (Seney et al. 2000a), have demonstrated increased capabilities. Both the SND and ND showed strong correlations with combined water in the soil (0.5 to 2.0 cm in depth) and in plants for corn (Senay et al., 2000a, 2000b).

Water in a Crop Environment

Most of our research has included an emphasis on water resource characteristics as they influence crops, soils, and water quality. These efforts have involved analyses of surface soil moisture, near surface drainage conditions, and vadose zone and groundwater conditions in support of watershed modeling (Ward and Elliot, 1995; Gowda et al., 1999). Remotely sensed measures have been valuable for a number of analyses. These include use of remote sensors or photography to document: wet soil conditions that may influence planting or germination; remote measurements of the impact of drought (Thenkabail et al., 1992); identification of land cover types in watersheds to help parameterize watershed models including GIS-based models (Moore et al., 1993; Gowda et al., 1999); calculation of watershed elevations for modeling (Garbrecht and Martz, 1993); remote measures of soil surface moisture (Van Deventer et al., 1997; Senay et al., 2000a, 2000b); use of remote sensing and GIS for nonpoint source modeling (Jakubauskas et al., 1992; Kang and Bartholic, 1994; Gowda et al., 1999); the accuracy of the remote sensor data or GIS products (Bolstad and Smith, 1995; Congalton and Green, 1998); and others.

Low altitude, 35 mm camera data can also identify drainage problems and the presence of subsurface drainage (Figure 19.6). During an early season overflight in Van Wert County, the weather and hydrology conditions were favorable for identification. Drainage systems were evident as light-toned areas which experienced better drainage than adjacent wet soil areas. The overflight occurred after a particularly wet spring, and about two days after the last rain. The drained areas were light toned because of the low relative moisture and hence increased reflectance as compared to adjacent, wetter soil areas (Figure 19.6).

General Characteristics of Fields

Statistical procedures and GIS technologies can be used to categorize spectral data into classes showing different crop characteristics, soils, topography and/or hydrological conditions. Figure 19.1 (see color section) shows more than 30 classes for the MSEA site which are based on the fine resolution visible, near infrared and middle infrared and elevation difference from the Digital Elevation Model (DEM). The classes themselves include measures of different levels (stalking, bio-

Figure 19.6. Low altitude aerial photographs of farm fields in Van Wert county showing wet areas, and the influence of subsurface drainage as lighter-toned linear features.

mass) of crop, differences in soils and elevation, and difference in plant and soil moisture conditions. The soybean crop is senescent and the corn is mature. The variety of spectrally distinct classes shows that a lot of detail can be sensed with the MSS sensor at one meter resolution and the additional discrimination power supplied by the data on relative elevation. Figure 19.7b is a false color composite of the fine resolution data and shows many of the same features that can be seen in Figure 19.1 (see color section). However, it provides less information on why these features occurred. These sorts of products can make a great amount of detail more obvious, and that same quantity of detail can be analyzed and evaluated as to accuracy (Lyon and McCarthy, 1995; Congalton and Green, 1998).

DISCUSSION AND CONCLUSIONS

The value of remotely sensed measurements lies in the utility of the data to identify normal and abnormal conditions in fields, and to do so rapidly over large areas. Producers know the characteristics of their fields, but things sometimes change rapidly and additional within-field data of a quantitative nature can only help the thoughtful management of crops and fields. The advent of Differential GPS positioning (Blackmer and Schepers, 1996) and on-the-go yield monitoring and more soil sampling allows the collection of powerful data sets. Combining these measurements with remotely sensed data and processing it all with GIS maps and modeling bodes well for detailed field management at low relative costs. Remotely sensed data can take the form of very low cost data either from shared satellite data collection and analysis efforts, or from low cost and rapid aerial photography from available local aircraft or commercial services (Lyon et al., 1986, 1995).

The best way to get started using precision agriculture technologies is to use some of the tech-

nologies to solve a specific field problem. The problem may be weeds, or variable fertility of soils, or poor germination. Use the technologies available to identify the locations of problems within the field and devise a solution and apply the solution. It is best to conduct the effort over several years to be sure that the result is not a chance situation with weather or some practice conducted in one certain year.

Infrared, color, or multispectral data acquired a few weeks after germination should provide information on weeds, soil spatial variability, and crop responses to this variability. Remote sensing data acquired during senescence is useful in identifying variability in crop densities and biomass. Data acquired during these two times and at other times during the growing season is also helpful in determining what caused the variability in crop responses.

Many experiments to characterize the potential of a given technology tend to be relatively expensive (Falkner, 1995; Lyon et al., 1995). As the technology and its utility become known, the researcher and industry then seeks to deliver this knowledge or technology at a cost of a few dollars per hectare. If the solution improves the yield in absolute terms, or removes the year to year variability of field yields, costs of this magnitude are economically viable.

A variety of results reported here or elsewhere have demonstrated that available remote sensor and GIS technologies can supply good information for management of crops, soils and waters on a within-field basis. Precision agriculture approaches can be implemented using a suite of these technologies, over time, to potentially improve yields and maintain soil and water quality.

ACKNOWLEDGMENTS

The authors wish to thank a number of people and sponsors for assisting in the development of these results.

The U.S. Environmental Protection Agency has helped to supply funding and aircraft remote sensor data through the National Exposure Research Laboratory Division at Research Triangle Park, NC and Las Vegas, NV, and through the assistance of EPA scientists Ross Lunetta and Clay Lake.

Some of the analysis of aerial photographs and farm conditions was conducted as a student Service Learning project at OSU, as part of a larger national effort called Learn and Serve America by the Ohio Campus Compact via the national Campus Compact at Brown University, as funded by the Corporation for National and Community Service.

Our thanks also go to our Departments at Ohio State University, OSU Extension, Van Wert County air service, Larry Lotz, Nathan Watermeier, Prasanna Gowda, and Jay Johnson.

BIBLIOGRAPHY

Blackmer, T., and J. Schepers, 1996. Using GPS to improve corn production and water quality. *GPS World*, March, pp. 44–52.

Bolstad, P., and J. Smith, 1995. Errors in GIS, assessing spatial data accuracy. In Lyon, J. and J. McCarthy, 1995.

Congalton, R., and K. Green, 1998. *Accuracy Assessment of Remotely Sensor Data: Principles and Practices*. CRC/Lewis Publishers, Boca Raton, FL.

Eidenshink, J., and R. Haas, 1992. Analyzing vegetation dynamics of land system with satellite data. *Geocarto International,* 1:53–61.

Falkner, E., 1995. *Aerial Mapping Methods and Applications*. Lewis/CRC Publishers, Boca Raton, FL, 352 pp.

Garbrecht, J., and L. Martz, 1993. Network and subwatershed parameters extracted from digital elevation models: The Bills creek experience. *Water Resources Bulletin,* 29:909–916.

Gowda, P., A. Ward, D. White, J. Lyon, and E. Desmond, 1999. The sensitivity of ADAPT model predictions of streamflows to parameters used to define hydrologic response units. *Transactions of the American Society for Agricultural Engineers,* 42:381–389.

Hatfield, J., and P. Pinter, 1993. Remote sensing for crop protection. *Crop Protection,* 12:403–413.

Huete, A., and C. Tucker, 1991. Investigation of soil influences in AVHRR red and near-infrared vegetation index imagery. *International Journal of Remote Sensing,* 12:1223–1242.

Jakubauskas, M., J. Whistler, M. Dillworth, and E. Martinko, 1992. Classifying remotely sensed data for use in an agricultural nonpoint source pollution model. *Journal of Soil and Water Conservation,* 47:179–183.

Kang, Y., and J. Bartholic, 1994. A GIS-based agricultural nonpoint source pollution management system at the watershed level. *Proceedings of the Annual ACSM/ASPRS Convention*, Reno, NV, pp. 281–289.

Lunetta, R., and C. Elvidge, 1998. *Remote Sensing Change Detection*. Ann Arbor Press, Chelsea, MI.

Lyon, J., 1993. *Practical Handbook for Wetlands Identification and Delineation.* CRC Lewis Publishers, Boca Raton, FL, 157 pp.

Lyon, J., 2003. *GIS for Water Resources and Watershed Management*. Taylor & Francis, London.

Lyon, J., and J. McCarthy, 1995. *Wetland and Environmental Applications of GIS*. Lewis Publishers, Boca Raton, FL.

Lyon, J., E. Falkner, and W. Bergen, 1995. Cost estimating photogrammetric and aerial photography services. *Journal of Surveying Engineering,* 121:63–86.

Lyon, J., J. McCarthy, and J. Heinen, 1986. Video digitization of aerial photographs for measurement of wind erosion damage on converted rangeland. *Photogrammetric Engineering and Remote Sensing,* 52:373–377.

Lyon, J., D. Yuan, R. Lunetta, and C. Elvidge, 1998. A change detection experiment using vegetation indices. *Photogrammetric Engineering and Remote Sensing,* 64:143–150.

Messer, J., R. Linthurst, and W. Overton, 1991. An EPA program for monitoring ecological status and trends. *Environmental Monitoring and Assessment,* 17:67–78.

Moore, I., A. Turner, J. Wilson, S. Jenson, and L. Band, 1993. GIS and land surface-subsurface process modeling. In M. Goodchild, et al., Eds., *GIS for Modeling*, Oxford University Press, Oxford, UK, pp.196–230.

Moran, M., Y. Inoue, and E. Barnes, 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment,* 61:319–346.

Price, J., 1992. Estimating vegetation amount from visible and near infrared reflectances. *Remote Sensing of Environment,* 41:29–34.

Price, J., 1993. Estimating leaf area index from satellite data. *IEEE Transactions on Geoscience and Remote Sensing,* 31:727–734.

Qi, J., A. Huete, M. Moran, A. Chehbouni, and R. Jackson, 1993. Interpretation of vegetation indices derived from multi-temporal SPOT images. *Remote Sensing of Environment,* 44:89–101.

Senay, G., A. Ward, J. Lyon, N. Fausey, and S. Nokes, 1998. Manipulation of high spatial resolution aircraft remote sensing data for use in site-specific farming. *Transactions of the American Society of Agricultural Engineers,* 41:489–495.

Senay, G., J. Lyon, A. Ward, and S. Nokes, 2000a. Using high spatial resolution multispectral data to classify corn and soybean crops. *Photogrammetric Engineering and Remote Sensing,* 66:319–327.

Senay, G., A. Ward, J. Lyon, N. Fausey, and L. Brown, 2000b. The relations between spectral data

and water in a crop production environment. *International Journal of Remote Sensing,* 21:1897–1910.

Singh, A., 1989. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing,* 10:989–1003.

Thenkabail, P., A. Ward, J. Lyon, and C. Merry, 1994. Thematic Mapper vegetation indices for determining soybean and corn growth parameters. *Photogrammetric Engineering and Remote Sensing,* 60:437–442.

Thenkabail, P., A. Ward, J. Lyon, and P. Van Deventer, 1992. Landsat Thematic Mapper data (TM) indices for evaluating management and growth characteristics of soybeans and corn. *Transactions of the American Society of Agricultural Engineers,* 35:1441–1448.

Van Deventer, P., A. Ward, P. Gowda, and J. Lyon, 1997. Using Thematic Mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing*, 63:87–93.

Van Sickle, J., 2001. *GPS for Land Surveyors*. Ann Arbor Press, Chelsea, MI.

Ward, A., and W. Elliot, 1995. *Environmental Hydrology*. Lewis Publishers, Boca Raton, FL.

Wu, Q., A. Ward, and S. Workman, 1996. Using GIS in simulation of nitrate leaching from heterogeneous unsaturated soils. *Journal of Environmental Quality,* 25:526–534.

Yang, C., G. Anderson, and J. Everitt, 1998. A view from above: Characterizing plant growth with aerial videography. *GPS World*, April, pp. 34–37.

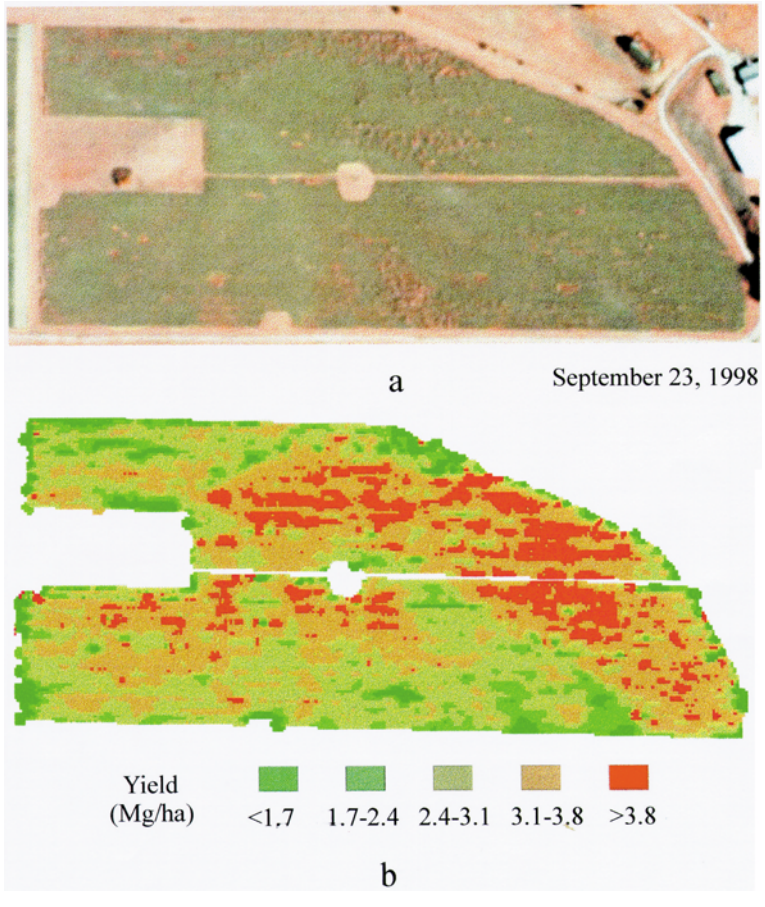


Figure 19.4. Digital infrared image (a) for a soybean field at the MSEA site (1998). Green areas in the original infrared image are senescent soybeans and red-brown areas are Canada Thistle, Giant Ragweed, and Johnson Grass weeds. Image (b) is an on-the-go yield monitor results from the same field.

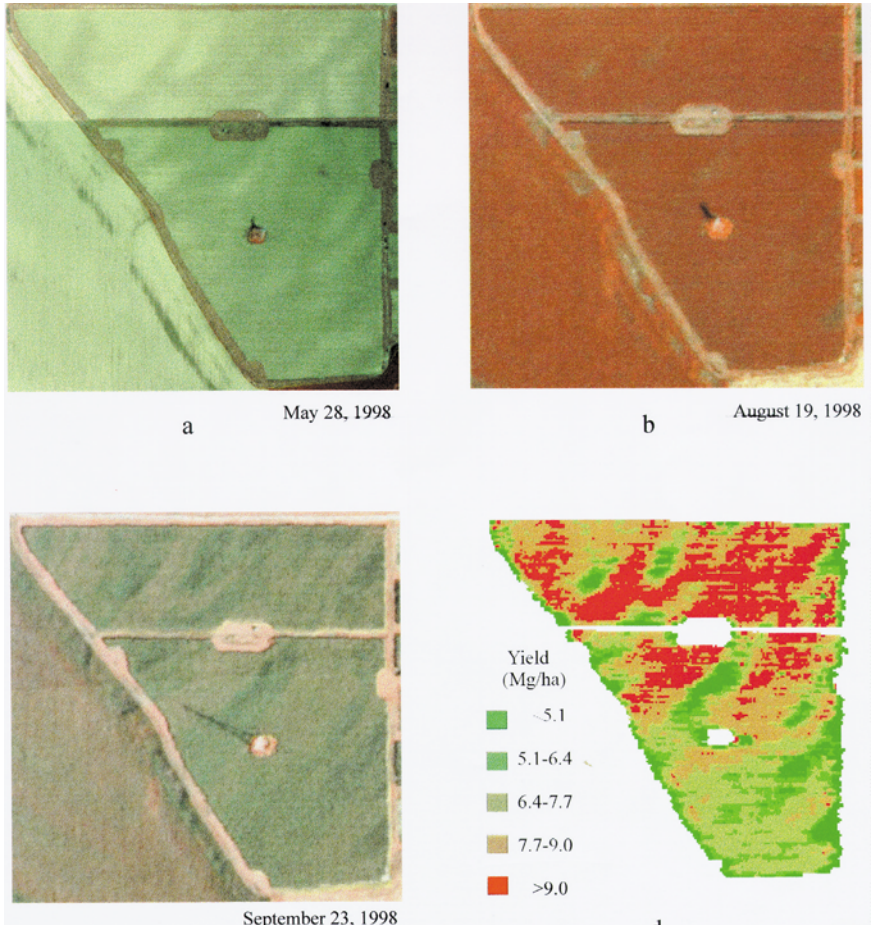


Figure 19.5. Digital infrared images and a yield map for a cornfield at the MSEA site. Note that green vegetation is red or magenta in color on the original and bare soils or areas with poor canopy development are light-toned and blue-green in color.

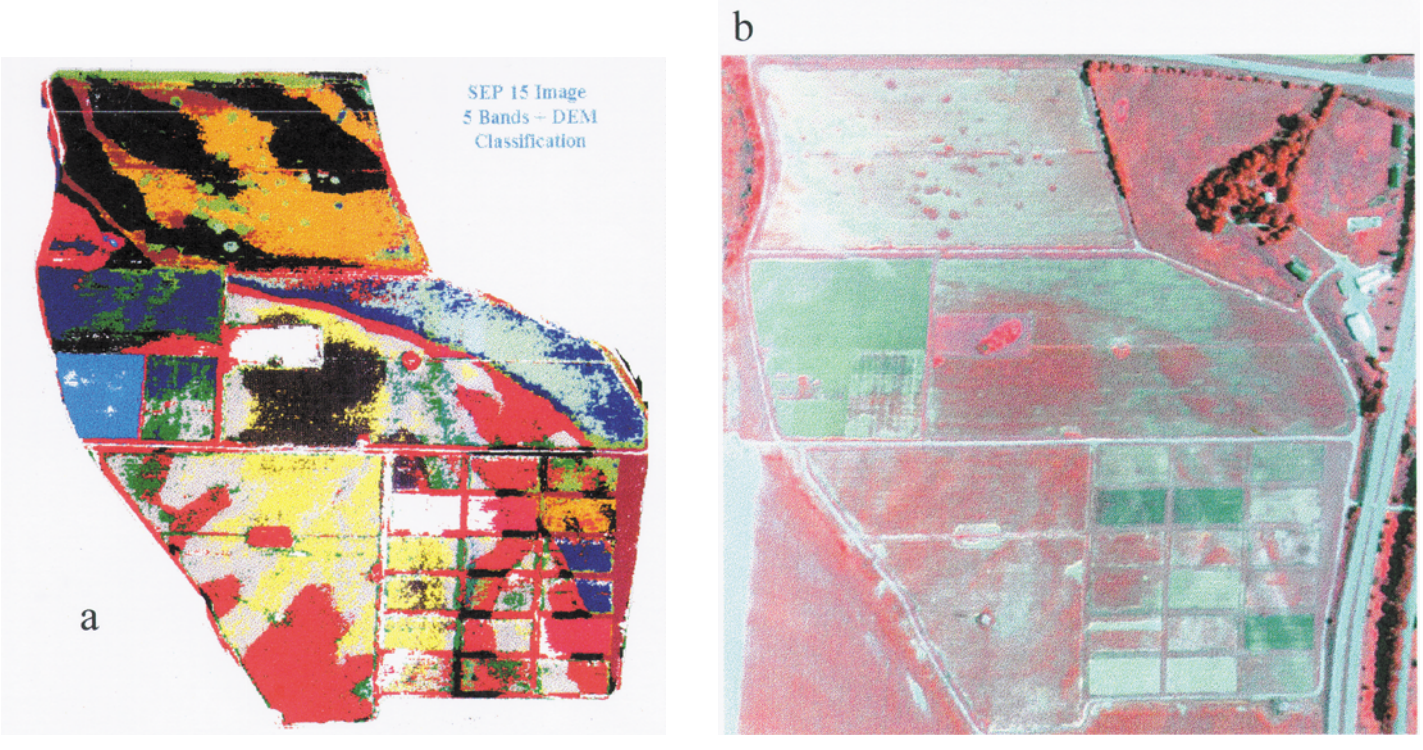


Figure 19.7a & b. Figure 19.7(a). Computer categorized images of crop, soils, and hydrological conditions. These images have more than 30 classes of types based on the light reflectance in the visible, near infrared and middle infrared and use of the Digital Elevation Model (DEM). Figure 19.7(b). False color composite of the MSEA site during senescence. Note dark spots in the top left field that are perennial weeds that stand out against the light-toned senescent soybeans.

Index

Accuracy *See* Validation data
Acrisol 87, 91
Active Microwave Instrument Synthetic Aperture (AMI-SAR) 195
Advanced Very High Resolution Radiometer (AVHRR) 192
Aerial photography 122, 124, 162, 219–240, fig. 11.5, 16.3a (color section)
  color infrared 193
  EDC 192
  land cover classification 184
  in hydrologic models 191
  remotely sensed data 244, 250
  signatures 222–223
  St. Marys River 145, 147–148, 151, 155–156
  See also Rio Guanajibo study
AEZ 77, 78, 96, 97
  in Africa 77–99
  agroecological zones 84, 87
Africa, West and Central 77–97, fig. 8.6 (color section)
Agriculture
  inland valleys 96
  nutrient loading 129
  technologies 241, 250–251
Agroecological zones 84, 87
Agroecosystems 77–96, 97, fig 8.6 (color section)
Airborne multispectral scanner data 244–245
Airborne Terrestrial Applications Scanner (ATLAS) 195
Albemarle-Pamlico Estuarine Study 109–117
Algorithm, 17
  land surface parameters 207–208, 210–212
  overland area parameters 214–215
  remote sensing 194, 195–196
AMI-CAR 195
AMI-SAR 195
AML 39, 42, 44, 51, 54, 68
Animation, geographic 67
APES 109–117
Ara Experimental Basin 33, 34–35, 36
Arc Macro Language *See* AML
ARC/INFO 39, 40, 42, 45, 50, 54, 182, 198
  GRID program 63, 64, 66
  RDBMS 132, 135–136, 137, 138
  stream channel visualization 67
  vector maps 56
ARC/INFO GRID 40, 165, 182, 183
Arc-node topology 56
ArcView interface platform 67, 132, 137, 139
AREA 104
AREA4 104
AREA7 104
AREA10 104
Areal cover 242
Arenosol 87
Aridland 53
ASLP100 104
ASLP200 104
ASP100 104
ASP200 104
ASPECT 104
Aspect parameter 101
ATLAS 195
Atmospheric/hydrologic modeling systems 189
Attribute data 72–73, 137, 138–139
Attribute tables 19
Automated segmentation 17
Avenue scripting langauge139
Average slope 56
AVHRR 192
AVIRIS sensor 194–195, 245

BA 40, 42
Bas-fond *See* Inland valley
Basin azimuth 40, 42
Basin characteristics 53–54
Basin Geomorphic Information Systems (BGIS) 25–36, 36–37
Basin length (BL) 40, 42
Basin perimeter (BP) 40, 42
Basin relief (BR) 40, 42
Basin slope, average (BS) 40, 42, 48, 51
Basin stream order (BSO) 41, 42, 48
Basin width, effective (BW) 40, 42
Basin-area quantifications 40
Basin-aspect quantifications 40
Basin-length quantifications 40
Basinsoft 29, 39–51, fig. 4.3, 4.5 (color section)
  development of 39–41
  drainage basin characteristics 39, 40–41
  graphics 51
  output formats 46
  program structure 42, 43, 50–51
  requirements of 43–45
  verification 46–50

Beach wetland class 147, 152–153, 154, 155
Bedrock 229, 234–235
BGIS 25–36, 36–37
Bills Creek (OK) watershed 18–23, fig. 2.1(color section)
Biological Species class 135
Biological Species module 133, 134–135
Biomass 242, 243, 246, fig. 8.21 (color section)
BL 40, 42
Blue-green algae 129–130
Blue-line network 15, 18, 19, 20, 122
Bottom *See* Inland valley bottom
Boundaries, drainage area 25
BP 40, 42
BR 40, 42
Breaching length (BL) 209
BS 40, 42, 48, 51
BSO 41, 42, 48
Buffer *See* Stream buffer
Buffer zones, of stream channels 124, 125
Burn in technique 64
BW 40, 42

Cachí Reservoir 179
Calibrated Airborne Multispectral Scanner (CAMS) 195
CAMS 195, 197–198, fig. 16.4, 16.5 (color section)
Caribbean National Forest 101, 102
Cascade morphology reach 160
CASI 194–195
Catchment area 7, 8, 11, 12
CCM 41, 42
CDA 40, 42, 44, 48
CGIA 109
CHAN 104
Channel, permanent 19
Channel, stream *See* Stream channel
Channel area 58
Channel computations 41
Channel cross-sectional area, 56–57, 58
  analysis of 58–59
  flow length 58
  response to watershed characteristics 58
  survey 53, 54–55
Channel depth
  measurement of 55, 57
  watershed characteristics 58
Channel geometry *See* Channel variables
Channel length, main
  minimum 12, 18–19
  tributaries 101
Channel link 214
Channel morphology 160, 171, 174
Channel network 7, 12–13, 18, 19, 20, fig 2.2, 2.5, 2.6 (color section)
  Bills Creek 19
  composition 22–23
  data set 26–27, 28, 30, 32, 36
  definition 17,
  identification 17
  map 55
  polygonal 28, 29
  topology of 207–208, 212–213
Channel ordering 207
Channel pattern 160
Channel reach slope 160
Channel routing 57
Channel shape 53, 54, 56, 57
Channel slope 160–161
Channel systems 53
Channel tracing 8
Channel variables, 55, 56, 57, 58, 59 *See also* Channel cross-sectional area
Channel width, 55, 56, 57, 58
Channel-length quantifications 40
Channel-relief measurement 41
CHANSLP 104, 106
CHATRI 103, 104, 106
Chlorophyll 194, 242
Chloropleth maps 67
Chowtan River Basin 109, 111, 114, 117
CHSLP, defined 104
CI 48
Clean Water Act 228–229
Climate 101, 104
  Basinsoft 50
  drainage area 19
  surface water runoff 39
  water level fluctuations 145
CLUSTR 90
Coho *See* Salmon
Color data 251
Command Programming Language (CPL) 39
COMP 104
Compactness ratio (CR) 40, 42
Comprehensive Environmental Response, Compensation, and Liability Act *See* Superfund
Concave areas 17
Connectivity patterns 17
Constant of channel maintenance (CCM) 41, 42
Conterminous United States Soils (CONUS) database 192
Continuous simulation hydrologic models 190
Contour crenulation 18
Contour interval (CI) 48
Contour-based network 25
Contributing drainage area (CDA) 40, 42, 44, 48
CONUS 192

Coordinates class 135
Correlation analysis, watershed parameters 104
Costa Rica 179, 180
Cover complexity 161
Coverage coordinates 137–138
CPL 39
CR 40, 42
CREAMS model 99
Critical source area (CSA) 18–19
  concept 71, 72
  criteria 19
  value 71, 72
Crop canopy 242, 243
Crop characteristics
  remote sensor measurements of 246–247
Crop moisture 247
Crop residue cover 242, 247–248
Crop responses 251
Crop types 79, fig. 8.11 (color section)
Crop yield variability 247
Cultivation patterns 97
  Bobo-Dioulasso study area 94
  Save study area 89
  Sikasso study area 94

D8 method 71
DAAC 192
Daedalus Mutlispectral Scanner (MSS) 244–245, 249
Dambo *See* Inland valley
Data backup procedures 138–139
Data layer, 39, 40, 43, 50
  drainage-divide 43, 44
  hydrography 43, 44–45
  hypsography 43, 45
  lattice 43, 45, 51
  See also Drainage basin, characteristics
Data management standardization 139
Data model design 133–137, 139
Data point set 134
Data server system 132–133, 139
Data sets 26–28, 29–33
Data sources, on-line 192
Database implementation 137–138, 139
Database management 138–139
Database model 133–137
Databases, online 192
DD 101, 103, 105, 106, 183, 184, defined 104
DEDNM 7–9, 14–15, 17–23
  code 9
  compared to topographic maps 18–23
  flow vector 10–11
  low relief terrain, 18
  outputs 8
  sinks 9–10
  subwatersheds 14
  See also TOPAZ
Defense Mapping Agency (DMA) 102
DEM 7–9, 14–15, 17–23, 64, 70–72, fig. 7.2 (color section)
  channel cross-sectional properties 59
  channel network 22–23, 55
  depressions 208–209
  flow vector 10–11
  format and EDC 192
  greater resolutions and watershed characterization in 105
  in low flow studies 102
  low relief terrain 17
  prediction of channel shape 54
  preprocessing 9–10, 11
  rainfall-runoff model 25
  raster 211–212
  regular grid 29, 30
  resolution 71, 171
  square-grid networks 25
  vertical resolution of 17
  of WRIA 164–165
  See also Basinsoft
Depressions 8, 9–10, 11, 17, 19, 22, 70–71, 72, fig 2.2 (color section)
  algorithm 209
  breaching length 209
  DEMs 207, 208–209, 215
  elevation 208–209
Derived Vegetation Index (DVI) 86
DF 41, 42
Differential GPS positioning 241, 250
Digital Elevation Drainage Network Model *See* DEDNM
Digital Elevation Model *See* DEM
Digital Elevation Terrain Data (DTED) 182, 183–184
Digital landscape analysis system 7
Digital Line Graph format *See* DLG
Digital remote sensing 193
Digital satellite data 182
Digital terrain elevation data processing 182
Digital Terrain Model (DTM) 191
Directory structure 137–138
Distributed Active Archive Centers (DAAC) 192
Distributed hydrologic models 190
DLG 25, 26, 39, 45, 48, 102, 192, 197
DMA 102
Downslope flow path 208
Downslope flow routing concept 71
Drainage analysis 7
Drainage area 25
Drainage basin
  characteristics of 39, 40–42

and effects on surface runoff 39
quantification *See* Basinsoft
Drainage density *See* DD
Drainage divide 7, 71, fig. 4.3a (color section)
Drainage features *See* Channel network
Drainage frequency (DF) 41, 42
Drainage network 71, 72
and DEDNM 7, 8, 15
Drainage network density 72, 74
Drainage parameters 19
Drainage-divide data layer 43, 44
DTED 182, 183–184
DTM 191

Earth Observing System Data and Information System (EOSDIS) network 192
Ecological Variable module 133, 134
Ecosystem studies,
of Lake Okeechobee (FL) 129–130
soft points and polygons 135–137
spatial data 135–137
EDC 192
EL 104
ELAS 198
Elevation 9–11
contours 49, 51
data 70–71
data layers 182, 185
quality of data 183–184
spatial distribution of 25
See also Depressions, Flat surfaces
Elevation-contour length (CL) 43, 45, 48, fig 2.4 (color section)
ELONG 104
Elongation ratio (ER) 40, 42
ELRANGE 104, 105, 106
EMAP 236–237
Emergency response 228
Emergent wetland (marsh) wetland class 147, 150–152, 153, 154–155, 156
Environmental Monitoring and Assessment Program (EMAP) 236–237
Environmental Photographic Interpretation Center *See* EPIC
Environmental Protection Agency (EPA) 219 *See also* EPIC
EOSDIS 192
EPA 219 *See also* EPIC
EPIC 219, 222, 223, 224, 228, 229, 237, 239
ER 40, 42
ERDAS 90, 131, 147, 148, 198
ERDAS BSTATS 148, 152–153, 154, 155
ERDAS IMAGINE 183
EROS Data Center (EDC) 192
Erosion problems *See* Nemadji River Watershed, Soil mass wasting
ERS-1 195
ET 119, 126
European Remote-Sensing Satellite (ERS-1, ERS-2) 195
Evaporation measurement 66, 68
Evapotranspiration 196
Evapotransporation (ET) rate 119, 126, 196

Fadama *See* Inland valley
False pits 29, 30–31, 36
FAO classification 78
Farming 241, 242
Farms Service Agency 220
Ferrasol 87
Field management 250
File naming convention 137, 138
Film transparencies 220
First-order streams, number of *See* FOS
Flat area 8, 9–10, 11, 17, 70–71, 72, fig. 2.2 (color section)
and DEMS 207, 209–211, 212
Flat file format 46
Flooding
temporal aspects of 67
and visualization 67–68
See also 1993 Midwest Flood
Floodplain management 68
Flow accumulation grid 64
Flow blockages 71
Flow direction 56, 63, 64
Flow downstream accumulation 63
Flow drainage, and flat surfaces 210–211, 212
Flow energy, and channel shape 58
Flow length 56, 58
Flow paths, 9–10, 11, 17
Flow routing concepts 17
Flow routing management 207–208
Flow simulation algorithms 8
Flow trajectories 25
Flow vector10–11, 12, 13–14, fig. 2.5 (color section)
Flow volume 58
Forced pool-riffle channel 160, 161, 168, 174
Forest cover
age 160
classification problems 112, 117
and soil mass wasting 120, 126–127
types 125
and water yield 120–121
See also Forest seral stage, Trees
Forest seral stage 159, 160, 161, 162, 166, 168
Forested wetland (swamp) wetland class 147, 153, 154–155

FORM 104, 105, 106
FORTRAN 9, 39
FOS 41, 42, 44
Fracture trace analysis 229, 234–235
Frequency analysis 102
Fringe *See* Inland valley fringe

Gagnoa, Côte d'Ivoire 87, 90–93, 96, 97, fig. 8.10, 8.16 (color section)
Gauged watersheds 106
Gauging station zone 62–63, 64, 65, 66
Geographic data management 138
Geographic location 135–137
Geographic Resources Analysis Support System (GRASS) 56
Geographic visualization 67
Geology 101, 102, 104, 229
Geomorphology 101
GIS
  arc-node topology 56
  maps 250, 251
  models 103, 105, 106
  types of *See* ARC/INFO
GIS Data Management Guidelines 138
GIS IDRISI 183, 185
GIS-RDBMS database 129–144
Gleysol 87
Global Positioning System (GPS) data 80, 87, 97, 237, 241, 245
GPS Garmin 100-SRVY 87
Graphic data 138–139
GRASS 56
Great Lakes
  water level fluctuations of 145–147
  wetland class areas156
Greenness vegetation index algorithm 74
GRID 63
Ground-truth data 73, 80, 82, 83, 85, 86–87, 88, 97, 222
Guinea 96, 97

Habitat
  freshwater 159
  preservation 159
  restoration 171
  stream 159, 171
  See also Salmon
Hard point 136
Hazardous waste disposal sites
  analyses 223–228, 230–236, 238–239
  characterization 220–223
  inventories of 228
  mapping 229, 236
  photointerpretation of 219–220, 222, 223
  site features 223, 228

HEC-1 214
Historical aerial photograph 219–220 *See also* Aerial photography
Holdridge classification system, and low flow study 102
HRU 199
HYDICE 194–195
Hydraulic discharge 160
Hydroelectric power, and watershed management 179
Hydrographic data 102, fig. 4.3B, 14.1 (color section)
  layer 43, 44–45, 70, 74
  of stream network 162, 165,166, 167
Hydrologic cycle
  data 196
  MSFC 192
  See also Hydrologic models
Hydrologic model 179, 182, 212
  accuracy of 191
  channel routing 57, 59
  continuous 190
  data requirements of 190–191
  distributed 190
  GRID 63
  macroscale 191, 192, 198, 200
  on-line databases for 192
  overland area 214–215
  physical data required 190
  remotely sensed data 189–201
  single event 190
  See also Rio Guanajibo study
Hydrologic response units (HRU) 199
Hydrologic system 62–63
Hydrology 122
HYP1 104
HYP2 104
HYP3 104
Hyperspectral Digital Imagery Collection Experiment (HYDICE) 195
Hyperspectral instruments 194–195
Hyperspectral sensors 193, 195–196
Hypsography data layer 43, 45, fig. 4.3c (color section)

IITA 77–79, 80, 96
Image analysis *See* Photointerpretation, aerial
Ina Basin 36
INFO table format 46
Infrared data
  crops 246, 247, 251
  fields, 249–250
Infrared imagery 237, 238, fig. 19.4, 19.5 (color section)

Infrared reflectance 242
Inland valley 77–97
  defined 82
  mapping 82– 84, 91
  model 85
  morphology 80
  and upland cultivation, 97
  and watersheds 79, 80
  See also Land cover, Land use
Inland valley bottom 82, 83, 84, 85, 89, fig. 8.7, 8.8, 8.10, 8.11, 8.12, 8.13, 8.19 (color section)
  land use 86, 87, 88, 90, 91, 92, 93, 94, 95, 96
Inland valley fringe 82, 83, 84, 85, fig. 8.9 (color section)
  land use 86, 87, 88, 90, 91, 92, 93, 94, 95, 96
Inland valley upland 93
  land use 86, 87, 88, 90, 91, 92, 94, 95, 96
International Institute of Tropical Agriculture (IITA) 77–79, 80, 96
INTRU 104
IVs *See* Inland valley

Japanese Earth Resources Satellite, Synthetic Aperture Radar (JERS-1 SAR) 195
JERS-1 SAR 195

KINEROS 214

Lake ecosystem study 129–144
Lake Okeechobee (FL) 129–130
Lake Okeechobee Ecosystem Study *See* LOES
Land cover 70, 73–74, 112, 162, 163, 164 185, fig. 11.5,14.1, 16.3a/b, 16.4 (color section)
  categories of 73, 163, 168,
  data 124, 125,
  for hydrologic models 191
  inland valley 84, 86, 87, 90
  types 86, 87
  See also Land use maps, Rio Guanajibo study
Land ownership 162
Land use 97, 162, 169, 174–176
  classes 8, 90, 95, 96–97
  classification 112–113, 113–114,
  correlation with population 114–117
  coverage 101, 102, 103
  data 111
  erosion 119, 126–127
  Gagnoa study area 92
  inland valley 79, 84, 85, 86, 87, 88–93, fig. 8.15, 8.16, 8.17 (color section)
  IVs 77, 79–80
  mapping 96–97, 109
  multispectral data 192
  practices 68
  uplands 79, 84
  See also Rio Guanajibo study, Soil mass wasting
Land Use and Land Cover (LULC) 192
Land Use Data Analysis (LUDA) 112, 113, 115
Landsat satellite imagery 162, 163, 164, 184, fig. 8.6, 8.7, 8.12, 8.14, 8.17, 8.20, 15.3, 16.3b (color section)
  and AEZs 79
  for hydrologic modeling 193, 194
  inland valley 84
  land cover identification 70, 73, 78, 79, 81, 96. 112
  land use 86–87, 109, 111, 113, 115
  and remote sensing of crops and soils 244, 247
Landscape analysis 17 *See also* DEDNM, TOPAZ
Landscape properties 71
Large woody debris (LWD) 160–161, 169, 171–172, 174
Latitude/longitude coordinates 237
Lattice data layer 43, 45, 51, fig. 4.3d (color section)
LDP 237
Leaf area index (LAI) 242
Level I map, 77–78, 79, 89, fig. 8.6 (color section)
Level II map 77, 78–79, 86
Level III map 77
Life Zone Map 101, 102
Light reflectance 242, 243, 244
Line data 122
Lithologic mapping 229
Lithosol 87
Litigation 219–220, 222, 223
Little Washita River Research Watershed (OK) 69–74, fig. 7.2, 7.4 (color section)
Locational coordinates 135–137
Locational Data Policy (LDP) 237
Locational module 133–134
Locational records 131
LOES 129–144
Low flow
  GIS 101–102, 106
  measurements 103
  prediction 101, 102
Low relief terrain 17
Lower Roanoke River Basin 109, 111
LSEU 104
LUDA 112, 113, 115
LULC 192
Lumped parameter hydrologic models 190
Lumped unit hydrograph method 190
Luquillo Experimental Forest 102
Luvisol 84, 88, 89
LWD 159, LWD 160–161, 169, 171–172, 174

M#DNR 104
Main channel length (MCL) 40, 42, 45, 50
Main channel sinuosity ratio (MCSR) 41, 42
Main channel slope proportion (MCSP) 41, 42
Main channel slope (MCS) 41, 42, 48, 50
Map algebra 63, 66
Map scale 19,
  drainage area 19
  hydrologic models 191
Map, raster 19
Mapping units, soil 72–73, 74
Maps
  chloropleth 67
  land use 109
  quadrangle 122, 197
  raster 19
  topographic 7
  UTM 122
  and visualization 67,68
  water balance 62
Marsh wetland class 147, 150–152, 153, 154–155, 156
Marshall Flight Center (MSFC) 192
Mass wasting *See* Soil mass wasting
Matrix, transition 148–149, 152, 153, 154, 155
Mayaguez Bay 196–199
MCL 40, 42, 45, 50
MCS 41, 42, 48, 50
MCSP 41, 42
MCSR 41, 42
Mean slope (SLP) 101, 106
Measurement techniques 7
Metadata 137, 138, 141–144
METAMENU 138, 141–144
Midwest Systems Evaluation Area (MSEA) 244–247, 249, fig. 19.4, 19.5,19.7b (color section)
MMCDNR 104
MMS 179–180
MMYR 104
Modeling Response Units (MRU) 179, 182, 183, 185, 186
Models, database 133–137 *See also* DEM, DEDNM
Modular Modeling System (MMS) 179–180
Modules 42
  of database model 133–135
  TIN-DEMs generating system, 29–32
Montgomery and Buffington channel classification system 160
MRU 179, 182, 183, 185, 186
MSEA 244–247, 249, fig. 19.4, 19.5,19.7b (color section)
MSFC 192
MSLP100 104
MSLP200 104
MSS sensor 195, 249, 250
Multiple regression analysis *See* Regression analysis
Multispectral data 191, 241, 247, 251
  and land-use determination 192
  MSS 195, 249, 250
Multispectral Scanner (MSS) 195, 249, 250
Multispectral sensors 193

Naming convention 137, 138
NASA *See* EOSDIS
National Imagery and Mapping Agency (NIMA) 182
National Priorities List (NPL) 229
National Resources Conservation Service (NRS) 192
National Wetlands Inventory 148, 156
Navarro watershed 179–186, fig. 15.3 (color section), map 181
NCDA 40, 42, 44, 46
NDVI 195, 246, 247, 249, fig. 8.8, 8.21 (color section)
Nemadji River Watershed 119–127
  basin characteristics 123
  soil mass wasting 119, 126
Network
  blue line 15, 18, 19, 20, 122
  connectivity 17
  contour-based 25
  and DEDNM 19
  parameters 18
  segments 17
  square-grid 25
  triangulated irregular 25
Neuse River Basin 109, 111, 114, 116
NIMA 182
1993 Midwest Flood 61–68
Nitosol 87
Node indexing 207–208, 213
Noncontributing drainage area (NCDA) 40, 42, 44, 46
Normalized Difference Vegetation Index (NDVI) 195, 246, 247, 249
North Carolina Center for Geographic Information System (CGIA) 109
NPL 229
NRCS 192
Nutrient loading, excessive 129–130

Object-Oriented Modeling Technique (OMT) 133
OMT 133
On-line databases 192
ORACLE database 132, 133, 138, 139

Organizational module 133, 134–135
Overland area, parameterization 214–215, 216

PA90 103, 104, 106
Pacific Northwest (PNW) 159, 160, 171
Parallax, stereoscopic 220, 221
Parameter
  hydrologic surface runoff 7, 8
  subcatchment 7
Pasquotank River Basin 109, 111, 117
PERC, defined 104
PERM, defined 104
Permeability 73
PET 66
Photo analysis *See* Photointerpretation
Photogeology 229
Photogrammetric mapping 229, 237
Photography, aerial *See* Aerial photography
Photointerpretation, aerial 219–220, 222, 223
Pits, false 29, 30–31, 36
Plane-bed channel, 160, 161
Plant population density 242
Plants, marsh 151–152
PLASTI 104
Point data 122
Point-attribute data 53
Polygon 25, 122, 135–137
Pool formation 160
Pool-riffle channel 160
Pool-riffle stream bed morphology 159, 160, 168, 176
Population data 109–110, 114–117
Potential evapotranspiration (PET) 66
Precipitation 65–66, 68
  data layers 182, 185
  land surface 66–67
Preciptation Runoff Modeling System (PRMS) 179, 185
Principle Responsible Parties (PRPs) 219
PRMS 179, 185
Probablilty 135–137
PRODUS 179
Programs, information systems *See* Basinsoft, FORTRAN
PRPs 219
Puerto Rico 101, 102, 196–199, fig. 16.3a/b, 16.4 color section)

QA, defined 104
Quad maps 122, 197
Quadrangle maps 122, 197
Quantifications, Basinsoft 40

Radar systems 219
Rainfall-runoff model 25
RASPRO 14
Raster data 7, 8, 9, 15, 17, 19, 63, 64, 164, 214
Raster processing program (RASPRO) 14
RB 40, 42
RDBMS 129–144
Reach average slope 160
Reach morphology 160
RECODE 148
Red Clay Area 119
Red Clay Project 119, 122, 126
Reflectance 193–194, 196, fig. 19.7a (color section)
Region class 135
Regional characterization study (level II) 78–79, 79–80
Regression, linear 110, 113, 115, 117, 124, 125
Regression analysis
  low flow study 102,
  prediction of low flows 101, 102, 103
  watershed parameters 104
  wetland classes 149–155
Relational database management system (RDBMS) 129–144
Relative relief (RR) 40, 42
Relative stream density (RSD) 41, 42
RELIEF 104
Remote sensing 193–196, 219, 241
  agroecosystems of inland valleys 77–96
  biomass 246
  crop moisture conditions 244
  defined 193
  field conditions 250
  in hydrologic models 193–196
  instruments 194–195
  methodology 244–246
  ratio indices 246–247
  related to field data collection 246
  spatial databases 189
  water 249
  See also Rio Guanajibo study
RER 104
Resolution 191
Response reach, fig. 14.1 (color section)
  analysis 167–169, 172–176
  stream 160
Restoration strategy, habitat 159, 171–176
Reventazón River basin 179, map 181
Rio Guanajibo (PR) study 196–199
Rio Rosario 196–200, fig. 16.5 (color section)
Riparian habitat 159, 172, 174
River bottom wetland class 147, 148
River reach files 162
RN 41, 42
Roads 162, 168, 169
  inland valley 79, 84, 86, 88

soil mass wasting 120, 124, 125, 126
Rotundity of basin (RB) 40, 42
RR 40, 42
RSD 41, 42
Ruggedness number (RN) 41, 42
Runoff conditions 18
Runoff modeling 19
Runoff simulation 201, 202

Saddle topography
Salmon habitat 159, 160–161, 168–169, 172–174, 176, 210–211
Salmonid habitat utilization 171–172
Sampling 165, 167–168, 71
SAR systems 194, 195
Satellite imagery *See* DAAC, Landsat, SPOT HRV
Save, Republic of Benin 87–90, 96, 97, fig. 8.7, 8.15 (color section)
Scanners 219
Scrub-shrub wetland (scrub-shrub) wetland class 147, 153, 154–155
Scrub-shrub wetland class 147, 153, 154–155
SD 41, 42, 104
SE 105, 106
Section 404 (of Clean Water Act) 228–229
Sediment, fig. 11.3 (color section)
deposition via Nemadji River 119
land cover 125
load 160
slumping 119
supply 160
turbidity 122
Segmentation 17 *See also* Watershed segmentation
Semidistributed hydrologic models 190
Sensors 193
Seral-stage forest 159, 160, 161, 162, 165, 166, 168–171, 174
SERS-2 1895
SF 40, 42
SFWMD 129, 130, 131
Shape factor (SF) 40, 42
Shear stress 119
Shreve's classification of magnitude 18, 20, 23
Side-attribute index 26–27, 30
Side-component index 26–27, 30
Signatures 222–223, 238
Sikasso, Mali/Bobo-Dioulasso, Burkino Faso 87, 93–96, 97, fig. 8.11, 8.17 (color section)
Simple Normalized Difference (SND) 249
Simple Vegetation Index (SVI) 249
Single event hydrologic models 190
Sinks 9–10, 165
SINUO 104
Slope 73
analysis 18
properties 7
and soil mass wasting 120, 125, 126, 127
and turbidity 124
watershed 101, 106
See also Stream channel slope, Terrain slope
Slope ratio (SR) 41, 42
SLORAT 104
SLP 103, 104, 106
Slump sites 124, 125–126
Slumping *See* Soil mass wasting
SND 249
Soft geographic features, modeling 139
Soft points 135–137
Soft polygons 135–137
Soil 101, 102, fig. 7.4 (color section)
attribute data 72–73, 74
characteristics 19
class 135
depth 73
digital data 192
estimates of fertility 249
moisture of 119, 122, 126, 194, 248–249
movement of 119
reflectance 194, 243
root systems 119
stress 126
surface water runoff 39
texture 73
transportability rate potential 119
zones 82, 84, 87, 96, 97, fig. 8.6 (color section)
See also Rio Guanajibo study
Soil grouping 78, 88, 84, 87, 91
Soil mapping units 72–73, 74
Soil mass wasting 119–122, 124, 125, 126–127, fig. 11.6 (color section)
Soil source channels 19
Soil surveys, county 192
Soil Survey Geographic Data Base (SSURGO) 192
Source data layer *See* Data layer
Source data sets 28, 36
Source reach 160
South Florida Water Management District (SFWMD) 129, 130, 131
Spatial data
analysis 165–166
preprocessing 164–165
Spatial databases 189, 190, 200
Spawning gravels 161
Spectral responses of crops 243
SPOT HRV satellite 9, 79, 80, 81, 84, 86–87, 90, 96, 195, fig. 8.6, 8.10, 8.13, 8.16 (color section)
Square-grid network 25
SR 41, 42
SSURGO 192
St. Marys River (MI) 145–156

description of 147
map 146
Standard error (SE) 105, 106
State Soils Geographic Database (STATSGO) 192
Station class 134–135
STATSGO data files 72–73, 74, 192
Step-pool reach 160
Stereoscopic parallax 220, 221
Stereoscopy 220–222
Strahler order 8, 12–13, 18, 19, 20, 22, 23, 41, 53, 56, 207, 212–213
STRAT, defined 104
Stream, low flow 101
Stream "burn in" technique 64
Stream bed morphology 168
Stream buffer 163, 164, 166
Stream channel 159
buffer zones 124, 125
classification systems 159–160
channel coverage 55
gradient 120
network 66
sampling 167–169
slope 162, 165, 166
stability 126
visualization 67
Stream computations 41
Stream gauge data 102–103
Stream gauge grid 64–65
Stream gauge stations 102–103
Stream gradient 122, 124
Stream length 101, 120, 124, 167, 176
Stream network 64, 67, 101, 102, 104, 105, fig 7.2 (color section)
Stream order 41
channel shape variables 56, 57,
as predictive tool 57
statistical analysis of 56
vectors 56
Stream reach 166
Streambank degradation 121, 126
Streamflow 68, fig. 11.3 (color section)
effects on water yield 121, 125
gauging stations 182, 196, 198
values of 64–65
Stream-length quantifications 40
Subcatchment 72
drainage boundaries 71
parameters 7, 208, 213–214, 215
SUBET 148
Subwatershed 13–14, 15, 19, 23, 56, 121, 124–125, figures. 2.6, 11.5 (color section)
definition 17
and measurement of channel shape characteristics 53
Succession, trends in wetland classes 153, 154, 155
Superfund program 219, 220, 236
Surface water runoff 39
model 7, 14–15
simulation of 25
SVI 249
Swamp wetland class 147, 153, 154–155
Synthetic Aperture Radar (SAR) systems 194
Système Probatoire d'Observation de la Terre (SPOT) 195

Tabular output of DEDNM 8
Tar-Pamlico River Basin 109,111
TDA 40, 42
10YRCDNR 104
Terrain aspect 71
Terrain slope 19, 71
Thematic Mapper ™ 195
Thermal imagery 196
Thermal sensors 219, 237
3-D 219, 220, 221, fig. 2.1(color section)
Time of concentration
defined 124–125
and soil mass wasting 120, 122
Time Series module 134
TIN 25, 45, 191 *See also* TIN-DEM
TIN-DEM
of Ara Experimental Basin 35–36
data structure 25–27, 28, 33, 35–36
generating system 29–33
TM data 88, 195, 198 *See also* Landsat
TOPAZ 7, 10, 14, 15, 17, 23, 70–71, 72, 208, 214, fig. 7.2 (color section)
Topographic map 7, 18, 55
of Ara experimental basin 34
and Basinsoft 51
See also USGS
Topographic model, TIN-based 36
Topographic Parameterization *See* TOPAZ
Topographic properties 7–8
TOPOGRID command 45, 56
TOPSCALE 191
Total drainage area (TDA) 40, 42
Total stream length (TSL) 40, 42
Transition matrix, wetland 148–149, 152, 153, 154, 155
Transport reach 160, 174
Transportability rate, soil 119
Trees
roots 119, 122, 126
species 120, 121, 122, 126, 127
and water level 154
See also Forest cover
Triangle network data set 26–27, 28, 30, 31, 36

Triangular Irregular Network (TIN) *See* TIN, TIN-DEM
TRIB 104
TRIPER 104
TSL 40, 42
Turbidity
  classifications of 122
  measures of 122, 124, 125, 127
Typography 74

U.S. Geological Survey *See* USGS
UNCON 104
Unconsolidated bottom (river bottom) wetland class 147, 148, 152–153, 154, 155
Unconsolidated shore (beach) wetland class 147, 152–153, 154, 155
Uniform coordinate system 176
Uniform map themes 176
Unit hydrograph method 190
Universal Transverse Mercator (UTM) map system *See* UTM map system
Upland 152, 153, 154, 155
Upland, inland valley 84, 86
User interface development 139
USFWS Classification System 147, 148
USGS 19
  land use classifications 112, 113–114
  and Basinsoft 39
  and topographic maps 18–19, 20, 55
USGS 7.5′ DEM 17, 18 19
UTM map system 87, 102, 122, 164–165, 198

Validation data 162, 165, 166, 170, 171, 173
Valley, inland *See* Inland valley
Valley segments, discontinuous 31–32
Vector data 55, 56, 122, 164
Vegetation cover 19
Vertex data set 26–27, 28, 30, 31, 36
Vertices *See* Vertex data set
Virginia (southeastern) 109, 111
Virtual data server 137, 139
Visualization 61
  as component of GIS-based hydrologic modeling 67
  of floods 67–68
  processes of 67
  streamflow 67
  of water balance 68

Walnut Gulch Experimental Watershed (AZ) 53–59
WAMS 148
Washington Department of Natural Resources (WDNR) 161, 162
Waste disposal sites *See* Hazardous waste disposal sites
Water balance 61, 62–63, 68
Water demand 102
Water flow direction 25, 34, 36
Water holding capacity 73
Water level fluctuations
  and marsh plants 151–152
  and wetlands 147, 149–155
Water quality 129–130
Water resources
  in crops 249
  in soils 248–249
  and urban fringe growth 179
Water Resource Inventory Area *See* WRIA
Water storage depth 63, 66, 67, fig 6.4 (color section)
Water yield 120, 121, 126
Watershed
  area 56, 58, 101, 103
  aridland 53
  basin 25, 36
  Bills Creek 18–23
  boundaries 8, 11–12, 71
  channel cross-sectional area 58
  channel network 12–13
  channel shape 54
  DEDNM 8
  delineation 63
  DEMs 105
  inland valleys 79, 80
  lag 190
  Little Washita River Research 69–74
  low flow study 102–103
  modeling 7, 70
  parameters 56, 59, 101
  relief 104
  restoration 159
  segmentation 17, 18–19, 71, 72, 74
  slope 101, fig. 11.4 (color section)
  soil mass wasting 120, 122, 125
  surface topography 191
  See also Climate, DEM, DEDNM, Geology, Soil, Stream network, TOPAZ
Watershed Administrative Unit *See* WAU
WAU 161, 162, 165–166, 168–176, fig. 14.3 (color section)
WDNR 161, 162
Wetland
  change detection study 155
  classes 147
  inland valleys 82
  land use classification 112–113, 117, fig. 8.11 (color section)
  mapping 156, 228–229, 236
  water level fluctuations 145, 156

Wetland Analytical Mapping System (WAMS) 148
White Oak River Basin 109, 110, 117
Wildlife habitat studies, LOES 130, 131
Wooding catchment 214
Woody debris *See* LWD
WRIA 161, 162, 165–169, 170, fig. 14.3 (color section)

Zonalaverage function 66
Zonalperimeter 183