



UNIVERSITATEA ECOLOGICĂ DIN BUCUREȘTI

FACULTATEA DE ECOLOGIE ȘI

PROTECȚIA MEDIULUI



Program de studii universitare de licență: Ecologie și Protecția Mediului

Forma de învățământ: învățământ cu frecvență redusă

Suport de curs pentru disciplina:

STATISTICA ECOLOGICA

Titular de disciplină:

Profesor dr. ing. Daniel SCRĂDEANU

București

2022

CUPRINS

Introducere	3
a. Date privind titularul de disciplină	3
b. Date despre disciplină.....	3
c. Obiectivele disciplinei	3
d. Competențe acumulate după parcurgerea cursului.....	4
e. Resurse și mijloace de lucru	4
f. Structura cursului	5
g. Cerințe preliminare.....	5
h. Durata medie de studiu individual	6
i. Evaluarea	6
Capitolul 1. ANALIZA VARIABILITATII GLOBALE A PARAMETRILOR DE MEDIU ...	7
1.1. Introducere	7
1.2. Conținut	8
1.3. Rezumat	18
1.5. Bibliografie recomandată	19
Capitolul 2. ANALIZA VARIABILITATII SPATIALE A PARAMETRILOR DE MEDIU .	20
2.1. Introducere	20
2.2. Conținut	21
2.3. Rezumat	30
2.4. Test de autoevaluare a cunoștințelor	30
2.5. Bibliografie recomandată	30
Capitolul 3. EVALUAREA DISTRIBUTIEI SPATIALE A PARAMETRILOR DE MEDIU	31
3.1. Introducere	31
3.2. Conținut	32
3.3. Rezumat	36
3.4. Test de autoevaluare a cunoștințelor	36
3.5. Bibliografie recomandată	37
Răspunsuri la testele de autoevaluare a cunoștințelor	38

Introducere

Cursul este destinat, în principal, studenților Facultății de Ecologie și Protecția Mediului din cadrul Universității Ecologice din București – învățământ cu frecvență redusă (IFR), dar poate constitui un util material de studiu și pentru studenții la învățământ cu frecvență (IF).

Cursul este destinat analizei și modelării extraordinarei *variabilități* a caracteristicilor ecologice care se manifestă sub două aspecte: *global*, exprimat prin distribuția valorilor caracteristicii în jurul unei valori centrale (*media* sau *mediana*) și *spațial*, exprimat prin variația *valorii* caracteristicii (v_i) în funcție de *poziția* în spațiul geometric de desfășurare a procesului în care este implicată (poziție exprimată de obicei prin coordonatele x_i, y_i, z_i , i fiind indicii punctului de observație în care este măsurată valoarea v).

a. Date privind titularul de disciplină

Nume și prenume:	SCRĂDEANU Daniel
E-mail:	daniel.scradeanu@ueb.education

b. Date despre disciplină

Anul de studiu:	II
Semestrul:	3

c. Obiectivele disciplinei

Obiectivul general al disciplinei	analiza și modelarea <i>variabilități</i> a caracteristicilor ecologice
Obiectivele specifice	<i>Analiza variabilității</i> sub două aspecte: • Variabilitatea globală, exprimat prin distribuția valorilor caracteristicii în jurul unei valori centrale (<i>media</i> sau <i>mediana</i>); • Variabilitatea spațială, exprimat prin variația <i>valorii</i> caracteristicii (v_i) în funcție de <i>poziția</i> în spațiul geometric de desfășurare a procesului în care este implicată

d. Competențe acumulate după parcurgerea cursului

Competențe profesionale	Utilizarea aplicațiilor specifice pentru prelucrarea, reprezentarea și stocarea datelor de mediu • Utilizarea tehnicii de calcul în achiziția, prelucrarea, reprezentarea și stocarea datelor experimentale în concordanță cu cerințele din domeniul Mediu. • Compararea critică a datelor achiziționate, analizate și prelucrate cu estimările teoretice sau cu date furnizate de literatura de specialitate. • Implementarea de software specific în cadrul aplicațiilor practice ca instrument auxiliar în elaborarea proiectelor și rapoartelor profesionale
Competențe transversale	• Documentarea în limba română și cel puțin într-o limbă străină (engleza/franceza), pentru dezvoltarea profesională și personală, prin formare continuă și adaptarea eficientă la noile descoperiri științifice.

e. Resurse și mijloace de lucru

Pentru o pregătire temeinică, vă sugerăm să efectuați toate testele de evaluare, astfel încât rezultatul pregătirii dumneavoastră să fie cât mai obiectiv. De asemenea, vă sugerăm să vă alcătuiți propriile dumneavoastră planuri și scheme, acestea ajutându-vă la o mai bună sistematizare a cunoștințelor dobândite.

Fiecare curs debutează cu prezentarea obiectivelor pe care trebuie să le atingeți – din punctul de vedere al nivelului de cunoștințe – prin studierea temei respective. Vă recomandăm să le citiți cu atenție și apoi, la sfârșitul cursului, să le revedeți pentru a verifica dacă le-ați atins în întregime.

Testele de evaluare prezente la sfârșitul fiecărui curs vă vor ajuta să verificați modalitatea specifică de învățare și să vă îmbunătățiți cunoștințele. Timpul recomandat rezolvării testelor este de 30 de minute. Rezolvați testele numai după studierea în întregime a unui curs și nu vă uitați la răspunsuri decât după rezolvarea testului și numai dacă nu reușiți să găsiți răspunsul prin recitirea cursului.

La fiecare curs există indicată o bibliografie selectivă pe care vă sfătuim să o parcurgeți.

f. Structura cursului

Cursul de STATISTICA ECOLOGICA este alcătuit din 3 capitole (fiecare capitol fiind aferent unei **unități de învățare**):

Capitolul_1. ANALIZA VARIABILITATII GLOBALE A PARAMETRILOR DE MEDIU: univariata si multivariata

Capitolul_2. ANALIZA VARIABILITATII SPATIALE A PARAMETRILOR DE MEDIU: modele deterministe si modele stocastice

Capitolul_3. EVALUAREA DISTRIBUTIEI SPATIALE A PARAMETRILOR DE MEDIU: modele analitice, modele topo-probabiliste

La sfârșitul fiecărui capitol beneficiați de modele de teste de autoevaluare care vă vor ajuta să stabiliți singuri ritmul de învățare și necesitățile proprii de repetare a unor teme. La sfârșitul suportului de curs sunt prezentate răspunsurile corecte la testele de autoevaluare a cunoștințelor.

Cursul de **STATISTICA ECOLOGICA** poate fi studiat atât în întregime, potrivit ordinii prestabilite a capitolelor, dar se poate și fragmenta în funcție de interesul propriu mai accentuat pentru anumite teme. Însă, în vederea susținerii examenului este obligatorie parcurgerea tuturor celor TREI capitole și efectuarea testelor prezentate.

g. Cerințe preliminare

Pentru o mai bună înțelegere a materiei, este necesară coroborarea cunoștințelor dobândite în cadrul acestei discipline cu cele acumulate la disciplinele

generale din anul I (BILOGIE VEGETALA SI ANIMALA, FIZICA SI CHIMIA MEDIULUI, GEOGRAFIA SI GEOLOGIA GENERALA, ATMOSFERA SI CALITATEA AERULUI, HIDROLOGIE SI HIDROGEOLOGIE, ECOLOGIE GENERALA etc.), deoarece STATISTICA ECOLOGICA opereaza cu variabile de natura calitativa si cantitativă care descriu cele doua componente ale mediului: BIOTOPUL SI BIOCEENOZA.

h. Durata medie de studiu individual

Durata medie de învățare estimăm a fi de aproximativ 28 de ore, iar pentru examen ar fi necesar un studiu de o săptămână.

Studierea fiecărui capitol necesită un efort estimat la 9 ore, astfel încât să se fixeze temeinic cunoștințele specifice acestui domeniu.

i. Evaluarea

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

Forma de evaluare (E-examen, C-colocviu/test final, LP-lucrări de control)		
Stabilirea notei finale (procentaje)	- evaluarea finală	60%
	- activități aplicative /laborator/lucrări practice/proiect etc.	20%
	- teste pe parcursul semestrului	-
	- teme de control	20%

Capitolul 1.

ANALIZA VARIABILITĂȚII GLOBALE A PARAMETRILOR DE MEDIU



1.1. Introducere

Analiza variabilității *globale* a parametrilor de mediu vizează asigurarea *reprezentativității* evaluărilor și se realizează prin: analiza *modului de distribuție*, a *eterogenității* și a *valorilor extreme* ale selecțiilor de date.

Metodologia evaluării *geostatistice a unei caracteristici* este elaborată pentru distribuția normală a acestora și din acest motiv *neconcordanța* dintre distribuția valorilor prelucrate și cea normală conduce la *supraestimări* sau *subestimări* proporționale cu *gradul de asimetrie* al acestei distribuții. Prin urmare, *modul de distribuție* al valorilor unei caracteristici geologice în jurul *mediei* sau *mediane* de selecție influențează în mod determinant rezultatele prelucrărilor geostatistice.

Analiza *modului de distribuție* al frecvenței valorilor, în cazul pregătirii acestora pentru prelucrări geostatistice, poate conduce la două rezultate:

- **repartiția valorilor este normală** și în consecință prelucrarea lor prin modele geostatistice conduce la rezultate corect interpretabile;
- **repartiția valorilor nu este normală**, în acest caz fiind necesară transformarea lor (*normalizarea*) în scopul eliminării erorilor introduse prin subestimări sau supraestimări.

Analiza *eterogenității* selecției de date disponibile, realizată de obicei prin *analiza dispersională multifactorială* (D.Scrădeanu, 1995), are ca obiectiv separarea selecției de date în funcție de dispersia diferențiată a valorilor, diferențiere determinată de factori fizico-chimici cu *acțiune divergentă*. Rezultatul analizei eterogenității selecției de date poate conduce la două variante de continuare a estimărilor geostatistice :

- **selecția de date este omogenă**, variantă în care estimările geostatistice se realizează asupra întregului set de date utilizându-se un singur model de variabilitate spațială ;
- **selecția de date este eterogenă**, variantă în care selecția de date trebuie separată în grupuri omogene, pentru fiecare din acestea identificându-se modele distincte de variabilitate spațială.

Analiza *valorilor extreme* ale selecțiilor de date asigură estimarea corectă a intervalului de încredere al parametrilor statistici prin corectarea valorilor exagerate ale dispersiei. Includerea *valorilor extreme* în prelucrare modifică semnificativ *dispersia de selecție*, conducând la creșterea artificială a gradului de incertitudine al

evaluărilor statistice și geostatistice. Analiza valorilor extreme poate conduce și ea la două situații distincte:

- **valorile extreme se elimină** deoarece sunt puțin numeroase și din punct de vedere statistic nu sunt reprezentative pentru caracteristica studiată (sunt fie rezultatul unor erori de măsurare fie al unor variații bruște ce nu sunt definitorii pentru variabilitatea spațială a caracteristicii studiate);
- **valorile extreme nu se elimină** deoarece sunt suficient de numeroase pentru a putea forma o selecție de date căreia i se aplică metode specifice de prelucrare (P.Bomboe, 1979).

Durata medie de parcurgere a acestui capitol este de aproximativ 9 ore.



1.2. Conținut

1.2.1. Parametrii statistici descriptivi

- media aritmetica
- dispersia
- abaterea standard
- coeficient de asimetrie
- coeficient de boltire
- erorare de estimare a amediei (interval de incredere al mediei aritmetice)
- link-uri utile:
 - <http://statisticasociala.tripod.com/parametri.htm>
 - <http://onlinestatbook.com/2/estimation/mean.html>

1.2.2. Testarea normalitatii unei selectii de date

- *Testarea analitică* a concordanței repartiției valorilor cu modelul repartiției normale se realizează de cele mai multe ori cu testul χ^2
- Calculul frecvențelor teoretice (np_i), corespunzătoare repartiției și al statisticii χ^2_{exp} se efectuează pe baza relațiilor (D.Scrădeanu, 1995):

$$np_i = \frac{\Delta \cdot n}{s \cdot \sqrt{2 \cdot \pi}} \cdot \text{Exp} \left[-0,5 \cdot \left(\frac{x_{ci} - m}{s} \right)^2 \right] \quad \text{și} \quad \chi^2_{\text{exp}} = \sum_{i=1}^{i=k} \frac{(n_i - np_i)^2}{np_i}$$

- Aplicație: *Analiza unei repartiții normale și omogene*

Ce mod de repartiție au grosimile depozitelor daciene măsurate în 124 de foraje (Tabelul 1.1) distribuite relativ uniform în Bazinul Dacic (Fig.1.1) ?

Rezolvare:

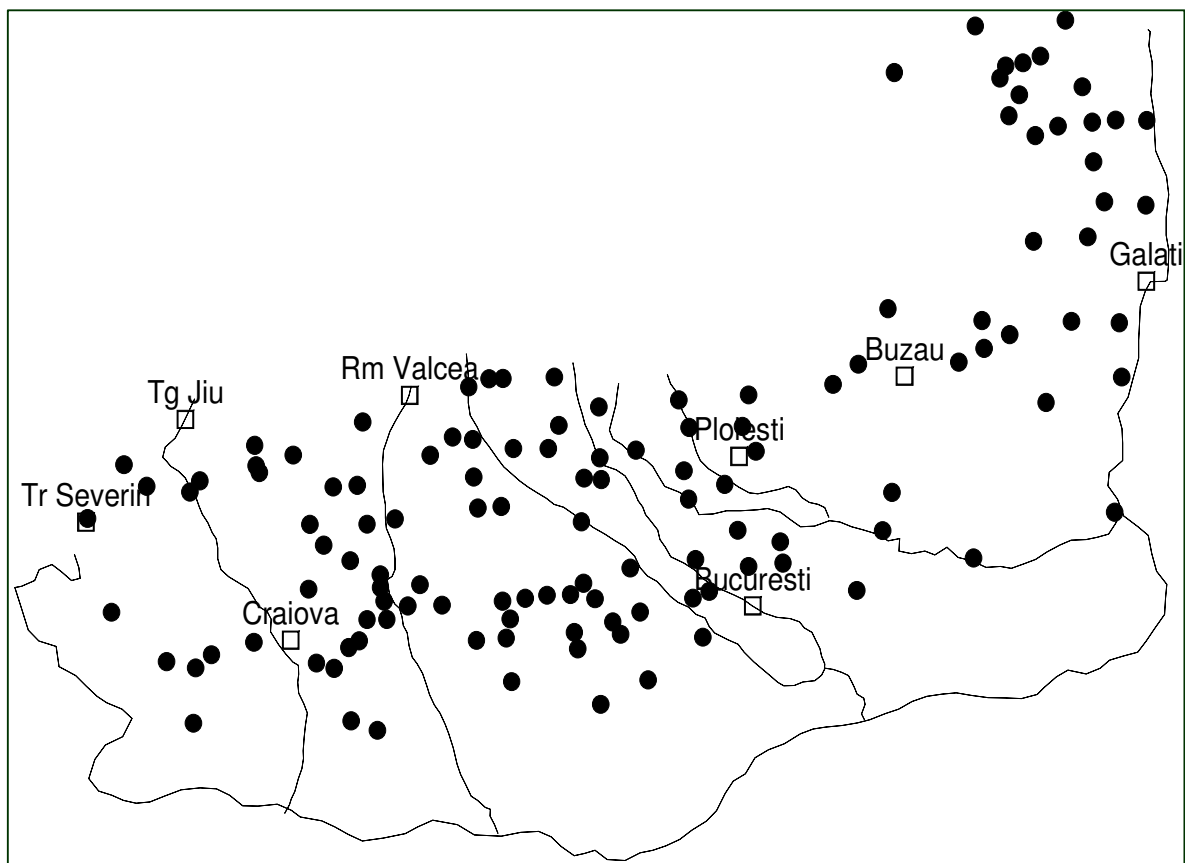


Fig.1.1. Distribuția celor 124 de foraje care au traversat formațiunile daciene

Compararea bazi-nelor de sedimentare, identificarea direcțiilor de transport a sedimentelor, zonarea acestor bazine, utilizează frecvent ca date primare *grosimile formațiunilor* depuse în anumite intervale de timp (Scrădeanu et.al.,1998). Bazinul Dacic, unitatea sedimentară lacustră din estul Paratethisului, a fost intens cercetat prin foraje, în legătură cu zăcămintele de petrol.

Stabilirea modului de repartiție al grosimii formațiunilor daciene se realizează în două etape :

- *analiza grafică* a modelului de distribuție ;
- *testarea analitică* a concordanței cu modelul de repartiție identificat.

Tabelul 1.1. Grosimi ale depozitelor daciene din Bazinul Dacic

Nr. crt.	Grosime g[m]	Nr. crt.	Grosime g[m]	Nr. crt.	Grosime g[m]	Nr. crt.	Grosime g[m]
1	602.47	32	555.46	63	551.54	94	244.11
2	607.20	33	406.81	64	539.97	95	83.55
3	927.18	34	750.22	65	622.09	96	192.70
4	450.37	35	649.77	66	638.18	97	216.03
5	662.22	36	528.71	67	383.54	98	235.44
6	737.94	37	627.29	68	559.45	99	367.14
7	799.48	38	496.16	69	588.81	100	503.28
8	617.00	39	536.19	70	575.85	101	410.55
9	391.45	40	780.16	71	612.05	102	310.32
10	499.71	41	453.90	72	571.67	103	543.78
11	532.44	42	492.62	73	683.04	104	485.56
12	675.76	43	325.98	74	824.11	105	478.52
13	859.18	44	726.91	75	643.87	106	330.93
14	584.43	45	698.84	76	428.93	107	259.91
15	371.32	46	707.51	77	354.17	108	15.27
16	395.34	47	632.66	78	521.33	109	547.64
17	482.04	48	517.69	79	580.10	110	315.68
18	928.00	49	506.85	80	525.01	111	425.30
19	716.81	50	432.54	81	690.71	112	510.45
20	563.46	51	471.50	82	439.71	113	399.20
21	567.54	52	475.00	83	178.82	114	274.13
22	597.83	53	446.83	84	414.27	115	320.89
23	514.06	54	421.65	85	443.27	116	299.09
24	358.55	55	304.80	86	464.47	117	293.21
25	340.51	56	162.71	87	436.13	118	387.51
26	457.43	57	403.02	88	349.70	119	345.15
27	118.71	58	467.99	89	362.88	120	252.24
28	379.52	59	460.95	90	267.21	121	226.13
29	764.08	60	417.98	91	204.98	122	143.42
30	668.83	61	489.08	92	280.74	123	335.78
31	593.28	62	655.88	93	287.09	124	375.44

Analiza grafică a modelului de distribuție se bazează pe *histograma (Fig.1.2.a)* și *diagrama de probabilitate (Fig.1.2.b)* ale grosimilor depozitelor daciene.

Analiza *histogramei* conduce la următoarele observații :

- histograma are *un singur modul* - indicație clară a caracterului *omogen* (din punct de vedere statistic) al selecției; consecința acestei observații este că din punct de vedere statistic toți factorii care au determinat distribuția grosimilor depozitelor dacianului au acționat convergent ;

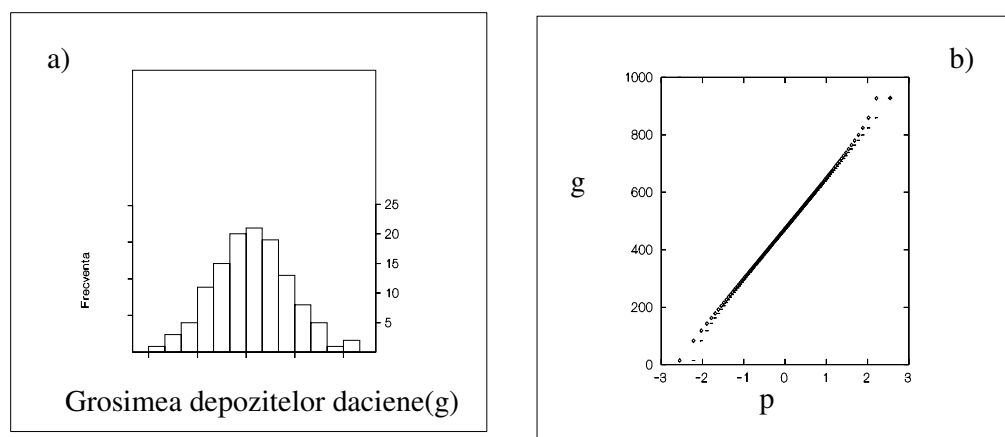


Fig.1.2. Histograma (a) și diagrama de probabilitate (b) ale grosimii depozitelor daciene (Bazinul Dacic)

- caracterul *simetric al histogramei* (coeficientul de asimetrie: $\alpha_1 = 0,01$) sugerează o *distribuție normală* a valorilor; corectitudinea acestei ipoteze urmează să fie testată prin metode analitice .

Analiza *diagramei de probabilitate*, întocmită pentru repartitia normală, permite sesizarea următoarelor caracteristici :

- punctele ce reprezintă valorile grosimilor se *coliniarizează* - indicație clară a *distribuției normale* a acestora ;
- la extremitățile graficului se plasează un număr de 4 *valori extreme* (două la limita inferioară și două la cea superioară) a căror apartenență la selecție trebuie testată (printr-un test analitic; ex.: Chouvenet, Irwin, Romanovski etc.) .

Testarea analitică a concordanței repartii valorilor cu modelul repartii normale se realizează de cele mai multe ori cu testul χ^2 .

Aplicarea testului χ^2 se bazează pe histograma celor $n = 124$ valori de grosimi, histogramă care s-a realizat prin gruparea datelor pe $k = 13$ intervale, fiecare având amplitudinea valorică $\Delta g = 75$ m.

Media aritmetică ponderată a celor 124 de valori este $m = 475,16$ m, abaterea lor standard $s = 176,24$ m iar valorile centrale ale intervalelor de grupare (x_{ci}) și frecvențele absolute experimentale (n_i) sunt grupate în **Tabelul 1.2**.

Datele necesare testului χ^2 sunt parametrii statistici elementari și valorile din primele patru coloane din **Tabelul 1.2**.

Calculul frecvențelor teoretice (np_i), corespunzătoare repartii normale (coloana a cincea din **Tabelul 1.2**), și al statisticii χ^2_{exp} se efectuează pe baza relațiilor (D.Scrădeanu,1995):

$$np_i = \frac{\Delta \cdot n}{s \cdot \sqrt{2 \cdot \pi}} \cdot \text{Exp} \left[-0,5 \cdot \left(\frac{x_{ci} - m}{s} \right)^2 \right] \quad \text{și} \quad \chi^2_{\text{exp}} = \sum_{i=1}^{i=k} \frac{(n_i - np_i)^2}{np_i}$$

Prin eliminarea intervalelor de grupare extreme, cu frecvențe absolute mai mici de doi (P.Bomboe,1979), se obține:

$$\chi^2_{\text{exp}} = 0.46$$

Pentru un risc al erorii de genul I, $\alpha = 1\%$, la un număr al gradelor de libertate $\nu = 7$ (după eliminarea celor trei intervale de grupare extreme) rezultă (din tabelul funcției χ^2):

$$\chi^2(0,99;7) = 1,23$$

Conform criteriului testului de concordanță χ^2 , deoarece :

$$(\chi^2_{\text{exp}} = 0,46) < (\chi^2(0,99;7) = 1,23)$$

rezultă că, din punct de vedere statistic, cu o eroare admisibilă $\alpha = 1\%$, distribuția grosimilor dacianului este conformă cu modelul repartiției normale.

Tabelul 1.2. Calculele asociate testului χ^2

Nr. interval	Lim. inf. interval	Lim. sup. interval	x_{ci}	n_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
1	0	75	15.3	1	0.70	0.13
2	75	150	115.2	3	2.62	0.06
3	150	225	191.0	5	5.74	0.10
4	225	300	265.4	11	10.37	0.04
5	300	375	339.6	15	15.66	0.03
6	375	450	411.9	20	19.74	0.00
7	450	525	485.7	21	21.02	0.00
8	525	600	560.2	19	18.74	0.00
9	600	675	633.8	13	14.04	0.08
10	675	750	704.7	8	9.01	0.11
11	750	825	783.6	5	4.55	0.04
12	825	900	859.2	1	1.96	0.47
13	900	975	927.6	2	0.78	1.91
$\chi^2_{\text{exp}} =$						0.46

COMENTARIU :

Grosimile depozitelor daciene, măsurate în cele 124 de foraje din Bazinul Dacic (Fig.1.1) au o repartiție normală.

1.2.3. Analiza corelațiilor globale

Modelele cantitative statistice exprimă interdependențele dintre componentele ecosistemelor și sunt construite pe baza prelucrării unui mare număr de măsurători experimentale realizate pe parcursul unui program complex de monitorizare.

Elaborarea modelelor statistice se realizează în trei etape principale:

- **Cuantificarea intensității corelațiilor** de diferite tipuri prin intermediul coeficienților de corelație, coeficienți diferențiați în funcție de tipul variabilelor factoriale și al variabilelor independente (x, y, t);
- **Factorizarea corelațiilor** care are ca scop ierarhizarea și selectarea corelațiilor reprezentative din punct de vedere statistic.
- **Modelarea matematică a corelațiilor** de diferite tipuri.

Modelele statistice au un **domeniu de aplicare** restrans la **spațiul** și **intervalul de timp** în care s-a realizat programul de monitorizare pe baza căruia s-au obținut datele necesare elaborării acestora.

1.2.3.1. Coeficienti de corelație

Această categorie de coeficienți este definită pentru cuantificarea intensității legăturii dintre caracteristicile ecologice cantitative dar pot fi adaptați și pentru studiul caracteristicilor calitative.

Caracteristica lor comună este adimensionalitatea și domeniul valoric restrâns ($[-1;1]$ sau $[0;1]$). Valorile extreme indică o intensitate maximă sau minimă a intensității corelației.

Coeficientul corelației lineare

Coeficientul corelației lineare este cel mai des întâlnit în cercetarea ecologică a corelațiilor și din nefericire este utilizat în general fără absolut nici o precauție legată de caracteristicile statistice ale variabilelor implicate.

Definit pentru două variabile cu repartiție normală (x, y), coeficientul corelației lineare (= coef. lui PEARSON = coeficientul corelației totale) este definit cu relația:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^n (x_i - m_x)^2 \sum_{i=1}^n (y_i - m_y)^2}}$$

Valorile coeficientului de corelație lineară sunt cuprinse între -1 și 1 iar dacă x și y sunt independente, $r_{xy} = 0$.

Abaterile de la repartiția normală a variabilelor x și y antrenează modificări ale interpretării valorilor coeficientului de corelație lineară. Valoarea minimă a coeficientului Pearson ($r_{xy} = 0$) nu este un indicator al independenței celor două

caracteristici, ci numai de necorelare liniară a lor. Acestea pot fi corelate printr-o relație funcțională de tip parabolic, logaritmic etc.

Pentru interpretarea valorilor nenule ale coeficienților de corelație, o explicare grafică este mult mai sugestivă pentru cei neacomodați cu statistica matematică. Valoarea coeficientului de corelație liniară este în dependență directă cu distribuția perechilor de valori (x_i, y_i) într-un sistem rectangular de referință XOY . Corespunzător configurației geometrice a distribuției punctelor, se disting următoarele cazuri:

a) alinierea perfectă a punctelor de-a lungul unei drepte - fie ascendentă ($r_{xy} = 1$; Fig. 1.3a), fie descendentă ($r_{xy} = -1$; Fig. 1.3.b) - care indică o dependență liniară perfectă între cele două variabile. O astfel de situație este foarte rar întâlnită în studiul unor relații funcționale între două caracteristici geologice;

b) punctele sunt dispersate aleator, norul de puncte neavând nici o orientare preferențială (Fig. 1.3.c). În circumstanțele amintite anterior, cele două variabile sunt independente sau necorelate ($r_{xy} = 0$);

c) configurația tranzitorie între cele două extreme, în care norul de puncte are o orientare preferențială corespunzătoare valorilor lui r_{xy} aparținând intervalului $[-1, 1]$ (Fig. 1.3.d).

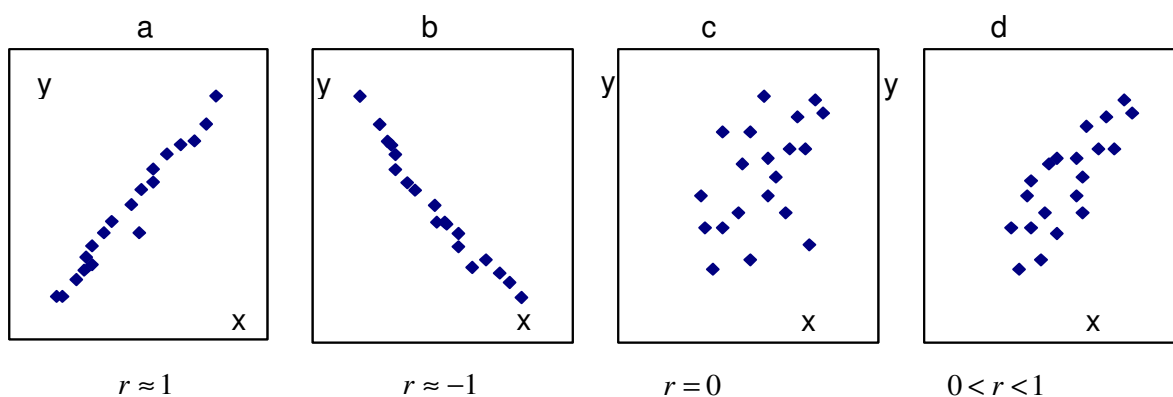


Fig. 1.3. Semnificația geometrică a coeficientului Pearson

O analiză mai detaliată a coeficientului de corelație liniară este reluată la analiza modelului liniar de o singură variabilă independentă .

Valorile coeficientului de corelație lineară, în cazul în care repartiția celor două variabile se abate de la cea normală, nu mai exprimă în mod obligatoriu intensitatea corelației lineare între cele două variabile x și y . În cazul frecvent al repartițiilor lognormale, pentru calculul coeficientului de corelație lineară se operează cu valorile logaritmice ale caracteristicilor analizate.

1.2.3.2. Modelarea corelațiilor globale

Cel mai simplu model pentru corelația între două variabile geologice este cel liniar, în care se presupune că dependența poate fi descrisă prin ecuația unei drepte:

$$y = \alpha_0 + \alpha_1 x + e$$

în care

y - variabila dependentă (= rezultativă);

x - variabila independentă (= factorială);

α_0, α_1 - parametrii modelului;

e - eroarea de estimare a modelului.

Există două modele liniare limită pentru dependența dintre două variabile geologice x și y :

- ambele variabile (x și y) sunt afectate de erori întâmplătoare (**Fig. 1.4**);
- variabila independentă (x) este cunoscută riguros, iar variabila dependentă (y) este afectată de erori distribuite normal (**Fig. 1.5**).

Modelul a) este adecvat studierii corelației conținuturilor de Au și Ag dintr-un zăcămintă sau dintre granulozitate și porozitate într-un acvifer nisipos, iar modelul b) se recomandă pentru studiul corelației între adâncime (x) și conținutul în Au (y) sau între adâncimea (x) și gradul de saturare (y) din zona de aerare a unui acvifer freatic.

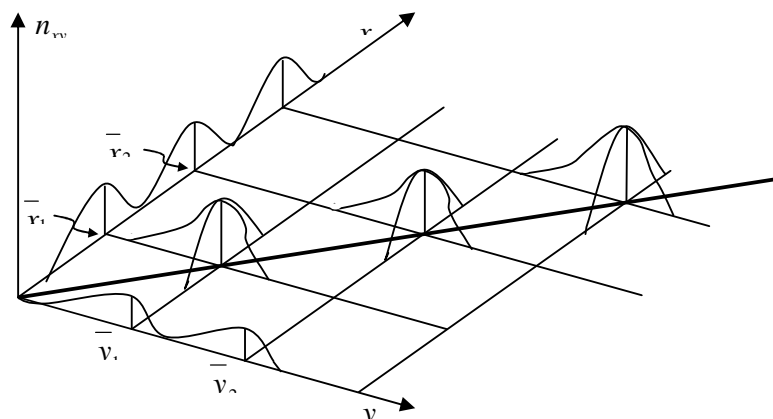


Fig. 1.4 Model liniar cu ambele variabile (x, y) afectate de erori aleatoare

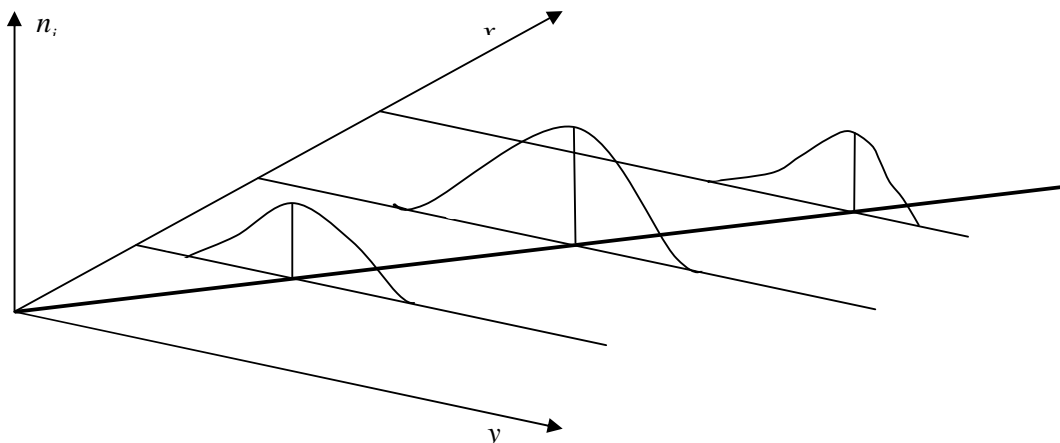


Fig. 1.5. Model liniar cu o singură variabilă (y) afectată de erori

Calculul parametrilor a_0 și a_1 ca estimări de selecție ale parametrilor (α_0 și α_1) se realizează prin metoda celor mai mici pătrate care constă în minimizarea sumei pătratelor abaterii valorilor selecției de la ecuația generală.

Notând suma pătratelor abaterilor de la modelul liniar:

$$SPA = \sum_{i=1}^n [y_i - (a_0 - a_1 x_i)]^2 \quad \text{prin}$$

derivare în raport cu a_0 și a_1 se obține sistemul de ecuații normale

$$\begin{cases} a_0 r + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

Prin rezolvarea sistemului (III.290) se obțin soluțiile:

$$\begin{cases} a_0 = m_y - \frac{s_{xy}}{s_{xx}} m_x \\ a_1 = \frac{s_{xy}}{s_{xx}} \end{cases}$$

în care:

m_x - media valorilor variabilei x : $m_x = \sum_{i=1}^n x_i / n$

m_y - media valorilor variabilei y : $m_y = \sum_{i=1}^n y_i / n$??

$$s_{xy} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

$$s_{xx} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$$

b) Evaluarea incertitudinii

Evaluarea intervalului de încredere pentru parametrii modelului (α, α_1) se bazează pe amplexarea fluctuațiilor variabilei y în jurul modelului determinată de parametrii calculați a_0 și a_1 :

$$\sigma_y^2 \approx s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - m_y)^2$$

Parametrul a_0 , ce estimează parametrul necunoscut α_0 , are o distribuție $N(\alpha_0, \sigma_0)$ în care:

$$\sigma_{\alpha_0}^2 = \sigma_y^2 \sum_{i=1}^n x_i^2 / \left[n \sum_{i=1}^n (x_i - m_x)^2 \right]$$

Variabila:

$$t_{\text{exp}} = (\alpha_0 - a_0) / s_{\alpha_0}$$

are o distribuție t cu $\nu = n - 2$ grade de libertate în care

$$s_{\alpha_0}^2 = s_y^2 \sum_{i=1}^n x_i^2 / \left[n \sum_{i=1}^n (x_i - m_x)^2 \right]$$

Pentru un nivel de semnificație α , intervalul de încredere pentru parametrul α_0 se scrie:

$$a_0 - t \left(1 - \frac{\alpha}{2}; \nu \right) s_{\alpha_0} < \alpha_0 < a_0 + t \left(1 - \frac{\alpha}{2}; \nu \right) s_{\alpha_0}$$

În condițiile aceluiași ipoteze, valoarea α_0 nu se acceptă ca o estimatie a valorii α_0 dacă

$$|t_{\text{exp}}| > t \left(1 - \frac{\alpha}{2}; \nu \right)$$

Parametrul a_1 ce estimează parametrul necunoscut α_1 are o distribuție $N(\alpha_1, \sigma_{\alpha_1})$ în care:

$$\sigma_{\alpha_1} = \sigma_y^2 / \left[\sum_{i=1}^n (x_i - m_x)^2 \right]$$

Variabila

$$t_{\text{exp}} = (a_1 - \alpha_1) / s_{\alpha_1}$$

are deci o distribuție t cu $\nu = n - 2$ grade de libertate, abaterea standard de estimatie calculându-se cu relația:

$$s_{\alpha_1}^2 = s_y^2 / \left[\sum_{i=1}^n (x_i - m_x)^2 \right]$$

Intervalul de încredere pentru parametrul a_1 corespunzător unui nivel de semnificație α este deci:

$$a_1 - t\left(1 - \frac{\alpha}{2}; \nu\right) s_{\alpha_1} < \alpha_1 < a_1 + t\left(1 - \frac{\alpha}{2}; \nu\right) s_{\alpha_1}$$

În mod analog, valoarea a_1 este acceptată ca estimatie a parametrului α_1 numai în cazul în care:

$$|t_{\text{exp}}| < t\left(1 - \frac{\alpha}{2}; \nu\right)$$

APLICATIE : laborator : modelul corelației dintre temperatura și presiunea aerului



1.3. Rezumat

ANALIZA VARIABILITĂȚII GLOBALE (A.V.G.) are ca obiective:

- Identificarea **cele mai probabile valori** a parametrilor de mediu: media aritmetică, în cazul unei repartitii normale a valorilor disponibile;
- Identificarea **intensității corelațiilor** dintre caracteristicile mediului (numerice/alfanumerice)
- **Modelarea corelațiilor** dintre caracteristicile mediului în scopul simulării evoluției acestuia în condiții naturale sau antropice.

1.4. Test de autoevaluare a cunoștințelor



1) De ce este necesară eliminarea valorilor extreme pentru identificarea celei mai probabile valori a caracteristicii studiate?

- a. Reducerea efortului de calcul
- b. Reducerea erorilor de estimare a mediei
- c. Eliminarea erorilor de estimare

2) Ce tip de parametri de mediu sunt corelați prin intermediul coeficientului Pearson?

- a. Calitativi (alfanumerici)
- b. Cantitativi (numerici)
- c. Biofizici

3) Care este efectul distributiei asimetrice a valorilor unei selectii asupra celei mai probabile valori estimate?

- a. Supraestimarea
- b. Subestimarea
- c. Reducerea erorilor de estimare



1.5. Bibliografie recomandată

- a. **Bomboe, P.**, *Geologie matematică (vol. I, Analiza statistică a datelor geologice)*, Editura Universității din București, 1979.
- b. **Brown, S.R.**, *A note on the description of surface roughness using fractal dimension*, Geophys. Res. Lett. 14, 1095-8, 1987.
- c. **Cennini, C.**, *Tratatul de pictură*, Ed.Meridiane, 1977.
- d. **Chauvet, P.**, *Aide memoire de Geostatistique Lineare*, Fascicule 2, Cahiers de Geostatistique, Centre de Geostatistique, Ecole de Mines de Paris, 1991.
- e. **Cheaney, R.F.**, *Statistical methods in geology*, George Allen & Unwin (publishers) Ltd, London, 1983.
- f. **Clarke, G.P.Y. and Dane, J.H.**, *A simplified theory of point kriging and its extension to cokriging and sampling optimization*, Bulletin 609, Alabama Agricultural Experiment Station, Auburn University, Alabama, february 1991.
- g. **Craiu, V., Enache, R., Bâscă, O.**, *Teste de concordanța cu programe în Fortran*, Editura științifică și enciclopedică, București, 1986.

Capitolul 2.

ANALIZA VARIABILITĂȚII SPAȚIALE A PARAMETRILOR DE MEDIU



2.1. Introducere

Analiza variabilității spațiale integrează în prelucrare, pe lângă valorile caracteristicilor geologice (v_i) care au constituit obiectul prelucrării în cadrul *analizei variabilității globale*, o a doua categorie de date: *coordonatele spațiale* (x_i, y_i, z_i) ale punctelor în care au fost determinate valorile acelor caracteristici.

Cele n valori v_i ($i = 1 \dots n$) pe care până acum le-am ținut într-un sac pe care l-am scuturat până ce valorile au fost *normalizate* ($t_i = T(v_i)$) sunt acum împrăștiate pe suprafața de unde au fost colectate pentru a obține elementele necesare realizării desenului ce redă forma obiectelor din hărțile și secțiunile geologice.

Variabilitatea spațială, de mare complexitate pentru caracteristicile geologice, este obiectul unor metode de analiză și sinteză foarte laborioase. Eficiența acestor metode, proporțional cu gradul lor de sinteză, este determinată de experiența celor care le aplică și de parcurgerea într-o succesiune strictă a următoarelor etape:

- *reprezentarea grafică* a datelor, utilizată pentru formarea unei imagini generale asupra distribuției valorilor variabilei în spațiul cercetat;
- *analiza parametrică* a datelor ce sintetizează în trei tipuri de *funcții de distanță* caracteristicile variabilității spațiale atât pentru *o singură variabilă* (*covarianța*, *corelograma* și *variograma*) cât și pentru *o pereche de variabile* (*intercovarianța*, *intercorelograma* și *intervariograma*) probate în aceleași puncte de observație;
- *analiza staționarității* caracteristicilor geologice studiate, primul pas dificil al analizei variabilității spațiale, care pune la încercare experiența cercetătorului privind circumstanțele acceptării unor aproximări și a consecințelor acestor aproximări asupra rezultatelor finale ale prelucrărilor geostatistice;
- *analiza variografică*, ultima și cea mai dificilă etapă a analizei variabilității spațiale în care cuantificarea variabilității spațiale se face sub forma *cele mai probabile legi de variație spațială* a caracteristicii studiate.

Durata medie de parcurgere a acestui capitol este de aproximativ 9 ore.

Link-uri utile:

- http://www.ahgr.ro/specialisti/daniel-scradeanu/3_geostatistica/processing.aspx



2.2. Conținut

2.2.1. Reprezentarea grafică

Proprietăți precum *localizarea valorilor extreme* (minime sau maxime), *tendința* de evoluție regională, *gradul de continuitate*, sunt de mare interes pentru studiul proceselor geologice.

Bazată pe un număr minim de instrumente și prelucrări, *reprezentarea grafică* are ca obiectiv sintetizarea caracteristicilor topologice ale datelor trecute deja prin filtrul analizei *variabilității globale univariate* (*tip de repartiție, valori extreme, dispersie*) și *multivariate* (*analiză discriminant, analiză factorială, analiză corelatorie și spectrală* etc.; D.Scrădeanu, 1995).

Ca și *histograma*, pentru *tipul de repartiție*, sau *dreapta de regresie*, pentru *corelația* dintre două variabile, cele mai eficiente instrumente pentru descrierea variabilității spațiale sunt cele grafice. Principalele caracteristici structurale ale datelor primare se exprimă în mod curent prin: *hărți punctuale, hărți simbolice și indicatoare, diagrame de continuitate și variabilitate*.

Harta punctuală se realizează prin simpla dispunere într-un sistem de coordonate a punctelor de observație lângă care se înscriu sau nu, în funcție de densitatea punctelor de observație, valorile variabilei studiate. Harta punctuală se realizează în prima etapă a studiului caracteristicilor spațiale, ea fiind utilizată pentru:

- identificarea erorilor în amplasarea punctelor de observație;
- calculul densității punctelor de observație;
- localizarea valorilor extreme, determinate fie de erori de măsură, fie de anomalii

locale, care solicită un interes special (ex.: prezența unor *pepite*).

Harta punctuală este utilă pentru situația în care *numărul de puncte de observație este redus*; în caz contrar se apelează la *harta simbolică*.

Harta simbolică presupune o primă filtrare a datelor primare prin reducerea variabilității spațiale și se realizează în cazul unui număr mare de puncte de observație, număr care face inexpresivă *harta punctuală* prin suprapunerea punctelor de observație sau a etichetelor atașate acestora.

Suprafața pe care se realizează harta simbolică este acoperită cu o *rețea rectangulară/pătratică* în celulele căreia se calculează valoarea medie a caracteristicii studiate. *Numărul de simboluri* (alfanumerice, tonuri de gri sau culori) utilizate se stabilește în funcție de gradul de detaliere necesar. Când se utilizează numai două simboluri *harta simbolică* poartă denumirea de *hartă indicatoare*.

Dimensiunile celulelor rețelei rectangulare/pătratice, procedeul de calcul al valorii medii pentru fiecare celulă și numărul de simboluri influențează în mod determinant aspectul hărții simbolice.

Hărțile simbolice oferă o imagine simplificată a distribuției spațiale a valorilor caracteristicii studiate și permite sesizarea tendințelor și localizarea zonelor cu valori maxime și minime.

Diagrama de continuitate este reprezentarea grafică prin care se face o primă evaluare a gradului de precizie cu care se va putea calcula distribuția spațială a unei variabile, probată prin intermediul unei rețele de observație cu o geometrie oarecare. *Buna continuitate* a variabilei (adică o variație lentă a variabilei de la un punct la altul) conduce la o precizie crescută în evaluarea distribuției spațiale. Diagrama de continuitate sintetizează într-o reprezentare rectangulară binară, similaritatea valorilor măsurate în puncte vecine.

Este ușor de intuit că similaritatea a două valori vecine depinde de: *distanța* care separă punctele în care au fost determinate valorile;

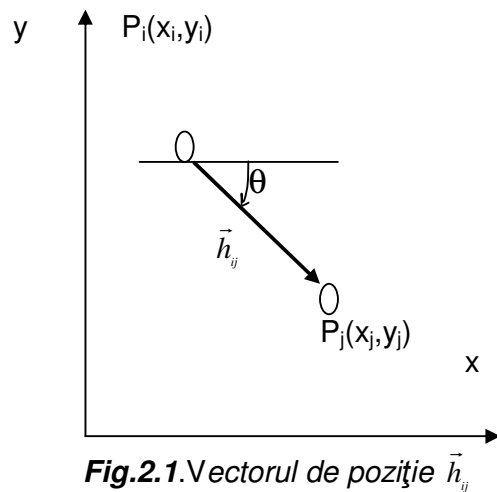


Fig.2.1. Vectorul de poziție $\vec{h}_{ij}$

orientarea spațială a dreptei care unește aceste puncte.

Pentru a construi *diagrama de continuitate* se utilizează un *vector de poziție* care este caracterizat prin:

modul, numeric egal cu lungimea segmentului care unește cele două puncte ($|\vec{h}_{ij}|$ pentru punctele P_i și P_j din **Fig.2.1**);

orientarea acestui segment (θ din **Fig. 2.1**), măsurată prin unghiul dintre direcția axei abscisei și a segmentului care unește cele două puncte.

Vectorul de poziție ales este utilizat pentru a identifica toate perechile de puncte din spațiul cercetat care se află la distanța și orientarea aleasă. Perechile de valori ($v_i(P), v_j(P + \vec{h}_{ij})$) se reprezintă prin puncte într-un sistem de referință rectangular (**Fig.2.2**).

Distribuția punctelor din diagrama de continuitate exprimă gradul de continuitate al variabilei studiate:

- *gruparea* “strânsă” a punctelor în jurul bisectoarei unghiului dintre axele sistemului de referință indică o bună continuitate a variabilei pentru direcția și distanța aleasă; la limită, când distanța pentru care se întocmește diagrama de continuitate este *zero*, toate punctele se află plasate pe bisectoarea unghiului dintre cele două axe ale sistemului de referință deoarece, pentru orice punct, valoarea din acel punct este egală cu ea însăși;
- *dispersarea* punctelor în spațiul dintre cele două axe ale sistemului rectangular de referință indică o slabă continuitate; cu cât norul de puncte este mai difuz, cu atât similaritatea valorilor este mai mică și deci continuitatea mai slabă.

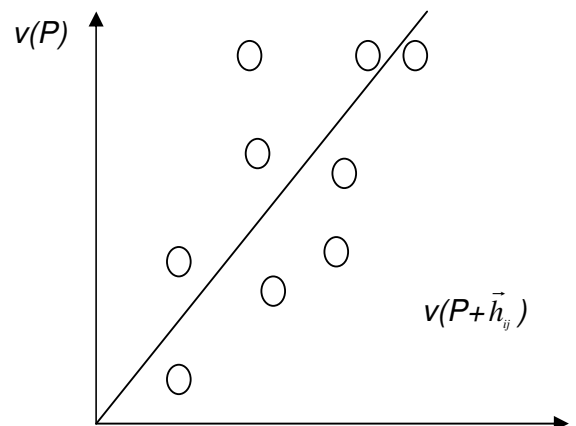


Fig.2.2. Diagrama de continuitate

Diagramele de continuitate sunt afectate în mod semnificativ de valorile *extreme* ale variabilei studiate, de *modulul* și *direcția* vectorului de poziție ($\vec{h}_{ij}$). Pentru etapa evaluării *modelului de structură spațială* (variograma), diagramele de continuitate sunt singurele instrumente care permit identificarea valorilor *nodale* (*valori nodale* - valori care sunt determinante în stabilirea legii de variație spațială a unei caracteristici regionalizate) ale structurilor spațiale. Numai *diagramele de continuitate* permit eliminarea *valorilor extreme*, care nu se încadrează în modelul structural fiind “*accidente structurale*”.

Compararea diagramelor de continuitate pe diferite direcții permite identificarea *anizotropiei structurilor spațiale*. Structurile *anizotrope* sunt caracterizate de diagrame de continuitate diferite pe direcții diferite de calcul. Identitatea diagramelor de continuitate calculate pe orice direcție indică *izotropia* structurii.

Diagrama de variabilitate exprimă corelația dintre *valoarea medie* (m) a unei caracteristici într-o anumită zonă și *eroarea* cu care ea poate fi estimată. *Eroarea* este calculată pe baza *abaterei standard* (s) corespunzătoare. Existența corelației între *valoarea medie* și *abaterea standard* este cunoscută sub denumirea de *efect de proporționalitate*. Efectul de *proporționalitate directă* (**Fig.2.3.a**) indică faptul că:

- în zonele în care au fost determinate *valori mari* ale caracteristicii studiate *variabilitatea este mare* și ca urmare erorile de estimare vor fi mari;
- în zonele în care au fost determinate *valori mici* ale caracteristicii studiate *variabilitatea este mică* și ca urmare erorile de estimare vor fi mici.

Este frecvent și efectul de *proporționalitate inversă* (**Fig.2.3.b**), adică:

- în zonele în care au fost determinate *valori mari* *variabilitatea este mică* și ca urmare erorile de estimare vor fi mici;
- în zonele în care au fost determinate *valori mici* *variabilitatea este mare* și ca urmare erorile de estimare vor fi mari.

Lipsa efectului de proporționalitate nu permite prognozarea mărimii relative a erorilor de estimare. Este situația unei variabilități cu amplitudine mare și cu o distribuție spațială neuniformă care conduce în general la erori mari de estimare a distribuției spațiale a caracteristicii studiate.

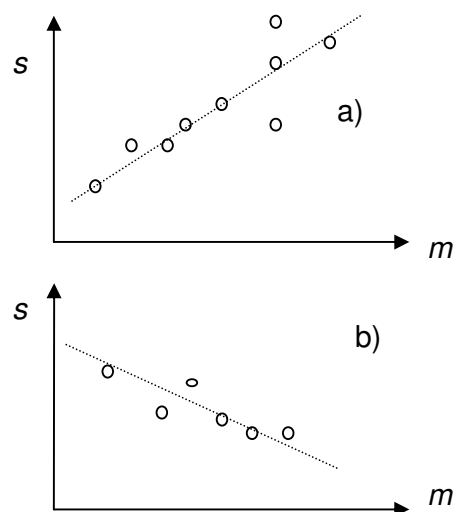


Fig.2.3. Diagrame de variabilitate cu efect de proporționalitate directă(a) și inversă(b).

Pe parcursul realizării reprezentării grafice a datelor începem să “simțim” caracteristicile structurii spațiale ale variabilei geologice studiate prin intermediul *văzului*. Dintre toate simțurile *văzul* ne face să cunoaștem mai bine și descoperă mai multe diferențe. Dacă îi oferim imagini *reprezentative* pentru structura studiată el ne permite să ne formăm o imagine corectă asupra acesteia.

În etapa reprezentării grafice a datelor ne familiarizăm cu particularitățile unei *anumite structuri spațiale*; altfel spus, acumulăm *experiență* prin confruntare cu *individualul* reducând riscul aplicării greșite a instrumentelor geostatistice în etapele ulterioare de prelucrare.

Analiza în paralel a *hărților punctuale*, *hărților simbolice*, a *diagramelor de continuitate* și a *diagramelor de variabilitate* este absolut necesar să preceadă orice încercare de estimare spațială

a unei variabile geologice deoarece elimină riscul interpretărilor greșite determinate de erori grosolane ce pot interveni pe parcursul prelucrărilor:

- la realizarea fișierelor de date (prin greșeli comise la introducerea valorilor);
- funcționarea greșită a programelor de calcul etc.

Reprezentările grafice realizate trebuie permanent consultate și confruntate cu rezultatele intermediare și cele finale ale estimărilor distribuției spațiale.

2.2.2. Analiza parametrică a datelor

În mod uzual instrumentele de prelucrare care se utilizează în această etapă de analiză parametrică sunt trei funcții de distanță: funcția de covarianță ($c(h)$), corelograma ($\rho(h)$) și funcția de variogramă ($\chi(h)$).

Toate cele trei funcții de continuitate sunt puternic influențate de valorile extreme care nu se încadrează în variabilitatea globală a selecției de date cu care se operează. Dacă forma uneia dintre cele trei funcții nu este clar definită, este de mare utilitate examinarea diagramelor de continuitate pentru identificarea acestor valori și eliminarea lor.

Calculate în raport cu o singură variabilă, funcțiile de continuitate permit cuantificarea continuității acesteia în raport cu direcția și distanța.

Funcția de covarianță $c(h)$ reprezintă variația similitudinii valorilor din două puncte în raport cu distanța dintre ele. Ea se calculează cu formula:

$$c(h) = \frac{1}{N(h)} \sum_{(i,j) \mid h_{ij}=h}^{N(h)} v_i v_j - m_{-h} m_{+h}$$

în care:

$N(h)$ este numărul perechilor de puncte separate prin vectorul $|\vec{h}|$;

(i, j) - perechea de puncte (p_i, p_j) separate prin vectorul $|\vec{h}_{ij}|$;

v_i - valoarea variabilei din originea vectorului $|\vec{h}_{ij}|$;

v_j - valoarea variabilei în vârful vectorului $|\vec{h}_{ij}|$;

m_{-h} - media valorilor situate în originea celor $N(h)$ vectori $|\vec{h}|$.

$$m_{-h} = \frac{1}{N(h)} \sum_{h_{ij}=h}^{N(h)} v_i$$

m_{+h} - media valorilor situate în vârful celor $N(h)$ vectori $|\vec{h}|$.

$$m_{+h} = \frac{1}{N(h)} \sum_{h_{ij}=h}^{N(h)} v_j$$

Corelograma ($\rho(h)$) este o funcție de covarianță standardizată și se calculează cu relația:

$$\rho(h) = \frac{c(h)}{\sigma_{-h} \sigma_{+h}}$$

în care:

σ_{-h} este abaterea standard a tuturor valorilor aflate în originea celor $N(h)$ vectori $\vec{h}$:

$$\sigma_{-h}^2 = \frac{1}{N(h)} \sum_{i|_{h_{ij}=h}}^{N(h)} v_i^2 - m_{-h}^2$$

σ_{+h} este abaterea standard a tuturor valorilor aflate în vârful celor $N(h)$ vectori $\vec{h}$.

$$\sigma_{+h}^2 = \frac{1}{N(h)} \sum_{j|_{h_{ij}=h}}^{N(h)} v_j^2 - m_{+h}^2$$

Funcția de variogramă ($\gamma(h)$) reprezintă variația varianței erorii de estimare în raport cu distanța dintre punctul în care se cunoaște valoarea variabilei și cel în care aceasta se estimează. Altfel spus, valoarea variogramei pentru un anumit vector $\vec{h}$ exprimă eroarea care se comite atunci când se atribuie variabilei în punctul $p+h$ valoarea sa din punctul p . Ea se calculează cu relația:

$$\gamma(h) = \frac{1}{2 \cdot N(h)} \sum_{(i,j)|_{h_{ij}=h}}^{N(h)} (v_i - v_j)^2$$

Toate cele trei *funcții univariate* de distanță nu sunt afectate de sensul vectorului $\vec{h}$, fiind *funcții pare*; dacă se schimbă indicele i cu j în toate formulele de calcul ale funcțiilor de distanță valorile $c(h)$, $\rho(h)$ și $\gamma(h)$ nu se schimbă:

$$c(h) = c(-h); \quad \rho(h) = \rho(-h); \quad \gamma(h) = \gamma(-h)$$

Ideea de continuitate poate fi extinsă și la două variabile. Corelația spațială între două variabile u și v depinde de continuitatea fiecăreia. Continuitatea poate fi exprimată grafic printr-o *diagramă de continuitate* în care pe cele două axe se reprezintă $v(p)$ și $u(p + \vec{h})$. Este evident că pentru $|\vec{h}| = 0$ într-o diagramă de continuitate bivariată (u, v) nu toate punctele se află pe bisectoarea unghiului făcut de cele două axe de coordonate, așa cum se întâmplă în cazul diagramelor de continuitate univariate.

Aceleași funcții, utilizate pentru o singură variabilă, se utilizează cu modificările corespunzătoare pentru descrierea continuității bivariate. Pentru evidențierea aspectului bivariat le vom numi: *funcția de intercovarianță*, *funcția de intercorelație* și *funcția de intervareogramă* și vom introduce doi indici corespunzători celor două variabile ($c_{uv}(h)$, $\rho_{uv}(h)$, $\gamma_{uv}(h)$).

Funcția de intercovarianță se calculează cu relația:

$$c_{uv}(h) = \frac{1}{N(h)} \sum_{(i,j)|_{h_{ij}=h}}^{N(h)} u_i v_j - m_{u-h} m_{v+h}$$

în care:

u_i sunt valorile variabilei u ;

v_i - valorile variabilei v ;

$N(h)$ - numărul perechilor de puncte separate prin vectorul h ;
 m_{u-h} - media valorilor variabilei u situate în originea celor $N(h)$ vectori.

$$m_{u-h} = \frac{1}{N(h)} \sum_{i|h_{ij}=h}^{N(h)} u_i$$

m_{u+h} - media valorilor variabilei v situate în vârful celor $N(h)$ vectori:

$$m_{v+h} = \frac{1}{N(h)} \sum_{j|h_{ij}=h}^{N(h)} v_j$$

Funcția de intercorelație este dată de ecuația:

$$\rho_{uv}(h) = \frac{c_{uv}(h)}{\sigma_{u-h} \sigma_{v+h}}$$

în care:

σ_{u-h}^2 - abaterea standard a tuturor valorilor variabilei u , aflate în originea celor $N(h)$ vectori h :

$$\sigma_{u-h}^2 = \frac{1}{N(h)} \sum_{i|h_{ij}=h}^{N(h)} u_i^2 - m_{u-h}^2$$

σ_{u+h}^2 - abaterea standard a tuturor valorilor aflate în vârful celor $N(h)$ vectori $|\vec{h}|$:

$$\sigma_{v+h}^2 = \frac{1}{N(h)} \sum_{j|h_{ij}=h}^{N(h)} v_j^2 - m_{v+h}^2$$

Funcția de intervareogramă ($\gamma_{uv}(h)$) se calculează cu relația:

$$\gamma_{uv}(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{ij}=h}^{N(h)} (u_i - u_j) \cdot (v_i - v_j)$$

2.2.3. Modelarea variogramelor experimentale

Modelarea variogramelor experimentale este necesară deoarece toate operațiunile de estimare de tip topo-probabilist se bazează pe variogramă și acestea nu pot fi efectuate cu o funcție tabelară așa cum este variograma experimentală. Realizarea calculelor de estimare implică identificarea unui model teoretic care să interpoleze cât mai bine valorile variogramei experimentale.

În alegerea *modelelor teoretice*, două caracteristici ale variogramei experimentale sunt luate în considerare: comportarea în vecinătatea originii (parabolică, lineară sau efect de pepită total) și prezența sau absența palierului.

Funcție de aceste caracteristici modelele teoretice utilizate în mod curent pot fi separate în trei categorii:

- *modele cu palier* (modele tranzitive) cu comportament linear în vecinătatea originii (*modelul sferic* și *modelul exponențial*) și cu comportament parabolic în vecinătatea originii (*modelul gaussian*);
- *modele fără palier* (*modelul putere* și *modelul logaritmic*);
- *modelul efect de pepită*.

Modelele teoretice cu care se operează sunt normate, adică au varianța unitară. Pentru a obține un model cu palier diferit de unu este suficient să se multiplice expresia normată a modelului cu o constantă.

Modelul sferic, cel mai frecvent utilizat, are ecuația:

$$\gamma(h) = \begin{cases} \frac{3}{2} \cdot \frac{h}{r} - \frac{1}{2} \cdot \frac{h^3}{r^3}, \forall h \in [0, r] \\ 1, \forall h \in [r, \infty) \end{cases} \quad (1.53)$$

Modelul exponențial are ecuația:

$$\gamma(h) = 1 - \exp\left(-\frac{h}{r}\right) \quad (1.54)$$

De remarcat că *modelul sferic* atinge valoarea maximă (palierul) pentru o distanță finită $h = r$ în timp ce *modelul exponențial* tinde asimptotic la aceasta (**Fig.2.5**). În practică, pentru *modelul exponențial* se adoptă o rază de influență efectivă $r' = 3r$, pentru care $(r') = 0,951$.

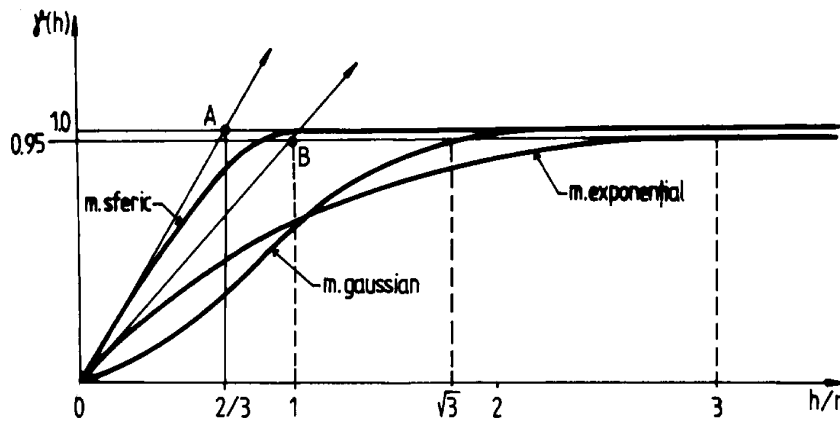


Fig.2.5. Modele cu palier (tranzitive)

Ceea ce diferențiază *modelul sferic* de cel *exponențial* sunt abscisele punctelor de intersecție dintre tangentele la origine și palier; pentru *modelul sferic*, două treimi din raza modelului ($x_A = 2r/3$), iar pentru *modelul exponențial* o treime din raza de influență efectivă ($x_B = r'/3$).

Modelul *gaussian* este valabil pentru un comportament foarte regulat în vecinătatea originii și are ecuația:

$$\gamma(h) = 1 - \exp\left(-\frac{h^2}{r^2}\right)$$

Palierul este atins asimptotic, raza de influență efectivă fiind $r' = r\sqrt{3}$

Comportamentul parabolic din vecinătatea originii al modelului gaussian nu trebuie confundat cu efectul unei derivate deoarece la distanțe mai mari valoarea variogramei se stabilizează în jurul palierului. Pentru analize locale această confuzie este posibilă și ea trebuie evitată pentru corecta alegere a modelului.

Modele fără palier corespund variogramelor experimentale a căror creștere este continuă în limitele domeniului de observație.

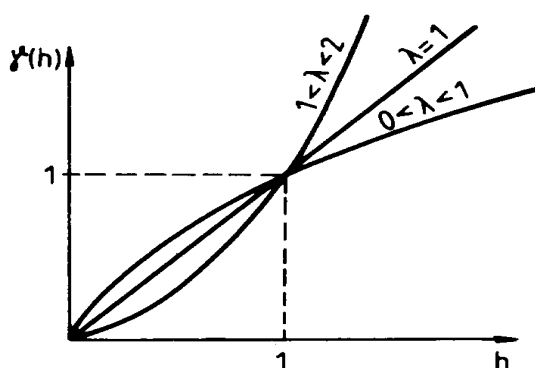


Fig.2.4. Model putere

Modelul *putere* (**Fig.2.4**) este de ecuație:

$$\gamma(h) = h^\lambda, \lambda \in (0, 2)$$

În practică, modelul linear este cel mai utilizat, el putând servi la estimarea în vecinătatea originii a tuturor modelelor cu comportament linear (modelul sferic și modelul exponențial). Pe măsură ce *puterea* crește comportamentul în vecinătatea originii este mai regulat, dar pentru puterea > 2 modelul nu mai este pozitiv definit fiind incompatibil cu evaluările geostatistice lineare.

Modelul *logaritm* (cunoscut și sub numele de modelul De Wijs) are ecuația:

$$\gamma(h) = \log h$$

Modelul logaritm nu poate descrie structurile de suport punctual, dar această condiție nu este deranjantă deoarece în practică datele experimentale care definesc variabila regionalizată sunt relative la un suport nepunctual (nenul), adică la o probă de o dimensiune finită.

Discontinuitatea în vecinătatea originii a variogramei (efectul de pepită) poate fi interpretată ca un model tranzitiv care atinge palierul la o rază de influență mai mică în raport cu distanțele dintre punctele de observație disponibile.

Modelele adecvate acestor situații poartă numele de *modele efect de pepită* și au ecuația:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ 1, & h > 0 \end{cases}$$

Modelul efect de pepită nu este considerat în mod uzual un model elementar dar apare ca o constantă în ecuația majorității modelelor de variogramă de diferite tipuri.

În *mediile izotrope* modelul variogramei depinde numai de distanță și este independent de direcție. În astfel de situații este necesară modelarea variogramei numai pe o direcție și de cele mai multe ori se preferă modelarea variogramei omnidirecționale (cu toleranța direcțională = 90°).

Deși de cele mai multe ori se poate modela satisfăcător o variogramă experimentală cu ajutorul unui model *elementar*, pentru o calare mai riguroasă se preferă un model *complex* obținut prin combinarea lineară a mai multor modele elementare:

$$\gamma(h) = \sum_{i=1}^n |w_i| \cdot \gamma_i(h)$$

Pentru calarea unei combinații de modele elementare la o variogramă experimentală direcțională trebuie identificat în primul rând modelul care dă caracteristica esențială (cu palier, fără palier, etc.). Deseori este necesară combinarea modelelor elementare din categorii diferite. Pentru o variogramă experimentală care nu atinge un palier dar are o comportare parabolică în vecinătatea originii este necesară combinarea unui model gaussian cu altul linear.

Începătorii în analiza variografică sunt tentați să complice modelele pentru o calare foarte exactă a variogramelor experimentale care în etapa de estimare spațială nu se justifică. Principiul economiei în definirea modelelor este un bun ghid în modelarea variogramelor experimentale.

În selectarea caracteristicilor variogramelor experimentale ce trebuiesc modelate este înțelept de luat în considerare existența unei explicații fizice a caracteristicii respective. Dacă informațiile calitative asupra cauzelor fenomenului a cărui structură spațială este studiată explică sau confirmă o anumită caracteristică a variogramei experimentale, atunci este necesar ca aceasta să fie conținută în model, în caz contrar ea poate fi considerată ca rezultat al unui fapt accidental și ignorată în procesul de modelare.

După alegerea modelelor elementare, modelarea variogramei experimentale se transformă într-un simplu exercițiu de calare în care se determină parametrii modelelor teoretice.



2.3. Rezumat

ANALIZA VARIABILITATII SPATIALE are ca obiectiv identificarea legii de variație spațială a parametrilor de mediu de natură alfanumerică/numerică și se realizează în trei etape:

- reprezentarea grafică a datelor prin:
 - harti punctuale
 - harti simbolice
 - diagrama de continuitate
 - diagrama de variabilitate
- analiza parametrică a datelor finalizată sub formă
 - funcția de covarianță ($c(h)$)
 - corelograma ($\rho(h)$)
 - funcția de variogramă ($\gamma(h)$).
- Modelarea matematică a variogramei prin:
 - Modele cu palier
 - Modele fără palier



2.4. Test de autoevaluare a cunoștințelor

- 1) Care sunt legile de variație spațială a parametrilor de mediu
 - a. covarianța
 - b. corelograma
 - c. variograma
- 2) Ce reprezentări grafice se utilizează în analiza spațială a datelor
 - a. Harta punctuală
 - b. Harta simbolică
 - c. Corelograma
- 3) Care sunt modelele fără palier ale variogramei?
 - a. Modelul sferic
 - b. Modelul logaritmic
 - c. Modelul exponențial



2.5. Bibliografie recomandată

Capitolul 3.

EVALUAREA DISTRIBUTIEI SPAZIALE A PARAMETRILOR DE MEDIU



3.1. Introducere

Estimarea distribuției spațiale este o operațiune laborioasă și rezultatul ei este totdeauna discutabil. Obiectivul ei este realizarea *imaginii* distribuției spațiale a unei variabile *pe baza valorilor acelei variabile determinate în diferite puncte din spațiu*. Pentru ecolog aceste *imagini* realizate prin *metode geostatistice* sunt .

Hărțile sunt în același timp suportul și rezultatul cercetărilor ecologice. Tehnica de realizare a hărții se alege în funcție de obiectivul pentru care este utilizată. Într-o clasificare simplă se pot separa trei categorii de astfel de tehnici:

- *estimarea globală*, prin care se obține *valoarea medie reprezentativă* a unei variabile pentru *toată suprafața hărții*. Metoda este utilizată de obicei în faza de prospecțiune pentru evaluarea aproximativă a rezervelor și se aplică în cazul unui număr redus de puncte de observație;
- *estimarea punctuală* este cea mai flexibilă și performantă tehnică de realizare a tuturor categoriilor de hărți cu izolinii. Ea dispune de numeroase variante de prelucrare geostatistică (*kriging punctual ordinar, k.p.universal, k.p.indicator, k.p.disjunctiv* etc.);
- *estimarea zonală* se utilizează pentru *calculul rezervelor* în etapa de explorare detaliată și permite vizualizarea distribuției conținuturilor de substanțe minerale utile pe suprafețe/volume delimitate de contururi cu geometrie cunoscută (ex.: panouri de exploatare, orizonturi miniere etc.).

După ce am colectat *probe* din *n puncte de observație* (puncte marcate cu x-uri în **Fig.2.1**), am determinat prin diferite metode *n valori* ale unei *variabile* ($v_i; i=1...n$) și le-am *analizat variabilitatea* (*globală și spațială*), avem la dispoziție informațiile și instrumentele necesare pentru a putea răspunde la oricare din următoarele întrebări:

- Care este *valoarea medie a variabilei pe toată suprafața*

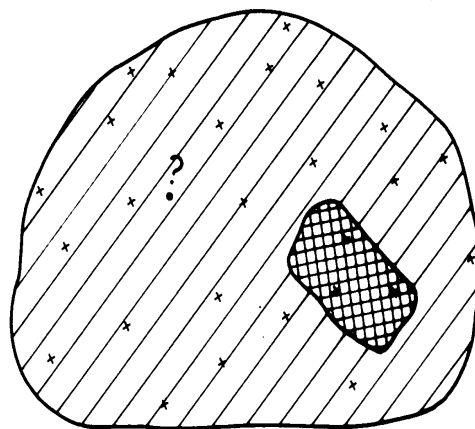


Fig. 3.1. Obiectivele estimării spațiale.

probată (suprafața marcată cu hașură simplă din **Fig.3.1**)?

- Care este *valoarea medie a variabilei probate pe suprafețe delimitate* de contururi închise, mai mici decât întreaga suprafață probată (suprafața marcată prin hașură încrucișată în **Fig.3.1**)?
- Care este *valoarea variabilei într-un punct* oarecare al suprafeței studiate (de exemplu punctul marcat cu semnul întrebării în **Fig.3.1**, punct în care nu am avut acces pentru a măsura variabila).

Indiferent de întrebarea care se pune, răspunsul îl dau metodele *estimării distribuției spațiale univariate*, metode care se bazează de cele mai multe ori pe o combinație lineară de forma:

$$v_o^* = \sum_{i=1}^n w_i v_i$$

în care:

- v_o^* - valoarea estimată a variabilei într-un punct oarecare p_0 ;
- v_i - valorile cunoscute ale variabilei în punctele p_i ($i=1,2,...,n$);
- w_i - ponderile acordate fiecărei valori măsurate.

Combinațiile lineare operează și asupra datelor transformate (*normalize*). Dacă se transformă datele originale și se operează asupra lor o combinație lineară ponderată, se obține un estimator al valorilor transformate:

$$t_o^* = \sum_{i=1}^{i=n} T(v)_i \cdot w_i$$

în care:

- t_o^* - estimatorul valorii transformate a variabilei într-un punct oarecare p_0 ;
- $T(v)$ - funcția de transformare a valorilor variabilei v ;
- w_i - ponderile acordate fiecărei valori transformate.

În astfel de situații, pentru obținerea valorii estimate a variabilei (v_o^*) este necesară transformarea inversă a estimatorului valorilor transformate (t_o^*):

Durata medie de parcurgere a acestui capitol este de aproximativ 9 ore.



3.2. Conținut

3.2.1. Estimare globală

Declusteringul poligonal, pentru eliminarea influenței grupării neuniforme a punctelor de observație atribuie fiecărei valori din setul de date o pondere proporțională cu suprafața *poligonului de influență*.

LINK : <http://alexbeutel.com/webgl/voronoi.html>

Poligonul de influență pentru un punct de observație rezultă din intersecția mediatoarelor segmentelor ce unesc centrul poligonului de influență (O) cu punctele vecine (1,2,3,4,5,6). Punctul 7 nefiind în vecinătatea imediată a punctului O nu modifică forma poligonului de influență al acestuia (**Fig.3.2a,...,f**).

Definitivarea poligoanelor de influență pentru punctele din *vecinătatea limitelor ariei explorate* implică introducerea unor reguli suplimentare. Există două astfel de reguli: închiderea poligonului pe o *limită fizică* (**Fig.3.2a**) sau *convențională* a ariei explorate (**Fig.3.2b**) conturată printr-un arc de cerc a cărui rază este media aritmetică a apotemelor OA , OB și OC .

Separarea în poligoane de influență pe baza metodei prezentate este *unică*.

Fiecărui punct din rețeaua de observație îi este asociat un poligon cu suprafața mai mare în zonele cu densitate mică de puncte de observație și cu suprafață mai mică în zonele cu densitate mai mare de puncte (**Fig.3.3**).

Pentru calculul ponderilor standardizate ale fiecărei valori se utilizează relația:

$$w_i = \frac{a_i}{\sum_{i=1}^n a_i} ,$$

în care a_i este aria fiecărui poligon de influență ($i = 1, 2, \dots, n$; n - numărul punctelor de observație).

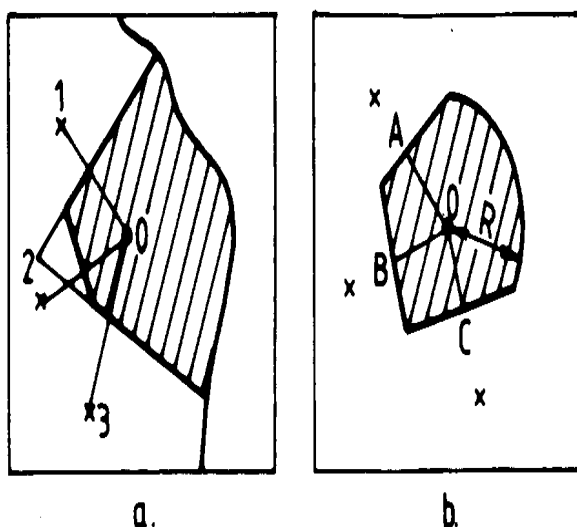


Fig. 3.2. Construirea poligonului de influență pentru un punct din vecinătatea limitei ariei explorate

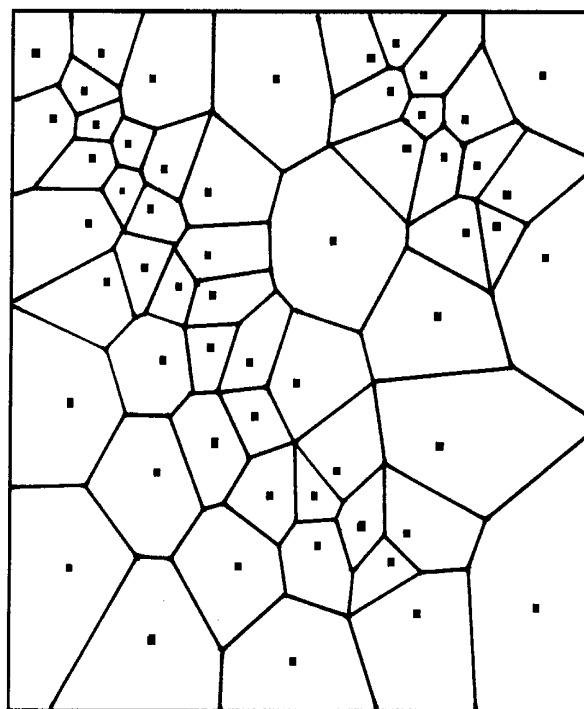


Fig.3.3. Poligoanele de influență ale punctelor de observație dintr-o rețea neuniformă.

Declusteringul celular aplică pe întreaga suprafață a ariei explorate o rețea rectangulară. Fiecare valoare primește o pondere invers proporțională cu numărul de puncte din celula căreia îi aparține. În acest fel, valorile plasate în zone cu densitate mai mare a punctelor de observație primesc o pondere mai mică, iar valorile plasate în zonele cu densitate mai mică a punctelor de observație primesc o pondere mai mare.

Pentru dispoziția punctelor de probare din **Fig.2.5**, plasate în celulele unei rețele rectangulare cu dimensiunea, pe direcția V-E, de 100 m și pe direcția N-S, de 50 m, ponderile acordate valorilor sunt:

- pentru celula A3 : $w = 1/4$;
- pentru celula A2 : $w = 1/6$;
- pentru celula A1 : $w = 1/2$;
- pentru celula B3 : $w = 1$;
- pentru celula B2 : $w = 1$;
- pentru celula B1 : $w = 1/5$.

Deoarece toate valorile din fiecare celulă au ponderi egale și ponderea valorilor din fiecare celulă este unitară, estimarea mediei globale prin metoda declusteringului celular se realizează în două etape:

a - calculul mediei aritmetice simple a valorilor din fiecare celulă;

b - calculul mediei aritmetice simple a valorilor medii din toate celulele rețelei.

Media estimată prin *declustering celular* depinde de dimensiunea celulelor rețelei rectangulare. Dacă celulele sunt prea mici, în fiecare celulă va fi un singur punct de observație și fiecare valoare va avea aceeași pondere. Dacă celulele sunt așa de mari încât toate valorile intră într-o singură celulă, din nou toate valorile vor avea aceeași pondere. Între aceste două situații extreme se află dimensiunile celulelor care permit evaluarea *mediei reprezentative*. Alegerea dimensiunilor optime ale celulei rețelei rectangulare se bazează pe analiza *hărții conturale a mediei globale* construită pentru diferite valori ale dimensiunilor celulei de discretizare a ariei explorate și în corelație cu *obiectivul estimării*.

3.2.2.Estimare punctuala

Ecuatiile sistemului de kriging în funcție de variogramă. În aceleași ipoteze care au permis deducerea expresiei varianței erorii de estimare poate fi utilizată și *variograma* a cărei relație de definiție este:

$$\gamma_{ij} = \frac{1}{2} E\{(V(p_i) - V(p_j))^2\},$$

care pentru evaluarea varianței erorii de estimare poate fi scrisă sub forma:

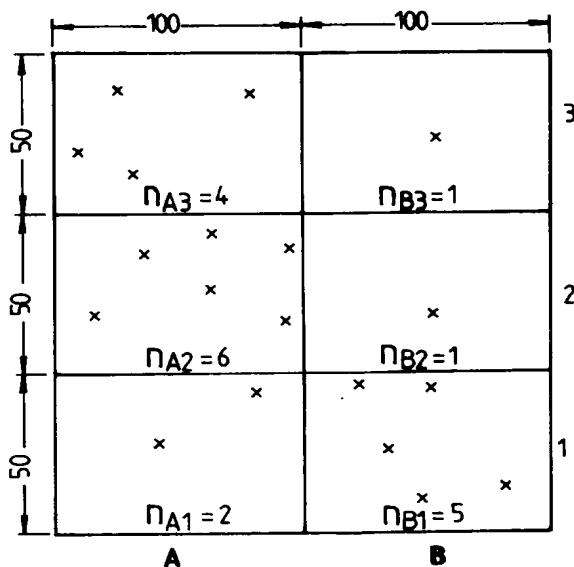


Fig. 2.5.Declustering celular

$$\begin{aligned}
\gamma_{ij} &= \frac{1}{2} E\{[(V(p_i) - V(p_0)) - (V(p_j) - V(p_0))]^2\} = \\
&= \frac{1}{2} E\{[V(p_i) - V(p_0)]^2\} + \frac{1}{2} E\{[V(p_j) - V(p_0)]^2\} - \\
&- E\{[V(p_i) - V(p_0)][V(p_j) - V(p_0)]\} = \\
&= \tilde{\gamma}_{i0} + \tilde{\gamma}_{j0} - E\{[V(p_i) - V(p_0)][V(p_j) - V(p_0)]\}
\end{aligned}$$

Varianța erorii de estimare în această variantă este dată de relația:

$$\begin{aligned}
\tilde{\sigma}_R^2 &= E\{[V^*(p_0) - V(p_0)]^2\} = E\left\{\left[\sum_{i=1}^n w_i \cdot V(p_i) - \sum_{i=1}^n w_i \cdot V(p_0)\right]^2\right\} = \\
&= \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot E\{[V(p_i) - V(p_0)][V(p_j) - V(p_0)]\}
\end{aligned}$$

în care, utilizând forma variogramei din ecuația (2.42), se obține:

$$\begin{aligned}
\tilde{\sigma}_R^2 &= E\{[V^*(p_0) - V(p_0)]^2\} = -\sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \tilde{\gamma}_{ij} + \\
&+ \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \tilde{\gamma}_{i0} + \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \tilde{\gamma}_{j0} = \\
&= -\sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \tilde{\gamma}_{ij} + \sum_{j=1}^n w_j \left(\sum_{i=1}^n w_i \cdot \tilde{\gamma}_{i0} + \sum_{i=1}^n w_i \cdot \tilde{\gamma}_{j0} \right) = \\
&= -\sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \tilde{\gamma}_{ij} + 2 \sum_{i=1}^n w_i \cdot \tilde{\gamma}_{i0}
\end{aligned}$$

Pentru minimizarea varianței erorii de estimare, în condiția de estimare nedeviată (2.21), utilizând parametrul lui Lagrange (μ), relația corespunzătoare ecuației (2.23) scrisă pentru variogramă este:

$$\tilde{\sigma}_R^2 = -\frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \tilde{\gamma}_{ij} + 2 \sum_{i=1}^n w_i \cdot \tilde{\gamma}_{i0} \right) - \mu \left(\sum_{i=1}^n w_i - 1 \right)$$

Minimizarea varianței erorii de estimare scrisă sub forma (2.45) se realizează prin anularea derivatelor în raport cu cele $(n+1)$ necunoscute: $w_1, w_2, \dots, w_n, \mu$.

În mod analog cu ecuațiile (2.26-2.28) se obțin derivatele în raport cu ponderile w sub forma:

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_i} = \frac{1}{2} \left(-2 \sum_{j=1}^n w_j \cdot \tilde{\gamma}_{ij} + 2 \tilde{\gamma}_{i0} \right) - \mu, \forall i = 1, 2, \dots, n$$

iar pentru derivata în raport cu μ

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial \mu} = -\sum_{i=1}^n w_i + 1$$

Sistemul de kriging în raport cu variograma se obține prin anularea celor $(n+1)$ derivate parțiale din ecuațiile (2.46) și (2.47) și are forma:

$$\begin{cases} \sum_{j=1}^n w_i \cdot \tilde{\gamma}_{ij} + \mu = \tilde{\gamma}_{i0}, \forall i = 1, 2, \dots, n \\ \sum_{i=1}^n w_i = 1 \end{cases}$$

care sub formă matricială poate fi scris:

$$\begin{bmatrix} \tilde{\gamma}_{11} & \tilde{\gamma}_{12} & \dots & \tilde{\gamma}_{1n} & 1 \\ \tilde{\gamma}_{21} & \tilde{\gamma}_{22} & \dots & \tilde{\gamma}_{2n} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \tilde{\gamma}_{n1} & \tilde{\gamma}_{n2} & \dots & \tilde{\gamma}_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} \tilde{\gamma}_{10} \\ \tilde{\gamma}_{20} \\ \vdots \\ \tilde{\gamma}_{n0} \\ 1 \end{bmatrix}$$

Valoarea minimă a *varianței erorii de estimare* în raport cu valorile măsurate este dată de relația:

$$\tilde{\sigma}_R^2 = \sum_{i=1}^n w_i \cdot \tilde{\gamma}_{i0} + \mu$$



3.3. Rezumat

Estimarea distribuției spațiale se realizează în două variante distincte:

- Estimarea globală prin:
 - Declustering polygonal
 - Declustering cellular
- Estimarea punctuală prin:
 - Kriging ordinar



3.4. Test de autoevaluare a cunoștințelor

- 1) Care sunt metodele de estimare spațială globală?
 - a. Kriging
 - b. Declustering celular
 - c. Declustering poligonal
- 2) Care este instrumentul de bază în aplicarea kriging-ului?
 - a. covarianța
 - b. corelograma
 - c. variograma

3) Ce efect are asimetria distributiei datelor asupra rezultatelor kriging-ului?

- a. Supraestimarea
- b. Subestimarea
- c. Reducerea erorii de estimare



3.5. Bibliografie recomandată

David, M., *Geostatistical ore reserve estimation*, Elsevier, Amsterdam, 1977.

Davis, J. C., and McCullagh, M. J., *Display of analysis data*, Wiley, New York, 1975.

Delfiner, P., Matheron, G., *Les fonction Aleatoires Intrinseques d'ordre k*, Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau, Ecole de Mines de Paris, 1980.

Delhomme, J.P., *Les variables regionalisees dans les sciences de l'eau*, B.R.G.M., Deuxieme serie, no4, Section III, Hydrogeologie-geologie de l'ingeneur, Paris, 1978.

Deutsch, C.V., Journel, A.G., *GSLIB: Geostatistical Software Library*, New York, Oxford University Press, 1992.

Scrădeanu, D., *Informatică geologică*, Editura Univ.Bucuresti, 1995.

Scrădeanu, D., *Modele geostatistice în Hidrogeologie*, vol.I, Editura didactică și Pedagogică, R.A.-București, 1996.

Scradeanu Daniel, Popa Roxana- *Geostatistica aplicata*, Editura Universitatii Bucurestim 2001

Răspunsuri la testele de autoevaluare a cunoștințelor

Capitolul 1. ANALIZA VARIABILITATII GLOBALE A PARAMETRILOR DE MEDIU

- 1) b
- 2) b
- 3) a și b

Capitolul 2. ANALIZA VARIABILITATII SPATIALE A PARAMETRILOR DE MEDIU

- 1) b
- 2) d
- 3) b

Capitolul 3. ESTIMAREA DISTRIBUTEI SPATIALE

- 1) b și c
- 2) a și b
- 3) b