



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:

MODELAREA PROCESELOR

ECOLOGICE

Titular de disciplină:

Profesor dr. ing. Daniel SCRĂDEANU

București

2022

CUPRINS

Introducere.....	5
a. Date privind titularul de disciplină	5
b. Date despre disciplină.....	5
c. Obiectivele disciplinei.....	6
d. Competențe acumulate după parcurgerea cursului.....	6
e. Resurse și mijloace de lucru	6
f. Structura cursului.....	7
g. Cerințe preliminare	8
h. Durata medie de studiu individual.....	9
i. Evaluarea	9
Capitolul 1. MODELE DINAMICE CALITATIVE.....	10
1.1. Introducere	10
1.2. Conținut	10
Cursul 1. 1. Modele calitative	14
1.1. Principii pentru elaborarea modelelor calitative	14
1.2. Metodologia de elaborare a modelelor calitative	14
1.3. Simboluri standard pentru modele calitative	15
1.3.1. Sursă	15
1.3.2. Depozit	16
1.3.3. Interacțiune	17
1.3.4. Consumator	18
1.3.5. Producător	18
1.3.6. Amplificator	19
1.3.7. Consum energie	20
1.3.8. Tranzacție	20
1.3.9. Simbol “cutie neagră”	20
1.3.10. Conexiuni, forțe, fluxuri.....	21
1.3.11. Ciclu condițional	22
1.3. Rezumat	23
1.4. Test de autoevaluare a cunoștințelor	23
1.5. Bibliografie recomandată	23
Capitolul 2. MODELE DINAMICE CANTITATIVE.....	24
2.1. Introducere	24
2.2. Conținut	24
Cursul 2. 2. Modele cantitative dinamice	24
2.1. Modelul dinamic NETPROD.....	27
2.1.1. Exemple	27

2.1.2. Ecuațiile modelului	28
2.1.3. Aplicație.....	28
2.2. Modelul dinamic RENEW	30
2.2.1. Exemple	30
2.2.2. Ecuațiile modelului	30
2.2.3 Aplicație.....	31
2.3. Modelul dinamic SLOWRENEW	32
2.3.1. Exemple	32
2.3.2. Ecuațiile modelului	33
2.3.3. Aplicație.....	33
2.4. Modelul dinamic EXCLUS	34
2.4.1. Exemple	34
2.4.2. Ecuațiile modelului.....	35
2.4.3. Aplicație numerică	35
2.5. Modelul dinamic INTERACT	36
2.5.1. Exemple	37
2.5.2. Ecuațiile modelului	37
2.5.3. Aplicație.....	37
2.6. Modelul dinamic COOP	38
2.6.1. Exemple	38
2.6.2. Ecuațiile modelului	38
2.6.3. Aplicație.....	39
2.7. Modelul dinamic DESTRUCT	40
2.7.1. Exemple	40
2.7.2. Ecuațiile modelului	40
2.7.3. Aplicație.....	41
2.3. Rezumat	42
2.4. Test de autoevaluare a cunoștințelor	42
2.5. Bibliografie recomandată	42
Capitolul 3 MODELE STATISTICE.....	43
3.1. Introducere	43
3.2. Conținut	43
Cursul 3. 3. Modele cantitative statistice	43
3.1. Cuantificarea intensității corelațiilor	44
3.1.1. Coeficienții de corelație.....	45
3.1.2. Coeficienții de corelație a rangurilor	54
3.1.3. Coeficienți de asociere	60
3.1.4. Coeficienți de corelație temporală	63
3.2. Factorizarea corelațiilor.....	74
3.2.1. Valori proprii și vectori proprii	75
3.2.2. Standardizarea.....	78
3.2.3. Analiza în componente principale.....	79
3.2.4. Analiza factorială R-MOD.....	86
3.2.5. Rotatia factorilor.....	91
3.2.6. Analiza factorială Q-MOD	94

3.3. Modelarea matematică a corelațiilor substanțiale	97
3.3.1. Model liniar de o singură variabilă independentă	97
4.1.2. Evaluarea intensității corelației liniare	102
4.1.3. Testarea adecvării modelului liniar	104
4.1.4. Parametrii modelului	105
4.1.5. Aplicație	107
4.2. Model liniar multiplu	109
4.2.1. Analiza grafică a corelației multiple	110
4.2.2. Evaluarea intensității corelației	111
4.2.4. Parametrii modelului	114
4.2.5. Aplicație	116
3.3. Rezumat	117
3.4. Test de autoevaluare a cunoștințelor	118
n.5. Bibliografie recomandată	118
Răspunsuri la testele de autoevaluare a cunoștințelor	119

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).

Modelarea proceselor ecologice este știința reprezentării organizării și funcționării ecosistemelor în scopul înțelegerii creșterii și evoluției acestora.

Modelarea și **simularea** sunt instrumentele de **conectare** a **ideilor**, ce constituie formularea **abstractă** a **interacțiunii** factorilor biotici și abiotici în **procesele** specifice **ecosistemelor** cu **realitatea** reprezentată de ecosisteme.

Obiectivele modelării și simulării ecosistemelor sunt:

- o înțelegerea completă și a funcționării ecosistemelor
- o controlul riguros cantitativ al creșterii și evoluției spațio-temporale a ecosistemelor.

MODELAREA are ca obiectiv specific reprezentarea sintetică a organizării și funcționării ecosistemelor (sub forma grafică și cantitativă).

SIMULAREA dă “viață” modelelor în scopul controlului cantitativ al creșterii și evoluției ecosistemelor (prin intermediul ecuațiilor și sistemelor de ecuații asociate proceselor cercetate).

a. Date privind titularul de disciplină

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

b. Date despre disciplină

Anul de studiu:	III, Disciplina OPTIONALA
Semestrul:	5

c. Obiectivele disciplinei

Obiectivul general al disciplinei	Familiarizarea studenților cu orientările moderne și nivelul actual de modelare cantitativa a proceselor ecologice
Obiectivele specifice	Evaluarea cantitativa a proceselor ecologice prin cuplarea modelor statistice cu cele numerice: Modele analitice cuplate cu modele de tip statistic Modele de tip topo-probabilist cuplate cu modele numerice

d. Competențe acumulate după parcurgerea cursului

Competențe profesionale	C1. Identificarea și utilizarea principalelor legități, noțiuni și concepte specifice Ecologiei și protecției mediului. C1.1. Definirea principiilor și legilor aplicabile în Ecologie și protecția mediului pentru abordarea problemelor specifice Ecologiei C1.2. Formularea de probleme de ecologie în termeni clari prin asocierea de modele experimentale și teoretice fenomenelor de bază din Ecologie și C4. Identificarea alternativelor optime în vederea caracterizării ecologice corespunzătoare a factorilor de mediu și elaborarea de măsuri privind protejarea acestora. C4.1. Interpretarea datelor achiziționate pe parcursul procesului de investigare și alegerea alternativelor optime pentru caracterizarea ecologică a apei, aerului, solului, biotei și a relațiilor dintre acestea. C4.2. Evaluarea gradului de incertitudine al măsurărilor și compararea rezultatelor cu date bibliografice.
Competențe transversale	CT2. Aplicarea tehnicilor de muncă eficientă în echipă multidisciplinară pe diverse paliere ierarhice. CT3. 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 ***MODELARE A APROCESELOR ECOLOGICE*** este alcătuit din 3 capitole (fiecare capitol fiind aferent unei **unități de învățare**):

Capitolul I. **MODELE DINAMICE CALITATIVE**

Capitolul II. **MODELE DINAMICE CANTITATIVE**

Capitolul III. **MODELE STATISTICE**

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.

Totodată, sunt formulate **3 teme de control**, în vederea evaluării pe parcurs, astfel:

Tema 1 – Capitolul 1: „Realizați un referat de maximum 5 pagini în care să tratați *semnele conventionale standard utilizate pentru elaborarea modelelor dinamice*”;

Tema 2 – Capitolul 2: „Realizați un referat de maximum 5 pagini în care să tratați minimum *trei modele dinamice cantitative* (ex.: SLOWRENEW, NETPROD, EXCLUS”;

Tema 3 – Capitolul 3: „Realizați un referat de maximum 5 pagini în care să tratați *modelul statistic cantitativ liniar bivariat*”.

Temele de control se prezintă de către fiecare student, iar rezultatele acestei evaluări se comunică de către cadrul didactic în mod direct și nemijlocit studenților.

Cursul de **MODELARE A APROCESELOR ECOLOGICE** 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 STATISTICA ECOLOGICA și la BILOGIE VEGETALA SI ANIMALA, FIZICA SI CHIMIA MEDIULUI, GEOGRAFIA SI GEOLOGIA GENERALA, ATMOSFERA SI CALITATEA AERULUI, HIDROLOGIE SI HIDROGEOLOGIE, ECOLOGIE GENERAL, având în vedere că disciplina se studiază în anul III, semestrul 5.

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

La încheierea studierii cursului, este obligatorie realizarea celor trei teme de control.

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%
	- proiect	20%
	- verificare pe parcursul semestrului	
	- teme de control	20%

Capitolul 1.

MODELE DINAMICE CALITATIVE



1.1. Introducere

În cuprinsul acestui capitol sunt prezentate cunoștințele referitoare la **MODELE DINAMICE CALITATIVE**, precum și conceptele, principiile și legitățile fundamentale ale modelării cantitative a proceselor ecologice

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



1.2. Conținut

Strategie de modelare a ecosistemelor

Modelarea ecosistemelor se bazează pe o *analiză* detaliată a acestora cu două obiective principale:

- stabilirea *factorilor* biotici și abiotici ai ecosistemului cercetat;
- identificarea *interacțiunilor* dintre factorii biotici și abiotici în ambianța proceselor ecosistemului.

Rezultatele analizei ecosistemului sunt *sintetizate* în două *modele* realizate succesiv:

- *Modelul calitativ* (conceptual) al ecosistemului, realizat de regulă sub forma unor diagrame în care sunt figurate prin *simboluri* componentele ecosistemului și prin *linii* legăturile dintre ele.
- *Modelul cantitativ* realizat prin completarea diagramelor cu *numere, de același tip sau de tipuri diferite*.

În etapă finală este utilizat *modelul cantitativ calibrat* pe datele experimentale obținute în programul de monitorizare al ecosistemului, pentru *simularea evoluției spațio-temporale a ecosistemului* în două situații distincte:

- *evoluția ecosistemului în condițiile naturale*, în lipsa unui stress extern care să modifice condițiile în care s-a elaborat și calibrat modelul cantitativ;

- *evoluția spațio-temporală a ecosistemului în condiții de stress* natural sau antropic, stress care modifică parametrii energetici și materiali ai acestuia.

Etapele metodologiei de modelare

Metodologia de realizare a celor două tipuri de modele, calitativ și cantitativ, poate fi separată în patru etape:

- definirea *frontierelor modelului calitativ*, prin simbolizarea suprafeței în care vor fi reprezentate toate componentele și interconexiunile sistemului (**Fig.1**):
- plasarea *componetelor* ecosistemului (**Fig.2**):
 - sursele de energie și de materie
 - componentele de stocare
 - producători și consumatori
- trasarea *interconexiunilor* dintre componentele *modelului calitativ* al ecosistemului (**Fig.3**.)
- inițializarea numerică a *modelului cantitativ* prin plasarea *numerelor* în diagrama *modelului calitativ* (particularizare pentru ciclul fosforului: valorile sunt exprimate în [grame/m².an], **Fig.4**)

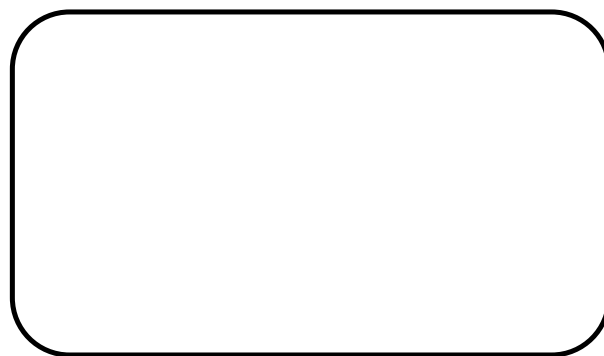


Fig.1. *Frontierele ecosistemului*

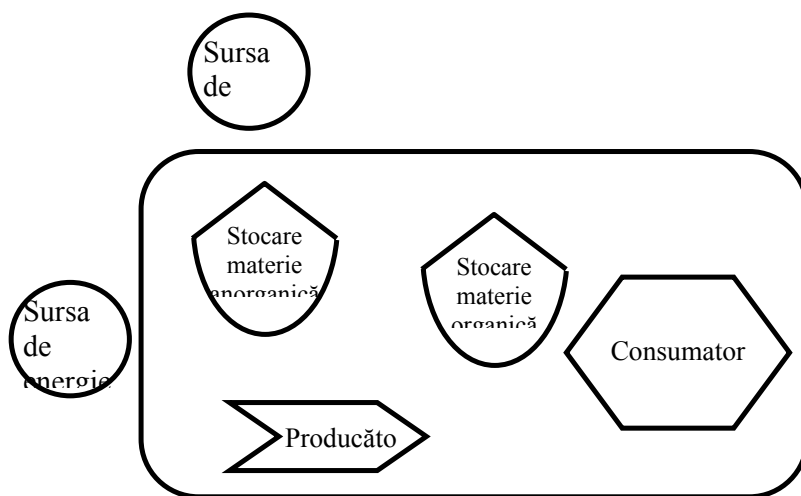


Fig.2. Componentele unui process ecologic

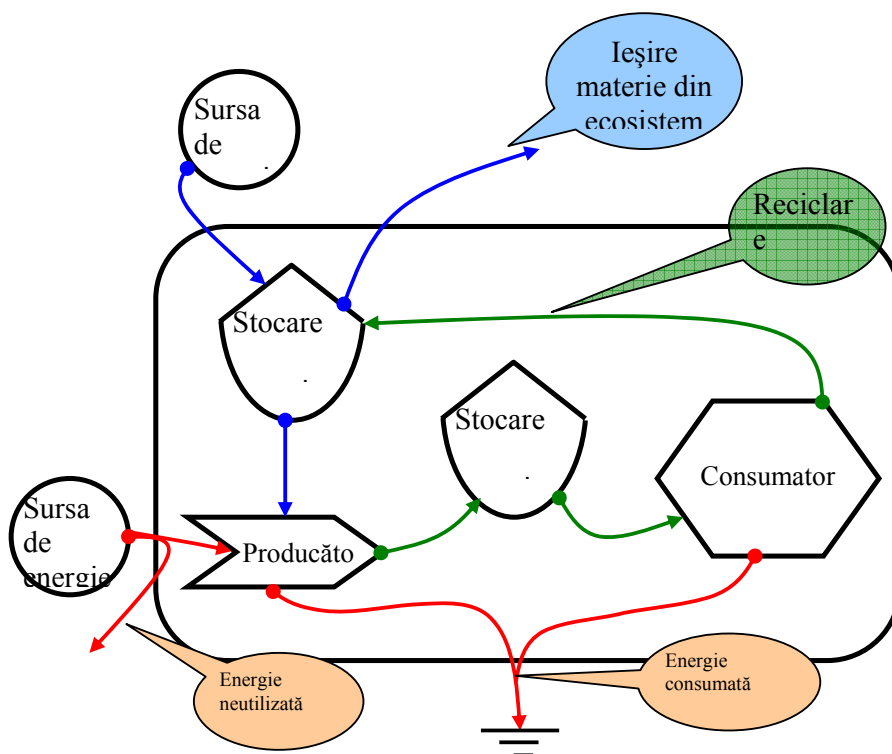


Fig.3. Trasarea interconexiunilor dintre componentelor ecosistemului .

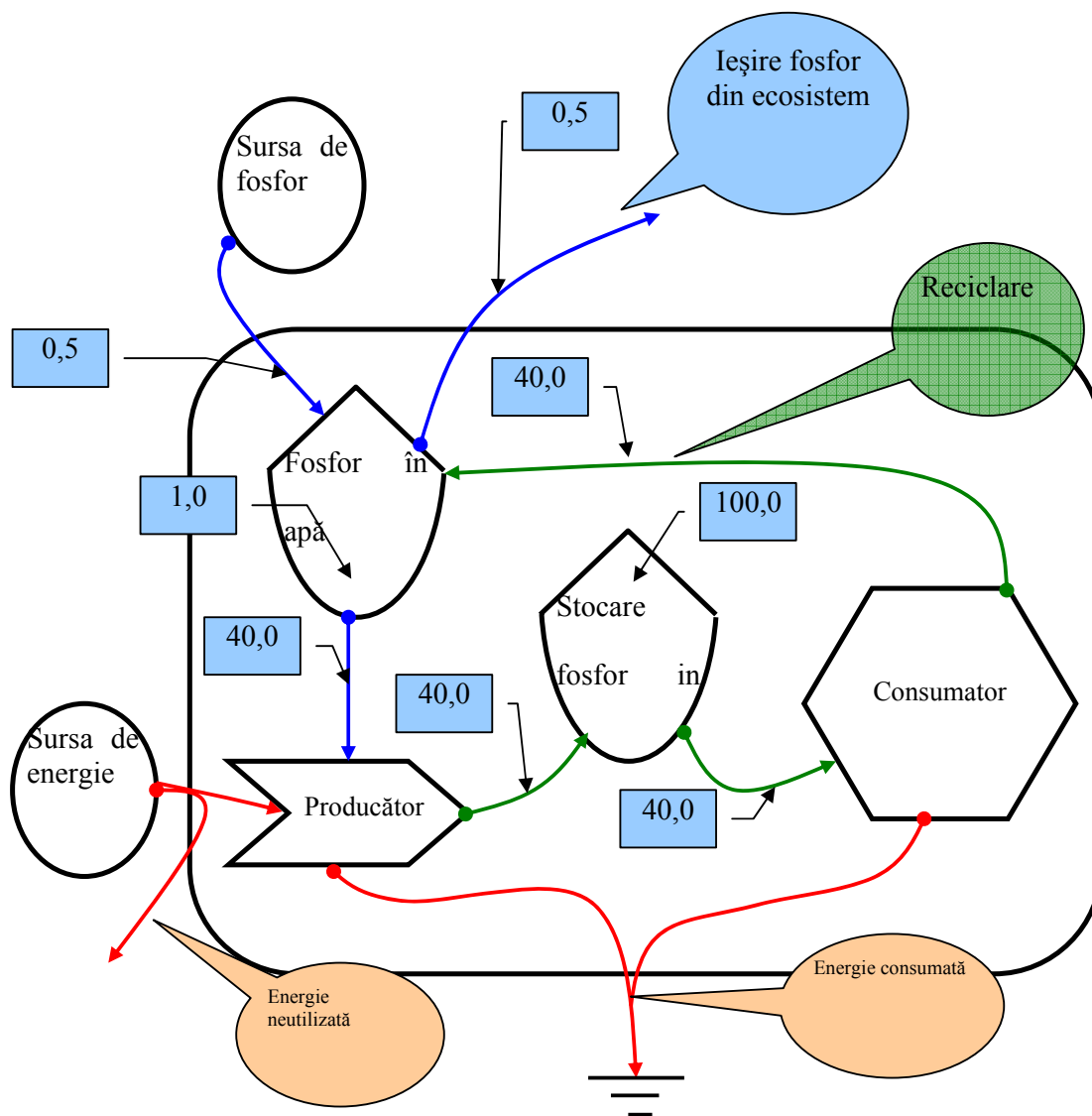


Fig.4. Obținerea **modelului cantitativ** prin plasarea numerelor în diagrama **modelului calitativ** (exemplificare pentru ciclul fosforului într-un ecosistem; valorile sunt exprimate în [grame/metru pătrat și an])

Cursul 1. 1. Modele calitative

1.1. Principii pentru elaborarea modelelor calitative

Elaborarea modelelor calitative, în variantă grafică, are avantajul că vizualizează într-un mod intuitiv componentele și relațiile dintre acestea, fără a înlocui realitatea palpabilă cu simboluri abstracte specializate, de tipul ecuațiilor matematice, care presupun o anumită specializare pentru o manipulare eficientă.

Realizarea diagramelor grafice pentru modelele calitative respectă câteva principii generale:

- **sursele principale de energie** se amplasează în afara sistemului modelat și liniile de legătură traversează frontierele acestuia;
- consumul de energie se face în general de la stânga la dreapta și de sus în jos;
- fiecare sistem are o pierdere de energie pe frontiera inferioară, pierdere inevitabilă conform cu principiul al doilea al termodinamicii.;

1.2. Metodologia de elaborare a modelelor calitative

Elaborarea **modelelor calitative** este prima etapă obligatorie în analiza numerică și simularea oricăror procese.

Rezultatul acestei prime etape este **modelul conceptual** pe care se fundamentează toate evaluările cantitative. CORECTITUDINEA MODELULUI CONCEPTUAL ESTE CHEIA EVALUĂRII CORECTE A PROCESELOR ECOLOGICE.

Etapele realizării modelului calitativ sunt :

1. Trasarea limitelor sistemului
2. Consemnarea tuturor traseelor care traversează limitele sistemelor (**intrări** și **iesiri**)
 - Plasarea fiecărui **inceput** într-o sursă plasată în afara sistemului studiat
 - Marcarea simbolurilor de sursă cu **Cuvinte** sugestive

3. Consemnarea *componentelor* sistemului:
 - *Lista* completa a componentelor cercetate
 - *Plasarea* lor in interiorul sistemului de la stanga la dreapta in ordinea intrarii in actiune;
4. Consemnarea *proceselor* din system
 - O *lista* cu procesele importante
 - *Conexiunile* intre componentele implicate de fiecare process
5. Marcarea *conservarii masei* prin evidentierea clara de-a lungul proceselor pentru:
 - Intrari;
 - Stocari
 - Iesiri
6. Verificarea circuitului *banilor* in system
7. Marcarea *circuitului energiei* prin:
 - Intrari
 - Consum
 - Iesiri
8. Utilizarea *culorilor* standardizate pentru intocmirea diagramelor:
 - **roșu** pentru circuitul *energiei*
 - **albastru** pentru circuitul *materiei din biosfera*: aer, apa, nutrienți
 - **maron** pentru componentele *geologice*, combustibil, minereu
 - **verde** pentru zona ambientală, *producatori, productie*
 - **portocaliu** pentru *consumatori*: animale, oameni, industrie etc.
 - **purpuriu** pentru bani
9. Definitivarea modelului pentru obiectivul studiului
 - *detaliere* pentru studiul știintific detaliat
 - *sinteza* pentru discutii cu beneficiari ai rezultatelor (public, politicieni)

1.3. Simboluri standard pentru modele calitative

1.3.1. Sursă

Simbolul utilizat în mod curent pentru **sursă** este circular, se plasează în exteriorul frontierelor ecosistemului și simbolizează aportul de informație, materie sau energie în ecosistem (**Fig.5**). În interiorul cercului este precizat prin text tipul de aport (sursă de materie anorganică/organică, energie solară etc.).

Plasamentul surselor este de regulă în partea stângă a frontierelor ecosistemului, consumul de energie, materie sau informație realizându-se de la stânga spre dreapta (**Fig.4**).

1.3.2. Depozit

Stocarea, sub diferite forme (energiei, materiei, informație, structură) se reprezintă cu un **simbol de bază** (**Fig.6a**).

Stocarea într-un depozit fiind nelimitată trebuie să existe nu numai căi de **intrare** dar și căi de **ieșire** (difuzie, dispersie sau depreciere), ambele fiind de același tip.

Stocările specializate sunt reprezentate prin diferite conexiuni asociate simbolului de bază:

- stocarea energiei cu indicarea deprecierei acestei (**Fig.6b**);
- stocarea energiei și materiei cu consumul energiei și deprecierea materiei (**Fig.6c**).

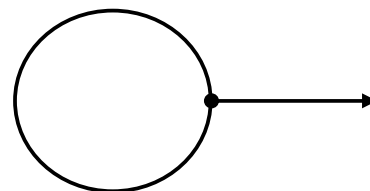


Fig.5. Simbol utilizat pentru surse de informație, materie sau energie.

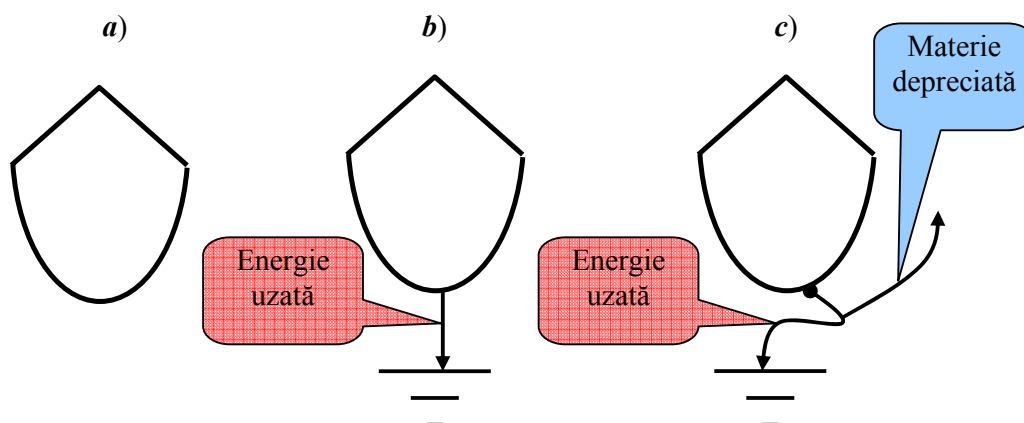


Fig.6. Simboluri pentru stocare elementară(**a**), stocare de energie cu deprecierea acesteia (**b**) și stocare de energie și materie cu depreciere energiei și materiei (**c**).

1.3.3. Interacțiune

Simbolul pentru **interacțiune** reprezintă o transformare și conține (**Fig. 7**):

- căi prin care sunt simbolizate afluxurile de materie sau energie;
- casetă în care se produce **transformarea**;
- una sau mai multe ieșiri pentru **produsul** rezultat, energia consumată etc.

Există diferite **tipuri** de interacțiuni pentru care se atașează diferite atribute suplimentare simbolului elementar de interacțiune:

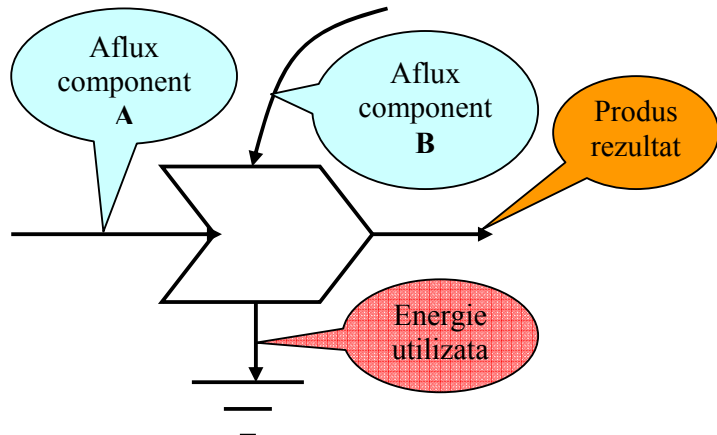


Fig. 7. Simbolul utilizat pentru interacțiunea din care rezultă diverse produse.

- **interacțiune cu niveluri de intensitate variabilă**

a transformărilor, poziționate în ordinea crescătoare de la stânga la dreapta în diagrama modelului calitativ (**Fig. 8a**).

- **interacțiune cu diluție** (**Fig. 8b,c**), în care produsul rezultat este proporțional cu afluxul de materie și energie, **divizat** sau **reduc** proporțional cu ponderea factorilor care sunt

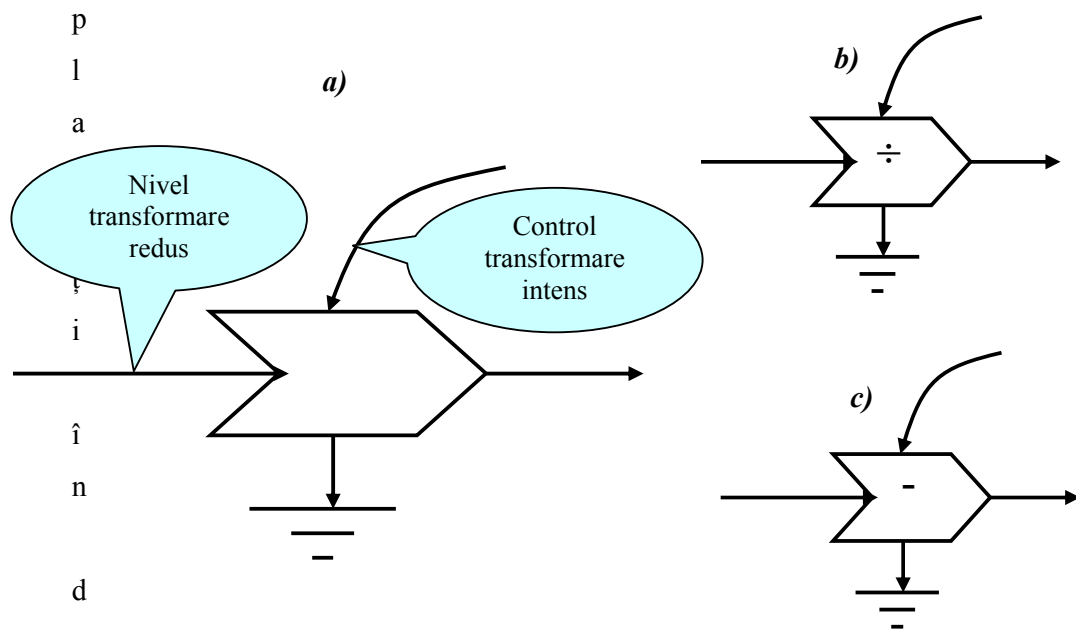


Fig. 8. Interacțiuni cu niveluri de transformare diferențiate (**a**), cu diluție divizată (**b**) sau redusă (**c**)

apta simbolului de interacțiune (exemplu: cantitatea de plancton dintr-un lac este redusă prin diluția apei rezultată din alimentarea lacului);

1.3.4. Consumator

Simbolul pentru **consumator** se referă la un grup de acțiuni, în mod uzual reprezentate prin cuplul transformare-stocare, încadrate într-un hexagon (**Fig.9a**).

Procesul de transformare din de interacțiune (**Fig.8a**) este un proces de transformare **primar** și devine **secundar** plasat într-un simbol grup de consumator.

Diversele variante de proces consumator se diferențiază prin simbolurile plasate în interiorul hexagonului de bază:

- flux de consum proporțional cu factorii determinanți (ex.: consumul microbilor proporțional cu zahărul disponibil) (**Fig.9b**);

- fluxul de consum proporțional cu productiv determinat de doi factori (ex.: descompunerea substanțelor organice proporțional cu concentrația materiei organice și concentrarea oxigenului) (**Fig.9c**);

- fluxul de consum este proporțional cu sursele de materie și energie precum și semnalul de feedback dat de stocarea proprie (ex.: creșterea zooplanctonului proporțional cu cantitatea de hrană și concentrația de oxigen (**Fig. 9d**).

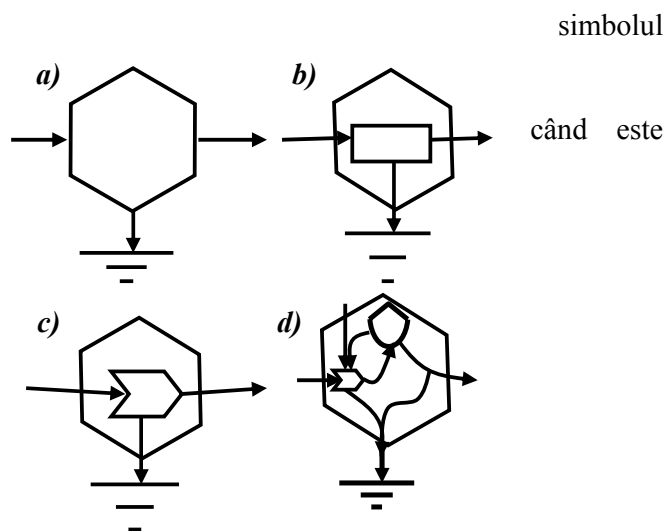


Fig.9 .Utilizarea simbolului de grup consumator.

1.3.5. Producător

Simbolul pentru **producător** implică o unitate de producere și de cele mai multe ori una de stocare

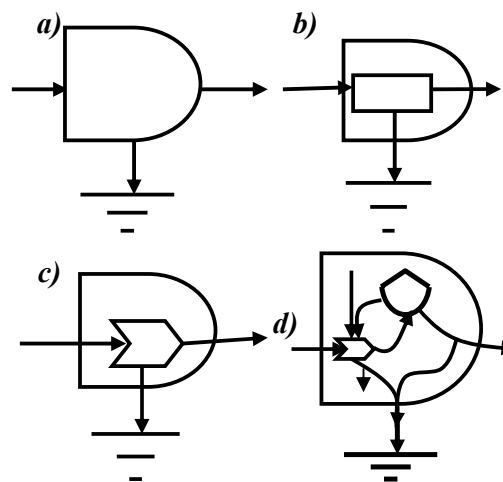


Fig.10 .Utilizarea simbolului de grup producător.

a produsului creat. Pentru simbolizarea unui producător se utilizează, în cel mai general caz, un cadru care maschează o structură internă detaliată (**Fig.10a**) iar pentru precizarea unor caracteristici ale structurii interne se adaugă atributele necesare:

- producător influențat proporțional cu concentrația aportului de energie (ex.: producere de materie organică prin procesul de fotosinteză, proporțional cu concentrația luminii) (**Fig.10b**);
- producător “stimulat” simultan de două aporturi (ex.: stimularea fotosintezei de concentrația luminii și a nutrienților) (**Fig.10c**);
- producător stimulat proporțional cu aportul de energie/materie și controlat prin feedback-ul rezultat de stocarea produsului (ex.: producția de fitoplancton stimulată de concentrația de lumină și nutrienți, și inhibată de cantitatea de produs stocată (**Fig.10d**).

1.3.6. Amplificator

Acest operator simbolizat printr-un **triunghi** (**Fig. 11a**) controlează aportul de materie/energie din diferite surse, aport care **aplică** intensitatea unui proces de consum/producție (ex.: reproducerea organismelor care poate fi stimulată de o cantitate suplimentară de hrană) (**Fig.11b**).

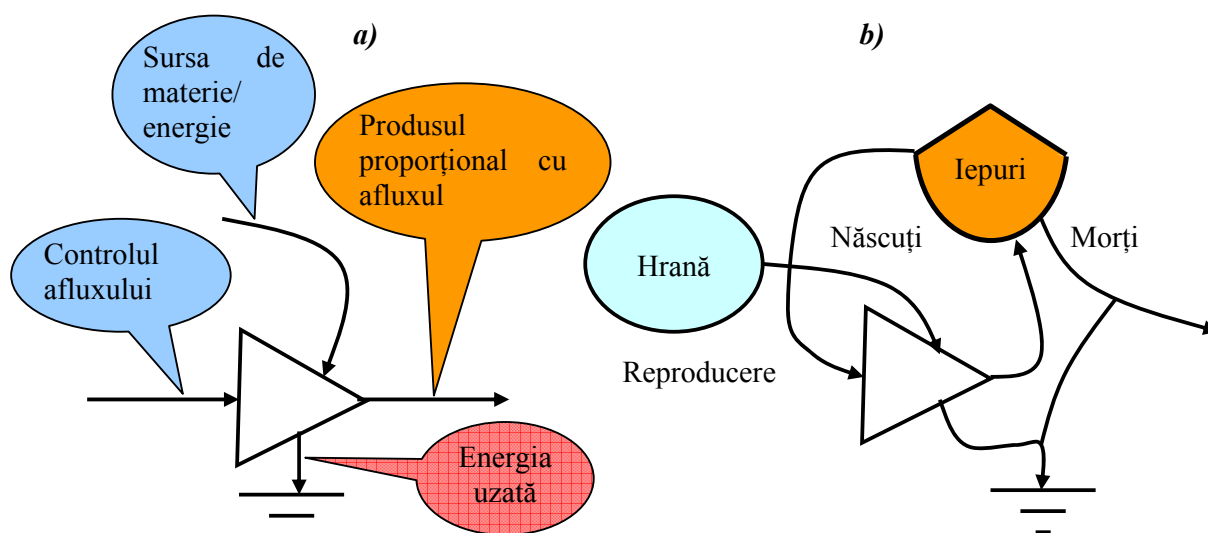


Fig.11. Amplificator cu rată constantă (a) cu un exemplu de reproducere cu amplificator stimulat de aport de hrană nelimitat (b).

Fig.12. Consum ireversibil de energie

1.3.7. Consum energie

Fiecare ecosistem trebuie să aibă, pentru ca modelul să respecte legea a doua a termodinamicii, poziționat pe frontiera de la bază, un simbol care să figureze pierderea/consumul/dispersia de energie în afara sistemului, nerecuperabilă și neregenerabilă (**Fig.12**).

Simbolul nu trebuie confundat cu cel de legare la pământ al unei surse electrice.

1.3.8. Tranzacție

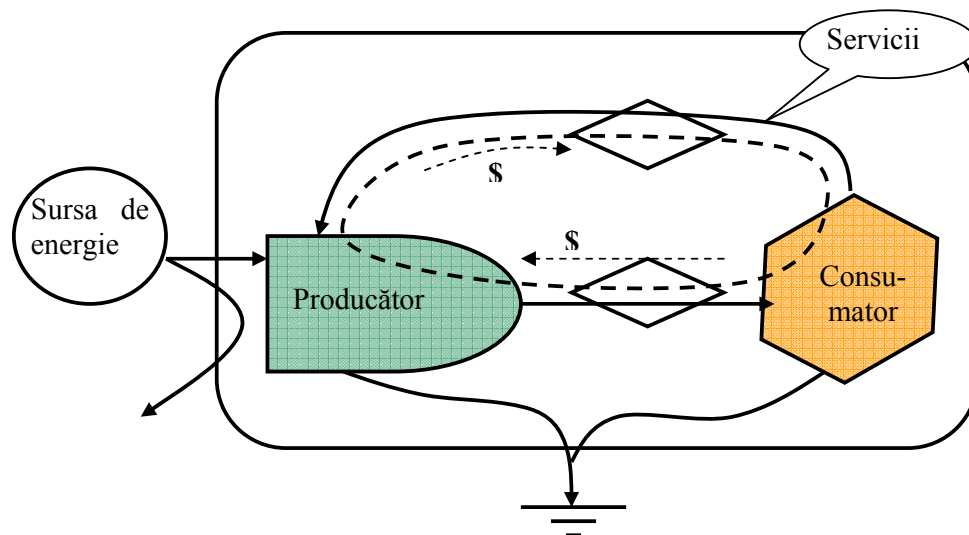


Fig.13. Circulația banilor într-un ciclu de producere și consum

Circulația **banilor** în cadrul **tranzacțiilor** asociate diferitelor procese de producție și consum este în general în **sens contrar** sensului de consum al energiei și materiei și se reprezintă prin linie întreruptă (**Fig.13**).

Pentru situații particulare se completează circuit banilor, a căror valoare se conservă în circuitul proceselor asociate, cu simboluri suplimentare (**Fig.14**).

1.3.9. Simbol “cutie neagră”

Simbolul de cutie neagră este utilizat pentru a reprezenta componente cu structură internă

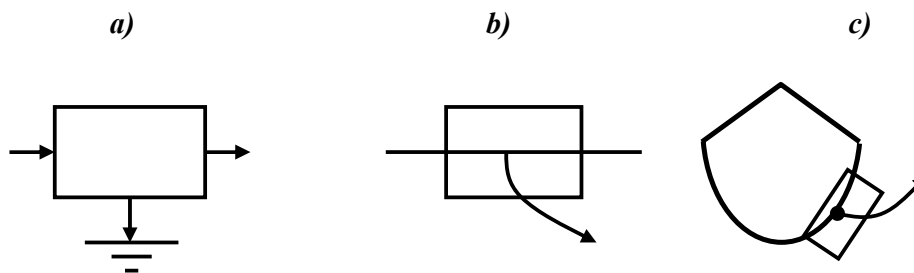


Fig.15. Simbolul cutie neagră utilizat pentru: **a)** componente cu structură internă necunoscută; **b)** forțe rezultate dintr-un flux principal; **c)** senzori pentru identificarea unor componente secundare rezultate dintr-un anumit proces.

necunoscută (**Fig.15a**), sau simboluri pariculare ale unor ecosisteme (cu apariție extrem de rară; **Fig.b,c**).

1.3.10. Conexiuni, forțe, fluxuri

Structura ecosistemelor este constituită din simbolurile componentelor legate prin linii de diferite tipuri: conexiuni, forțe, fluxuri.

O linie de legătură poate fi utilizată pentru: material, informație, organisme, populație, energie etc.

Fluxurile sunt activate de **forțe**, forțe reprezentate prin: forțe fizice, concentrație chimică, sau orice alte proprietăți ce au energia necesară întreținerii unui flux. Forțele provin dintr-o sursă exterioară sau dintr-un stocaj intern.

Fluxurile sunt diferențiate grafic în funcție de particularitățile de circulație și de numărul

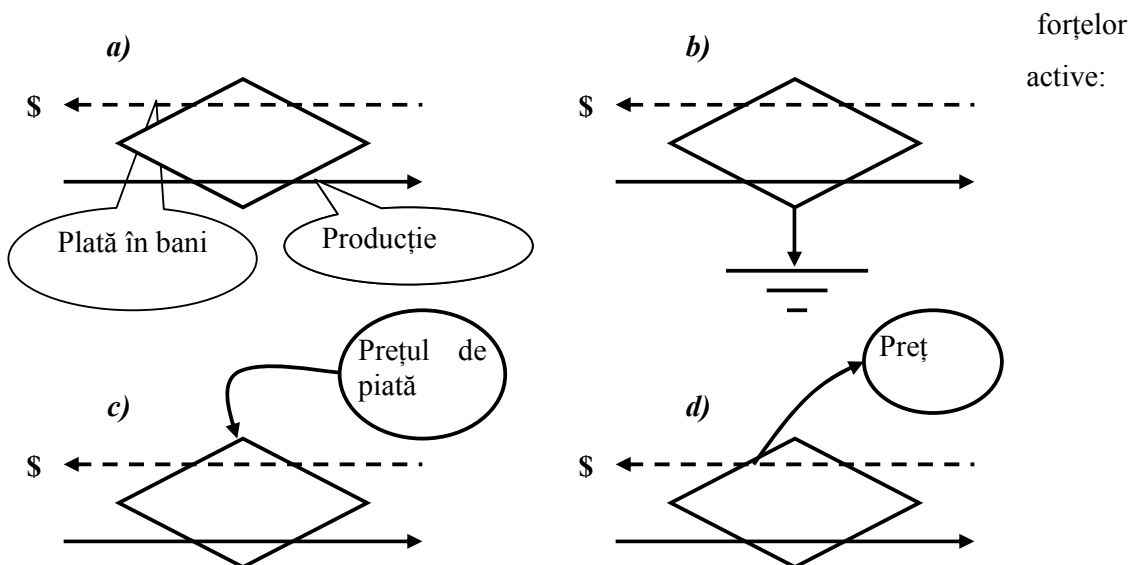


Fig.14. Circulația banilor în diferite tipuri de tranzacții: **a)** cumpărare; **b)** tranzacție cu pierdere de energie; **c)** tranzacție cu preț dictat de un sistem mai mare; **d)** flux dintr-un sistem mai mare care stabilește prețul de piață.

- fluxul proporțional cu o singură forță, de tip linear, este reprezentat printr-o linie cu o singură săgeată, indiferent de prezența sau absența unei pierderi sau transformări de energie (**Fig. 16a,b,c**);
- flux divizat sau combinat din două fluxuri de același tip (**Fig. 16d,e**);
- flux dependent de diferența de forțe de la cele două capete ale circuitului (**Fig. 16f**).

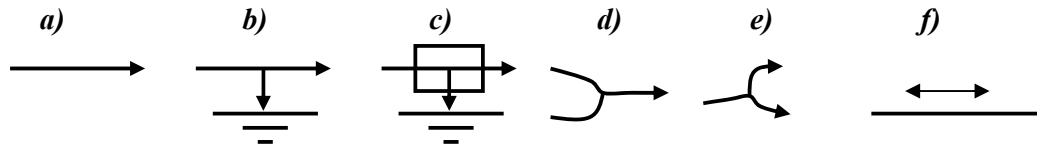


Fig.16. Diferite tipuri de fluxuri din structura ecosistemelor: **a)** flux linear cu o forță; **b)** flux linear cu pierdere de energie; **c)** flux linear cu transformare de energie; **d)** combinarea a două fluxuri de același tip; **e)** divizarea în două fluxuri de același tip; **f)** flux dependent de diferența dintre forțele de la capetele circuitului.

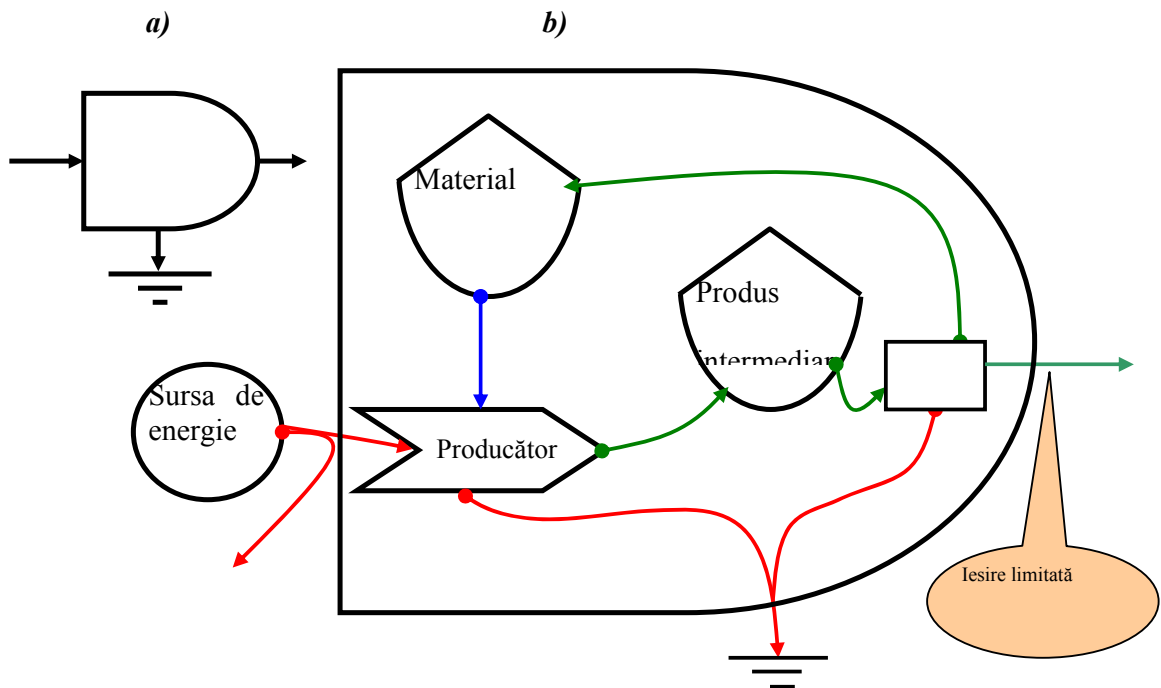


Fig.17. Ciclu limitativ: **a)** fără structură internă cunoscută ("black box") sau cunoscută, dar nereprezentată, din rațiuni de simplificare a diagramei ("white box"); **b)** cu structura internă cunoscută și reprezentată.

1.3.11. Ciclu condițional

Ciclul condițional limitativ/de maxim (**Fig.17a**) este un simbol de grup care limitează ieșirea dintr-un sistem la creșterea energiei provenite dintr-o sursă internă.

Ciclul condițional limitativ este utilizat pentru un flux energetic al unei unități cu un ciclu intern propriu. Este cazul procesului de producere de oxigen și substanța organică prin fotosinteză (*Fig.17b*):

- în primul pas clorofila primește energie (lumina de la soare) și produce sarcina pozitivă sau negativă;
- al doilea pas se produce oxigen și substanță organică și se “resetează” clorofila ca să poată primi din nou energie pentru un nou ciclu de producere, declanșat numai dacă mai există materie primă disponibilă.



1.3. Rezumat

Realizarea modelelor dinamice calitative se bazează pe un set redus de simboluri care se conectează în scheme grafice care materializează sensul și tipul de corelații dintre componentele ecosistemului.



1.4. Test de autoevaluare a cunoștințelor

- 1) Care este succesiunea etapelor de modelare a procedurilor ecologice?
 - a. definire frontiere-plasare componente-trasare interconexiuni
 - b. plasare componente - definire frontiere- trasare interconexiuni
 - c. trasare interconexiuni- plasare componente- definire frontiere.
- 2) Unde se plasează sursele de energie ale procesului ecologic?
 - a. în exteriorul frontierelor sistemului
 - b. în interiorul frontierelor sistemului
 - c. Între producător și depozitul de stocare
- 3) Ce culori se utilizează pentru marcarea circuitului din biosferă, aer, apă, nutrienți?
 - a. roșu
 - b. albastru
 - c. verde



1.5. Bibliografie recomandată

- a. Ventsel H, 1973: Théorie des probabilités. - Moscou: Edit. MIR
- b. Chaston I, 1971: Mathematics for Ecologists. - London: Butterworths, 132 pp
- c. Daget J, 1976: Les modèles mathématiques en écologie. – Paris etc.: Masson,.
- d. Digby PGN & Kempton RA, 1987: Multivariate Analysis of Ecological Communities. - London & New York: Chapman & Hall, viii + 206 pp.
- e. Patil GP, Pielou EC & Waters WE (edit.), 1971: Statistical ecology. - Penn. State Univ. Press

Capitolul 2. MODELE DINAMICE CANTITATIVE



2.1. Introducere

În cadrul acestui capitol, studenților le sunt prezentate.....

Durata medie de parcurgere a acestui capitol este de aproximativ 2 ore.



2.2. Conținut

Cursul 2. 2. Modele cantitative dinamice

Modelele *cantitative dinamice* se construiesc pe structura modelului conceptual reprezentat de modelul calitativ al ecosistemului prin:

- introducerea *numerelor* în diagrama modelului calitativ;
- atașarea *ecuațiilor* modelului calitativ.

Introducerea *numerelor* în diagrama modelului calitativ îl transformă în model cantitativ. Cu ajutorul numerelor introduse în diagramele modelelor calitative se poate sesiza unde stocarea sau fluxul sunt mai mari sau mai mici. Diagramele cu numere au calitatea de a reprezenta sintetic și sugestiv caracteristicile cantitative generale ale ecosistemului.

Ecuațiile asociate modelului calitativ permit construirea unui model cantitativ care permite:

- analiza detaliată a *evoluției* componentelor ecosistemului;
- *prognoza* evoluției ecosistemului în etapa de *simulare*, pentru diverse condiții (cele monitorizate sau generate de situații excepționale: catastrofe naturale, poluări accidentale).

Inițializarea numerică a modelelor calitative se bazează pe date obținute prin *monitorizarea* componentelor ecosistemului cercetat, pe o perioadă îndelungată de timp în care pot fi sesizate tendințele de variație temporală și spațială.

Introducerea numerelor în diagramele modelelor se face, în funcție de complexitatea ecosistemului studiat, în două variante:

- cu numere de același tip;
- cu numere de tipuri diferite.

Diagramele care urmăresc fluxul unui singur component sunt completate cu numere exprimate în aceeași unitate de măsură.

Știile biochimice, de cele mai multe ori, urmăresc fluxul unui singur component chimic și în această situație pe toate liniile de conexiune ale componentelor sunt plasate valorile componentului respectiv în aceeași unitate de măsură.

Reprezentarea cantitativă prin numere a ciclului pentru fosfor într-un ecosistem (**Fig.18**), poate fi exprimată prin numere care iată cantitatea de fosfor în *[grame/metru pătrat și an]* și trebuie completată pe toate conexiunile cu excepția conexiunii cu sursa de energie primară și conexiunea care indică pierderea de energie din baza diagramei ecosistemului (energia pierdută)

O diagramă similară poate fi completată cu **energia** consumată pe fiecare tronson și exprimată în *[10⁶ Joule/metru pătrat și an]* (**Fig.19**).

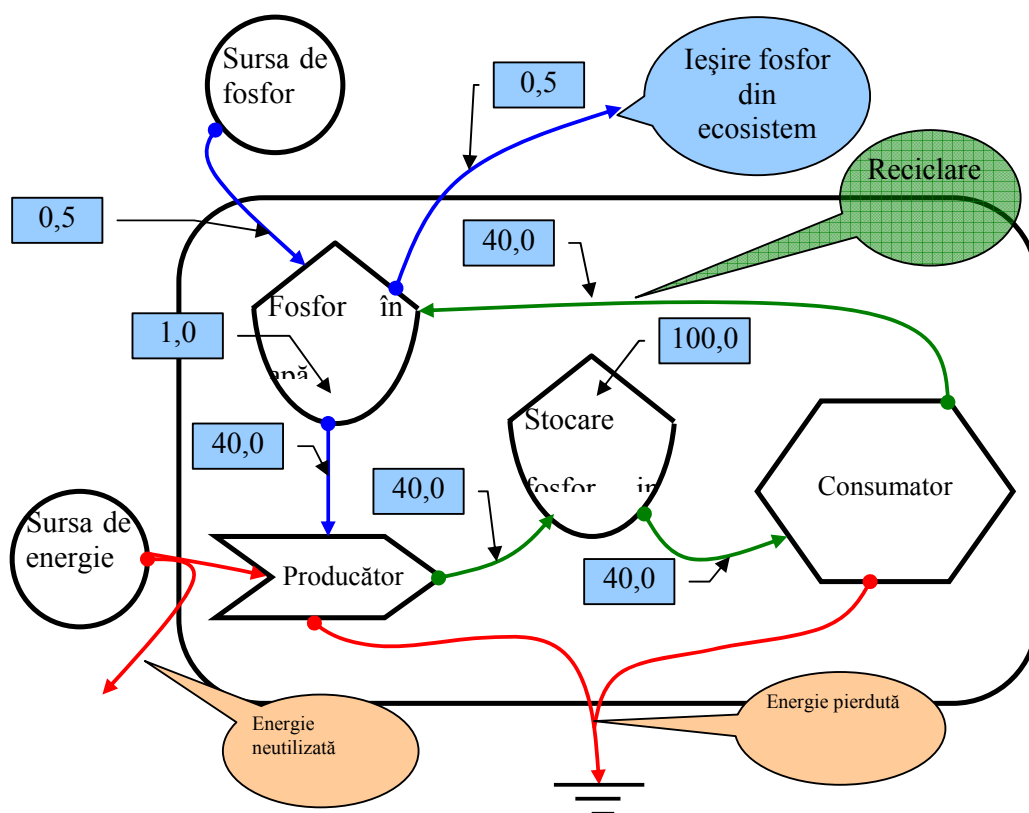


Fig.18. Obținerea **modelului cantitativ** prin plasarea numerelor în diagrama **modelului calitativ** (exemplificare pentru ciclul fosforului într-un ecosistem; valorile sunt exprimate în *[grame/metru pătrat și an]*)

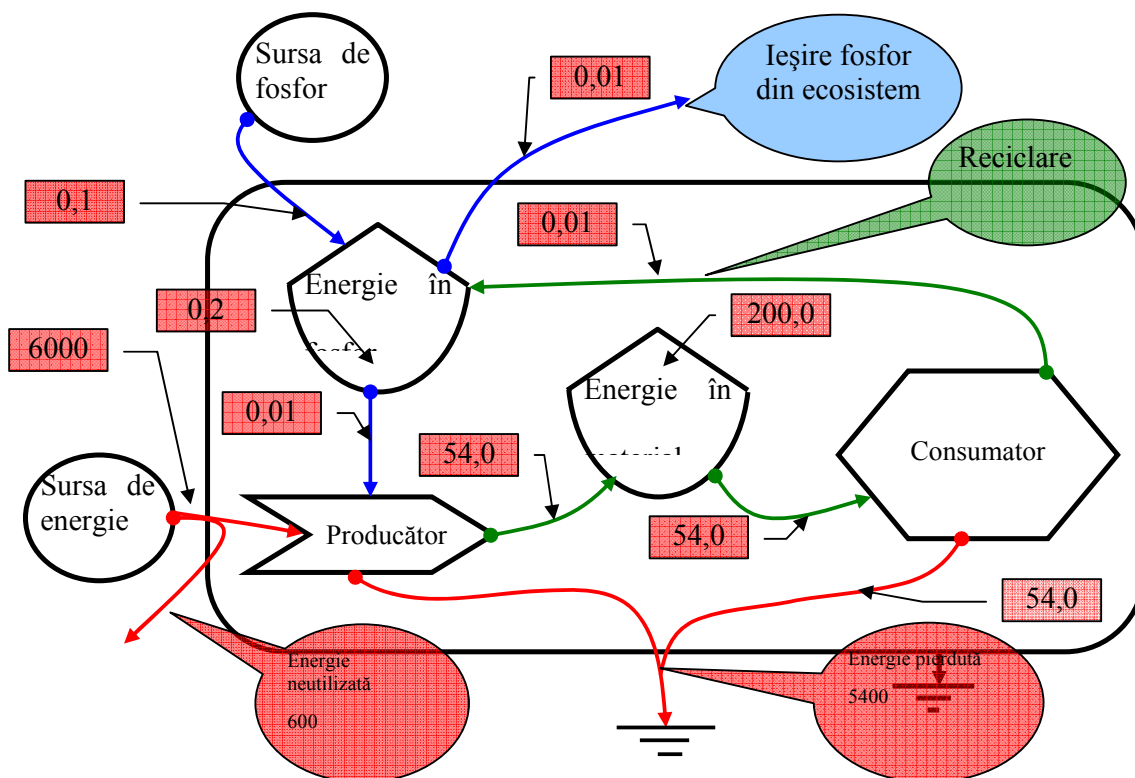


Fig.19. Obținerea **modelului cantitativ** prin plasarea numerelor în diagrama **modelului calitativ** (exemplificare pentru ciclul fosforului într-un ecosistem; valorile sunt exprimate în $[10^6$ Joule/metru pătrat și an])

Cele două diagrame cu numere, **materie**(Fig.18) și **energie** (Fig.19), pot fi combinate și rezultă o diagramă cu tipuri diferite de numere, unele exprimate în $[grame/metru pătrat și an]$ și altele exprimate în $[10^6$ Joule/metru pătrat și an]. Pentru a elimina confuziile într-o astfel de diagramă este util să se noteze unitatea de măsură lângă fiecare număr.

2.1. Modelul dinamic NETPROD

Modelul **NETPROD** ilustrează conceptul de **producție netă**, ca diferență dintre **producția totală** și **consum**.

2.1.1. Exemple

Ilustrarea conceptului de producție netă poate fi realizată într-un sistem cu o sursă permanentă de energie (S), o unitate de producție (P), una de stocare a produsului creat (Q) și una de consum (C) (**Fig.20**).

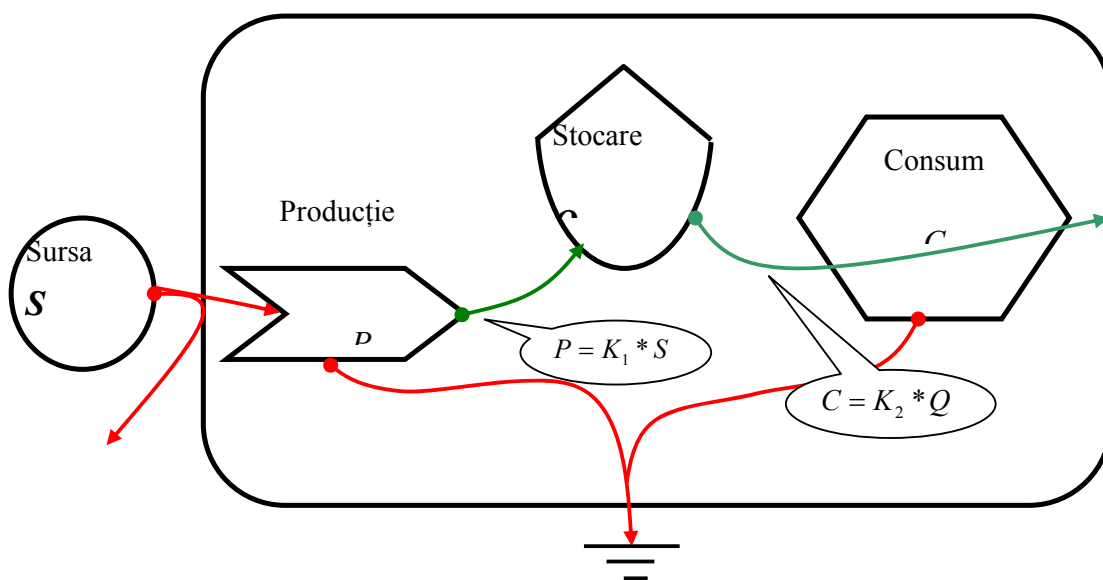


Fig.20. Modelul NETPROD.

În procesul de **fotosinteză** plantele produc materie organică (P) care se acumulează într-un deposit (Q). Din materia organică produsă (P) o parte este consumată (C) de plante și animale. Diferența dintre producția totală (P) și consum (C) constituie producția netă ($P-C$).

Producția P este proporțională cu energia solară care este variabilă în funcție de sezon, iar consumul este proporțional cu cantitatea de materie organică produsă și stocată (Q).

Procese similare de producție se desfășoară în orice ecosisteme: *lacuri, exploatări forestiere, bazine hidrografice* etc. Pentru fiecare din aceste ecosisteme pot fi trasate cu claritate diagramele care reprezintă producția și consumul din a căror diferență rezultă producția netă.

2.1.2. Ecuațiile modelului

Ecuațiile modelului sunt de tip linear și sunt construite pe principiul proporționalității dintre sursă, stocare și consum. Coeficienții de proporționalitate se obțin pe baza măsurărilor și prin calarea unor modele analitice simple, în cazul acesta fiind ales modelul linear.

- Producția: $P = K_1 * S$
- Consumul: $C = K_2 * Q$
- Producția netă pe un interval de timp: $DQ = P - C$
- Cantitatea stocată la un moment dat: $Q = Q + DQ$

2.1.3. Aplicație

Aplicația numerică este construită pe un proces de producție sezonier care ține seama de variația ciclică a energiei solare, energie care este sursa continuă pentru procesul de producție al materiei organice stocate în interiorul sistemului.

Modelul de calcul se poate realiza într-un spreadsheet de tip excel și poate fi ilustrat cu variația parametrilor de intrare și ieșire în funcție de tip:

- $S=f(T)$
- $P=f(T)$
- $C=f(T)$
- $P-C=f(T)$

Datele utilizate sunt:

$K_1 = 0,0225$ coeficientul de transformare al energiei (S) în biomasă(Q);

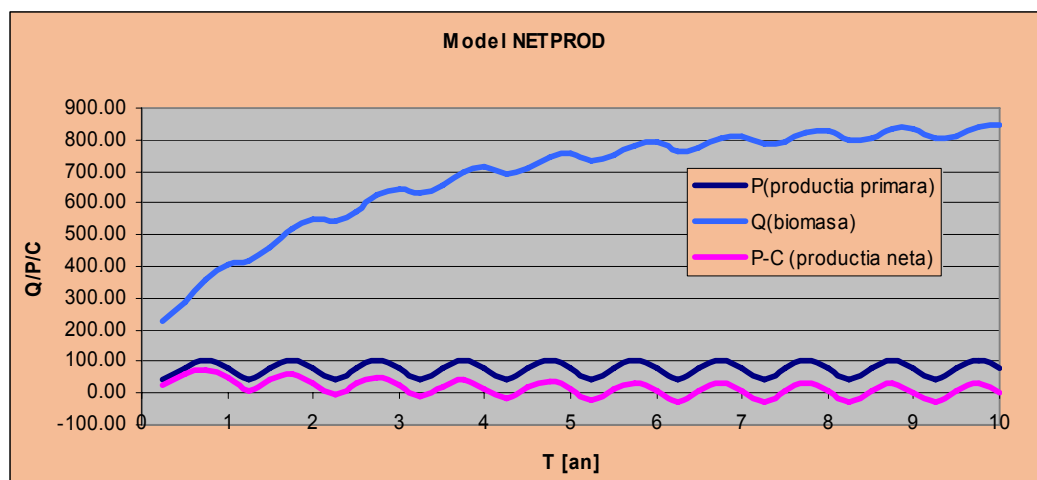
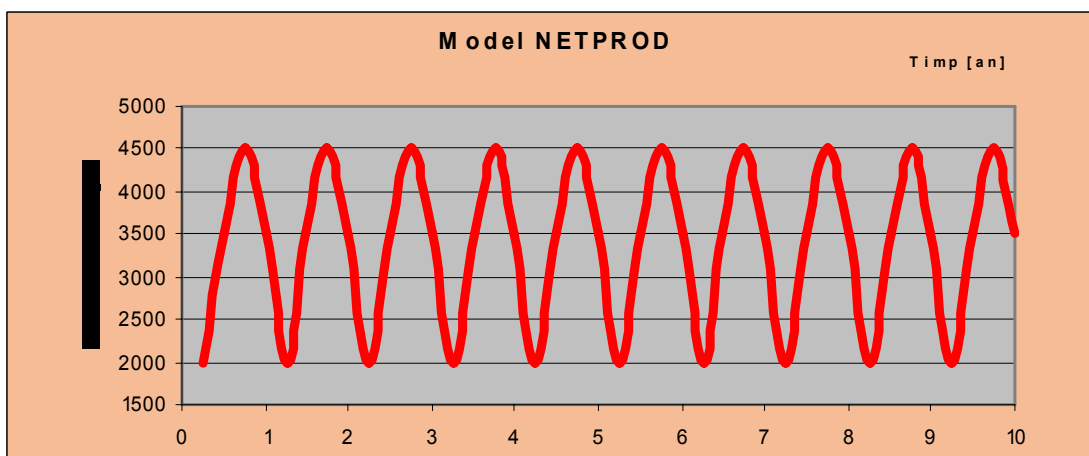
$K_2 = 0,09$ coeficientul de consum (C) al biomasei(C);

$S1 = 2000$, $S2 = 3500$, $S3 = 4500$, $S4 = 3500$; sursele sezoniere de energie furnizată de soare (1-iarna, 2-primavara, 3-vara, 4-toamna);

$Q_{initial} = 200$ -cantitatea inițială de biomasă stocată în ecosistem;

Tabel.1. Table de calcul pentru modelul iynamic NETPROD

T[an]	(N)	S1	S2	S3	S4	(S)	(P)	(C)	(DQ)	Q	P-C
0,00	0									200.00	
0.25	1	2000	0	0	0	2000	45.00	18.00	27.00	227.00	27.00
0.5	2	0	3500	0	0	3500	78.75	20.43	58.32	285.32	58.32
0.75	3	0	0	4500	0	4500	101.25	25.68	75.57	360.89	75.57
1	4	0	0	0	3500	3500	78.75	32.48	46.27	407.16	46.27
1.25	1	2000	0	0	0	2000	45.00	36.64	8.36	415.52	8.36
1.5	2	0	3500	0	0	3500	78.75	37.40	41.35	456.87	41.35
1.75	3	0	0	4500	0	4500	101.25	41.12	60.13	517.00	60.13



2.2. Modelul dinamic RENEW

Modelul RENEW are o unitate autocatalitică bazată pe un flux de energie limitată, din afară, care limitează creșterea cantității de materie organică stocată, la un regim staționar.

2.2.1. Exemple

Modelul poate fi realizat într-un sistem cu o unitate de producție și una de stocare, sistem în care avem o sursă exterioară de energie și pierdere de energie pe două căi (**Fig.21**).

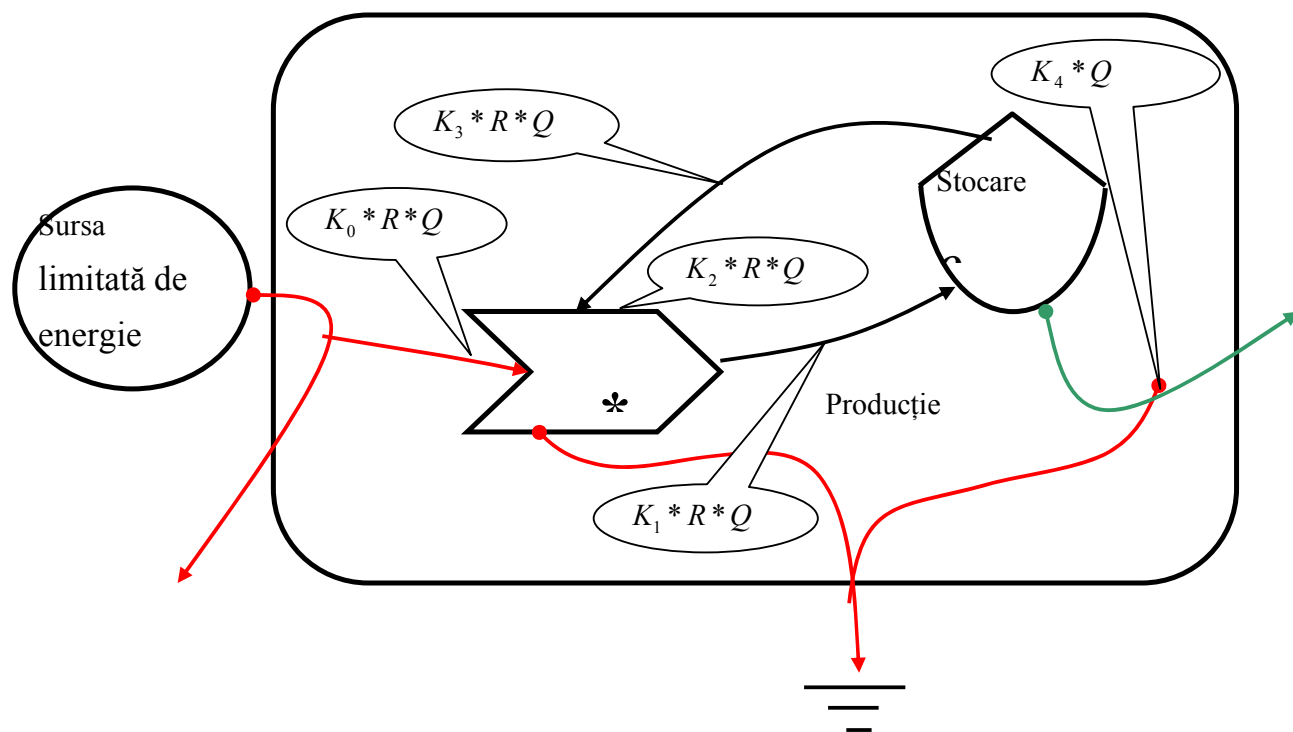


Fig.21. Modelul RENEW.

Un astfel de sistem este o pădure în creștere, care creează biomasă (frunze, trunchiuri, rădăcini, animale, bacterii) pe baza energiei solare regenerabilă dar limitată.

În acest proces de creștere, la un moment dat, atunci când cantitatea de biomasă creată este în echilibru cu cea descompusă, se intră într-o stare de echilibru.

2.2.2. Ecuațiile modelului

Notățiile utilizate pentru scrierea ecuațiilor sunt:

Q: biomasă

J: afluxul de energie în situație de echilibru (stare staționară)

$K_0 * R * Q$: energia utilizată pentru producerea de biomasă ;

R: energia rămasă disponibilă pentru utilizare în continuare

- $R_i = J - K_0 * R_i * Q_{i-1}$ din care rezultă $R_i = \frac{J}{1 + K_0 * Q_{i-1}}$

DQ: schimbarea de biomasă din pădure la fiecare iterație

- $DQ_i = K_3 * R_i * Q_{i-1} - K_4 * Q_{i-1}$

cu care se poate estima cantitatea de biomasă după fiecare iterație:

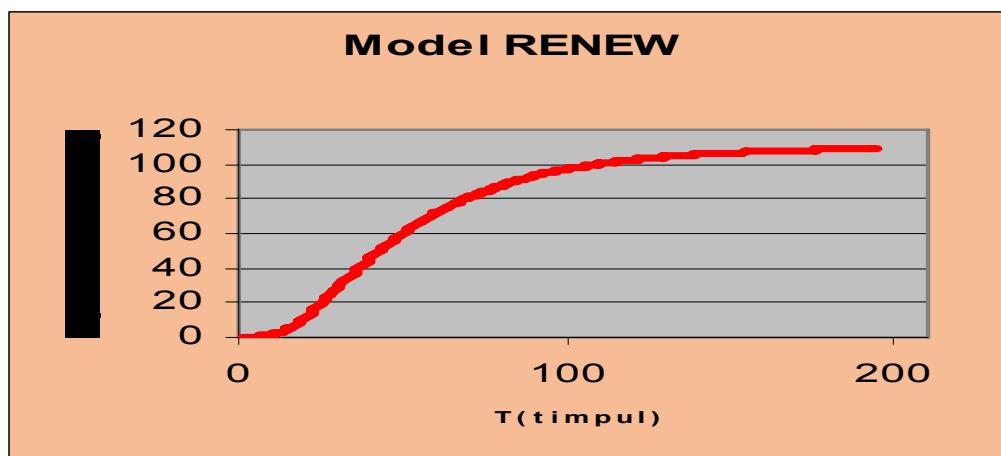
- $Q_i = Q_{i-1} + DQ_i * DT$

DT-modificarea de timp de la o iterație la alta.

2.2.3 Aplicație

Valorile utilizate pentru aplicație sunt sintetizate în tabel, iar pentru reprezentarea grafică a variației biomasei stocate (Q) sunt calculate valorile acestora pentru o perioadă de 200 unități de timp.

J	45	K ₃	0.008	T ₀	1
Q	0.1	K ₄	0.03	Q ₀	1
K ₀	0.1	DT	1		
T	R	DQ	Q		
0	"_"	"_"	0.1		
1	44.55446	0.032644	0.132644		
2	44.41092	0.043147	0.175791		



2.3. Modelul dinamic SLOWRENEW

Modelul **SLOWRENEW** are ca obiectiv evaluarea cantității de biomasă creată în condițiile existenței unei surse de energie limitată și două depozite de stocare interne.

2.3.1. Exemple

Multe procese biologice, geologice și economice au inclus un stocaj intermediar pentru energia provenită dintr-o sursă limitată (Fig.22).

Modelul SLOWRENEW este o bună reprezentare și pentru modul în care se procedează cu resursele energetice în economia mondială, mare consumatoare de energie. În lume există depozite mari de carbune, petrol, minereuri, apă, utilizate pentru realizarea diverselor produse. Reglarea consumului este legată de necesitatea produselor și de resursele disponibile, resurse energetice și materiale.

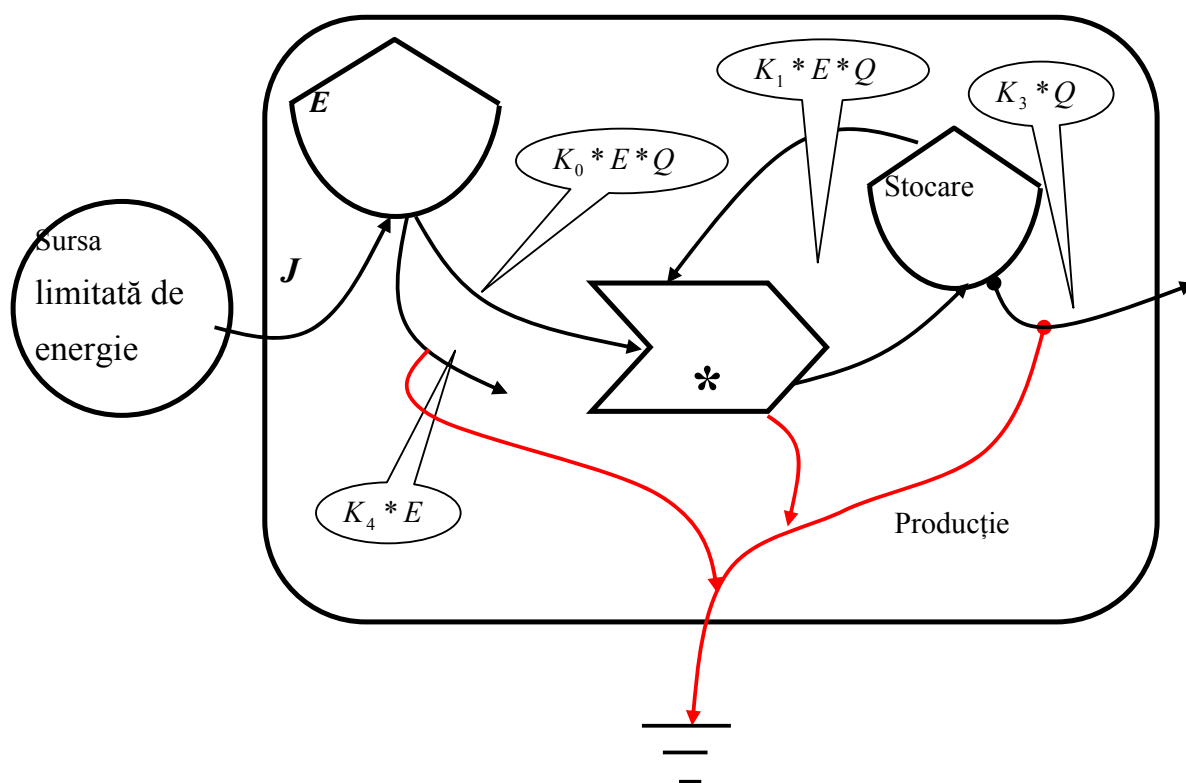


Fig.22. Modelul SLOWRENEW.

2.3.2. Ecuațiile modelului

Afluxul J de energie din exteriorul sistemului este într-o primă etapă stocat într-un rezervor (E) de unde este folosit pentru dezvoltarea unui proces autocatalitic care acumulează produsul într-un al doilea depozit (Q).

Ecuațiile modelului :

- $DE = J - K_4 * E - K_0 * E * Q$
- $DQ = K_1 * E * Q - K_3 * Q$

J : afluxul din exterior

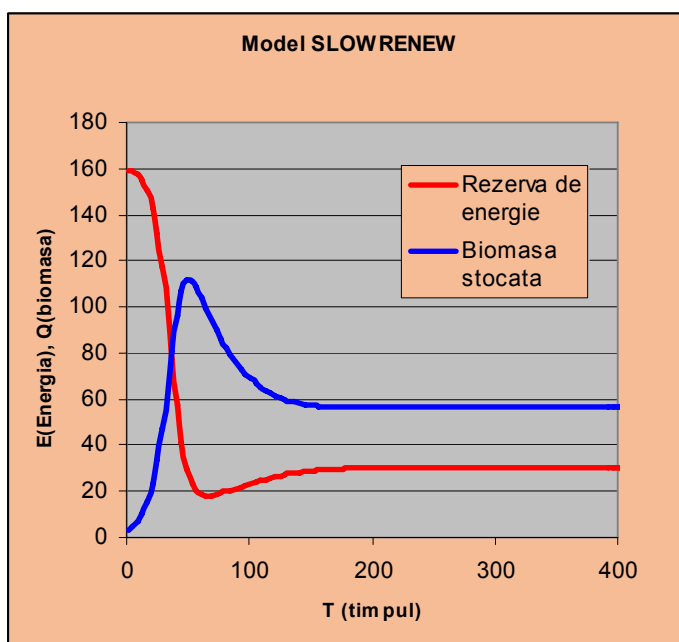
E : primul depozit de energie din sistem

DE : modificarea de energie din depozitul intern:

Q : stocarea de biomasă creată, al doilea depozit din interiorul ecosistemului

2.3.3. Aplicație

J	2	K1	0.001	DT	4
E	159	K3	0.03	Q	3
K0	0.001	K4	0.01		
T	DE	DQ	E	Q	
			159	3	
1	-0.067	0.387	158.732	4.548	
5	-0.30923	0.585473	157.4951	6.889893	



2.4. Modelul dinamic EXCLUS

Model **EXCLUS** conține două cicluri concurente alimentate de aceeași sursă de energie.

2.4.1. Exemple

Modelul poate fi aplicat în orice ecosistem unde există două sau mai multe specii care se hrănesc dintr-o sursă limitată de hrană. Dacă una dintre specii este mai puternică și manancă mai mult decât cealaltă, atunci cea slabă moare din lipsă de hrană.

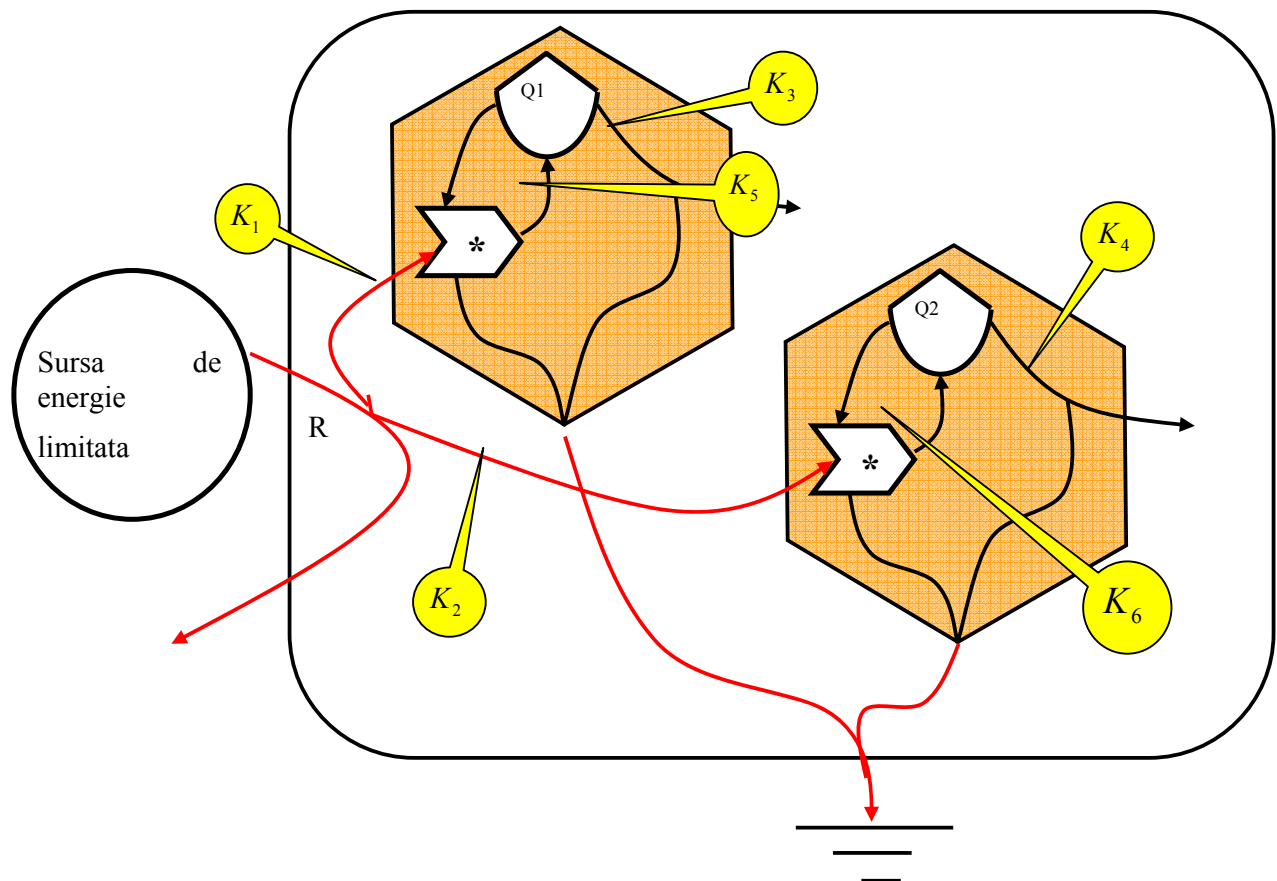


Fig.23. Modelul EXCLUS

2.4.2. Ecuațiile modelului

Ecuațiile modelului descriu cele trei componente principale:

- Energia disponibilă: $R = I - K_1 * R * Q_1 - K_2 * R * Q_2$

I - energia disponibilă inițial;

- Creșterea populației Q1 într-un interval de timp DT:

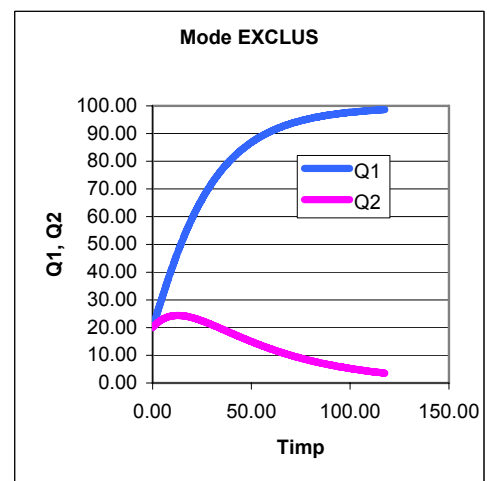
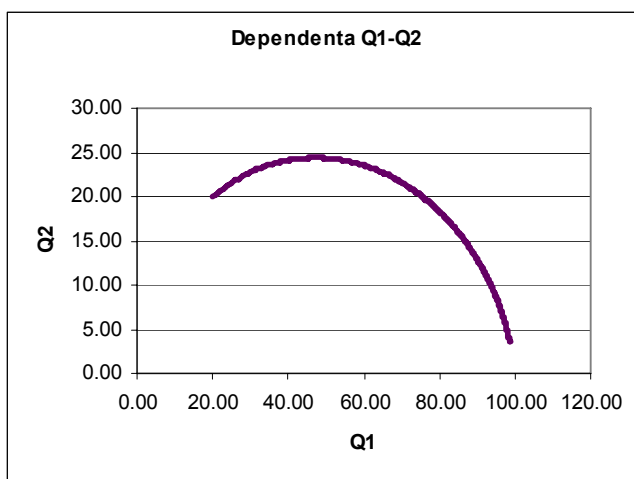
$$DQ_1 = K_5 * R * Q_1 - K_3 * Q_1$$

- Creșterea populației Q2 într-un interval de timp DT:

$$DQ_2 = K_6 * R * Q_2 - K_4 * Q_2$$

2.4.3. Aplicație numerică

I	5Q1		20		
DT	0.3Q2		20		
K1	0.08	K2	0.01		
K3	0.05	K4	0.05		
K5	9.000001E-02	K6	0.05		
T	R	DQ1	DQ2	Q1	Q2
0.00				20.00	20.00
0.30	1.79	2.21	0.79	20.66	20.24
0.60	1.75	2.22	0.76	21.33	20.46
0.90	1.72	2.23	0.73	22.00	20.68



2.5. Modelul dinamic INTERACT

Modelul **INTERACT** reprezintă competiția dintre două populații **Q1** și **Q2**. Fiecare populație are propriul ciclu **autocatalitic**, și au la dispoziție o sursă nelimitată de hrană **E**, constantă (**Fig.24**).

Dinamica modelului este posibilă în două variante:

- **lipsa interacțiunii** care determină o creșterea a ambelor populații până la o valoare maximă la care se stabilizează.
- **interacțiunea negativă (concurența)** una din populații se dezvoltă și ajunge în regim stabilizat în timp ce cealaltă populație dispare la un moment dat din lipsa de resurse, consumate de populația concurentă.

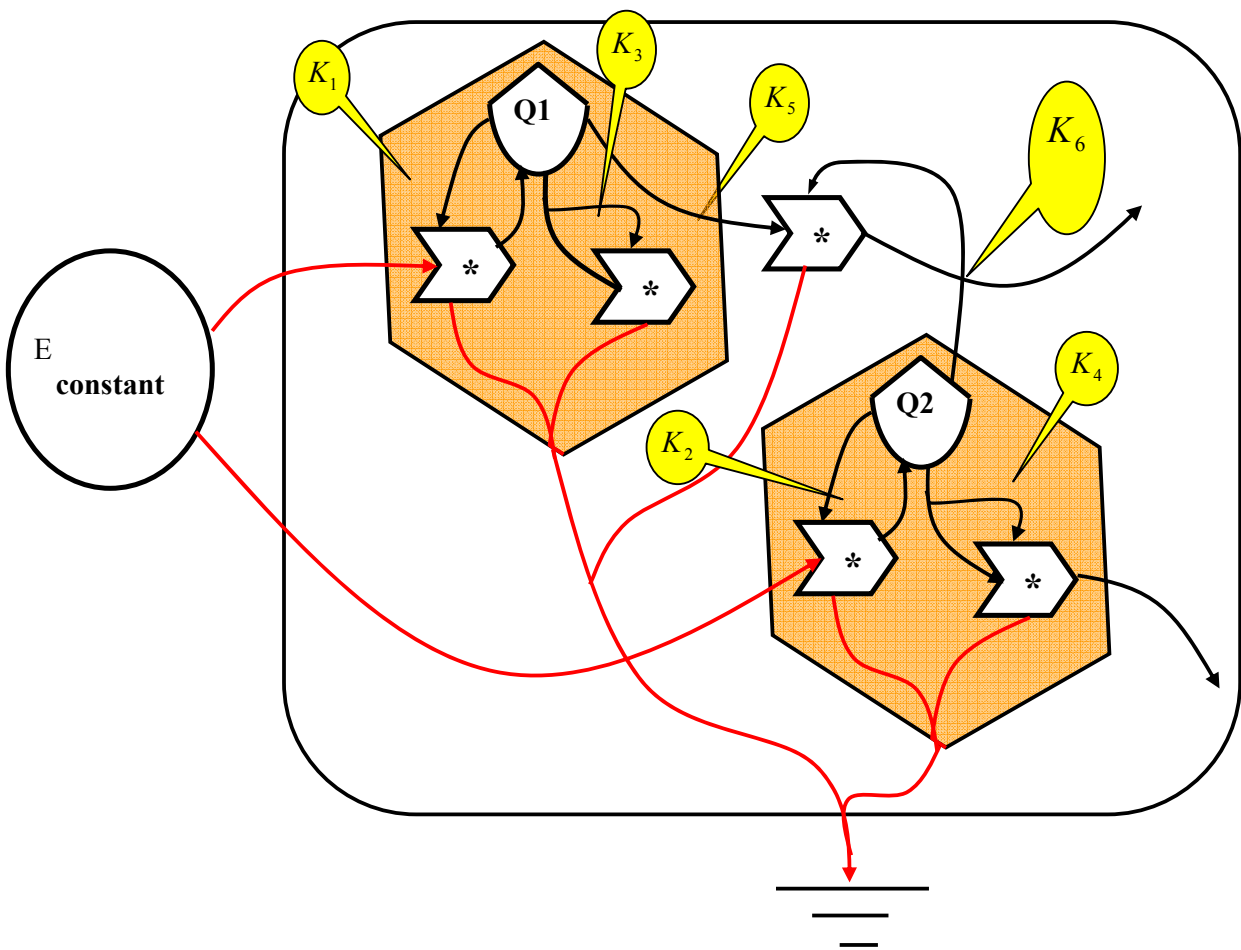


Fig.24. Modelul INTERACT

2.5.1. Exemple

Un exemplu de competiție este cea dintre două specii de cărabuși care se hrănesc din aceeași făină pusă într-un borcan. Dacă o singură specie este prezentă atunci aceasta se dezvoltă numeric până ajunge la un regim de stabilizare. Dacă în același borcan sunt puse două specii, una dintre specii o distruge pe cealaltă.

Există situații speciale când o specie acționează direct pentru eliminarea speciei concurente la utilizarea aceleiași surse de hrană. Sunt plante care secretă o substanță toxică, ce inhibă dezvoltarea rădăcinilor speciei concurente.

2.5.2. Ecuatiile modelului

Ecuatiile modelului INTERACT exprimă a doua variantă, a interacțiunii **concurențiale**:

- Energia disponibilă: E -constantă;
- Creșterea populației 1 într-un interval de timp DT :

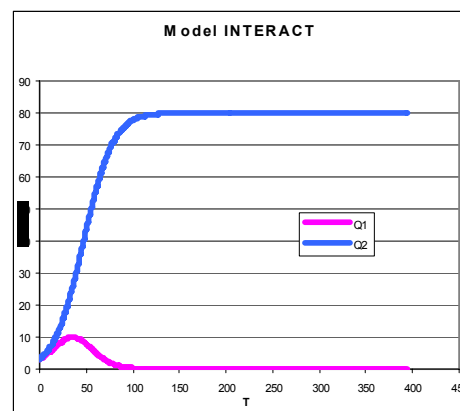
$$DQ_1 = K_1 * E * Q_1 - K_3 * Q_1 * Q_1 - K_5 * Q_1 * Q_2$$

- Creșterea populației 2 într-un interval de timp DT :

$$DQ_2 = K_2 * E * Q_2 - K_4 * Q_2 * Q_2 - K_6 * Q_1 * Q_2$$

2.5.3. Aplicație

E	1	K1	0.07	K4	0.001
Q1	3	K2	0.08	K5	0.002
Q2	3	K3	0.002	K6	0.001
		DT	1		
T	DQ1	DQ2	Q1	Q2	
0.000			3.000	3.000	
1.000	0.174	0.222	3.174	3.222	
2.000	0.182	0.237	3.356	3.459	
3.000	0.189	0.253	3.545	3.712	



2.6. Modelul dinamic COOP

Modelul COOP este construit pentru cooperare mutuală între cele două populații care se dezvoltă pe aceeași sursă de hrană, limitată dar regenerabilă.

Modelul conține două cicluri autocatalitice care spre deosebire de modelul EXCLUS, în care cele două sunt concurente, aici cooperează pentru o coexistență utilizând pentru creșterea fiecărui component produsele create de celălalt (**Fig.25**).

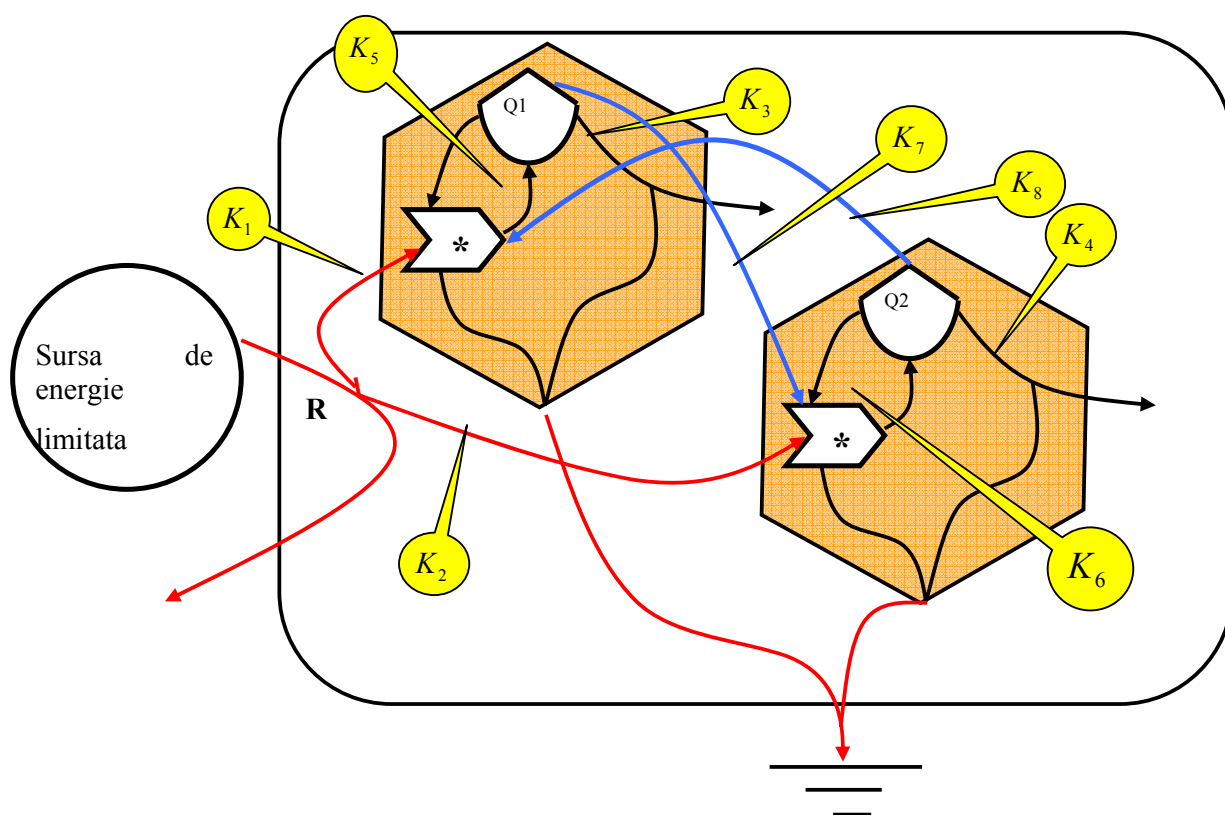


Fig.25. Modelul COOP

2.6.1. Exemple

Exemplul clasic de cooperare în procesele ecologice este *simbioza*:

- Insectele polenizează florile, iar florile produc polenul cu care se hrănesc insectele;
- Veverițele plantează ghindă, din ghindă cresc copacii care produc ghindă pentru următoarele generații de veverițe;
- Comerțul între țări este un exemplu de cooperare.

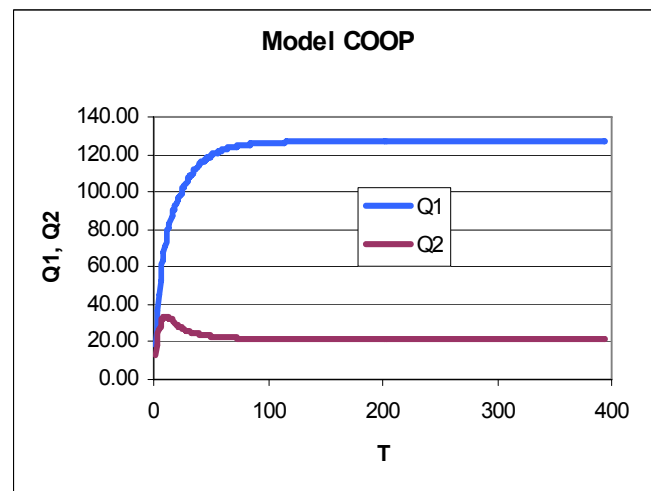
2.6.2. Ecuațiile modelului

Ecuațiile modelului sunt:

- Sursa inițială de energie: I
- Energia disponibilă după un pas DT: $R = \frac{I}{1 + K_1 * Q_1 * Q_2 + K_2 * Q_1 * Q_2}$
- Creștere populație 1: $DQ_1 = K_5 * R * Q_1 * Q_2 - K_3 * Q_1 - K_7 * R * Q_1 * Q_2$
- Creștere populație 2: $DQ_2 = K_6 * R * Q_1 * Q_2 - K_4 * Q_2 - K_8 * R * Q_1 * Q_2$

2.6.3. Aplicație

I	10	K1	0.08	K5	0.09	
Q1	8	K2	0.04	K6	0.05	
Q2	8	K3	0.05	K7	0.002	
DT	1	K4	0.05	K8	0.002	
T	R 1	R2	DQ1	DQ2	Q1	Q2
1					8	8
2	1.79	1.79	9.77	5.19	17.77	13.19
3	0.54	0.54	10.11	5.24	27.88	18.44
4	0.25	0.25	9.23	4.52	37.11	22.96



2.7. Modelul dinamic **DESTRUCT**

Modelul **DESTRUCT** ilustrează o proprietate importantă a unui ecosistem care se manifestă atunci când un produs (A) este în exces.

Este vorba de utilizarea unei cantități de energie care conduce la distrugerea produsului A și în felul acesta este pus din nou la dispoziția sistemului materialul M utilizat la crearea produsului A.

Acest proces de dispersie și deterioare reduce cantitatea de produs (ordinea), reciclează materia (dezordine).

Cantitatea de energie necesară deprecierii produsului A, utilizată în sensul creșterii dezordinii în sistem, este mult mai mică decât cea necesară creerii produsului A, adică creșterea ordinii din sistem.

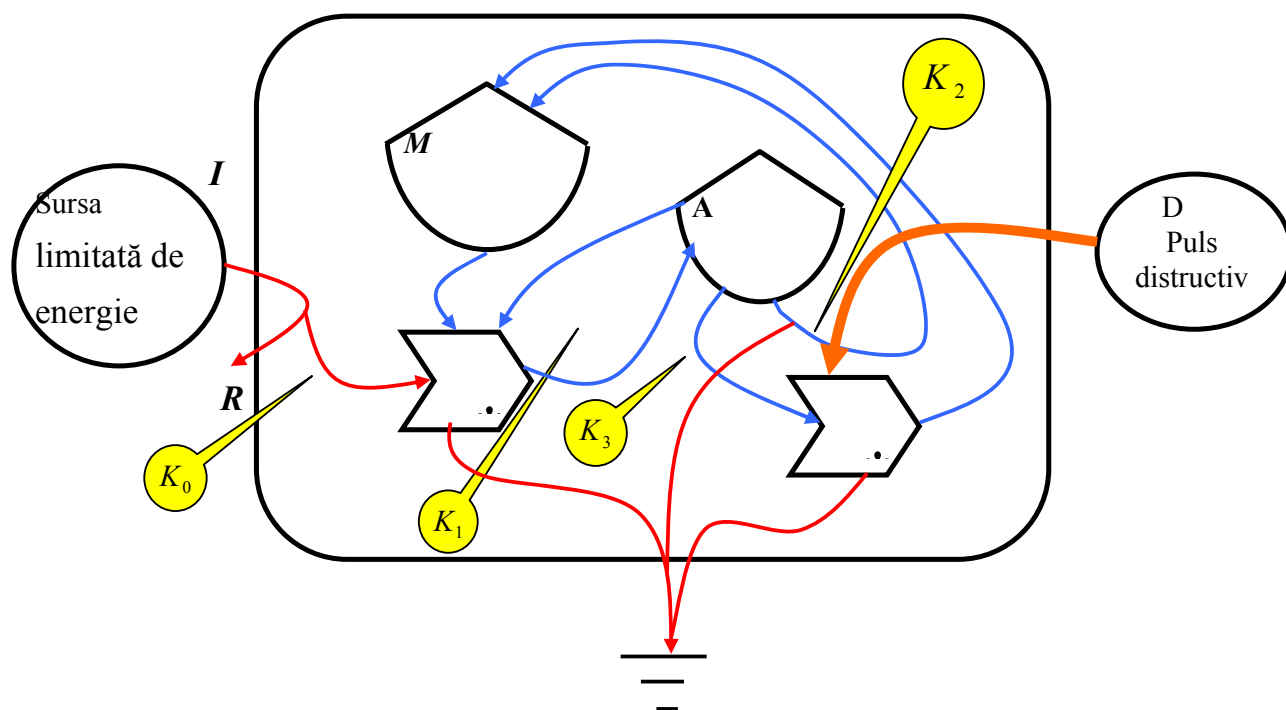


Fig.26. Modelul **DESTRUCT.**

2.7.1. Exemple

Uraganele și incendiile care distrug copacii din pădure și reciclează materia pentru alt ciclu de creștere.

În ecosisteme, bolile distrug populațiile care sunt prost adaptate condițiilor noi create.

2.7.2. Ecuațiile modelului

Ecuațiile modelului exprimă cantitativ energia valabilă, materia disponibilă și cantitatea de produs creată:

- Energia valabilă: $R = \frac{I}{1 + K_0 * M}$
- Materialul disponibil: $M = M_t - F * A$;

F -fracția din materialul total disponibil utilizat pentru crearea produsului A

- Cantitatea de produs A creată:

$$DA = K_1 * R * M * A - K_2 * A - X * K_3 * A * D$$

X -energie utilizată pentru distrugerea produsului A și eliberarea materialului M.

D - energia disponibilă pentru distrugerea produsului A

2.7.3. Aplicație

I	4	A	1	AO	3
F	0.2	DT	0.5	K1	0.001
D	1	TO	1	K2	0.01
MT	100	MO	3	K3	0.02
				KO	0.0009
T	X	R	DA	A	M
1				3	3
1.5	0	3.96786	0.035711	3.017855	99.4
2	0	3.149662	0.944819	3.490265	99.39643

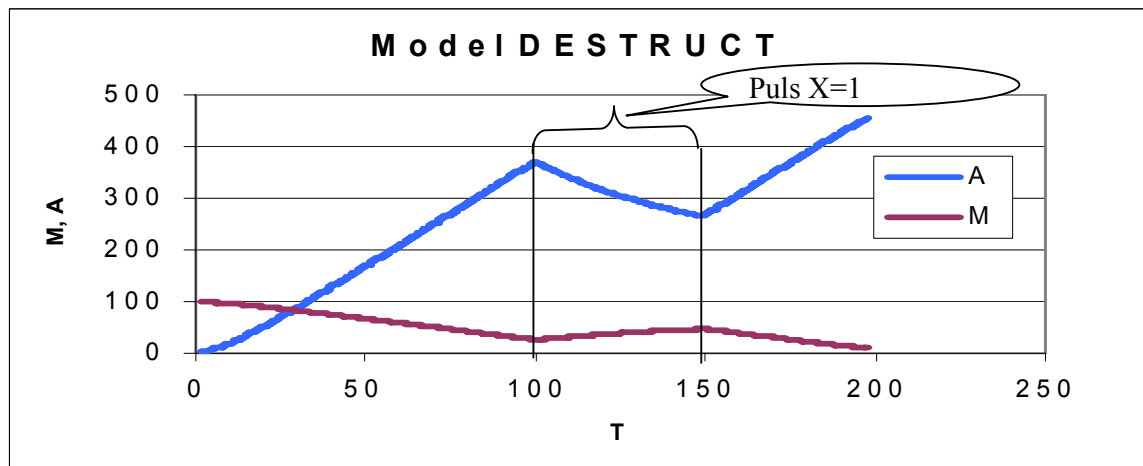
2.2.1.

2.2.2.

2.2.3.

2.2.3.1.

2.2.3.2



Să ne reamintim

2.2.4.

2.2.5.



2.3. Rezumat

Modelele dinamice cantitative exprima prin modele analitice simple relatiile „statistice” dintre componentele procesele ecologice si sunt exemplificate prin:

- Modelul NETPROD
- Modelul RENEW
- Modelul SLOWRENEW
- Modelul EXCLUS
- Modelul INTERACT
- Modelul COOP
- Modelul DESTRUCT



2.4. Test de autoevaluare a cunoștințelor

1) Cum se calculeaza PRODUCTIA NETApe un interval de timp?

- CANTITATEA STOCATA -CONSUMUL
- CONSUMUL + PRODUCTIA
- PRODUCTIA - CONSUMUL

2) Ce contine modelul EXCLUS?

- Doua surse de energie si doi consumatori
- Doi consumatori si o sursa de energie
- Trei consumatori si trei surse de energie

3) Ce tip de surse de energie contine modelul INTERACT ?

- LIMITATA SI CONSTATA
- NELIMITATA SI CONSTANTA
- LIMITATA SI VARIABILA



2.5. Bibliografie recomandată

- Odum, Howard T., Odum, Elisabeth C., (2000), Modeling for all Scales, An Introduction to system Simulation, Academic Press, London.
- Ford , David, E., ,(200), Scientific Method for Ecological Research, Cambridge, University Press.

Capitolul 3

MODELE STATISTICE



3.1. Introducere

În cadrul acestui capitol sunt prezentate modelele statistice cantitative elementare utilizate pentru modelarea corelațiilor dintre variabilele cantitative și calitative ale ecosistemelor/proceselor ecologice

Durata medie de parcurgere a acestui capitol este de aproximativ 10 ore.



3.2. Conținut

Cursul 3. 3. Modele cantitative statistice

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.

3.1. Cuantificarea intensității corelațiilor

Utilizarea termenului corelație în ecologie are o semnificație mult mai largă decât cea matematică. În sens statistic, corelația reprezintă un anumit grad de legătură evaluat prin diferite tehnici matematice, fiecare caracter fiind tratat ca o variabilă aleatoare. Ansamblul caracterelor studiate formează o variabilă aleatoare cu mai multe componente iar ipoteza normalității acestei variabile în spațiul multidimensional este la baza tehnicilor de evaluare a intensității corelației. În ecologie o mare parte a cercetării este consacrată identificării relațiilor dintre caracteristicile măsurabile.

Natura corelațiilor în ecologie este determinată de structura fizico-chimică și biologică a “obiectelor” de studiu care este constituită dintr-un ansamblu de variabile care formează biotopul și biocenoza. De aici rezultă natura *substantială* a corelațiilor care se realizează pe baza compoziției fizice, chimice, pe baza speciilor sau a calității fizico-chimice a câmpurilor terestre (magnetic, gravimetric etc).

Ecologia se ocupă, de asemenea, cu analiza proceselor ce se desfășoară în *timp* și *spațiu* ; în acest fel se completează spectrul naturii corelațiilor ecologice cu trei componente principale:

- corelații substanțiale;
- corelații temporale.
- corelații spațio-temporale sau topo-probabiliste;

Cercetarea corelațiilor poate fi realizată cu instrumente diferite în funcție de dimensiunea și natura fenomenelor studiate. În literatura există încă o mare confuzie în terminologia utilizată pentru instrumentele cu ajutorul cărora evaluăm intensitatea legăturilor/corelațiilor dintre caracteristicile ecologice.

Vom adopta în continuare pentru instrumentele de cuantificare a intensității *corelațiilor substanțiale* dintre două variabile următoarele categorii:

- *coeficient de corelație* utilizat pentru variabile cantitative (numerice) și adaptabil, în anumite circumstanțe, pentru variabile calitative (alfanumerice);
- *coeficient de corelație a rangurilor* utilizați pentru variabile ordonabile (numerice/alfanumerice);
- *coeficient de asociere* utilizați pentru variabile calitative (alfanumerice)

Cuantificarea *corelațiilor temporale* se bazează pe o formalizare particulară a serilor de timp și se exprimă prin:

- coeficienți de autocorelație
- coeficienți de intercorelație

Cuantificarea *corelațiilor spațio-temporale* presupune o prelucrare complexă și un volum mare de date cu o structură spațială și temporală complexă. Metodologia de evaluare a acestor corelații este de o deosebită complexitate constituind o direcție specială (Scrădeanu, D., 2003, Geostatistică aplicată).

3.1.1. Coeficienții 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.

a) Raportul de corelație

Raportul de corelație permite evaluarea intensității și sensului corelației dintre două variabile geologice (y, x) independent de modelul de corelație. Raportul de corelație realizează această evaluare prin intermediul gradului de împrăștiere al valorilor y_i măsurate în jurul mediilor condiționate $\bar{y}_{xi}$.

Analizând intensitatea dependenței variabilei y (rezultative) în raport de variabila x (factorială), dispersia acesteia poate fi exprimată sub forma:

$$s_y^2 = s_{y(x)}^2 + s_{y0}^2 \quad (\text{III.169})$$

în care

s_y^2 - dispersia totală a variabilei y în raport cu toți factorii cunoscuți sau necunoscuți;

$s_{y(x)}^2$ - dispersia condiționată a variabilei y în raport cu variabila x ;

s_{y0}^2 - dispersia reziduală a variabilei y în raport cu celelalte variabile care-i condiționează variabilitatea și care nu sunt specificate în model.

Separarea dispersiei totale în cele două componente necesită gruparea datelor într-un tabel de corelație a cărei configurație este condiționată de sensul corelației. Pentru evaluarea

gradului de dependentă al variabilei y în raport cu variabila x , tabelul de corelație (Tabelul III.19) conține:

m_{yxi} - mediile variabilei y pentru fiecare interval x_i ;

n_{xi} - frecvențele marginale ale valorilor y_i pentru fiecare interval x_i ;

în timp ce tabelul de corelație al variabilei x în raport cu y ($x = f(y)$) (Tabelul III.20):

m_{xyi} - mediile variabilei x pentru fiecare interval y_i ;

n_{yi} - frecvențele marginale ale variabilei x_i pentru fiecare interval y_i .

Tabelul III.19 Corelație $y = f(x)$

x	y - var. dependentă	n_{xi}	m_{yxi}
x_1	$y_{11}, y_{12}, \dots, y_{1n1}$	n_{x1}	m_{yx1}
x_2	$y_{21}, y_{22}, \dots, y_{2n2}$	n_{x2}	m_{yx2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	$y_{k1}, y_{k2}, \dots, y_{knk}$	n_{xk}	m_{yxk}

Tabelul III.20 Corelație $x = f(y)$

y	x - var. dependentă	n_{yi}	m_{xyi}
y_1	$x_{11}, x_{12}, \dots, x_{1n1}$	n_{y1}	m_{xy1}
y_2	$x_{21}, x_{22}, \dots, x_{2n2}$	n_{y2}	m_{xy2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_k	$x_{k1}, x_{k2}, \dots, x_{knk}$	n_{yk}	m_{xyk}

Dispersiile s_y^2 și $s_{y(x)}^2$ se evaluează cu relațiile:

$$s_y^2 = \frac{\sum_{i=1}^k (y_i - m_y)^2}{k-1} \quad (\text{III.170})$$

$$s_{y(x)}^2 = \frac{\sum_{i=1}^k n_{xi} (m_{yxi} - m_y)^2}{k-1} \quad (\text{III.171})$$

pentru analiza intensității corelației $y = f(x)$, iar dispersiile s_x^2 și $s_{x(y)}^2$ cu relațiile:

$$s_x^2 = \frac{\sum_{i=1}^k (x_i - m_x)^2}{k-1} \quad (\text{III.172})$$

$$s_{x(y)}^2 = \frac{\sum_{i=1}^k n_{yi} (m_{xyi} - m_x)^2}{k-1} \quad (\text{III.173})$$

pentru analiza intensității corelației $x = f(y)$.

Intensitatea corelației dintre cele două variabile se măsoară cu ajutorul raportului dintre dispersia ($s_{y(x)}$ sau $s_{x(y)}$) și dispersia totală ($s^2 y$ sau $s^2 x$). Pentru exprimarea cantitativă a acestei corelații se definește raportul de corelație cu:

$$\eta_{y(x)} = \sqrt{\frac{s_{y(x)}^2}{s_y^2}} \quad (\text{III.174})$$

$$\eta_{x(y)} = \sqrt{\frac{s_{x(y)}^2}{s_x^2}} \quad (\text{III.175})$$

Valoarea maximă a raportului de corelație este 1 și exprimă o corelație maximă între cele două variabile, iar lipsa de corelație dintre cele două variabile corespunde valorii zero, valoarea minimă a raportului de corelație.

În analiza corelației dintre două variabile geologice, nu întotdeauna este evident care din variabile este rezultativă și care este factorială, motiv pentru care este necesar să se determine valoarea raportului de corelație în ambele variante (III.174) și (III.175). Analiza ambelor valori poate conduce la următoarele variante extreme de interpretare:

a) variabila y este dependentă de x iar x este independentă;

$$\eta_{y(x)} = 1 \text{ și } \eta_{x(y)} = 0 \quad (\text{III.176})$$

b) variabila x este dependentă de y iar y este independentă;

$$\eta_{y(x)} = 0 \text{ și } \eta_{x(y)} = 1 \quad (\text{III.177})$$

c) variabilele x și y sunt independente;

$$\eta_{y(x)} = 0 \text{ și } \eta_{x(y)} = 0 \quad (\text{III.178})$$

d) variabilele x și y se intercondiționează sau ambele sunt condiționate de o a treia variabilă neidentificată:

$$\eta_{y(x)} = 1 \text{ și } \eta_{x(y)} = 1 \quad (\text{III.179})$$

În practica analizei corelațiilor dintre variabilele geologice, raportul de corelație ia valori cuprinse între 0 și 1 iar semnificația lor statistică se poate testa cu ajutorul factorului F pe baza inegalității:

$$F_{\text{exp}} = \frac{\eta^2}{1-\eta^2} \times \frac{n-k}{k-1} > F(\alpha, \nu_1; \nu_2) \quad (\text{III.180})$$

în care $\nu_1 = k-1$, $\nu_2 = n-k$ (n = perechi de valori, k = număr de intervale de grupare, α = nivelul de semnificație al testului).

Verificarea inegalității (III.180) indică o valoare semnificativă statistic a raportului de corelație, deci existența unei corelații între variabilele analizate.

b) Coeficientul corelației lineare

Coeficientul corelației lineare este cel mai des întâlnit în cercetarea geologică 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:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^n (x_i - m_x)^2 \sum_{i=1}^n (y_i - m_y)^2}} \quad (\text{III.181})$$

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

Abaterea 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, logaritm 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 lineară 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. 58a), fie descendentă ($r_{xy} = -1$; Fig. 58b) - care indică o dependență lineară perfectă între cele două variabile. O astfel de situație este foarte rar întâlnită în studiul unor relații functionale între două caracteristici geologice;

b) punctele sunt dispersate aleator, norul de puncte neavând nici o orientare preferențială (Fig. 58c). 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. 58d).

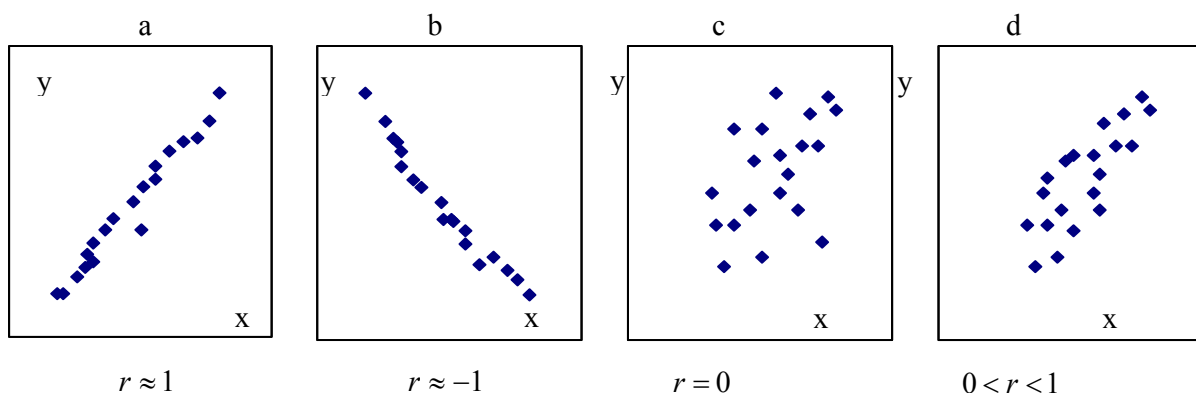


Fig. 58 Semnificația geometrică a coeficientului Pearson

O analiză mai detaliată a coeficientului de corelație lineară 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.

c) Coeficientul cosinus θ

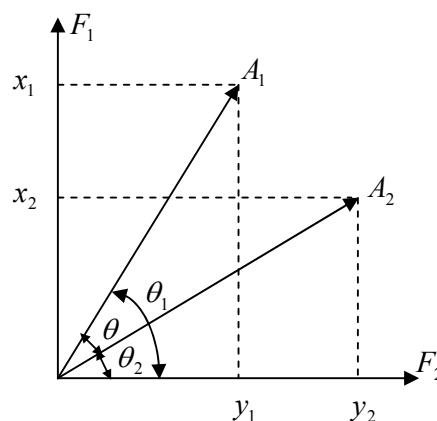


Fig. 59 Coeficientul cosinus θ pentru un spațiu bidimensional

Coeficientul cosinus θ este o măsură a distanței unghiulare, utilizat pentru estimarea similarității între obiecte geologice de studiu (ex.: aflorimente, zăcămintele, bazine de sedimentare, acvifere etc), reprezentate în spațiul variabilelor măsurabile (ex.: compoziție chimică, compoziție granulometrică, parametri hidrogeologici etc). Estimarea lui implică ortogonalitatea axelor sistemului de referință, motiv pentru care este preferat în analiza factorială Q - MOD.

Într-un spațiu bidimensional definirea coeficientului cosinus θ se bazează pe relațiile trigonometrice elementare ale cosinusului unghiului unei diferențe de unghiuri (Fig.59):

$$\cos \theta_{A_1 A_2} = \cos(\theta_1 - \theta_2) = \frac{x_1 \times y_1 + x_2 \times y_2}{\sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)}} \quad (\text{III.183})$$

Generalizând pentru n dimensiuni (n factori independenți $F_1, F_2, \dots, F_n$, spre exemplu n aflorimente probate în cazul analizei Q-MOD) se obține formula:

$$\cos \theta_{A_1 A_2} = \frac{\sum_{i=1}^k x_i y_i}{\sqrt{\sum_{i=1}^k x_i^2 \sum_{i=1}^k y_i^2}} \quad (\text{III.183})$$

Acest coeficient de corelație indică o similaritate completă între două obiecte geologice A_1 și A_2 pentru $\cos \theta = 1$ și o disimilaritate totală pentru $\cos \theta = 0$ (corespunzător unui unghi $\theta = 90^\circ$ echivalent cu ortogonalitatea vectorilor de poziție).

d) Coeficientul distanței taxonomice

Ca măsură a similarității între două obiecte geologice, coeficientul distanței taxonomice își are originea în modelul geometric al distanței euclidiene între două puncte A și B într-un spațiu n -dimensional. Distanța taxonomică între cele două obiecte geologice este invers proporțională cu similaritatea, n fiind numărul de caracteristici proprii celor două obiecte geologice studiate.

În cazul distanței taxonomice dintre două eșantioane A și B reprezentate prin două caracteristici x_1 și x_2 (Fig.60) formula de calcul este:

$$D_{AB} = \sqrt{(x_{1A} - x_{1B})^2 + (x_{2A} - x_{2B})^2} \quad (\text{III.184})$$

în care:

x_{1A} - caracteristica x_1 determinată în eşantionul A (ex.: conţinutul în zinc);

x_{1B} - caracteristica x_1 determinată în eşantionul B;

x_{2A} - caracteristica x_2 determinată în eşantionul A (exemplu: conţinutul în plumb);

x_{2B} - caracteristica x_2 determinată în eşantionul B.

Dacă pentru cele două obiecte geologice (A şi B) se determină mai multe caracteristici ($x_1, x_2, \dots, x_n$) se utilizează o generalizare a distanţei taxonomice:

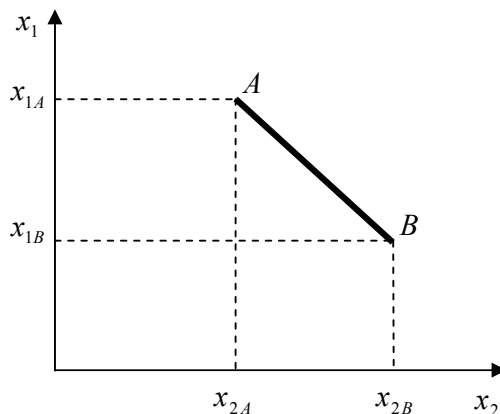


Fig. 60 Distanța taxonomică în spațiu bidimensional

$$D_{AB} = \sqrt{\sum_{i=1}^n (X_{iA} - X_{iB})^2} \quad (\text{III.185})$$

Creșterea numărului de caracteristici utilizate reduce posibilitatea interpretării valorii distanței taxometrice în comparație cu a altor coeficienți de corelație datorită diversității unităților de măsură și a amplitudinilor de selecție. Eliminarea acestor inconveniente se realizează prin standardizarea valorilor caracteristicilor măsurate, normarea lor pe intervalul $[0,1]$ și definirea coeficientului distanței taxonomice:

$$d_{AB} = 1 - \sqrt{\frac{1}{n} \sum_{i=1}^n (XS_{iA} - XS_{iB})^2} \quad (\text{III.186})$$

în care:

XS_{iA} - valoarea standardizată și normată a caracteristicii "i" din eşantionul A;

XS_{iB} - valoarea standardizată și normată a caracteristicii "i" din eşantionul B.

În aceste condiții, valorile extreme ale coeficientului de distanță sunt: zero, când cele două esantioane sunt identice, deci similaritatea este maximă și unu, când cele două eşantioane A și B sunt total diferite.

e) Coeficientul corelației binare

Coeficientul corelației binare (r_D) a fost propus de Derec, Sarcia și Troly (1964) pentru cercetări metalogenetice și este definit prin relația:

$$r_D = \frac{ne_{ab} - ab}{\sqrt{ab(n-a)(n-b)}} \quad (\text{III.188})$$

în care:

n - numărul total de cazuri analizate (Fig. 61);

a - numărul de cazuri analizate care prezintă caracteristica A;

b - numărul de cazuri analizate care prezintă caracteristica B;

e_{ab} - numărul de cazuri analizate care prezintă ambele caracteristici A și B, a căror corelație se analizează.

Coeficientul de corelație binară este o măsură a intensității legăturii între caracteristicile A și B. Cu cât coeficientul r_D este mai mare (valori pozitive) legătura este mai puternică. Valorile

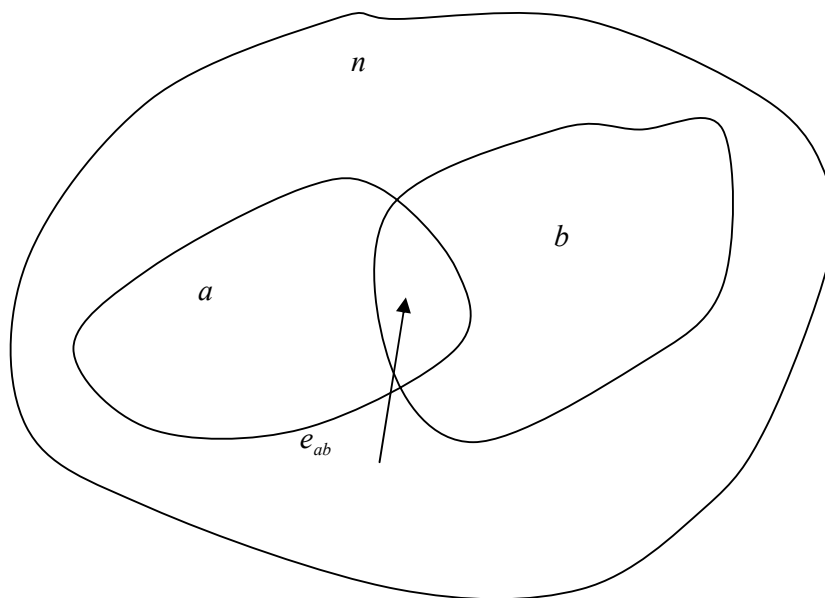


Fig. 61 Relația dintre elementele coeficientului de corelație binară

negative indică o "respingere" a caracteristicilor, iar valoarea nulă o independență totală.

Interpretarea naturalistă a valorilor lui r_D permite ierarhizarea corelațiilor într-un sistem multivariat pe baza coeficienților corelației binare calculați pentru toate perechile de caracteristici măsurabile. Asamblate într-o matrice de similaritate, toate valorile coeficientului de corelație pot forma o imagine sintetică a ierarhiilor corelaționale din sistemul studiat. În tabelul III.21 este prezentată configurația unei astfel de matrici ce va constitui obiectul unor prelucrări ulterioare în scopul factorizării corelaționale.

Tabelul III.21 Matricea coeficienților r_D pentru mineralele caracteristice ale pegmatitelor cu beril din Madagascar și Mozambic (după P. Lafitte, 1972)

	1	2	3	4	5	6	7	8	9	10	11
1	1	0.31	0.34	-0.16	0.18	0.31	0.17	0.18	-0.06	-0.57	0.26
2	0.31	1	-0.17	-0.46	0.13	0.1	0.05	0.13	0.13	-0.55	-0.19
3	0.34	-0.17	1	-0.16	0.18	-0.28	-0.31	-0.06	0.18	0.01	0.26
4	-0.16	-0.46	-0.16	1	0	0.14	-0.11	0.29	0	0.14	0.15
5	0.18	0.13	0.18	0	1	-0.14	-0.13	-0.07	-0.33	0.08	0.24
6	0.31	0.1	-0.28	0.14	-0.24	1	0.55	0.08	0.08	-0.18	0.06
7	0.17	0.05	-0.31	-0.11	-0.13	0.55	1	-0.13	-0.13	-0.1	-0.29
8	0.18	0.13	-0.06	0.29	-0.07	0.08	-0.13	1	0.73	-0.24	0
9	-0.06	0.13	0.18	0	-0.33	0.08	-0.13	0.73	1	-0.24	0
10	-0.57	-0.55	0.01	0.14	0.08	-0.18	-0.1	-0.24	-0.24	1	-0.23
11	0.26	-0.19	0.26	0.15	0.24	0.06	-0.28	0	0	-0.23	1

1 - minereuri de Nb și Ta; 2 - mică litiniferă; 3 - amfibolit și spodumen; 4 - fosfați de Mn și Fe; 5 - minerale de Bi; 6 - casiterit și wolframit; 7 - molibdenit; 8 - minerale de U; 9 - pământuri rare; 10 - minerale de Cs; 11 - granat.

3.1.2. Coeficienții de corelație a rangurilor

Ordonarea valorilor unei caracteristici geologice într-o succesiune ascendentă sau descendentă este realizabilă atât pentru caracteristicile cantitative cât și pentru cele calitative. Operațiune extrem de ieftină din punct de vedere al prelucrării, ordonarea asociază fiecărei valori a caracteristicii studiate un număr natural, cunoscut sub denumirea de rang.

Analiza corelației rangurilor este o tehnică neparametrică pentru studiul legăturilor dintre variabilele geologice care nu ține seama de diferența dintre valorile numerice ale proprietăților, ci numai de ordinea lor.

Coeficienții definiți pentru cuantificarea intensității corelației rangurilor au valori cuprinse în intervalul $[-1,1]$ și permit analiza corelațiilor pentru două sau mai multe variabile. Ei pot fi utilizați cu deosebit succes pentru corelarea secvențelor sedimentare investigate prin carotaj geologic complex în structuri sedimentare cu numeroase alternanțe litologice pe unitatea de adâncime.

a) Coeficientul lui Spearman

Coeficientul lui Spearman (ρ_{SP}) este definit pe baza coeficientului corelației lineare al lui Pearson între două variabile v_1, v_2 și are formula:

$$\rho_{SP} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (\text{III.189})$$

în care:

n - numărul de perechi de valori ordonate crescător;

d_i - diferența rangurilor celor două variabile :

$$d_i = \text{rang } x_i - \text{rang } y_i$$

$\text{rang } x_i$ - rangul valorii x_i în sistemul ordonat crescător;

$\text{rang } y_i$ - rangul valorii y_i în sistemul ordonat crescător.

Aplicație. Analiza corelației între valoarea economică a unei roci și indicele ei de duritate pe baza valorilor din tabelul III.22 conduce la o valoare a coeficientului lui Spearman:

$$\rho_{SP} = 1 - \frac{6 \cdot 150}{10 \cdot (100 - 1)} = 0,9$$

Valorile ρ_{SP} sunt cuprinse în intervalul $[-1,1]$ iar interpretarea este similară cu a coeficientului lui Pearson din care este dedus. Pentru aplicația precedentă se poate concluziona pe baza valorii $\rho_{SP} = 0,9$ că există o bună concordanță între valoarea economică a rocii și tăria ei rezultată dintr-un ansamblu de proprietăți elementare (compoziție mineralogică, structură, textură etc.).

Tabelul III.22 Calculul coeficientului lui Spearman

Nr. crt.	Proba	Rangul		d_i	d_i^2
		Valoare economică	Tărie		
1	P1	10	5	5	25
2	P2	2	3	-1	1
3	P3	3	1	2	4
4	P4	1	10	-9	81
5	P5	5	8	-3	9
6	P6	4	2	2	4
7	P7	6	9	-3	9
8	P8	7	4	3	9
9	P9	8	6	2	4
10	P10	9	7	2	4

b) Coeficientul lui Kendall

Coeficientul lui Kendall (τ_k) are aceleași proprietăți cu coeficientul Spearman, fiind egal cu zero când cele două variabile analizate sunt independente și cu +1 și -1 când dependența dintre cele două variabile este maximă, pozitivă sau negativă.

Relația de definiție este:

$$\tau_k = \frac{2S}{n(n-1)} \quad (\text{III.190})$$

în care:

n - numărul de perechi de valori ordonate;

S - suma concordantelor posibile, calculate prin consemnarea cu +1 a "consensului" și cu -1 a variației inverse.

Aplicatie. Pentru o serie de $n = 5$ perechi de valori [densitate (ρ), coeziune (c)] (Tabelul III.23a), succesiunea operațiunilor necesare calculului coeficientului τ_k este:

Tabelul III.23 Elementele de calcul pentru coeficientul Kendall

a)			b)		
Proba	Rangul		Proba	Rangul	
	ρ	c		ρ	c
1	5	4	3	1	3
2	2	1	2	2	1
3	1	3	4	3	2
4	3	2	5	4	5
5	4	5	1	5	4

1. Ordonarea probelor după rangul unei caracteristici, de exemplu ρ (Tabelul III.23b).
2. Realizarea perechilor de ranguri prin combinarea probelor disponibile (Tabelul III.24).
3. Calculul lui S prin însumarea algebrică a variațiilor relative.
4. Calculul lui τ_k cu formula (III.190):

$$\tau_k = \frac{2 \cdot 4}{5(5-1)} = 0,4$$

Tabelul III.24 Calculul parametrului S pentru coeficientul Kendall

Nr. crt.	ρ	c	Consens +1 Contrasens -1
1	$1 \Rightarrow 2$	$3 \Leftarrow 1$	-1
2	$1 \Rightarrow 3$	$3 \Leftarrow 2$	-1

3	$1 \Rightarrow 4$	$3 \Rightarrow 5$	+1
4	$1 \Rightarrow 5$	$3 \Rightarrow 4$	+1
5	$2 \Rightarrow 3$	$1 \Rightarrow 2$	+1
6	$2 \Rightarrow 4$	$1 \Rightarrow 5$	+1
7	$2 \Rightarrow 5$	$1 \Rightarrow 4$	+1
8	$3 \Rightarrow 4$	$2 \Rightarrow 5$	+1
9	$3 \Rightarrow 5$	$2 \Rightarrow 4$	+1
10	$4 \Rightarrow 5$	$5 \Leftarrow 4$	-1
$S =$			4

În practică, frecvent, selecțiile de date conțin grupuri de k valori cu același rang. Pentru astfel de situații se calculează un rang mediu prin media aritmetică a rangurilor celor k valori. Vor apare astfel în seria ordonată a selecției k valori cu același rang. Tranzitiile între valori cu același rang sunt consemnate cu valoarea zero în calculul parametrului S .

Aplicatie. Dacă ordonarea a $n = 5$ probe după gradul de alterare este realizată de doi specialiști (A, B) obținându-se situația din tabelul III.25, rangul mediu al probelor P3 și P4 după clasificarea obținută de specialistul A este:

$$rangA_{P3} = rangA_{P4} = \frac{2+3}{2} = 2,5$$

Conform tabelelor de calcul (tabelul III.26 și tabelul III.27):

$$\tau_k = \frac{2 \cdot 1}{5(5-1)} = 0,1$$

Tabelul III.25 Coef. Kendall

Proba	RANG	
	A	B
P1	1	3
P2	4	1
P3	2-3	2
P4	2-3	4
P5	5	5

Tabelul III.27 Coef. Kendall

Nr. crt.	A	B	+1/-1
1	$1 \rightarrow 2,5$	$3 \leftarrow 2$	-1
2	$1 \rightarrow 2,5$	$3 \rightarrow 4$	1
3	$1 \rightarrow 4$	$3 \leftarrow 1$	-1
4	$1 \rightarrow 5$	$3 \rightarrow 5$	1

Tabelul III.26 Coef.Kendall

Proba	RANG	
	A	B
1	1	3
2	2,5	2
3	2,5	4
4	4	1
5	5	5

5	2,5 → 2,5	2 → 4	0
6	2,5 → 4	2 ← 1	-1
7	2,5 → 5	2 → 5	1
8	2,5 → 4	4 ← 1	-1
9	2,5 → 5	4 → 5	1
10	4 → 5	1 → 5	1
			$S = 1$

c) Coeficientul OMEGA-Kendall

Corelarea simultană a rangului mai multor variabile poate fi cuantificată prin coeficientul definit cu relația:

$$\Omega_K = \frac{12S}{m^2(n^3 - n)} \quad (\text{III.191})$$

în care:

S - suma concordanțelor multiple:

$$S = \sum_{i=1}^m (S_i - \bar{S})^2 \quad (\text{III.192})$$

S_i - suma concordanțelor binare;

$\bar{S}$ - media concordanțelor binare;

m - numărul variabilelor comparate;

n - numărul cuplurilor de valori ale selecției.

Aplicatie. Analiza corelației rangurilor a trei variabile V_1 , V_2 și V_3 , a căror clasificare este consemnată în tabelul III.28a, conduce la următoarele etape de calcul (Tabelul III.28b):

1 - media concordanțelor binare

$$\bar{S} = \frac{4+0+2}{3} = 2$$

2 - suma concordanțelor multiple

$$S = (4-2)^2 + (0-2)^2 + (2-2)^2 = 8$$

3 - coeficientul Ω_K

$$\Omega_K = \frac{12 \cdot 8}{3^2(5^3 - 5)} = 0,1$$

Valoarea 0,1 indică o corelație nesemnificativă între cele trei variabile (V1, V2 și V3).

Tabelul III.28 Elementele de calcul pentru coeficientul OMEGA-Kendall

a)				b)						
Nr. probă	Rang			Nr. crt.	Tranziții			+1/-1		
	V1	V2	V3		V1	V2	V3	V1:V2	V1:V3	V2:V3
P1	1	2	3	1	$1 \Rightarrow 2$	$2 \Leftarrow 1$	$3 \Leftarrow 2$	-1	-1	+1
P2	2	1	2	2	$1 \Rightarrow 3$	$2 \Rightarrow 4$	$3 \Rightarrow 5$	+1	+1	+1
P3	3	4	5	3	$1 \Rightarrow 4$	$2 \Rightarrow 5$	$3 \Leftarrow 1$	+1	-1	-1
P4	4	5	1	4	$1 \Rightarrow 5$	$2 \Rightarrow 3$	$3 \Rightarrow 4$	+1	+1	+1
P5	5	3	4	5	$2 \Rightarrow 3$	$1 \Rightarrow 4$	$2 \Rightarrow 5$	+1	+1	+1
				6	$2 \Rightarrow 4$	$1 \Rightarrow 5$	$2 \Leftarrow 1$	+1	-1	-1
				7	$2 \Rightarrow 5$	$1 \Rightarrow 3$	$2 \Rightarrow 4$	+1	+1	+1
				8	$3 \Rightarrow 4$	$1 \Rightarrow 5$	$5 \Leftarrow 1$	+1	-1	-1
				9	$3 \Rightarrow 5$	$4 \Rightarrow 5$	$5 \Leftarrow 4$	+1	-1	+1
				10	$4 \Rightarrow 5$	$5 \Leftarrow 3$	$1 \Rightarrow 4$	-1	+1	-1

Dacă în selecțiile analizate există și valori identice, deci cu același rang, formula (III.191) se modifică sub forma:

$$\Omega_K = \frac{12S}{m^2(n^3 - n) - m \sum_{i=1}^n (t_i^3 - t_i)} \quad (III.193)$$

semnificațiilor notațiilor fiind aceleași cu cele menționate anterior:

3.1.3. Coeficienți de asociere

Asocierea caracteristicilor calitative este o problemă de importanță deosebită în cercetarea geologică fundamentală. Compararea rocilor pe baza asociațiilor mineralogice, a nivelurilor stratigrafice pe baza speciilor fosile determinate, a zăcămintelor pe baza caracteristicilor petrografice, toate solicită existența unui instrument pentru ierarhizarea asocierii caracteristicilor calitative funcție de intensitatea ei. Aproape jumătate din datele obținute prin prospecțiune și explorare geologică sunt de natură calitativă și ignorarea acestora în etapa de analiză corelațională echivalează cu pierderea contactului cu ambianța geologică a fenomenului studiat.

Coeficienții de asociere permit descrierea cantitativă a celor două tipuri de relații fundamentale ce se stabilesc între două caracteristici calitative A și B (ex.: A=tipul petrografic: granit, dacit, bazalt etc.; B=caracterul mineralogic: ortoză, albit, olivină etc.): independența și asocierea.

Independența a două caracteristici calitative A și B este exprimată cantitativ prin identificarea aceleiași proporții de elemente A, atât printre elementele B cât și nonB. Exprimat prin intermediul frecvențelor de grupă, forma clasică a criteriului de independență pentru cele două caracteristici A și B este:

$$\frac{(AB)}{(B)} = \frac{(A\beta)}{\beta} \quad (\text{III.194})$$

Pentru identificarea comodă a independenței, indiferent de forma în care au fost sistematizate datele din cele N puncte de probare, criteriul exprimat prin relația (III.194) poate fi formulat în diferite variante echivalente :

$$\frac{(AB)}{(B)} = \frac{(A)}{N} \quad (\text{III.195})$$

$$\frac{(AB)}{(A)} = \frac{(B)}{N} \quad (\text{III.196})$$

$$(AB) = \frac{(A)(B)}{N} \quad (\text{III.197})$$

$$\frac{(AB)}{N} = \frac{(A)}{N} \frac{(B)}{N} \quad (\text{III.198})$$

Ecuția (III.198) exprimă simbolic regula fundamentală a independenței:

"Dacă caracteristicile calitative A și B sunt independente, proporția elementelor (AB) este egală cu proporția elementelor A înmulțită cu proporția elementelor B."

Asocierea exprimă existența unei legături între caracteristicile calitative, iar funcție de sensul, intensitatea și numărul de variabile implicate poate fi: pozitivă sau negativă, completă sau incompletă, totală sau parțială.

Asocierea pozitivă a două caracteristici A și B atrage creșterea numărului de elemente B o dată cu creșterea numărului de elemente A și este exprimată de inegalitatea:

$$(AB) > \frac{(A)(B)}{N} \quad (\text{III.199})$$

Asocierea negativă, opusă celei pozitive, exprimă dezasocierea caracteristicilor comparate, adică reducerea numărului de elemente B proportional cu creșterea numărului de elemente A, și este exprimată de inegalitatea:

$$(AB) < \frac{(A)(B)}{N} \quad (\text{III.200})$$

Proportional cu creșterea intensității legăturii între cele două caracteristici calitative implicate, asocierea pozitivă și negativă tind să devină complete ((A)=(B) - asociere completă; (AB)=0 – dezasociere = asociere negativă completă).

Analiza corelațională a unui sistem geologic, fie el bazin de sedimentare, zăcămint polimetalic sau de petrol, implică în mod obligatoriu studiul simultan al mai multor variabile calitative. Numai din considerente operationale, în anumite etape ale prelucrării datelor se ignoră ansamblul de corelații, lunându-se în considerare numai informațiile referitoare la două caracteristici calitative A și B, definindu-se asocierea totală între acestea. Definirea asocierii totale, presupune ipoteza că în sistemul studiat nu există o altă variabilă care să condiționeze variabilele luate în studiu.

Pentru cuantificarea intensității asocierii, presupuse totale, se utilizează în mod uzual coeficientul de asociere (Q), coeficientul de interdependentă (Y) și coeficientul de corelație calitativă (r_{AB}).

a) Coeficientul de asociere Yule și Kendall

Coeficientul Yule și Kendal, (Q), are relația de definiție:

$$Q = \frac{(AB)(\alpha\beta) - (\alpha B)(A\beta)}{(AB)(\alpha\beta) + (\alpha B)(A\beta)} \quad (\text{III.201})$$

Coeficientul de asociere Q este zero când cele două caracteristici A și B sunt independente, +1 când există asociere pozitivă completă și -1 când cele două caracteristici sunt dezasociate (= asociere completă negativă).

Coeficientul de asociere Q este independent de proporțiile relative ale elementelor A și α în selecția de date, proprietate ce-l face adecvat cazurilor în care proporțiile sunt arbitrare.

b) Coeficientul de interdependență

Coeficientul de interdependență Y , cu proprietăți similare coeficientului de asociere Q este definit cu relația:

$$Y = \frac{1 - \sqrt{\frac{(AB)(\alpha B)}{(AB)(\alpha \beta)}}}{1 + \sqrt{\frac{(AB)(\alpha B)}{(AB)(\alpha \beta)}}} \quad (\text{III.202})$$

c) Coeficientul de corelație asociativă

Coeficientul de corelație asociativă (r_{AB}) este definit (Sarapov, 1968) pe structura coeficientului corelației lineare, având aceleași proprietăți cu acesta :

$$r_{AB} = \frac{(AB)(\alpha \beta) - (A\beta)(\alpha B)}{\sqrt{(A)(\alpha)(B)(\beta)}} \quad (\text{III.203})$$

Testarea caracterului total al asocierii caracteristicilor A și B necesită verificarea influenței unei alte caracteristici C asupra asocierii acestora. Pentru aceasta se definește asocierea parțială a caracteristicilor A și B în raport cu C.

Asocierea parțială ca și cea totală poate fi pozitivă dacă se verifică inegalitatea:

$$(ABC) > \frac{(AC)(BC)}{C} \quad (\text{III.204})$$

sau negativă dacă:

$$(ABC) < \frac{(AC)(BC)}{C} \quad (\text{III.205})$$

Prin adaptarea formulelor (III.201), (III.202) și (III.203) se definesc coeficienții de asociere parțială corespunzători:

$$Q_{AB.C} = \frac{(ABC)(\alpha\beta C) - (\alpha BC)(A\beta C)}{(ABC)(\alpha\beta C) + (\alpha BC)(A\beta C)} \quad (\text{III.206})$$

$$Y_{AB.C} = \frac{1 - \sqrt{\frac{(A\beta C)(\alpha BC)}{(ABC)(\alpha\beta C)}}}{1 + \sqrt{\frac{(A\beta C)(\alpha BC)}{(ABC)(\alpha\beta C)}}} \quad (\text{III.207})$$

$$r_{AB.C} = \frac{(ABC)(\alpha\beta C) - (\alpha BC)(A\beta C)}{\sqrt{(AC)(\alpha C)(BC)(\beta C)}} \quad (\text{III.208})$$

Testarea influenței caracteristicii C asupra asocierii caracteristicilor A și B se bazează pe compararea coeficienților calculați pentru asociere în raport atât cu caracteristica C cât și cu caracteristica nonC ($=\gamma$). Egalitatea $Q_{AB.C} = Q_{AB\gamma}$ indică independența asocierii caracteristicilor A și B în raport cu caracteristica C, altfel spus, între caracteristicile A și B este o asociere totală.

Proportional cu creșterea numărului de caracteristici luate în studiu crește numărul asociațiilor parțiale care se pot analiza pentru precizarea ansamblului de corelații din sistemul studiat.

3.1.4. Coeficienți de corelație temporală

În cercetarea ecologică se operează frecvent cu serii de valori ale unor variabile v_{ij} ($i = 1, 2, 3, \dots, nv$; $j = 1, 2, 3, \dots, ni$; nv - numărul de variabile; ni - numărul de valori pentru fiecare variabilă) obținute prin determinări realizate la intervale mai mult sau mai puțin egale.

Astfel de serii de valori cunoscute sub denumirea generică de serii de timp pot fi constituite

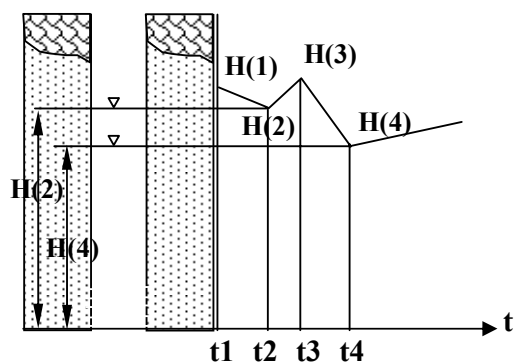


Fig. 62 Serie de timp a nivelurilor piezometrice măsurate într-un acvifer freatic

din: cote ale nivelului piezometric ale unui acvifer măsurate la intervale de timp egale (Fig.62), succesiunea litologică a unei secvențe sedimentare separată în intervale egale ca grosime (Fig.63), număr de microfosile identificate pe o direcție oarecare de probare (Fig.64).

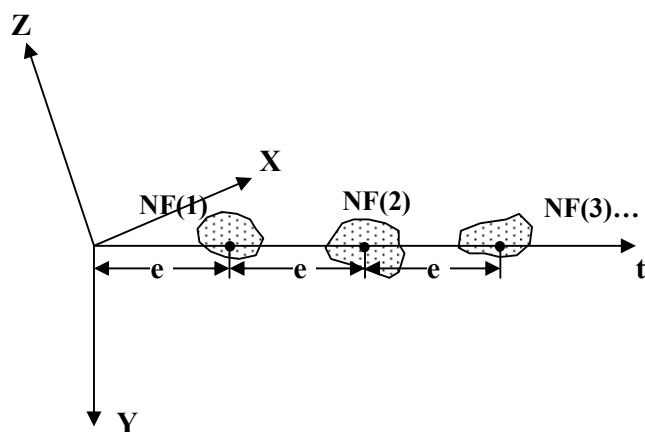


Fig. 64 Număr de microfosile identificate în puncte de probare plasate pe o direcție oarecare de probare

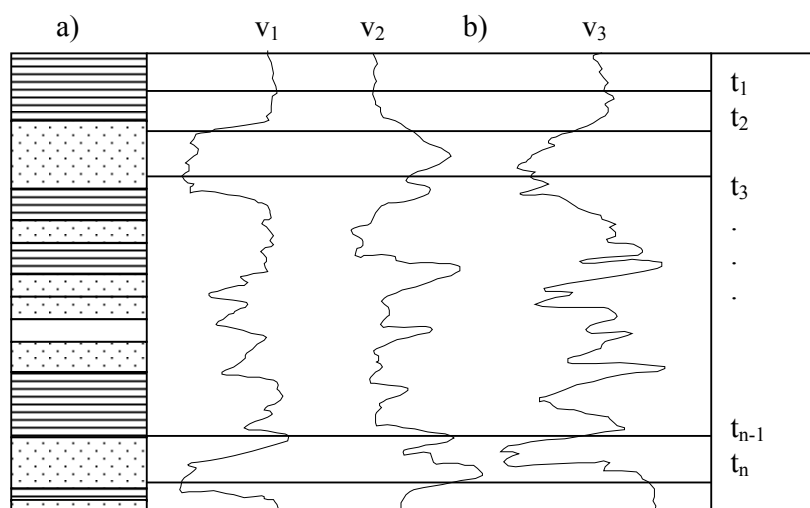


Fig. 63 Serii de timp rezultate din cercetarea unei succesiuni sedimentare

a) serie de timp litologică univariată;

b) serie de timp multivariată ($v_1 = \rho, v_2 = PS; v_3 = \gamma\gamma$) obținută din diagrafa geofizică complexă

Timpul, într-o astfel de serie de valori sau stări ale procesului studiat este echivalent fie cu grosimea stratigrafică, fie cu adâncimea măsurată într-un foraj, fie cu distanța de-a lungul unei direcții oarecare din spațiu.

Studiul seriilor de timp beneficiază de o amplă și sofisticată metodologie (Tertisco M.et.al.,1985) care nu poate fi utilizată cu eficiență maximă în geologie din două motive principale:

- a)volumul mare de date necesar calculului parametrilor caracteristici analizei seriilor de timp univariate, cu semnificație relativ redusă în studiul proceselor geologice complexe, multivariate;
- b)complexitatea metodologiei care introduce dificultăți de interpretare în analiza seriilor de timp multivariate, adecvate studiului proceselor geologice complexe.

a)Formalizarea stocastică a seriilor de timp

Existența unui volum minim de date pentru studiul unei serii de timp în scopul estimării stocastice a corelațiilor presupune o formalizare care asociază caracteristicile studiate (ex.: litologia, nivelul piezometric, numărul de fosile identificate etc.) o variabilă aleatoare de obicei discretă (caracterul discret fiind determinat de modul de colectare a datelor și nu de natura variabilei studiate), iar conținutului variabilei, un ansamblu de stări (ex.: variate tipuri litologice: calcar, argilă, gresie; sensul evoluției: ascendent, descendent, constant).

O serie de timp este din punct de vedere formal o succesiune de stări exclusive, iar instrumentul operational care permite identificarea probabilistică a ponderii componente deterministe (=corelaționale) a procesului este matricea de tranziție.

Matricea de tranziție sacrifică toate informațiile referitoare la poziția stărilor în secvența de date, în favoarea identificării tendinței unei stări de a fi urmată sau precedată de alta.

Există două tipuri principale de matrici de tranziție: matrici de tranziție unitară (de un pas) și matrici de tranziție multiplă, fiecare dintre ele putând fi exprimate numeric în trei forme diferite: 1) matricea frecvențelor de tranziție, 2) matricea proporției perechilor de tranziții, 3) matricea proporțiilor de tranziție.

1) Matricea frecvențelor de tranziție este formată din numărul tranzițiilor de la o stare la alta determinată pe baza seriei de observații disponibile.

Pentru seria de $n = 31$ stări:

ABACDCDABCBADCDCBACABDABCDABACDA

matricea frecvențelor celor $n - 1 = 30$ tranziții (MFT) este:

$$\begin{array}{ccccc}
 & A & B & C & D \\
 & & & & \text{Total} \\
 MFT = \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 4 & 3 & 1 \\ 4 & 0 & 2 & 1 \\ 1 & 2 & 0 & 5 \\ 3 & 1 & 3 & 0 \end{bmatrix} & \begin{array}{c} 8 \\ 7 \\ 8 \\ 7 \end{array} \\
 & & & & 30 \\
 & \text{Total} & 8 & 7 & 8 & 7
 \end{array} \quad (III.209)$$

2) Matricea proporției perechilor de tranziții ($MPPT$) se obține din MFT prin divizarea fiecărei valori cu numărul total de tranziții și exprimă ponderea unei tranziții în totalul acestora:

$$\begin{array}{cccc}
 A; & B; & C; & D; \\
 & & & \text{Total}
 \end{array}$$

$$\begin{array}{c}
 A \\
 B \\
 C \\
 D
 \end{array}
 \begin{bmatrix}
 0,00 & 0,13 & 0,10 & 0,03 \\
 0,13 & 0,00 & 0,07 & 0,03 \\
 0,03 & 0,07 & 0,00 & 0,17 \\
 0,10 & 0,03 & 0,10 & 0,00
 \end{bmatrix}
 \begin{array}{c}
 0,26 \\
 0,23 \\
 0,27 \\
 0,23
 \end{array}
 \quad (III.210)$$

1,00

Total 0,26 0,23 0,27 0,23

3) Matricea proporțiilor de tranziție (*MPT*) exprimă proporția în care o stare poate fi urmată de alta fără a ține seama de ponderea stării inițiale în totalul acestor tranziții. Ea se calculează prin divizarea fiecărui element dintr-un rând al MFT prin suma frecvențelor din rândul respectiv.

	A	B	C	D	Total
<i>A</i>	0,000	0,500	0,375	0,125	1,000
<i>B</i>	0,571	0,000	0,286	0,143	1,000
<i>C</i>	0,125	0,250	0,000	0,625	1,000
<i>D</i>	0,428	0,143	0,428	0,000	1,000

$$\begin{array}{c}
 A \\
 B \\
 C \\
 D
 \end{array}
 \begin{bmatrix}
 0,000 & 0,500 & 0,375 & 0,125 \\
 0,571 & 0,000 & 0,286 & 0,143 \\
 0,125 & 0,250 & 0,000 & 0,625 \\
 0,428 & 0,143 & 0,428 & 0,000
 \end{bmatrix}
 \begin{array}{c}
 1,000 \\
 1,000 \\
 1,000 \\
 1,000
 \end{array}
 \quad (III.211)$$

Cele trei forme de exprimare ale matricii de tranziție pot fi construite pentru o tranziție unitară când procesul studiat operează la momente consecutive, exprimate formal de indicele superscris al probabilității de tranziție de la starea "j" la starea "k".

$$p_{jk}^{(1)} = P\{V_{m+1} = k | V_m = j\} \quad (III.212)$$

Pentru o tranziție multiplă (*n* pași), probabilitatea de tranziție de la starea "j" la starea "k" se scrie:

$$p_{jk}^{(n)} = P\{V_{m+n} = k | V_m = j\} \quad (III.213)$$

În cazul în care probabilitățile p_{jk} depind numai de pasul *n* și sunt independente de poziția inițială "m" (situație valabilă pentru un lanț Markov omogen) matricea de tranziție multiplă se calculează pe baza matricilor de tranziție unitară.

Relația de recurență a prognozei stării sistemului pentru orice "moment" este:

$$p^{(m)} = p^{(0)} \times P^{(m)} \quad (III.214)$$

în care $P^{(m)}$ este matricea constituită din probabilitățile de tranziție multiplă $p_{jk}^{(m)}$.

Aplicatie. Pentru matricea proporției de tranziție unitară:

$$MPT1 = \begin{matrix} GRESIE \\ ARGILA \\ CALCAR \end{matrix} \begin{bmatrix} 0,70 & 0,20 & 0,10 \\ 0,16 & 0,50 & 0,34 \\ 0,50 & 0,25 & 0,25 \end{bmatrix}$$

se obține prin calcule succesive:

$$MPT1^{(2)} = \begin{bmatrix} 0,57 & 0,27 & 0,16 \\ 0,36 & 0,37 & 0,27 \\ 0,52 & 0,29 & 0,11 \end{bmatrix} \quad MPT1^{(4)} = \begin{bmatrix} 0,50 & 0,30 & 0,20 \\ 0,48 & 0,31 & 0,21 \\ 0,50 & 0,30 & 0,20 \end{bmatrix}$$

$$MPT1^{(6)} = \begin{bmatrix} 0,50 & 0,30 & 0,20 \\ 0,50 & 0,30 & 0,20 \\ 0,50 & 0,30 & 0,20 \end{bmatrix} \quad MPT1^{(8)} = \begin{bmatrix} 0,50 & 0,30 & 0,20 \\ 0,50 & 0,30 & 0,20 \\ 0,50 & 0,30 & 0,20 \end{bmatrix}$$

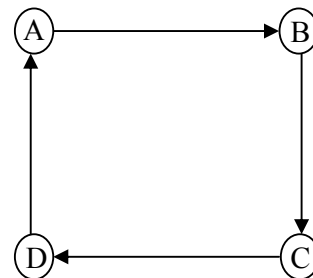
o matrice de echilibru, care nu se modifică peste o anumită valoare a exponentului și care prin structura numerică exprimă intensitatea corelațiilor care există în seria de timp analizată.

Pentru exemplificarea modului în care se reflectă gradul de determinare în structura unei matrici de tranziție prezentăm în continuare:

a) matricea unui proces determinist de tipul MPTD:

...ABCDABCDABCDABCD...

$$MPTD = \begin{matrix} & A & B & C & D \\ A & \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \\ B & \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \\ C & \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \\ D & \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$



cu exprimarea grafică a tranzițiilor în fig. 65. **Fig. 65** Tranzițiile într-un proces determinist

b) matricea unui proces aleator de tip MPDA:

...DBABCDCAABCABDCDCBCDDBAD...

$$MPDA = \begin{matrix} A & \begin{bmatrix} 0,000 & 0,390 & 0,450 & 0,160 \end{bmatrix} \\ B & \begin{bmatrix} 0,360 & 0,000 & 0,320 & 0,320 \end{bmatrix} \\ C & \begin{bmatrix} 0,370 & 0,100 & 0,000 & 0,530 \end{bmatrix} \\ D & \begin{bmatrix} 0,150 & 0,460 & 0,390 & 0,000 \end{bmatrix} \end{matrix}$$

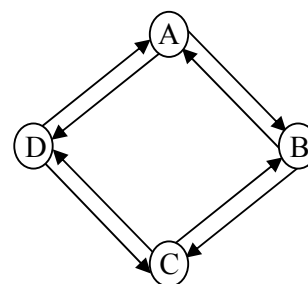


Fig. 66 Tranzițiile în MPDA

cu exprimarea grafică a tranzițiilor în fig. 66.

La un număr mare de valori ale unei serii de timp aleatoare, probabilitățile devin egale (ex.: $P(A|B) = P(A|C) = P(A|D) = 1/3$) în cazul unui sistem cu patru stări distincte A,B,C,D). Între cele două extreme (model determinist și aleator) există o infinitate de variante differentiate prin intensitatea corelațiilor.

Descrierea statistică a seriilor de timp este realizată prin patru funcții elementare: dispersia, densitatea de probabilitate, coeficientul de autocorelație sau intercorelație și densitatea spectrală. Dacă primele două sunt utilizate pentru orice variabilă cu comportament aleator, ultimele două sunt specifice seriilor de timp.

b) Coeficientul de autocorelație

Autocovarianța este covarianța a două realizări ale aceleiași variabile (V) care este determinată în două puncte separate prin intervalul h . Covarianța, ca o funcție de h poate fi scrisă sub forma:

$$C_V(h) = E(V_n, V_{n+h}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N V_n \times V_{n+h} \quad (\text{III.215})$$

în care

h - "distanța" dintre cele două valori ($h = 0, 1, 2, \dots, N-1$);

N - numărul de valori ale seriei de timp.

Funcția de covarianță este simetrică în jurul valorii zero:

$$C_V(-h) = C_V(h) \quad (\text{III.216})$$

iar dacă $h = 0$ covarianța se reduce la dispersie (=variantă) și se poate scrie :

$$C_V(0) = \text{var}(V) = \frac{1}{N} \sum_{n=1}^N V_n^2 \quad (\text{III.217})$$

Coeficientul de autocorelație se obține prin divizarea covarianței la variantă și poate fi scris sub forma:

$$R_V(h) = \frac{C_V(h)}{C_V(0)} \quad (\text{III.218})$$

Estimatorul coeficientului de corelație se calculează cu relația:

$$r_V(h) = \frac{(N-h) \sum_{i=1}^{N-h} v_i v_{i+h} - \sum_{i=1}^{N-h} v_i \sum_{i=1}^{N-h} v_{i+h}}{\sqrt{(N-h) \sum v_i^2 - \left(\sum v_i\right)^2} \sqrt{(N-h) \sum v_{i+h}^2 - \left(\sum v_{i+h}\right)^2}} \quad (\text{III.219})$$

Valorile coeficientului de autocorelație sunt cuprinse în intervalul $[-1,1]$ și evident $R_V(0) = 1$ este valoarea care indică o corelație maximă. Valoarea $R_V(0) = -1$ indică o corelație maximă inversă. Valorile estimate ale coeficientului de autocorelație permit identificarea ciclicităților dintr-o serie de timp.

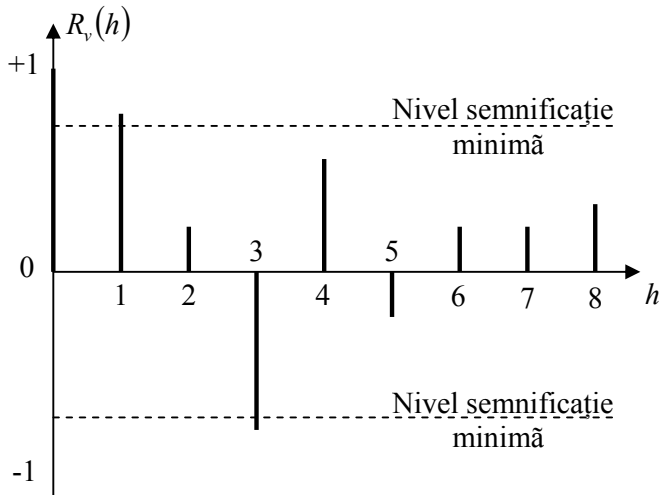


Fig. 67 Corelograma unei serii de timp

Reprezentarea grafică a variației coeficientului de autocorelație în funcție de h poartă denumirea de corelogramă (Fig. 67) și ilustrează într-o formă sintetică semnificația statistică a componentelor ciclice ale seriei studiate.

Selectarea componentelor cu semnificație statistică se face prin alegerea unui nivel de semnificație minimă care filtrează valorile coeficientului de autocorelație. Într-un model pentru

reproducerea și prognoza seriei de timp sunt reprezentate numai componentele al căror coeficient de autocorelație depășește nivelul de semnificație minim.

Aplicatie. Ca un exemplu simplu se poate calcula corelograma unui proces geologic de tip markovian descris printr-o matrice de tranziție. Acest lucru se poate realiza prin asocierea unei valori numerice fiecărei stări a sistemului .

Pentru un proces cu două stări distincte, asociind unei stări valoarea unu și celei de-a doua valoarea zero matricea de tranziție va fi notată:

$$MPT = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

în care p_{00}, p_{01}, p_{10} și p_{11} sunt probabilitățile de tranziție din sistemul studiat.

Conform relației (III.215):

$$C_V(h) = E(V_n = V_{n+h} = 1) = P(V_n = 1) \times P(V_{n+h} = 1 | V_n = 1)$$

și deoarece

$$P(V_n = 1) = E(V_n) = p_1$$

în care p_0, p_1 sunt probabilitățile stabile ale matricii MPT:

$$C_V(h) = p_1 \times p_{11}^h$$

și

$$R_V(h) = p_{11}^h$$

Corelograma unui astfel de proces markovian corespunde puterilor probabilităților de tranziție p_{11} și în general, pentru orice lanț markov va fi o funcție simplă de $MPT^{(h)}$.

Dacă se calculează corelograma unui proces aleator "pur" în care $E(V_n) = 0$, atunci $R_V(h) = 0$ pentru $h = 1, 2, 3, \dots$ având un singur maxim de $R_V(h) = 1$ pentru $h = 0$. Acest lucru este în acord cu definiția unui proces aleator în care se presupune că nu există corelații între V_n și V_{n+h} pentru orice n și orice h diferit de zero.

c) Coeficientul de intercorelație

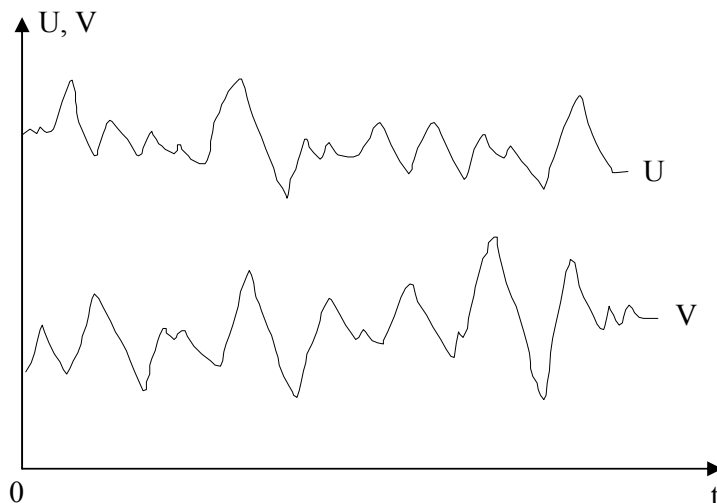


Fig. 68 Variația în “timp” a două caracteristici geologice cu comportament aleator înregistrat într-un carotaj etc.) (Fig. 68).

Coeficientul de intercorelație este utilizat pentru evaluarea intensității corelației dintre două serii de timp ce măsoară variația a două variabile disticte U, V (ex.: U = precipitațiile, V = cota nivelului piezometric al unui acviferului freatic); U = porozitatea, V = valoarea PS-ului corespunzător

Relația de calcul pentru coeficientul de intercorelație este:

$$r_{UV}(h) = \frac{(N-h) \sum_{i=1}^{N-h} V_i U_{i+h} - \sum_{i=1}^{N-h} V_i \sum_{i=1}^{N-h} U_{i+h}}{\sqrt{(N-h) \sum_{i=1}^{N-h} V_i^2 - \left(\sum_{i=1}^{N-h} V_i \right)^2} \sqrt{(N-h) \sum_{i=1}^{N-h} U_{i+h}^2 - \left(\sum_{i=1}^{N-h} U_{i+h} \right)^2}} \quad (\text{III.220})$$

Domeniul de variație și semnificația coeficientului de intercorelație sunt analoage cu cele ale coeficientului de autocorelație. Referindu-se la două variabile $R_{UV}(0)$ este identic cu coeficientul lui Pearson și numai în cazul unei corelații liniare perfecte între U și V va avea valoarea unitară, pozitivă sau negativă după cum corelația este directă respectiv inversă.

Corelograma coeficientului de intercorelație este utilizată în scopul identificării periodicității seriilor de timp multivariate, a decalajelor cu semnificație statistică pentru cupluri de două variabile.

Prin analiza corelației dintre variația precipitațiilor și a nivelului piezometric din acviferele freatice se poate evalua, spre exemplu, cu ajutorul coeficientului de intercorelație, durata de tranzit a apei prin zona de aerare și implicit vulnerabilitatea la poluare a acviferelor.

Atât pentru coeficientul de autocorelație cât și pentru cel de intercorelație seriile de timp sunt presupuse lineare și stationare. Dacă aceste condiții nu sunt îndeplinite, evaluarea corelațiilor temporare presupune o preprocesare care să realizeze:

- a) linearizarea datelor (prin logaritmare, ridicare la putere, extragerea rădăcinii de un ordin oarecare) sau separarea datelor într-un număr oarecare de subdomenii pe care să se comporte linear;
 - b) eliminarea tendințelor neperiodice care maschează componentele ciclice ale seriilor de timp.
- Această operațiune se realizează prin identificarea modelului analitic al tendinței și eliminarea ei din datele brute. Evaluarea coeficienților se operează asupra valorilor "reziduale" (M.Tertisco et.al., 1985).

3.2. Factorizarea corelațiilor

Rezultată din complexitatea proceselor ecologice, necesitatea identificării factorilor principali care determină evoluția fenomenelor este obiectivul final al descrierii multivariate a proceselor ecologice. Unul din cele mai adaptate instrumente pentru soluționarea acestei probleme este *analiza factorială*.

Analiza factorială a fost privită în general ca o metodă misterioasă de o mare complexitate. O parte din misterul care o înconjoară provine din bogata terminologie utilizată. Analiza factorială a fost dezvoltată de psihologii experimentalisti în anii 1930-1940 și mare parte din terminologie are semnificație numai în contextul acestui domeniu.

Obiectivul original al analizei factoriale a fost să dea un sistem corect de evaluare a inteligenței prin corelarea punctajelor obținute din diferite teste relative la abilitatea mentală. Este în general acceptat faptul că punctajul dintr-un singur test nu poate da o măsură reală a inteligenței unei persoane. O persoană bine înzestrată intelectual va obține rezultate mai bune la majoritatea testelor de inteligență decât o persoană considerată inferioară mental. Diferențele la testele specifice nu reflectă diferențele mentale ci de educație, cultură generală și circumstanțiale, legate de condițiile în care se desfășoară testele. Psihologii au considerat analiza factorială capabilă să extragă coeficientul corect de evaluare a inteligenței din rezultatele tuturor testelor chiar dacă nici unul dintre aceste teste, individual, nu este capabil să o facă corect.

Aplicată în cercetări biologice și geologice analiza factorială studiază relațiile dintre un număr mare de variabile măsurabile, cu scopul evidentierii unor noi variabile, teoretice, numite factori.

Aceste noi variabile (=teoretice =factori) sunt într-un număr mai mic decât variabilele măsurabile și sunt în același timp funcții lineare de variabilele măsurabile.

Noile variabile sunt astfel stabilite încât să explice într-un procent cât mai mare varianta variabilelor originale. Se caută prin analiza factorială găsirea unui număr cât mai mic de factori (=variabile teoretice) care să exprime variabilitatea observată prin intermediul valorilor măsurate.

Variabilitatea reziduală, rămasă neexprimată este o pierdere de informație compensată prin numărul redus de variabile teoretice cu care se operează în continuare pentru modelarea procesului studiat.

Variabilele teoretice (=factorii) vor putea reflecta fenomene naturale care sunt la originea variabilității observate și astfel se vor putea interpreta într-o optică naturalistă rezultatele calculului cantitativ.

Fundamentate pe aceleași principii, factorizarea corelațiilor sistemelor multivariate poate fi abordată prin trei variante ale analizei factoriale: analiza în componenți principali, analiza factorială R-MOD și analiza factorială Q-MOD.

Separarea tipurilor de sedimente pe baza variabilității compoziției granulometrice și identificarea fracțiunilor caracteristice diferitelor tipuri de sedimente pot fi realizate prin aplicarea analizei componentilor principali. Dacă se studiază un corp plutonic, pentru stabilirea numărului factorilor care condiționează distribuția elementelor chimice și mineralelor se utilizează analiza factorială R-MOD. Gruparea taxonomică a unui lot de esantioane prelevate din diferite tipuri de roci (ex.: sienit, monzonit, diorit, quartit, gabrou, norit, diabaz) pe baza oxizilor continuți (ex.: SiO_2 , Al_2O_3 , Fe_2O_3 , FeO , MgO , CaO , Na_2O , K_2O) se poate realiza printr-o analiză factorială Q-MOD.

Toate variantele analizei factoriale vor fi luate în studiu în acest capitol, punctul de plecare fiind obligatoriu analiza în componenți principali.

Obiectivul operational al analizei factoriale este interpretarea structurii matricilor de varianță-covarianță pentru un ansamblu multivariat de date. Tehnica utilizată este extragerea valorilor proprii și a vectorilor proprii din aceste matrici care exprimă sintetic ansamblul de relații dintre variabilele măsurate.

3.2.1. Valori proprii și vectori proprii

Determinarea valorilor proprii și vectorilor proprii este privită ca fiind cea mai dificilă operație în algebra matricială. Dificultatea nu constă în metoda de calcul, care nu este mai dificilă decât alte procedee matematice, ci în perceperea semnificației acestor instrumente în mod intuitiv.

Pentru o clară percepere a acestor semnificații vom utiliza o interpretare geometrică deosebit de clară aplicabilă matricei coordonatelor a două puncte plasate într-un spațiu bidimensional și vom interpreta valorile proprii, vectorii proprii și funcțiile asociate ca proprietăți geometrice ale aranjamentului acestor puncte.

Această abordare ne limitează la matrici mici (2X2) dar rezultatele obținute pot fi extrapolate la sisteme mai mari chiar dacă calculul manual devine impracticabil. Trebuie notat cu acest prilej că suntem într-un domeniu în care puterea de calcul chiar a celor mai moderne calculatoare deseori este inadecvată pentru soluționarea problemelor reale.

a) Valori proprii

Considerăm sistemul matricial ipotetic:

$$[A][X] = \lambda[X] \quad (\text{III.258})$$

care formal este similar cu

$$[A][X] = [B] \text{ în care } [B] = \lambda[X] \quad (\text{III.259})$$

Ecuția poate fi rescrisă sub forma:

$$([A] - \lambda[I])[X] = [O] \quad (\text{III.260})$$

în care I este matricea identitate.

Pentru matrici [2X2], ecuația matricială (III.260) poate fi scrisă sub forma sistemului:

$$\begin{cases} (A_{11} - \lambda)X_1 + A_{12}X_2 = 0 \\ A_{21}X_1 + (A_{22} - \lambda)X_2 = 0 \end{cases} \quad (\text{III.261})$$

Presupunând că sistemul are și alte soluții decât cea banală $X_1 = X_2 = 0$ atunci trebuie ca:

$$\det|A - \lambda * I| = 0 \quad (\text{III.262})$$

care prin dezvoltare devine ecuația:

$$\lambda^2 - (A_{11} + A_{22})\lambda + A_{11}A_{22} - A_{21}A_{12} = 0 \quad (\text{III.263})$$

cu două soluții reale în cazul unei matrici A simetrice.

Aplicatie. Pentru două puncte $P_1(4,8)$ și $P_2(8,4)$ matricea coordonatelor este:

$$A = \begin{bmatrix} 4 & 8 \\ 8 & 4 \end{bmatrix}$$

iar matricea pentru calculul valorilor proprii

$$A = \begin{bmatrix} 4 - \lambda & 8 \\ 8 & 4 - \lambda \end{bmatrix}$$

Soluțiile ecuației de gradul doi care rezultă prin dezvoltarea determinantului sunt:

$$\lambda_1 = -4 \text{ și } \lambda_2 = 12$$

Punctele P_1 și P_2 pot fi imaginate ca fiind plasate pe conturul unei elipse al cărei centru este plasat în centrul sistemului de referință. Elipsa este ca o anvelopă care cuprinde ambele puncte iar valorile proprii pot fi interpretate ca semiaxele elipsei. Raportul axelor poate fi o expresie numerică a gradului de împrăștiere a punctelor. Cu cât punctele sunt mai apropiate, lungimea axelor diferă mai mult și elipsa tinde spre o dreaptă. Dacă cele două puncte se află pe doi vectori perpendiculari elipsa devine cerc.

Ca exemplificare se calculează valorile proprii pentru matricile coordonatelor a două puncte situate pe două axe care fac un unghi de: a) 90° ; b) 45° ; c) 30° ; d) 0° (Fig. 69).

$$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix}; \quad \text{a) } \begin{bmatrix} -4 & 8 \\ 8 & 4 \end{bmatrix} \quad \text{b) } \begin{bmatrix} 4 & 8 \\ 8 & 4 \end{bmatrix} \quad \text{c) } \begin{bmatrix} 6 & 8 \\ 8 & 6 \end{bmatrix} \quad \text{d) } \begin{bmatrix} 4 & 8 \\ 4 & 8 \end{bmatrix}$$

a)

$$\lambda_1 = 8,95$$

$$\lambda_2 = -8,95$$

b)

$$\lambda_1 = 12$$

$$\lambda_2 = 4$$

c)

$$\lambda_1 = 14$$

$$\lambda_2 = -2$$

d)

$$\lambda_1 = 12$$

$$\lambda_2 = 0$$

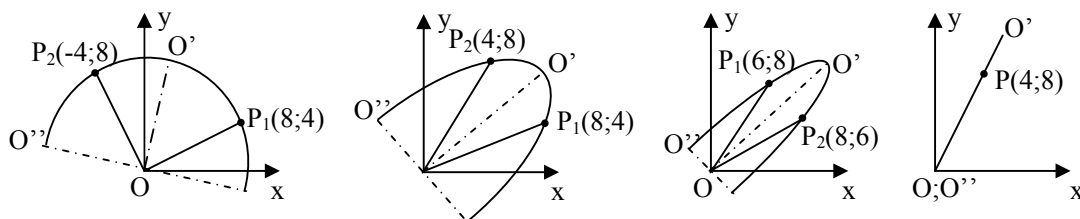


Fig. 69 Semnificația geometrică a valorilor proprii și vectorilor proprii

Ca regulă de verificare a corectitudinii calculului valorilor proprii se reține că suma valorilor proprii este egală cu urma matricii inițiale (suma valorilor de pe diagonala principală).

Valorile proprii reprezintă lungimile celor două semiaxe ale elipsei pe care sunt plasate cele două puncte sau, generalizând, la "n" dimensiuni, "n" semiaxe ale elipsoidului care înglobează toate punctele într-un spațiu cu "n" dimensiuni.

b) Vectori proprii

Revenind la ecuația $([A] - \lambda[I])[X] = [O]$, dacă după calculul valorilor proprii acestea sunt utilizate pentru calculul soluției nebanale, se obțin vectorii proprii ai matricii inițiale.

Pentru matricea $[2 \times 2]$ dezvoltând ecuația (III.260) se obține:

$$\begin{bmatrix} A_{11}-\lambda & A_{12} \\ A_{21} & A_{22}-\lambda \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (\text{III.264})$$

Vectorul $[X_1, X_2]$ se numeste vector propriu (=caracteristică proprie =caracteristică latentă =vector principal) asociat valorii proprii.

Pentru a concluziona relativ la partea operatională, trebuie mentionat că pentru a afla vectorii proprii și valorile proprii ale unei matrici $[n \times n]$ trebuie să-i găsim determinantul, rădăcinile ecuației polinomiale caracteristice și să soluționăm un set de n ecuații cu n necunoscute.

Aplicatie. Revenind pentru interpretare la matricea

$$A = \begin{bmatrix} 4 & 8 \\ 8 & 4 \end{bmatrix}$$

ecuația de calcul pentru vectorul propriu al valorii proprii $\lambda_1 = 12$ este:

$$\begin{bmatrix} 4-12 & 8 \\ 8 & 4-12 \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

cu soluția

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Pentru ecuație există o infinitate de vectori proprii pentru că sistemul este satisfăcut de

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \beta \times \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

unde β este o constantă oarecare. Practic este insuficient să ne limităm la $\beta = 1$ deoarece, așa cum se va vedea, suntem interesați de valorile rapoartelor dintre elementele vectorului care nu se schimbă prin multiplicare cu o constantă.

Pentru cea de-a doua valoare proprie $\lambda_2 = -4$, soluția pentru al doilea vector propriu este:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \beta \times \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Revenind la figura 69, vectorii proprii pot fi interpretați ca pantele celor două axe ale elipsei. Primul vector propriu definește bisectoarea unghiului determinat de cele două puncte și centrul elipsei și a cărei lungime este egală cu prima valoare proprie ($\lambda_1 = 12$), iar ce-l de-al doilea vector propriu definește axa ortogonală cu prima.

De reținut că matricile simetrice au toate valori proprii reale iar vectorii proprii corespondenți sunt ortogonali.

3.2.2.Standardizarea

Analiza factorială este deseori **confruntată** cu interpretarea unei matrici de varianță-covarianță obținută dintr-o colecție de caracteristici geologice exprimate în unități de măsură diferite.

Valorile exprimate în unități de măsură diferite nu pot fi comparate direct necesitând o transformare a datelor originale prin standardizare.

Standardizarea se realizează prin extragerea din fiecare valoare originală a valorii medii a variabilei și divizarea diferenței prin abaterea standard. Se obține astfel un nou set de valori cu media zero și dispersia unu .

Standardizarea permite compararea variabilelor exprimate în unități de măsură diferite, altfel spus permite compararea "merelor" cu "perele".

Dacă se operează cu matricea de corelație a variabilelor studiate, cum este cazul în analiza factorială Q-MOD sau R-MOD, nu este necesar să se standardizeze valorile pentru că de fapt matricea de corelație este matricea de varianță-covarianță a datelor standardizate.

Standardizarea poate avea o influență determinantă asupra structurii matricii de varianță-covarianță și în consecință asupra rezultatelor analizei factoriale dacă amplitudinile de selecție ale variabilelor diferă semnificativ și distribuțiile sunt puternic asimetrice. **Când unitățile de măsură nu diferă se recomandă din acest evitarea standardizării.**

Pentru ilustrarea efectului standardizării să considerăm reprezentările grafice ale datelor brute (Fig. 70) și ale celor standardizate (Fig. 71) pentru care au fost calculate separat matricile de covarianță, valorile proprii și vectorii proprii.

Efectul standardizării este extinderea ambelor variabile pe același interval valoric cu modificarea raportului de împrăștiere a valorilor pe cele două axe și rotirea axelor principale cu 45° (cu 45° pentru toate matricile binare și cu valori diferite în cazul matricilor mai mari).

De asemenea, se remarcă o reducere slabă a varianței de-a lungul primului vector propriu (de la 96% la 93%), reducere care se accentuează proportional cu diferența dintre domeniile de variație ale variabilelor originale.

Tabelul III.32 Elementele de standardizare

Valori nestandardizate	Valori standardizate
MEDIA	
$m(X_1) = 5$	$m(XS_1) = 0$
$m(X_2) = 10$	$m(XS_2) = 0$
VARIANȚA	
$s^2(X_1) = 6,08$	$s^2(XS_1) = 1$
$s^2(X_2) = 27,54$	$s^2(XS_2) = 1$
MATRICE DE COVARIANȚĂ	MATRICE DE CORELAȚIE
$\text{cov} = \begin{bmatrix} 6,08 & 11,08 \\ 11,08 & 27,54 \end{bmatrix}$	$R = \begin{bmatrix} 1,00 & 0,86 \\ 0,86 & 1,00 \end{bmatrix}$
VALORI PROPRII	
$\lambda_1 = 32,23$ (96%)	$\lambda_1 = 1,86$ (93%)
$\lambda_2 = 1,39$ (4%)	$\lambda_2 = 0,14$ (7%)
VECTORI PROPRII	
$V_1[0,39;0,92]$	$V_1[0,707;0,707]$
$V_2[0,92;-0,39]$	$V_2[-0,707;0,707]$

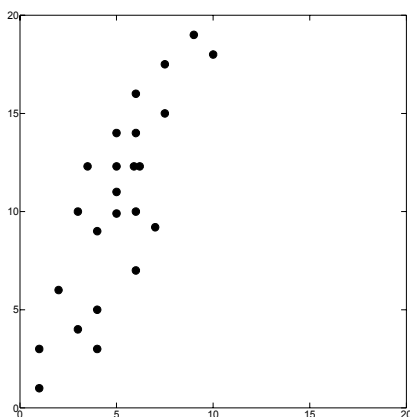


Fig. 70 Reprezentarea grafică a datelor nestandardizate

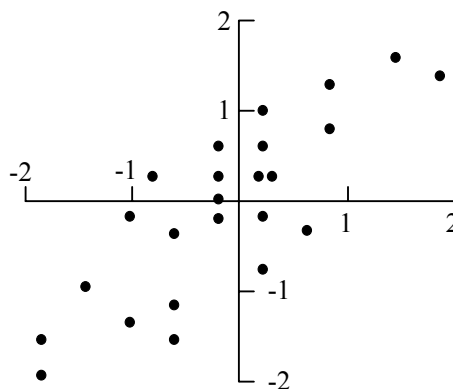


Fig. 71 Reprezentarea grafică a datelor standardizate

3.2.3. Analiza în componenți principali

Analiza în componenți principali constă în transformarea liniară a m variabile măsurabile corelate, în n variabile teoretice care sunt combinații liniare ale celor vechi. Fiecare nouă variabilă este astfel creată încât să înglobeze cât mai mult din varianța totală a datelor originale.

Componenții principali nu sunt altceva decât vectorii proprii ai matricii de varianță-covarianță. În calcule nu este implicată nici o ipoteză probabilistică sau testare astfel încât A.C.P., strict vorbind, este doar o prelucrare matematică și nu o procedură statistică. Utilitatea A.C.P. este apreciată după performanțe și nu după considerații teoretice.

a) Metodologia de lucru

Presupunând că dispunem de o colecție de 25 de exemplare de brahiopode și măsurăm pentru fiecare exemplar lungimea X_1 și lățimea X_2 (tabelul III.32) matricea de varianță-covarianță obținută prin calcul este

$$\text{cov} = \begin{bmatrix} 20,3 & 15,60 \\ 15,60 & 24,10 \end{bmatrix}$$

Reprezentând grafic această matrice, considerând-o ca fiind alcătuită din coordonatele a două puncte cu abscisele pe prima linie și cu ordonatele pe a doua, se obține o reprezentare vectorială care exprimă grafic corelația dintre cele două variabile X_1 și X_2 (Fig. 72 și 73).

Calculul vectorilor proprii și al valorilor proprii conduc la obținerea elementelor elipsei ce înglobează toate cele 20 de puncte din tabelul III.32: $Vector I = [0,66; 0,75]$, $Vector II = [0,75; -0,66]$ cu $\lambda I = 37,9$ și $\lambda II = 6,5$ (Fig. 74).

Tabelul III.32 Elemente ale analizei în componente principale

VALORILE SELECTIEI			VALORILE FACTORIZATE		DATELE ORDONATE	
Nr.	X_1	X_2	Y_1	Y_2	X_1	X_2
1	3	2	3.49	0.92	3	2
2	4	10	10.14	-3.64	4	2
3	6	5	7.72	1.18	6	5
4	6	8	9.97	-0.81	6	5
5	6	10	11.46	-2.14	6	6
6	7	2	6.14	3.91	7	7
7	7	13	14.37	-3.38	7	7
8	8	9	12.04	3.32	8	8
9	9	5	9.71	3.42	9	8
10	9	8	11.96	1.43	9	9
11	9	14	16.45	-2.45	9	10
12	10	7	11.87	2.84	10	10
13	11	12	16.28	0.28	11	10
14	12	10	15.44	2.35	12	11
15	12	11	16.19	1.69	12	12
16	13	16	13.11	5.75	13	13
17	13	14	19.1	0.45	13	13
18	13	15	19.85	-0.22	13	13
19	13	17	21.35	-1.54	13	14
20	14	7	14.52	5.84	14	14
21	15	13	19.68	2.6	15	15
22	17	13	21	4.1	17	17
23	17	17	24	1.45	17	17
24	18	19	26.16	0.87	18	19
25	20	20	28.23	1.7	20	20

Se poate defini varianța totală a setului de date ca sumă a varianțelor individuale și deoarece valorile acestor varianțe se află pe diagonala principală a matricii de varianță-covarianță ea va fi numeric egală cu urma acestei matrici și implicit cu suma valorilor proprii ale matricii:

$$\text{Varianța totală} = 20,3 + 24,1 = 44,4$$

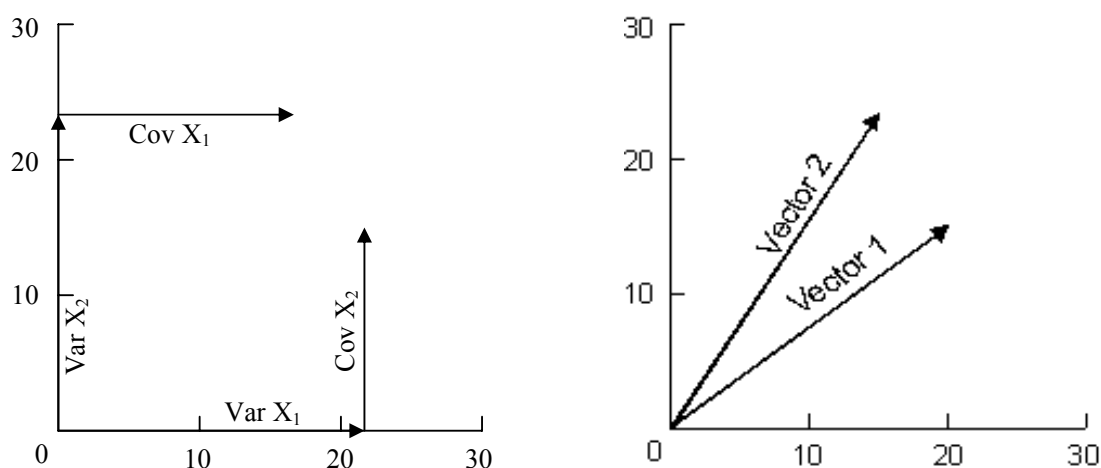


Fig. 72 & 73 Reprezentarea grafică a matricii de varianță-covarianță

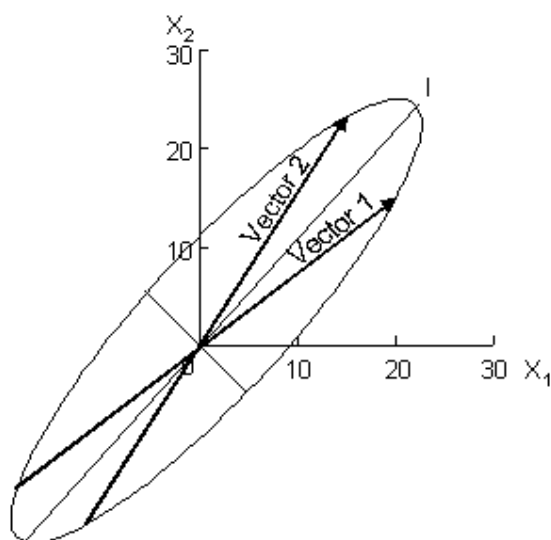


Fig. 74 Elipsa definită de varianța și covarianța datelor din tabelul III.32

La această varianță totală variabila X_1 contribuie cu $20,3/44,4 = 46\%$ iar X_2 cu $24,1/44,4 = 54\%$.

Varianța totală fiind egală cu suma valorilor proprii ale matricii de varianță-covarianță rezultă că axele elipsei ce înglobează toate perechile (X_i, Y_i) reprezintă varianța totală, iar fiecare axă exprimă o anumită parte din ea. Pentru matricea utilizată, axa principală reprezintă $37,9/44,4 = 86\%$ din varianța totală în timp ce a doua axă, corespunzătoare celei de-a doua valori proprii ($\lambda_2 = 6,5$) $6,5/44,4 = 14\%$.

Astfel spus, dacă măsurăm varianța setului de date de-a lungul primei axe principale putem reprezenta 86% din totalul varianței totale. Este evident că cel puțin una din axele principale va fi mai eficientă în exprimarea varianței decât oricare din axele originale și implicit, printre celelalte axe principale se va găsi una mai puțin eficientă decât oricare din axele originale.

Dacă se realizează transformările liniare de forma:

$$Y_1(i) = V_{11}X_1(i) + V_{12}X_2(i) \quad Y_2(i) = V_{21}X_1(i) + V_{22}X_2(i)$$

în care $V_{11}, V_{12}, V_{21}, V_{22}$ sunt elementele celor doi vectori proprii, se crează două noi variabile factorizate: Y_1 care reprezintă $37,9/44,4 = 86\%$ și Y_2 numai $6,5/44,4 = 14\%$ din varianța totală (Tabelul III.32)

Deoarece noile variabile proprii Y_1 și Y_2 sunt măsurate de-a lungul celor doi vectori, ortogonali, corelația dintre ele va fi zero.

Componentele vectorilor proprii ($V_{11}, V_{12}, V_{21}, V_{22}$), coeficienții numerici ai ecuațiilor liniare de generare a noilor variabile sunt ponderile fiecărei variabile pe un anumit factor (ex.: V_{11} este ponderea variabilei X_1 pe "factorul" Y_1).

Dacă este obligatoriu din considerente de eficiență a prelucrării datelor să reducem sistemul nostru la numai o variabilă: dacă renunțăm la una din variabilele originale X_1 sau X_2 pierdem 46% sau 56% din varianța totală. Dacă convertim variabilele originale prin proiectarea pe axele componentilor principali, operând cu Y_1 păstrăm 86% din varianța totală pierzând doar 14%.

b) Influenta covariantei asupra A.C.P.

Eficiența repartizării varianței totale pe un număr de factori mai mic decât cel al variabilelor originale este determinată de intensitatea corelației dintre ele.

Pentru exemplificare, în setul de date brute se realizează o ordonare și o randomizare a valorilor (Tabelul III.32). Se obțin două noi serii de 20 de perechi de valori fiecare cu aceeași varianță dar cu covarianțe diferite.

Reprezentările grafice ale celor două serii de valori ilustrează în raport cu seria inițială a valorilor creșterea corelației în cazul ordonării și reducerea ei în cazul randomizării (Fig. 75 și 76).

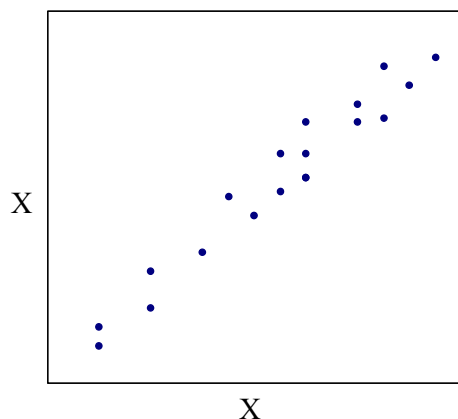


Fig. 75 Datele ordonate

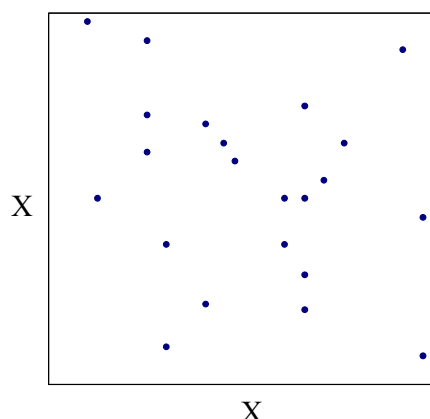


Fig. 76 Datele randomizate

Rezultatele calculului pentru cele două noi seturi de date conduc la următoarele rezultate:

VALORI ORDONATE	VALORI RANDOMIZATE
$\text{cov} = \begin{bmatrix} 20,3 & 21,9 \\ 21,9 & 24,1 \end{bmatrix}$	$\text{cov} = \begin{bmatrix} 20,3 & -0,05 \\ -0,05 & 24,1 \end{bmatrix}$
VALORI PROPRII	
$\lambda_1 = 44,2(99\%)$	$\lambda_1 = 24,3(54,7\%)$
$\lambda_2 = 0,2(1\%)$	$\lambda_2 = 20,1(45,3\%)$
VECTORI PROPRII	

$V_1 = [0,68;0,74]$	$V_1 = [-0,22;0,98]$
$V_2 = [0,74;-0,68]$	$V_2 = [0,98;0,22]$

Reprezentările grafice sunt sugestive pentru ilustrarea eficienței cu care componentii principali pot exprima varianța în cele două cazuri (fig. 77 și 78).

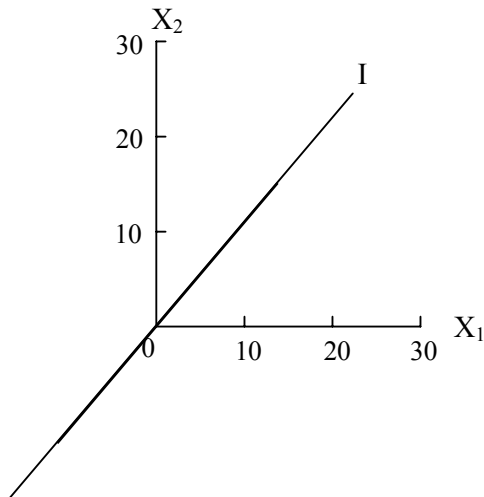


Fig. 77 “Elipsa” valorilor ordonate

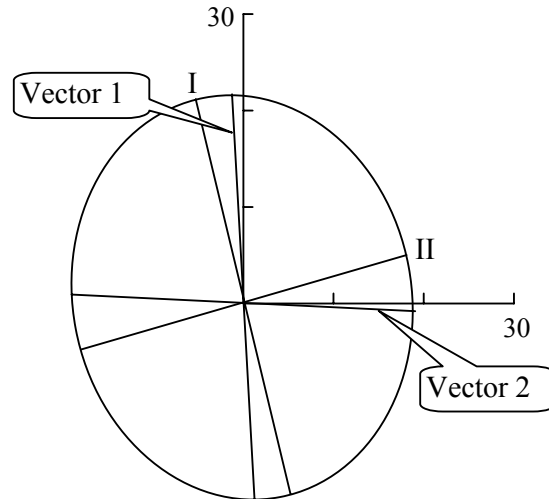


Fig. 78 “Cercul” valorilor randomizate

În cazul valorilor ordonate (Fig. 77), axa principală poate exprima 99% din varianța totală, cea de-a doua fiind așa de scurtă încât practic este imposibil de reprezentat grafic. Dacă renunțăm la ceașta a doua componentă pierderea de varianță a datelor originale este foarte mică.

Se poate reduce deci dimensionalitatea setului de date originale de la doi la unu prin proiectarea pe prima axă principală cu o pierdere de varianță totală de 1%, utilizând relația: $Y_1(i) = V_{11}X_1(i) + V_{12}X_2(i)$.

În cazul valorilor randomizate (Fig. 78), cele două valori proprii sunt practic identice, elipsa devenind cerc. Nici una din axele principale, în aceste condiții, nu va capta mai bine varianța totală în comparație cu variabilele originale. În această situație A.C.P. nu își găsește utilitatea și factorizarea corelației nu își are obiect, corelația lipsind între variabile.

c)Aplicatie

Aplicarea analizei în componente principale este exemplificată prin separarea tipurilor de sedimente pe baza analizelor granulometrice realizate pe 50 de probe recoltate din cinci domenii distincte (I, II, III, IV, V) pentru care s-au determinat șapte fracțiuni granulometrice ($x_1, x_2, x_3, x_4, x_5, x_6, x_7$).

Calculul matricii de varianță-covarianță se face pe date originale, nestandardizate deoarece toate sunt măsurate în aceleași unități de măsură.

Deoarece matricea de covarianță este supradeterminată (suma tuturor fracțiunilor granulometrice este 100), una din valorile proprii teoretic trebuie să fie nulă. Practic ea va fi foarte mică și nu nulă deoarece nu în toate probele suma fracțiunilor componente dau 100 din cauza erorilor de determinare.

Tabelul III.33 Matricea de varianță-covarianță a celor 7 fracțiuni

x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	4,8443					
x_2	-2,6234	468,848				
x_3	-0,0011	81,3941	353,1255			
x_4	-1,5449	-200,2109	-84,6165	130,2741		
x_5	-0,5972	-84,2597	-73,0435	44,7616	30,4350	
x_6	-0,3805	-71,2097	-65,5433	34,9927	23,7565	22,4189
x_7	-0,0222	-57,8578	-56,1533	23,9136	19,3907	17,967

Tabelul III.34 Valorile proprii ale matricii de varianță-covarianță

Vector	Valoare proprie	Varianță totală	Varianță totală cumulată %
I	659,7759	64,18	64,19
II	318,4384	30,98	95,17
III	35,1959	3,42	98,59
IV	6,7528	0,66	99,25
V	3,8193	0,37	99,62
VI	2,3763	0,23	99,85
VII	1,5540	0,15	100,00

Tabelul III.35 Vectori proprii

Var	I	II	III	IV	V	VI	VII
x_1	-0,0019	0,0039	-0,0689	-0,5829	0,7554	0,2793	0,0818
x_2	0,7710	-0,4777	0,3194	0,1885	0,1169	0,1581	0,0326
x_3	0,4167	0,8647	0,0531	0,2119	0,1123	0,1294	0,0421
x_4	-0,3907	0,0761	0,8844	0,0704	0,0490	0,2280	0,0028
x_5	-0,1895	-0,0794	-0,0775	0,6308	0,6255	-0,3240	-0,2401
x_6	-0,1618	-0,0813	-0,1629	0,3330	0,0526	0,2510	0,8723
x_7	-0,1308	-0,0735	-0,2750	0,2570	-0,0815	0,8107	-0,4146

Pe baza elementelor calculate în tabelele III.33, III.34, III.35 se deduc elementele necesare interpretării.

Primii doi componenți principali acumulează 95,17% din varianța totală, încărcarea principală aparținând fracțiunii fine și foarte fine (factorul I: (x_2), (x_3) și (x_4); factorul II: (x_2) și (x_3)).

Diferența dintre cele cinci medii de sedimentare poate fi complet descrisă prin numai doi factori principali. Prin reprezentarea variabilelor transformate în sistemul de referință al factorilor I și II separarea lor este evidentă (Fig. 79).

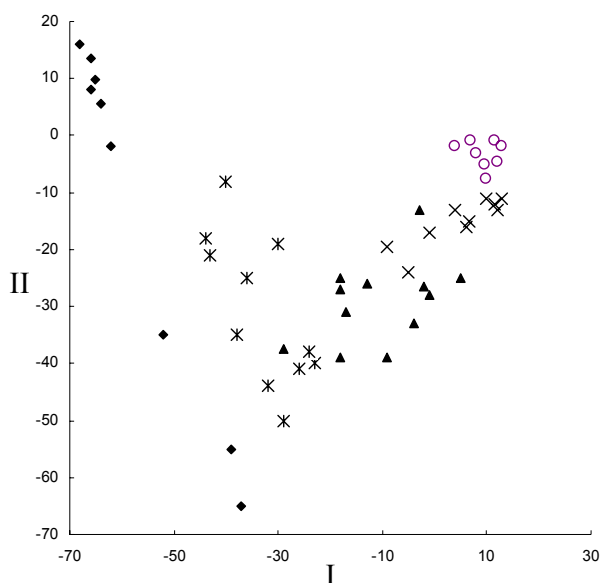


Fig. 79 Reprezentarea valorilor funcție de factorii I, II

Relațiile de transformare sunt:

1) pentru factorul I:

$$YI(i) = -0,0019X_1(i) + 0,7710X_2(i) + 0,4167X_3(i) - 0,3907X_4(i) - 0,1895X_5(i) - 0,1618X_6(i) - 0,1308X_7(i)$$

2) pentru factorul II:

$$YII(i) = -0,0039X_1(i) - 0,4777X_2(i) + 0,8647X_3(i) + 0,0761X_4(i) - 0,0794X_5(i) - 0,0813X_6(i) - 0,0735X_7(i)$$

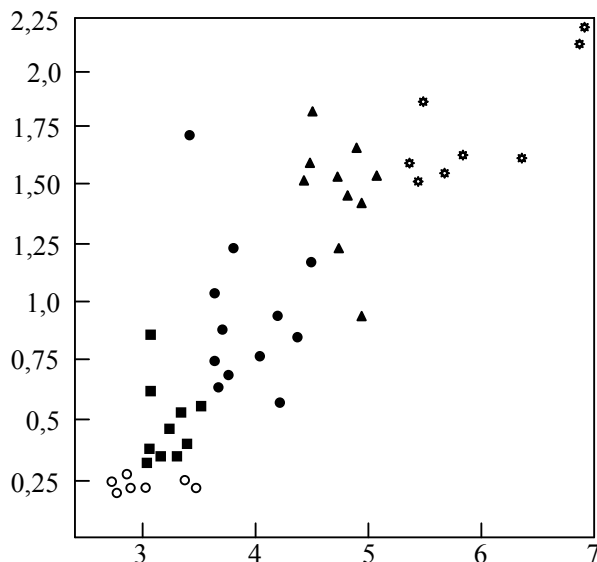
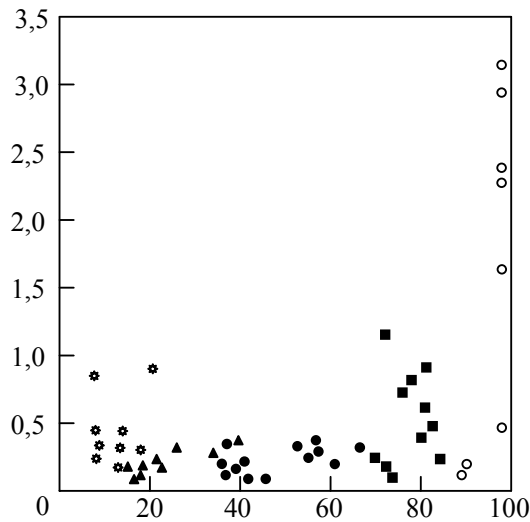


Fig. 80 Separarea funcție de mediană (OX) și gradul de sortare (OY)

Eficiența celor doi factori poate fi comparată cu puterea de separare a tipurilor de sedimente pe baza medianei și gradului de sortare (Fig. 80) sau a procentajului de nisip și raportului dintre nisip fin și nisip foarte fin (Fig. 81).

Fiecare din aceste diagrame sunt aproximativ la fel de eficiente în separarea tipurilor de sedimente.

Avantajul A.C.P este implicată de faptul că din analiza încărcărilor factorilor pentru fiecare variabilă se poate concluziona că sedimentele analizate pot fi considerate o mixtură de material nisipos și silt argilos. Această observație sugerează nu numai un alt mod de a privi sedimentele dar indică și o posibilitate de reducere a fracțiunilor granulometrice la trei, suficiente pentru a permite separarea clară a celor cinci tipuri de sedimente.



Analiza în componenți principali poate fi utilizată în acest mod pentru testarea eficienței relative în separarea tipurilor de sedimente și a altor coeficienți sau parametri statistici (ex.: media, mediana, coeficientul de sortare).

Fig. 81 Separarea tipurilor de sedimente funcție de conținutul în nisip (OX) și raportul nisip fin/nisip foarte fin (OY)

3.2.4. Analiza factorială R-MOD

În analiza factorială R-MOD (R este simbolul matematic al matricii de corelație) relațiile dintre m variabile măsurabile sunt privite ca o reflectare a corelației acestora cu p factori necorelați. Presupunerea uzuală este că $p < m$.

Rezultă că varianța totală are două componente: una determinată de p factori comuni și alta individuală/specifică fiecărei variabile.

Modelul matematic poate fi exprimat sub forma:

$$X_j = \sum_{r=1}^p l_{jr} f_r + \varepsilon_j \quad (\text{III.264})$$

în care:

f_r - factorul comun;

p - numărul de factori;

l_{jr} - încărcarea factorului r pe variabila j ;

ε_j - variația aleatoare specifică variabilei X_j ;

Presupunând o distribuție normală multivariată a variabilelor X_j , varianța și covarianța formează o matrice $[m \times m]$ ale cărei elemente diagonale sunt de forma:

$$s_{jr}^2 = \sum_{r=1}^p l_{jr}^2 + \text{var } \varepsilon_j \quad (\text{III.265})$$

iar restul elementelor de forma:

$$\text{cov}_{jk} = \sum_{r=1}^p l_{jr} l_{kr} \quad (\text{III.266})$$

Dacă notăm matricea varianță-covarianță cu s^2 , cu L matricea $[m \times p]$ a încărcărilor factoriale și cu $[\text{var}(\varepsilon_j)]$ matricea diagonală $[m \times m]$ cu variantele aleatoare specifice fiecărei variabile, avem relația:

$$[s^2] = [L] * [L]^T + [\text{var}(\varepsilon_j)] \quad (\text{III.267})$$

Produsul $[L] * [L]^T$ conduce la o matrice $[m \times m]$ cu p valori proprii pozitive și cu vectorii proprii asociați. Dacă $p = m$, matricea $[\text{var}(\varepsilon_j)] = 0$ și problema este echivalentă cu Analiza în Componente Principali.

Analiza Factorială cere ca numărul de factori să fie mai mic decât numărul de variabile și să fie cunoscut înainte de începerea analizei. Acest lucru presupune deținerea unor informații suplimentare față de datele numerice ce vor fi prelucrate și din care să rezulte numărul de factori ce trebuie extrași. Dacă p nu este cunoscut, împărțirea varianței între factorii comuni și factorii specifici poate fi rezolvată într-un număr practic nelimitat de variante.

a) Diferența operațională dintre A.C.P. și A.F.R.-MOD

Calculul valorilor proprii și vectorilor proprii în analiza factorială R-MOD se face plecând de la matricea de corelație. Acest lucru implică transformarea componentelor principale ale vectorilor în factori.

Vectorii proprii obținuți din matricea de corelație sunt normalizați (adică suma ponderilor este unitară) și pentru a putea realiza analiza factorială trebuie convertită valoarea unitară a vectorului într-o valoare a cărei lungime să reprezinte valoarea proprie corespunzătoare. Acest lucru se face prin multiplicarea fiecărei componente a vectorului propriu normalizat cu rădăcina pătrată a valorii proprii corespunzătoare. Rezultatul este un factor, adică un vector care este ponderat proporțional cu mărimea varianței totale pe care o reprezintă.

Pentru matricea de corelație:

$$COV = \begin{bmatrix} 1,00 & 0,86 \\ 0,86 & 1,00 \end{bmatrix}$$

cu valorile și vectorii proprii:

$$\begin{aligned} \lambda_1 &= 1,86 \text{ și } V_1 = [0,707 \quad 0,707] \\ \lambda_2 &= 0,14 \text{ și } V_2 = [-0,707 \quad 0,707] \end{aligned}$$

factorii ce înglobează varianța ansamblului sunt:

$$\begin{aligned} FACTOR1 &= \begin{bmatrix} 0,707 \times \sqrt{1,86} \\ 0,707 \times \sqrt{1,86} \end{bmatrix} = \begin{bmatrix} 0,964 \\ 0,964 \end{bmatrix} \\ FACTOR2 &= \begin{bmatrix} -0,707 * \sqrt{0,14} \\ 0,707 * \sqrt{0,14} \end{bmatrix} = \begin{bmatrix} -0,264 \\ 0,264 \end{bmatrix} \end{aligned}$$

Verificarea corectitudinii convertirii vectorilor proprii standardizați în factori se face prin însumarea pătratelor ponderilor factoriale care trebuie să fie egale cu valorile proprii:

$$0,964^2 + 0,964^2 = 1,86 \quad \text{și} \quad (-0,264)^2 + 0,264^2 = 0,14$$

Primul factor reprezintă $1,86/2,00=93\%$ din varianța totală a variabilelor originale. Din această varianță $0,964^2/1,86=50\%$ este ponderea variabilei 1 și $0,964^2/1,86=50\%$ este ponderea variabilei 2.

Al doilea factor reprezintă $0,14/2,0=7\%$ din varianța totală a datelor cu $(-0,264)^2/0,14=50\%$ pondere pentru prima variabilă și $0,264^2/0,14=50\%$ pentru a doua.

Cei doi factori redau 100% din varianța totală iar scrierea matricială utilizată pentru exprimarea ponderilor factoriale este:

$$\begin{array}{c} \text{VARIABLE:} \\ 1 \\ 2 \end{array} \begin{array}{c} \text{FACTORI} \\ \text{I} \quad \text{II} \\ \left[\begin{array}{cc} 0,964 & 0,264 \\ 0,964 & 0,264 \end{array} \right] \end{array}$$

Prin însumarea pătratelor ponderilor factoriale pentru fiecare variabilă se obține mărimea totală a varianței reținută de factori care poartă numele de comunalitate. Pentru matricea $[2 \times 2]$ luată ca exemplu, comunalitățile pentru ambele variabile sunt unitare:

$$\text{Variabila 1: } h^2_1 = (0,964)^2 + (-0,264)^2 = 1$$

$$\text{Variabila 2: } h^2_2 = (0,964)^2 + (0,264)^2 = 1$$

Dacă numărul factorilor extrași coincide cu numărul variabilelor, comunalitățile sunt egale cu varianța originală și pentru că se lucrează cu variabile standardizate ea va fi egală cu unitatea.

Dacă se extrag mai puțin de m factori ($m = \text{nr. variabile}$) comunalitățile vor fi subunitare și vor fi un coeficient al eficienței setului de factori relativ la exprimarea varianței setului original de date. Spre exemplu, dacă se retine numai primul factor comunalitățile matricii factorilor sunt:

$$h^2_1 = 0,964^2 = 0,93 \text{ pentru variabila 1;}$$

$$h^2_2 = 0,964^2 = 0,93 \text{ pentru variabila 2.}$$

Mărimea comunalității este dependentă de numărul de factori aleși și aceasta ridică marile probleme ale analizei factoriale.

b) Câți factori trebuie aleși?

Problema alegerii factorilor nu are soluție unică fiind o problemă de opțiune:

- a) psihologii experimentalisti extrag atâția factori cât cere teoria acceptată pentru studiul esanțonului de date;
- b) se extrag atâția factori cât pot fi reprezentați grafic (2 sau 3);
- c) se extrag toți factorii proprii care au valori proprii mai mari ca 1, adică factorii care contin varianțe mai mari decât cele ale variabilelor standardizate.

Dacă pentru reținerea unei mari părți din varianța totală a sistemului este nevoie de mulți factori, modelul analizei factoriale se consideră neadecvat analizei esanțonului de date disponibil.

c) Aplicații

Un exemplu clasic pentru aplicarea analizei factoriale R-MOD este separarea a 25 prisme rectangulare (Tabelul III.35) după formă și mărime (cei doi factori) pe baza unui număr de 7 variabile:

X1 = axa lungă;

X2 = axa intermediară;

X3 = axa scurtă;

X4 = cea mai lungă diagonală;

X5 = (raza sferei circumscrise)/(raza sferei înscrise)

$X6 = (\text{axa lungă} + \text{axa intermediară})/(\text{axa scurtă})$

$X7 = (\text{aria totală}/\text{volumul})$

În tabelele III.35b și III.36 sunt prezentate matricea de corelație, valorile proprii și matricea vectorilor proprii, pentru prelucrare și interpretare fiind reținuți doar primii doi factori (corespunzători forme și mărime) pentru care valorile proprii corespunzătoare sunt supraunitare.

Etapele de prelucrare ale căror rezultate intermediare sunt sintetizate în tabelele III.35, 36 și 37 sunt:

Tabelul III.35 Dimensiunile a 25 de prisme generate aleator

Nr.crt.	X1	X2	X3	X4	X5	X6	X7
1	3,760	3,660	0,540	5,275	9,768	13,741	4,782
2	9,840	9,270	1,510	13,604	9,017	12,668	1,745
3	8,390	4,920	2,540	10,053	3,956	5,237	1,432
4	4,940	4,380	1,030	6,678	6,494	9,059	2,807
5	7,230	2,300	1,770	7,790	4,393	5,374	2,274
6	9,460	7,310	1,040	11,999	11,579	16,182	2,415
7	9,550	5,350	4,250	11,742	2,766	3,509	1,054
8	4,940	4,520	4,500	8,067	1,793	2,103	1,292
9	8,210	3,080	2,420	9,097	3,753	4,657	1,719
10	9,410	6,440	5,110	12,495	2,446	3,103	0,914
11	5,900	5,760	1,550	8,388	5,395	7,497	1,973
12	1,660	1,610	1,570	2,799	1,783	2,087	3,716
13	5,510	1,340	1,270	5,808	4,566	5,382	3,427
14	4,690	3,010	2,170	5,983	2,760	3,554	2,013
15	7,120	5,490	3,680	9,716	2,642	3,430	1,189
16	8,590	2,980	1,170	9,170	7,851	9,909	2,616
17	9,730	1,330	1,000	9,871	9,871	11,064	3,704
18	9,640	9,490	1,030	13,567	13,133	18,519	2,354
19	8,740	7,000	3,310	11,675	3,529	4,757	1,119
20	3,270	0,620	0,440	3,357	7,629	8,838	8,389
21	5,510	3,980	1,300	6,924	5,326	7,304	2,403
22	9,030	7,080	2,590	11,762	4,539	6,217	1,276
23	7,570	7,280	7,070	12,662	1,791	2,101	0,822
24	6,220	6,140	4,520	9,842	2,175	2,732	1,089
25	8,590	4,990	1,340	10,022	7,500	10,162	2,130

Tabelul III.35b Matricea de corelație

Variabilele	X1	X2	X3	X4	X5	X6	X7
X1	1,000						
X2	0,580	1,000					
X3	0,201	0,364	1,000				
X4	0,911	0,834	0,439	1,000			
X5	0,283	0,166	-0,704	0,163	1,000		
X6	0,287	0,261	-0,681	0,202	0,990	1,000	
X7	-0,533	-0,609	-0,649	-0,676	0,427	0,357	1,000

Tabelul III.36 Valorile proprii

Vector	Valoare proprie	Varianță totală	Var.cumulată [%]
I	3,3946	48,4949	48,4949

II	2,805	40,0783	88,5731
III	0,4373	6,2473	94,8204
IV	0,2779	3,9707	98,7911
V	0,0810	1,1565	99,9476
VI	0,0034	0,0487	99,9963
VII	0,0003	0,0037	100,0000

Tabelul III.37 Vectorii proprii

Variabile	I	II	III	IV	V	VI	VII
X1	0,4053	-0,2929	-0,6674	0,0888	-0,2267	0,4098	-0,2782
X2	0,4316	-0,2224	0,6980	-0,0338	-0,4366	0,1443	-0,2540
X3	0,3854	0,3559	0,1477	0,6276	0,5121	0,1875	-0,1081
X4	0,4939	-0,2323	-0,1186	0,2103	-0,1054	-0,5878	0,5359
X5	-0,1277	-0,5751	0,0294	0,1108	0,3890	-0,4232	-0,5562
X6	-0,0968	-0,5800	0,1743	-0,0061	0,3549	0,5003	0,4975
X7	-0,4809	-0,1303	0,0176	0,7353	-0,4553	0,0332	0,0489

1. Calculul ponderilor factorilor comuni prin multiplicarea ponderilor normalizate cu radicalul valorilor proprii:

$$[L]^T = \begin{bmatrix} \text{FactI} \\ \text{FactII} \end{bmatrix} \begin{matrix} X1 & X2 & X3 & X4 & X5 & X6 & X7 \\ \begin{bmatrix} 0,747 & 0,795 & 0,710 & 0,910 & -0,235 & -0,178 & -0,886 \\ 0,491 & 0,373 & -0,596 & 0,389 & 0,963 & 0,971 & 0,218 \end{bmatrix} \end{matrix}$$

2. Calculul comunalităților prin însumarea pătratelor ponderilor factoriale pentru fiecare variabilă prin luarea în considerare a primilor doi factori conduce la:

$$H^2 = \begin{bmatrix} 0,747^2 + 0,491^2 \\ 0,795^2 + 0,373^2 \\ 0,710^2 + (-0,596)^2 \\ 0,910^2 + 0,389^2 \\ (-0,235)^2 + 0,963^2 \\ (-0,178)^2 + 0,971^2 \\ (-0,886)^2 + 0,218^2 \end{bmatrix} = \begin{bmatrix} 0,798 \\ 0,771 \\ 0,860 \\ 0,979 \\ 0,983 \\ 0,976 \\ 0,833 \end{bmatrix} \text{ pentru } \begin{bmatrix} X1 \\ X2 \\ X3 \\ X4 \\ X5 \\ X6 \\ X7 \end{bmatrix}$$

3. Calculul varianței reziduale care exprimă ponderea componentei specifice (ε_j):

$$\text{Rez} = \begin{bmatrix} 1 - H_1^2 \\ 1 - H_2^2 \\ 1 - H_3^2 \\ 1 - H_4^2 \\ 1 - H_5^2 \\ 1 - H_6^2 \\ 1 - H_7^2 \end{bmatrix} = \begin{bmatrix} 0,202 \\ 0,229 \\ 0,140 \\ 0,021 \\ 0,017 \\ 0,024 \\ 0,167 \end{bmatrix} \text{ pentru } \begin{bmatrix} X1 \\ X2 \\ X3 \\ X4 \\ X5 \\ X6 \\ X7 \end{bmatrix}$$

Dacă sunt reținuți m factori dintr-un set de m variabile matricea de covarianță originală $[s^2]$ poate fi generată prin multiplicarea tuturor perechilor de ponderi factoriale și însumarea acestora pentru toți factorii.

Când $p < m$ matricea originală nu poate fi reprodusă exact. Pentru variabilele j și k covarianța reproductibilă este dată de relația:

$$s_{jk}^2 = l_{j1} \times l_{k1} + l_{j2} \times l_{k2} + \dots + l_{jp} \times l_{kp} \quad (\text{III.268})$$

în care l_{j1} este încărcarea variabilei j pe factorul 1. Notând cu L matricea încărcărilor factoriale rezultă că matricea reproductibilă pe baza celor p factori se poate calcula prin:

$$[s^2] = [L] \times [L]^T$$

Reziduul matricii varianță-covarianță poate fi calculat prin diferența:

$$[s^2] - [L] \times [L]^T = [s^2_{\text{rezidual}}] \quad (\text{III.269})$$

Analiza factorială este aplicată cu eficiență în separarea faciesurilor calcaroase. Toomey (1966) a determinat pentru calcarele de Leavenworth (Pensilvanian superior = Carbonifer superior) din nordul regiunii Midcontinent 19 tipuri de constituenți petrografici: calcit spatic, micrit, pellete, trilobiti, ostracode, moluste, brachiopode, spiculi de spongieri, echinoderme, fusulinide, foraminifere mobile, foraminifere încrustate, Tubiphytes, Epimastopore, alge cu structură laminară, granule cu învelis algal și particule de schelete necunoscute. Datele au fost determinate în 33 de probe și pe baza lor au fost delimitate cinci grupuri bine individualizate: grupul fusulinide calcit, grupul micrit, grupul foraminifere mici, grupul cochilii-briozoare și grupul granulelor cu învelis algal, din care primele patru formează un cluster cu coeziunea internă mai mare.

Analiza factorială R-MOD poate fi utilizată pentru separarea cu eficiență maximă și totală obiectivitate a tipurilor de cărbune pe baza parametrilor fizico-chimici care se determină în mod clasic: grosime, greutate specifică, cenușă, umiditate, substanțe volatile, sulf, conținut în carbon, putere calorică etc.

3.2.5. Rotatia factorilor

Deși analiza factorială poate reduce dimensionalitatea unei probleme pentru a o face mai ușor de studiat, semnificatia factorilor poate fi dificil de dedus. Această dificultate poate fi determinată de faptul că poziția a p axe factoriale ortogonale într-un spațiu m dimensional ($p < m$) sunt fortate de $m - p$ axe inutile care de asemenea trebuie plasate ortogonal în spațiul de probare.

Deoarece avem nevoie numai de p axe factoriale, după eliminarea axelor inutile pare posibil și avantajos să rotim axele factoriale pentru a găsi o poziție care să maximizeze varianța încărcărilor factoriale.

Metoda KAISER-VARIMAX are ca obiectiv rotirea fiecărei axe în poziția în care proiecția fiecărei variabile să se plaseze în vecinătatea extremității sau originii sistemului de axe factoriale. Metoda operează prin ajustarea încărcărilor factoriale astfel încât ele să fie ori aproape de ± 1 , ori aproape de zero. În acest mod pentru fiecare factor vor fi câteva ponderi semnificative iar restul aproximativ nule.

Totuși, în unele cazuri, rotirea rigidă a axelor prin păstrarea ortogonalității nu va îmbunătăți sau chiar poate conduce la rezultate confuze. Aceste situații pot indica o corelare a factorilor (factori oblici) sau neadecvarea modelului factorial pentru analiza sistemului.

Criteriul VARIMAX implică maximizarea varianței încărcărilor factoriale. Se poate defini varianța încărcărilor pe factorul k sub forma:

$$s_k^2 = \frac{p \sum_{j=1}^m \left(\frac{l_{jp}^2}{h_j^2} \right) - \left(\sum_{j=1}^m \frac{l_{jp}^2}{h_j^2} \right)^2}{p^2} \quad (\text{III.270})$$

Cantitatea care trebuie minimizată este:

$$V = \sum_{k=1}^p s_k^2 \quad (\text{III.271})$$

Varianța este calculată din încărcările factoriale l_{jp} care sunt corectate prin divizarea lor cu comunalitățile h_j^2 , astfel încât numai partea comună a varianței fiecărei variabile este luată în considerare îndepărtând constrângerile impuse de cele $m - p$ componente (necesare pentru luarea în considerare a întregii varianțe a sistemului).

Maximizarea varianței implică mărirea domeniului încărcărilor care conduce la "extremizarea" ponderilor.

Rotatia factorilor se face iterativ. Două axe sunt ajustate simultan considerând restul axelor stationare. După ce toate axele au fost ajustate procesul este reiterat până când creșterea varianței încărcărilor la fiecare iteratie rămâne sub o anumită valoare.

Aplicatie. Rotatia axelor cu metoda Varimax. Considerăm cazul ponderilor factoriale pentru cei doi factori utilizați în separarea prismelor (notate cu 1,2,...) pe baza formei și mărimei.

După rotație, poziția relativă a variabilelor nu se schimbă ci numai raportul față de axele factoriale. Lungimea vectorilor este funcție de proporția și varianța originală a fiecărei variabile preluată de axele factoriale. În exemplul prezentat, cei doi factori preluând 88,59% din varianța sistemului, lungimea vectorilor de poziție este aproape unitară.

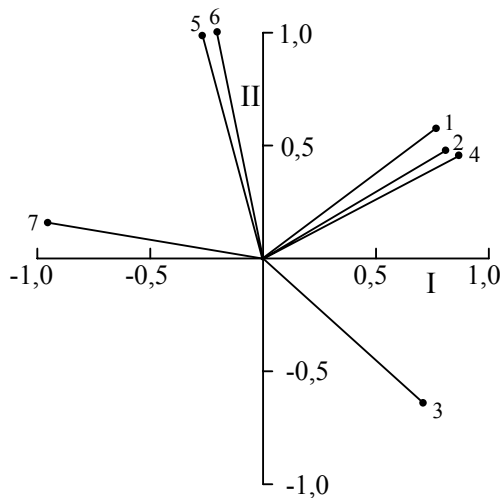


Fig. 82 Încărcările factoriale înainte de rotirea axelor

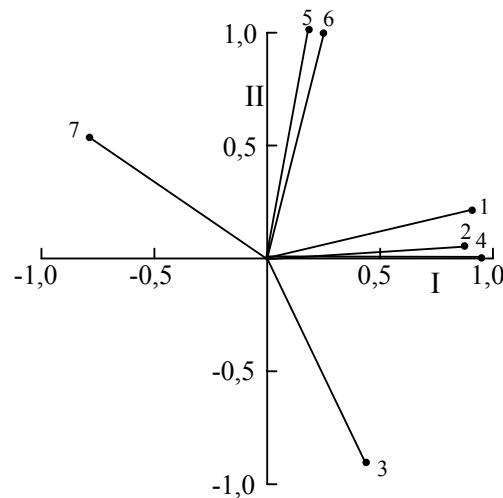


Fig. 83 Încărcările factoriale după rotirea axelor

Reprezentarea grafică a proiecțiilor factoriale (rotite sau nerotite) este mult mai complicată decât proiectarea pe axele componentelor principali. Componentii principali sunt transformări liniare și deci putem proiecta datele originale pe axele principale.

În analiza factorială proiecțiile datelor originale (=variabile măsurabile) pe axele factoriale reprezintă estimările contribuțiilor diferitelor factori asupra fiecărei observație (=proba în care se

execută determinarea celor m variabile). Deoarece factorii înșiși sunt estimați din aceleași date, calculul proiecțiilor factoriale este un proces circular, iar rezultatele nu sunt unice.

Calculul proiecțiilor factoriale este esențial pentru studiile geologice. Pentru explicitarea modului de calcul ne vom referi la setul inițial de date $[X]$ care este o matrice $[m \times n]$ (m - număr variabile; n - număr de probe).

În cazul ACP se poate calcula o matrice a proiecțiilor factoriale $[F]$ prin multiplicarea matricii de date $[X]$ cu matricea încărcărilor factoriale $[L]$:

$$[X] \times [L] = [F] \quad (\text{III.272})$$

Dacă reținem p factori, matricea încărcărilor $[L]$ va fi $[m \times p]$, iar matricea proiecțiilor va fi $[n \times p]$.

Se știe că variabilele originale nu reprezintă numai efectul factorilor comuni dar au și o componentă specifică $\varepsilon_{(j)}$. Matricea proiecțiilor calculată în acest mod va reflecta parțial structura covarianței datelor originale, în măsura în care factorii preiau această covarianță.

Influența variației specifice $\varepsilon_{(j)}$ trebuie eliminată pentru realizarea proiecțiilor factoriale. Acest lucru se realizează prin multiplicarea ecuației (III.273) cu inversul matricii de covarianță:

$$[X] \times [s^2]^{-1} \times [L] = [F'] \quad (\text{III.273})$$

Deoarece inversarea matricii de covarianță este laborioasă calculul nu se realizează direct din această ecuație. Se calculează în primul rând matricea $[s]$ prin înmulțirea matricii încărcărilor factoriale cu transpusa ei:

$$[L]^T \times [L] = [S] \quad (\text{III.274})$$

Matricea obținută se inversează și se multiplică cu $[L]$ obținându-se matricea coeficienților proiecțiilor factoriale $[B]$:

$$[L] \times [S]^{-1} = [B] \quad (\text{III.275})$$

Matricea proiecțiilor factoriale se obține din produsul cu matricea datelor originale:

$$[X] \times [B] = [F'] \quad (\text{III.276})$$

Sintetizând în termenii matricilor încărcărilor factoriale, operația se poate scrie:

$$[X] \times [B] = [F'] \quad (\text{III.277})$$

$$[X] \times [L] \times [S]^{-1} = [F'] \quad (\text{III.278})$$

$$[X] \times [L] \times ([L]^T \times [L])^{-1} = [F'] \quad (\text{III.279})$$

Aceeași procedură este utilizată pentru a obține proiecțiile factoriale în cazul axelor rotite sau nerotite. De reținut că matricea $[X]$ conține variabilele standardizate și nu pe cele initiale din selecția de valori ca în A.C.P., deoarece A.C.P. calculează încărcările componentilor principali plecând de la matricea de varianță-covarianță în timp ce încărcările factoriale se calculează plecând de la matricea de corelație.

Problema specificării numărului de factori p care trebuie reținuți este critică. Numărul lor afectează mărimea matricii reproduse și reziduale, comunalitățile și încărcările factoriale specifice (ε_j). Încărcările factoriale comune nu sunt afectate.

Astfel, dacă $p = 2$ și factorii sunt extrași din datele originale, încărcările pe factorii I și II nu sunt modificate dacă se extrage și un al treilea factor. Totuși, dacă extragem și rotim doi factori, ponderile factoriale pot fi radical diferite de cele obținute dacă extragem și rotim trei factori din setul de date. Când sunt extrași doi factori ei nu introduc constrângeri la rotație ca atunci când sunt extrași trei. Metoda Varimax păstrează orogonalitatea factorilor.

Există metode de rotație a axelor factoriale care nu păstrează ortogonalitatea, conducând la rezultate mai ușor de prelucrat deoarece se pot obține mai multe ponderi factoriale extreme. Din punct de vedere interpretativ apar contradicții cu principiile metodei care presupune că factorii comuni sunt necorelați, adică ortogonali. Renunțând la restricția ortogonalității se admite intercorelația dintre factori.

Dacă factorii sunt corelați între ei, relațiile între variabilele originale și factorii identificați sunt mult mai complexe decât în modelul adoptat deoarece interacțiunile sunt atât între perechile de variabile cât și între perechile de factori. Prezența corelațiilor între factori conduce la ideea că există alți SUPERFACTORI independenți care acționează asupra variabilelor măsurate și factorilor comuni separați la primul nivel. Soluțiile de rotație oblice introduc mai multă subiectivitate în interpretare și trebuie abordate cu multă atenție.

3.2.6. Analiza factorială Q-MOD

Analiza factorială Q-MOD, introdusă în geologie de Imbrie și Purdy (1962), este o a doua formă de analiză factorială în care rolul valorilor (sau probelor) și al variabilelor se schimbă. Prin această analiză se urmărește evidențierea corelațiilor dintre probe, având ca obiectiv gruparea lor într-o structură dendritică din care să poată fi deduse relațiile dintre ele.

În 1962, când au introdus analiza Q-MOD în cercetarea geologică, Imbrie și Purdy au utilizat-o pentru realizarea unui sistem obiectiv de clasificare a sedimentelor carbonatice actuale din Great Bahama Bank. Metoda a mai fost utilizată de Harbaugh și Demirmen (1964) pentru a discerne limitele de facies din calcarele de Americus.

Primul pas în analiza factorială Q-MOD este crearea unei matrici de similaritate $[n \times n]$ în care n este numărul de probe în care se face determinarea diferitelor "m" caracteristici geologice, calitative sau cantitative. Măsura similarității poate fi oricare dintre coeficienții de similaritate definiți în capitolul III.2. cu valori cuprinse în intervalul $[-1, +1]$. Cel mai utilizat coeficient de similaritate în analiza Q-MOD este coeficientul cosinus θ .

Analiza factorială Q-MOD are ca obiectiv identificarea unui hiperelipsoid n -dimensional care este definit prin corelațiile dintre cei n vectori care reprezintă cele n probe. Fiecare vector este determinat prin cele m variabile care au fost măsurate în fiecare probă și din acest motiv dimensionalitatea problemei nu depășește numărul variabilelor (m).

Al doilea pas este identificarea principalelor axe ale hiperelipsoidului prin extragerea valorilor și vectorilor proprii. Deoarece vor fi reținute, de fiecare dată, mai puțini factori decât numărul probelor, nu este necesară extragerea tuturor valorilor și vectorilor proprii, acest lucru reducând mult din timpul de calcul.

În al treilea pas se realizează maximizarea încărcărilor factoriale prin rotația axelor factoriale. Rotația axelor se poate face până ce fiecare factor coincide cu una din probele ce alcătuiesc selecția de date. Pe lângă tehnicile ce păstrează ortogonalitatea axelor factoriale după rotație, analiza factorială Q-MOD apelează și la rotația ce conduce la oblicitatea axelor factoriale cu implicațiile semnalate în paragraful anterior.

Aplicație. Ca un exemplu al aplicării analizei Q-MOD, prezentăm în continuare o analiză petrografică. Tabelul II.37 conține componenții chimici majori a 20 de eșantioane (1-Sienit, 2-Sienit, 3-Sienit, 4-Monzonit, 5-Diorit, 6-Diorit, 7-Diorit, 8-Diorit cuarțitic, 9-Gabrou, 10-Gabrou, 11-Norit, 12-Norit, 13-Gabrou cu hipersten, 14-Gabrou cu hipersten, 15-Sienit, 16-Sienit cuarțitic, 17-Sienit alterat, 18-Monzonit, 19-Monzonit, 20-Diabaz). Prin analiza Q-MOD se urmărește plasarea fiecărei probe în poziția proprie a seriei diferențiate de roci magmatice.

Plasarea probelor în succesiunea firească, determinată de compoziția chimică, se realizează prin utilizarea încărcărilor factoriale ce exprimă varianța ansamblului petrografic probat. Deoarece

valorile vor fi standardizate, vectorii definiți vor avea lungimi unitare și probele vor fi plasate pe circumferința unui cerc cu rază unitară. Unghiurile dintre acești vectori sunt o măsură a similarității dintre probe. Pentru evaluarea matricii de similaritate, ca rezultat al primei etape de prelucrare se utilizează coeficientul de cos θ , rezultatul fiind consemnat în tabelul III.38 (ANEXA 1).

Identificarea axelor este limitată la primii doi factori care asigură în etapa finală o reprezentare grafică simplă. Încărcările factoriale pentru fiecare probă sunt sintetizate în tabelul III.39.

Tabelul III.39 Încărcările factoriale pentru primii doi factori (I și II)

Proba	I	II	Proba	I	II
1	0,9948	-0,0910	11	0,9833	0,1202
2	0,9918	-0,1223	12	0,9890	0,1259
3	0,9958	-0,0587	13	0,9721	0,1719
4	0,9989	-0,0126	14	0,9561	0,02323
5	0,9963	-0,0191	15	0,9918	-0,1257
6	0,9904	0,1188	16	0,9844	-0,1665
7	0,9959	-0,0838	17	0,9866	0,0783
8	0,9996	0,0010	18	0,9950	-0,0870
9	0,9983	0,0204	19	0,9945	-0,0946
10	0,9978	0,0223	20	0,9981	-0,0161

Rotirea axelor prin metoda Varimax maximizează varianța încărcărilor factoriale (Tabel III.40) care permit reprezentarea grafică cea mai sugestivă a grupării celor 20 de probe funcție de afinitățile lor chimice (Fig. 84).

Tabelul III.40 Încărcările factoriale după rotație (pentru factorii I și II)

Proba	I	II	h^2	Proba	I	II	h^2
1	0,7851	0,6177	0,9980	11	0,6316	0,7632	0,9814
2	0,8044	0,5959	0,9986	12	0,6319	0,7712	0,9940
3	0,7636	0,6418	0,9950	13	0,5879	0,7930	0,9745
4	0,7342	0,6774	0,9980	14	0,5348	0,8259	0,9681
5	0,7368	0,6709	0,9929	15	0,8068	0,5904	0,9995
6	0,6377	0,7671	0,9950	16	0,8295	0,5556	0,9968
7	0,7809	0,6236	0,9988	17	0,6628	0,7350	0,9796
8	0,7254	0,6878	0,9993	18	0,7825	0,6207	0,9976
9	0,7111	0,7009	0,9970	19	0,7873	0,6148	0,9979
10	0,7094	0,7020	0,9960	20	0,7360	0,6744	0,9965

În final trebuie remarcat că analiza factorială Q-MOD are același obiectiv ca orice analiză a grupărilor însă cu o eficiență mai mare datorată reducerii timpului de calcul, în condițiile în care se apelează la mijloacele automate.

Eficiența metodei este sporită și de faptul că ea este aplicabilă și în condițiile în care matricea de similaritate conține și coeficienți negativi, caz în care analiza factorială R-MOD nu este utilizabilă.

Tabel III.37 Principalii oxizi din 20 de eșantioane recoltate dintr-o serie magmatică

Nr. probă	X1=SiO ₂	X2=Al ₂ O ₃	X3=Fe ₂ O ₃	X4=FeO	X5=MgO	X6=CaO	X7=Na ₂ O	X8=K ₂ O
1	61,7	15,1	2,0	2,3	3,7	4,6	4,4	4,5
2	58,3	17,9	3,2	1,7	1,5	3,7	5,9	5,3
3	51,2	17,6	3,5	4,3	3,2	4,5	5,7	4,4
4	54,4	14,3	3,3	4,1	6,1	7,7	3,4	4,2
5	58,0	15,7	0,7	2,8	5,0	10,9	3,0	3,2
6	46,6	15,9	2,9	10,0	7,0	9,6	2,7	0,7
7	58,0	17,3	2,2	3,8	2,2	4,3	4,3	4,1
8	55,5	16,5	1,7	4,6	6,7	6,7	3,2	2,5
9	55,4	15,3	2,7	5,5	5,8	9,9	2,9	1,5
10	55,9	13,5	2,7	5,9	6,5	8,9	2,4	1,7
11	47,2	14,5	1,6	13,8	5,2	8,1	3,1	1,2
12	48,2	18,3	1,3	6,1	10,8	9,4	1,3	0,7
13	44,8	18,8	2,2	4,7	11,3	14,6	0,9	0,1
14	47,0	14,1	0,8	15,0	16,0	2,3	0,4	1,7
15	59,8	17,3	3,6	1,6	1,2	3,8	5,0	5,1
16	66,2	16,2	2,0	0,2	0,8	1,3	6,5	5,8
17	50,0	9,9	3,5	5,0	11,9	8,3	2,4	5,0
18	57,4	18,5	3,7	2,1	1,7	6,8	4,5	3,7
19	59,8	15,3	3,8	3,3	2,2	3,9	3,0	4,4
20	52,2	18,2	3,3	4,4	4,7	6,5	4,6	1,9

3.3. Modelarea matematică a corelațiilor substanțiale

Exprimarea într-o formă sintetică a sistemului de corelații între caracteristicile unui proces este obiectivul final al oricărei cercetări sistematice. Modelul operational rezultat din formalizarea matematică a sistemului de corelații este o construcție intelectuală care înlocuiește "vizibilul complicat" (procesele fizico-chimice studiate) cu "invizibilul" (ecuații, sisteme etc.) ușor de manevrat.

În funcție de calitatea descrierii (completă sau de tendință), scara modelului (atomică, macroscopică), caracterul intrinsec (determinist, probabilist, linear, nelinear), structura matematică (algebric, în diferențe finite sau element finit, diferențial) există o diversitate de modele aplicabile studierii proceselor geologice. În continuitate imediată cu demersul statistic de prelucrare a informațiilor geologice prezentăm cea mai simplă modalitate de formalizare empirică a relațiilor dintre variabilele unui proces geologic complex: modelarea lineară a corelațiilor substanțiale.

3.3.1. Model liniar de o singură variabilă independentă

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:

$$y = \alpha_0 + \alpha_1 x + e \quad (\text{III.277})$$

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 :

- a) ambele variabile (x și y) sunt afectate de erori întâmplătoare (**Fig. 85**);
- b) variabila independentă (x) este cunoscută riguros, iar variabila dependentă (y) este afectată de erori distribuite normal (**Fig. 86**).

Modelul a) este adecvat studierii corelației conținuturilor de Au și Ag dintr-un zăcământ 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.

Pentru studiul complet al corelației liniare între două variabile este necesară parcurgerea unui număr de patru etape de prelucrare.

4.1.1. Analiza grafică a corelației

Reprezentarea grafică a bidimensionale a variabilelor este cea mai rapidă formă de identificare calitativă a corelației. Ea se poate analiza variante: diagrama de împrăștiere, stereograma și corelație.

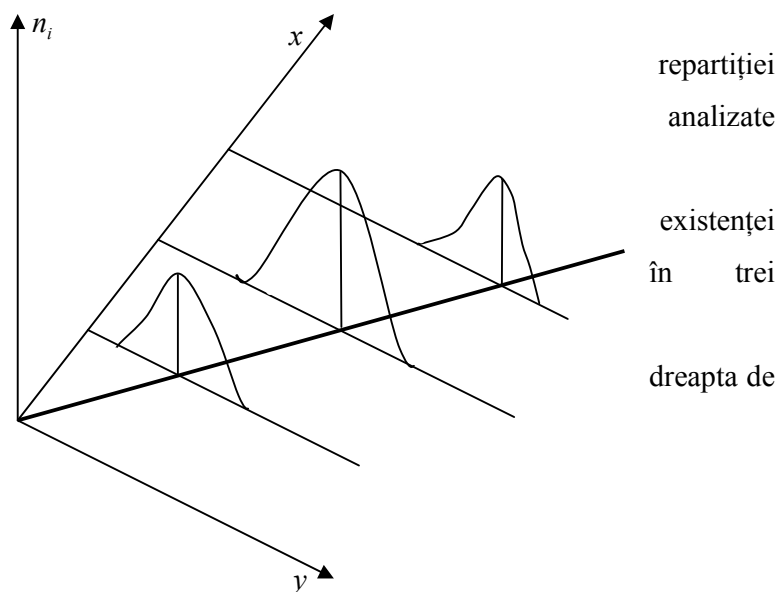


Fig. 86 Model liniar cu o singură variabilă (y) afectată de erori

a) Diagrama de împrăștiere

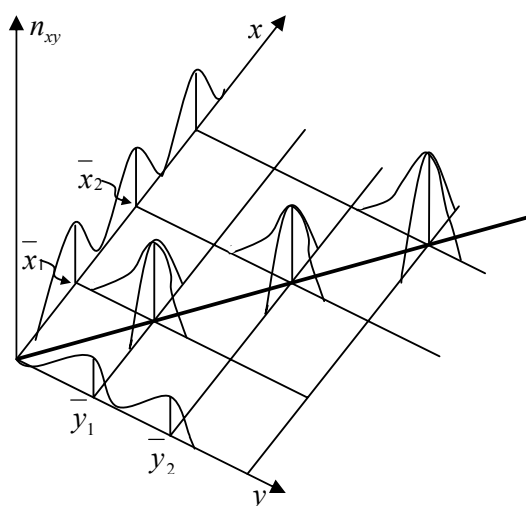


Fig. 85 Model liniar cu ambele variabile (x,y) afectate de erori

Diagrama de împrăștiere este cea mai simplă formă de reprezentare grafică în care utilizând un sistem de referință rectangular, fiecare pereche de valori măsurată (x_i, y_i) se materializează printr-un punct. Se obține în acest mod o mulțime de puncte a cărei configurație geometrică sugerează prezența sau absența corelației între cele două variabile (**Fig. 87**).

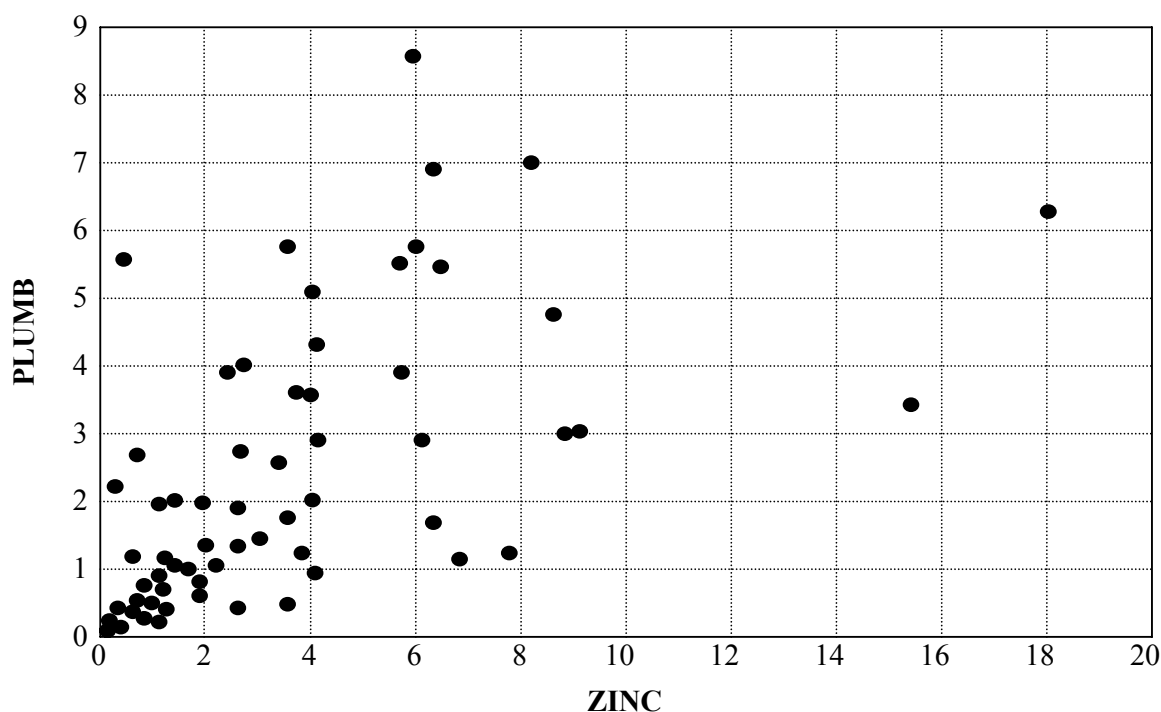


Fig. 87 Diagrame de împrăștiere

Punctele pot avea o distribuție: haotică - corelația între cele două variabile fiind nulă, concentrată pe o zonă alungită rectilinie - corelația fiind de tip liniar sau concentrată pe o zonă alungită curbilinie, situație în care se presupune existența unei corelații neliniare între cele două variabile.

Diagrama de corelație poate fi realizată și cu valori standardizate, variantă recomandată atunci când valorile sunt exprimate în unități de măsură diferite și au amplitudini de selecție disproporționate.

b)Stereograma

Stereograma este o reprezentare tridimensională care se bazează pe gruparea bidimensională a valorilor celor două variabile după aranjarea în ordine crescătoare a variabilei independente. Intervalele de grupare care formează compartimentele tabelului de corelație (Tabel III.41), pentru ambele variabile se stabilesc după aceleași criterii ca cele stabilite pentru descrierea univariată.

Tabelul III.41 Tabel de corelație pentru două variabile (x, y)

y	y_1	y_2	...	y_k	...	y_n
x						
x_1	$n_{x_1y_1}$	$n_{x_1y_2}$	...	$n_{x_1y_k}$	...	n_{x_1}
x_2	$n_{x_2y_1}$	$n_{x_2y_2}$	...	$n_{x_2y_k}$	...	n_{x_2}
...	...	...	...	...	...	...
x_l	$n_{x_ly_1}$	$n_{x_ly_2}$	...	$n_{x_ly_k}$	...	n_{x_l}
xy	n_{y_1}	n_{y_2}	...	n_{y_k}	...	n

În tabelul de corelație apar trei tipuri de frecvențe:

- 1) frecvența valorilor perechi ($n_{x_iy_i}$) reprezintă numărul de perechi pentru fiecare interval de grupare.
- 2) frecvențe parțiale după variabila X (n_{x_i}) care reprezintă numărul de valori ale variabilei Y corespunzătoare unei valori x_i sau valorii centrale a intervalului i , x_{c_i} , care se calculează însumând frecvențele perechilor de valori de pe un rând al tabelului III.41.

$$n_{x_i} = \sum_{j=1}^k n_{x_iy_j} \quad (i = 1, 2, \dots, l) \quad (\text{III.278})$$

3) frecvențele parțiale după variabila Y (n_{y_i}) se evaluează în mod analog pe coloanele tabelului III.41.

$$n_{y_i} = \sum_{j=1}^l n_{x_j y_i} \quad (i = 1, 2, \dots, k) \quad (\text{III.279})$$

Stereograma se obține prin construirea pentru fiecare compartiment al tabelului de corelație a unui paralelipiped având înălțimea proporțională cu frecvențele perechilor de valori. Suprafața care îmbracă stereograma poartă denumirea de suprafață de frecvență și oferă o imagine globală a corelației între cele două variabile într-un spațiu tridimensional.

c) Dreapta de corelație

Dreapta de corelație reprezintă grafic tendința pe care o urmează media unei variabile în comparație cu valorile celeilalte variabile. Se construiesc două drepte de corelație pentru fiecare cuplu de două variabile (x, y):

a) dreapta de corelație corespunzătoare modelului $y = f(x)$ în care pentru fiecare x_i se determină și se reprezintă valoarea medie (**Fig. 89**).

b) dreapta de corelație corespunzătoare modelului $x = f(y)$ în care pentru fiecare valoare y_i se calculează și se reprezintă grafic (**Fig. 90**).

Linia în jurul căreia se grupează punctele se numește linie de regresie și pentru foarte multe caracteristici geologice este rectilinie. Raporturile spațiale dintre cele două drepte de regresie ($x = f(y)$ și $y = f(x)$) exprimă intensitatea corelației dintre variabilele analizate:

- 1) independența, dacă cele două linii de regresie sunt ortogonale (Fig. 91a);
- 2) dependența totală, dacă cele două linii de regresie coincid (Fig. 91b);
- 3) dependența intermediară, dacă cele două linii de regresie formează un anumit unghi, unghi a cărui mărime este invers proporțională cu intensitatea corelației (nulă când unghiul este de 90°).

Cele trei modele de reprezentare grafică a distribuției bidimensionale a unui cuplu de variabile geologice exprimă doar calitativ intensitatea corelației, care poate fi cuantificată prin intermediul unor parametri.

4.1.2. Evaluarea intensității corelației liniare

Din reprezentările grafice se pot deduce la nivel calitativ inexistența corelației sau existența unei corelații directe sau inverse. Cele două variabile sunt corelate direct dacă valorile mari ale uneia tind să se asocieze cu cele mari ale celeilalte. In rocile poroase, porozitatea și permeabilitatea sunt un exemplu tipic de variabile pozitiv corelate. Două variabile geologice sunt corelate negativ dacă valorile mari ale uneia tind să se asocieze cu valorile mici ale celeilalte. Corelații negative se stabilesc de obicei între concentrațiile a două elemente majore, de exemplu în rocile dolomitice continutul în calciu este în mod normal corelat negativ cu continutul de magneziu.

Sub aspect cantitativ, intensitatea corelației liniare se poate cuantifica prin intermediul coeficientului de corelație Pearson și a coeficientului de corelație a rangurilor.

a) Coeficientul de corelație Pearson

Coeficientul de corelație este cel mai utilizat parametru pentru cuantificarea intensității corelației liniare a două variabile și se calculează cu relația:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \approx \frac{\sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^n (x_i - m_x)^2 \sum_{i=1}^n (y_i - m_y)^2}} = r \quad (\text{III.280})$$

Coeficientul de corelație (ρ) are valori cuprinse între -1 și +1, indiferent de amplitudinea selecției de date. Valorile extreme ale coeficientului de corelație liniară indică o aliniere perfectă a punctelor într-o diagramă de împrăștiere de-a lungul unei drepte fie cu panta pozitivă ($\rho = 1$; corelație pozitivă), fie cu panta negativă ($\rho = -1$; corelație negativă).

Pentru valori $|r| < 1$ (r fiind estimatorul lui ρ), distribuția punctelor se abate de la linia dreptei devenind din ce în ce mai difuză cu cât $|r|$ descrește de la 1 spre 0.

Valoarea coeficientului de corelație este puternic influențată de existența perechilor aberante de puncte. O bună aliniere a câtorva valori extreme poate crește foarte mult valoarea coeficientului de corelație pentru două variabile slab corelate și invers, o bună corelație poate fi "distrușă" de slaba aliniere a câtorva valori extreme.

Aplicatie. Pentru analiza corelației între continuturile în Au și Ag din zăcămintul Căvnic filonul 80 s-a evaluat un coeficient de corelație $r_1 = 0,64$ cu luarea în considerare a tuturor valorilor selecției în care era inclusă și o pereche de valori afectată de erori de măsurare (Fig. 92). Prin eliminarea acestei singure perechi de valori și recalcularea coeficientului de corelație s-a obținut $r_2 = 0,84$.

Dacă relația dintre două variabile nu este lineară, coeficientul de corelație (r) poate avea o valoare foarte mică. Din acest motiv este deseori util să se suplimenteze utilizarea lui cu cea a coeficientului de corelație a rangurilor.

b) Coeficientul de corelație a rangurilor

Coeficientul de corelație a rangurilor (ρ_r) se calculează aplicând formula de calcul a coeficienților de corelație Pearson rangurilor valorilor variabilelor.

$$\rho_r = \frac{\sigma_{R_{xy}}}{\sigma_{R_x} \sigma_{R_y}} \approx \frac{\sum_{i=1}^n (R_{x_i} - m_{R_x})(R_{y_i} - m_{R_y})}{\sqrt{\sum_{i=1}^n (R_{x_i} - m_{R_x})^2 \sum_{i=1}^n (R_{y_i} - m_{R_y})^2}} = r_i \quad (\text{III.281})$$

în care:

R_{x_i}, R_{y_i} - rangul valorii x_i respectiv y_i ;

$\sigma_{R_x}, \sigma_{R_y}$ - abaterea standard a rangurilor valorilor variabilelor x , respectiv y ;

m_{R_x}, m_{R_y} - media rangurilor valorilor $R_{x_1}, \dots, R_{x_n}$, respectiv $R_{y_1}, \dots, R_{y_n}$.

O mare diferență între ρ_r și ρ poate fi deseori determinată de prezența unei perechi de valori extreme. Spre deosebire de coeficientul de corelație (r), coeficientul de corelație a rangurilor (r_r) nu este atât de sensibil la perechi extreme de valori. O valoare mare a coeficientului de corelație a rangurilor și una mică a coeficientului de corelație Pearson poate fi datorată faptului că un număr redus de perechi aberante afectează buna corelație a variabilelor studiate. Dacă coeficientul de corelație a rangurilor este mare și coeficientul de corelație Pearson mic este posibilă o "îmbunătățire" falsă a corelației prin prezenta câtorva valori extreme bine "aliniată".

Pentru situația prezentată anterior valorile corespunzătoare ale coeficientului de corelație a rangurilor sunt: $r_{r_1} = 0,80$ înainte de eliminarea valorii extreme și $r_{r_2} = 0,79$, eliminarea valorii aberante având o influență mult mai mică asupra coeficientului de corelație a rangurilor decât asupra coeficientului de corelație r .

Diferența dintre r și r_r poate fi revelatoare și asupra altui aspect al corelației între cele două variabile: cel al liniarității. Dacă $r_r = +1$, adică rangurile celor două variabile sunt identice, valorilor mari ale variabilei x le corespund valori mari ale variabilei y , corelația are intensitate maximă dar

ea nu este obligatoriu de tip linear. Neliniaritatea corelației este evidentiată de valorile mici ale coeficientului de corelație (r).

4.1.3. Testarea adecvării modelului liniar

Adecvarea unui model liniar este sintetizată în evaluarea semnificației statistice a coeficientului de corelație care se poate realiza în două etape succesive: cea a acceptării (funcție de valoarea calculată) existenței unei corelații liniare și cea de evaluare a incertitudinii asupra intensității acesteia.

Testarea statistică a existenței corelației liniare se poate realiza cu ajutorul testului STUDENT aplicat ipotezelor statistice:

$$\begin{cases} H_0 : \rho = 0 & (\text{absența corelației liniare}) \\ H_1 : \rho \neq 0 & (\text{prezența corelației liniare}) \end{cases}$$

Pentru testarea inexistenței corelației ($\rho = 0$) se calculează valoarea:

$$t_{\text{exp}} = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}} \quad (\text{III.282})$$

care se compară cu valorile repartiției STUDENT $t(\alpha, \nu)$ cu $\nu = n - 2$.

În alternativă $t_{\text{exp}} < t(\alpha, \nu)$ se acceptă ipoteza absenței corelației liniare între cele două variabile. Dacă $t_{\text{exp}} > t(\alpha, \nu)$, din punct de vedere statistic se admite existența unei corelații liniare între cele două variabile și se trece la etapa de evaluare a incertitudinii asupra valorii r calculate.

Calculul intervalului de încredere pentru valoarea coeficientului de corelație ρ se poate realiza utilizând variabila cu repartiție normală propusă de Fisher:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r} \quad (\text{III.283})$$

Pentru calculul intervalului de încredere al coeficientului de corelație (ρ) se utilizează relațiile:

$$r_{\text{inf}} = \frac{e^{2z_1} - 1}{e^{2z_2} + 1} < \rho < r_{\text{sup}} = \frac{e^{2z_1} - 1}{e^{2z_2} + 1} \quad (\text{III.284})$$

în care:

$$z_1 = z - nps_z \quad (\text{III.285})$$

$$z_2 = z + nps_z \quad (\text{III.286})$$

np - argumentul funcției inverse Laplace (Φ^{-1}) pentru o anumite probabilitate (p) de verificare a ipotezei testate.

$s_z = \frac{1}{\sqrt{n-3}}$ - abaterea standard a variabilei z .

Pe baza abaterii standard a coeficientului de corelație

$$s_r = \frac{1-r^2}{\sqrt{n}} \quad (\text{III.287})$$

intervalul de încredere al coeficientului de corelație pentru o probabilitate p se calculează cu relația:

$$r - np \frac{1-r^2}{\sqrt{n}} < \rho < r + np \frac{1-r^2}{\sqrt{n}} \quad (\text{III.288})$$

4.1.4. Parametrii modelului

Evaluarea parametrilor modelului statistic liniar parcurge cele două etape clasice de calcul al parametrilor pe baza eșantionului de date disponibile și de evaluare a incertitudinii acestor parametri.

a) Calculul parametrilor

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:

$$SPA = \sum_{i=1}^n [y_i - (a_0 + a_1 x_i)]^2 \quad (\text{III.289})$$

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

$$\begin{cases} a_0 r + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases} \quad (\text{III.290})$$

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

$$\begin{cases} a_0 = m_y - \frac{s_{xy}}{s_{xx}} m_x \\ a_1 = \frac{s_{xy}}{s_{xx}} \end{cases} \quad (\text{III.291})$$

în care:

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

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

$$s_{xy} = \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i \quad (\text{III.292})$$

$$s_{xx} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (\text{III.293})$$

b) Evaluarea incertitudinii

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

$$\sigma_y^2 \approx s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - m_y)^2 \quad (\text{III.294})$$

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

$$\sigma_{\alpha_0}^2 = \sigma_y^2 \sum_{i=1}^n x_i^2 / \left[n \sum_{i=1}^n (x_i - m_x)^2 \right] \quad (\text{III.295})$$

Variabila:

$$t_{\text{exp}} = (\alpha_0 - a_0) / s_{\alpha_0} \quad (\text{III.296})$$

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

$$s_{\alpha_0}^2 = s_y^2 \sum_{i=1}^n x_i^2 / \left[n \sum_{i=1}^n (x_i - m_x)^2 \right] \quad (\text{III.297})$$

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

$$a_0 - t \left(1 - \frac{\alpha}{2}; \nu \right) s_{\alpha_0} < \alpha_0 < a_0 + t \left(\frac{\alpha}{2}; \nu \right) s_{\alpha_0} \quad (\text{III.298})$$

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

$$|t_{\text{exp}}| > t \left(1 - \frac{\alpha}{2}; \nu \right) \quad (\text{III.299})$$

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

$$\sigma_{\alpha_1} = \sigma_y^2 / \left[\sum_{i=1}^n (x_i - m_x)^2 \right] \quad (\text{III.300})$$

Variabila

$$t_{\text{exp}} = (a_1 - \alpha_1) / s_{\alpha_1} \quad (\text{III.301})$$

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

$$s_{\alpha_1}^2 = s_y^2 / \left[\sum_{i=1}^n (x_i - m_x)^2 \right] \quad (\text{III.302})$$

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

$$a_1 - t \left(1 - \frac{\alpha}{2}; \nu \right) s_{\alpha_1} < \alpha_1 < a_1 + t \left(1 - \frac{\alpha}{2}; \nu \right) s_{\alpha_1} \quad (\text{III.303})$$

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

$$|t_{\text{exp}}| < t \left(1 - \frac{\alpha}{2}; \nu \right) \quad (\text{III.304})$$

4.1.5. Aplicație

Diagrama de împrăștiere pentru masa în stare umedă (M_w) și masa în stare uscată (M_d) a depozitelor recoltate din iazul de decantare Baia Sprie sugerează o corelație lineară între acești doi parametri (**Fig. 93**).

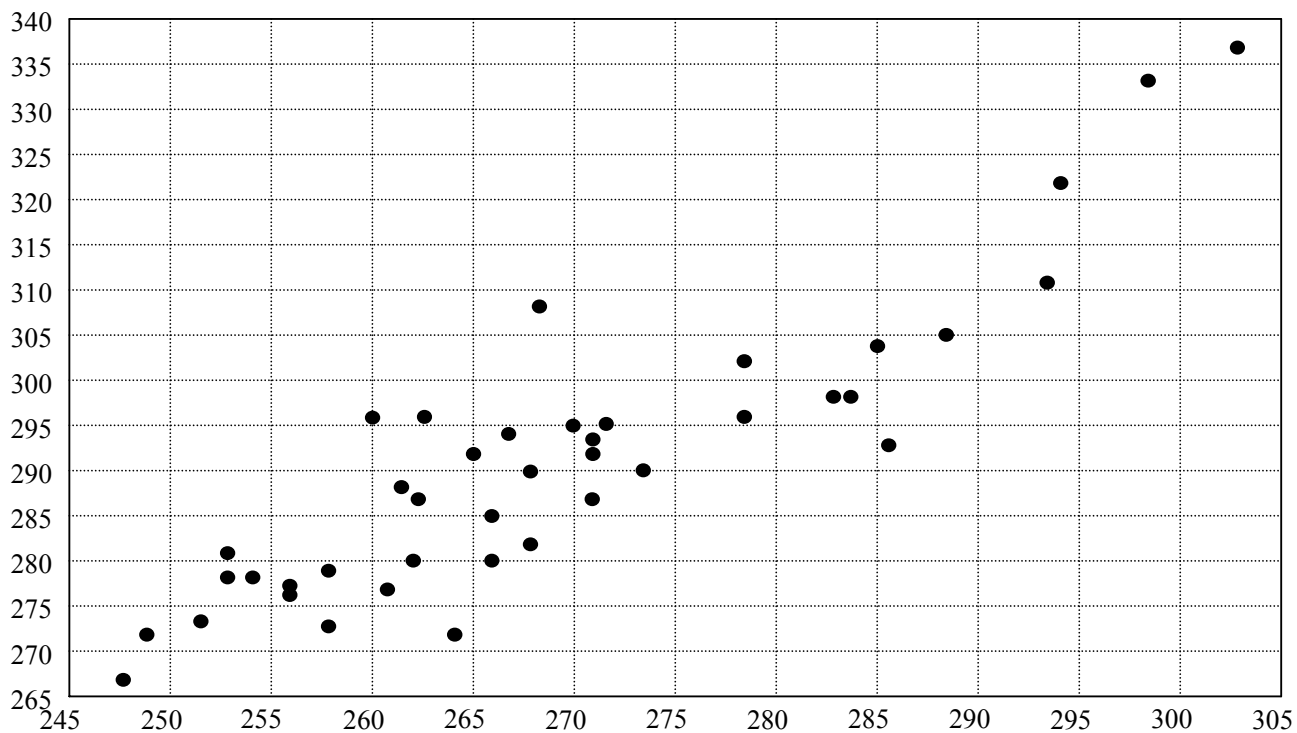


Fig. 93 Diagrama de împrăștiere pentru M_w și M_d

Pe baza celor 49 de valori prelucrate se vor parcurge în continuare principalele etape ale obținerii modelului:

$$M_w = \alpha + \beta \times M_d$$

Realizarea stereogramei evidențiază într-un mod sugestiv două aspecte determinante pentru strategia aplicării metodologiei clasice:

- existența unui număr de valori extreme aberante ce trebuie eliminate înaintea evaluărilor numerice;
- caracterul normal al repartiției bidimensionale a variabilelor M_w și M_d care asigură interpretarea corectă atât a valorilor coeficientului de corelație cât și a parametrilor modelului.

Intensitatea corelației între cele două variabile este evaluată prin intermediul coeficientului de corelație:

- 1) înaintea eliminării valorilor extreme: $r_1 = 0,32$, valoare care contrazice flagrant aspectul diagramei de împrăștiere și al stereogramei;
- 2) după eliminarea a opt valori extreme: $r_2 = 0,889$.

Testarea adecvării modelului devine formală la o valoare a coeficientului de corelație $r_2 = 0,889$ și într-adevăr prin calcul se obține:

$$t(\alpha = 0,05; \nu = 39) = 0,021 < t_{\text{exp}} = 12,12$$

criteriu care confirmă din punct de vedere statistic adecvarea modelului linear.

Intervalul de încredere al coeficientului corelației lineare este:

$$0,81 < \rho < 0,93$$

Parametrii modelului estimați în condițiile aceleiași precizii sunt:

$$20,36 < \alpha < 40,68 \text{ cu estimatorul } a = 18,17$$

$$-0,812 < \beta < 1,217 \text{ cu estimatorul } b = 0,781$$

Modelul estimat al corelației lineare este deci:

$$M_w = 18,17 + 0,781 \times M_d$$

Acest model poate fi utilizat cu o bună aproximare pentru deducerea unuia dintre parametrii pe baza celuilalt reducând la jumătate efortul de determinare realizat în laborator pentru depozitele iazului Baia Sprie. Desigur că pentru alte amplasamente coeficienții și poate chiar structura modelului vor fi alții deoarece acest model este un model empiric valabil doar pentru domeniul **valor** (valoric??) al selecției pe baza căreia a fost construit.

4.2. Model liniar multiplu

Complexitatea proceselor geologice implică frecvent analiza influenței simultane a mai multor variabile, aparent independente, asupra unei variabile considerată dependentă (rezultativă) de acțiunea acestora.

Modelarea lineară a acestei corelații multiple este cea mai simplă soluție adoptată într-o etapă preliminară de studiu. Formal ea se exprimă prin ecuația:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n + e_i \quad (\text{III.304})$$

în care:

y - variabila rezultativă (dependentă);

$x_1, x_2, \dots, x_n$ - variabilele factoriale;

$\alpha_1, \alpha_2, \dots, \alpha_n$ - parametrii modelului;

e_i - eroarea de estimare.

Din punct de vedere metodologic, utilizarea acestui model pune două probleme specifice aplicării ei în studiul variabilelor geologice:

- 1) alegerea variabilei rezultative;
- 2) stabilirea numărului de variabile factoriale.

Caracterul rezultativ sau factorial al unei variabile poate fi bine precizat în contextul geologic în care se realizează studiul sau rezultă după rularea tuturor variabilelor sistemului pe poziția variabilei rezultative.

Dacă spre exemplu, caracterul rezultativ al cotei nivelului piezometric într-un acvifer freatic, în raport cu variabilele factoriale: precipitații, grad de acoperire cu vegetație, modul de infiltrare și porozitate, pare evident, nu același lucru se poate spune despre analiza corelației dintre conținuturile de Au, Ag, Pb, Zn, Cu dintr-un zăcământ polimetalic. În acest al doilea caz stabilirea variabilei rezultative poate fi aleasă după criterii statistice pe baza valorii maxime a coeficientului corelației multiple sau pragmatice, de exemplu, necesitatea prognozei conținutului unui anumit metal (Au) funcție de conținutul celorlalte.

Numărul variabilelor factoriale ale modelului este controlat de criterii operaționale (capacitatea de prelucrare a instrumentului de calcul) precum și de necesitățile interpretării rezultatelor. De cele mai multe ori în modelarea statistică se preferă un număr minim de variabile pentru ca efectele numerice să nu estompeze caracteristicile intrinseci ale procesului modelat.

Precizarea configurației modelului liniar multiplu este obligatoriu să fie precedată de o analiză factorială care să simplifice și să ierarhizeze la nivel statistic importanța variabilelor în reflectarea ansamblului de corelații propriu sistemului studiat.

4.2.1. Analiza grafică a corelației multiple

Diagrama de împrăștiere este singura dintre reprezentările grafice utilizate în cazul modelului liniar de o singură variabilă independentă care poate fi generalizat pentru cazul a trei dimensiuni, corespunzător unei corelații multiple cu două caracteristici independente și una factorială.

În cazul a trei variabile X_1, X_2 și X_3 , tripletele (x_1, x_2, x_3) pot fi considerate ca determinând un punct ale cărei coordonate sunt valorile x_1, x_2 și x_3 . Reprezentate într-un sistem de referință ortogonal, toate punctele vor forma o mulțime cu o anumită dispoziție geometrică în raport cu diferite "suprafețe de corelație". Gruparea punctelor în vecinătatea unei astfel de suprafețe poate fi o măsură calitativă a intensității corelației între cele trei variabile.

Pentru mai mult de trei variabile, reprezentări grafice care să rezume în mod sugestiv corelația între variabile nu se poate realiza decât după prelucrări speciale de tipul celor prezentate în cadrul analizei factoriale.

Datele brute nu mai pot fi examinate prin aceleași procedee prezentate la modelul liniar de o singură variabilă independentă (stereograma, dreapta de regresie) decât formând perechi din variabila rezultativă și fiecare variabilă factorială, metodă care ignoră însă tocmai efectul ansamblului de intercorelații pe care tinde să-l exprime modelul corelației multiple.

4.2.2. Evaluarea intensității corelației

Calitatea modelului liniar multiplu se evaluează sub două aspecte:

- a) intensitatea corelației între variabila rezultativă și toate variabilele factoriale, cuantificată cu ajutorul raportului corelației multiple și coeficientului corelației multiple;
- b) intensitatea corelației între variabila rezultativă și fiecare variabilă factorială, exprimată prin coeficientul de corelație parțială.

a) Raportul corelației multiple

Raportul corelației multiple se calculează cu formula:

$$R_{y(x_1 \dots x_n)} = \sqrt{1 - \frac{\sum_{i=1}^k (y - y_{x_1 \dots x_n}^*)^2}{\sum_{i=1}^k (y_i - \bar{y})^2}} \quad (\text{III.305})$$

în care

y_i - valoarea măsurată a variabilei rezultative;

$y_{x_1 x_2 \dots x_n}^*$ - valoarea estimată a variabilei rezultative;

$\bar{y}$ - media valorilor măsurate ale variabilei rezultative;

k - numărul de probe în care se măsoară cele n variabile.

Valoarea $R_{y(x_2, \dots, x_n)}$ depinde deci de raportul dintre dispersia valorilor determinate pe baza ecuației de regresie liniară și dispersia valorilor măsurate ale variabilei rezultative. Cu cât valorile măsurate se abat mai puțin de la valorile calculate, cu atât coeficientul de corelație are o valoare mai mare și ca atare corelația este mai intensă.

b) Coeficientul corelației multiple

Coeficientul corelației multiple între variabilele $y, x_1, x_2, \dots, x_n$ măsoară gradul de precizie cu care y poate fi reprezentată prin modelul liniar multiplu. Relația de calcul a coeficientului corelației multiple este:

$$R_{y(x_1x_2\dots x_n)} = \sqrt{\frac{a_0 \sum_{i=1}^k y_i + a_1 \sum_{i=1}^k x_{1i} y_i + \dots + a_n \sum_{i=1}^k x_{ni} y_i - \frac{1}{n} \left(\sum_{i=1}^k y_i \right)^2}{\sum_{i=1}^k y_i^2 - \frac{1}{n} \left(\sum_{i=1}^k y_i \right)^2}} \quad (\text{III.306})$$

utilizabilă după evaluarea parametrilor modelului prin intermediul coeficienților $a_0, a_1, a_2, \dots, a_n$.

Coeficientul corelației multiple se poate calcula și cu formula:

$$R_{y(x_1x_2\dots x_n)} = \sqrt{1 - (1 - r_{y1}^2)(1 - r_{y2.1}^2) \dots (1 - r_{yn.12\dots(n-1)}^2)} \quad (\text{III.307})$$

în care $r_{y1}^2, r_{y2.1}^2, \dots, r_{yn.12\dots(n-1)}^2$ sunt coeficienții de corelație parțială.

Dacă $R_{y(x_1x_2\dots x_n)} = 1$, variabila rezultativă y poate fi perfect reprezentată prin modelul liniar multiplu. Se poate demonstra că $R_{y(x_1x_2\dots x_n)}$ este mai mare decât coeficientul de corelație între y și orice funcție liniară de $x_1, x_2, \dots, x_n$ diferită de cea din expresia (III.304).

Coeficientul corelației multiple este mai mare sau egal cu zero și deci în mod evident este mai mare (sau egal) decât oricare din coeficienții de corelație parțială care aparțin modelului. Ca o consecință a acestui fapt, dacă $R_{y(x_1x_2\dots x_n)} = 0$ toți coeficienții de corelație referitori la y sunt zero și deci y este independentă față de toate variabilele factoriale ale modelului.

c) Coeficienții de corelație parțială

Coeficienții de corelație parțială exprimă intensitatea corelației între variabila rezultativă (y) și o variabilă factorială oarecare ($x_1, x_2, \dots, x_n$) când restul variabilelor modelului rămân constante.

Pentru un model liniar multiplu cu n variabile calculul coeficienților de corelație parțială se face funcție de coeficienții de ordin inferior cu relația de recurență:

$$r_{y1.23\dots n} = \frac{r_{y1.23\dots(n-1)} - r_{yn.23\dots(n-1)} \times r_{1n.23\dots(n-1)}}{\sqrt{(1 - r_{yn.23\dots(n-1)}^2)(1 - r_{1n.23\dots(n-1)}^2)}} \quad (\text{III.308})$$

Pentru un model liniar cu două variabile independente:

$$y = a_0 + a_1x_1 + a_2x_2 \quad (\text{III.309})$$

aplicând formula (III.308) se obține relația de calcul a coeficientului corelației parțiale între y și x_1 :

$$r_{y_1.2} = \frac{r_{y1} - r_{y2} \times r_{12}}{(1 - r_{y2}^2)(1 - r_{12}^2)} \quad (\text{III.310})$$

în care r_{y1}, r_{y2} și r_{12} sunt coeficienții de corelație binară calculați cu formula (III.280) utilizată pentru evaluarea intensității modelului liniar cu o singură variabilă independentă.

Coeficienții corelației parțiale au valori cuprinse între -1 și +1 semnificația fiind cea a coeficientului de corelație Pearson analizată în detaliu la paragraful IV.2.1.

4.2.3. Testarea adecvării modelului liniar multiplu

Adecvarea modelului liniar multiplu este condiționată de semnificația statistică a coeficientului corelației multiple.

Pentru modelul liniar multiplu, suma pătratelor abaterilor valorilor observate ale lui y față de media lor este egală prin definiție cu

$$k \times s_y^2 \quad (\text{III.311})$$

având $\nu = k - 1$ grade de libertate și două componente:

a) suma pătratelor abaterilor valorilor măsurate față de cele date de ecuația modelului și care este egală cu:

$$k \times s_y^2 \times (1 - R_{y(x_1 x_2 \dots x_n)}^2) \quad (\text{III.312})$$

cu $k - n$ grade de libertate;

b) suma pătratelor abaterilor valorilor calculate prin ecuația modelului față de media valorilor măsurate:

$$k \times s_y^2 \times R_{y(x_1 x_2 \dots x_n)}^2 \quad (\text{III.313})$$

cu $n - 1$ grade de libertate.

Dacă y (valoarea măsurată) și y^* (valoarea estimată prin model) sunt complet necorelate, abaterile lui y față de valorile modelului (y^*), vor fi independente de abaterile valorilor calculate față de media valorilor măsurate și deci dispersiile celor două componente vor fi practic identice ($R = 0$).

Testarea semnificației statistice a diferenței celor două componente poate fi realizată cu ajutorul repartiției Z calculând factorul experimental:

$$Z_{\text{exp}} = \frac{1}{z} \ln \frac{R_{y(x_1 x_2 \dots x_n)}^2}{1 - R_{y(x_1 x_2 \dots x_n)}^2} \times \frac{k - n}{n - 1} \quad (\text{III.314})$$

cu $\nu = n - 1$ și $\nu_2 = k = n$ grade de libertate.

Dacă

$$Z_{\text{exp}} < Z(\alpha, \nu_1, \nu_2) \quad (\text{III.315})$$

valoarea coeficientului de corelație $R_{y(x_1 x_2 \dots x_n)}$ este nesemnificativă și modelul liniar multiplu nu este adecvat modelării corelației între $n + 1$ variabile.

În caz contrar, din punct de vedere statistic, corespunzător nivelului de semnificație ales, modelul liniar multiplu este adecvat modelării relației între variabila rezultativă (y) și variabilele factoriale: $x_1, x_2, \dots, x_n$.

Semnificația coeficientului corelației multiple este puternic afectată de numărul de valori disponibile (k) și numărul de variabile ale modelului (n). În cazul limită în care numărul de variabile este egal cu numărul de observații disponibile, toate corelațiile parțiale de cel mai ridicat grad posibil vor fi egale cu valoarea unitară și în consecință R va indica o corelație totală indiferent de ansamblul real de corelații din sistemul studiat.

4.2.4. Parametrii modelului

Evaluarea parametrilor modelului corelației multiple parcurge aceleași două etape cu cele prezentate în paragraful precedent pentru modelul liniar cu o singură variabilă independentă.

a) Calculul parametrilor

Evaluarea parametrilor $a_0, a_1, \dots, a_n$ se face prin aplicarea modelului

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (\text{III.316})$$

În mod analog cu procedeul aplicat modelului liniar de o singură variabilă independentă se minimizează suma abaterii pătratelor:

$$SPA = \sum_{i=1}^k [y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_n x_{ni})]^2 \quad (\text{III.317})$$

prin derivare în raport cu $a_0, a_1, \dots, a_n$ obținându-se sistemele :

$$\begin{cases} \frac{\partial SPA}{\partial a_0} = -2 \sum_{i=1}^k [y_i - (a_0 + a_1 x_{1i} + \dots + a_n x_{ni})]^2 = 0 \\ \frac{\partial SPA}{\partial a_1} = -2 \sum_{i=1}^k x_{1i} [y_i - (a_0 + a_1 x_{1i} + \dots + a_n x_{ni})]^2 = 0 \\ \frac{\partial SPA}{\partial a_n} = -2 \sum_{i=1}^k x_{ni} [y_i - (a_0 + a_1 x_{1i} + \dots + a_n x_{ni})]^2 = 0 \end{cases} \quad (III.318)$$

$$\begin{cases} a_0 k + a_1 \sum_{i=1}^k x_{1i} + \dots + a_n \sum_{i=1}^k x_{ni} = \sum_{i=1}^k y_i \\ a_0 \sum_{i=1}^k x_{1i} + a_1 \sum_{i=1}^k x_{1i}^2 + \dots + a_n \sum_{i=1}^k x_{ni} x_{1i} = \sum_{i=1}^k x_{1i} y_i \\ a_0 \sum_{i=1}^k x_{ni} + a_1 \sum_{i=1}^k x_{1i} x_{ni} + \dots + a_n \sum_{i=1}^k x_{ni}^2 = \sum_{i=1}^k x_{ni} y_i \end{cases} \quad (III.319)$$

prin a căror rezolvare se obțin valorile parametrilor.

Fiecare dintre parametrii modelului $(a_1, a_2, \dots, a_n)$ reprezintă variația medie a variabilei rezultative (y) corespunzătoare unei variații unitare a variabilei factoriale, considerându-le pe celelalte constante.

Termenul liber (a_0) reprezintă nivelul de referință al variabilei rezultative fără a avea o semnificație geologică precizată.

b) Evaluarea incertitudinii

Pentru parametrii modelului corelației multiple intervalul de încredere se evaluează pe baza inegalității:

$$a_j - t(\alpha, \nu) \times \frac{S_{yy_i}}{\sqrt{n}} < \alpha_j < a_j + t(\alpha, \nu) \times \frac{S_{yy_i}}{\sqrt{n}} \quad (III.320)$$

pentru coeficienții variabilelor factoriale $(j = 1, 2, \dots, n)$ iar pentru termenul liber pe baza inegalității:

$$a_0 - t(\alpha, \nu) \times \frac{S_{yy_i}}{\sqrt{n}} < \alpha_0 < a_0 + t(\alpha, \nu) \times \frac{S_{yy_i}}{\sqrt{n}} \quad (III.321)$$

în care

s_{yy_i} - abaterea medie pătratică a valorilor observate față de valorile calculate prin model:

$$s_{yy_i} = \sqrt{\frac{\sum_{i=1}^k (y_i - y_i^*)^2}{k - n - 1}} \quad (III.321)$$

s_{a_j} - abaterea standard introdusă de fiecare variabilă factorială:

$$s_{a_j} = \frac{s_{yy_i}}{\sqrt{\sum_{i=1}^k (x_{ij} - m_{xj})}} \quad (\text{III.322})$$

4.2.5. Aplicație

Dintr-un acvifer freatic s-a exploatat pe o perioadă de 10 ani un debit ce variază de la 1000 la 6000 m³/zi. Acviferul este alimentat prin infiltrații rezultate din precipitații care în zonă au valoarea medie de 350 mm/an.

Pentru optimizarea regimului de funcționare a forajelor de drenaj s-a elaborat un model statistic de tip linear pe baza valorilor medii lunare ale debitelor exploatate și precipitațiilor pe perioada 1970 - 1980.

Elaborarea modelului a cuprins trei etape: identificarea variabilelor modelului, evaluarea parametrilor și evaluarea performanțelor.

a) Identificarea variabilelor modelului s-a realizat pe baza corelogramelor calculate pentru cele două variabile principale (Q-debit și P-precipitații). Din corelogramele calculate se remarcă o autocorelare importantă a debitului de exploatare pentru un decalaj de 1 lună și 4 luni (**Fig. 96**) și o

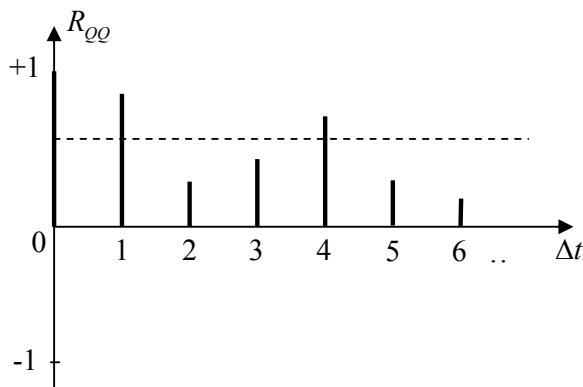


Fig. 96 Autocorelograma Q-Q

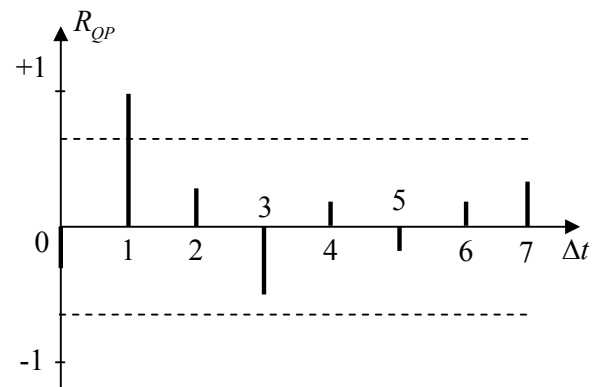


Fig. 97 Autocorelograma Q-P

corelare importantă între precipitații și debitul de exploatare cu un decalaj de o lună (**Fig. 97**).

În aceste condiții modelul identificat optim este de forma:

$$Q(t) = a_0 + a_1 Q(t-1) + a_2 Q(t-4) + a_3 P(t-1)$$

b) Evaluarea parametrilor modelului prin minimizarea abaterilor a condus la coeficienții:

$$a_0 = 2077,5; \quad a_1 = 0,3299; \quad a_2 = 0,2128; \quad a_3 = 0,9648.$$

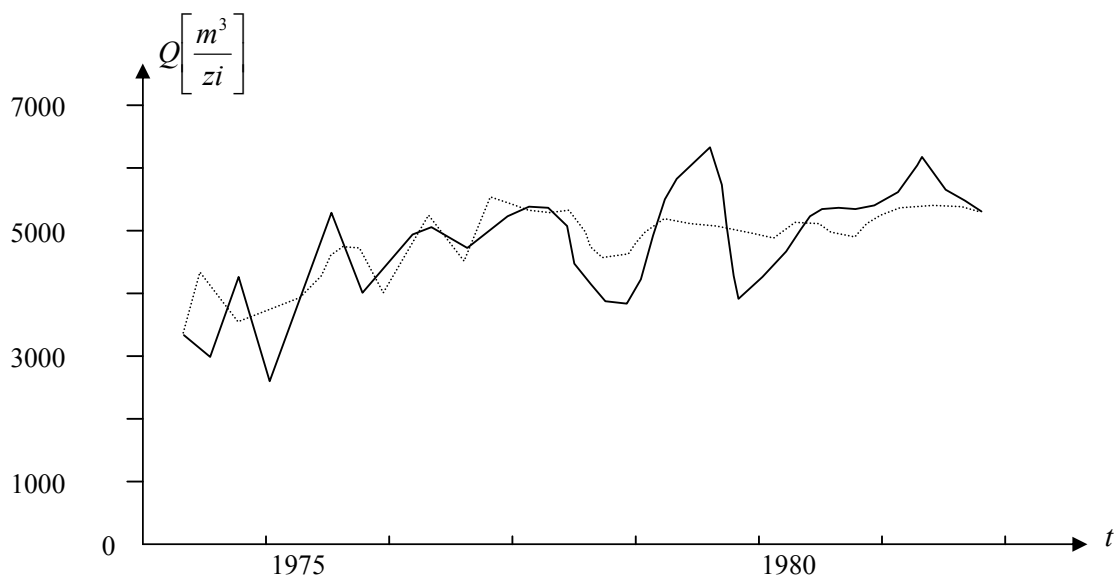


Fig. 98 Relația dintre debitul calculat (modelat) și cel măsurat

c) Performanțele modelului exprimate prin coeficientul corelației multiple și a coeficienților de corelație parțială sunt:

corelația totală între $QQ(t-1)Q(t-4)P(t-1)$: $R = 0,65$

corelația parțială între Q și $Q(t-1)$: $r_{QQ(t-1)} = 0,16$;

corelația parțială între Q și $Q(t-4)$: $r_{QQ(t-4)} = 0,14$;

corelația parțială între Q și $P(t-1)$: $r_{QP(t-1)} = 0,63$.

Grafic relația dintre valorile observate și cele calculate prin model (Fig. 98) exprimă o bună adecvare a modelului pentru corelațiile între debitul de exploatare și precipitații.



3.3. Rezumat

Modelele statistice cantitative prezintă:

- Tipuri de coeficienți de corelație pentru variabile cantitative și calitative
- Metoda analizei factoriale pentru ierarhizarea intensității tipurilor de corelații
 - o R-MOD
 - o Q-MOD
- Modelul analitic al corelațiilor de tip linear/linearizabil
 - o Model liniar binar
 - o Model liniar multiplu



3.4. Test de autoevaluare a cunoștințelor

- 1) Care sunt valorile extreme ale coeficientului PEARSON?
 - a. -1 și +1
 - b. -1 și 0
 - c. 0 și +1
- 2) Care sunt valorile extreme ale coeficientului de intercorelație?
 - a. zero și +infinit
 - b. -infinit și zero
 - c. -1 și +1
- 3) Cum este intensitatea corelației dintre două variabile (A și B), pentru un coeficient al distanței taxonomice $D(A,B)=ZERO$?
 - a. MAXIMA
 - b. MINIMA
 - c. MEDIE



n.5. Bibliografie recomandată

13. Tufescu M, 1981: "Populația biologică privită ecologic". – Pp. 60-78 în: Tufescu V & Tufescu M, 1981: "Ecologia și activitatea umană". - București: Editura Albatros, 406 pp.
14. Watt AD, Leather SR, Hunter MD & Kidd NAC, 1990: Population Dynamics of Forest Insects. Andovers, Hampshire: Intercept, xxii + 408 pp.
15. Bănărescu P, 1973: Principiile și metodele zoologiei sistematice. – București: d.Acad.Rom.
16. Générmont J & Lamotte M, 1980: "Les problèmes de l'espèce dans le règne animal", Tome III. (Mémoire de la Société Zoologique de France #40). Paris: Société Zoologique de France
17. Dallwitz MJ, Paine TA & Zurcher EJ, 1995: User's Guide to the DELTA System. A General System for Processing Taxonomic Descriptions, Edition 4.03. – [152 + 14 pp.]

Răspunsuri la testele de autoevaluare a cunoștințelor

Capitolul 1 MODELE DINAMICE CALITATIVE.

- 1) a
- 2) b
- 3) b

Capitolul 2. MODELE DINAMICE CANTITATIVE.

- 1) a
- 2) b
- 3) b

Capitolul 3. MODELE STATISTICE.

- 1) a
- 2) c
- 3) a