

INTERPOLAREA SPAȚIALĂ A DATELOR UTILIZÂND PACHETUL DE PROGRAME ZAZA

Cristian RUSU

**Studiu de caz: modelarea cantităților de precipitații după un
accident nuclear**

Modelarea asistată de calculator a suprafețelor

Modelarea unei suprafețe presupune determinarea unei funcții de două variabile $z=f(x,y)$ care să aproximeze cel mai bine un set de date de intrare de forma (x_i, y_i, z_i) , prezentate sub formă tabelară. Datele de forma (x_i, y_i, z_i) reprezintă **date spațiale**, z fiind parametrul observat în puncte a căror localizare spațială x, y se cunoaște. Modelarea suprafețelor face parte din domeniul mai larg al **analizei datelor spațiale**. Scopul de bază al analizei datelor spațiale este obținerea unui model al unei întregi arii, plecând de la valorile z măsurate în câteva puncte de control.

Problema modelării suprafețelor se pune în multe domenii: geologie, topografie, meteorologie, hidrologie, ecologie, agricultură, științe biologice, matematică, marketing, statistică, inginerie etc. Datele inițiale (x_i, y_i, z_i) sunt, în marea majoritate a cazurilor, neregulat distribuite. Mai mult, datele sunt adesea distribuite neuniform sau foarte neuniform. În foarte puține cazuri punctele de cunoaștere sunt localizate în nodurile unei rețele regulate. Transformarea datelor punctuale în variabile funcționale se face prin interpolare. Interpolarea poate avea două semnificații:

- interpolare **globală** – încearcă să descopere și să explice comportamentul unei funcții pornind de la premiza că aceasta are caracteristici ce se repetă cu o anumită periodicitate; dacă parametrii ce caracterizează comportamentul spațial al funcției pot fi determinați, ei se pot utiliza pentru reconstrucția funcției și determinarea valorilor ei în orice punct;
- interpolare **locală** – încearcă să reconstruiască funcția spațială printr-o abordare locală, doar pe baza datelor din vecinătate; se pleacă de la premiza că funcția, deși este global heterogenă, este relativ simplă local.

Interpolarea globală se preferă atunci când semnificația datelor este cunoscută și se caută o expresie analitică care să exprime comportamentul acestora. Interpolarea locală se face prin metode de tip medieri ponderate.

Prin **gridding** se generează o rețea regulată, de obicei pătratică, în nodurile căreia se estimează valorile z pe baza setului de date cunoscute în punctele de măsură (x_i, y_i, z_i) . Rețeaua regulată determină un set de patrulatere strâmbe în spațiu. Aproximarea suprafeței se face prin modelarea ei pe fiecare astfel de "petec" în parte, prin interpolări ce nu mai iau în considerare setul de date inițiale ci doar nodurile rețelei regulate.

Toate metodele de interpolare utilizate în gridding trebuie să răspundă aceluiași probleme:

- reprezentări cât mai precise în zonele cu puține date;
- reprezentări ale unor caracteristici locale, evitându-se totodată generarea unor false caracteristici ale suprafeței;
- interpolări în zonele marginale.

Performanțele obținute prin gridding depind de gradul de complexitate al suprafeței modelate, distribuția punctelor de observație și metoda de gridding aleasă. Alegerea metodei de gridding depinde de o mulțime de factori, cei mai importanți fiind natura și caracteristicile datelor inițiale, precum și scopul urmărit prin gridding. În orice metodă de gridding intervin **parametri** pe care, în general, utilizatorul îi poate alege sau modifica. Modificarea parametrilor poate produce modificări importante în valorile suprafeței în nodurile rețelei regulate.

Comparații între metodele de gridding se pot face pe baza unor criterii de evaluare a fiecărei metode. Câteva dintre criteriile ce pot fi avute în vedere sunt: acuratețea aproximării suprafeței, volumul de date inițiale ce poate fi prelucrat, ușurința implementării algoritmului, viteza de calcul, influența distribuției datelor.

Zaza - program de modelare a suprafețelor de uz general

Experiența mea în utilizarea dar și proiectarea unor pachete de modelare a suprafețelor s-a concretizat în pachetul integrat pentru modelarea suprafețelor (**ZAZA**). **ZAZA** reprezintă un pachet complet, de uz general, de software pentru modelarea suprafețelor, dezvoltat sub mediul de programare Borland Pascal 7.0, în tehnologie obiectuală, pentru sistemul de operare MS-DOS. Programul realizează modelarea suprafețelor prin gridding, toate tipurile de reprezentări grafice uzuale, operații cu suprafețe, estimări de arii, volume și rezerve. Utilizatorul poate controla parametrii de lucru în toate fazele. Toate fișierele de intrare/ieșire sunt compatibile cu cel mai cunoscut program similar existent, **SURFER** (Golden Software Inc.).

În pachetul **ZAZA** au fost implementate până acum **metode locale** (medii ponderate mobile, aproximări locale polinomiale, algoritm de aproximare liniară cu corector), **metode globale** (analiza suprafețelor polinomiale de tendință) și **metode speciale** de gridding (algoritmi de eșantionare, de distribuție spațială și de densitate). Metodele implementate oferă o gamă largă de posibilități de modelare a suprafețelor. Utilizatorul poate alege metoda cea mai potrivită, în funcție de natura și caracteristicile datelor inițiale și de scopul urmărit în modelare. El poate controla procesul de gridding nu numai prin alegerea algoritmului dorit, dar și prin alegerea parametrilor rețelei, tipului de funcție pondere utilizată și caracteristicilor acesteia. Algoritmii implementați acoperă cea mai mare parte a situațiilor întâlnite în practică. Programul poate fi oricând completat prin implementarea altor algoritmi sau clase de algoritmi, după necesități. Programul **ZAZA** a fost utilizat în prelucrarea datelor geologice, meteorologice și de statistică sanitară.

Modelarea cantităților de precipitații după un accident nuclear

Pentru exemplificarea metodelor de gridding implementate în pachetul ZAZA am ales un set de date format din 100 de măsurători ale precipitațiilor observate în Elveția, în 1986, în urma catastrofei nucleare de la Cernobîl. Acest set de date a fost utilizat în cadrul programului *Spatial Interpolation Comparison (SIC'97)*, la care am participat. SIC'97 a fost un program internațional on-line, prin Internet, pentru evaluarea metodelor de interpolare spațială, sub egida Radioactivity Environmental Monitoring (R.E.M), din cadrul Environment Institute, Joint Research Centre (Ispra, Italy). Scopul programului SIC'97 a fost tocmai cel de evaluare a metodelor de interpolare a datelor spațiale, a software-ului utilizat și de stabilire a unei strategii de atenționare imediată în cazul unor accidente ce pot avea urmări catastrofale asupra mediului. Am ales acest set de date pe de o parte pentru că el a fost conceput special pentru evaluarea metodelor de interpolare spațială și pe de altă parte pentru importanța deosebită pe care o are modelarea acestui tip de date. În prelucrarea datelor am utilizat exclusiv pachetul ZAZA.

Distribuția spațială a punctelor de observație este redată în fig. 1. Cercurile au centrele în punctele de observație, iar razele lor sunt proporționale cu valorile suprafeței în acele puncte. Aria totală a datelor este de 340 x 220 km.

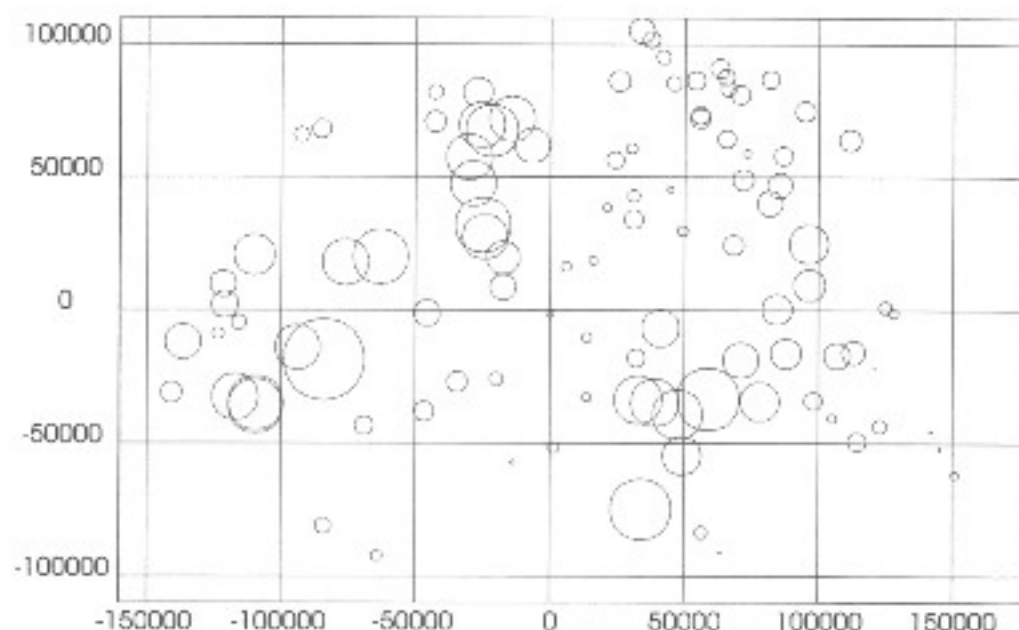


Fig. 1. Distribuția punctelor de observație

Parametrii rețelei regulate au fost aleși astfel:

- pași egali de rețea pe ambele direcții: $X_{pas} = Y_{pas} = 10$ km;
- număr de linii: 23;
- număr de coloane: 35.

Raza zonei circulare de selecție (raza de interpolare) a fost aleasă $R_s = 200$ km. Raza de extrapolare a fost aleasă $R_e = 100$ km. Ea a fost intenționat mărită exagerat pentru a se putea face modelare și în zonele marginale și pentru a se putea compara rezultatele extrapolării prin diverse metode de gridding. Au fost trasate izoliniile cu valori cuprinse între 0 și 600, cu pasul 50.

Ca regulă generală s-au utilizat, în cazul fiecărui algoritm, atât funcții pondere deterministe cât și statistice. Alura funcțiilor pondere statistice a fost aleasă fixă pentru toate nodurile de rețea, iar alura funcțiilor pondere deterministe a fost aleasă variabilă, în funcție de punctele de observație utilizate pentru interpolare în acel nod de rețea.

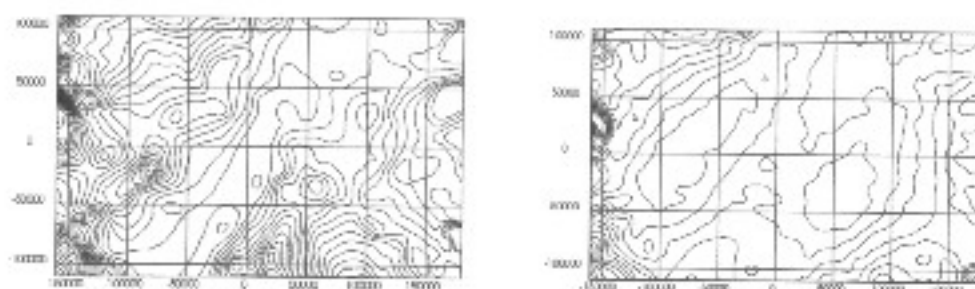


Fig. 2. Aproximare liniară (funcție pondere deterministă, respectiv funcție pondere statistică)

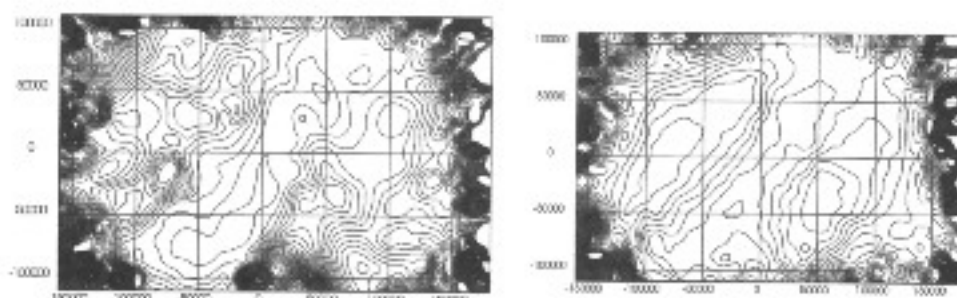


Fig.3. Aproximare pătratică (funcție pondere deterministă, respectiv funcție pondere statistică)

În fig. 2 sunt prezentate rezultatele interpolării prin **aproximare locală liniară**. În primul caz (stânga) s-au utilizat funcții pondere deterministe iar în al doilea (dreapta) funcții pondere statistice. În acest din urmă caz medierea este mai pronunțată, fapt ce se reflectă pe harta cu izolinii prin mai puține detalii interceptate. În ambele cazuri, extrapolările în zonele marginale (fără date) produc tendințe exagerate (atât negative cât și pozitive). Acestea se reflectă pe ambele hărți prin concentrarea de izolinii în zonele marginale. Tendințele sunt ceva mai puțin pronunțate atunci când se utilizează funcții pondere statistice și medierea este mai accentuată.

Interpolarea prin **aproximare locală pătratică** s-a utilizat pentru obținerea hărților din fig. 3. Și aici medierea este mai pronunțată în cazul utilizării funcțiilor pondere statistice. Tendințele

sunt mult exagerate, în zonele marginale, față de cazul aproximării liniare.

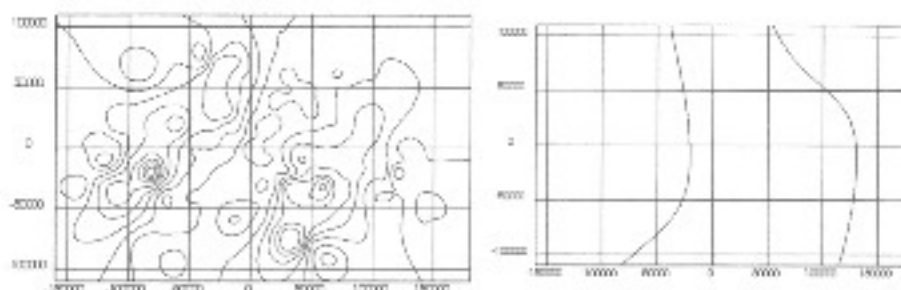


Fig. 4. Mediere ponderată deterministă, respectiv statistică

Hărțile din fig. 4 s-au obținut prin **medii ponderate mobile**. În ambele cazuri la interpolare se iau în calcul toate punctele de observație din zona circulară de selecție. Prin utilizarea funcțiilor pondere deterministe (stânga) medierea este mult mai mare decât în cazurile anterioare. Zonele minimale și maximale sunt sesizate, dar amplitudinile suprafeței sunt subestimate în aceste zone. Prin utilizarea funcțiilor pondere statistice (dreapta) medierea este aproape totală.

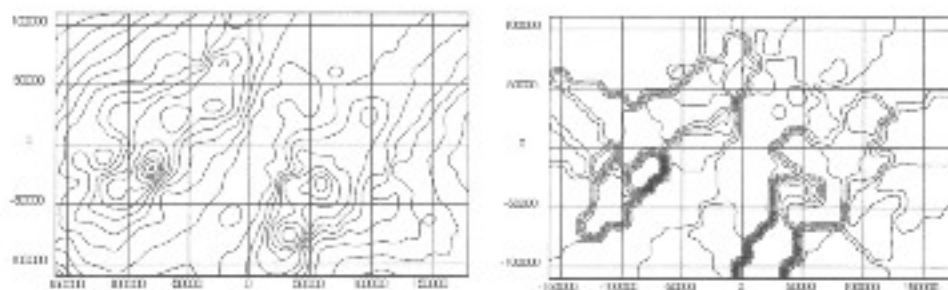


Fig. 5. Aproximare liniară cu corector, respectiv algoritm de eșantionare

Cea mai precisă aproximare a suprafeței descrisă de punctele de observație inițiale se obține prin utilizarea **algoritmului liniar cu corector** (fig. 5, stînga). Aproximarea este precisă în apropierea punctelor de observație iar în rest este obținută prin mediere. În zonele

marginale nu mai sunt prezente tendințe exagerate. În fig. 5 (dreapta) sunt reprezentate și rezultatele aplicării **algoritmului de eșantionare**, prin care fiecărui nod de rețea îi este atribuită valoarea celui mai apropiat punct de control. Treptele prin care este aproximată suprafața sunt evidente pe harta cu izolinii.

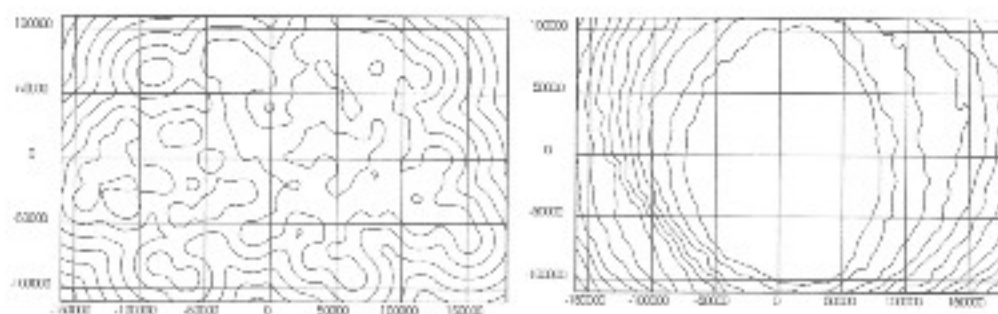


Fig. 6. Algoritm de distribuție spațială, respectiv algoritm de densitate

Prin aplicarea **algoritmului de distribuție spațială** (fig. 6, stînga) se obține o zonă relativ constantă, în zonele cu puncte de observație, și izolinii de valori tot mai mici, în zonele marginale, pe măsura depărtării de regiunile cu date. Acest algoritm atribuie fiecărui nod de rețea valoarea distanței până la cel mai apropiat punct de control. Harta din fig. 6 (dreapta) a fost obținută prin aplicarea **algoritmului de densitate**. Zona centrală are o densitate de puncte de observație aproape constantă. Densitatea acestora scade spre zonele marginale.

Interpretarea rezultatelor

În eventualitatea producerii unui accident nuclear, obținerea cât mai rapidă a unor hărți cu distribuția spațială a materialului poluant este esențială, pentru identificarea teritoriilor celor mai expuse și evacuarea populației din zonele respective. Radioactivitatea se măsoară în cadrul unei rețele de monitorizare, iar prin interpolarea spațială a acestor măsurători se obțin hărți cu continuitate spațială. Metodele folosite în interpolare trebuie să fie capabile să furnizeze valorile extreme (minime și maxime) precum și probabilitatea confirmării acestor predicții.

În programul SIC'97 participanții au avut la dispoziție un set de 100 de valori alese aleator dintr-un set de 467 de măsurători efectuate. Pe baza acestor 100 de date disponibile ei au trebuit să estimeze celelalte 367 de valori și să identifice, dintre acestea, cele mai mici și respectiv mari 10 valori. În final au fost disponibile toate cele 467 de măsurători, pentru ca participanții să poată aprecia corectitudinea estimărilor efectuate.

În estimarea celor 367 de valori am folosit programul ZAZA, în care am modificat procedura de gridding, astfel încât interpolarea spațială să se realizeze nu în nodurile unei rețele regulate, ci în cele 367 de locații. Algoritmul folosit a fost liniar cu corector. Statistica datelor inițiale cunoscute, a datelor reale ce s-au cerut a fi estimate, respectiv a estimărilor realizate este prezentată în tabelul tab. 1. Histograma valorilor estimate este prezentată în fig. 7.

Toți indicatorii statistici au valori similare pentru măsurătorile reale ce s-au cerut a fi estimate și pentru estimările realizate. Abaterea medie pătratică, eroarea absolută și eroarea relativă pentru valorile estimate sunt:

$$RMSE = 59,01; MAE = 40,95; MRE = 0,31.$$

STATISTICA	Nr. de valori	Min	Max	Medie	Mediană	Deviația standard
Valori inițiale cunoscute	100	10	585	180,15	141	110,70
Valori reale ce se cer a fi estimate	367	0	517	185,35	162	111,01
Valori estimate	367	0	504,6	179,42	155,95	94,32

Tab. 1. Indicatori statistici

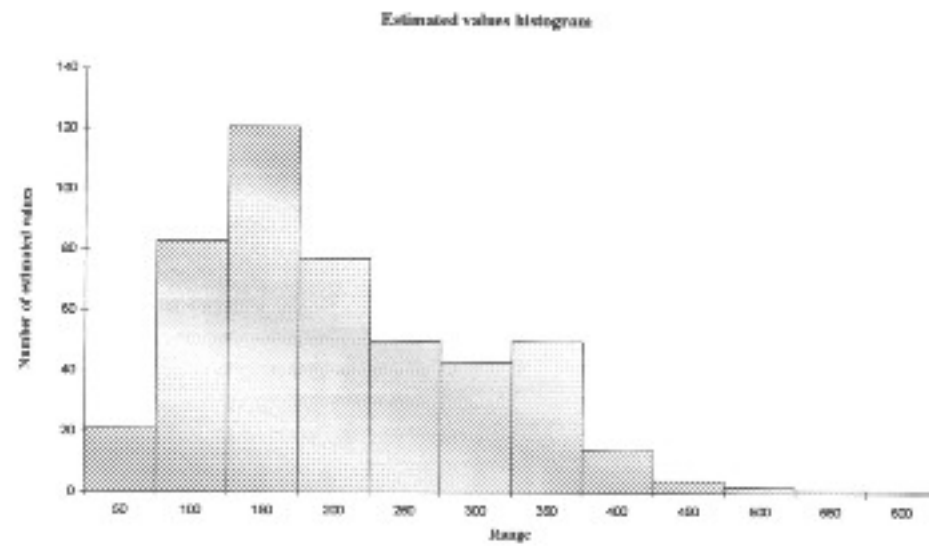


Fig. 7. Histograma valorilor estimate

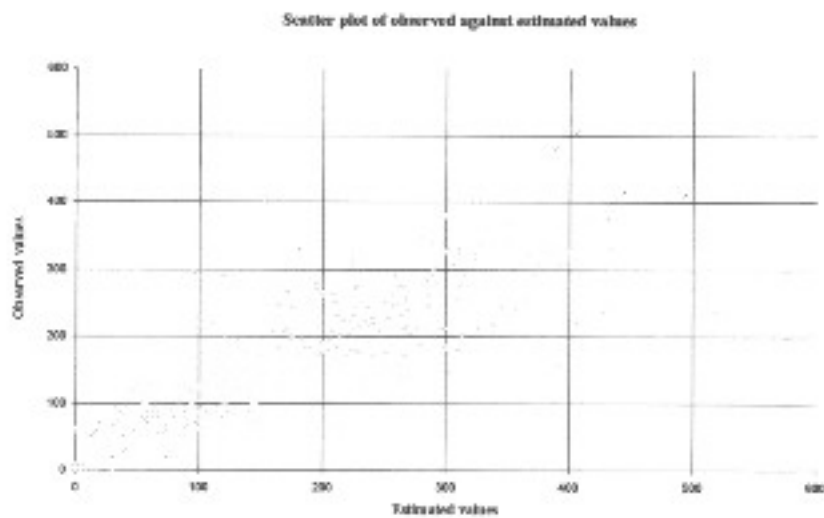


Fig. 8. Diagrama de corelație între valorile reale și cele estimate

Media erorilor are o mică valoare negativă (-5.93), ceea ce înseamnă că valorile reale au fost ușor subestimate.

În fig. 8 se prezintă diagrama de corelație între valorile estimate și cele reale. Covarianța și respectiv coeficientul de corelație între aceste două seturi de date sunt:

$$\text{COV} = 8887,54; \text{CORR} = 0,85.$$

Aceste valori indică o puternică corelație directă între valorile estimate și cele reale.

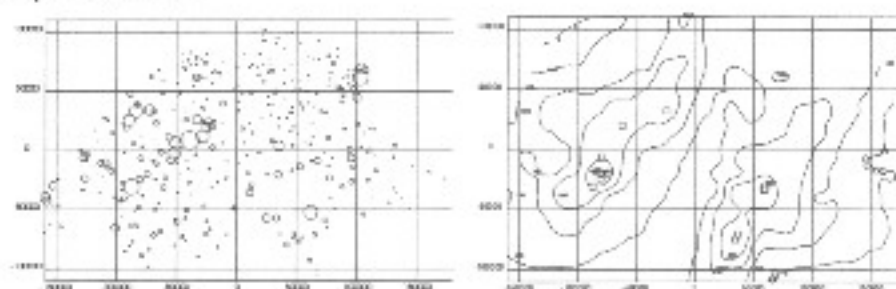


Fig. 9. Harta erorilor (stânga) și harta cu izolinii a valorilor estimate (dreapta), cu precizarea celor mai mici și mai mari 10 valori

Harta erorilor pentru cele 367 valori estimate este prezentată în fig. 9 (stânga). Raza cercurilor este proporțională cu valorile absolute ale erorilor. În fig. 9 (dreapta) se prezintă și harta cu izolinii realizată pe baza celor 100 de date inițiale cunoscute și a celor 367 de valori estimate. Pasul izoliniilor este de 100. Cele mai mari 10 valori sunt marcate cu pătrate, iar cele mai mici 10 valori estimate sunt marcate cu cercuri. Valorile maxime sunt concentrate în zonele cu valori înalte, iar valorile minime sunt localizate în apropierea izoliniei de valoare minimă. Dintre cele 10 valori maxime reale au fost identificate 5, iar dintre cele 10 valori minime reale au fost identificate 8.

Metoda utilizată în estimare este relativ simplă, dar corectitudinea rezultatelor obținute este comparabilă cu cea obținută prin metode mult mai complexe (krigaj simplu, funcții cu bază radială, rețele neuronale, interpolatori cu logică Fuzzy, krigaj pe clase etc.). O parte din rezultatele obținute de participanții la programul SIC'97 au fost publicate în *Journal of Geographic and Decision Analysis*, Vol. 2, No. 2, 1998.

Bibliografie

1. Dubois G.: Spatial Interpolation Comparison 97: Foreword and Introduction, Special Issue of Journal of Geographic and Decision Analysis, Vol. 2, No. 2, 1998, pp.1-10.
2. Rusu C.: Situația actuală și perspectivele modelării asistate de calculator a suprafețelor în geologie, Referat pentru doctorat, Universitatea Tehnică din Cluj-Napoca, Facultatea de Electrotehnică, 1998.
3. Rusu C.: Metode de calcul al suprafețelor în rețea regulată, Referat pentru doctorat, Universitatea Tehnică din Cluj-Napoca, Facultatea de Electrotehnică, 1998.

SPATIAL INTERPOLATION OF DATA USING ZAZA SOFTWARE. A CASE STUDY: RAINFALL DATA MODELING AFTER A NUCLEAR ACCIDENT

Abstract. ZAZA is a software package that solves all surface modeling aspects. The program particularly implements local gridding techniques (weighted average, local polynomial surface fitting) but also trend surface analysis with polynomials. A set of 100 daily rainfall measurement made in Switzerland on the 8th of May 1986, randomly extracted from a dataset of 467 measurements, were used to test the interpolation methods implemented in ZAZA. The rainfalls measured at the remaining 367 locations were estimated. The best results were obtained using an algorithm based on piecewise linear least square. A corrector was applied on calculated values. All negative calculated values were replaced by zero. Computation was made directly at the given locations. Statistics were made using Microsoft Excel.

Keywords: Spatial interpolation, surface modeling, gridding algorithms, contouring, statistics, ZAZA software.

North University of Baia Mare
Department of Mathematics and Computer Science
76, Victoriei
4800 Baia Mare
Romania
chrisro_98@yahoo.com