

# Analiza datelor spațio-temporale

## Curs

Asistent universitar doctorand **Alexandru I. Petrișor**  
**University of South Carolina**

Analiza datelor (seriilor) spațiale și temporale reprezintă un capitol al statisticii care răspunde numeroaselor preocupări din biologie (răspândirea spațială a populațiilor biologice, analiza datelor prelevate din stații amplasate în diferite regiuni sau/ și la diferite momente), dar și din epidemiologie (răspândirea spațială a bolilor, evoluția unor boli cu caracter epidemic) și alte științe ale Pământului.

### 1. Date spațiale

#### 1 a. Testul DAC

Testul DAC (Drane- Aldrich- Creangă) a fost prezentat pentru prima dată în literatura statistică printr-un studiu efectuat de Drane, Aldrich, Creangă și Hudson [2]. Scopul acestui test este acela de a identifica aglomerări spațiale (*cluster*), sau, în sens mai larg, zone cu probleme. Pentru calculul testului DAC se consideră distribuția bidimensională de frecvențe cumulate (extrapolarea în două dimensiuni a distribuției de frecvențe cumulate, în care acumularea se face pe direcția primei bisectoare, respectiv SV→NE), sau, în termeni matematici:

$$F_n(x_1, x_2) = m(x_1, x_2) / n$$

unde  $m(x_1, x_2)$  este numărul de puncte din eșantionul de volum  $n$  care satisfac condiția:  $x_{1,i} \leq x_1$  și  $x_{2,j} \leq x_2$ .

Testul DAC este:

$$DAC(x_1, x_2) = F_m(x_1, x_2) - F_n(x_1, x_2)$$

unde  $F_m$  este distribuția bidimensională de frecvențe cumulate a "cazurilor" din eșantion, iar  $F_n$  este distribuția bidimensională de frecvențe cumulate a tuturor punctelor din eșantion.

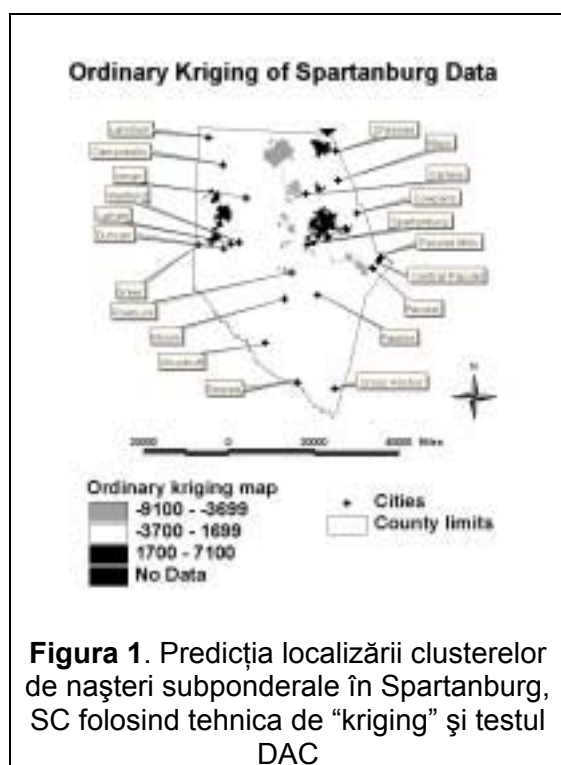
Valorile pozitive ale acestui test indică eventuale probleme, "cazurile" acumulându-se cu o frecvență mai ridicată decât a celorlalte puncte din eșantion.

Petrișor și colaboratorii au arătat că testul DAC nu depinde de poziția originii axelor de coordonate, dar este influențat de orientarea acestora [4]. În alte studii, Petrișor și colaboratorii au investigat utilitatea acestui test în detectarea clusterelor de nașteri subponderale în regiunea Spartanburg din Carolina de sud, SUA [5, 6, 7]. Autorii

au arătat că utilitatea acestui test crește dacă este folosit în combinație cu alte metode [5, 6], în special tehnica de “kriging” care va fi prezentată ulterior [7].

## 1 b. Tehnicile de “kriging”

Numită în onoarea geologului **D. G. Krige**, tehnica de “kriging” reprezintă o metodă de predicție (interpolare) spațială, în care se consideră că variațiile aleatoare ale unei variabile aleatoare sunt determinate de localizarea în spațiu a stațiilor de prelevare a eșantioanelor [8]. Diversele tehnici de “kriging” prezentate în continuare presupun îndeplinirea unor condiții mai mult sau mai puțin restrictive de către variabila considerată [3].



1. O primă tehnică este cunoscută în literatura anglo-saxonă sub numele de “**ordinary kriging**” (kriging obișnuit) și presupune că media variabilei aleatoare este cunoscută și constantă, iar fluctuațiile aleatoare ale acesteia depind doar de amplasarea geografică a stațiilor de prelevare a eșantioanelor.

2. O altă tehnică este numită “**simple kriging**” (kriging simplu) și presupune că media variabilei aleatoare este constantă, dar necunoscută, și supusă unor fluctuații aleatoare determinate de amplasarea geografică a stațiilor de prelevare a eșantioanelor.

3. Cea mai complexă tehnică, denumită “**universal kriging**” (kriging universal) presupune că media variabilei aleatoare depinde de cel puțin o altă variabilă independentă, dar și de amplasarea geografică a stațiilor de prelevare a eșantioanelor.

Cele trei definiții prezentate se regăsesc în funcțiile produsului ArcGIS, utilizat de geografi pentru analiza datelor spațiale. O extensie denumită “Geostatistical Analyst” permite analiza datelor folosind diverse metode de “kriging”.

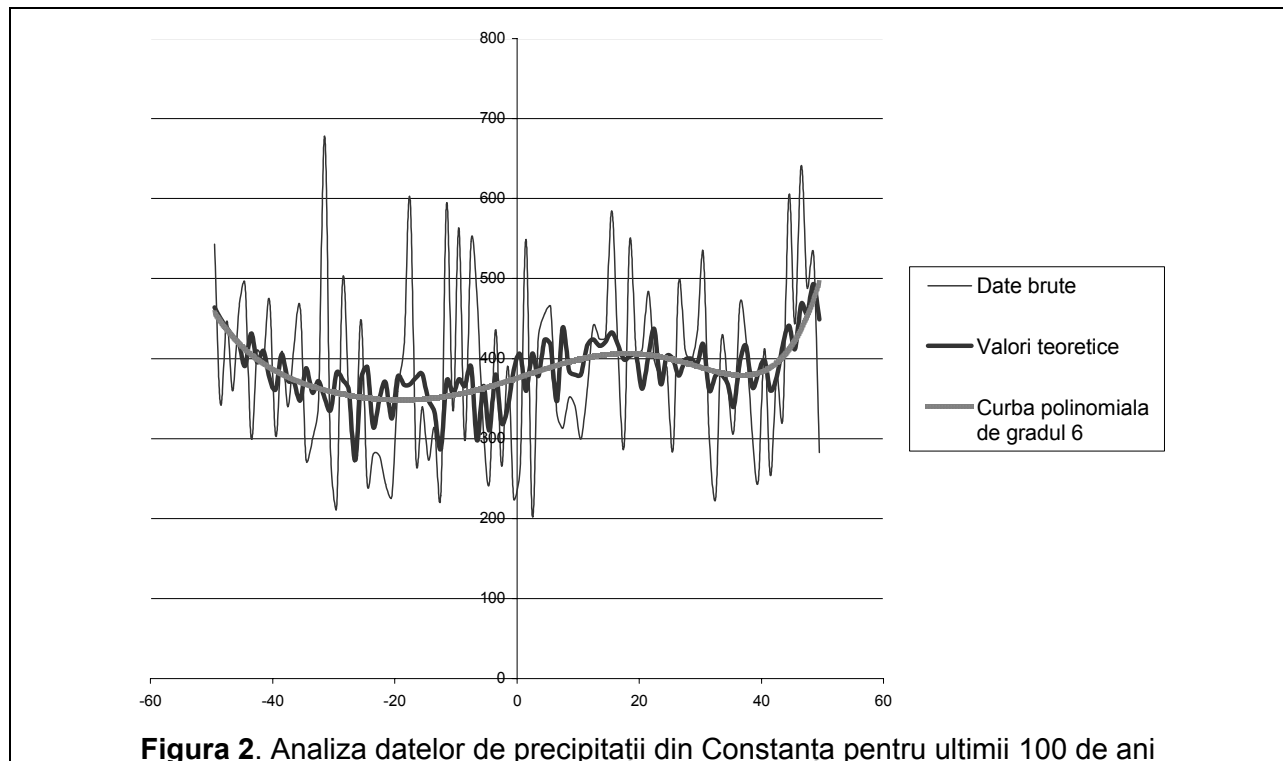
Într-un studiu privind utilitatea testului DAC în identificarea și predicția localizării clusterelor de nașteri subponderale, Petrișor și colaboratorii au utilizat tehnicile de “kriging” descrise mai sus pentru predicția localizării valorilor maxime ale testului [7]. În acest caz, toate metodele prezentate au produs rezultate similare, prezentate în **Figura 1** pentru prima tehnică.

## 2. Date temporale

Analiza datelor temporale (cunoscute și sub numele de **serii cronologice**) și-a găsit numeroase aplicații în ecologie. Spre exemplu, în analiza impactului asupra mediului literatura anglo-saxonă citează exemplul studiilor “BACI” (“*before- after-control- impact*”- înainte- după- control- impact), în care se prelevează eșantioane din două situri (unul de referință- “control” și unul impactat) la diverse momente de timp înainte și după producerea impactului [9]. O astfel de abordare produce două serii cronologice în care variabilitatea spațio-temporală joacă un rol important.

Analiza seriilor cronologice este cunoscută și sub numele de “analiză exploratorie” pentru că, în cele mai multe cazuri, nu urmărește testarea unor ipoteze statistice, ci identificarea unei anumite structuri a datelor. Spre exemplu, renumitul ciclu de 11 ani (presupus a fi datorat exploziilor solare) a fost detectat folosind această metodă în evoluția unor date privind sistemul pradă-prădător (râs și iepure) [8].

Una dintre primele utilizări a analizei seriilor cronologice este, deci, detectarea periodicității datelor. Repetarea unor fenomene la intervale mai scurte (de exemplu, ciclurile datorate variațiilor nocturne/ diurne sau sezoniere, datorate anotimpurilor) sau/ și mai lungi (anuale sau multianuale) poate fi surprinsă cu ajutorul acestei metode. În formele sale mai elaborate, această metodă poate fi folosită și pentru detectarea unei tendințe generale, simple (lineare) sau complexe (modele polinomiale sau mult mai complexe) “în spatele” variațiilor periodice.



Pentru a ilustra acest aspect, să ne imaginăm datele care ar putea surprinde efectul de seră: în afară variațiilor nocturne/ diurne și sezoniere, a celor multianuale (din nou, ciclul de 11 ani), se poate constata o tendință de creștere a mediei temperaturii pentru ultimii 100 de ani.

Cheval și colaboratorii au studiat datele legate de precipitațiile din Constanța pentru ultimii 100 de ani [1]. Rezultatele prezentate în **Figura 2** au indicat că, în acest caz, cel mai important ciclu este cel de 0,2 ani, iar tendința generală este cel mai bine aproximată de o curbă polinomială de gradul 6 (coeficientul de regresie fiind, în acest caz,  $R^2 = 14,02$ ).

## Bibliografie

- [1] **Cheval S., A. I. Petrișor.** *The Variability of Monthly, Seasonal and Yearly Precipitation at Constanta (Romania) during the 20th Century.* Climatic Change (trimis spre publicare)
- [2] **Drane J. W., D. L. Creangă, T. E. Aldrich, M. B. Hudson.** *Detecting Adverse Health Events via Empirical Spatial Distributions.* Proceedings of the Symposium on Statistical Methods. Atlanta, GA: USDHHS, PHS, CDC. 1995
- [3] **Johnston K., J. M. Ver Hoef, K. Krivoruchko, N. Lucas.** *Using ArcGIS Geostatistical Analyst.* ESRI, Redlands, CA. 2001
- [4] **Petrișor A. I.** *Empirical Spatial Distributions.* Teză de Master. University of South Carolina. 2000
- [5] **Petrișor A. I., J. W. Drane, L. Dragomirescu, K. L. Jackson.** *A Computer-Based Application to Detect Spatial Clusters Using the DAC Statistic.* Proceedings of the Medical Infobahn for Europe Special Topic Conference "Healthcare Telematics Support in Transition Countries", Bucharest, Romania. 2001
- [6] **Petrișor A. I., J. W. Drane, L. Dragomirescu.** *Using Kriging and the DAC Statistic to Predict Low Birthweight Clusters in Spartanburg County, SC.* Proceedings of the 2002 Joint Statistical Meetings, New York, NY. 2002
- [7] **Petrișor A. I., L. Dragomirescu, J. W. Drane, K. L. Jackson, D. J. Cowen.** *Using a Quick-Basic Application Based on the DAC Statistic to Detect Spatial Clusters.* Control Engineering and Applied Informatics (în curs de apariție)
- [8] **Piegorsch W. W., D. Edwards.** *Notes on Temporal and Spatial Analysis for a Course in Environmetrics.* University of South Carolina. Columbia, SC. 2001
- [9] **Stewart-Oaten A.** *Problems in the Analysis of Environmental Monitoring Data.* În: Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats, Editori: R. J. Schmitt, C. W. Osenberg. Academic Press. 1996