GEOSTATISTICAL ANALYSIS OF TIME SERIES: AN EXPLORATORY ANALYSIS OF CLIMATOLOGIC AND ECOLOGICAL DATA USING A SPATIAL INSTRUMENT

Alexandru I. PETRIŞOR
PhD Lecturer, Facultatea de Urbanism,
Universitatea de Arhitectură şi Urbanism "Ion Mincu", Bucureşti
Institutul Naţional de Cercetare - Dezvoltare pentru Urbanism şi Amenajarea Teritoriului URBANPROIECT, Bucureşti

E-mail: alexandru_petrisor@yahoo.com

Sorin CHEVAL
Senior Researcher, Deputy-Head of the Dept. of Climatology
National Meteorological Administration

E-mail: sorin.cheval@meteoromania.ro

Victor ZINEVICI
Institute of Biology

E-mail: victor.zinevici@ibiol.ro

Laura PARPALĂ
Institute of Biology

E-mail: lauraparpala@yahoo.com

Abstract:
The extensive use of the Geographical Information Systems, but also of their extensions, in fields that do not necessary connect to geography, but allow for the implementation of methodologies based on the analysis of spatially referenced data, had resulted into new applications and methodological approaches to answer specific research questions. This paper proposes such an extension from the real space define by longitude and latitude to a virtual space defined by two time parameters, month and year, in order to reveal specific temporal patterns of climatologic and ecological phenomena. The methodology is applied in two studies, one aiming to analyze temperature and precipitation data in Romania over a long period, and the other to analyze the effects of anthropic impact and ecological restoration on the density of specific groups of aquatic organisms. In the first study, the method was able to detect climate hazards, while in the second the results had shown different behaviors of analyzed groups, relevant to their ecology. In both cases, the proposed methodology proves its potential for being used as an exploratory research tool in climatology, ecology, and other fields as well.

Key words: *exploratory data analysis; geostatistics; kriging; GIS; time series*

1. Introduction

Within inferential statistics, authors distinguish two chapters: confirmatory and exploratory statistics. Most applied statistics professionals are familiar with the confirmatory data analysis, dealing with the confirmation of statistical hypotheses, estimation, and prediction under very specific circumstances [3, 12]. However, this paper could be placed into the framework of the second chapter, exploratory data analysis (EDA). EDA means simply looking at the data to see what it seems to say, *i.e.*, offer partial descriptions and new insights beneath them, regardless of the statistical criteria used in confirmatory settings [12]. Testing statistical hypotheses is not always possible in EDA, as the aim is to explain and not to confirm [9].

Particular attention is given within the EDA to the analysis of temporal and spatial data. Spatial analysis had been placed under the generic name of kriging, attributed to G. Matheron, and also known under the initial name "theory of regional variables" or as geostatistics [7]. The technique had been developed by D. G. Krige as an optimal interpolation method in the mining industry and named in his honor. Kriging relies on the rate at which the variation between locations changes in space, represented in a graph called semivariogram [9]. Therefore, the mathematical model of kriging is:

$$Z(s_i) = \mu(s_i) + \varepsilon(s_i),$$

where $\mu(s_i)$ is a deterministic term, and $\varepsilon(s_i)$ is the sampling error, depending on the coordinates where the value $Z(s_i)$ is measured [9].

The semivariogram is:

$G(s, u) = ½ \times Var(Z(s)–Z(u))$ for any two locations s and u.

Normally, the semivariogram is assumed to be known and approximated by different mathematical models (linear, spherical, exponential, Gaussian etc.). However, even in these cases parameters must be known. In reality, the semivariogram is not known; therefore, the empirical semivariogram is used instead:

$$\hat{\gamma}(h) = \frac{1}{2N_h} \sum_{(i,j)\in G_h} \left[z(s_i) - z(s_j)\right]^2,$$

where G_h is the set of locations for which $s_i - s_j = h$, and N_h is the number of distinct pairs of elements from G_h [9].

It appears to be simpler to describe the empirical semivariogram than to calculate it: for some value h, identify all locations situated at the distance h and compute half of their mean square error. However, not every h could be represented for all locations (unless exhaustive data is used), and the semivariogram is limited to available values. Problems appear when there are too few or too many locations for a given value of h [9]; in these situations, computations are limited to a set of discrete values $h_1, h_2, \ldots, h^*$, where:

$$h^* = \frac{1}{2} \max_{i,j} \left|s_i - s_j\right|$$

The SAS® program groups observed distances $||s_i – s_j||$ in B classes ranging between 0 and h^*, such that $10 \le B \le 30$, *i.e.*, choose $B \ge 10$ such that there are at least 30 distances in each class used cu compute $\hat{\gamma}(h)$ [11].

Currently, three types of kriging are used: (a) ordinary - assumes that the average of the random variable is known and constant, and its fluctuations depend only on the location of the sampling stations; (b) simple - assumes that the average of the random variable is unknown and constant, and its fluctuations depend only on the location of the sampling stations; and (c) universal - assumes that the average of the random variable depends of at least another variable, but also on the location of the sampling stations [8]. All techniques are available in an ArcGIS extension called "Geostatistical Analyst", used for the analysis of spatially referenced data.

Geographical Information Systems (GIS) can represent, correlate and assess the relationships between large amounts of data, referenced to a known spatial and temporal framework. Hence, besides being a powerful cartographic tool, a GIS can store, retrieve and combine data to create new representation of geographic space, provides tools for spatial analysis and performs simulations [10].

All the techniques presented above had been developed for georeferenced data, *i.e.*, data for which two coordinates are recorded: X, the longitude, and Y, the latitude, defining the real space. This paper proposes an extension of the real space to a virtual space, in which X and Y are any two variables defining the virtual space. We analyze two examples (one from climatology and one from ecology) in which time periodicity is explored by defining a space where the two coordinates are the year and month when observations are recorded.

2. Methods

Both examples provided used the same methodology. Ordinary kriging was used employing the Geostatistical Analyst available with ArcGIS (version 8.x in the first study and 9.x in the second). Default values proposed by the software were accepted. The first study used data on monthly precipitations and temperature collected in Romania at five stations during 1961-2000 and grouped in arrays having the year on the vertical axis and the month on the horizontal one. The second study used data on the density of primary and secondary consumers from two lakes in the Danube Delta (Matiţa and Merhei) collected at five stations in each lake during 1980-2007 and grouped in arrays having the month on the vertical axis and the year on the horizontal one.

3. Results and Discussion

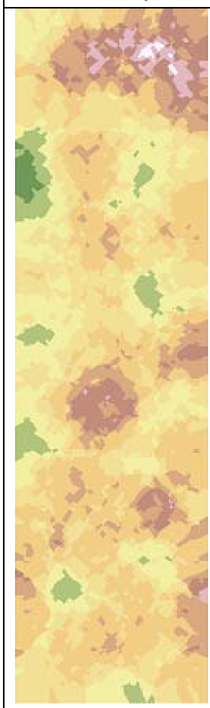
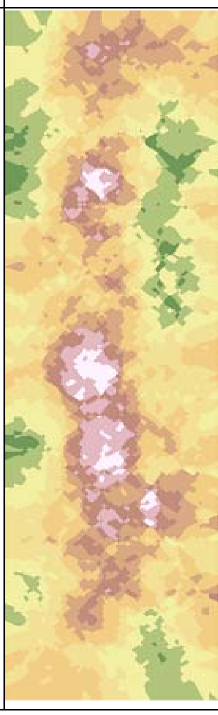
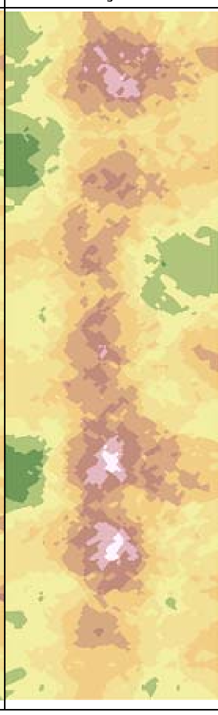
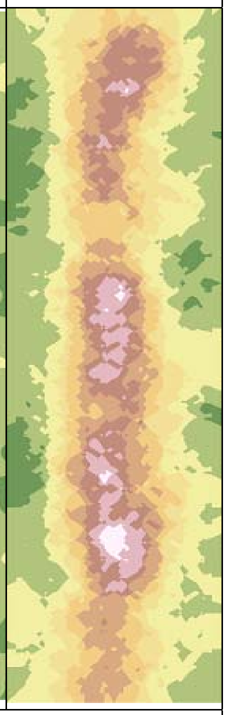
3.1. Example #1: Analysis of climate time series using the Geostatistical Analyst

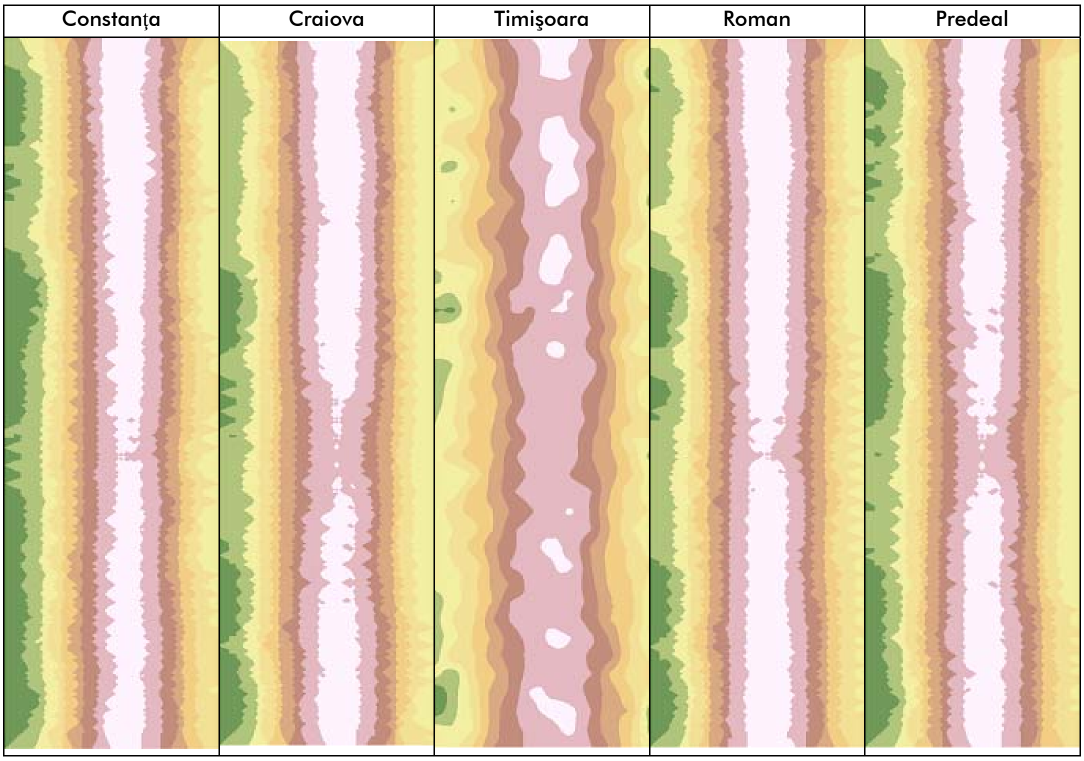
The temporal dimension of climatologic phenomena is one of the most important factors that bias the scientific approach, therefore the integration, analysis, and visualization of spatial data that has a significant temporal component has become a priority in environmental sciences [5, 6].

The first study aimed to create a prediction model for weather hazards (periods with extremely low or high temperatures, and with very low or high amounts of precipitations) based on data collected in Romania at five stations, located in the cities Constanţa, Craiova, Timişoara, Roman, and Predeal (*Fig. 1*), during 1961-2000 [1, 2, 4].

Figure 1. Location of the five sampling stations described in the climate time series analysis using the Geostatistical Analyst (gray dots, black text) and of the two lakes described in the analysis of the effects of anthropic environmental impact and ecological restoration on biodiversity (red/orange dot, red text)

Table 1. Ordinary kriging prediction maps of the precipitations and temperatures in five Romanian cities, based on 1961-2000 data. The color ramp ranges from dark green (abundant precipitations, respectively low temperatures) to bright pink (low precipitations, respectively high temperatures) in the order dark green - light green - yellow - beige - orange - brown - purple - bright pink.

Precipitations				
Constanţa	Craiova	Timişoara	Roman	Predeal
				
Temperatures				



In addition to kriging, the first study used several other interpolation methods: spline fitting, inverse distance weighting, and filtering. Out of all interpolation methods, given its mathematical foundation, kriging yielded the most significant and sound results. The methodology was able to pinpoint one of the most debated issues, climate changes (*e.g.*, cycles of "green" spots for Predeal temperatures, while "bright pink" is apparently wider towards the end of the 1990s).

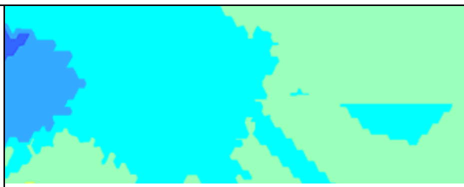
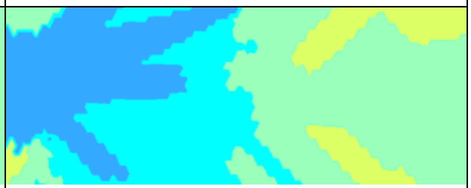
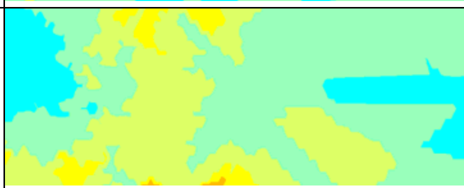
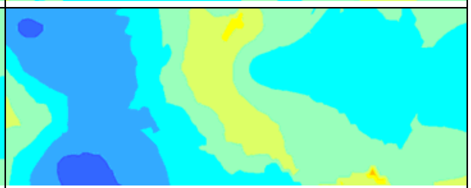
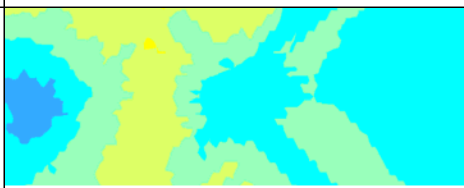
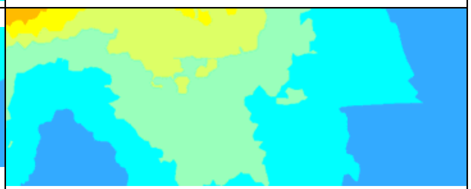
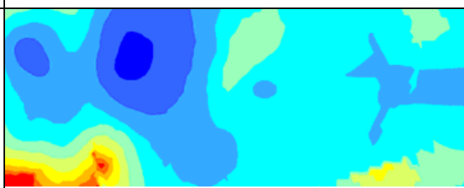
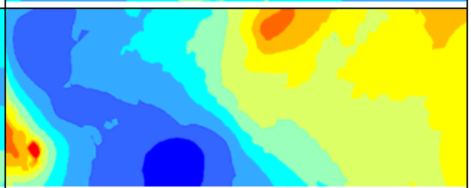
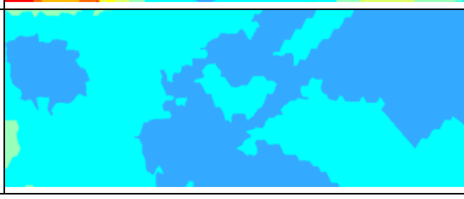
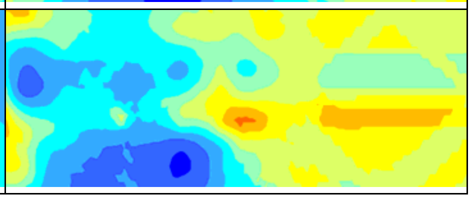
The study indicated that temperatures follow an easily predictable pattern, with slight variations corresponding to lengths of the cold or the warm season. The four seasons characteristic to Romania are easily noticeable: white and purple indicate the summer and green points to the winter, whereas the other colors correspond to the transition seasons.

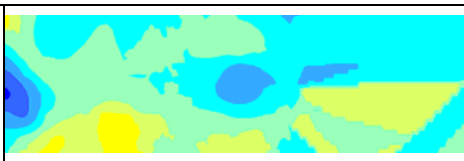
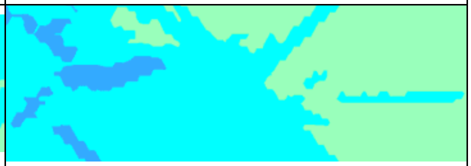
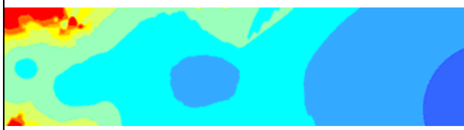
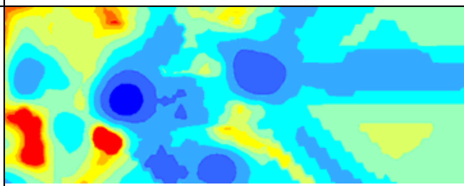
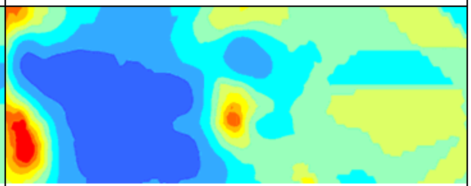
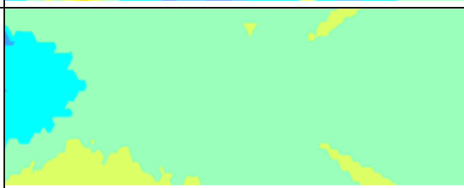
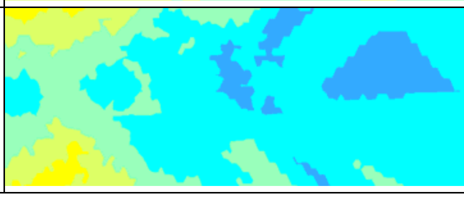
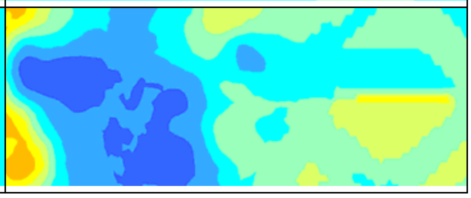
Precipitation clusters appeared more relevant. In our setting, they pointed out dry and wet spells, suggesting the occurrence of hazard events such as floods or droughts. This was particularly obvious in naturally dry/wet regions.

3.2. Example #2: Analysis of the effects of anthropic environmental impact and ecological restoration on the biodiversity of lakes in the Danube Delta

Data were the result of a study aiming to look at the dynamics of several environmental variables in two Danube Delta lakes (Matiţa and Merhei) over a period starting in 1980 and ending in 2007 (*Fig. 1*). The period was marked by three moments. 1980 is the only year in the reference period; 1981 marks the beginning of a deficit in the circulation of water and severe eutrophication due to nitrates and phosphates used in agriculture. The situation improved starting in 1995, when ecological restoration started naturally. Data analyzed in this paper refers to the density of specific groups of organisms, placed from an ecological standpoint on two trophic levels (primary and secondary consumers). Density changes in time were analyzed using the Geostatistical Analyst and the results are displayed in *Table 2*, showing 11 figures for each lake. The first nine figures correspond to the taxonomic groups (first six to the primary consumers or phytophagous organisms, and last three to the secondary consumers or zoophagous organisms), while the last two summarize all groups by trophic levels.

Table 2. Ordinary kriging prediction maps of the densities of primary and secondary consumers in Lakes Matiţa and Merhei, Danube Delta, based on 1980-2007 data. The color ramp ranges from red (low densities) to dark blue (high densities) in the order red - orange - yellow - green - light blue - dark blue.

Taxonomic Group		Lake Matiţa	Lake Merhei
		Year	Year
Ciliata, phytophagous	Month		
Testacea, phytophagous	Month		
Lamellibranchia, phytophagous	Month		
Rotatoria, phytophagous	Month		
Cladocera, phytophagous	Month		
Copepoda, phytophagous	Month		

Taxonomic Group		Lake Matiţa	Lake Merhei
		Year	Year
Rotatoria, zoophagous	Month		
Cladocera, zoophagous	Month		
Copepoda, zoophagous	Month		
All phytophagous	Month		
All zoophagous	Month		

The results suggest that some groups were more abundant in the beginning, but anthropic impact had caused their extinction (phytophagous *Ciliata*, *Lamellibranchia*, and all phytophagous organisms together), while the restoration had favored other groups (phytophagous *Testacea* and *Rotatoria*, and zoophagous *Cladocera*, especially in Lake Matiţa). Some other species seem not to be affected by either the impact or restoration (phytophagous *Lamellibranchia*, and all zoophagous organisms together), or seem even to strive during the impact period (zoophagous *Copepoda* in Lake Matiţa). With respect to the month, results show that most groups reach their maximum density during the warm season, but some others are more abundant in the cold periods (phytophagous *Copepoda*, *Testacea*, and *Lamellibranchia* in Lake Merhei). Also, differences between the time distributions could easily be seen between the two lakes; these differences are attributable to the geography and biophysical parameters of the two sites.

4. Conclusions

The proposed methodology emphasized in the first study the temporal variability of the climatologic phenomena, and could provide an approach to the identification and qualitative prediction of one of the most debated issues, climate changes. In the second study, the proposed approach allowed for the exploration of the reaction of analyzed species to both anthropic impact and ecological restoration, leading to conclusions relevant to the ecology of the groups. Therefore, the methodology had shown a significant potential for being used as an exploratory research tool in climatology, ecology, and other fields as well.

References

1. Cheval, Sorin, and Petrişor, Alexandru-Ionuţ, **"Predicting Average Monthly Temperatures and Precipitations Using the Geostatistical Analyst"**, 33rd Annual Meeting of the South Carolina Chapter of the American Statistical Association, Columbia, SC, 2003.
2. Cheval, Sorin, Baciu, Mădălina, and Breza, Traian, **"An investigation into the precipitation conditions in Romania using a GIS-based method"**, *Theoretical and Applied Climatology* 76, no. 1-2, 2003, 77-82.
3. Dragomirescu, Liviu, **Biostatistica pentru începatori,** Bucureşti: Editura Constelaţii, 1998), 102.
4. Dumitraşcu, Monica, Cheval, Sorin, Baciu, Madalina, and Breza, Traian, **"Analysis of Monthly Temperatures and Precipitations in Oltenia (Romania): GIS Application"**, *Forum geografic* 1, 2002, 82-96.
5. Edsall, Robert M., Kraak, Menno-Jan, MacEachren, Alan M., and Peuquet, Donna J., **"Assessing the Effectiveness of Temporal Legends in Environmental Visualization",** Proceedings of GIS/LIS '97 Annual Conference and Exposition, Cincinnati, OH, 1997, 677-685.
6. Tufte, Edward R., **The visual display of quantitative information**, Cheshire: Graphics Press, 1983, 28-39.
7. Ionita, Angela, and Moise, Eugen, **Dicţionar bilingv englez-roman de termeni GIS and TIC**, Bucureşti, Editura ICI, 2001.
8. Johnston, Kevin, Ver Hoef, Jay M., Krivoruchko, Konstantin, and Lucas Neil, **Using ArcGIS Geostatistical Analyst,** Redlands, CA: ESRI, 2001, 131-166.
9. Piegorsch, Walter W., and Edwards, Don, **Notes on Temporal and Spatial Analysis for a Course in Environmetrics**, Statistics Technical Report no. 206, 62P12-01, Columbia, SC: University of South Carolina, 2001, 35-79.
10. Rigaux, Philippe, Scholl, Michel, and Voisard, Agnès, **Spatial Databases with Application to GIS,** San Diego, CA: Morgan Kaufman Publishers, 2001, 28.
11. SAS Institute Inc., **Spatial Prediction Using the SAS® System**, SAS/STAT® Technical Report, Cary, NC: SAS Institute, 1996, 9-12, 49-52.
12. Tukey, John W., **Exploratory Data Analysis**, Reading, MA: Addison-Wesley Publishing Co., 1977, v-vii.