TESTING THE PERFORMANCE OF DIFFERENT SPATIAL INTERPOLATION TECHNIQUES ON MAPPING SHORT DATASERIES OF PRECIPITATION PROPRETIES

COJOCARU ŞTEFANA[1], BREABĂN IULIANA GABRIELA[1]

ABSTRACT. Testing the performance of different spatial interpolation techniques on mapping short dataseries of precipitation properties. Spatial interpolation, in the context of spatial analysis, can be defined as the derivation of new data from already known information, a technique frequently used to predict and quantify spatial variation of a certain property or parameter. In this study we compared the performance of Inverse Distance Weighted (IDW), Ordinary Kriging and Natural Neighbor techniques, applied in spatial interpolation of precipitation parameters (pH, electrical conductivity and total dissolved solids). These techniques are often used when the area of interest is relatively small and the sampled locations are regularly spaced. The methods were tested on data collected in Iasi city (Romania) between March – May 2013. Spatial modeling was performed on a small dataset, consisting of 7 sample locations and 13 different known values of each analyzed parameter. The precision of the techniques used is directly dependent on sample density as well as data variation, greater fluctuations in values between locations causing a decrease in the accuracy of the methods used. To validate the results and reveal the best method of interpolating rainfall characteristics, leave-one – out cross-validation approach was used. Comparing residues between the known values and the estimated values of pH, electrical conductivity and total dissolved solids, it was revealed that Natural Neighbor stands out as generating the smallest residues for pH and electrical conductivity, whereas IDW presents the smallest error in interpolating total dissolved solids (the parameter with the highest fluctuations in value).

Keywords: Interpolation, IDW, Ordinary Kriging, Natural Neighbor, Cross-validation.

1. INTRODUCTION

The overall quality of the atmosphere in an area can be appreciated by analyzing the physico –chemical characteristics of precipitation. Rain has the ability to retain certain air pollutants, their contact with water molecules producing new compounds, potentially harmful to the environment. Detailed knowledge of the chemical composition of rainwater is important for many studies regarding the prevention, understanding or elimination of pollution.

[1] Alexandru Ioan Cuza University, Faculty of Geography and Geology, Iaşi, Romania
e-mail: stefana_g_cojocaru@yahoo.com

At these levels, point sampling data does not offer a clear, complete, image, current studies needing spatially continuous data. This kind of data can be obtained by spatial interpolation, a process by which unknown values can be predicted based on known values of adjacent points within the same region (Li & Heap, 2014).

Spatial interpolation of precipitation data set is difficult because of the many particularities of the natural and urban environment, which generate differences, even in neighboring areas. Therefore, it is important to identify the best Spatial Interpolation Methods (SIMs) which can be used in mapping the qualitative parameters of precipitation, different SIMs being suitable for different data types (Chiu, Lin, & Lu, 2009). The aim of the present study is to identify the best SIM for mapping a small data set of chemical parameters of precipitation

There are many factors that affect the final result of each mapping method, including sampling density, spatial distribution of samples, sample clustering, surface type, data variance, data normality, grid size and resolution, as well as the different ways in which they combine and interact (Li & Heap, 2011). The methods chosen for testing are Inverse Distance Weighted (IDW), Natural Neighbor (NaN) and Ordinary Kriging (OK), techniques which usually generate good results when applied to short and irregularly sparsed data sets. Validation of the results and identifying the best method for modeling precipitation was achieved using leave-one-out cross validation technique.

2. STUDY AREA

The study was conducted in Iasi (95060 km^2), a city located in north - eastern Romania, between 47^{o}13' and 47^{o}05' north latitudes and 27^{o}38' and 27^{o}41' east longitudes (Fig. 1). The city is located at the contact area between a plain and a plateau, the terrain being a very hilly one.

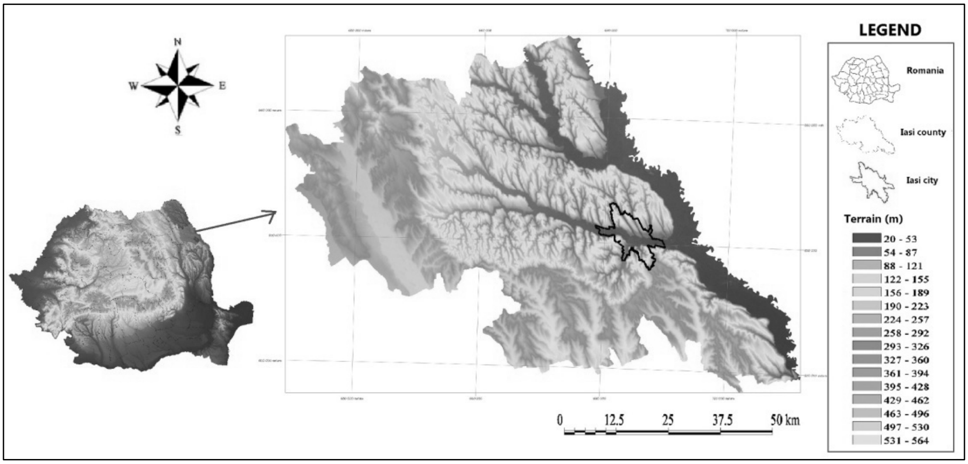

Fig. 1. Location of Iasi city in Romania and in Iasi county

Bahlui river runs thought the center of Iasi, generating different forms of terrain on each of its sides. On the left side the relief consists in hills of lower heights, whereas the right side of the river is defined by high plateaus (reference Barbu N. et. al., 1987). The vertical deviation is 342 m, varying from 30 m (in the river valley) to 372 m.

The city is located in a temperate climate area. This type of climate is characterized by cold winters, hot summers (with temperatures reaching up to 20°C) and springs with moderate temperatures (10.6°C). The average air temperature between the years 1961 and 2011 was 10.6°C (Alexe C., 2012). The average yearly rainfall in the area is about 574.3 l/m^2, with maximum rainfall quantities in June and minimum quantities in February The most considerable quantities of rain usually fall during the summer months (36.9% of the total), followed by spring (25.8%), fall (21%) and winter (16.3%).

3. PRECIPITATION DATA

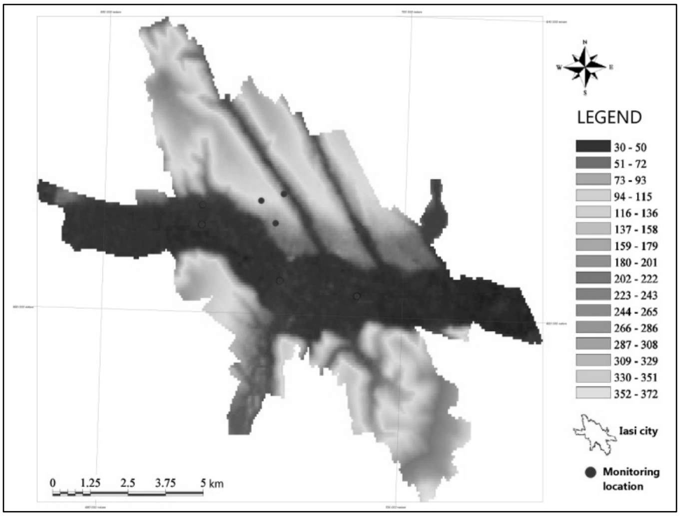

Fig. 2. Distribution of the 7 sampling sites in Iasi city

Iaşi does not have a dense network of precipitation monitoring stations; the data is collected from only 2 or 3 locations in the city by Iasi Environmental Protection Agency.
The precipitation data used in the study was collected in March – May 2013 using home constructed rain gauges placed in 7 different locations around the city (Fig. 2). The devices have a 201 cm^2 surface area and were positioned in open areas, away from interfering objects or vegetation, at a height of 30 cm or more above ground. Seeing as the analysis performed was qualitative and not quantitative, parameters regarding the total amount of precipitation produced at each event were not registered. The only rainfall events taken into account were those that produced a minimum of 2 mm rain, a total of 13 such events taking place in March – May 2013.

Water samples were collected in 50 ml sealed containers and analyzed in the laboratory. The investigated parameters were pH, electrical conductivity and the total amount of dissolved solids in water. Fluctuations in the values of each of those characteristics depend on the amount of pollutants in the atmosphere. To ensure accurate data, quality control of the rain gauges was performed, the devices were cleaned using distilled water less than 24 hours before each rainfall. The water samples were analyzed immediately after being collected, if this was not possible, they were refrigerated at 2°C and examined the following day. From the 13 episodes of rain, a total of 91 samples were validated for analysis.

The final result of the spatial interpolation process can be affected by a number of different factors, among which is the sample density. In the past it was believed that the accuracy of the final result is higher with increasing the density of the sampling sites. This theory has been dismissed in recent studies, Li, J., and Heap, A. D. (2011) arguing that the impact of sample density is insignificant in the overall performance of the interpolation method. The distance for two neighboring locations in the study varies from 60 to 2500 m.

4. DATA INTERPOLATION

Tested were some of the methods most used methods in spatial interpolation of climatic data sequences, Inverse Distance Weighted (IDW), Ordinary Kriging (OK) and Natural Neighbor (NaN). Among them there are both similarities and differences in terms of how they work with the data. All three are local interpolation methods, which means that in assessing the unknown values, the methods use the nearest, already known values (Li & Heap, 2014). Also, while OK is a stochastic method, IDW and NaN are deterministic methods. That means that, in modeling, OK uses statistical as well as mathematical methods and can address the concept of randomness, and IDW and NaN do not take into account the idea of uncertainty.

The same paper states that IDW and NaN methods are suitable for a quick interpolation of sparse, regular or irregularly dispersed data samples, while kriging methods work best when data series are larger and a variogram can be computed. The data interpolation methodology used in the present study is shown in Fig. 3.

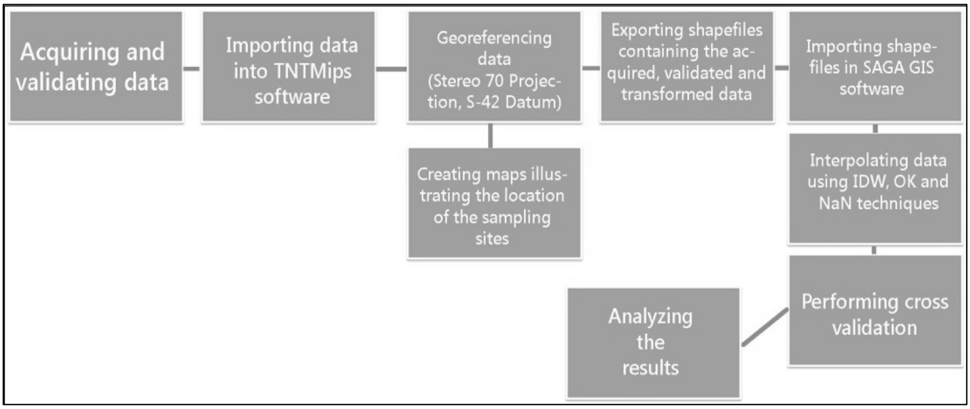

Fig. 3. Data interpolation methodology

The algorithm used by IDW interpolation method assigns cell values based on measured values around the point, relying on the assumption that, as the elements are closer together, the more similar they are. Thus, it is presumed that the influence of an already known value is inversely proportional with the distance to the value that is to be determined. IDW method interpolates a value for each cell in the resulting raster using a series of vector – type input data (“Z”). The “weigh” of each Z depends on the proximity of the data we wish to estimate. The weights are measured in base 10 and the greater the power of ten, the more the influence of farther points diminishes. In the present study, IDW was estimated using the power of 2.

Natural Neighbor is a deterministic method used to estimate values based on modeling a network of polygons, known as Thiessen polygons or Voronoi diagrams (Li & Heap, 2014). It applies well on unevenly distributed data series, an advantage being the fact the search distance for each neighboring point and the weight of their value is calculated automatically.

Kriging is an interpolation method based on the idea that the spatial variation of a parameter is oftentimes so uneven that it can’t be modeled using just a single function. This method is a statistical approach to interpolation, which assumes that the input values generate a continuous surface where the closest points will have similar values and the farthest ones will be independent. Hence, kriging interpolates a value for unknown points by calculating a weighted average of the closest known values (Robinson & Metternicht, 2006).

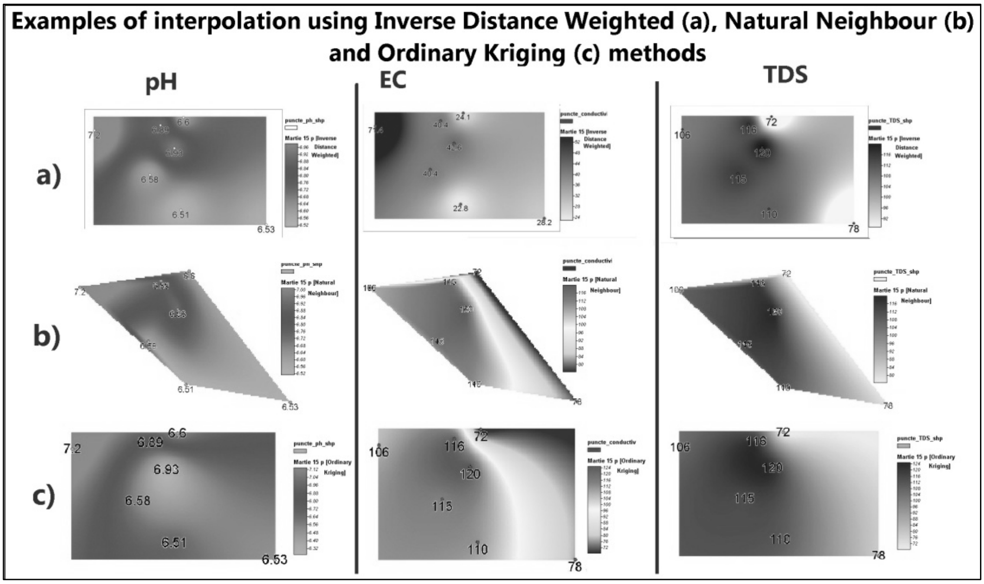

Fig. 4. Examples of each interpolation technique used, for pH, electrical conductivity and total dissolved solids

5. RESULTS AND VALIDATION

In order to certify the results of each interpolation method, cross validation method was applied, which excludes during the interpolation process, by rotation, each observed point (Joseph, Sharif, Sunil, & Alamgir, 2013). A raster was generated for each rainfall event (13 in total) and for all three precipitation parameters (pH, EC and TDS) using all known data. Afterwards the rasters were re-generated, leaving out, one at a time, each of the 7 known values. The estimated values were then compared with the observed ones, determining the error for each of the methods used.

To estimate the overall performance of each method, a common used practice is calculating the bias (or mean absolute error, MAE) and the root mean squared error (RMSE) (Wagner, Fiener, Wilken, Kumar, & Schneider, 2012). Both MAE and RMSE are statistical measures which assesses the errors between the values observed on site and the ones predicted by the interpolation method (Arun, 2013),(Teegavarapu, Meskele, & Pathak, 2012). The formulas used for calculating MAE and RMSE are presented below:

$$MAE = \frac{1}{n}\sum_{i=1}^{N}|p_i - o_i|$$

and

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2\right]$$

where: n is the number of samples with a known value; o are the observed values; p are the predicted values; ō is the mean of observed values; o′i is oi−ō; and p′i is pi−ō (Li & Heap, 2011). The errors that resulted for each method and each analyzed parameter, calculated by applying the MAE and RMSE formulas are presented below (Table 1).

Precipitation parameter	Error	IDW	NaN	OK	Best performance
pH	MAE	0.0463	0.0483	0.008	OK
	RMSE	0.0832	0.0606	0.0873	NaN
EC	MAE	3.1617	2.6156	5.659	NaN
	RMSE	6.7703	5.3772	9.7484	NaN
TDS	MAE	3.9706	4.3013	2.1833	OK
	RMSE	11.9016	11.7216	22.1062	IDW

Table 1. MAE and RMSE calculated errors

The analysis of the errors calculated using the two statistical techniques shows that Natural Neighbor performs best when applied to the electrical conductivity (EC) collected data, and that it is also the best overall performer. Ordinary Kriging had the smalles errors for MAE for both pH and TDS, whereas

Inverse Distance Weighted was found to generate the smallest error only in RMSE applied to TDS.

6. CONCLUSIONS

Knowing the qualitative parameters of precipitation is important to numerous studies, a complete and spatially continuous set of data being critical. This can be achieved using different interpolation techniques, suited for each particular type of data wished to be modeled. Three of the most used methods in interpolating climate data are Inverse Distance Weighted, Natural Neighbor and Ordinary Kriging. The difficulty in mapping climate elements, especially when working with short data series, makes it imperative to test and determine the best methods that can be used.

The data used for testing in the present study was collected in Iasi city, from 7 different locations over a period of 3 months. A total of 91 water samples were validated for analysis. The resulting data was processed and modeled using each of the 3 interpolation techniques.

Cross validation was used in assessing the performance of each method. Mean absolute error (MAE) and root mean squared error (RMSE) statistical methods were then applied to determine the errors between what was estimated and what was observed. The results concluded that, overall, Natural Neighbor had the best performance in the interpolation of climate data, followed by Ordinary Kriging and, finally, Inverse Distance Weighted.

REFERENCES

1. Alexe C. (2012). *Clima şi topoclima municipiului Iaşi şi a ariei metropolitane*. Editura Universitatea Alexandru Ioan Cuza. Iaşi.
2. Arun, P. V. (2013). *A comparative analysis of different DEM interpolation methods*. The Egyptian Journal of Remote Sensing and Space Science. doi:10.1016/j.ejrs.2013.09.001
3. Barbu N. et al. (1987). *Geografia Municipiului Iaşi*. Editura Univ. Al. I. Cuza. Iaşi.
4. Chiu, A. C., Lin, P., & Lu, K. (2009). *GIS-based Tests for Quality Control of Meteorological Data and Spatial Interpolation of Climate Data GIS-based Tests for Quality Control of Meteorological Data and Spatial Interpolation of Climate Data*, 29 (4), 339–349.
5. Joseph, J., Sharif, H. O., Sunil, T., & Alamgir, H. (2013). *Application of validation data for assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents*. Environmental pollution (Barking, Essex : 1987), 178, 411–8. doi:10.1016/j.envpol.2013.03.035
6. Li, J., & Heap, A. D. (2011). *A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors*. Ecological Informatics, 6 (3-4), 228–241. doi:10.1016/j.ecoinf.2010.12.003

7. Li, J., & Heap, A. D. (2014). *Spatial interpolation methods applied in the environmental sciences: A review*. Environmental Modelling & Software, 53, 173–189. doi:10.1016/j.envsoft.2013.12.008
8. Robinson, T. P., & Metternicht, G. (2006). *Testing the performance of spatial interpolation techniques for mapping soil properties.* Computers and Electronics in Agriculture, 50(2), 97–108. doi:10.1016/j.compag.2005.07.003
9. Teegavarapu, R. S. V., Meskele, T., & Pathak, C. S. (2012). *Geo-spatial grid-based transformations of precipitation estimates using spatial interpolation methods*. Computers & Geosciences, 40, 28–39. doi:10.1016/j.cageo.2011.07.004
10. Wagner, P. D., Fiener, P., Wilken, F., Kumar, S., & Schneider, K. (2012). *Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions*. Journal of Hydrology, 464-465, 388–400. doi:10.1016/j.jhydrol.2012.07.026 .