Comparative Study Between Geostatistical Models Applied to Analyze the Phenomena Caused by Geomorphological Processes

ANA-CORNELIA BADEA

Surveying and Cadastre Department
Technical University of Civil Engineering Bucharest, Faculty of Geodesy
Lacul Tei Blvd., 124, 2nd District, Bucharest
ROMANIA
badeacadastru@gmail.com   http://www.geodezie.utcb.ro

Abstract: In this paper we want to emphasize the importance and the main application of geostatistical analysis. We want to highlight the link between hazard maps and geostatistical methods. In the case study were applied some geostatistical methods in surface erosion. In conclusion we have a comparative study about these methods.

Key-Words: -geostatistical methods, GIS, risk maps, ArcGIS Geostatistical Analyst, IDW, Kriging, Radial Basis method

1 Introduction

Geostatistical representations can be defined as analytical maps based on field observations and auxiliary information created with a software that calculates the locations of interest of the study area.

Exists many software [9] that can be used to make qualitative geostatistical analyzes. In table 1 are highlighted the main geostatistical software products.

A software package can offer all geostatistical methods developed in the last 20 years, but itcan scare away potential users by its price or design.

In particular for academics,price and transparency typically drive the choice of geostatistical software. Consultingcompanies and federal agencies are likely to favor products that do not requireadvanced statistical background and provide all necessary functionalitieswithin a single package. [9]

Geostatistics is composed of statistical techniques adapted for application to spatial data. [10]

The aim of geostatistical analysis consists of the systematic errors, which leads to a realistic interpretation of the data and then by choosing the model that reflects in a very appropriate way the information about variability of the studied entities.

Geostatistics is closely related to the interpolation process, which is actually the art of design results available data, providing a clue of corresponding hypothesis. In addition, it is also used to predict variable values on the basis of point-sampled from an area of interest, a method known as spatial prediction. [8]

Today geostatistics is not only used to analyze the data point type, but more than that, in combination with various GIS layers is used to explore spatial variation of remote sensing data to quantify and image noise filtering (ie filling holes / pixel) to improve the generation DEM's and their simulation.

Table 1 – The main geostatistical software products – adapted from [9]
Cost: H high, M moderate, L low, F free

Name	Code	Cost[a]	Reference
Agromet	C++	F	Bogaert et al. (1995)
AUTO-IK	Fortran	F	Goovaerts (2009)
BMELib	Matlab	F	Christakos et al. (2002)
COSIM	Fortran	F	ai-geostats website
EVS (C-Tech)		H	C Tech Development Corporation
GCOSIM3D/ISIM3D	C	F	Gomez-Hernandez and Srivastava (1990)
Genstat		F,L	Payne et al. (2008)
GEO-EAS	Fortran	F	Englund and Sparks (1988)
GeoR	R	F	Ribeiro and Diggle (2001)
Geostat Analyst		H	Extension for ArcGIS
Geostatistical Toolbox		F	Froidevaux (1990)
Geostokos Toolkit		H	ai-geostats website
GS+		M	Robertson (2008)
GSLIB	Fortran	F	Deutsch and Journel (1998)
Gstat	C,R	F	Pebesma and Wesseling (1998)
ISATIS (Geovariances)		H	www.geovariances.com
MGstat	Matlab	F	ai-geostats website
SADA (UT Knoxville)		F	Spatial analysis and decision assistance
SAGE2001		M	Isaaks (1999)
SAS/STAT		H	SAS Institute Inc. (1989)
S-GeMS	C++	F	Remy et al. (2008)
SPRING		F	Camara et al. (1996)
Space-time routines	Fortran	F	De Cesare et al. (2002)
STIS (TerraSeer)		M	AvRuskin et al. (2004)
Surfer		M	Golden Software, Inc.
Uncert	C	F	Wingle et al. (1999)
Variowin		F	Pannatier (1996)
VESPER		F	Minasny et al. (2005)
WinGslib	Fortran	L	www.statios.com

Table 2 – List of functionalities for main geostatistical software – adapted from [9]

Name	Data	V	K	CK	IK	MG	S	G
Agromet	2D	X	X	X				
AUTO-IK	2D	X			X			
BMELib	3D, ST	X	X	X			X	
COSIM	2D						X	
EVS (C-Tech)	3D	X	X		X			X
GCOSIM3D/ISIM3D	3D						X	
Genstat	3D	X	X	X				
GEO-EAS	2D	X	X					
GeoR	2D	X	X				X	
Geostat Analyst	2D	X	X	X	X	X		X
Geostatistical Toolbox	3D	X	X	X				
Geostokos Toolkit	3D	X	X	X	X		X	
GS+	2D	X	X	X			X	
GSLIB	3D	X	X	X	X	X	X	
Gstat	3D	X	X	X			X	
ISATIS	3D	X	X	X	X	X	X	X
MGstat	3D, ST	X	X					
SADA	3D	X	X		X			X
SAGE2001	3D	X						
SAS/STAT	2D	X	X					
S-GeMS	3D	X	X	X	X	X	X	
SPRING	2D	X	X		X		X	X
Space-time routines	2D, ST	X	X					
STIS (TerraSeer)	2D, ST	X	X			X	X	X
Surfer	2D	X	X					
Uncert	3D	X	X				X	
Variowin	2D	X						
VESPER	2D	X	X					
WinGslib	3D	X	X	X	X	X	X	

Notes: V variography, K kriging, CK cokriging, IK indicator kriging, MG multi-Gaussian kriging, S simulation, G GIS interface

From all these software possibilities we made the choice of ArcGIS Geostatistical Analyst.

Today geostatistics is not only used to analyze the data point type, but more than that, in combination with various GIS layers is used to explore spatial variation of remote sensing data to quantify and image noise filtering (ie filling holes / pixel) to improve the generation DEM's and their simulation.

A geostatistician must ask themselves the following problems to analyze a case study:

- How does an entity in space?
- What controls its variation in space?
- Where should be located evidence points to describe the spatial variability?
- How many samples are required to represent spatial variability?
- What is the value of the variable in a new location?
- What is the uncertainty of the estimate?

An important difference between the conventional and geostatistical representation is that the former is based on statistical techniques, quantitative. Unlike traditional approaches of representation that is based solely on the use of empirical knowledge, representation geostatistics is based on actual measurements and algorithms. [8]

Geostatistical analysis provides two groups of interpolation techniques: deterministic and geostatistical. [1]

Based on these geostatistical analyzes can be made risk maps. The necessity and importance of studying the risk phenomena and processes arises from the impact of events and/or natural phenomena have on man and his activities. The concept of risk includes several components: the thing that can happen (phenomenon) or environmental context, the disaster, the consequences that can occur, the relative uncertainty of the event itself. [2]

Identification and location of geomorphological processes plays an important role in defining, designing, developing and implementing local development strategies at regional and national level. Once identified geomorphological processes can proceed to achieve the necessary risk maps in urban studies and planning, from which it will be identifed the optimal direction of expanding settlements, local economic (restrictions, conditioning), real estate investment, transport networks etc..

The study conducted by this paper is based on the workflow with Geostatistical Analyst extension provided byESRI. Geostatistical Analyst is a link between geostatistics and GIS, as an important step in achieving:

- exploratory analyzes of spatial data;
- structural analysis;
- predictions.

2 The problem of quantitative estimation of surface erosion

Surface erosion is a geomorphological process which is difficult to identify in terms of visual impact, being visible as yellowish-white spots on a darker background.

The importance of surface erosion lies in the fact that this geomorphological process is extensively during heavy rains and is widespread in Romania. The effects are important and, therefore, is perfectly justified the application of empirical methods of surface erosion amount for calculation of material eroded by the action of rainfall.

Among the mathematical models applied over time, the most used is the Universal Soil Loss Equation (USLE), proposed by Wischemeier and Smith.

The factors on which depends the erosion of the surface are:

- landform (slope, slope shape and slope aspect);
- specific climatic elements (related to precipitation and temperature);

- lithology (rock classification based on the resistance to erosion);
- soil type structure, texture;
- vegetation (being important as soil protection factor);
- anthropic (human activities, with the role of increased or decreased erosion).

The formula behind this model is:

$$A = R * K * L * C * P \qquad (1)$$

in which:

- A - soil loss in tons/ha /year, amount to be calculated;
- R - factor of aggressiveness climate;
- K - factor of soil erosion;
- L - length of slope
- C - factor which highlights the influence of the land use
- P - factor which highlights the influence of anti-erosion works
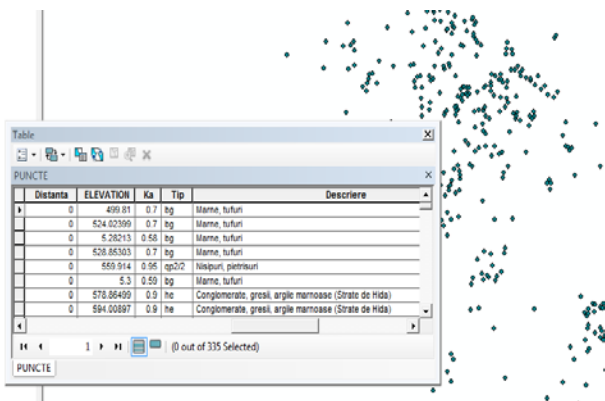
Fig. 1 – Erosion factor K in the study area – west of Romania
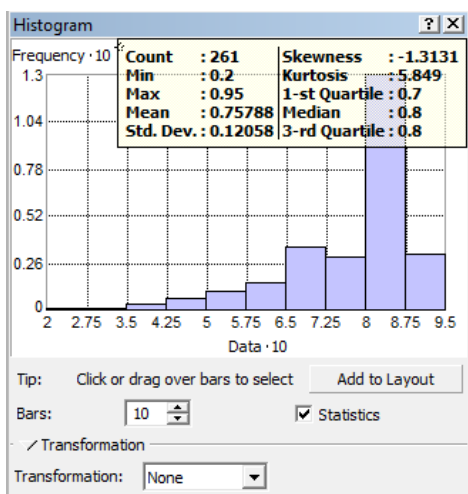
Fig. 2 – Histogram with erosion factor K in the study area – west of Romania

Erosion factor K is the rate of erosion per unit of erosion index of land considered standard (benchmark). This factor is influenced primarily by the structure and texture of soil organic matter. In the figure 1 and 2 is emphasized the K factor values for a pilot area in the west of Romania.

3 Analysis of land erosion in the area working with ArcGIS Geostatistical Analyst

The followed workflow consists in the phases:

- Data representation which represents converting of the information into spatial information through a GIS.
- Exploring data is the use of statistical tools such as frequency distribution graphs to search for a pattern in the data.
- Identification of the appropriate model represents the choice of surface interpolation method to estimate the area of the sites where no measurements were made.
- Execution of diagnosis consists by statistical tests such as Cross-Validation to assess the quality of estimates.
- Comparing models is made to assess the quality of a set of estimates in comparison with others.

3.1 Geostatistical analyzes

The first step to make a proper geostatistical representation is to understanding the variability of the data source. As noted previously, the variability is the result of deterministic and stochastic processes with noise additionally. In other words, the data variability is the sum of two components: the spatial natural variation and inherent noise (due to measurement error). [3]
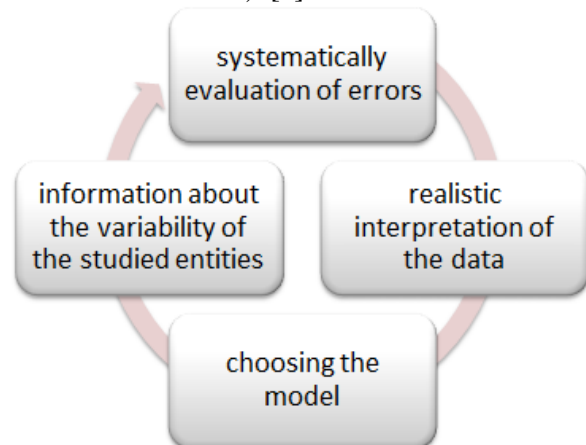

Fig. 3 – Geostatistical analyzes

Typical measurement errors occur during positioning in the field, during the of sampling or laboratory analysis. Ideally, these errors are minimized because they are not the main concern of specialists. They are interested in natural spatial variation which is due to the physical processes that can be explained by a mathematical model.

The second step for a reliable modeling consists in taking into account of all aspects of the natural variation. Although the spatial prediction of environmental variables is primarily focused on geographic variability, there are many other aspects of the natural soil variation that are often overlooked by cartographers: vertical aspects, temporal aspects and scale aspects.

First of all, we are looking for geographical variation (2D). The spatial prediction results are being viewed as 2D maps or cross sections. Some environmental variables such as soil horizons thickness, plant species and soil type have not a third dimension (they refer only to the surface of the earth). Others as temperature, population density, etc.. could be measured at different altitudes even above Earth's surface. The geographic variation could be modeled using a continuous model, discrete or complex. [4]

Another specific variation aspects are vertical variations (3D) Many environmental variables vary with the depth and altitude. Considering variables like temperature, to explain their vertical distribution can often be more difficult than horizontal distribution. The transition between different soil layers can also be gradual or sudden, which requires a double mixed model for prediction of soil variation in 3D space. [5]

Another kind of variatioation ise the temporal variation. The environmental variables connected with plant and animal species vary not only with the season, but often by smaller time periods. Even soil characteristics such as pH, the nutrients, water saturation levels and their content may vary over several years, a season or only a few days. The temporal variability make the geostatistical mapping to be more complex and expensive. The maps of environmental variables produced for two different time periods may vary significantly. This means that most maps are valid only for a period of time. In many cases, seasonal periodicity of environmental variables is regular, so that prediction not necessarily requires new samplings. Another important factor is the size of the mesh that represents the geographic area and is linked to the concept of scale. For spatial prediction, there are two support sizes: the size of the sample and the grid resolution of auxillary maps.[6]

Field observations are collected as sample points. The size of the auxiliary paper support is often much larger than the sample area (eg auxiliary variables are typically the average, while the environmental variables may describe local features at micro level). As a result, the correlation between the paper and the measured auxiliary variables is often low or insignificant.

If modeling predictions are focusing only on geographical component (2D), then samples must be taken on certain conditions: the same season / season, same depth, same access area. This means that each 2D map including environment variables should always indicate a time reference (range), a vertical applicable size and a support size like scale. [11]

3.2 IDW Method

IDW interpolation method explicitly implements the assumption such that the points that are close together are more similar than those situated at greater distances. To predict a value for any unmeasured location, IDW will use the measured values surrounding the prediction location. These closest measured values to the prediction location will have more influence on it than those which are situated far away.

Thus, IDW assumes that each measured point has a local influence that diminishes with distance. The points situated near the location of the prediction are weighted more than the others more far away hence where the name Inverse distance weight.

The general formula is:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \quad (2)$$

where:
- $\hat{Z}(s_0)$ is the value that would be predicted for the location s_0
- N is the number of measured points surrounding the prediction location and will be used for prediction
- λ_i are the weights associated with each measured point that will be used. These weights will decrease along with the distance
- $Z(s_i)$ is the observed location at si location

The results can be observed in figure 4.

3.3 Radial Basis Method

Being a method for accurate interpolation, the RBF (Radial Basis Functions) techniques differ from global and local polynomial methods (interpolated inaccurately) that does not require that the surface to

pass through all the points. Comparing RBF with IDW, that is a more accurate interpolation method, IDW will not predict the measured values above the maximum measured or below the minimum measured. However, the RBF can predict values above the maximum and below the minimum of the measured sample points. Optimal parameters are determined using Cross Validation. The results obtained in the study area are those from figure 5.
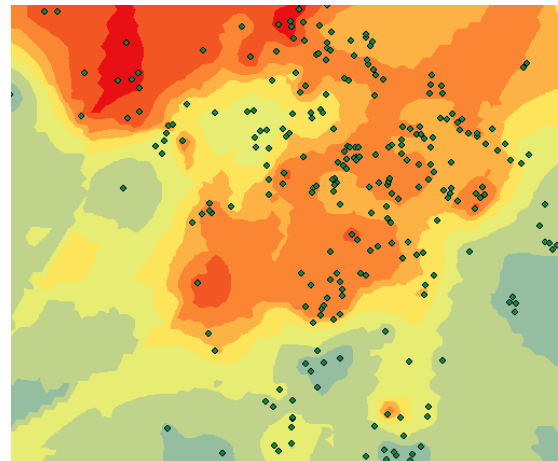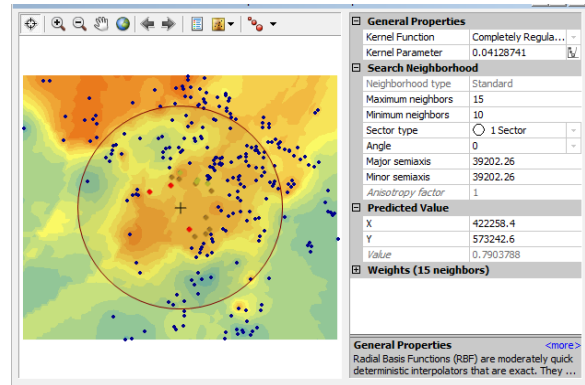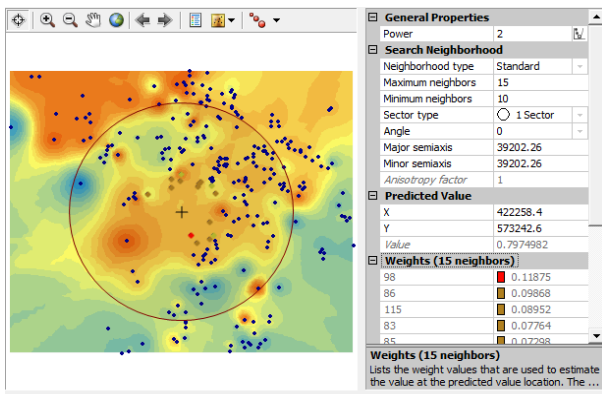

Fig. 4 - Results obtained by IDW method

Fig. 5 - Results obtained by Radial Basis method
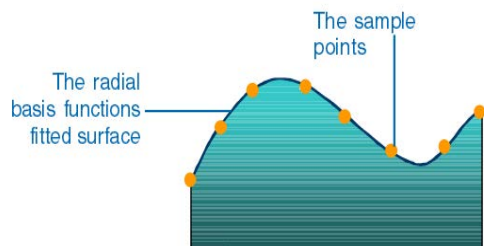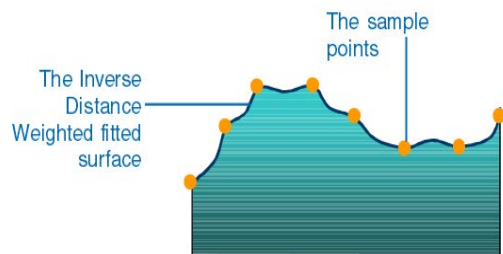
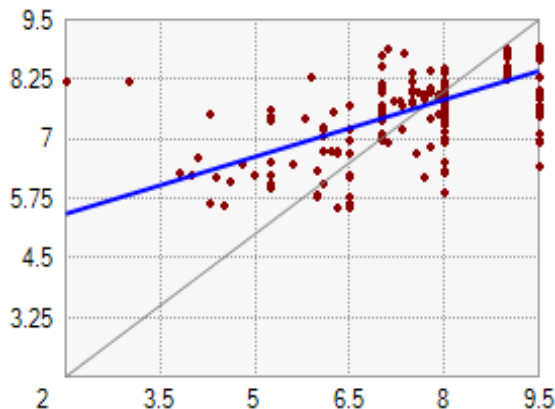Fig. 6 – Differences between IDW interpolation method and Radial Basis interpolation method (adapted from [7])

3.4 Kriging Method

The kriging interpolator is considered the most sophisticated and accurate way to determine the intensity of a phenomenon at unmeasured locations.

Kriging weights surrounding measured values are based not only on the distance between measured points and the prediction location but also on the overall spatial arrangement of the measured points. Except for generating an estimated prediction, kriging can provide a measure of an error, or uncertainty of the estimated surface. Since the estimation variances can be mapped, a confidence placed in the estimates can be calculated and their spatial distribution can be presented on a map to assist in the decision-making process. The prediction standard error maps show a distribution of a square root of a prediction variance, which is a variation associated with differences between the measured and calculated values. The prediction standard error quantifies an uncertainty of a prediction. [12]

Kriging models depend on mathematical and statistical models. The difference is made by adding a statistical model that includes probabilities – so the Kriging methods are seprarated from the presented above deterministic methods. This method is based on the concept of autocorrelation.

This type of interpolation involves the following model:

$$Z(s) = \mu(s) + \varepsilon(s) \qquad (3)$$

where μ is an unknown constant.

One of the major problems regarding Ordinary Kriging is whether the assumption of a constant average is reasonable. However, it is a simple method of prediction and has a remarkable flexibility. In figure 6 we have the results obtained by applying the Kriging method.
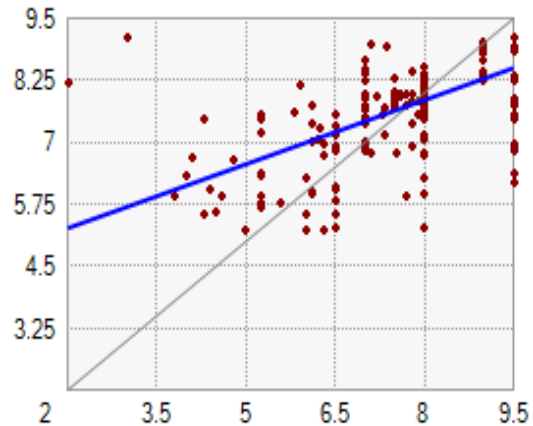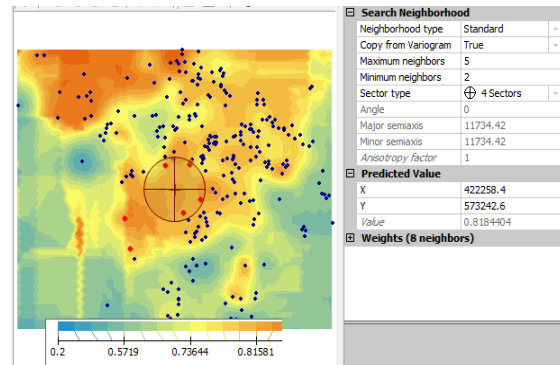
Fig. 6 – Results obtained by Kriging method

# 4 Comparative studies - the problem of comparing models

Comparative studies between the different used models help us to determine how good is a model by comparison with another one. The two geostatistical compared layers can be created from two different models or from the same model but using different parameters. In the first case we are comparing which method is better for the data included in the study and in the second case we are examining the effects of various input parameters on the model when creating the output surface.

This type of comparison using cross validation statistics, placed side by side.

The best model is the one that has:
- standard average close to 0,
- the smallest RMS error of prediction,
- error standard average close to prediction RMS error,
- standard prediction error close to 1.

It is common practice to create more surfaces to identify the better one to solve an existing problem. We can compare systematically different surfaces, by removing the least appropriate and that reaching the model most suitable choice.
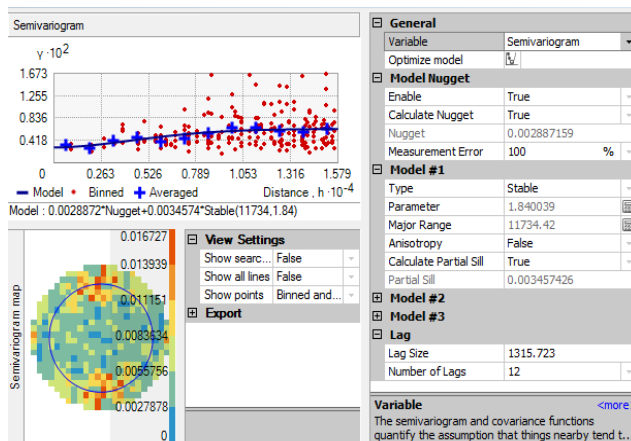
There are two problems when comparing the results of two methods or models: optimality and validity.

For example, the RMS prediction error can be smaller for a model. Therefore, we can conclude that that is the right model. However when we are comparing with another model, RMS prediction error can be closer to the estimated average of standard prediction error. So this is a much more valid model. When the average of estimated standard prediction is close to the RMS from Cross Validation we can consider that we have a good prediction of the standard errors.

In the following we made comparisons for the study area between some models. Taking into account the RMS, we can conclude that the Radial Basis Function model is the most suitable in this case study.

In conclusion we have a comparison from many criteria (surface, working time, assumptions) between the applied methods in the following (adapted from [7]).
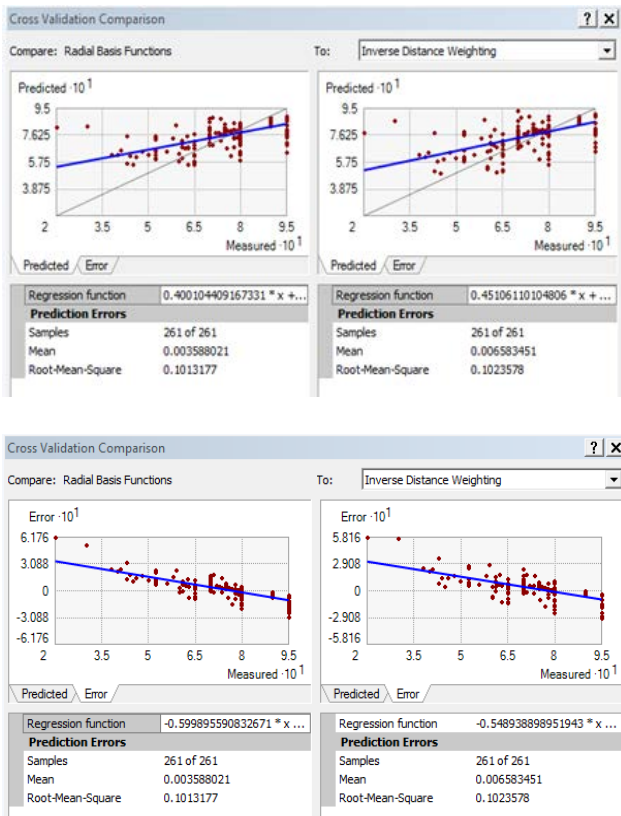

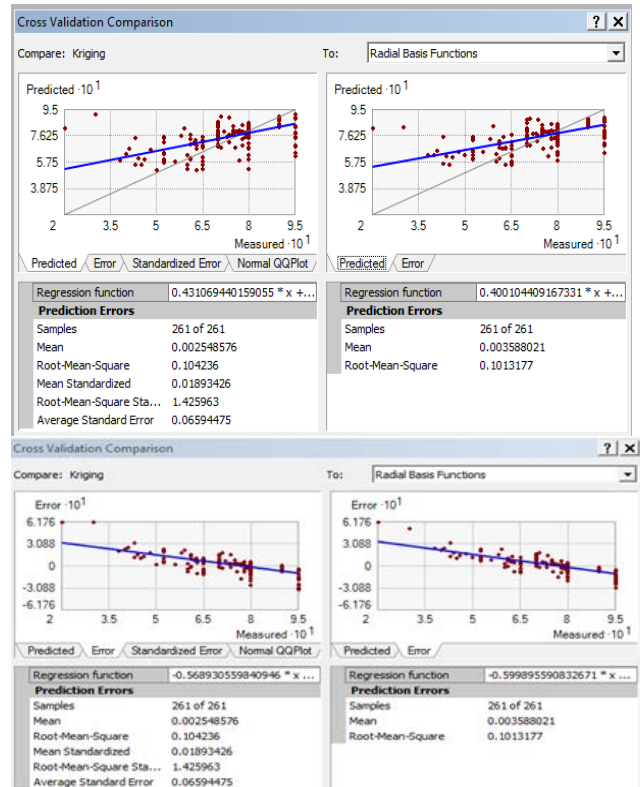Fig.7 – Comparison between Radial Basis method and IDW method

Fig. 8 – Comparison between Radial Basis method and Kriging method

Table 3 - Comparative Study through Geostatistical Analyst analysis methods

Method	Method type	Obtained surface	Working time/ Modeling time
IDW	Deterministic	Prediction	Fast/ Fast
Radial Basis Functions	Deterministic	Prediction	Medium/ Medium
Kriging	Stochastic	Prediction Standard Prediction Errors	Medium/ Slow

Method	Advantages	Disadvantages	Assumptions
IDW	Few decision parameters	Doesn't exist an assesment of prediction errors	-
Radial Basis Functio	Flexible and automatic – few	Doesn't exist an assesment of prediction	-

ns	decision parameters	errors – authmatic method	
Kriging	Very flexible Many decision parameters	It is meeded a decision about transformations, trends, neighborhoods	A normal data distributio n

4 Conclusion

Modeling soil erosion is quite complicated because landslides vary spatially and temporally depending on many factors and the interaction between them.

It is necessary to know both estimation and prediction for unknown locations.

In this study we analyzed the factor K - the erosion assessment is an important step in understanding soil quality and susceptibility to erosion and to predict soil erosion.

Area that provides the best prediction was that created by Ordinary Kriging model.

Contours represented with gradual colors indicate areas of erosion and some intensity factor K for erodabilitate. Prediction over the overlapping area containing soil type and type of texturing can be concluded about the validity, if the areas of a certain intensity of erosion of the surface of the prediction is given in accordance with the type of ground over which they overlap.

Both in the North and in the South there is a higher intensity of erosion (dark blue) and lower in center. Checking soil types and their texture areas in the North and South, it is observed that textures are sandy loam and clay so the land is prone to erosion. In the East, West and central is observed that the intensity factor of erosion is lower, so the soil has a varied surface texture prediction is consistent with the study area.

This study proves that is very good to combinegeostatistical calculation of risk maps with GIS. This is an important and useful tool for the study of spatial changes in environmental sciences.

References:

[1] Costinescu C., Popescu S., Mierluş-Mazilu I.– *Probabilităţi şi statistică tehnică*, editura Conspress, Bucureşti, 2005

[2] Stupar, T., Badea, A. C. – „*Elemente privind analiza geostatistică pentru predicţia alunecărilor de teren din zona Transilvaniei*", Sesiunea Natională de Comunicări Stiintifice Studentesti, Cluj-Napoca 24 Mai 2013, http://sens-group.ro/sncs/

[3] Stupar, T. - *"Elemente privind analiza geostatistică pentru predicţia erodabilităţii terenurilor din zona Transilvaniei"*, Lucrare de Disertaţie, coordonator ştiinţific Ana-Cornelia Badea, Biblioteca UTCB, 2013;

[4] Stângă, I.C., Minea, I. – "Riscuri şi catastrofe Vol IV -*"Vulnerabilitatea la eroziune a unor soluri zonale din Carpaţii Orientali* ", Cluj-Napoca 2005 http://riscurisicatastrofe.reviste.ubbcluj.ro/Volume/IV_Nr_2_2005/PDF/Stanga_Minea.pdf

[5] R. Mihăilescu, T. Man, M. Oncu– Riscuri şi catastrofe Vol III -*"Evaluarea risuclui de eroziune a solului în bazinul Someşului Mic prin aplicarea modelării GIS* ", Cluj-Napoca, 2004 http://riscurisicatastrofe.reviste.ubbcluj.ro/Volume/III_Nr_1_2004/PDF/Mihaiescu_Man_Oncu.pdf

[6] Danci, I. –*"Riscurile asociate proceselor geomorfologice din aria metropolitană clujeană"*, Teza de Doctorat, Universitatea "Babeş-Bolyai" Cluj Napoca, Scoala Doctorală de Geografie, 2012

[7] www.esri.com - *Geostatistical Analyst Tutorial, Using ArcGIS Geostatistical Analyst Book, Concepts and applications of Kriging, Geostatistical Analyst Tutorial Index* - accesed July 2013

[8] http://dlsr.com.au/2012/06/about-geostatistics/

[9] Goovaerts, P. - *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications, Chapter 7, Springer-Verlag, Berlin, 2010*

[10] Goovaerts, P. -*A coherent geostatistical approach for combining choropleth map and field data in the spatial interpolation of soil properties*. European Journal of Soil Sciences, 62(3), 371-380, 2011

[11] Ortiz, B.V., Perry, C., Goovaerts, P., Vellidis, G., and D. Sullivan.. *Geostatistical modeling of the spatial variability and risk areas of Southern Root-Knot nematodes in relation to soil properties*, Geoderma, 156, 243-252., 2010

[12] http://www.esri.com/news/arcuser/0807/geostatistics.html - accesed in July 2013