

GEOSTATISTICAL ANALYSIS FOR LANDSLIDE PREDICTION IN TRANSILVANIA, ROMANIA

C.-B. Vilceanu¹, A. C. Badea², S. Herban¹

¹ Department Overland Communication Ways, Foundations and Cadastral Survey, Civil Engineering Faculty, Politehnica University Timisoara, Traian Lalescu 2, postal code 300223, Timisoara, Romania - beatrice.vilceanu@upt.ro, sorin.herban@upt.ro

² Surveying and Cadastre Department, Faculty of Geodesy, Technical University of Civil Engineering Bucharest, Bd. Lacul Tei no. 122 – 124, postal code 020396, sector 2, Bucharest, Romania - badeacadastu@yahoo.com

KEY WORDS: landslide monitoring, data processing, geostatistics, geomatics, GIS

ABSTRACT:

Management of spatial data by means of Geographic Information System (GIS) plays an essential role based on the latest achievements in Geomatics domain. Geomatics offers the possibility to share, compare, and exchange data between researcher and users in unambiguous and accessible ways for map production and user-friendly technologies for results communication.

Although sometimes the geodesist's contribution to certain projects for landslide monitoring meant to develop early-warning-systems or risk maps is not adequately appreciated and he is only seen as supplier of measured geometric data, the geodesist has a significant contribution through his abilities regarding the modelling of dynamic systems, like strategic constructions (dams, tall buildings etc.) or landslides and data processing and interpretation.

This study focuses on using geomorphological characteristics to detect the changes and the effects of landsliding using the ArcGIS 10.1 extension, Geostatistical Analyst.

One of the main uses of geostatistics is to predict values of a sampled variable over the whole area of interest, which is referred to as spatial prediction or spatial interpolation.

The extension allows creating a surface from data measurements occurring over an area where collecting information for every possible location would be impossible and gives us the possibility to fully understand the qualitative and quantitative aspects of the data.

By providing us with the opportunity to predict and model spatial phenomena based on statistics and incorporating powerful exploration tools, ArcGIS Geostatistical Analyst effectively bridges the gap between geostatistics and geographic information system (GIS) analysis.

1. INTRODUCTION

Geostatistics represents a branch of statistics specializing in analysis and interpretation of data with geographic reference. In other words, geostatistics consists of statistical techniques adjusted to spatial data.

This discipline has gradually developed and has been applied over time in many areas of the world. Originally it was created to estimate the probability of distributing certain existing resources and their volumes, such as ore classes for mining. Over the years, geostatistics has been applied in various areas such as: oil geology, hydrology, meteorology, geography, forestry, oceanography, and geochemistry.

Geostatistics is closely linked to the interpolation process, which actually represents the art of designing the results of the available data, giving a hint about the appropriate hypothesis.

Geostatistics implies more than data interpolation, being used to predict values based on sampled variables in an area of interest, method known as spatial prediction.

Accurate results can be obtained by using ArcGIS Geostatistical Analyst extension that provides a set of tools for exploring spatial data and generating surfaces by using statistical methods. This extension allows users to create surfaces based on measured data of a surface, where the collection of information for each individual location would be impossible.

The present paper proposes a comparative analysis of the main methods of prediction of the glimes in the Transylvanian area, Romania using Ka – the erosion factor.

2. THE PRINCIPLES OF GEOSTATISTICAL ANALYSIS

Geostatistics analysis actually represents an interpolation of the surface, part of the spatial analysis, and relies on spatial self-correlation. Geostatistical techniques focus on the spatial structure of a variable that represents the mathematical relation between it, as a measured value at a given point and the same type of variable measured at other points at certain distances from the first point. The first law of geography, formulated by Waldo Tobler, assumes that everything is related to any other thing, but the close ones are much more similar as characteristics than the other ones.

Geostatistics analysis uses sample points taken from different locations of a surface and creates (interpolates) a continuous surface. The sample points represent the measurement of a phenomenon e.g. radiation leakage from a nuclear power plant, an oil spill, depth to the water table or altitudes. This type of analysis creates a surface using the values in the measured locations to predict values for each location in that area.

The above mentioned extension, Geostatistical Analyst, provides two groups of interpolation techniques: deterministic and geostatistical. The deterministic ones are based on mathematical functions for interpolation, and the geostatistics are based on both mathematical and statistical methods and can be used to create surfaces and assess the uncertainty of predictions.

To create surfaces, both methods are based on the similarity of the vicinity of the sample points. In addition to providing various interpolation techniques, Geostatistical Analyst also provides support tools. These tools allow users to explore and gain better understanding of the data in order to create the best surface based on the available information.

2.1 The determinist interpolation technique

As mentioned earlier, this technique uses mathematical functions. The most used surface for this technique is the digital model of the terrain based on elevation, but also other measurements can be used (measurements at ground level, basement, atmospheric) to generate continuous surfaces. A major challenge to which a large proportion of GIS modeling is to be faced, is to generate the most accurate surface of the sample data and also to characterize the error and variability of the projected area. New generated surfaces are further used for modeling and analysis in 3D views. By understanding the quality of the data, users can improve the usefulness and purpose of the GIS modeling. This is the Geostatistical Analyst's role.

This technique can be divided into two groups: global and local. Global techniques calculate predictions for the entire dataset, and the local ones calculate predictions from the points measured in the vicinities, which are smaller surfaces than the ones taken into consideration in the study. The Geostatistical Analyst provides global polynomial as the global interpolator and the Inverse Distance Weighted (local polynomial) and the radial base as local interpolators.

A deterministic interpolation can either force the resulting surface to pass through the given values or not. An interpolation technique that predicts an identical value to that measured in the sample location is defined as the exact interpolator. An inaccurately interpolator predicts a value that is different from the measured value. The latter can be used to avoid the sharpening of the bed or gutters in the resulting surface. IDW and radial base are precise interpolators, while global and local polynomials are inaccurate.

As deterministic methods that can be applied are: Inverse Distance Weighted (IDW), the Global Polynomial, the Local Polynomial and the functions of the radial base.

2.2 The geostatistical interpolation technique

Unlike the deterministic interpolation technique, the geostatistical assumes that all values in the study are the result of a random process. A random process does not mean that all events are independent just as throwing a coin. Geostatistics is based on random process with dependency. For example, throwing three coins and seeing if they are heads or tails. The rule of determining how the fourth coin is to be put is: if the second and third are the head, then the fourth should be the same as the first; contrary the fourth will have to be different from the first.

In the spatial or temporal context, such dependency is called self-correlation (Figure 1).



Figure 1. Geostatistics' principle – self-correlation

Until now, deterministic interpolation techniques were based on surrounding measured values or on specific mathematical formula that determines the smoothness of the resultant surface. A second family of interpolation methods consists of geostatistical methods that are based on statistical models that include autocorrelation (statistical relations between measured points). These techniques do not only have the capability to produce predictive surfaces, but can provide measurements of uncertainty or accuracy of predictions.

As added scientific value, geostatistics is synonymous with kriging, which represents the statistical version of interpolation. Kriging meets two requirements: the quantification of the spatial structure and making predictions. The quantification of the structure, known as Variography, is the place where a spatial dependency model can fit the data from the study. To make a prediction for an unknown value of a specific location, kriging will use the right model in variography, spatial configuration of the data, and the measured values of sample points around the prediction location.

Thus, in this category are included the following techniques: Ordinary kriging, Simple Kriging, Universal Kriging, Indicator Kriging, Probability Kriging, Disjunctive Kriging, Cokriging.

3. CASE STUDY

This case study implies the analysis of the erodibility factor influence on a region in Transylvania, Romania and the estimation of the areas susceptible to landsliding. In order to carry out the proposed study, the previously discussed interpolation techniques were applied, and after obtaining the corresponding surfaces, they were compared and chosen the one that gave the best prediction.

The erodibility factor K, represents a quantitative description of the erodibility inherent to a particular type of soil. It also constitutes a measure of the soil particles' susceptibility to the detachment and their transport by precipitation and landslides. For a specific type of soil, the erodibility factor is the erosion rate per unit of erosion index from a land considered standard (standard sample). This factor reflects the fact that different types of soil are eroding in different percentages when other factors affecting erosion (infiltration rate, permeability, total water capacity, dispersion, rain smears, abrasion) are the same. Texture is the main factor that affects the erodibility factor, but also contributes to structure and organic matter. The erodibility of the soil varies between 0.02 and 0.69. (Goldman et al. 1986; Mitchell and Bubenzer 1980)

There are several ways to determine the erodibility factor, but the most commonly used is the diagram method based on the work by Wischmeier et al. (1971) and can be represented by the mathematical formula:

$$K = \frac{[2.1 \cdot 10^{-4} (12 - MO) M^{1.14} + 3.25(S - 2) + 2.5(P - 3)]}{100} \quad (1)$$

where K – soil erodibility factor ($t \cdot ha \cdot hour \cdot ha^{-1} \cdot Mj \cdot l \cdot mm^{-1}$);

M – percentage of granulometric fractions (% common clay + very fine sand fractions) $\cdot (100 - \% \text{ clay})$;

MO – organic matter content (humus) (%);

S – soil structure (structure code);

P – soil's permeability to water (permeability class).

K=0.1 most resistant to erosion;

K=0.8 common clay-sandy – common clay-clay texture soils;

K=0.9 s common clay-sandy – common clay texture soils;
 K=1 highest predisposition to erosion class.

The following diagram (Figure 2) describes the workflow for accomplishing the proposed purpose:

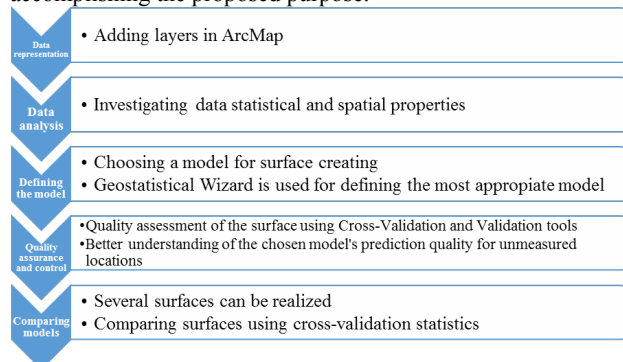


Figure 2. Workflow of geostatistical analysis for landslide prediction in Transilvania, Romania

- 1) **Data representation** consists in converting information to spatial information by means of Geographic Information Systems.

A sample of points surveyed in different locations of Transilvania area, Romania (Figure 3) are to be used. Attributive data of the sample points consists in erodibility factors, points' heights and terrains stratifications.

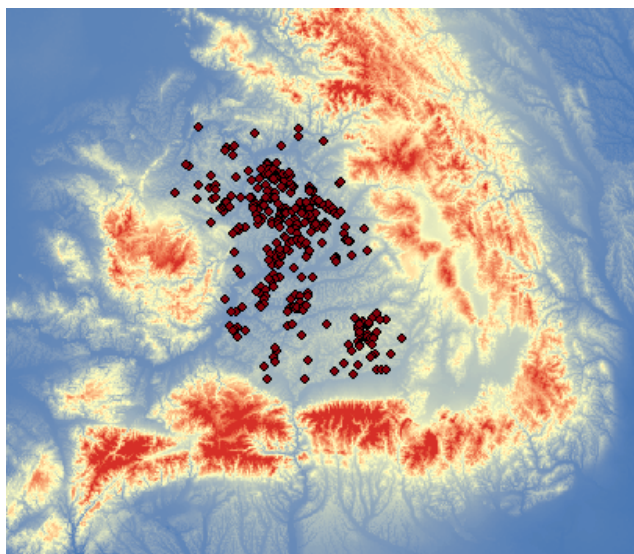


Figure 3. ArcMap data representation

- 2) **Data Analysis** is represented by using statistical methods such as frequency distribution graphics.

This exploration is accomplished through the Exploratory Spatial data Analysis (ESDA) that allows data examination in different ways. Before creating the surface, ESDA provides tools for a better understanding of the investigated phenomenon to make the best decisions about the data. Each tool allows the data to be viewed so that it can be explored and manipulated, facilitating its better understanding. Each view is interlinked to the others so that selecting items in a view will appear selected into the other windows. These types of views are: Histogram (Figure 4), Voronoi Map (Figure 6), Normal QQ Plot (Figure 5), Trend Analysis, Semivariogram/Covariance Cloud, CrossCOVariance Cloud.

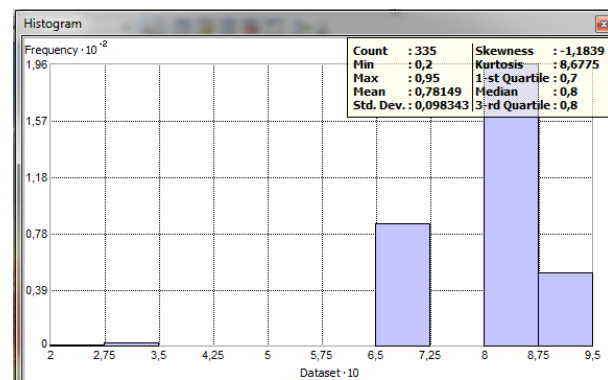


Figure 4. Histogram

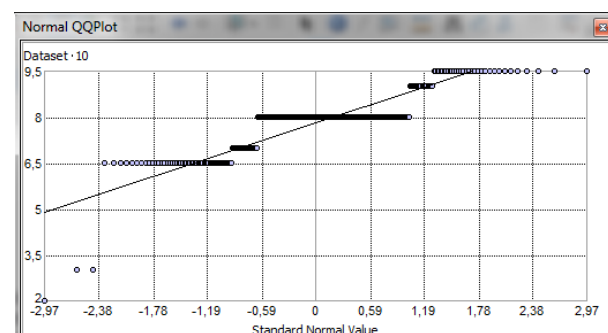


Figure 5. Normal QQ Plot

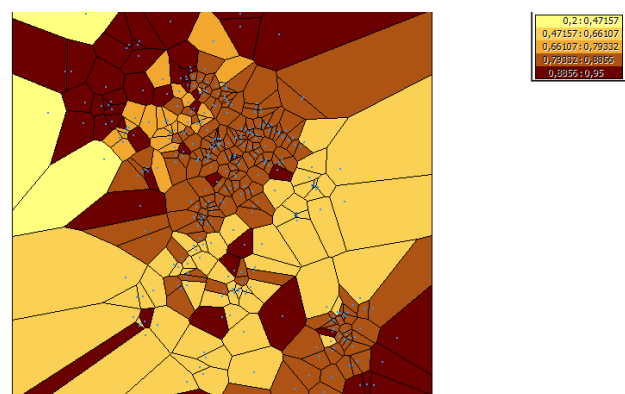


Figure 6. Voronoi Map

- 3) **Defining the model** implies choosing the surface's interpolation method in order to estimate the values where measurements have not been performed.

This step is achieved by means of Geostatistic Wizard which is based on the above mentioned techniques: deterministic methods that do not allow an assessment of the prediction error and do not require data evaluation and statistical methods which evaluate the prediction error and assume that the data comes from a stationary statistical process or that they have a normal distribution.

The surfaces resulting from this process may be:

- a) Prediction maps – estimate values at locations where measurements have not been carried out.
- b) Maps with standard errors – show the distribution of prediction errors for a certain surface. The error tends to be higher in areas where there is little sample data or not found at all.
- c) Quantile maps – show the 100% probability values of real values being lower than the quantile map values. Probability maps – show the chance that the actual value at a particular location to be greater than

a threshold value. The probability of exceeding the threshold is determined by the expected values, error distribution and specified threshold value.

- 4) **Diagnosis** represents statistical test such as Cross-validation for assessing the estimation quality.

Before realizing the final surface, it should be know how well the model predicts values in unknown locations. Cross validation and Validation facilitate the process of obtaining a conclusion regarding the model that offers best predictions. Cross validation uses all the data to estimate the self-correlation model. It then deletes the data from each location and predicts the values associated with the data, and then they are compared to the measured values.

Firstly, Validation deletes part of the data, and then uses the remaining data to develop the trend or autocorrect patterns that will be used for the prediction.

In both methods, the graphs and summary statistics used for the diagnosis are the same: prediction, predictive error (expected-measured), standard error (error/standard kriging error), normal Qqplot (standard error and normal standard distribution).

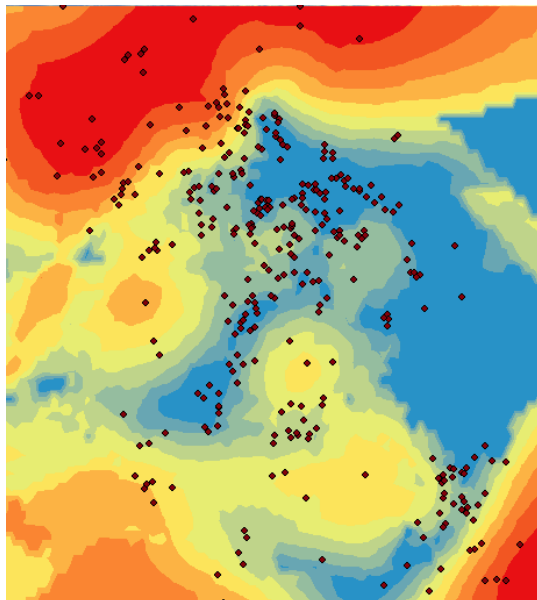


Figure 7. Surfaces resulted after Ordinary Kriging

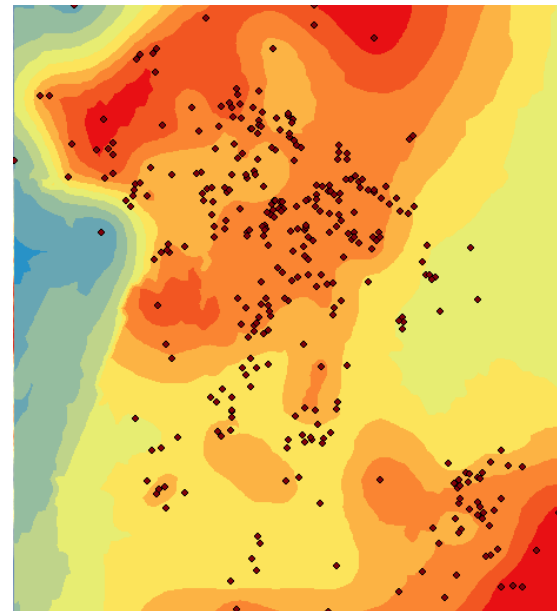


Figure 8. Surfaces resulted after Simple Kriging

- 5) **Comparing models** represents assessing the quality of an estimation set by comparison to other estimation sets.

Through this comparison conclusions regarding how accurate a surface is by comparison to others created can be drawn. The two comparative geostatistical layers can be created using two different models or using the same model, but with different parameters. In the first case, it is compared which method best fits the dataset, and in the second case, the effects of using different input parameters of a model for creating a surface are examined. Generally, the best model is the one whose average standard is close to 0, the smallest average quadratic error of prediction, the average standard error close to the mean quadratic error of prediction and the standard quadratic error close to 1.

Thus, surfaces will be compared and, by elimination, will remain the model that best lends itself to the factor used and will contain the least errors in the interpolation process.

4. CONCLUSIONS

Soil erosion modeling is quite complicated because landslides vary spatial and temporal depending on certain factors and interaction between them. It is necessary to determine both the estimation and the prediction for the unknown locations. In this study, the erodibility factor K was analysed. The assessment of erodibility represents an important step for understanding both the soil quality and its susceptibility to erosion and to predict soil erosion.

This study proves that merging geostatistics with the calculation of GIS maps represents an important and useful tool for the study of spatial changes in environmental sciences.

Method	Method type	Resulted surface	Operation period/ Modeling period	Advantages	Disadvantages	Hypothesis
IDW	Deterministic	Prediction	Rapid/ Rapid	Few decision parameters	Assessing the prediction error does not exist	No
Global Polynomial	Deterministic	Prediction	Rapid/ Rapid	Few decision parameters	Assessing the prediction error does not exist; boundary points may have a great influence	No
Local Polynomial	Deterministic	Prediction	Rapid Moderate/ Moderate	Lots of decision parameters	Assessing the prediction error does not exist; it may be a too automated method	No
Radial basis functions	Deterministic	Prediction	Rapid Moderate/ Moderate	Flexible and automated with several decision parameters	Assessing the prediction error does not exist; it may be a too automated method	No
Kriging	Statistical	Prediction; Standard prediction; Errors; Probability; Quantile	Rapid Moderate/ Slower	Very flexible; Allows spatial self-correlation assessment; Standard prediction errors can be obtained; Lots of decision parameters	Decisions regarding models, transformations, trends, parameters and vicinities are needed	Data come from stationary stochastic process and some of the methods need data to have a normal distribution
Cokriging	Statistical	Prediction; Standard prediction; Errors; Probability; Quantile	Moderate/ Slower	Very flexible; Information from different sets of data can be used; Allows cross-validation spatial self-correlation assessment; Lots of decision parameters	Decisions regarding models, transformations, trends, parameters and vicinities are needed	Data come from stationary stochastic process and some of the methods need data to have a normal distribution

Figure 9. Comparative study between Geostatistical Analyst techniques

REFERENCES

- Badea, A.C., 2013. *Comparative Study between Geostatistical Models Applied to Analyze the Phenomena Caused by Geomorphological Processes*, Recent Advances in Geodesy and Geomatics Engineering, ISBN: 978-960-474-335-3, pp. 164-171, available online at: <http://www.wseas.us/e-library/conferences/2013/Antalya/GENG/GENG-20.pdf>
- Badea, A.C., 2014. *Geostatistical Models Used in GIS for Geomorphological Processes*, International Journal of Geology, Vol. 8, ISSN: 1998-4499, online <http://www.wseas.us/e-library/conferences/2013/Antalya/GENG/GENG-20.pdf>
- Ferencz, Z., Hila, A. S., Dimen, L., Borsan, T., 2016. *Use of geoprocessing tools in hydrological and geomorphological analyses for the 'Frumoasa' lake area of the Harghita county, Romania*, J. Environmental Protection and Ecology, Vol. 17(4), pp. 1514-1522, ISSN: 1311-5065, WOS:000393355100028.
- Grigore, M., 1998. *Tipologia alunecărilor de teren din România*, Analele Universității București - Geografie, tom XLVII, Bucharest.
- Goovaerts, P., 2011. *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), pp. 371-380.
- Hârjoabă, I., 1968. *Alunecările de teren din colinele Tutovei. Comunicări de geografie*, Societatea Științifică Națională de Geografie, vol I, Bucharest.
- Ioniță, I., 1985. *Eroziunea solului și amenajarea terenurilor în Podișul Moldovei*, Vol. Cercetări geomorfologice pentru lucrări de îmbunătățiri funciare, Bucharest.
- Ortiz, B.V., Perry, C., Goovaerts, P., Vellidis, G., Sullivan, D., 2010. *Geostatistical modeling of the spatial variability and risk areas of Southern Root-Knot nematodes in relation to soil properties*, Geoderma, 156, pp. 243-252.
- Vilceanu, C.-B., Grecea, C., 2013. *Using Geographic Information System analysis in the management of flood risk areas*, Geoconference On Informatics, Geoinformatics and Remote Sensing - Conference Proceedings, Vol. I, ISSN 1314-2704, DOI:10.5593/sgem2013, WOS:000349067200134, pp. 1049-1056.
- Vilceanu, C.-B., Herban, S., Alionescu, A., Musat, C., 2015. *Processing of environmental data using Digital Terrain Models for the western part of Romania*, Informatics, Geoinformatics and Remote Sensing, Vol. II (SGEM 2015), pp. 1043-1050, ISBN:978-619-7105-35-3, ISSN: 1314-2704, WOS:000371600100130.
- Vilceanu, C.-B., Herban, S., Alionescu, A., 2016. *Using Open Source GIS for the Management of the Administrative Territorial Unit*, 15th National Technical-Scientific Conference on Modern Technologies for the 3rd Millennium, Romania, ISBN:978-88-7587-724-8, WOS:000378314000013, pp. 73-78

http://spatial-analyst.net/book/system/files/Hengl_2009_GEOSTATe2c1w.pdf

http://riscuriscatastrofe.reviste.ubbcluj.ro/Volume/IV_Nr_2_2005/PDF/Stanga_Minea.pdf

http://www.soilscience.uaic.ro/doc/SFFPTZ_No_4_p_247.pdf

<http://www.iwr.msu.edu/rusle/kfactor.htm>