Modelling the Uncertainty of Digital Elevation Models with Geostatistical Simulation and Application on Spatially Distribution of Soil Nitrogen

Mirko Borisov, PhD (Eng)[1]
Radoje Banković, MSc (Eng)
Siniša Drobnjak, MSc (Eng)

Digital representation of terrain has been an important focus in geography and surveying sciences for many decades. This article presents how the simulated digital elevation model (DEM) uncertainty affects common terrain analyses, such as slope and aspect calculations. In this article, the DEM of Avala test area was simulated on the basis of the Monte Carlo stochastic simulation. The purpose of this study is to generate multiple DEMs using error propagation techniques, then derive nitrogen storage for simulated DEM, and finally evaluate how the propagated uncertainty correlates to various topographic parameters, such as slope, aspect and wathershed delineation.

Key words: **geographical information system, topography, terrain analysis, stochastic analysis, Monte Carlo method.**

Introduction

THE most frequently used methods in the geographical information science (GIS) for representing terrain in the digital form are geometrical representations with regularly or irregularly distributed points, which may be supported by feature points and lines.

A geographical information system (GIS) is a computer system capable of capturing, storing, searching, processing and visualizing geographically-related information. It consists of technical equipment (hardware and software), geographical database, personal and organizational rules. The GIS analyzes and represents information based on spatial components, and their data management architectures permit an integrated view of extremely diverse types of information, such as digital cartography, remotely sensed imagery and survey data from a variety of sources.

Simulation models represent another category of tools, which may be used for planning or decision support. Several GIS packages offer a simulation capability for simple processes, but specialized models will be required to refine the level of analysis, or may represent consolidated tools that are pre-existing to the GIS in a given context. For this reason, the GIS and simulation models are often proposed to the decision-maker as an "integrated tool".

Simulations

A stochastic representation of uncertainty provides a distribution of potential answers, from which a "good" answer can be chosen, given some predefined criteria. The stochastic approach to modeling DEM uncertainty requires a number of equally probable realizations upon which selected statistics are performed. Uncertainty is quantified by evaluating the statistics associated with the range of outputs. Stochastic simulations provide a series of random equiprobable maps. Simulation does not ensure that a "real" map is generated from the process, but provides a distribution of results within which we can state the "true" map lies. Much research has focused on the use of stochastic simulation techniques to propagate error and quantify uncertainty in spatial data.

As a numerical computation method, the geostatistical simulation bypasses the analytical error propagation approach. This is particularly important in regional modeling tasks, in which the analysis result is a function of far reaching and unpredictable spatial interactions [1].
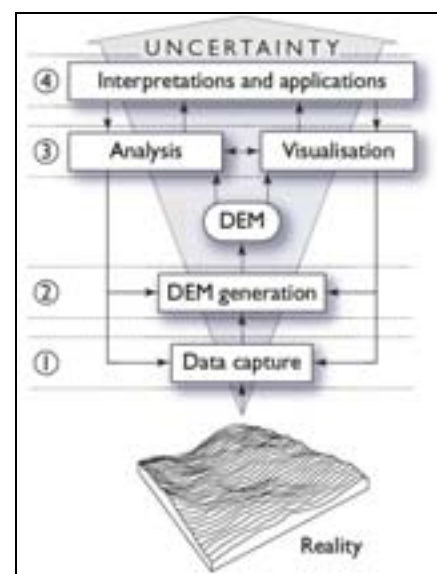

Figure 1. The main task of digital terrain modelling and the phases (1-4) in which uncertainty is introduced into the decision-making based on terrain analysis [2].

[1] Military Geographical Institute, Mije Kovačevića 5, 11000 Belgrade, SERBIA

There is some level of uncertainty in all simulation models. This uncertainty can come from a lack of complete understanding of the phenomena being modeled, error in the input data, or uncertainty in the model parameters. Some phenomena are not understood well enough, so they appear as random events such as locations for lightning strikes, but with a more sophisticated model and information, they are identified as not random.

The second source of uncertainty comes from the inherent error of the input data.

Another source of uncertainty comes from the inability to completely understand the phenomenon being modeled.

Vectorization of contour lines

For the simulation and generation of DEM from a digital raster image digitized from topographic maps, it is necessary to identify the contour lines and assign altitude values to them. At present, this pre-processing requires much time and numerous accumulated topographic maps to be digitized. It is important to develop an efficient method. However, it is difficult to develop a full automatic system including the editing process. If such a system were developed, it would be a high cost system which would need a powerful computer for the complex algorithm.

Therefore, we propose to combine the automatic processing with the interactive processing effectively. This approach enables that the automatic processing algorithm can be realized by utilizing the information obtained in the automatic processing. Fig.2 shows the digitalized contour lines with their attributes.
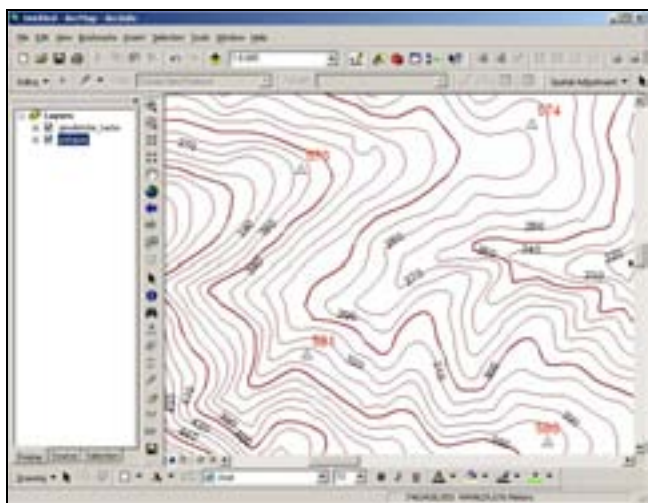

Figure 2. Contour lines with attributes

First, a contour image is produced by digitizing reproduction origins using the scanner. Second, all contour lines are automatically digitized. The contour attributes are stored in the attribute file. Finally, this file and the edited binary contour image will be the final output of the pre-processing after the modification of updating, recognizing principal lines, recognizing numerals represented altitude and assignment of altitude values to contour lines [9].

Principal contour lines and intermediate contour lines are drawn on a 1:25,000 scale contour map. The height interval of principal contour lines is 50 m and the intermediate contour lines are 10 m. Therefore, there are four intermediate contour lines between neighbouring principal contour lines. This contour map has some map symbols such as cliff, rocks, depression, in addition to the contour lines and numerals indicating the altitude. Supplementary contour lines have a height interval of 5 m and 2.5 m in the areas of low relief (Fig.3).
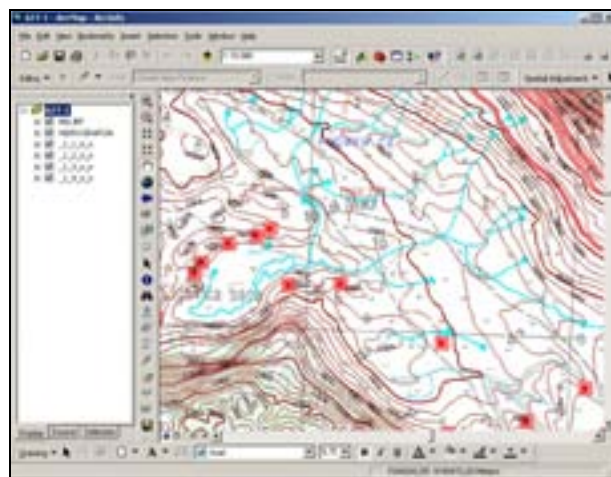

Figure 3. Relief, hydrography and their topology

Attributes of contours

After removing the isolated point noises and unnecessary branches, all contour lines are traced to collect the contour attributes. There are some attributes [9]:

1. Location - Coordinates of the start point and the end point of a contour line.
2. Length - Total number of pixels which compose a traced line.
3. Code – Classified contour lines on principal, intermediate and supplementary.
4. Attribute of numerals - This is a flag indicating the possibility of certain traced lines whether it is a numeral or not. In the line tracing process, lines on the contour image are not identified whether they are contours or numerals. But it can be said that broken or closed short lines have a possibility to be numerals.
5. Layer – Number of layers.
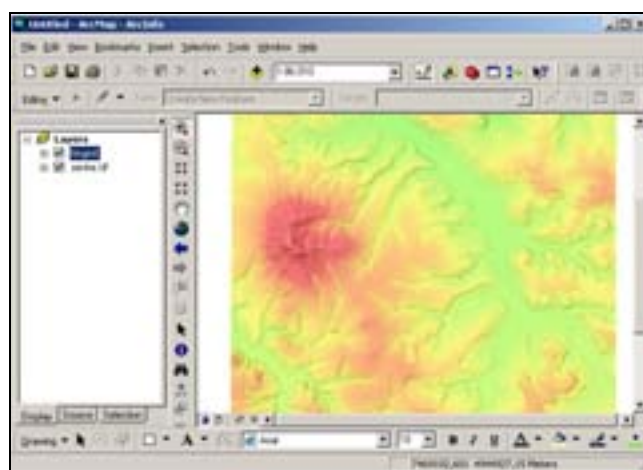
In Fig.4, he generated GRID model is shown.

Figure 4. Generated GRID model

Study area and used data

The "Avala hill" study area, located in the central part of the Republic of Serbia, has been mapped extensively over the years and several GIS layers are available at various scales. The study area corresponds to one sheet topographic map at a scale of 1:25 000 (TM25 430-3-1) which is

produced in the Military Geographical Institute (MGI). The bounding geographical coordinates of the study area are presented in Table 1. The main geomorphic features of the study area include hill summits and shoulders, eroded slopes of small valleys, valley bottoms, a large abandoned river channel, and river terraces (Fig.5). The elevation of the area ranges from 70 m to 500 m with an average of 195.05 m and a standard deviation of 56.7 m.

Table 1. Study area

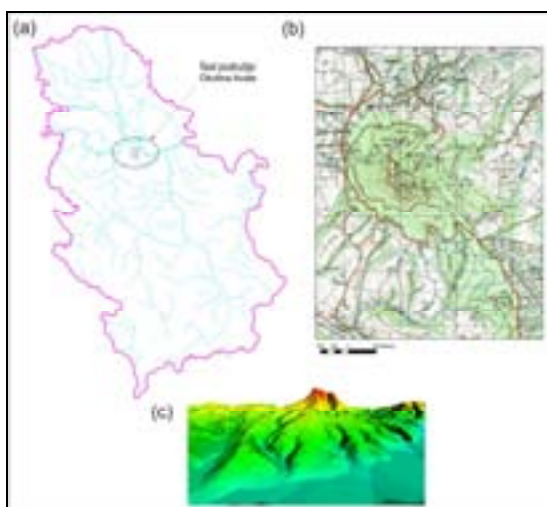|	Geographical Latitude	Geographical Longitude
1	44° 37' 30"	20° 30'
2	44° 37' 30"	20° 37' 30"
3	44° 45'	20° 37' 30"
4	44° 45'	20° 30'

Figure 5. Avala Test Area: (a) Location in the Republic of Serbia; (b) Topographic map 1:25 000; (c) 3D view of the test area

Contour lines were extracted from the topographic map at a scale of 1:25 000, with the contour interval of 10 m and supplementary 5 m and 2.5 m contours in the areas of low relief. In principle, the only inputs for this study are point and line vector data which represent relief. This data will be used to generate a DEM and then run spatially distributed simulation of nitrogen dynamics, as implemented in the SAGA GIS package [10].

Geostatistical simulations

Simulation is broadly defined as the process of replicating reality using a model. In geostatistics, simulation is the realization of a random function (surface) that has the same statistical features as the sample data used to generate it (measured by the mean, variance, and semivariogram). The Gaussian geostatistical simulation (GGS), more specifically, is suitable for continuous data and assumes that the data, or a transform of the data, has a normal (Gaussian) distribution. The main assumption behind the GGS is that the data is stationary — the mean, variance and spatial structure (semivariogram) do not change over the spatial domain of the data. Another key assumption of the GGS is that the random function being modeled is a multivariate Gaussian random function.

The Gaussian Geostatistical Simulations tool allows two types of simulations [6]:
- The conditional simulation honors the data values (unless the measurement error has been included in the kriging model). Some differences between the measured and the

simulated values for sample location may occur because simulation generates values at grid cell centers, which may not correspond exactly with the location of the sample points. The conditional simulation also replicates the mean, variance, and semivariogram of the data, on average (that is, averaged over many realizations). The simulated surfaces look like kriging prediction maps, but show more spatial variability.
- The unconditional simulation does not honor the data values, but does replicate the data mean, variance, and semivariogram (on average). The simulated surfaces show a spatial structure that is similar to a kriged map, but high- and low-value areas will not necessarily occur where high- and low-data values exist in the input data.
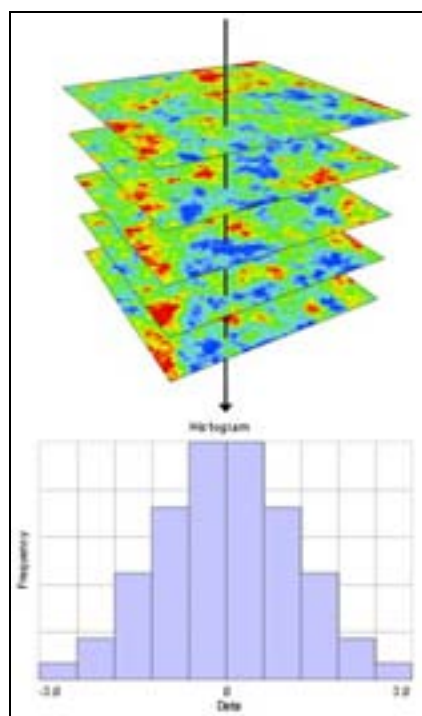

Figure 6. Variability of values simulated for a particular location [6].

Geostatistical simulations are used to generate simulations of a target variable using the same inputs as in the case of the spatial prediction system. An equiprobable realisation of an environmental variable can be generated by using the sampled values and their variogram model:

$$Z^{(SIM)}(s_0) = E\{Z|z(s_j), \gamma(h)\} \quad (1)$$

where $Z^{(SIM)}$ is the simulated value at the new location. The most common technique in geostatistics that can be used to generate equiprobable realisations is the Sequential Gaussian Simulation. It starts by defining a random path for visiting each node of the grid once. At the first node, simulation is used to determine the location-specific mean and the variance of the conditional cumulative distribution function [8].

A simulated value can then be drawn by using the inverse normal distribution:

$$z_i^{SIM} = \hat{z}_i + \hat{\sigma}_i \cdot \sqrt{-2 \cdot \ln(1-A)} \cdot \cos(2 \cdot \pi \cdot B) \quad (2)$$

where z_i^{SIM} is the simulated value of the target variable with the induced error, A and B are the independent random numbers within the $0 - 0.99\ldots$ range, $\hat{z}_i$ is the estimated value at i th location, and $\hat{\sigma}_i$ is the simulation error. The simulated value is then added to the original dataset and the procedure is repeated until all nodes have been visited.

Geostatistical simulations are used in many different fields to generate multiple realisations of the same feature or to generate realistic visualizations of a natural phenomenon.

Simulating DEM

Input data for simulating DEM is only point vector data which represents relief. We used this data to generate a DEM which we then used to run a spatially distributed simulation of nitrogen dynamics, toolkit implemented in the SAGA GIS package. In order to be able to run the Monte Carlo simulations, we need to first fit a variogram for elevations. We first fitted a variogram runing a script written in the R software for statistical computing and graphics to the sampled elevations dataset (elevations) which used to produce the DEM. We fitted a variogram using an exponential variogram with a nugget parameter $C_0=0$, a sill parameter of $C_1=560$ and the range parameter of 500 m. Elevations follow a normal distribution with a mean of 195.05 m and a standard deviation of 56.7 m (Fig.7) [4].

```
> elevations.or = variogram(Z ~ 1, elevations)
> elevations.ovgm = fit.variogram(elevations.or,
vgm(nugget = 0,
+ model = "Exp", range = sqrt(maparea)/3, sill =
var(elevations$Z)))
> elevations.ovgm
```

```
model psill range
1 Nug 0 0
2 Exp 560 500
```

```
> elevations.vgm.plt = plot(elevations.or,
elevations.ovgm,
+ plot.nu = T, pch = "+")
> print(elevations.vgm.plt)
```
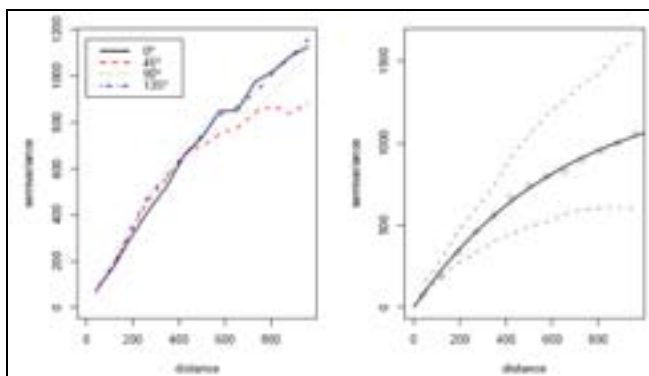

Figure 7. Omnidirectional variogram from the elevation data

Now we can simulate the *N* number of DEMs with the Monte Carlo simulations. The number of realisations (*N*) must be sufficiently large to obtain stable results, but exactly how large N should be depends on how accurate the results of the uncertainty analysis should be. For simulation we used a script written in the R software for statistical computing and graphics [11].

```
> N.sim = 100
> DEM.sim = krige (Z ~ 1, elevations, demgrid,
elevations.ovgm, + nmax = 100, nsim = N.sim)
```

```
drawing 100 GLS realizations of beta...
[using the conditional Gaussian simulation]
```

The accuracy of the Monte Carlo method is proportional to the square root of the number of runs N. Therefore, to double the accuracy one must quadruple the number of runs. This means that although many runs may be needed to reach stable and accurate results, any degree of precision can be reached by taking a large enough sample N. This implies that the Monte-Carlo method is computationally demanding, particularly when the GIS operation takes much computing time [4].
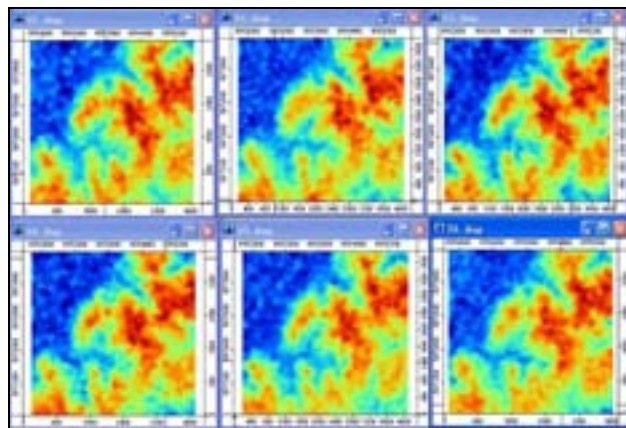

Figure 8. Different simulated DEM25

The realization of 100 simulations and averaging obtained results by different values of root mean squared errors (*RMSE)* we generate DEMs. Some of simulated DEMs can be seen in Fig. 8.
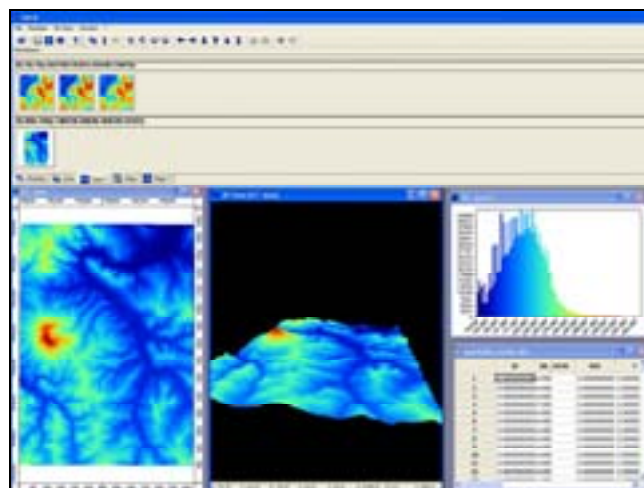

Figure 9. Software SAGA GIS

A window of the software SAGA GIS [10] can be seen in Fig.9.

Note that the simulated DEMs are quite noisy (we did not expect this considering that the variogram is very „smooth"). There are several explanations for this. Having a non-zero grid resolution implies that the correlation between adjacent grid cells is not equal to 1, so that grids may still appear to have noise [3, 7]. Still, this method creates too noisy DEMs, and the output values of nitrogen storage would be smoothed. To avoid this problem, we can simulate the same number of point maps and then re-interpolate them using thin-plate spline interpolation in SAGA.

The obtained results of the simulated DEMs were used for the production of maps of different topographic parameters like aspect, slope, local channel network, curvature, catchments area for propagation uncertainty correlates various topographic parameters with simulated DEMs (Fig.10).
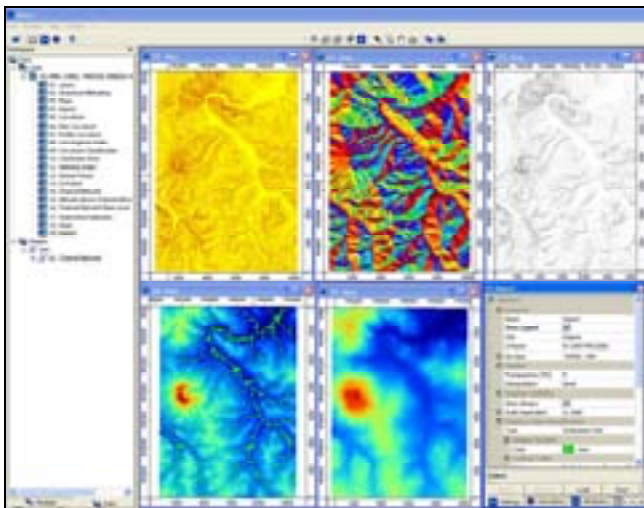
Figure 10. Different topographic parameters generated from the simulated DEM
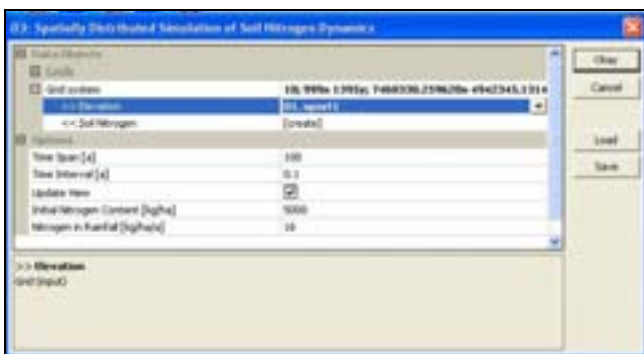

Figure 11. A window of the algorithm for the distributed simulation soil Nitrogen

Simulated and generated DEMs were used like input data to analyze the distribution soil nitrogen in kg/ha using the algorithm „*Spatially Distributed Simulation of Soil Nitrogen Dynamics*" which was implemented in the software SAGA GIS (Fig.11). For the initial Nitrogen content we used a value of 5000 kg/ha, and for the second parameter of the spatially distribution of Nitrogen soil, content of Nitrogen in rainfall, we used a value of 16 kg/ha/a. By realizing 100 simulations and averaging the obtained results by different values of root mean squared errors (*RMSE),* we generated the locations of distributed soil nitrogen for the Avala study area. We do not really need all maps but only the summary measures of the mapped values - mean, standard deviation, etc. Nitrogen distribution will be very skewed. Therefore, it might be more appropriate to log-transform the original values, or use the median value per pixel.

Fig.12 shows a 3D view of the study area with the locations of distributed soil nitrogen.

Figure 12. 3D view of locations of distributed soil nitrogen.

Now that we have estimated the propagated uncertainty of the mapping nitrogen deposition, we can try to understand how this model uncertainty corresponds to the geomorphology of terrain. We will derive only a few land surface parameters: aspect, slope, local channel network, etc. Obviously, the precision of mapping nitrogen is most critical for concave shapes. The coeficient of variation shows somewhat different images - the precision is most critical for locally dissected areas, while the valley bottoms and ridges present no problem to estimate nitrogen deposition.

Error propagation

Accuracy measures of DEM products quantify the accuracy or uncertainty associated with these products. Error propagation techniques such as the Monte Carlo simulation can be also used to analyse the propagation of errors.

In the Monte Carlo method, the derivation of variance equations is replaced with 'brute force' by repeating the terrain analysis a number of times on a DEM in which the realisations of the error model are added. The method offers flexibility at the cost of heavy computational load. In the terrain analysis, three factors that are critical from the performance perspective in the Monte Carlo simulation are:

1. the generation method of DEM error model realisation,

2. computation of the surface derivatives and

3. the simulation runs needed for converged analysis results.

This method was applied in DEM error propagation which is a numerical solution for resolving the convolution kernel from the a priori correlogram. The method was found to be functional for Gaussian spatial autocorrelation models, and the validity of the procedure was verified with a number of numerical experiments.

The convergence of the error propagation analysis results in the Monte Carlo simulation is still largely unknown, even though some application specific thresholds have been given. The analysis has often been done with less than 100 simulation runs, even though the risk of unreliable results is high. The problem is that the convergence analysis is always case-specific, as it is dependent on the terrain analysis task, error model parameters and topography.

Summary

The application of simulation in GIS analyses is significant for the production of novice attributes from existing characteristics entities or cells, which cannot be produced with the standard statistical theory of error propagation in environmental modeling.

The increased use of the GGS follows a trend in geostatistical practice that emphasizes the characterization of uncertainty for decision and risk analysis, rather than producing the best unbiased prediction for each unsampled location, which is more suited to showing global trends in the data. The Monte Carlo method offers a flexible and general solution for any type of error propagation analysis. This challenges the generalization that spatial autocorrelation of variables leads to the use of the Monte Carlo method.

Finally, the distribution of soil nitrogen is one of the biggest indicators for air and soil pollution, and their analyses by geoinformation tools integrate different science areas in one.

References

[1] BURROUGH,P.A., Mcdonnell,R.A.: *Principles of Geographical Information Systems, second edition*, Oxford University Press, 1998.

[2] OKSANEN,J.: *Digital elevation model error in terrain analysis*, Academic dissertation in Geography, Faculty of Science, University of Helsinki, 2006.

[3] HEUVELINK,G.B.: *Error Propagation in Environmental Modelling with GIS*, 1998.

[4] HENGL,T.: *Spatio-temporal data analysis (R+ILWIS/SAGA+Google Earth)*, PhD Course, Građevinski fakultet, Beograd 2008.

[5] HENGL,T.: *A Practical quide to Geostatistical mapping*, 2009.

[6] ESRI, *Geostatistical Simulation Concepts*, ArcGis software help.

[7] HENGL,T., GRUBER,S., SHRESTHA,D.P.: *Digital Terrain Analysis in Ilwis*, 2003.

[8] HENGL,T., BAJAT,B., REUTER,H.I., BLAGOJEVIĆ,D.: 2007. *Geostatistical modelling of topography using auxiliary maps*, Computers and Geosciences, 34: pp.1886-1899.

[9] BORISOV,M., BANKOVIĆ,P., TATOMIROVIĆ,C.: *Digitalni prostorni model Srbije i Crne Gore – Članak*, Zbornik radova Sa 4. Savetovanja "Bezbednost Putnog Saobraćaja U Sistemu Odbrane", Beograd, 2006, Str. 35-45.

[10] SAGA GIS - System for Automated Geoscientific Analyses (http://www.saqa-gis.org)

[11] R software for statistical computing and graphics: (http://www.r-project.org).

Modelovanje nesigurnosti digitalnih modela visine pomoću geostatističkih simulacija i njihova primena u prostornom modelovanju azota u zemljištu

Predstavljenje terena u digitalnom obliku već decenijama se nalazi u žiži interesovanja geografskih i geodetskih nauka. Ovaj članak prikazuje kako nepouzdanost simuliranih digitalnih modela visina (DMV) utiče na pojedine analize terena, kao što su računanje nagiba i ekspozicije terena. U ovom radu, simuliran je DMV test područja Avale baziran na Monte Karlo stohastičkoj simulaciji. Generisani DMV korišćeni su radi simulacije deponovanja soli nitrogena za svaku realizaciju DMV pojedinačno i na kraju pokazana je korelacija između dobijenih rezultata i različitih topografskih paramatara, kao što su nagib i ekspozicija terena i mreža rečnih slivova.

Ključne reći: geografski informacioni sistem, topografija, analiza terena, stohastička analiza, metoda Monte Karlo.

Моделирование изменчивости цифровых моделей высот при помощи геостатических симуляций

Изображение территории цифровым способом уже десятками лет находится в фокусе интересования географических и геодезических наук. Настоящая статья показывает каким способом изменчивость симулирующих цифровых моделей высот (ЦВМ) влияет на отдельные анализы территории, а в том числе и на расчёт крена (наклона) и показ (экспозицию) территории. В настоящей работе проведена симуляция ЦВМ испытания территории горы Авала, основана на стохастической симуляции Монте Карло. Генерирующие ЦВМ использованы ради симуляции хранения в складах соли азота для каждой реализации ЦВМ отдельно и в итоге показано соотношение между полученными результатами и различными топографическими параметрами, а в том числе и крена (наклона) и показа (экспозиции) территории и сети водоразделов.

Ключевые слова: географическая информационная система, топография, анализ местности, стохастический анализ, метод Монте Карло.

Modélisation de l'incertitude des modèles digitaux d'altitude par les simulations géostatiques et leur emploi dans la modélisation d'espace de l'azote dans le sol

La représentation du terrain en forme digitale se trouve au centre d'intérêt des sciences géographiques et géodésiques depuis des décennies. Cet article démontre de quelle manière l'incertitude des modèles digitaux simulés d'altitude influe sur les analyses du terrain, tels que le calcul de l'inclinaison et l'exposition du terrain. Dans ce travail on a simulé le test des modèles digitaux d'altitude pour la région de Avala, basé sur la simulation stochastique Monté Carlo. Les tests des modèles digitaux d'altitude générés ont été employés pour simuler le dépôt du sel nitrogène pour chaque réalisation du modèle digital d'altitude et à la fin on a démontré la corrélation entre les résultats obtenus et les paramètres topographiques variés, tels que l'inclinaison et l'exposition du terrain ainsi que le réseau des basins fluviaux.

Mots clés: système géographique d'information, topographie, analyse du terrain, analyse stochastique, méthode Monté Carlo.