NUMERICAL METHOD FOR UNCERTAINTY ASSESSMENT OF THE TOPOGRAPHIC PARAMETERS DERIVED FROM DEMs

B. Bajat[a], S.Rath[b], D. Blagojevic[a]

[a] University of Belgrade, Faculty of Civil Engineering, Department of Geodesy, Serbia & Montenegro.
Bajat@grf.bg.ac.yu

[b] Hamburg University of Technology, River & Coastal Engineering, D-21073 Hamburg.

ABSTRACT

Topography has a great influence on many characteristics which are connected to the distinct area on Earth such as hydrology, microclimate characteristics, vegetation cover properties, geological structures, mineral deposits, plant characteristic etc. Digital terrain models represent segments of spatial data bases related to presentation of terrain features and landforms. Square grid elevation models (DEMs) have emerged as the most widely used structure during the past decade because of their simplicity and simple computer implementation. They have become an important segment of Topographic Information Systems (TIS), storing natural and artificial landscape in forms of digital models. This kind of data structure is especially suitable for morphometric terrain evaluation and analysis which is very important in modeling Earth surface and atmospheric processes or environmental modeling applications. Treating topographic parameters like slope, aspects and terrain curvature as the first and second derivatives or terrain surface, they can easily be calculated in such data structures.

DEMs are used as proxies for the actual terrain surface and determined as a gathering of the measured terrain data and interpolation techniques. According to this statement, the quality of DEMs are characterized by the uncertainty being referred to data model-base uncertainty and data based uncertainty. According to the regular grid structure of points, uncertainties in DEMs are mostly treated through the height accuracy of the points. Well known methods to assess the uncertainty of derived parameters, such as Monte Carlo stochastic simulation techniques, are already used for these purposes. This method has both good and inferior characteristics. Production of numerous realizations of DEMs for particular terrain and subsequent statistical analysis may be time consuming and limited with available computer memory. Simple numerical solution of a variance propagation technique is presented in this paper as a possible alternative for uncertainty assessment of derived topographic parameters. Both uncorrelated and correlated uncertainty scenarios are taken into account. Obtained results are confirmed with Monte Carlo approach with high level of agreement.

Keywords: DEM, uncertainty, topographic parameters.

1 INTRODUCTION

Gathering, maintenance and visualization of relief data are mostly used GIS (Geographical Information System) functions. This segment of GIS applications is well known as Digital Terrain Model (DTM). During the last decade this has become convenient tool for representing terrain surface, not only in GIS applications, but in computer assisted software for projecting, as well. DTM bases are organized in different manners, like Triangular Irregular Networks (TIN) or in regular lattice (GRID). Term DEM is related to terrain height databases with regular grid structure or altitude matrix, and the term Digital Terrain Models (DTM) is mostly related to TIN structures.

DEMs approach for data base structures is suitable for National Heights Data Bases which cover the area of the whole country area as a part of nation-wide spatial data bases available from the national mapping agencies. This kind of data bases are very useful in various application areas in GIS, like ecological studies, 3D urban mapping, environmental monitoring, landscape planning, geological analysis, civil engineering, floodplain analysis, risk maps etc. They are also very suitable in complex analysis, especially in superimposing other kind of data, like satellite images or remotely sensing information. The new and improved methods of an analysis allow users to process even more complex application tasks with combinations of geometrical, topological and thematic aspects using hybrid data, i.e. raster data, as well as, vector data [1]

DTM concept had started as a stand alone application intended for different engineering tasks. Nowadays, it is a part of the spatial data bases which store the natural and artificial landscape in the form of digital models. DEM are compulsory component of one categories of spatial information systems known as Topographic Information System (TIS) [2].

Geomorphometry, defined as science 'which treats the geometry of the landscape' with quantitative parameters, attempting to describe the form of land surface, represents one of the disciplines where DEMs products are completely implemented.

Slope, aspect, profile (vertical) and contour (horizontal) curvature (convexity) are basic attributes of land surface at the vertex or its close neighborhood, and they form a coherent system for terrain analysis and description, easily calculated with such arranged databases of terrain heights[3].

DEM data is subject to errors such as any other spatial data source. DEM users must keep in mind that results obtained by analysis and processing of such data depend on the quality of this data. Quality assessment of obtained results indicates their reliability and suitability for particular applications.

Since a dataset is produced rather for various than just for one specific application, the quality of dataset can be assessed only by knowing the data quality elements, as well as by the data quality overview element [4].

2 DEM QUALITY ASSESSMENT

Terrain heights in DEM data bases are considered as point features in which heights components are stored as attributes. This concept is well known as 2.5 GIS data model. Attribute accuracy is the main element of quality of such spatial databases. Accuracy can be defined as closeness of agreement between the test result and the accepted reference or 'true' value, where reference values are data sources of higher accuracy. However, a true value or location is a luxury not often available in spatial data.

In the case of spatial data it refers to data of larger scale and resolution or usage of instrument of higher accuracy or more recent measurements. All those imply that definition of accuracy is relative and for that reason term accuracy is substituted with uncertainty.

Many words with similar connotation are used to express term *uncertainty,* like *reliability*, *confidence*, *accuracy*, *error* etc. This term is especially suitable for reporting the assertiveness of variables where *true* values are unknown. Under these circumstances it is not feasible to calculate exactly the error of the appraisal. A useful step in assessing the uncertainty is to consider the factors by which the error is influenced [5].

Two DEM-related uncertainty categories are commonly recognized. The primary is *data model-based* uncertainty, resulting from differences between the form of data model and actual elevation surface. The second is *data-based* uncertainty, referring to differences between the elevation of location specified in the data set and actual elevation at that location [6].

DEM altitudes are given as interval data, and root mean square error (RMSE) or standard deviation could be used as an uncertainty metrics. The RMSE calculated from residuals between models heights and 'ground truth' points is commonly used measure of the accuracy for the DEMs products [7]. Another powerful measure is the standard deviation of residuals, obtained from discrepancies between model heights and terrain heights at the particular locations.

$$\sigma_z = \sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1}} \quad , \quad \bar{d} = \frac{\sum d_i}{n} \qquad (1)$$

where:

d_i- residual between height in DEM and 'true' height $d_i = z_{DEM} - z_t$,

n- number of control points.

Accuracy evaluations often uses field survey data based on the Global Positioning System (GPS) because of high accuracy.

Performing control measurements is a standard procedure for evaluating the quality of DEMs products. It can also be performed by DEMs users if they are not acquainted with lineage of data. Locations of the control points have to be randomly chosen [8].

Developing adequate GIS methods and techniques which can take uncertainty of spatial data into account during their processing are crucial for errors handling in subsequent analysis [9]. The problem of error propagation through GIS operations can be undertaken by using statistical theory [10].

This analytical analysis can be achieved when algebraic relationships between input data and output results are simple, e.g. slope calculations in geomorphometry analysis (equation).

$$\sigma_{slope} = \sqrt{2\sigma_z^2 \frac{(1-r)}{d^2}} \tag{2}$$

where:

σ_{slope}- standard deviation of a slope,

d- grid spacing,

r- correlation between elevation errors.

Due to sufficiently complex analyses being performed in many GIS applications this analytical approach is usually impossible. Instead, uncertainty must be propagated by simulations and analyzed from the results, known as Monte Carlo approach [11]. This is more general approach where error is modeled stochastically through numerous simulations of probable realizations of DEM by producing a distribution of results which may be assessed statistically.

Two issues are important for this approach, the magnitude of error of the input data as well as the spatial structure of error. The spatial structure of error may be represented by semivariograms or covariograms.

Producing numerous realizations of DEMs for a particular terrain and subsequent statistical analysis may be time consuming and limited by the available computer memory.

3 NUMERICAL SOLUTION FOR ERROR PROPAGATION

An alternative of analytical approach of DEMs uncertainty assessment which avoids complicated formula derivations is offered in this section.

Value of primary geomorphological parameters for particular grid cell is calculated as function of heights of surrounding cells:

$$y=f(z_1,z_2,...z_n) \tag{3}$$

where:

y- value of a particular geomorphological parameter.

z_i- terrain height in surrounding grid cell.

n- number of surrounding cells, 4 for rook's case and 9 for queen's case (fig. 1.)

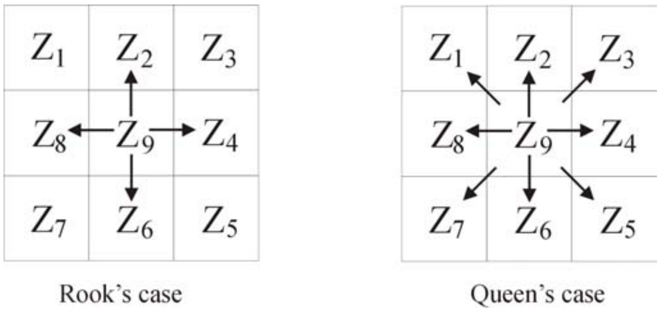

Figure 1. Chess rules for rook's and queen's motion

Since formula (3) is often nonlinear, the linearization is given by:

$$y = y_0 + \left(\frac{\partial f}{\partial z_1}\right)_0 \Delta z_1 + \left(\frac{\partial f}{\partial z_2}\right)_0 \Delta z_2 + \ldots + \left(\frac{\partial f}{\partial z_n}\right)_0 \Delta z_n \tag{4}$$

Partial derivatives:

$$a_i = \left(\frac{\partial f}{\partial z_i}\right)_0 \tag{5}$$

are calculated with aproximate values of the surrounding cells. Numerical procedure for calculation gives us the following equation:

$$a_i = \frac{f\left(z_1, z_2, \ldots z_i + \Delta z, \ldots z_n\right) - f\left(z_1, z_2, \ldots z_n\right)}{\Delta z} \tag{6}$$

where Δz is a small arbitrary increment.

Equation (4) can be present as:

$$y = y_0 + a_1 \Delta z_1 + a_2 \Delta z_2 + \ldots + a_n \Delta z_n \tag{7}$$

or in matrix form:

$$y = A^T \cdot \Delta Z + y_0, \quad A^T = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}, \quad \Delta Z^T = \begin{bmatrix} \Delta z_1 & \Delta z_2 & \ldots & \Delta z_n \end{bmatrix} \tag{8}$$

The variance of ***y*** can be determined by rules of error propagation law:

$$\sigma_y^2 = A^T \cdot C_z \cdot A \quad \Rightarrow \quad \sigma_y = \sqrt{A^T \cdot C_z \cdot A} \tag{9}$$

where σ_y^2 is variance of particular geomorphological parameter, $\boldsymbol{C_z}$ is covariance matrix of error cells heights:

$$C_z = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{n1} & \sigma_{22}^2 & \cdots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}^2 \end{bmatrix} \tag{10}$$

where σ_{ii}^2 is variance and σ_{ij} is covariance of heights errors. The matrix $\boldsymbol{C_z}$ is also symmetric, which means that, $\sigma_{ij}=\sigma_{ji}$.

When heights errors are uncorelated, matrix $\boldsymbol{C_z}$ has the following structure:

$$C_z = \begin{bmatrix} \sigma_{11}^2 & 0 & \ldots & 0 \\ 0 & \sigma_{22}^2 & \ldots & 0 \\ \vdots & & & \\ 0 & 0 & \ldots & \sigma_{nn}^2 \end{bmatrix} \tag{11}$$

In the case that variances of heights errors of surrounding cells are equal with σ_z^2 value, then:

$$C_z = \sigma_z^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma_z^2 \cdot I \tag{12}$$

where ***I*** represents the unit matrix. For given conditiones in equations (11) and (12) the variance of geomorphological parameters will be:

$$\begin{aligned} \sigma_y^2 &= a_1^2 \cdot \sigma_{11}^2 + a_2^2 \cdot \sigma_{22}^2 + \dots + a_n^2 \cdot \sigma_{nn}^2 \\ \sigma_y^2 &= \left(a_1^2 + a_2^2 + \dots + a_n^2\right) \cdot \sigma_z^2 \end{aligned} \tag{13}$$

It is necessary to determine non diagonal terms of matrix $\boldsymbol{C_z}$ for corellated heights errors. Under the assumption that corellation between heights errors becomes weaker with longer distance, one of possible models for correlation is exponential function:

$$\sigma_{ij} = \sigma_z^2 \cdot e^{-\frac{d_{ij}}{D}} \tag{14}$$

where

$\boldsymbol{d_{ij}}$- distance between ***i*** and ***j*** cells

D – correlation distance.

Figure 2. Flow chart for numerical solution for uncertainty assessment of topographical parameters

Presented procedure is suggested as practical solution with regard to time and computing resources for uncertainty assesment of calculated topographical parameters. Results obtained with this approach have been verifyied by comparing them with the results obtained from Monte Carlo stochastic simulations.

4 FORMULAS FOR CALCULATION OF TOPOGRAPHICAL PARAMETERS

It is straightforward to obtain primary geomorphological parameters in every grid cell according to the above mentioned scheme by calculating partial derivatives with method of finite differences:

$$\begin{aligned} z_x &= \frac{(z_3 + 2z_4 + z_5) - (z_1 + 2z_8 + z_7)}{8 \times d} \\ z_y &= \frac{(z_1 + 2z_2 + z_3) - (z_7 + 2z_6 + z_5)}{8 \times d} \\ z_{xx} &= \frac{\partial^2 z}{\partial x^2} \approx \frac{z_4 - 2z_9 + z_8}{d^2} \\ z_{yy} &= \frac{\partial^2 z}{\partial y^2} \approx \frac{z_2 - 2z_9 + z_6}{d^2} \\ z_{xy} &= \frac{\partial^2 z}{\partial x \partial y} \approx \frac{-z_1 + z_3 + z_7 - z_5}{4d^2} \end{aligned} \tag{15}$$

Slope S shows the rate of change of height of the terrain surface with distance:

$$S = \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} = \sqrt{z_x^2 + z_y^2} \tag{16}$$

Slope controls runoff and soil loss, thickness of soil horizons, some plant characteristics.

Aspect A is the orientation of the line of steepest descent, which is measured in clockwise direction starting from north:

$$A = 180^o - \arctan\left(\frac{z_y}{z_x}\right) + 90^o\left(\frac{z_x}{|z_x|}\right) \tag{17}$$

In association with slope, aspect controls isolation and evapotranspiration**,** it influences thickness of soil horizons and some plant properties, also.

Horizontal curvature K_c is the curvature in the horizontal plane of contour line. Vertical curvature K_p is the curvature in the vertical plane of a flow line. They represent the rate of change of a first deriviate such as slope or aspect in particular directions. Tangential curvature K_t is curvature in an inclined plane perpendicular to both the direction of flow and the surface. Tangential curvature represents horizontal curavture multiplied by the sine of the slope angle, and it is more appropriate than horizontal curvature for studing flow diveregence and convergence when terrain slope is too small[12]:

$$\begin{aligned} K_p &= \frac{z_{xx}z_x^2 + 2z_{xy}z_xz_y + z_{yy}z_y^2}{P\sqrt{Q^3}} & \qquad \mathrm{P} &= z_x^2 + z_y^2 \\ K_c &= \frac{z_{xx}z_y^2 - 2z_{xy}z_xz_y + z_{yy}z_x^2}{\sqrt{P^3}} & \qquad \mathrm{Q} &= \mathrm{P} + 1 \\ K_t &= \frac{z_{xx}z_y^2 - 2z_{xy}z_xz_y + z_{yy}z_x^2}{P\sqrt{Q}} \end{aligned} \tag{18}$$

Horizontal and vertical curvatures are the determining local factors of the dynamics of overland and intrasoil water, they influence soil moisture, pH, thickness of soil horizons, organic matter, plant cover distribution[13].

5 CASE STUDY

DEM was produced by digitizing contour layers from two adjacent sheets of the topographic maps of scale 1:5000, with contour interval of 5m, with total area of 13.5 square kilometers for the research purpose. Test location is the resort area Zlatibor in south western Serbia with minimum height of 850m and maximum height of 1174m. This area is hilly plateau, with the exception of the west and north-west part with greater slopes of terrain.

Digitized polylines with height attributes were broken into vertices and such big amount of obtained points was reduced by using the Douglas-Peucker algorithm for polyline simplification to 51847 points. A DEM was produced in two steps: An initial TIN was produced using a Delaunay triangulation, being subsequently converted into regular grid with 10m resolution (350 rows by 400 columns).

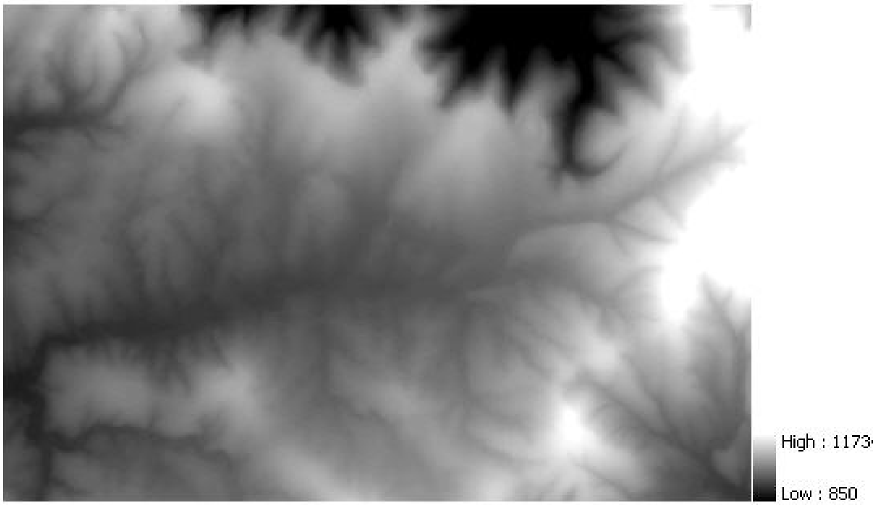

Figure 3. DEM of test area "Zlatibor"

A GPS survey of control points has been carried out, using fast static and real time kinematics (RTK). A seven parameters similarity transformation was used to bring GPS coordinates into official reference system. Ellipsoidal GPS heights were transformed into orthometric system by polynomial model of second degree. The achieved accuracy of heights is estimated to be better than 3 cm.

Calculated σ_z for data set was approximately 1.25m, and obtained result is in accordance with the expected accuracy of the cartographic source data used for DEM production [14]. Estimated correlation distance of errors was ***D***=110m.

The estimated uncertainties (standard deviation) of topographic parameters obtained by stochastic simulations σ_{sim} are compared with values σ_{an} of standard deviations obtained analytically by proposed algorithm. Both correlated and uncorrelated heights errors were taken into consideration.

Monte Carlo simulations were carried out by adding generated error fields to initial DEM [15]. Both uncorrelated (worst case scenario) and correlated error fields have been generated. For each scenario 25 simulations were made, as it was necessary to yield a statistically useful result at the 0.05 significance level [16].

The compliance of standard deviations was checked for each grid cell by testing of hypothesis:

H_0: estimated values of standard deviations are similar $\sigma_{an} = \sigma_{sim}$

H_1: estimated values of standard deviations are not similar $\sigma_{an} \neq \sigma_{sim}$

Analytically obtained value σ_{an} has $f = \infty$ degrees of freedom while simulated estimate has $f = n - 1 = 24$ degrees of freedom, with number of simulations n=25. Depending on test statistics, following critical values (for 99% probability) were used:

$$F \sim F_{0.99}(\infty, 24)\big|H_0 = 2.21,\ F \sim F_{0.99}(24, \infty)\big|H_0 = 1.79$$

Table 1. Acceptance of Fisher's test of similarity of simulated and analytically obtained uncertainty estimates for each topographic parameter

	F_σ_S	F_σ_A	F_σ_{kc}	F_σ_{kp}	F_σ_{kt}
Uncorrelated error fields	89.7%	61.9%	69.8%	97.7%	96.7%
Correlated error fields	84.8%	71.8%	73.0%	90.9%	85.9%

Legend: F_σ_i- Fisher's test for $\boldsymbol{i}^{th}$ topografical parameter, σ_S- standard deviation of a slope, σ_A- standard deviation of an aspect, σ_{kc}- standard deviation of contour curvature, σ_{kp}- standard deviation of profile curvature, σ_{kt}- standard deviation of tangenential curvature.

The acceptance of Fisher's test (acceptance of hypothesis H_0) is given in percentages which show the rate between the number of cell where hypothesis H_0 is accepted in regard to whole number of grid cells.

The agreement of these estimates varies from one parameter to the other, with a view that this agreement is somewhat lower for aspects, but which is quite expected considering specifies of those aspects closer to values of 0^o or 360^o respectively.

7 REMARKS

This analysis can not give an answer which one of both discusssed methods for uncertainty estimation gives more realible results. Nevertheless it confirms that simple numerical solutions for analitical method gives similar results as stochastic simulations. This procedure could be suggested as simple and economic regarding to software and time resources for an unceraintу analyses in geomorphometry.

REFERENCES

1 Glemser, M., Klein, U., Fritsch, D., Strunz G., 2000: Complex Analysis Methods in Hybrid GIS Using Uncertain Data. In: GIS - Geo-Informationssysteme, No. 2, pp. 34-40, Wichmann Verlag, Heidelberg.

2 Kraus K.,1995: From digital elevation model to topographic information system. In: Week, D. Fritsch and D. Hubbie (ed.): 45th.Photogram metric week, pp. 277-285. Stuttgart

3 Evans I.S, 1998: What do Terrain Statistic Rally Mean In: Lane S.N., Richards K.S. (ed): Landform Monitoring Modelling and Analysis, pp. 119-138. John Wiley &Sons, London

4 ISO/TC211, 2001: ISO 19113 Geographic information –Quality principles

5 Isaaks E., Srivastava R., 1989: An Introduction to Applied Geostatistics. Oxford University Press, pp. 561. New York, NY

6 Shortridge A., 2001: Characterizing Uncertainty in Digital Elevation Models., In: Hunsaker C.T., Goodchild M.F., Friedl M.A., Case T.J. (ed.): Spatial Uncertainty in Ecology,pp.238-257 Springer-Verlag New York.

7 Caruso V.M., 1987: Standards for digital elevation models. Proceesings of the ACSM/ASPRS Annual Convention, pp.159-166 Baltimore.

8 Li Z. 1991: Effects of Check Points on the Reliability of DTM Accuracy Estimates Obtained from Experimental Tests. *PERS*, 57, (10),pp. 1333-1340.

9 Openshaw S., 1994: Learning to live with errors in spatial databases, In: Goodchild M.F, Gopal S. (ed.): The accuracy of spatial databases, pp. 263-276. Taylor&Francis, London

10 Burrough P.A., 1986: Principles of Geographical Information Systems for Land Resources Assessment, Oxford University Press, New York, 194 pp.

11 Goodchild M.F., 1995: Attribute accuracy. In: Guptill S.C.& Morrison J.L (ed): Elements of spatial data quality, pp. 59-79. Elsevier.

12 Gallant J.C., Wilson J.P., 2000: Primary Topographic Attributes. In: Gallant J.C., Wilson J.P. (ed): *Terrain Analysis-Principles and Applications*,pp. 51-84. John Wiley&Sons, New York.

13 Florinsky I.V., Seminar: 'Digital Terrain Models: Theory and Practice' http://members.fortunecity.com/flor/seminar.htm

14 Merchant, D.C.,1987: Spatial accuracy specification for large scale topographic maps, *PERS* 53, pp. 958-961

15 Fisher P., 1991: First Experiments in Viewshed Uncertainty: The Accuracy of the Viewshed Area. PERS 57 (10), pp.1321-1327.

16 Hope A.C.A., 1968: A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society B*, 30, pp. 582-598