Geostatistical modelling of topography using auxiliary maps

T. Hengl[1], B. Bajat[2], H.I. Reuter[1], D. Blagojević[2]

[1]European Commission, Directorate General JRC, Insitute for Environment and Sustainability,
TP 280 JRC, Via E. Fermi 1, I-21020 Ispra (VA), Italy
Telephone: +39-0332-785535
Fax: +39-0332-786394
Email: tomislav.hengl@jrc.it; hannes.reuter@jrc.it

[2]Faculty of Civil Engineering, Bulevar kralja Aleksandra 73, 11 000 Belgrade, Serbia
Telephone: +381-11-3218579
Fax: +381-11-3370223
Email: bajat@grf.bg.ac.yu; bdragan@grf.bg.ac.yu

1. Introduction

A Digital Elevation Model (DEM) is a digital representation of the topography – the major input to quantitative analysis of topography, also known as geomorphometry. Typically, a DEM is a raster map (an image or an elevation array) that, like many other spatial features, can be efficiently modelled using geostatistics. The geostatistical concepts were introduced in geomorphometry by Fisher et al. (1998), then further elaborated by Kyriakidis et al. (1999), Holmes et al. (2000) and Oksanen (2006b). A review of the methodological developments and future trends of geostatistical modelling of topography can be especially followed in the Ph.D. thesis of Oksanen (2006a).

An important focus of using geostatistics to model topography is assessment of the errors in DEMs and observation of the impacts the DEM errors have on the results of spatial modelling. This is the concept of error propagation that commonly works as follows: a DEM is usually simulated to produce multiple equiprobable realisations of a DEM of an area; the spatial model is applied m times and then statistically analysed for mean values, standard deviations and similar; the results of analysis can be used to quantify DEM accuracy and observe impacts of uncertain information in various parts of the study area (Heuvelink 1998; Oksanen et al. 2005; Raaflaub and Collins 2006).

So far, DEMs have been modelled solely by using the field-sampled elevations. In most studies, no auxiliary (also known as additional or secondary) information on topography is employed directly in the geostatistical modelling. Unlike the standard geomorphometry where, for example, maps of streams are often used to produce hydrologically-correct DEMs, the geostatistical approach to modelling of topography has often been limited to analysis of points.

Our interest in this paper was to develop and test a more sophisticated methodology to model topography using geostatistics by including the auxiliary information directly into the geostatistical analysis. By auxiliary information, we consider all GIS layers that can explain spatial distribution of elevations and associated errors and which are produced independently from the DEM: hydrological networks, terrain complexity measures, land cover indices and similar. Our assumption is that, by including such information, we will be able to produce more accurate realisations of a DEM and, consequently, enhance the use of geostatistics in geomorphometry.

2. Methods and materials

2.1 Case study

We used a case study area "Zlatibor" located in the South-west part of Serbia (centred at 43º43'44.6"N and 19º42'37.8"E). The area is mainly hilly plateau, with exception of north-east part where the slopes are much steeper (Fig.1a). Elevations range from 850 m to a maximum of 1174 m; the total size of the area is 13.5 square kilometers. There are three major inputs into this analysis: (a) a set of 2051 height measurements used for generation of DEMs, (b) a set of 1020 very precise spot heights used for error assessment, (c) the original topo-map DEM at 30 m resolution and (d) 90 m SRTM DEM. The original topo-map DEM was produced by digitizing contour layers from two adjacent sheets of the 1:5000 topographic maps with contour interval of 5~m. Two sheets were scanned by ANATech Evolution scanner with 400 DPI resolution, then georeferenced to the Gauss-Kruger coordinate system (7th zone) and converted to a point map using a semi-automated digitalisation of contour lines.

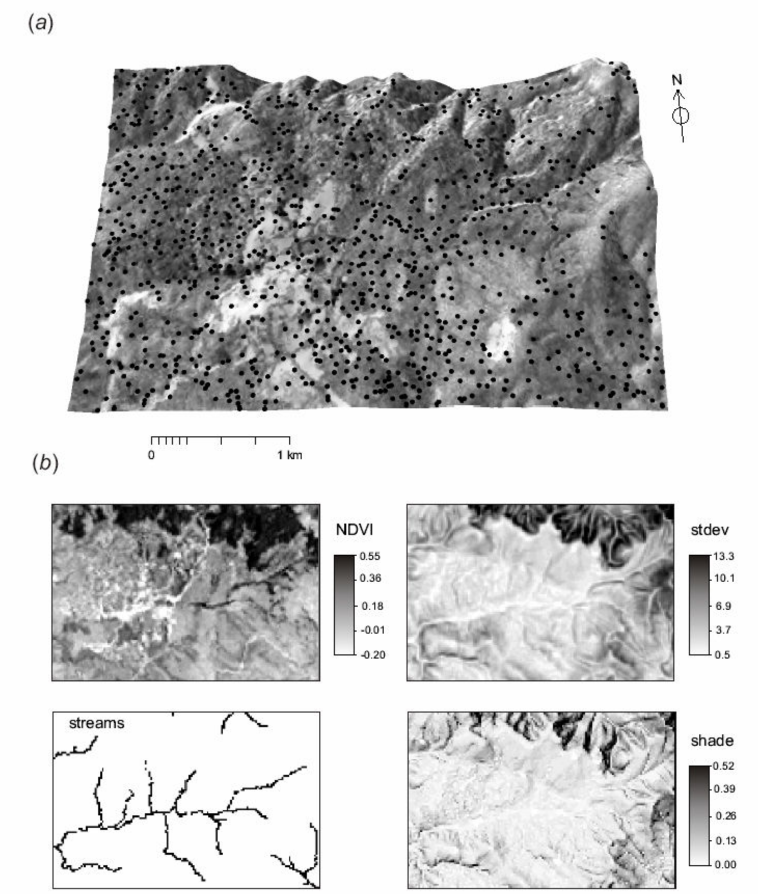

Fig. 1. The Study area Zlatibor: (a) perspective view on the area and location of 1020 error assessment points, (b) a preview of the auxiliary predictors used in geostatistical analysis.

We have considered four auxiliary variables that are of interest for locally adjusting the errors: (a) distance to streams (**dist.streams**), (b) terrain complexity (**stdev**), (c) analytical hillshading (**shade**) and (d) vegetation cover (**NDVI**). Distance from streams was derived by, first digitizing the stream lines from the topo map, and then running a distance operation in ILWIS GIS. The terrain complexity was estimated using a 5×5 window standard deviation filter in ILWIS GIS. This produces a map very similar to the slope gradient map and can be used to quantify the complexity of terrain. Analytical hillshading was derived in SAGA GIS using standard sun position settings for the geographic area. Both terrain complexity and analytical hillshading were derived from the existing 30 m DEM. NDVI was derived from the Landsat 7 ETM image. A preview of all auxiliary maps used in this exercises can be seen in Fig. 1b.

3. Results

We first fitted a variogram to the sampled elevations dataset (elevations) used to produce the DEM were fitted using a spherical variogram with nugget parameter C_0=0, a sill parameter of C_1=1561 and the range parameter of 1235 m. Elevations follow a normal distribution with a mean of 993 m and standard deviation of 46 m. The two predictors (dist.streams and NDVI) explain 34% of the total variation in the 2051 point measurements. Both predictors are significant at 0.001 probability levels. The variogram of the elevations, after using the predictors, has now 40% lower sill (968) and a similar range (1228 m). The results of generating predictions and simulations from sampled elevations without and with using the auxiliary maps can be seen in Fig. 2(a) and (b).

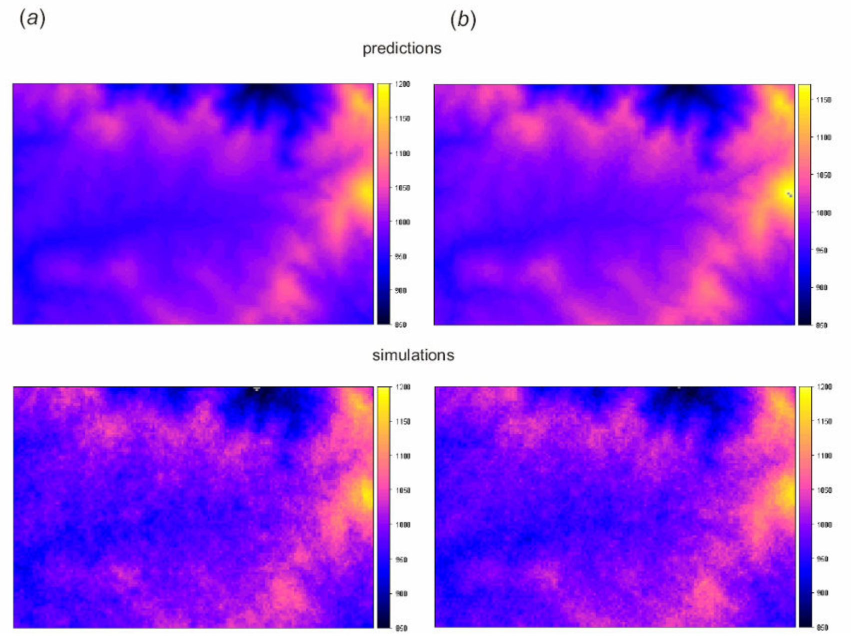

Fig. 2. DEM generation using geostatics: (a) without using auxiliary information: ordinary kriging (above) and simulations using ordinary kriging (below) and (b) with the help of auxiliary information: regression-kriging using auxiliary predictors – predictions (above) and simulations (below).

The deltas assessed at control points range from -7.2 to 7.8 m with an average of 0.18 m and a standard deviation (RMSE) of 1.2 m. The variogram of deltas was fitted with a nugget of 1.07, sill parameter of 0.42 and a range parameter of 332 m (Fig. 3a). In this case, the short range variation is significant, close to the pure nugget effect, which indicates that the measurement error of the original elevations indicated on the topo-map is about ±1 m. If we simulate these errors without any auxiliary information, the error surface will show almost no spatial structure (Fig 3c).

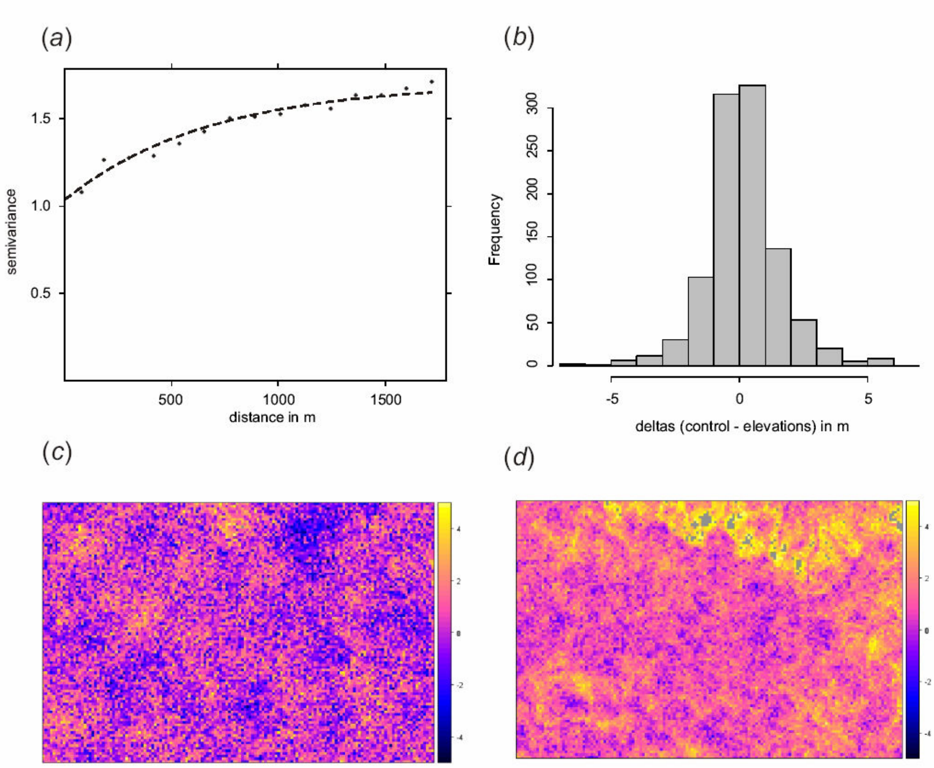

Fig. 3. DEM error assessment using the control dataset: (a) fitted variograms; (b) histograms of errors (deltas); (c) simulated error surface without auxiliary information and (d) with auxiliary information.

Regression analysis or errors versus auxiliary maps showed that the errors are significantly correlated with terrain complexity (stdev) and with terrain orientation (shade). The variogram model of the errors, after including the two predictors, changes drastically. The nugget parameter is now smaller, so is the sill and the range (114 m). This shows that the predictors are fairly useful in explaining the source of errors. Consequently, the resulting maps (Fig. 3b) reflect this dependency and the maps of errors are more realistic.

4. Conclusions and discussion

The proposed methodology is fit for use with any types of DEMs, provided that both auxiliary maps and a point datasets are available and suited for the area of interest (complementary effective scale, good coverage of space and features). We have demonstrated that there is indeed a benefit of using auxiliary maps that can help us generate more realistic DEMs (especially considering the hydrological features) and explain the spatial distribution of the inherent errors. In this case study, errors were positively correlated with terrain complexity (stdev) and with vegetation (NDVI). This confirms the previous results of Fisher (1998).

The major challenges for further refinement of these techniques will be a proper design and use of auxiliary data for such analysis. We especially anticipate that these techniques will be more and more attractive for handling rich LiDAR and radar-based topographic data, both to analyse their inherent geostatistical properties and generate DEMs fit-for-use in various environmental and earth science applications.

5. References

Fisher, P., October 1998. Improved Modeling of Elevation Error with Geostatistics. *GeoInformatica* 2 (3), 215–233.

Heuvelink, G.B.M., 1998, *Error propagation in environmental modelling with GIS* (New York: Taylor and Francis).

Holmes, K.W., Chadwick, O.A. and Kyriakidis, P.C., 2000, Error in a USGS 30m digital elevation model and its impact on digital terrain modeling. *Journal of Hydrology*, 233, pp. 154-173.

Kyriakidis, P.C., Shortridge, A.M. and Goodchild, M.F., 1999, Geostatistics for Conflation and Accuracy Assessment of Digital Elevation Models. *International Journal of Geographical Information Science*, 13, pp. 677-708.

Oksanen, J. and Sarjakoski, T., 2005, Error propagation of DEM-based surface derivatives. *Computers & Geosciences*, 31, pp. 1015-1027.

Oksanen, J., 2006a, *Digital elevation model error in terrain analysis* (PhD thesis, Faculty of Science, University of Helsinki: Publications of the Finnish Geodetic Institute).

Oksanen, J.J.A., 2006b, Uncovering the statistical and spatial characteristics of fine toposcale DEM error *International Journal of Geographical Information Science*, 20, p. 345.

Raaflaub, L.D. and Collins, M.J., 2006, The effect of error in gridded digital elevation models on the estimation of topographic parameters. *Environmental Modelling & Software*, 21, pp. 710-732.