Computers & Geosciences 34 (2008) 1886–1899

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Geostatistical modeling of topography using auxiliary maps

Tomislav Hengl [a,*], Branislav Bajat [b], Dragan Blagojević [b], Hannes I. Reuter [c]

[a] *Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*
[b] *Faculty of Civil Engineering, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia*
[c] *European Commission, Directorate General JRC, Institute for Environment and Sustainability, TP 280, Via E. Fermi 1, I-21020 Ispra (VA), Italy*

ARTICLE INFO

Article history:
Received 29 May 2007
Received in revised form
22 December 2007
Accepted 3 January 2008

Keywords:
DEM
R
Variogram
Regression-kriging
Error
Simulations
Elevation

ABSTRACT

This paper recommends computational procedures for employing auxiliary maps, such as maps of drainage patterns, land cover and remote-sensing-based indices, directly in the geostatistical modeling of topography. The methodology is based on the regression-kriging technique, as implemented in the R package gstat. The computational procedures are illustrated using a case study in the south-west part of Serbia. Two point data sets were used for geostatistical modeling: (1) 2051 elevation points were used to generate DEMs and (2) an independent error assessment data set (1020 points) was used to assess errors in the topo-DEM and the SRTM-DEM. Four auxiliary maps were used to improve generation of DEMs from point data: (1) distance to streams, (2) terrain complexity measured by standard deviation filter, (3) analytical hillshading map and (4) NDVI map derived from the Landsat image. The auxiliary predictors were significantly correlated with elevations (adj.$R^2 = 0.20$) and DEM errors (adj.$R^2 = 0.27$). By including auxiliary maps in the geostatistical modeling of topography, realizations of DEMs can be generated that represent geomorphology of a terrain more accurately. In addition, downscaling of a coarse 3 arcsec SRTM DEM using auxiliary maps and regression-kriging is demonstrated using the same case study. A methodological advantage of regression-kriging, compared to splines, is the possibility to automate the data processing and incorporate multiple auxiliary predictors. The remaining open issues are computational efficiency, application of local regression-kriging algorithms and preparation of suitable auxiliary data layers for such analyses.

1. Introduction

A digital elevation model (DEM) is a digital representation of the land surface—the major input to quantitative analysis of topography, also known as digital terrain analysis or geomorphometry (Evans, 1972; Pike, 1995; Wilson and Gallant, 2000; Hengl and Reuter, 2008). Typically, a DEM is a raster map (an image or an elevation array) that, like many other spatial features, can be efficiently modeled using geostatistics. The geostatistical concepts were introduced in geomorphometry by Fisher (1992, 1998) and Wood and Fisher (1993), then further elaborated by Kyriakidis et al. (1999), Holmes et al. (2000) and Oksanen (2006b). The methodological developments and future trends of geostatistical modeling of topography can be 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 analysis of effects that the DEM errors have on the results of spatial modeling. This is the principle of error propagation that commonly

* Corrresponding author. Tel.:+31 20 5257458; fax:+31 20 5257431.
E-mail addresses: hengl@science.uva.nl (T. Hengl), bajat@grf.bg.ac.yu (B. Bajat), bdragan@grf.bg.ac.yu (D. Blagojević), hannes.reuter@jrc.it (H.I. Reuter).

0098-3004/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2008.01.005

works as follows: simulations are generated from point-measured heights to produce multiple equiprobable realizations of a DEM of an area; a spatial model is applied m times and output maps then analyzed for mean values and standard deviations per pixel; the results of analysis can be used to quantify DEM accuracy and observe impacts of uncertain information in various parts of the study area (Hunter and Goodchild, 1997; Heuvelink, 1998; Oksanen and Sarjakoski, 2005; Raaflaub and Collins, 2006). An animated illustration of an error propagation study for derivation of slope, curvatures, solar insolation and northness maps is available at `www.geomorphometry.org` (under DEM simulations).

So far, DEMs have been generated by using solely point-sampled elevations. For example, ordinary kriging is used to generate DEMs (Mitas and Mitasova, 1999; Lloyd and Atkinson, 2002); conditional geostatistical simulations are used to generate equiprobable realizations of DEMs (Fisher, 1998; Kyriakidis et al., 1999). In most studies, no auxiliary (also known as *additional* or *secondary*) information on variation of relief is employed directly in the geostatistical modeling. Compared to the approach of Hutchinson (1989) or Hutchinson (1996) where auxiliary maps of streams are often used to produce hydrologically correct DEMs, the geostatistical approach to modeling of topography has often been limited to analysis of point-sampled elevations.

Our interest in this paper is to develop and test a more advanced methodology to model topography using geostatistics by including the auxiliary maps directly into the geostatistical analysis. By auxiliary maps, we consider all GIS layers that can explain spatial distribution of measured elevations and associated errors. Our assumption is that, by including such information, we will be able to produce more accurate realizations of DEMs and, consequently, advance the use of geostatistics in geomorphometry.

2. Theory

A DEM can be defined as an elevation array with a (large) number of grid nodes over the domain of interest:

$$\mathbf{Z} = \{Z(\mathbf{s}_j), j = 1, \ldots, N\}; \quad \mathbf{s}_j \in \mathbb{A} \tag{1}$$

where $\mathbf{Z}$ is the elevation array, $Z(\mathbf{s}_j)$ is the elevation at the grid node $\mathbf{s}_j$, $\mathbb{A}$ is the area of interest and N is the total number of grid nodes. DEMs are today increasingly produced using automated (mobile GPS) field sampling of elevations or airborne scanning devices (radar or LiDAR-based systems). In the case elevations are sampled at sparsely located points, a DEM can be generated using geostatistical techniques such as ordinary kriging (Wood and Fisher, 1993; Mitas and Mitasova, 1999). The elevation at some grid node ($\mathbf{s}_0$) of the output DEM can be interpolated using:

$$\hat{Z}(\mathbf{s}_0) = \mathbf{c_0^T} \times \mathbf{C}^{-1} \times \mathbf{z} \tag{2}$$

where $\mathbf{c_0^T}$ is the covariance vector at the prediction location, $\mathbf{C}$ is the covariance matrix at sampled locations $\mathbf{s}_i$ and $\mathbf{z}$ is an array of sampled points ($z(\mathbf{s}_i);\ i = 1, \ldots, n$). The covariances are determined by fitting a variogram model $\gamma(\mathbf{h})$, where $\mathbf{h}$ is the distance between the point pairs of sampled elevations.

The same technique (kriging) can be used to produce simulated DEMs. An equiprobable realization of a DEM can be generated by using the sampled elevations and their variogram model:

$$Z^{(\mathrm{SIM})}(\mathbf{s}_0) = E\{Z|z(\mathbf{s}_j), \gamma(\mathbf{h})\} \tag{3}$$

where $Z^{(\mathrm{SIM})}$ is the simulated value at the prediction location. The most common technique in geostatistics that can be used to generate equiprobable realizations is the Sequential Gaussian Simulation (Goovaerts, 1997, pp. 380–392). It starts by defining a random path for visiting each node of the grid once. At first node, kriging is used to determine the location-specific mean and variance of the conditional cumulative distribution function. A simulated value can then be drawn by using the inverse normal cumulative distribution function, e.g. (Box and Muller, 1958):

$$z^{\mathrm{SIM}}(\mathbf{s}_0) = \hat{z}(\mathbf{s}_0) + \hat{\sigma}(\mathbf{s}_0) \cdot \sqrt{-2 \cdot \ln(1 - A)} \cdot \cos(2 \cdot \pi \cdot B) \tag{4}$$

where z^{SIM} is the simulated elevation, A and B are the independent random numbers within the $0-0.99\ldots$ range, $\hat{z}$ is the predicted (mean) value at $\mathbf{s}_0$ location and $\hat{\sigma}$ is the prediction error at that location. The simulated value is then added to the original data set and the procedure is repeated until all nodes have been visited. Direct simulation of DEMs using the sampled elevations is discussed in detail by Kyriakidis et al. (1999).

If additional, auxiliary maps (drainage network, water bodies, physiographic break-lines) are available, a DEM can be generated from the point-measured elevations using the regression-kriging model (Christensen, 2001, p. 277):

$$\hat{Z}(\mathbf{s}_0) = \mathbf{q_0} \times \hat{\boldsymbol{\beta}}^{\mathbf{T}} + \mathbf{c_0^T} \times \mathbf{C}^{-1} \times (\mathbf{z} - \mathbf{q} \times \hat{\boldsymbol{\beta}}^{\mathbf{T}}) \tag{5}$$

where $\mathbf{q_0}$ is the vector of predictors at new location, $\hat{\boldsymbol{\beta}}$ is the vector of fitted regression coefficients and $\mathbf{q}$ is the matrix of predictor values at sampled locations.

The biggest advantage of using auxiliary maps is a possibility to more precisely model uncertainty of the sampled elevations and analyze which external factors cause this variability. Whereas, in pure statistical Monte Carlo approach where we work with global, constant parameters (Fig. 1a), in the case of geostatistical modeling, the DEM uncertainty can be modeled with a much higher level of detail (Fig. 1b).