A tool for assessing error in digital elevation models from a user's perspective

A. R. Darnell[1], N. J. Tate[2], and C. Brunsdon[2]

[1]School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ.
Tel: +44 (0)1603 591346, Fax: +44 (0)1603 458553
Email: a.darnell@uea.ac.uk
[2]Dept of Geography, University of Leicester, University Rd, Leicester, LE1 7RH

1. Introduction

A Digital Elevation Model (DEM) is a representation of geographic reality. The elevations recorded within DEMs have been shown to contain errors pertaining to sampling, measurement and interpolation (Fisher, 1998). Even a small amount of elevation error can greatly affect derivative products (Holmes et al., 2000). This can potentially have a significant impact on the application of DEMs in Geographical Information Systems (GIS) where first and second order derivatives are considered.

DEM vendors generally provide users with a measure of vertical accuracy in the form of the Root Mean Squared Error (RMSE) statistic only. However many papers have reported on the limitations of a single value of accuracy, stressing that DEM error is spatially autocorrelated (e.g. Kyriakidis *et al.*, 1999). Arguably, the best method for error modelling is based on conditional stochastic simulation (Fisher, 1998). Conditioning the simulation model includes sample observations of error and thus allows consideration of spatial autocorrelation. Unfortunately computation is complex.

The main aim of this research was to simplify existing procedures to enable the 'average' DEM user to perform his/ her assessment on the implications of choosing a particular dataset for their work. User requirements were identified and a methodology was designed to adhere to the essential requirements, whilst maintaining the option of modifying defaults (Table 1). As an application the use of a DEM for landslide susceptibility modelling was considered to add context to discussion and demonstrate the relevance of the propagation of error consideration.

2. Methods

The study area occupies approximately 25km^2 of north western Slovenia and focuses

Requirement type	*Description*	*Necessity*
Data:	Test DEM(s)	Essential
	Higher accuracy data for reference surface	Essential
Variables:	Sample size and point locations	Essential
	Grid resolution for display of results	Optional (default)
	Number of simulations (*N*)	Optional (default)
Knowledge/ skills:	Variogram interpretation	Essential basic
	Interpretation of results	Essential
External Software:	Manipulation of data (e.g. MS Excel)	Essential
	Further visualisation (e.g. Surfer, Golden Software)	Optional

Table 1 User requirements in the error assessment methodology

Two independent test DEMs were provided by the Environmental Agency of the Republic of Slovenia at 12.5m and 25m grid spacing. A Light Detection and Ranging (LiDAR) dataset was used as a surrogate for the ‘true’ elevation, from which 100 sample points were randomly selected to be representative of the reference elevation surface. Error is defined here as the disparity in the elevation value projected by a DEM and its true value, and is given by subtracting the DEM value from the reference surface (after Heuvelink, 1998). The geostatistical modelling and simulation was performed with the GSTAT package used in the R software environment (Pebesma, 1999). This is freely accessible software copyrighted under the General Public Licence (GPL). Prior to evaluation an algorithm was written in R that would automatically set up grid nodes for calculation, dependent on the bounds set by the minimum and maximum coordinate sample values (grid resolution defaults to $100m^2$). There were four main stages to represent the uncertainty and demonstrate the propagation of error to the landslide model.

1) *Modelling spatial dependence*
   A variogram was used to show the spatial autocorrelation of error. A model was achieved in GSTAT by a two-step procedure of first calculating the sample variogram from raw data and then fitting a model. Interpretation was required to select an appropriate model (Table 1).

2) *Stochastic simulation*
   The variogram model determined above was preserved and used with Inverse Distance Weighted (IDW) interpolation and sequential Gaussian simulation to generate *N* error map realisations. Each error realisation was added to the

4) *Propagation to landslide susceptibility modelling*

The landslide susceptibility model defined areas of slope angle greater than 14 degrees as susceptible to the hazard of landslide. Discrete valued categories (such as 'susceptible' and 'non-susceptible') can contain errors due to misclassification (Zhang and Goodchild, 2002). To examine the probable and possible uncertainties in classifying the landslide hazard, two frameworks were adopted: probabilistic and fuzzy.

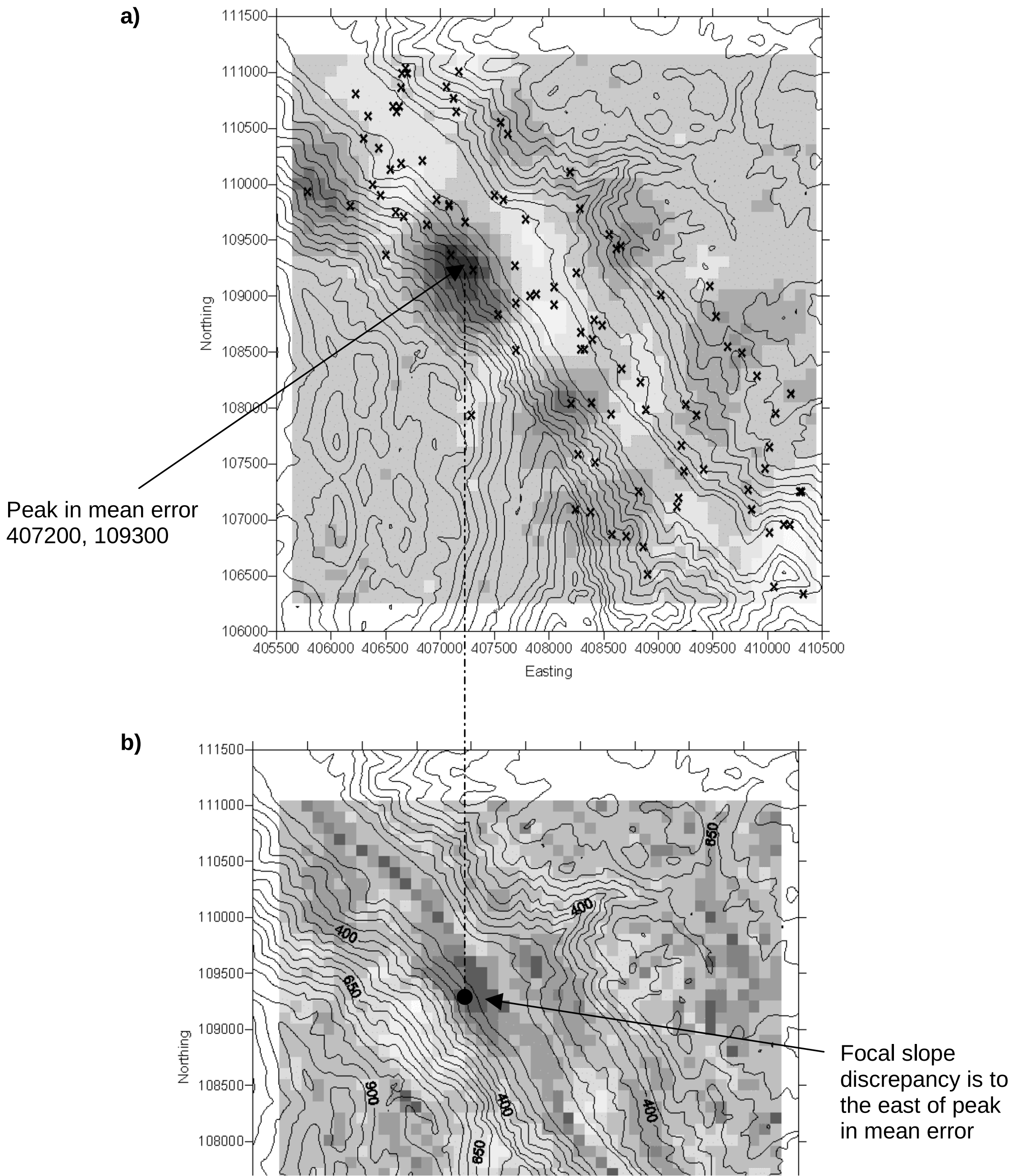
i. <u>Probability of susceptibility classification:</u> A nominal value of *true* was given to the simulated cell if its slope value exceeded the critical angle and *false* if it did not. Counting up the true and false declarations for each cell and dividing by *N* gave the probability of that cell being susceptible to landslide.

ii. <u>Fuzziness in susceptibility:</u> For the *N* simulations the minimum and maximum simulated values of slope (for each grid cell) were recorded. Each value was then tested against the critical angle as before and the cell attribute was set to *true* if it was greater than 14 degrees and *false* if it was equal to or less than. A three-tier classification system was used to define each cell. For each cell if both minimum and maximum values were *true* then that cell was given a value of 1 and was susceptible to landslide. If either minimum or maximum was *true* then the node was given a value of 0.5 and classified undecided. If neither was *true*, the cell was not susceptible and given a value of 0.

3. Results and analysis

The spatial distribution of error was similar for the DEMs of differing resolution, although the magnitude was slightly higher for the 25m DEM. The relative spatial variation of mean error for the 12.5m DEM is shown in Figure 1a. Graphics were created using the Lattice package in GSTAT and (optionally) projected onto a topographic map. The mode of formation was different for each DEM so a similar spatial distribution may imply a terrain-dependent error causal factor (e.g. Liu and Jezek, 1999). To demonstrate the propagation of error, slope values directly derived from each DEM (no consideration of uncertainty) were compared to the mean values of the slope realisations (Figure 1b). This provided an indication of how slope estimation would vary were it derived directly from the DEM or from the multiple equally likely realisations. The main region of slope discrepancy was adjacent to the peak in mean elevation error, thus supporting the hypothesis of error propagation and the previous work of authors such as Murillo and Hunter (1997).

comparison with the 25m DEM, the 12.5m DEM classifies a greater number of cells as ‘susceptible’ or ‘non susceptible’.



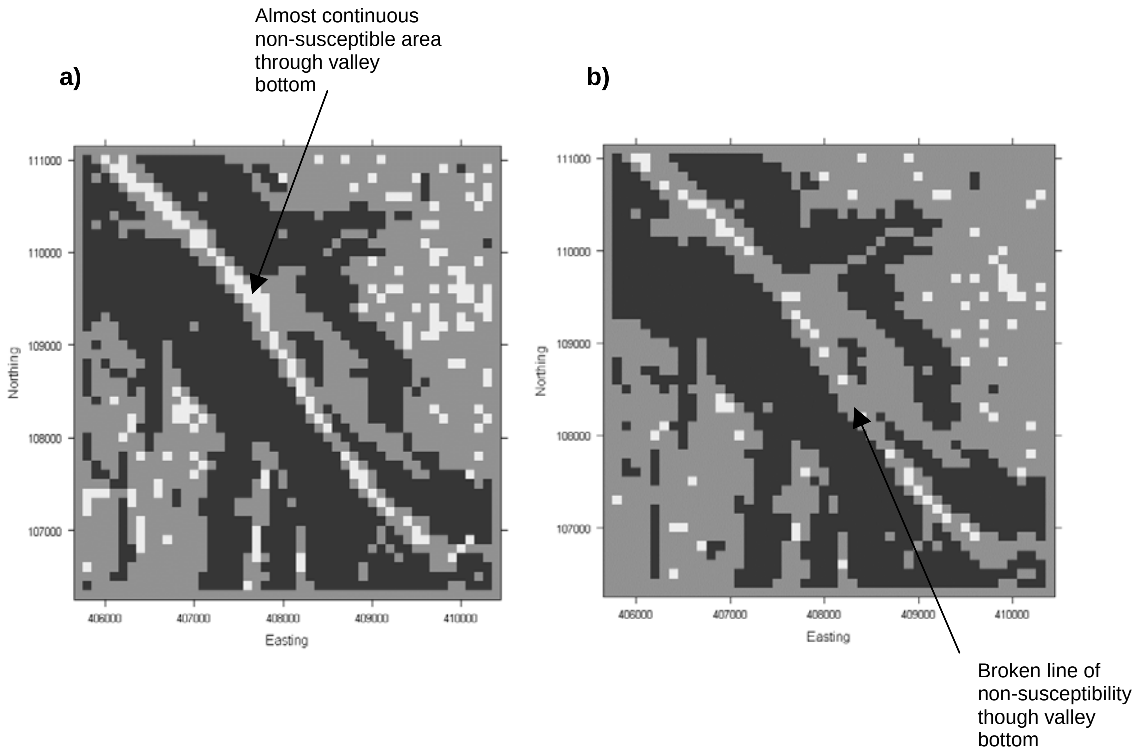
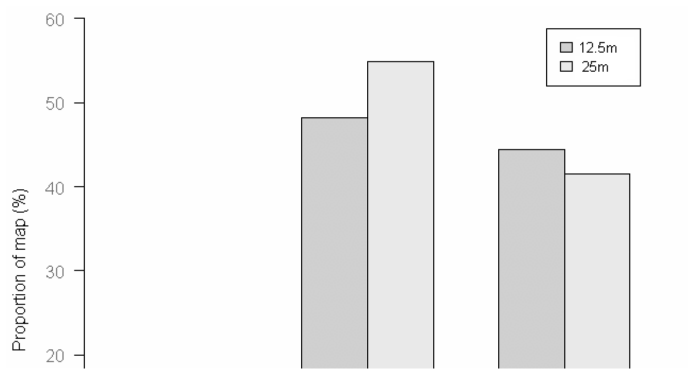


Figure 2 Fuzziness in defining cells as susceptible; darkest cells = 1 susceptible, lightest cells = 0 not susceptible, other = 0.5 undecided; a) 12.5m DEM; b) 25m DEM



4. Conclusion

The methodology developed here provides an instrument for error quantification, demonstration of propagation, and visualisation that is a simplification on existing techniques. Publicly available software has been used to facilitate a universally distributable and pliable tool. On the basis of this study a fuzzy framework proved to be the most useful approach for highlighting the consequences of using different DEMs for landslide hazard assessment. The existing code for stages 1 to 3 of the methodology could be used for any application. Following minor modifications, the work could be integrated into fitness for use assessment, risk management studies and cost benefit analyses etc. The necessity of user expertise was successfully minimised but it is obligatory for the user to have a general understanding of uncertainty analysis and variogram modelling. Uninformed interpretation of these aspects could add a further dimension to error propagation.

5. Acknowledgements

This work was undertaken in partial fulfilment of the requirements for the degree of Master of Science and submitted to the University of Leicester. The author would like to gratefully acknowledge Andej Gosar at the Environment Agency of the Republic of Slovenia and Marko Komac at the Geological Survey of Slovenia.

6. References

Aguilar, F. J., Agular, M. A., Aguera, F., and Sanchez, J., 2006, The accuracy of grid digital elevation models linearly constructed from scattered sample data. *International Journal of Geographical Information Science*, **20**, 169-192.

Fisher, P., 1998, Improved Modeling of elevation error with geostatistics. *GeoInformatica*, **2**, 215-233.

Heuvelink, G. B. M., 1998, *Error propagation in environmental modelling with GIS*. Research Monographs in GIS series. (London: Taylor and Francis) pp127.

Holmes, K. W., Chadwick, O. A., and Kyriakidis, P. C., 2000, Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology*, **233**, 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**, 677-707.

Liu, H., and Jezek, K. C., 1999, Investigating DEM error patterns by directional variograms and Fourier analysis. *Geographical Analysis*, **31**, 249-266.

Murillo, M. L., and Hunter, G. J., 1997, Assessing uncertainty due to elevation error in a landslide susceptibility model. *Transactions in GIS*, **2**, 289-298.

Pebesma, E. J., 1999, *GSTAT user's manual*, http://www.gstat.org/manual/gstat.html