

## Improving user assessment of error implications in digital elevation models

Amii R. Darnell \*, Nicholas J. Tate, Chris Brunsdon

Department of Geography, University of Leicester, University Road, Leicester LE1 7RH, UK

### ARTICLE INFO

#### Keywords:

Error  
Uncertainty  
Stochastic simulation  
Digital elevation models  
Propagation

### ABSTRACT

A digital representation of a terrain surface is an approximation of reality and is inherently prone to some degree of error and uncertainty. Research in uncertainty analysis has produced a vast range of methods for investigating error and its propagation. However, the complex and varied methods proposed by researchers and academics create ambiguity for the dataset user. In this study, existing methods are combined and simplified to present a prototype tool to enable any digital elevation model (DEM) user to access and apply uncertainty analysis. The effect of correlated gridded DEM error is investigated, using stochastic conditional simulation to generate multiple equally likely representations of an actual terrain surface. Propagation of data uncertainty to the slope derivative, and the impact on a landslide susceptibility model are assessed. Two frameworks are developed to examine the probable and possible uncertainties in classifying the landslide hazard: probabilistic and fuzzy. The entire procedure is automated using publicly available software and user requirements are minimised. A case study example shows the resultant code can be used to quantify, visualise and demonstrate the propagation of error in a DEM. As a tool for uncertainty analysis the method can improve user assessment of error and its implications.

© 2008 Elsevier Ltd. All rights reserved.

## 1. Introduction

### 1.1. The importance of considering error in DEMs

Terrain is a topographic surface that varies continuously over space. It is multiscale, including mountain ranges, river catchments and individual slopes. The conceptualisation of topographic reality can be represented in the form of a digital elevation model (DEM); however it is not feasible to sample infinite points or to create a continuous data structure from a finite data set. Thus approximations need to be made from discrete data that model the reality of terrain as closely as possible. Digital models of terrain take discrete data values of the elevation continuum and by using interpolation methods, reconstruct a surface. Error is inherent due to this approximation and can be considered the disparity in the elevation value projected by a DEM and a reference for the 'true' surface. In contrast, uncertainty is usually considered as the lack of knowledge about the reliability of a DEM's representation of the true surface (after Hunter & Goodchild, 1997).

The elevations recorded within DEMs have been shown to contain errors derived from a variety of sources: sampling, measurement and interpolation (Fisher, 1998; Fisher & Tate, 2006). Even

small elevation errors can greatly affect derivative products (Holmes, Chadwick, & Kyriakidis, 2000). This can potentially have a significant impact on the application of DEMs in Geographical Information Systems (GIS) where derivatives of elevation are frequently used. For example the propagation of error to slope can have a significant influence on soil erosion prediction models (Warren, Hohmann, Auerwald, & Mitasova, 2004). Therefore, after establishing that errors exist in DEMs it is necessary to acknowledge there is a risk that the outcome of any analysis based on this information will be incorrect (Fisher, 1998).

### 1.2. Spatial dependency and variability

DEM vendors such as the United States Geological Survey (USGS) generally provide users with a measure of vertical accuracy in the form of the root mean squared error (RMSE) statistic (USGS, 1997). A global accuracy measure such as this has its advantages: for example it is relatively quick to calculate, easy to report and it is a standard recognised worldwide. Furthermore, as it implies that error is uniform across the DEM (Burrough & McDonnell, 1998) comparison of different models is simple using this method. However many papers have reported on the limitations of a single value of accuracy, stressing that DEM error is spatially variable and autocorrelated (e.g. Fisher, 1998; Holmes et al., 2000; Kyriakidis, Shortridge, & Goodchild, 1999). Thus not only is error spatially variable throughout a DEM, but within the model the error value of an individual grid cell is related to errors from neighbouring cells.

\* Corresponding author. Present address: School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. Tel.: +44 (0)1603591346; fax: +44 (0)1603458553.

E-mail address: [a.darnell@uea.ac.uk](mailto:a.darnell@uea.ac.uk) (A.R. Darnell).

This spatial variability and spatial dependency (termed ‘spatial autocorrelation’) can be crucial for environmental applications of DEMs and its existence has been examined in many recent studies.

Several authors report the magnitude of DEM error is closely related to the characteristics of the terrain surface. For example slope will influence interpolation procedures (Kidner, 2003). The positive correlation of error and slope has been shown in numerous studies, e.g. Hunter and Goodchild (1997), Veregin (1997) and Fisher (1998); yet on mathematical surfaces (synthetic DEMs) smoother, flatter areas generate a greater amount of uncertainty in their derived parameters (Zhou, Liu, & Sun, 2006). However, a more recent study by Holmes et al. (2000) found only a weak correlation between calculated DEM error and various terrain attributes. This work largely adopted the geostatistical methods of Kyriakidis et al. (1999) but whereas the earlier study found high DEM error associated with areas of rough terrain, Holmes et al. (2000) found correlation negligible. In another study Liu and Jezek (1999) explored the spatial autocorrelation pattern of DEM error and found it to be closely reliant on the ruggedness of the terrain under investigation. However they stressed that surface slope cannot be simply used to predict the error magnitude; a linear relationship between the magnitude of DEM error and terrain surface slope exists for smooth terrain surfaces only.

Hence it is intuitive that certain types of terrain are more suited to the creation of accurate DEMs (Carlisle, 2005). A distinction must be made between the characteristics of the actual terrain and those represented by the digital model. There are two main ideas here: (1) actual terrain characteristics can induce error in the representation of elevation, and (independently) (2) error once incorporated in the terrain model can propagate to derivatives, giving a false representation of a terrain characteristic. Irrespective of which of these ideas prompts the research, many authors are agreed there is an intimate relationship between elevation error and terrain, and that this is influenced by the spatial autocorrelation of the error.

### 1.3. Modelling error and its propagation

There are two approaches to error modelling: analytical error models and (conditional and unconditional) stochastic simulation (Fisher & Tate, 2006). The most frequently cited analytical model of error was developed by Hunter and Goodchild (1995). This basic model was based on simple probability theory and the RMSE of the DEM. The RMSE was taken to be an estimate of the local error variance around a measured elevation value. This technique is useful, as it doesn’t require the comparison of DEM data with higher accuracy data however the limitations of a global accuracy measure have been presented (e.g. Wise, 2000). Alternatively geostatistics can be used to model the actual distribution of error with reference to a higher accuracy dataset.

Arguably, the best method for error modelling is based on conditional stochastic simulation (Fisher, 1998). Stochastic imaging involves the modelling of spatial uncertainty through alternative numerical representations (maps) of the same reality (Journel, 1996). Multiple versions of the derived products from these realisations can also be determined, demonstrating error propagation (Fisher, 1998). Kyriakidis et al. (1999) describe the mathematical details of the process of stochastic simulation. Conditioning the simulation model preserves sample observations of error and thus allows consideration of spatial autocorrelation. Having modelled the error there are a number of methods for tracing the propagation of quantitative error from source to surface derivatives, such as Taylor series approximation, Rosenblueth’s method and Monte Carlo simulation (see Heuvelink, 1998 for a discussion). The Monte Carlo method has been cited as the best method for determining the influence of error on DEM derivatives (Veregin, 1997). It is also

advantageous because it is not affected by the complexity or non-linearity of the model (Burrough, 2002) and is thus generally applicable. Monte Carlo simulation is a technique for producing estimates of ‘true’ outcomes of stochastic processes by simply running many iterations of the model process and comparing the outcomes. The main disadvantage is the numerical load as the process is repeated for typically 50–2000 simulation runs (Heuvelink, 1998), but the method remains widely recommended.

### 1.4. Statement of the problem

Many researchers advocate that the responsibility of error consideration lies with the data producer (e.g. Aguilar, Aguilar, & Aguera, 2007) however many datasets are provided with RMSE information only. This research moves towards enabling the DEM user to be more appreciative of the possible implications of error in their work. However, as the methods for uncertainty analysis evolve to be more comprehensive, they also become more difficult to understand and implement. Analytical procedures for error investigation can be very intellectually and resource-intensive. Furthermore the variety of methods proposed by academics and researchers has created ambiguity for the ‘average’ DEM user. This problem may be compounded by a general apathy or ignorance on the part of DEM users. A recent paper by Wechsler (2003) surveyed a spectrum of DEM users worldwide and found that only half of the respondents considered their work influenced by uncertainty. Furthermore a significant proportion (25%) had no awareness at all of whether DEM errors affected their work and many reported that they would spend a minimum amount of time to assess uncertainty. This contributes to the general lack of quality standards for GIS products derived by non-experts (Couclelis, 2003).

A DEM is usually created by interpolating a regularly gridded surface from a discrete number of points. Errors due to the interpolation process were not examined by this research but for a full review of processing errors the reader is referred to Rees (2000), Wise (2000) and Fisher and Tate (2006) among others. 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 error propagation to decision outcomes. This proposition had two principal objectives. First a methodology was designed to quantify, visualise and demonstrate the propagation of error. The consideration of spatial autocorrelation was identified as significant and thus conditional stochastic simulation was used, as supported above. Secondly the technique was designed to minimise user requirements and maximise simplicity. Fundamentally software needed to be accessible to all DEM users which suggested free software copyrighted under the General Public Licence (GPL).

## 2. Methods

### 2.1. Overview

This paper presents the results of a prototype design to assess the feasibility of a system for automating the quantification, propagation and visualisation of error. The application of a DEM for landslide susceptibility modelling was considered to add context to discussion and demonstrate the relevance of the propagation of error consideration. Two frameworks were adopted to examine the probable and possible uncertainties in classifying the landslide hazard: probabilistic and fuzzy.

Sample DEMs were acquired for a case study. Relative errors at sample locations within each DEM were calculated with reference to a higher accuracy dataset. To model the distribution of error geostatistical modelling and Monte Carlo simulation were executed by GSTAT in the R software environment (Pebesma, 1999). R and GSTAT were both copyrighted under GPL, easy to install

and freely available for download from the Internet (CRAN, 2007). A working script in R was developed to automate the procedure by linking new and existing R functions, and developing new code for visualisation. The script was designed to adhere to the essential requirements, whilst maintaining the option of modifying defaults (Table 1). The user would benefit from an automated yet flexible procedure and (owing to frequent annotation of the code) the user

would be provided with explanation and/or instructions throughout.

Fig. 1 outlines the overall methodology. The approach can be divided into four main stages: (A) modelling spatial dependence, (B) stochastic simulation, (C) propagation to surface derivatives, and (D) propagation to a landslide susceptibility model. These stages will now be discussed, following introduction to the study area and reference to the theoretical error model. The uncertainty analysis methodology was finally incorporated into a procedure for the DEM user.

**Table 1**  
User requirements in the error assessment methodology

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 ( <i>N</i> )	Optional (default)
Knowledge/skills	Variogram interpretation	Essential basic
	Interpretation of results	Essential
External software	Manipulation of data (e.g. MS Excel)	Essential
	Further (topographic) visualisation (e.g. Surfer, Golden Software)	Optional

2.2. Case study: The slovenian mountains

The study area occupied approximately 25 km<sup>2</sup> of north-western Slovenia and focused on a section of the Idrija River valley. At the intersection of the Alps and Dinaride mountain range this region is tectonically active and the strike-slip faults pose a serious seismic hazard (Cunningham, Grebby, Tansey, Gosar, & Kastelic, 2006). Landslides are frequent. Previous research had revealed that only slopes of 14 degrees or greater showed an association with landslide occurrence that was significantly different from the random occurrence of landslides (Komac, 2006 personal communication). This figure was taken as a critical angle, *a*, for slope susceptibility to landslide in the study area, and was the simple landslide model applied.

Two test DEMs were provided by the Survey and Mapping Authority of Slovenia at 12.5 m and 25 m grid spacing. They were derived independently using a combination of Interferometric Synthetic Aperture Radar (InSAR) data, digitisation of paper maps and geodetic data (roads, railways, buildings, etc.). The DEMs were chosen for their differing resolution, a factor that has been shown to affect derived slope (Warren et al., 2004). A Light Detection and Ranging (LiDAR) dataset (Cunningham et al., 2006) 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. The height data resulting from the LiDAR survey had an absolute RMS accuracy of better than 15 cm (Cunningham et al., 2006). Wehr and Lohr (1999) provide a general overview of LiDAR principles and Huising and Gomes Perreira (1998) consider accuracy issues of LiDAR data. One hundred corresponding elevation values were therefore taken from each DEM for comparison. Sample locations were determined by the LiDAR flight path and were thus largely confined to the valley and its immediate vicinity.

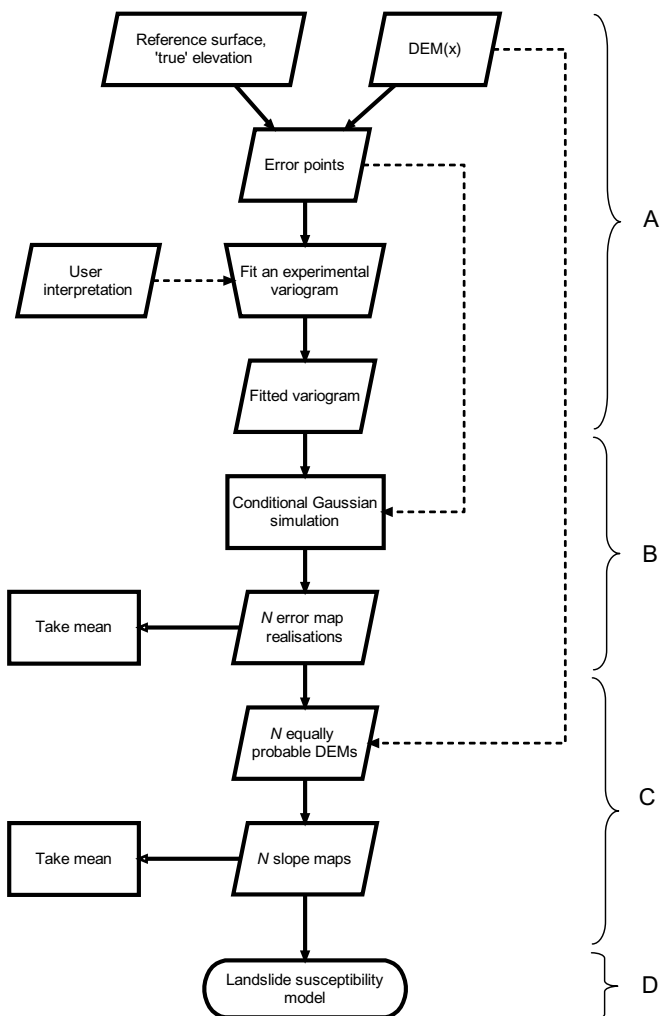
In addition to the higher accuracy LiDAR dataset, two DEMs of different resolutions were available. These were to test the prototype methodology and investigate the affect of differing resolution on derived slope. However multiple DEMs for the same area are not expected from the average DEM user; simply a DEM of unknown quality and a set of distributed sample elevation values of higher accuracy to act as a reference surface (for example elevation values from global positioning system (GPS) points or spot heights) are needed.

2.3. The error model

The first step was to define the error model. To find the error in a given DEM dataset the elevation given by the DEM may be subtracted from the ‘true’ elevation at a specific location. For a quantitative spatial attribute *A*(·), such as elevation, at some location *x* ∈ *D* its ‘true’ value is *A*(*x*) as defined by the model Heuvelink (1998):

$$A(x) = b(x) + V(x) \quad \text{for all } x \in D \tag{1}$$

where *b*(*x*) is the deterministic variable, i.e. the interpretation of *A*(·), and *V*(*x*) is the error.



**Fig. 1.** The four stages of error assessment. Solid arrows illustrate inputs and flow direction; dashed arrows indicate a process/object requiring further input.

2.4. Uncertainty analysis and error propagation

A grid was constructed dependent on the bounds set by the minimum and maximum coordinate sample values (grid resolution defaults to 100 m). For each DEM, error in elevation was calculated at each sample location. Where these coordinates did not correspond to grid nodes they were interpolated. The effect of correlated gridded DEM error was then investigated using stochastic conditional simulation to generate multiple equally likely representations of an actual terrain surface (as discussed Section 1.3). A raster error surface was then produced for each of the 100 simulations and the mean of these equally probable realisations was taken as being representative of the whole. Propagation of data uncertainty to the slope derivative and the impact on the landslide susceptibility model were finally assessed.

There were four main stages to represent the error and demonstrate the propagation of error to the landslide model.

2.4.1. Modelling spatial dependence

Geostatistical variograms have been employed to show the spatial autocorrelation of error (Fisher & Tate, 2006). A variogram summarises the relationship between differences in pairs of (error) measurements and the distance of the corresponding points from each other. A model was produced in GSTAT by a two-step procedure of first calculating the sample variogram from the raw error data and then fitting a variogram model (Pebesma & Wesseling, 1998). GSTAT supports many model types including linear, spherical, Gaussian, etc. Interpretation by the user was required to select an appropriate model, this was unavoidable. However the GSTAT user’s manual (Pebesma, 1999) provides some guidelines for this. The variogram model (as defined by the function, sill and/or range) was fitted to the sample variogram by weighted least squares estimation of the variogram model parameters (Pebesma & Wesseling, 1998).

2.4.2. Stochastic simulation

The variogram model determined above was preserved and used with sequential Gaussian simulation to generate *N* error map realisations. Each error realisation was added to the original test DEM (re-gridded to error nodes) to generate *N* equally probable DEMs. The basic steps of the conditional Gaussian simulation are detailed within Holmes et al. (2000) and Aerts, Goodchild, and Heuvelink (2003). The GSTAT program performed this simulation and required only the input data, the fitted variogram model and specification of *N*. *N* was specified to default to 100 after tests to optimize *N* whilst restraining numerical load for computer processing.

2.4.3. Propagation through to surface derivatives

For the next stage error propagation to slope was considered. Slope is the magnitude of the vector tangent to the topographic surface (Raaflaub & Collins, 2006) and the overall average slope was calculated according to the second-order finite difference method (Fleming & Hoffer, 1979). Gradient was computed from only the nearest four elevation points on the grid. Using the compass based notation adopted by Aguilar, Aguilar, Aguera, and Sanchez (2006) the average slope gradient, AS, was given by:

$$\tan AS \approx \sqrt{\left(\frac{Z_E - Z_W}{2\Delta x}\right)^2 + \left(\frac{Z_N - Z_S}{2\Delta y}\right)^2} \quad (2)$$

where *Z<sub>E</sub>*, *Z<sub>W</sub>*, *Z<sub>N</sub>*, and *Z<sub>S</sub>* were the elevations of the four cardinal points (east, west, north and south) and  $\Delta x$  and  $\Delta y$  represented the spacing of the grid DEM (resolution). Jones (1998) tested eight numerical slope algorithms (using synthetic and real data as test surfaces) and found that Fleming and Hoffer’s method had the lowest overall error for smooth surfaces. The slope algorithm calculation resulted in the generation of *N* equally probable slope maps.

2.4.4. Propagation to landslide susceptibility modelling

To test the methodology, a task was set within the context of a landslide susceptibility model. The landslide model itself was not a focus of the investigation and thus was very simple. The model defined areas of slope angle greater than a critical angle, *a*, as susceptible to the hazard of landslide. Discrete valued categories (such as ‘susceptible’ and ‘non-susceptible’) can contain errors due to misclassification (Zhang & Goodchild, 2002). To examine the probable and possible uncertainties in classification, two frameworks were adopted: probabilistic and fuzzy. Fuzzy logic deals with the imprecision of facts and possibilities, whilst probability deals with the likelihood of a precise event (Zadeh, 1999). Thus probabilities and possibilities can be used as different methods for conveying uncertainty in spatial data. Rebiasz (2007) applied both for uncertainty in investment project risk assessment. For a comprehensive discussion of these methods, the reader is referred to Zhang and Goodchild (2002). In this study the additional code was written as follows:

2.4.4.1. Probability of susceptibility classification. 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.

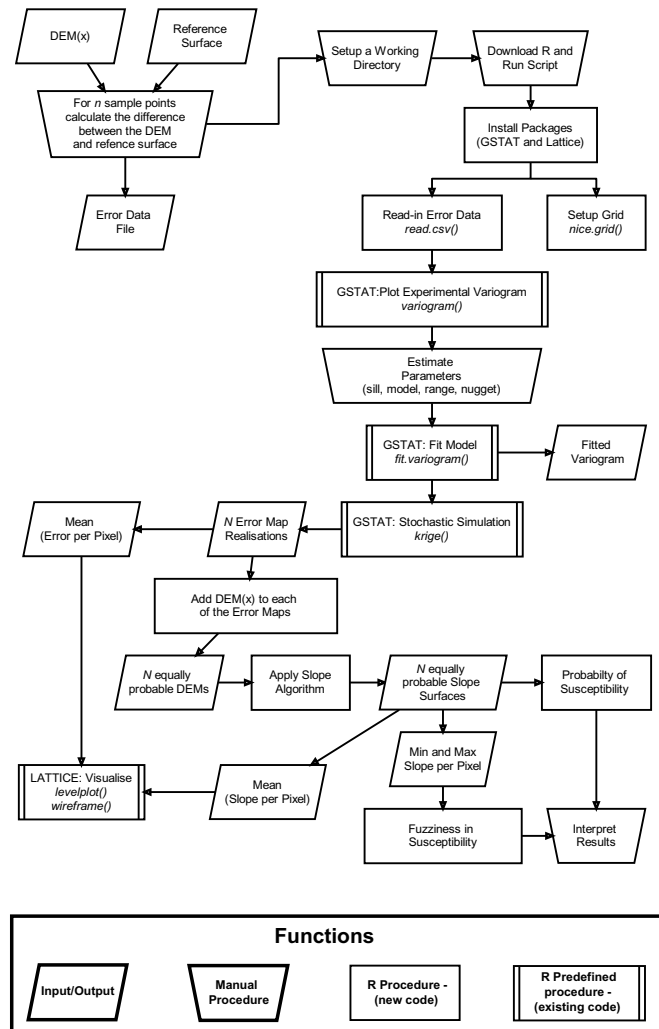


Fig. 2. Summary methodology for the user indicating different types of procedures and inputs.

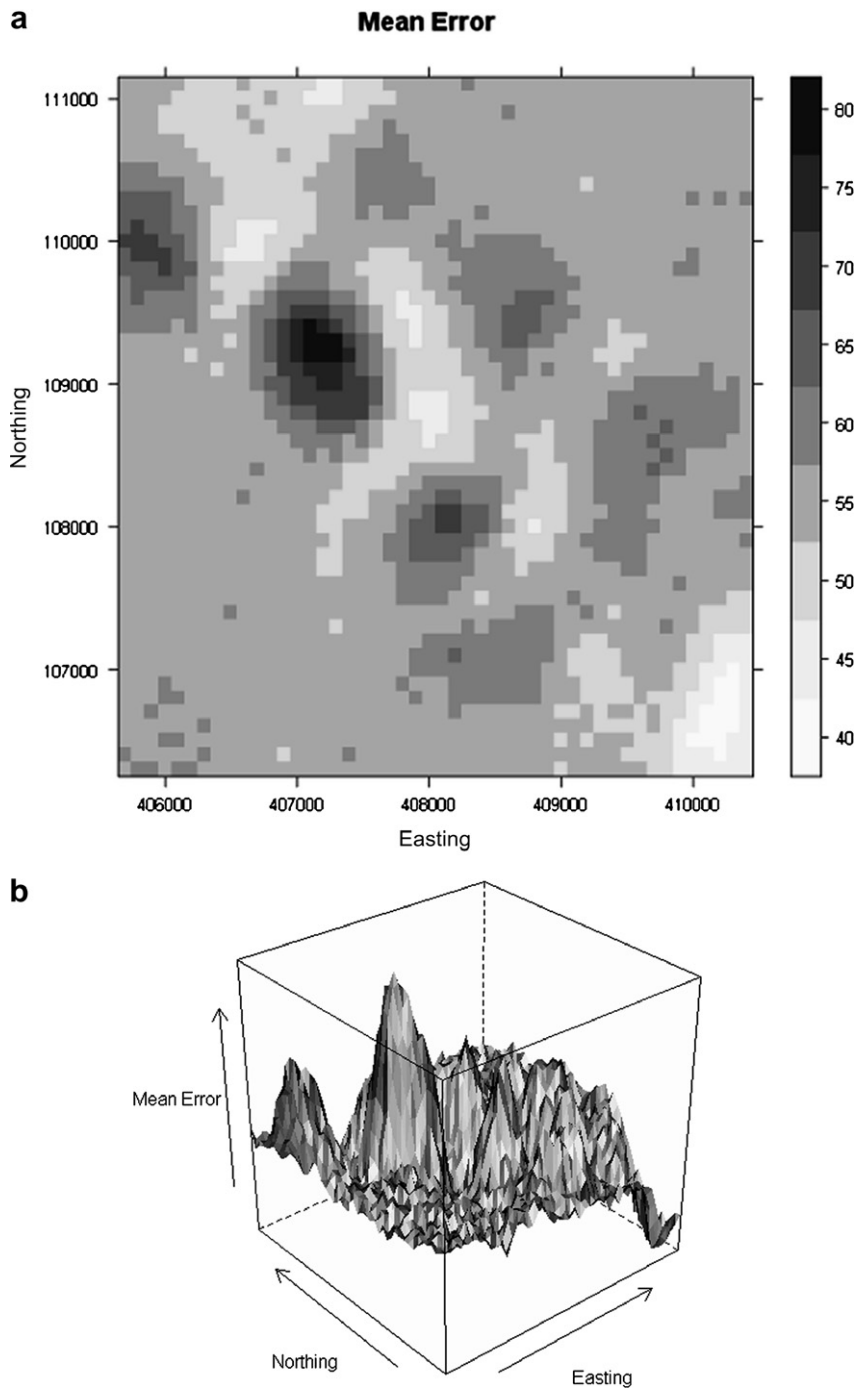


**2.4.4.2. Fuzziness in susceptibility.** For the  $N$  simulations the minimum and maximum simulated values of slope (for each grid cell) were recorded. Each value was then tested against the critical angle as before and the cell attribute was set to *true* if it was greater than  $a$  degrees and *false* 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 were *true* then the node was given a value of 0.5 and classified undecided. If neither were *true*, the cell was not susceptible and was given a value of 0. This was an adaptation of a fuzzy logic framework (see Zadeh, 1999). Davis and Keller (1997) used a more detailed fuzzy logic technique for slope stabil-

ity prediction, combining it with Monte Carlo simulation as adopted here.

### 2.5. Summary of methods: a procedure for DEM user

Stages A through D used existing GSTAT functions (e.g. *fit.variogram()*, *krige()*) and new algorithms for the slope, landslide susceptibility and visualisation. A fully annotated script was written to provide the user with full explanation of the processes at runtime. Fig. 2 shows the entire procedure for the DEM user, with new and existing functions highlighted. The procedure is also outlined below:



**Fig. 3.** Spatial distribution and magnitude of mean error for 12.5 m DEM: (a) levelplot using Lattice and (b) wireframe using lattice.

1. Download the R setup program (about 30 MB) and install.
2. Open the R environment and install two packages: GSTAT and Lattice.
3. Setup a working directory for the primary R script and those additional scripts called during runtime (e.g. the algorithm for slope calculation). There must also be at least one error data file (more if comparing DEMs). Error is the disparity between the reference surface and that projected by the DEM. These data should be stored as a comma separated variable file, three columns: coordinates  $x$  and  $y$ , and *error*.
4. Run the R script. This action prompts for the following user inputs where necessary:
  - (a) *Suitable grid resolution*. This defaults to 100 m but may be altered according to study area, sample size and simulation quantity.
  - (b) *Variogram parameters*. The sill, nugget and range need to be estimated from the experimental variogram in order to fit a model.
  - (c) *Number of simulations*. This defaults to 100 but can be changed.
  - (d) *Visualisation preferences and display of the results*.

Thus the program required minor setup but enough understanding to interpret (a) a variogram and (b) a visual display of the results.

### 3. Results

#### 3.1. Quantification and visualisation of error

An unexpected magnitude of error found in both case study DEMs was concerning. Metadata was provided and quoted a maximum error of 3.8 m and 13.8 m, for the 12.5 m and 25 m DEMs

respectively (courtesy of Survey and Mapping Authority of Slovenia). The mean error value of approximately +54 m for both DEMs suggested a systematic error uniformly added to both. A likely cause of this base error was an artefact of different sea level definitions. A uniform error did not impact on the significance of the results as relative error values were unaffected. Furthermore the case study was incidental, serving only to demonstrate the methodology, for which it performed satisfactorily. Therefore the baseline error was noted but not modified.

The spatial distribution of error was similar for DEMs of differing resolution, although the peak in mean error was slightly greater for the (coarser) 25 m DEM. The relative spatial variation of mean error for the 12.5 m DEM is shown in Fig. 3. Graphics were created using the Lattice package in R; these two visualisations show the flexibility of the package in offering a plan view, or 'Levelplot', (Fig. 3a) and a 3D 'Wireframe' (Fig. 3b) of the mean error surface. All graphics can be manipulated by the user (in terms of axes, colour schemes, labels, etc.) but default to those shown in the figures presented. Images (in addition to data) can be exported from R and saved.

Supplementary to the main enquiry, the data were also used for a quick investigation of the relationship between error and slope (derived using the second-order finite distance method). The results of this study appear to show a positive correlation of error and slope at a slope angle of 30° or greater (Fig. 4). At this angle of slope the error values appear to dramatically increase. For smaller angles of slope there appears to be little or no relationship. The possible baseline error is also clearly evident in Fig. 4.

#### 3.2. Error propagation

To demonstrate the propagation of error to slope, values of slope directly derived from each DEM were compared to the mean values of the slope realisations (Fig. 5b). Subtracting one dataset

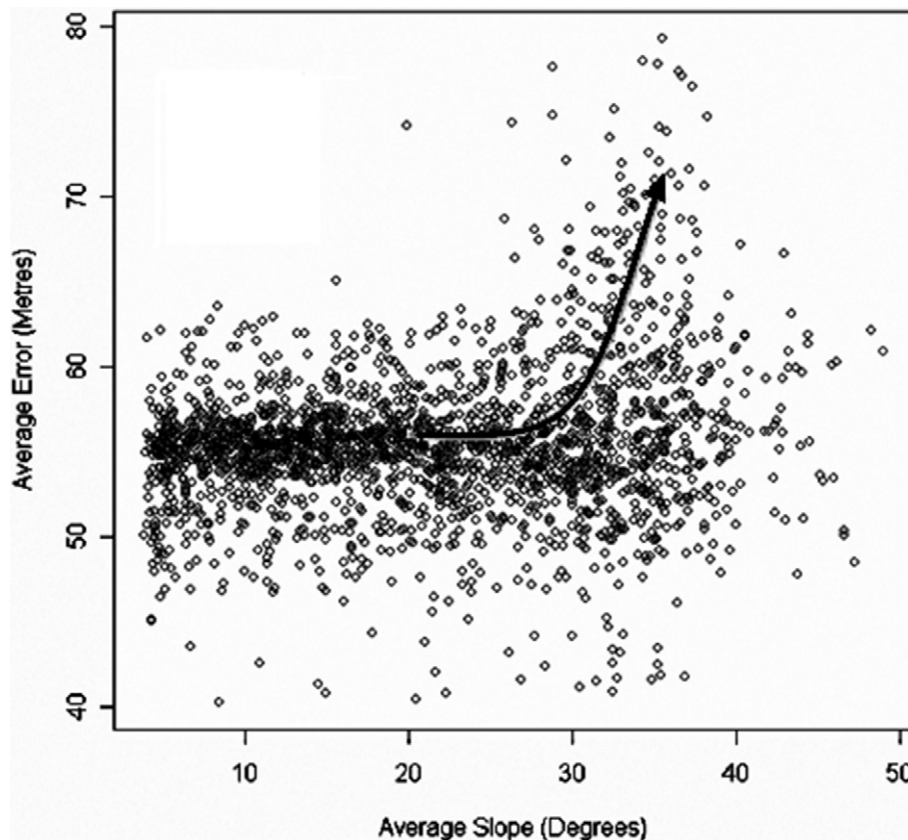


Fig. 4. Scatter plot showing the correlation of mean error and mean slope at every grid cell for the 100 realisations (12.5 m DEM).

from the other provided an indication of how slope estimation would differ were it derived directly from the DEM or from the multiple equally likely realisations. The main region of slope dis-

crepancy was adjacent to the peak in mean elevation error (Fig. 5a). A topographic layer has been overlain to allow easier comparison of the two mean output maps.

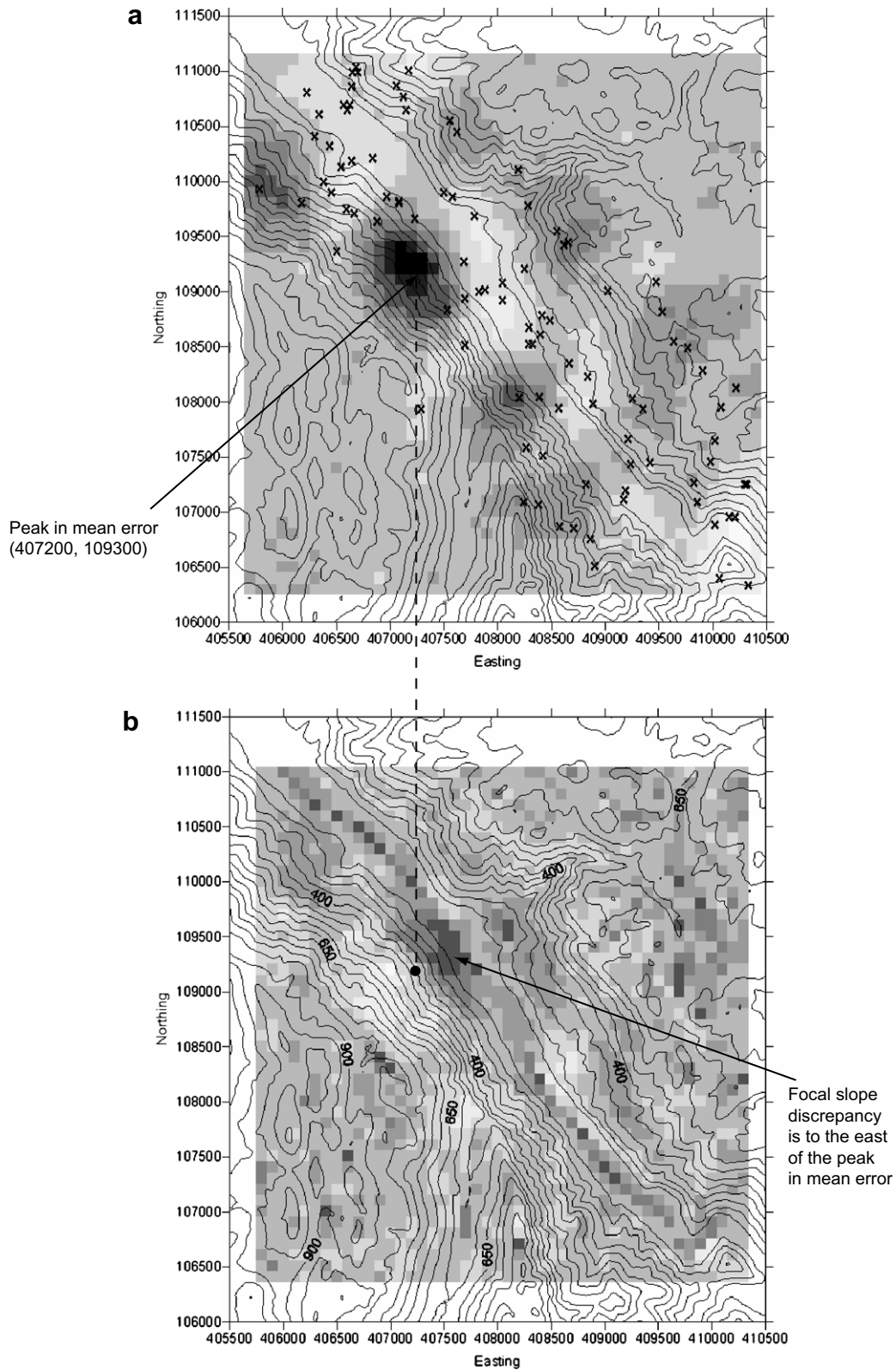


Fig. 5. Maps from GSTAT for 12.5 m DEM showing the location of (a) mean error in elevation and (b) propagation of error to slope. Crosses mark sample site locations and elevations on contours are in metres.

The likelihood of any cell being classified as ‘susceptible’ to landslide was calculated. This was used to demonstrate the propagation of error to a landslide susceptibility model and provided a measure of certainty about a decision such as landslide classification based on elevation error. A probability of 1 was given to those cells certain to be classified as ‘susceptible’ and zero to those certainly ‘not susceptible’, thus the cells most uncertain were about the point of inflection at 0.5. These regions of greatest uncertainty bordered the river valley in both DEMs (Fig. 6). Areas to the extreme NE and SW are omitted from analysis as there are sparse sample data in these areas (see Fig. 5a). Overall the test DEMs showed very similar spatial probability distributions.

The fuzzy classification resulted in a three-tier distinction between ‘susceptible’, ‘not susceptible’ and ‘undecided’. There were significant differences in the allocation of susceptibility areas from the two DEMs (Fig. 7). In particular the variation was evident in the valley bottom section (where highlighted). Fig. 8 summarises the consequences of landslide classification based exclusively on either the 12.5 m or the 25 m DEM. The coarser resolution DEM was more uncertain about classifying a cell as susceptible or non-susceptible to landslide. In comparison with the 25 m DEM, the 12.5 m DEM classified a greater number of cells as ‘susceptible’ or ‘non susceptible’.

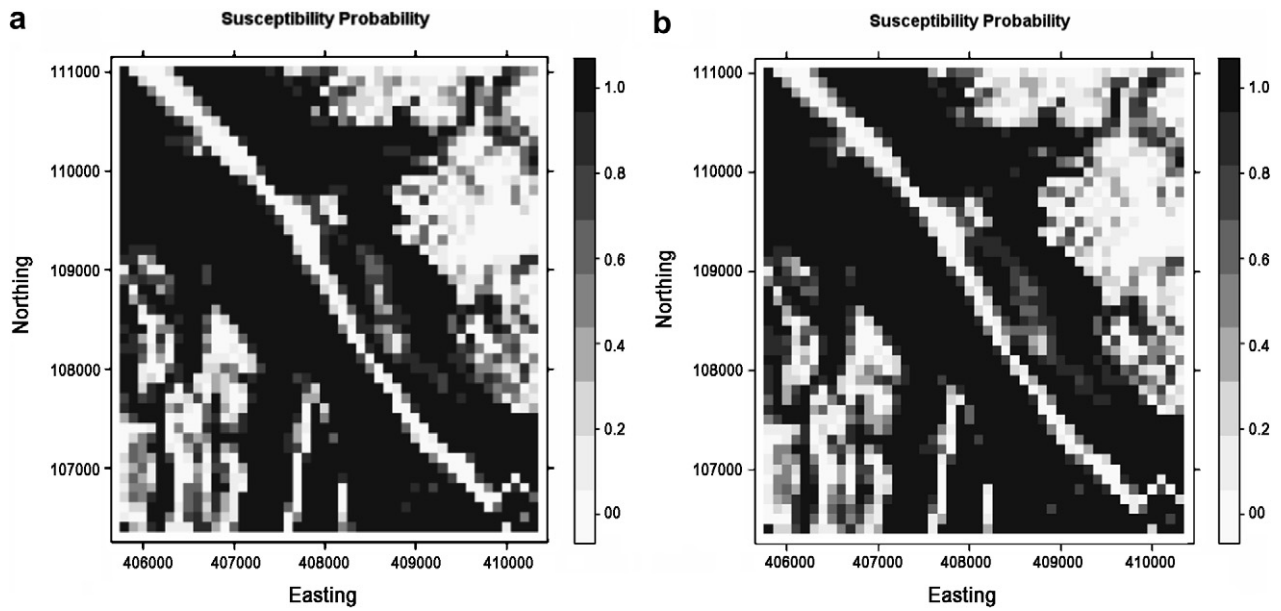


Fig. 6. The probability of a ‘susceptible’ classification, (a) 12.5 m DEM and (b) 25 m DEM.

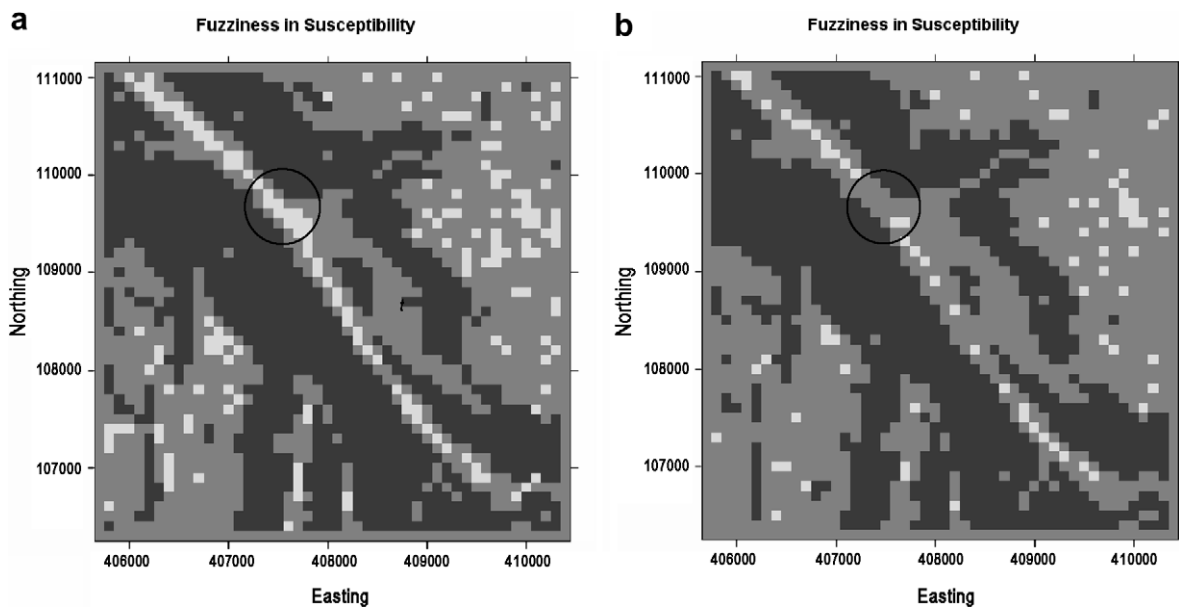


Fig. 7. Fuzziness in defining cells as susceptible: (a) 12.5 m DEM; (b) 25 m DEM.; darkest cells = ‘susceptible’, lightest cells = ‘not susceptible’, other = ‘undecided’.



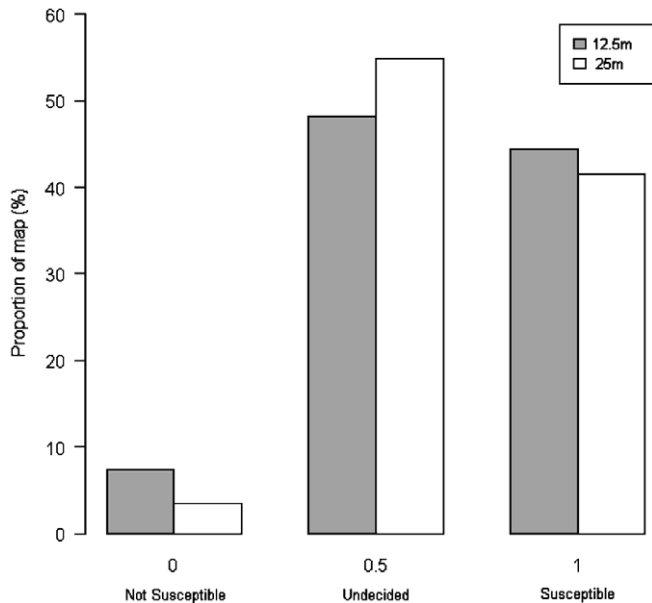


Fig. 8. Barplot produced in GSTAT showing the percentage of cells in each of the three classes.

## 4. Discussion

### 4.1. Factors contributing to error and its propagation

The application of the prototype to a case study example has demonstrated that error is spatially variable. Furthermore, the mode of formation was different for each DEM so a similar spatial distribution of error may support the hypothesis of a terrain-dependent error causal factor (e.g. Liu & Jezek, 1999). In terms of error in the slope derivative the main region of slope discrepancy was adjacent to the peak in mean elevation error (Fig. 5), thus supporting the hypothesis of error propagation and the previous work of authors such as Murillo and Hunter (1997). However to make any inferences on the contribution of terrain to error in the DEM it would be necessary to investigate the uncertainty inherent in the slope estimation. Many different algorithms are commonly used to generate aspect and gradient from gridded DEMs and these have differences in accuracy (Skidmore, 1989; Zhou & Liu, 2004). For example Horn's third-order finite difference method (Horn, 1981) uses all surrounding cells for analysis and could be better suited to the rugged terrain studied here. Other methods could include estimation of slope via cubic splines or weighted local regression. One reason that these alternative approaches to slope estimation may be useful is that they all apply a greater degree of smoothing than the basic method of slope estimation used here.

An estimate of slope is based on differences of height estimates, and in the situation where measurement errors are not correlated, the resultant error variances for estimates of difference in elevation equal the sum of individual errors in estimates of elevation. Thus, errors in slope can be notably larger than the errors in elevation estimation. Even in cases where measurement errors are positively correlated, error in estimate of height difference is still likely to be greater than individual errors of height estimation. Methods such as Horn's and the others mentioned above average a number of difference estimates, rather than just relying on one basic estimate in each direction. This averaging process reduces the variance in the final estimate, and to some extent counters the problem of higher error variance in slope estimation.

The cost of using either DEM for landslide susceptibility modelling was tested. 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. A method for a probabilistic assessment was also applied and demonstrated the certainty in a decision such as landslide classification. However the resultant maps were not adequate to illustrate the differences between the two datasets. In this respect the fuzzy framework was superior, its only disadvantage owing to it being a more complicated technique for the user to understand.

### 4.2. Feasibility and significance of the methodology

The methods developed allowed the investigation of uncertainty, as supported by the results. GSTAT provided the means for geostatistical simulation. Additional code was written to make the procedure as automated and as objective from user input as possible (Fig. 2). System defaults can be modified according to the familiarity of the user to general principles.

The use of GPL software enables any user to run and adapt the source code. The existing code for stages A and B can be used for any GIS application where DEM data and higher accuracy data are available. Following minor modifications, the work could be integrated into fitness for use assessment, risk management studies and cost benefit analyses etc. The studies of Hunter and De Bruin (2006), Agumya and Hunter (2002) and Aerts et al. (2003) provide interesting work on some of these themes. The implications and significance of such a generic approach are wide reaching as the hardware, software and knowledge required are as non-specialist as possible. Thus even a novice DEM user is provided with the tool to appreciate uncertainty in their data and further analyses. However 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.

### 4.3. Further work

A number of improvements to the general methodology are possible (e.g. slope algorithm) and due to the modular nature of the R code this would be simple to implement. The feasibility of an automatic variogram fitting procedure could be investigated. This would make the procedure more objective and also would result in a more statistically sound fitting (as each experimental variogram point can be weighted according to its own statistical uncertainty) (Pardo-Iguzquiza, 1999).

Finally an interesting area of further study would be to ask a variety of DEMs users to adopt the tool and survey their responses. Particular information should be gathered on the technique's ease of use and its ability to communicate the uncertainty in the relevant dataset.

## 5. Conclusion

The absence of simple yet robust uncertainty analysis procedures has perhaps led to the indifference or incapacity of the majority of DEM users to consider its influence. However the consequences of error can be severe as they can propagate from the DEM to decision outcomes, as demonstrated above. The methodology developed here provides an instrument for error quantification, demonstration of propagation, and visualisation that is a simplification on existing techniques and moves towards full automation. Publicly available software has been used to facilitate a universally distributable and pliable tool. Due to the modular nature of the methodology amendments and improvements can easily be made. Some interesting areas of further study have also been raised by this work.

## 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. This paper presents a revised and extended version of this work. LiDAR data were obtained by NERC ARSF grant MC04/09 and Royal Society Grant 28/D/23901/SM awarded to Dickson Cunningham at the University of Leicester. DEM data were used courtesy Surveying and Mapping Authority of Slovenia. The principle author would like to gratefully acknowledge Andrej Gosar at the Environment Agency of the Republic of Slovenia and Marko Komac at the Geological Survey of Slovenia. The paper also benefited from the input of Gary Hunter, Andrew Lovett and the comments of anonymous referees. Finally thanks to Kerry Allen at the University of Leicester, and Andrew Harwood at the University of East Anglia, for help with graphics.

## References

- Aerts, J. C. J. H., Goodchild, M. F., & Heuvelink, G. B. M. (2003). Accounting for uncertainty in optimization with spatial decision support systems. *Transactions in GIS*, 7, 211–230.
- Aguilar, F. J., Aguilar, M. A., Aguera, F., & 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.
- Aguilar, F. J., Aguilar, M. A., & Aguera, F. (2007). Accuracy assessment of digital elevation models using a non-parametric approach. *International Journal of Geographical Information Science*, 21, 667–686.
- Agumya, A., & Hunter, G. J. (2002). Responding to the consequences of uncertainty in geographical data. *International Journal of Geographical Information Science*, 16, 405–417.
- Burrough, P. A. (2002). GIS and geostatistics for environmental modelling. In W. Shi, P. F. Fisher, & M. F. Goodchild (Eds.), *Spatial data quality* (pp. 18–34). London: Taylor and Francis.
- Burrough, P. A., & McDonnell, R. A. (1998). *Principles of geographical information systems*. Oxford: Oxford University Press.
- Carlisle, B. H. (2005). Modelling the spatial distribution of DEM error. *Transactions in GIS*, 9, 521–540.
- Couclelis, H. (2003). The certainty of uncertainty: GIS and the limits of geographic knowledge. *Transactions in GIS*, 7, 165–175.
- CRAN (2007). *The comprehensive R archive network*. <http://cran.r-project.org/>.
- Cunningham, D., Grebby, S., Tansey, K., Gosar, A., & Kastelic, V. (2006). Application of airborne LiDAR to mapping seismogenic faults in forested mountainous terrain, southeastern Alps, Slovenia. *Geophysical Research Letters*, 33, 1–5.
- Davis, T. J., & Keller, C. P. (1997). Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction. *International Journal of Geographical Information Science*, 11, 409–434.
- Fisher, P. (1998). Improved modeling of elevation error with geostatistics. *Geoinformatica*, 2, 215–233.
- Fisher, P. F., & Tate, N. J. (2006). Causes and consequences of error in digital elevation models. *Progress in Physical Geography*, 30, 467–489.
- Fleming, M. D., & Hoffer, R. M. (1979). Machine processing of Landsat MSS data and LARS Technical Report 062879. Laboratory for applications of remote sensing, Purdue University, West Lafayette, IN, USA.
- Heuvelink, G. B. M. (1998). *Error propagation in environmental modelling with GIS. Research monographs in GIS series*. London: Taylor and Francis.
- Holmes, K. W., Chadwick, O. A., & 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.
- Horn, B. K. P. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, 69, 14–47.
- Huisling, E. J., & Gomes Perreira, L. M. (1998). Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53, 245–261.
- Hunter, G. J., & De Bruin, S. (2006). A case study in the use of risk management to assess decision quality. In R. Devillers & R. Jeansoulin (Eds.), *Fundamentals of spatial data quality. Geographical information systems series* (pp. 271–282). London: ITSE Ltd.
- Hunter, G. J., & Goodchild, M. F. (1995). Dealing with error in spatial databases: A simple case study. *Photogrammetric Engineering and Remote Sensing*, 61, 529–537.
- Hunter, G. J., & Goodchild, M. F. (1997). Modelling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis*, 29, 35–49.
- Jones, K. H. (1998). A comparison of algorithms used to compute hill slope as a property of the DEM. *Computers & Geosciences*, 24, 315–323.
- Journel, A. G. (1996). Modelling uncertainty and spatial dependence: Stochastic imaging. *International Journal of Geographical Information Systems*, 10, 517–522.
- Kidner, D. B. (2003). Higher-order interpolation of regular grid digital elevation models. *International Journal of Remote Sensing*, 24, 2981–2987.
- Kyriakidis, P. C., Shortridge, A. M., & 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., & Jezek, K. C. (1999). Investigating DEM error patterns by directional variograms and Fourier analysis. *Geographical Analysis*, 31, 249–266.
- Murillo, M. L., & Hunter, G. J. (1997). Assessing uncertainty due to elevation error in a landslide susceptibility model. *Transactions in GIS*, 2, 289–298.
- Pardo-Iguzquiza, E. (1999). VARFIT: A fortran-77 program for fitting variogram models by weighted least squares. *Computers & Geosciences*, 25, 251–261.
- Pebesma, E. J. (1999). *GSTAT user's manual*. <http://www.gstat.org/manual/gstat.html>.
- Pebesma, E. J., & Wesseling, C. G. (1998). GSTAT: A program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, 24, 17–31.
- Raaflaub, L. D., & Collins, M. J. (2006). The effect of error in gridded digital elevation models on the estimation of topographic parameters. *Environmental Modelling and Software*, 21, 710–732.
- Rebiasz, B. (2007). Fuzziness and randomness in investment project risk appraisal. *Computers & Operations Research*, 34, 199–210.
- Rees, W. G. (2000). The accuracy of digital elevation models interpolated to higher resolutions. *International Journal of Remote Sensing*, 21, 7–20.
- Skidmore, A. K. (1989). A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. *International Journal of Geographical Information Systems*, 3, 323–334.
- USGS (1997). *National mapping program standards: Standards for digital elevation models*. <http://rockyweb.cr.usgs.gov/nmpstds/demstds.html>.
- Veregin, H. (1997). The effects of vertical error in digital elevation models on the determination of flow-path direction. *Cartography and Geographical Information Systems*, 24, 67–79.
- Warren, S. D., Hohmann, M. G., Auerwald, K., & Mitasova, H. (2004). An evaluation of methods to determine slope using digital elevation data. *Catena*, 58, 215–233.
- Wechsler, S. P. (2003). Perceptions of digital elevation model uncertainty by DEM users. *URISA Journal*, 15, 57–64.
- Wehr, A., & Lohr, U. (1999). Airborne laser scanning – An introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 68–82.
- Wise, S. (2000). Assessing the quality for hydrological applications of digital elevation models derived from contours. *Hydrological Processes*, 14, 1909–1929.
- Zadeh, L. A. (1999). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100, 9–34.
- Zhang, J., & Goodchild, M. J. (2002). *Uncertainty in geographical information. Research monographs in GIS series*. London: Taylor and Francis.
- Zhou, Q., & Liu, X. (2004). Analysis of errors of derived slope and aspect related to DEM data properties. *Computers & Geosciences*, 30, 369–378.
- Zhou, Q., Liu, X., & Sun, Y. (2006). Terrain complexity and uncertainties in grid-based digital terrain analysis. *International Journal of Geographical Information Science*, 10, 1137–1147.