ERROR PROPAGATION ANALYSIS IN SLOPE ESTIMATION BY MEANS OF DIGITAL ELEVATION MODELS

Angel M. Felicísimo
INDUROT - Instituto de Recursos Naturales y Ordenación del Territorio
Universidad de Oviedo
E-33071 Oviedo Asturias (España)
E-mail : feli@gea.etsimo.uniovi.es

Abstract

This paper carries out an error propagation analysis in the creation process of the digital slope model (DSM) from the digital elevation model (DEM). This analysis uses the Monte Carlo methods for the error parameters calculation in the DSMs from the root mean square error values in the DEMs. Results allow us to establish the incertainty rate associated to the value calculated in accordance with the original DEM quality.

1 Introduction

The **digital terrain models** (DTMs) have been used for the extraction of different geomorphological and topographic features depending on altitude and on its spatial distribution and variation. The slope is enhaced as a topographic feature of great importance in the geomorphological processes understanding, many of which depend closely and directly on its value.

In the Geographical Information Systems (GIS), the slope values are estimated from the digital elevation model. Local methods are usually used for this purpose, that is to say, they are based on the information about a limited area around the problem point. The magnitude of the error inherent in the altitude data affects the results of these methods. This error, motivated by different causes, can be considered random and Gaussian, with a null mean value and a standard deviation equivalent to root mean square error (see section 2). The DEMs error has already been presented by some papers which, although rare, are sufficient for an adequate featuring thereof [1,2,3]. However, the influence of this error on the digital derived models (error propagation process) has received very little attention since the literature concerning this topic is sparse. Even in recent papers on Geographical Information Systems in general [4] with sections referring to digital terrain models in particular [5], critical analyses on error propagation are dramatically absent.

Error reduction in the DEMs can be only carried out to a certain extent. Due to this circumstance, it is very convenient to know very well the error values in order to estimate their influence on derived applications. This paper is aimed at analysing error propagation in the building process of the digital slope model (DSM) from the DEM. Numeric simulation methods are used there to allowing us to know the error scope accurately in the generation process of the DSM. These methods, generally known as Monte Carlo methods, have been widely used when analysing processes of complex analytic solution [6] and they are specially suitable for the foregoing analysis.

2 Error in the digital elevation model

Error quantification on a matrix digital elevation model is normally carried out comparing a set of known altitude values with the corresponding homologous points of the model and using conventional statistical tests. This fashion, the altitude error on a point i, δ_i, is defined as the difference between the altitude of the point in the model, z_i, and the "real" or reference altitude, z'_i, taken from an accurate data source: $\delta_i = z_i - z'_i$.

Error statisticals are calculated through an adequate number of homologous points suitably distributed over the chart. The mean error or mean deviation, $\overline{\delta}$, for this set of data is defined as $\overline{\delta} = \frac{1}{n} \cdot \sum_{i=1}^{n} \delta_i$, where n is the number of compared points. The error standard deviation, s_δ, is calculated through the expression:

$$s_\delta = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(\delta_{z_i} - \overline{\delta} \right)^2}$$ .Error quantification in the DEM is normally carried out by calculating the root mean square error or *RMSE* [7] through the expression: $RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \delta_i^2}$.

Mean error and error standard deviation values define the error distribution function in the DEM, a Gaussian distribution $N(\bar{\delta}, s_d)$. As formerly indicated, if the error of the altitude values is random, the mean error value is zero; in this case, the error standard deviation is similar to the mean quadratic error and the error function is $N(0, RMSE)$.

3 Building methods for the digital slope model

The slope calculation method in the DTMs has been almost regularly standardized. Its expression is based on the first altitude derivatives in the spatial distribution function $z=\zeta(x,y)$, generally represented as a_{10} y a_{01} [8]:

$$\gamma = \arctan\sqrt{\left(\delta_z/\delta_x\right)^2 + \left(\delta_z/\delta_y\right)^2} = \arctan\sqrt{a_{10}^2 + a_{01}^2}$$

For a_{10} and a_{01} calculation, several expressions have been presented although, in practice, the most usual in the GIS uses the 8 closest neighbours to the problem point and applies the Sobel operator to them as a gradient estimator. The expressions for a point placed in row i, and column j, are:

$$a_{10} = \left(z_{i-1,j+1} + 2z_{i,j+1} + z_{i+1,j+1}\right) - \left(z_{i-1,j-1} + 2z_{i,j-1} + z_{i+1,j-1}\right)/8d$$

$$a_{01} = \left(z_{i-1,j-1} + 2z_{i-1,j} + z_{i-1,j+1}\right) - \left(z_{i+1,j-1} + 2z_{i+1,j} + z_{i+1,j+1}\right)/8d$$

where d stands for the existing distance between rows and columns expressed in the same units as z.

4 Error analysis methods

Monte Carlo methods are usually used in problems of complex analytic solution where the obtention of accurate results is, though theoretically possible, excessively difficult. In these cases, Monte Carlo methods approach the problem from a different point of view, using statistical simulations which lead to useful results to solve the problem. Simulations are carried out by random number generators which, properly used, produce statistical distributions close to the searched solution arbitrarily accurately. This accuracy can be raised through the increase of simulations, aimed at achieving a balance between the computing cost and the error associated to statistical simulation.

In the present case, the error propagation problem is specially suitable for the use of this type of methods. The process consists of the following steps:

[1] DEM creation, **M**, with identical altitude values z_i, therefore through simulation of a plane terrain:

$$\forall z_i \in \mathbf{M} \mid z_i = k$$

[2] Generation of an error value $z_i \in \mathbf{M}$ for each point d_i in accordance with a Gaussian distribution $N(0, RMSE)$. Through this value, the altitude of the point is actualized in order to introduce the error factor:

$$\forall z_i \in \mathbf{M} \mid z_i \leftarrow z_{i+\delta i}$$

[3] Slope value calculation, p_i, for each point $z_i \in \mathbf{M}$ through the algorithm shown in section 3 which is symbolized as **S** in the foregoing expression. This produces a digital slope model **S** consisting of the same number of elements as **M**:

$$\forall p_i \in \mathbf{S} \mid p_i \leftarrow \mathbf{S}(z_i \in \mathbf{M})$$

[4] The distribution of values for p_i in **S** allows us to know the *RMSE* value influence on the creation process of the DSM. Conclusions on the error importance and, if any, the advantages or disadvantages of some or other algorithms are drawn from the different distribution analysis calculated for different combinations of *RMSE* values.

The former process has been carried out in order to simulate a DEM with 100,000 data. This amount has been considered adequate to create balanced and representative statistical distributions for different error values. Three distributions have been created corresponding to error values from 5, 10 and 20 per cent respectively for the algorithms shown in section 3 –corresponding to the Sobel operator–. The s_d values are shown as a percentage of the distance between data, that is to say, a 5% error means a *RMSE*=5 m for a d=100 m.

5 Results of simulations

In Figure 1 three distributions representing the results of the corresponding numeric simulations are shown. It is evident that the resulting distribution for a 5% *RMSE* is that one showing a lesser dispersion and, thus, a more reduced mean error. As the *RMSE* of the DEM is increased, the resulting slopes show a growing dispersion and greater error mean values.

However, the most remarkable effect is the fact that in this case distributions have no mean zero, that is, *on building the digital slope model from a DEM, we have underestimated the low slopes proportionally to the magnitude of the error in the DEM*. This underestimation is evident in the mean values of the distributions shown in Figure 1: for a 5% *RMSE*, the mean error is 1.4°, for 10% is 2.9°, and 5.8° for a 20% *RMSE*. These values must be interpreted as deviations of the expected slope value, which would be 0° for this simulation.
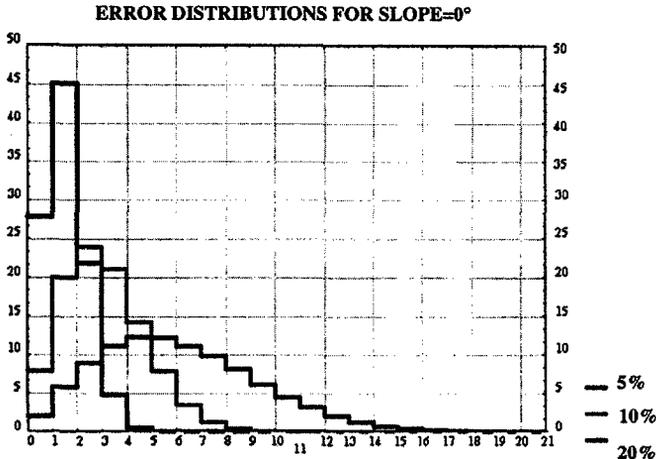
ERROR DISTRIBUTIONS FOR SLOPE=0°

Figure 1: Error distributions resulting from simulations for different *RMSE* values and an initial mean slope of 0°

There are other statisticals of interest which can be taken from the observed frequencies in the simulations. For instance, from the accumulated frequencies we can gather that for a plane DEM with a 10% *RMSE*, 50% of values will have an error equal or greater than 2°, 25% equal or greater than 3.5°, and 5% equal or greater than 5°. For a DEM with a 20% *RMSE* the results are obviously worse: 50% of values will have an error greater than 5°, 25% greater than 7° and 5% greater than 11°. In the most favourable case of those under study, errors are much lesser for a 5% *RMSE*: 95% of the values will be affected by an error lesser than 2°.

The former case has been carried out by means of a plane DEM, where all altitudes are originally similar and have been added a random error. The analysis has been repeated with a DEM whose values have been chosen so that the slope would be 10°. Figure 2 shows the slope error values deducible from the simulation results. This graph has been drawn from the absolute difference between the DSM slope and the slope mean value in the initial model (10°).
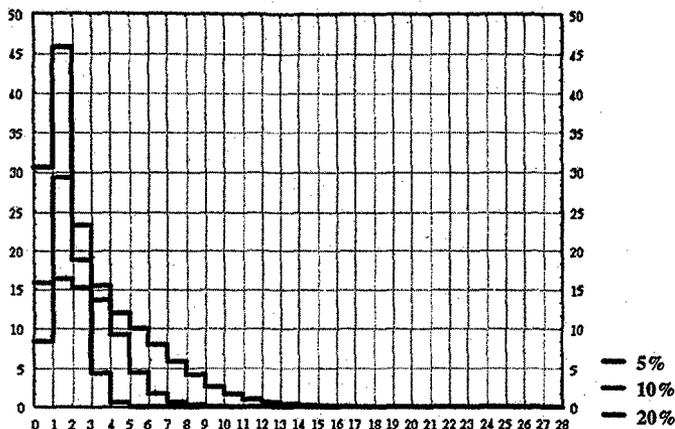
ERROR DISTRIBUTIONS FOR SLOPE = 10°

Figure 2: Error distributions resulting from simulations for different *RMSE* values and an initial mean slope of 10°.

As in the former case, we can estimate from the accumulated frequencies that, for a 10% *RMSE*, 50% of the values will have an error equal or greater than 2°, 25% equal or greater than 3.5° and 5% equal or greater than 5°. For a DEM with a 20% *RMSE*, 50% of the values will have an error greater than 4°, 25% greater than 6° and 5% greater than 9.5°. For a 5% *RMSE*, 95% of the values will be affected by an error lesser than 3°.

6   Conclusions and discussion

The following conclusions are drawn from the previous results:

- the error affecting the digital elevation models –expressed as root mean square error, *RMSE*–, also affects the derived models of the DEM, and in this case, the digital slope models or DSM;

- the DSMs show slope values whose statistical distribution presents a mean value greater than the real one and, therefore, they underestimate real terrain slopes at a different rate. The greater this underestimation is, the worse the original DEM quality will be;

- the DSM data will be proportionally erroneous in frequency and magnitude to the *RMSE* of the DEM and to be taken into account. The mean slope of the terrain represented by the DEM seems to operate as to diminish the error in stronger relief areas.

As a general consequence from the former conclusions, it should be enhanced that slope chart production from the DEMs should take into account the error distribution thereof. It would be nonsense, for instance, to produce charts where slopes are classed in excesivelly narrow groups for the expected accuracy. From a different point of view, when approaching a study which requires the building of a DEM, the minimum quality level should be established in order to achieve the aims of the work successfully.

In addition to the former conclusions, we should enhance the suitability of taking error measures on the DEMs, since without them, the quality of the results achieved through the geographical information systems cannot be evaluated.

Simulations carried out for this paper have led to a DEM corresponding to an area of slope 0 and, to another one of mean slope 10. These models are artificial sets and have been developed in order to study the problem from its simplest points. The error analysis on a real terrain, generally much more complicated, requires more completed analyses, where different error and mean slope combinations are studied. Nevertheless, this paper remains to be developed in future and we suggest that the geographical information systems should end up introducing new tools which simplify these quality tests.

Bibliografía

[1] Hannah, M.J., 1981. Error detection and correction in Digital Terrain Models. Photogrammetric Engineering and Remote Sensing, 47(1), pp. 63-69.

[2] Veregin, H., 1989. A taxonomy of error in spatial databases. NCGIA, National Center for Geographic Information and Analysis. Technical Paper 89-12, 115 p.

[3] Felicísimo, A.M., 1994. A parametric statistical method for error detection in digital elevation models. ISPRS Journal of Photogrammetry and Remote Sensing, 49(4), pp. 29-33.

[4] Maguire, D.J., Goodchild, M.F., Rhind, D.W. Eds., 1991. Geographical Information Systems. Principles and applications, 2 vol., 649+447 p. Longman Scientific & Technical, London.

[5] Weibel, R., Heller, M., 1991. Digital terrain modelling. In Maguire et al., 1991, vol. 1, pp. 269-297.

[6] Doll, Jim., 1994. Monte Carlo methods in chemistry. IEEE Computational Science & Engineering, 1(1), pp 22-32.

[7] Merchant, D.C., 1987. Spatial accuracy specification for large scale topographic maps. Photogrammetric Engineering and Remote Sensing, 53(7), pp. 958-961.

[8] Franklin, Steven E., 1987. Geomorphometric processing of digital elevation models. Computers & Geosciences, 13(6), pp. 603-609.

[9] James, Mike, 1988. Pattern recognition. John Wiley & Sons, 144 p. New York.