

Extended version of the work published in PROCEEDINGS ScanGIS 1997

## **On the improving of elevation accuracy of Digital Elevation Models: a comparison of some error detection procedures**

CARLOS LÓPEZ

Centro de Cálculo, Facultad de Ingeniería (11), Universidad de la República

Julio Herrera y Reissig 565, Montevideo, URUGUAY

internet:carlos@fing.edu.uy, <http://www.fing.edu.uy/~carlos>

**Abstract:** The widespread availability of powerful desktop computers, easy-to-use software tools and geographic datasets have raised the quality problem of input data to be a crucial one. Even though accuracy has been a concern in every serious application, there are no general tools for its improvement. Some particular ones exist however, and we are reporting here results for a particular case of quantitative raster data: Digital Elevation Models (DEM). We tested two procedures designed to detect anomalous values (also named *gross errors*, *outliers* or *blunders*) in DEM, but valid also for other quantitative raster datasets.

A DEM with elevations varying from 181 to 1044 m derived from SPOT data has been used as a contaminated sample, while a manually derived DEM obtained from aerial photogrammetry has been regarded as the ground truth. That allows a direct performance comparison for the methods with real errors.

We assumed that once an outlier location is suggested, a "better" value can be measured or obtained through some methodology. The options are different depending upon the user (DEM producers might go to the original data and make another reading, while end users might only interpolate). In this experiment we considered both choices.

Preliminary results show that for the available dataset, the accuracy might be improved to some extent with very little effort. Effort is defined here as the percentage of points suggested by the methodology in relation with its total number: thus 100 per cent effort implies that all points have been checked.

The method proposed by López (1997) gave poor results, because it has been designed for errors with low spatial auto correlation (which is not the case here). A modified version has been designed and compared also against the method suggested by Felicísimo (1994).

The three procedures can be applied both for error detection during the DEM generation and by end users, and they might be of use for other quantitative raster data. The choice of the best methodology is different depending on the effort involved. The conclusions have been derived for a photogrammetrically obtained DEM; other production procedures might lead to different results.

**KEYWORDS:** DEM, quality control, blunder location, gross error location, accuracy

### **1. Introduction**

Geographic Information Systems (GIS) is one of the fastest growing markets in software today (Anon 1994). That implies that more people have access to proper tools, and then are able to manipulate and produce data. Data availability will be assured in the future, through the operation of the so called *Clearinghouses*, which will distribute existing datasets to government, industry and the general public (Nebert 1995, 1996).

The combination of widespread data and ready made, easy to use software raises some critical points. John (1993) stated that "...very wrong answers can be derived using perfectly logical GIS analysis techniques, if the users are not aware of the particular peculiarities of data...". Data quality is emerging as one of the most important issues in GIS technology for the next years. Its management requires methods to describe, visualize and measure it properly (see Hunter *et al.* 1996). Standards for describe the quality are presently under development.

Thapa *et al.* (1992) remarked that when setting up a GIS, most of the costs (maybe up to 80 per cent) are related to acquiring and/or collect data. Once the dataset is obtained further efforts to improve accuracy should be as effective as possible. This paper reports some results on that subject. We will concentrate here on Digital Elevation Models (DEM). We will not consider errors in the intermediate steps in the process of DEM generation, but we will concentrate on the errors in the final product. According to Thapa *et al.* (1992) errors can be classified into three types: (1) gross errors and blunders, (2) systematic errors and (3) random errors. Gross errors and blunders are caused by carelessness or inattention of the observer in using equipment, reading scales or writing down readings, etc. They can also be caused by malfunctioning of the equipment. Observations affected by this kind of errors are useless, and should be eliminated. From a statistical point of view they cannot be considered as belonging to the same population as the other observations. This paper deals with such type of errors. Systematic errors occur in accordance with some deterministic system which, if known, may be represented by some functional relationship. In a statistical sense, systematic errors introduce bias in the observations, which need not to be constant in space. They might arise from improper orientation of the stereomodel or by neglecting lens distortion in photogrammetric procedures. Unlike gross errors, they cannot be detected or eliminated by repeated observations (the errors may be *precise*, but they will not be *accurate*). After removal of gross and systematic errors, differences still exist due to random errors. They cannot be removed by repeated

observation, and they cannot be modeled with a deterministic relationship. If sufficient observations are taken, random errors possess the following characteristics: a) positive and negative errors occur with almost the same frequency b) small errors occur more often than large errors and c) large errors rarely occur. They are usually assumed to belong to a gaussian distribution, with parameters valid for the whole DEM, and with negligible auto correlation in space.

Östman (1987a) pointed out the fact that there exists no unique criteria or single measure for the "quality" of a DEM. He suggested that one should at least, consider accuracy in elevation, slope and also curvature. However, accuracy reports in terms of slope are very unusual. An exception can be found in the work of Giles *et al.* (1996) who compared a 20 m resolution DEM derived from SPOT images with field measurements in terms of slope. They recognized that the elevation error might have two components at different scales. To filter out the small scale error they simply applied a 3 by 3 median filter and to remove the larger errors they used a 11 by 11 window, with a different filter. They claim that such filtering improve to some extent the accuracy in slope, without significantly degrading the accuracy in elevation. This has also been reported by Östman (1987a). He found that the RMS error in elevation decreases with decreasing grid size (as expected) but the effect in RMS error in slope is very limited. Förstner (1983) gave theoretical arguments for this fact.

Accuracy of photogrammetrically sampled DEM depends on the data sources and the procedures involved. It has been considered an important problem which led to collective efforts like the one summarized by Torlegård *et al.* (1986). They reported the results of DEM derived independently by a number of organizations working on the same set of aerial photographs. Six different terrain types have been chosen, ranging from smooth terrain to steep and rugged mountains. They found that the errors of the elevations in photogrammetrically measured DEM consist to a large extent of systematic components. Regarding gross error location, they applied a "rule of thumb" based on recursive filtering using a 5 by 5 window, and declared that everything located is an error. They conclude that the number of those so defined gross errors typically varies between 0 and 3 per cent. They noticed that error size is independent of terrain type and that errors are more frequent in difficult terrain.

A similar (deterministic) approach was used in an early paper by Hannah (1981), who detects errors by applying constraints to the slopes and to the changes in slope at each point. Felicísimo (1994) analyzed the

differences between the elevation and an interpolated value from the neighbors. Assuming gaussian distribution of the errors, he analyzed the differences by means of a standard Student t test.

Day *et al.* (1988) tested three methods for the generation of DEM based on SPOT data. The three results were compared with a very carefully, manually digitized 30 m grid DEM, in terms of elevation differences. Even though the goal of the work was to compare the operational behavior of the algorithms, they do not propose a solution for the location of the errors. The distribution function of the absolute size of such errors is also presented for each method. They also reported how many checkpoints lie outside the limit  $|\text{error}-\mu|>3\sigma$ .

Any method for locating the errors should make assumptions about size, location and spatial auto correlation. Bethel *et al.* (1984) proposed the method of maximum chi-squared ratio for on line quality control, and tested the methodology using uncorrelated in space, spike-like blunders of no more than 10 feet (about 3 m). López (1997) used two error models: one uncorrelated in space (spike-like blunders) and another weakly correlated (pyramid-like).

In the field of Image Processing the term *salt-and-pepper* has been coined for weakly auto correlated errors. They are routinely corrected using filters. The most popular and simple one is the median filter (Mitra *et al.* 1994) but it has the fundamental inconvenience that it smoothes out all the DEM; current efforts are directed towards a division of the problem: to separate error detection from error correction, and to use state variables for error detection (Abreu *et al.* 1996).

We will not discuss here the methods for obtaining the DEM itself. There are well established procedures based on photogrammetry, GPS, etc. However, if the equipment or the methods are at their limit today, there will be little chances to improve the final results by merely pointing out some locations likely to be in error. Fortunately, this is not the case. Ackermann (1995) points out that the trend in DEM production is towards a move from interpolation to approximation, because the new generation equipment is able to produce many elevation values, but possibly with less accuracy than traditional equipment. The surface is *approximated* using many points, instead of being *interpolated* from few, very carefully obtained values.

Summing up, accuracy is a concern for the data producer as well as for the end user. Accuracy is usually described using different statistics of the distribution of elevation error at some checkpoints.

This paper presents test results of some recently proposed methodologies for locating errors which can be applied both by the producer and the end user. The methods were tested in a DEM with real errors, and the results are presented. Also some guidelines for the error model for this particular case are presented.

The paper is organized in eight sections. Section 2 has a brief outline of both the already existing techniques and the modified technique. Section 3 describes the data and summarizes some of its statistics. In section 4 the performance of the three methods is compared for the test DEM under a *perfect inspector* assumption, while section 5 covers the *imperfect* case. Finally, section 6 contains a discussion and section 7 is devoted to conclusions, where the results and proposed future work are discussed. Acknowledgments are included under heading 8 and References appear at the end.

## 2. The error detection procedures in brief

For the sake of completeness we will describe briefly the methods of Felicísimo (1994) and López (1997), and a modification of the latter.

### 2.1 The method of Felicísimo (1994)

This is the simplest method available for this problem. Assuming that outliers are only locally correlated, the method analyzes the differences  $\delta_{i,j}$  between the elevation value  $z_{i,j}$  and an interpolated guess  $\hat{z}_{i,j}$  obtained from its immediate neighbors. Assuming that the difference has a Gaussian distribution with mean  $\bar{\delta}$  and standard deviation  $s_{\delta}$  (both obtained from the sample) a Student's t test can be applied to validate the hypothesis that  $\delta_{i,j}$  belongs to the population of deviations. Operationally, we analyze the statistics  $t_{i,j} = (\delta_{i,j} - \bar{\delta}) / s_{\delta}$  which can be regarded as a standardized deviation. Since the number of data points are usually large, we can assume a distribution  $t_{\alpha[\infty]}$  for  $t_{i,j}$ . For  $\alpha = 0.001$ , the statistical  $t_{\alpha[\infty]}$  has a value of 3.219 for a two-tail test, where the null hypothesis is  $H_0, \delta_{i,j} = \bar{\delta}$  and the alternative is  $H_1, \delta_{i,j} \neq \bar{\delta}$ .

The estimated height  $\hat{z}_{i,j}$  can be obtained by any weighting scheme and any definition of neighborhood. We used a best fit approximation with a biquadratic polynomial using the eight closest neighbors. Along the borders we assume a mirror symmetry, and in the corners we used a linear interpolation with the three closest values available. We point out as candidate to be in error any  $\delta_{ij}$  that makes  $|t_{i,j}| > 3.219$ . The author states that even though a significantly high value of  $t_{i,j}$  does not necessarily imply an error, it is an excellent alarm sign. We will analyze this topic later.

Once an error is located and corrected, both statistics  $\bar{\delta}$  and  $s_{\delta}$  change and new candidates appear. The method can be iterated and it might stop if no more "outlying" values remains. This is undesirable because we know that there still are errors in the dataset, so we proceed by lowering the limit 3.219 to 3.0 at least once. The new candidates once corrected modify the statistics, and new candidates with the limit 3.219 appear. Notice that this process is not equivalent to merely evaluate once  $t_{i,j}$ , and sort the values; the statistics  $\bar{\delta}$  and  $s_{\delta}$  are usually severely affected by gross errors, so  $t_{i,j}$ , should be re-evaluated after each step.

The method appears to be extremely simple and is parameter free (in the sense that the user need not to specify any parameter). In section 3 we will investigate if the test DEM fulfills the assumptions under which the Felicísimo's method can be applied (Gaussian distribution, low spatial auto correlation, etc.). Also the relationship of  $t_{i,j}$  and real errors (available in this experiment) will be presented.

## 2.2 *The method suggested by López (1997)*

The author proposed that any given raster dataset can be analyzed by means of subdividing it into elongated strips (figure 1). Each strip is assumed to have length  $n$  and width  $w$  ( $w \ll n$ ). The method considers the strip as a set of points in the  $R^W$  space. Each cross-section is represented by a point, where the original elevation values establish the  $w$  coordinates. The case of  $w=3$  is illustrated in the figure 2, where each point  $M_k$  represents a cross-section.

The error location procedure directly analyzes the cloud of points in  $R^W$ , disregarding any order among points. This is an important assumption, since the concept of spatial auto correlation loses completely all

significance in the cloud. Adjacent profiles (of length  $n$ ) need not to be in any special order, since they are coordinate axes in the space  $R^W$ .

The use of the cloud is common practice in statistics (Hadi 1992, 1994, Hawkins 1974, 1993a, 1993b), since the notion of "spatial auto correlation" and "precedence" is meaningless in most tabular data.

The procedure is based upon Principal Component Analysis (PCA), which attempts to find the direction  $e_1$  of the vector in  $R^W$  space which minimizes  $S$ , defined as the sum of distances  $M_k-H_k$  squared taken over all  $k$  (figure 2). The origin  $O$  is the centroid of the set of points. For the sake of clarity, points with negative coordinates are not shown in the figure.

The projection  $O H_k$ , which is also the scalar product of vector  $M_k-O$  with the unitary vector  $e_1$ , is called the score (after Richman 1986). Thus  $M_k-H_k$  is orthogonal to  $e_1$ . There is one score value associated with vector  $e_1$  for each point in  $R^W$ . Let us also assume that  $e_1$  is unique.

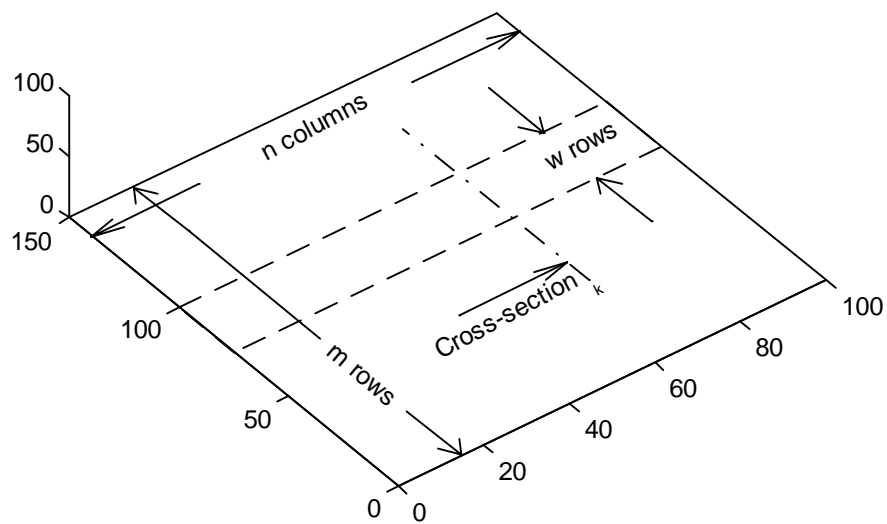


Figure 1 Sketch of the strip notation

If all the values  $M_k H_k$  are zero, we have reduced the problem of original dimension  $w$ , to a one-dimensional one. All the variability in the observations is explained by a single vector  $e_1$ . If this is not the case, we may try to repeat the procedure with the remaining variability  $M_k H_k$ , which belongs to a  $(w-1)$

subspace of  $R^W$  orthogonal to  $\mathbf{e}_1$ . The original measurements  $\mathbf{M}_k - \mathbf{O}$  can be replaced with the difference  $\mathbf{O}\mathbf{M}_k - \mathbf{O}\mathbf{H}_k$ , which is equal to  $\mathbf{M}_k - \mathbf{H}_k$ .

For the new cloud there should be a vector  $\mathbf{e}_2$  (orthogonal to  $\mathbf{e}_1$ ) which minimizes the distance  $S$  in the  $R^W$  space. The process continues until  $w$  vectors  $\mathbf{e}_p$  have been created; each new vector  $\mathbf{e}_p$  being orthogonal to all the previous ones. The vectors  $\mathbf{e}_p$  are called principal components (PC).

Each event  $\mathbf{M}_k - \mathbf{O}$  can be expressed as a linear combination of the PC's

$$\mathbf{M}_k - \mathbf{O} = a_1(k) * \mathbf{e}_1 + a_2(k) * \mathbf{e}_2 + a_3(k) * \mathbf{e}_3 + \dots + a_w(k) * \mathbf{e}_w \quad (1)$$

It can be shown that the scores  $a_j$  associated with vector  $\mathbf{e}_i$  are uncorrelated with those of vector  $\mathbf{e}_j$ . The vectors  $\mathbf{e}_i$  are the eigenvectors of the covariance matrix of the data, and its components are named *loadings* in the literature. The sum of the corresponding eigenvalues equals the sum of the squares of the distances  $M_k H_k$  (Lebart *et al.* 1987).

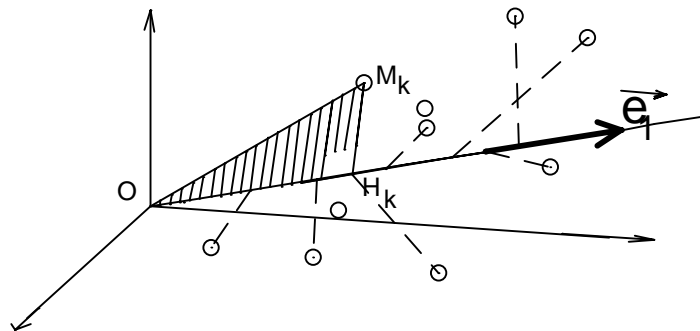


Figure 2 Sketch of the first principal component, for  $w=3$

PCA analysis generates a sequence of principal components, which explains most (or all, for  $p=w$ ) of the variance of the data. This implies that the RMS error in approximating the data with a linear combination of their first  $p$  vectors is a minimum for a given  $p < w$ ; ( $p=1$  in figure 2). It has been shown that



in most cases a good approximation of data is achieved for  $p \ll w$ . Since the  $w$  PC's form a basis in  $R^W$  space, they can replicate exactly any of the  $n$  points in the set, using the scores as weights. López (1997) claims that some of the scores contain essential information on the structure of the cloud, while others are more related to noise. Following Hawkins (1974) he suggested a rule to identify the noisy scores. Once identified, such scores were used to pinpoint those points in  $R^W$  space which are prone to hold an error.

However this is not the complete answer to the problem because each point depends on  $w$  elevation values. Which one is wrong?. Once a point in  $R^W$  space is selected, the elevation (or elevations) which make it unusual should be indicated. This is done using a weighted sum of the squared scores which are related to noise. Such statistics have been suggested for the first time by Hawkins (1974). It is a semi-distance, closely related to the Mahalanobis distance. Its sensitivity in terms of the elevations values is calculated and those elevations which generate the most important contribution to the distance value are considered as errors. The calculations are carried out independently for each outlying point in the  $R^W$  space.

We have briefly presented the procedure to find an error in a single strip. The method can be applied for all row-wise strips to cover the entire DEM. The candidates obtained can be grouped and designated here in after as row-wise candidates. However, the same procedure can be applied to column-wise strips, and a different set of column-wise candidates can be obtained.

The candidates belonging to both sets (row-wise and column-wise) represent the final result. The procedure can be applied iteratively, since, once an error is detected and "corrected", the cloud is modified to some extent, and so are the scores. Each iteration will be named "step" in the following discussion. We keep track of the point already checked in order to avoid to select them twice; we form the candidate set as the intersection of all previous row-wise candidates and all previous column-wise candidates.

The procedure involves five actions, and it can be outlined as follows:

*Given a DEM as a matrix of size  $m*n$*

*subdivide the DEM in row-wise and column-wise strips of width  $w$*

*repeat until criteria are satisfied:*

*a) increment the previous row-wise candidate set:*

*a.1.- locate the columns likely to have candidates*

*a.2.- within each column, find the rows that identify the candidates*

*b) increment the previous column-wise candidate set:*

*b.1.-locate the rows likely to have candidates*

*b.2.- within each row, find the columns that identify the candidates*

*c) intersect both sets*

*d) evaluate criteria*

*e) correct all errors*

*end*

Some remarks follows. In the pseudo code we have used a single strip width  $w$  for rows and columns. This simplifies the tuning process, as will be shown later.

The process is supposed to stop when some criterion is fulfilled. For his experiment, López (1997) suggested to stop if the type I error is too big (defined as the probability of missclassify as error a good value). This criterion is useless for real errors and as will be shown below.

### 2.3 *The modified version*

This variant has been specially designed in order to handle the problem of heavily correlated errors in space. Notice that the procedure of López (1997) has been tested with synthetic, weakly correlated errors. Its performance decays as the auto correlation increases. The procedure of Felicísimo suffers from the same problem, since the error at  $i,j$  is highly correlated with the one at the immediate neighbors. López's procedure does not require that the *along the strip* profiles are contiguous. Therefore we can skip some of them (the ones most correlated) for the analysis. The strip is chosen as before, but in the calculations we consider subsets created using every  $k$ -th row,  $k$  being related to the *range*, a geostatistical property (Samper *et al.* 1990) of the error field. In this paper we assume that the range can be estimated from an independent analysis: it might depend on DEM characteristics, method for obtaining it, scale of aerial photography, etc. The modified method resembles the multigrid approach (Strang 1989) used in scientific computing packages for the solution of differential equation.

### 3. The experiment

To test the method with real data we have selected two DEM of the Aix-en-Provence region in the South of France, both of 12.42 km by 6.9 km, 30 m spacing. A subset of 360 rows and 216 columns was used for all calculations. Both DEM have been described elsewhere (Day *et al.* 1988), and include as a significant feature Mount Sainte Victoire. The first DEM has been produced by photogrammetric measurement of spot elevations from aerial photography. Its accuracy has been estimated by multiple set-up and observation of several blocks within the DEM. An analysis of 830 duplicate points (i.e. set up and measured twice) is presented in table 1. The second DEM has been derived from a set of three SPOT images using an stereo matcher. It has been interpolated to a 30 m grid by using values within a window of size 21 pixels. Elevation values have been obtained using kriging with a spheric variogram of 4000 m<sup>2</sup> sill and 3000 m range, assuming an accuracy for the window of 11 m S.D. Table 2 shows the statistics for the difference between the interpolated DEM (obtained from the stereo matcher's output) and the one manually generated.

Figure 3 illustrates the main features of the DEM, and figure 4 shows the probability density function of the differences in elevation. It should be noticed that the probability of exactly zero error is negligible: only 6 out 95865 points have exactly the same elevation in both datasets.

Table 1 Comparison of 830 duplicate points of the manually derived DEM (From Day *et al.* 1988)

Mean abs. error	-0.026 m
S.D. error	1.837 m
RMSE	12.70 m
Max. (abs. size)	14.66 m
$ \text{error}-\mu >3\sigma$	1.7 %

Table 2 Comparison of 95865 points of the SPOT derived DEM against the manually derived one

Mean abs. error	0.93 m
S.D. error	12.67 m
RMSE	12.70 m
Max.	193.83 m
Min.	-86.22 m
$ \text{error}-\mu >3\sigma$	1.43 %

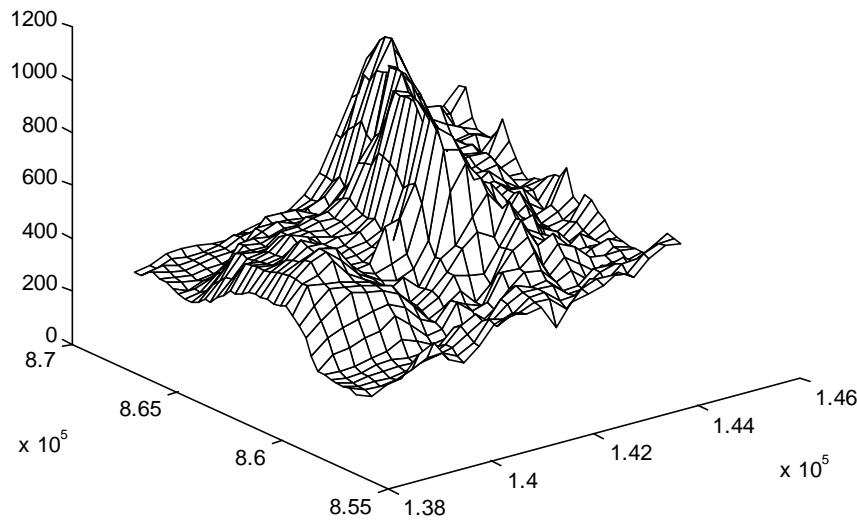


Figure 3 Illustration of the test DEM obtained using only every tenth grid value.

This leads to a paradox: since any choice for the locations will succeed in pointing a true error, the error type I will be identically zero disregarding the procedure, and the error type II will decrease linearly with the effort. This preclude to compare results with those presented in López (1997), since the author used such statistics to decide whether to stop the procedure or to continue. For real datasets a possible measure will be the RMSE between the original and the corrected elevations, and the procedure might go on as long as the RMSE exceeds a preset threshold.

Another interesting result regards some properties of the discrepancy field, i.e., the difference between both DEM. We used geostatistical techniques (Samper *et al.* 1990; Cressie 1993) to describe it. Figure 5 shows a plot of the sampled variogram. Even though the goal of the present paper is not to model the variogram itself, it can be noticed that the range can be roughly estimated as 300 m, i.e., 10 times the grid spacing. This numerical result is in agreement with results obtained by visual analysis of the discrepancy field and can be interpreted as a measure of the spatial auto correlation of the error field. Clearly, the occurrence of errors cannot be regarded as a local phenomenon, a hypothesis assumed by Felicísimo (1994) and López (1997).

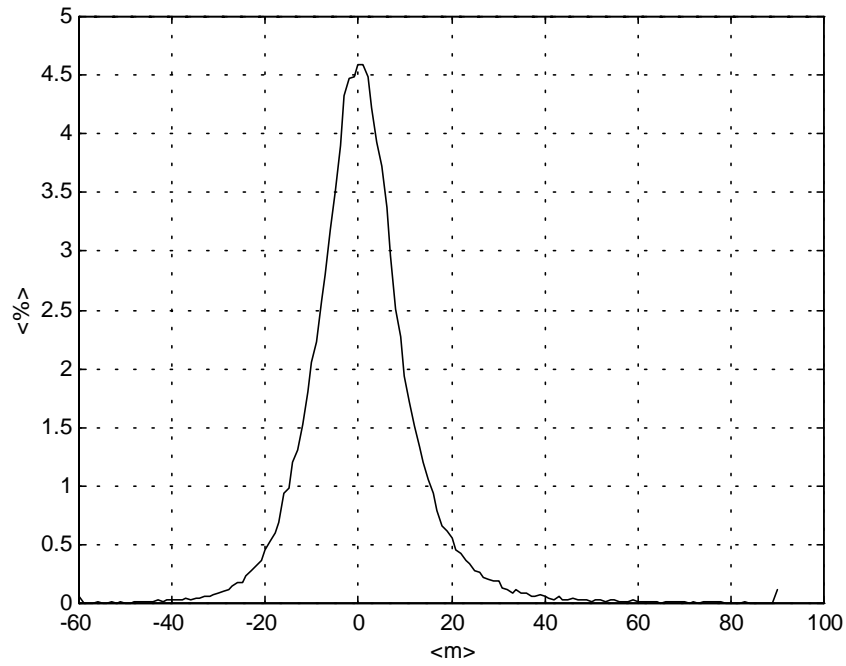


Figure 4 Sampled probability density function of the discrepancies in elevation between both DEM

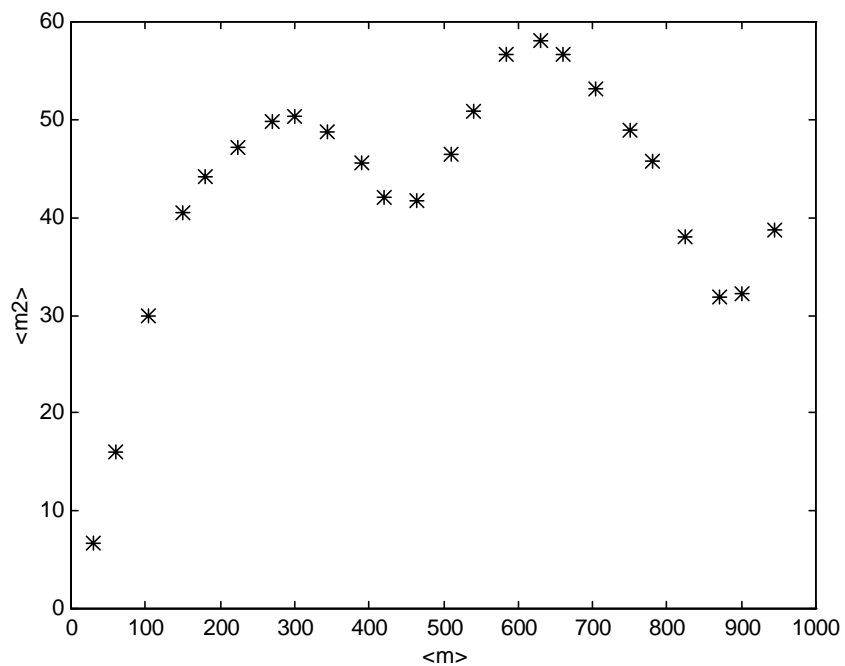


Figure 5 Sampled variogram of the difference of both DEM vs. distance. Results obtained from the GEOEAS software

Most errors are found in a smooth neighborhood regardless if they occur along breaklines or as isolated values. The errors typically influence the data over a distance of 10 pixels. At breaklines the decay should be considered across the line.

Before analyzing the accuracy results, we want to go a bit further into some hypothesis by Felicísimo (1992) a) Gaussian distribution of the errors and b) relationship between outlying values of the  $t_{i,j}$  population and the true errors. We show in figure 6 a QQ-plot of the distribution of the original  $t_{i,j}$  population. The QQ-plot produces a linear relationship when two distributions are of the same type (even though with different parameters). In this case the target distribution is Gaussian. The x-coordinate corresponds to the gaussian cumulative density function (cdf), with no units, while the y-coordinate is the sampled cdf of  $\delta_{i,j}$ , measured in m. As it can be seen, a gaussian distribution is hardly achieved.

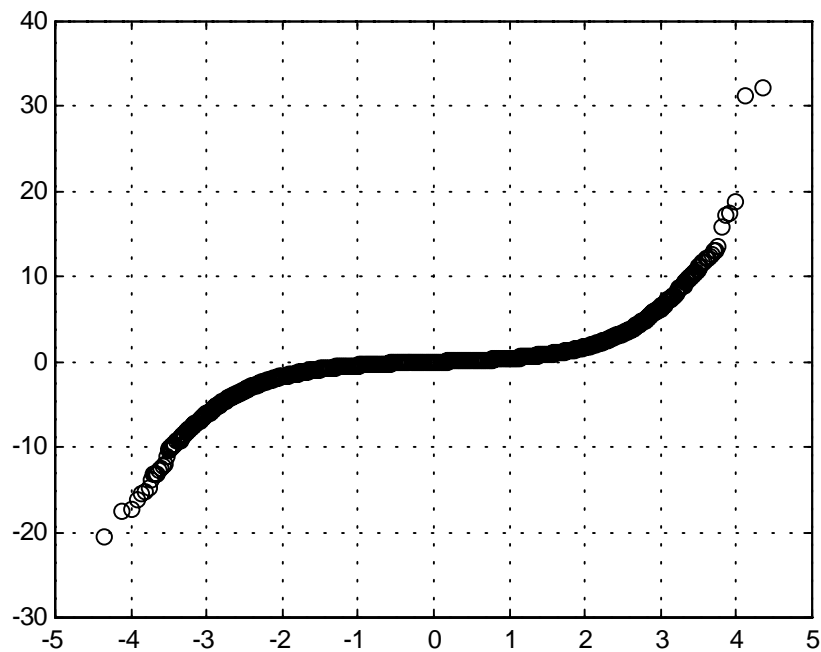


Figure 6 QQ-plot of the  $N(0,1)$  cumulative density function vs. the sampled cumulative density function of  $\delta_{i,j}$ , for the available DEM.

On the other hand figure 7 shows a QQ-plot comparing the distribution of the  $\delta_{i,j}$  against the real errors. The linear relationship shows that both cdf's belong to the same (unknown) class. However, this nearly linear relationships is not automatically related with the expected result (a tool to locate big errors by finding big values of  $\delta_{i,j}$ ). A strong correlation is requested, but unfortunately this is not the case. The regression of  $\delta_{i,j}$  vs. real errors (not shown) has an  $R^2$  coefficient is 0.0858, well below the ideal value of 1.00.

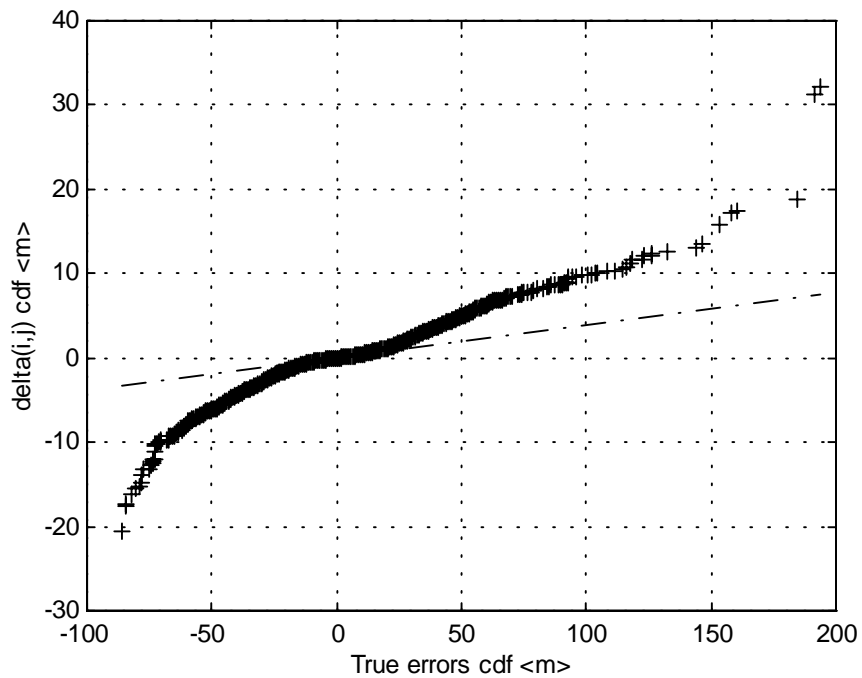


Figure 7 QQ-plot of the sampled cumulative density function (cdf) of the true errors vs. the  $\delta_{i,j}$  cdf for the available DEM.

#### 4. Results under the "perfect inspector" hypothesis

We may analyze the results using different statistics. The most interesting one will take into account the evolution of the elevation accuracy in terms of the editing effort. For our purposes we measure the editing effort as the number of elevation values checked divided by the total number of points in the DEM. We assumed that the user has a correction procedure and that procedure is perfect ("*perfect inspector*" assumption).

Normally the accuracy of a DEM is not directly known to the user; it can be estimated through sampling in isolated points if more precise measurements are available. For practical purposes it might be more meaningful to use statistics from the distribution of the errors detected while working with the dataset. For example, its RMSE will measure the size of the errors detected by the method for a given effort. We disregard the RMSE for each step, because its variability precludes any simple analysis.

We believe that a clear measure of the effort involved should be included. The effort per step (in turn) depends strongly on the choice of the margin level. It regulates how much of the tail of the distribution of the noisy scores will be regarded as being in error. Cutting out the tails might produce an empty set of candidates.

In order to avoid this we slightly increased the margin level to assure that there will be candidates to check in each step.

Figure 8 shows the evolution of the accuracy measured in terms of the RMSE for a strip width  $w=8$ . The boundaries of the dashed regions at the top and the bottom show the *worst* and *best* possible operation locus. The former is obtained by considering first the smallest discrepancies, while the latter corresponds to selecting the largest discrepancies first.

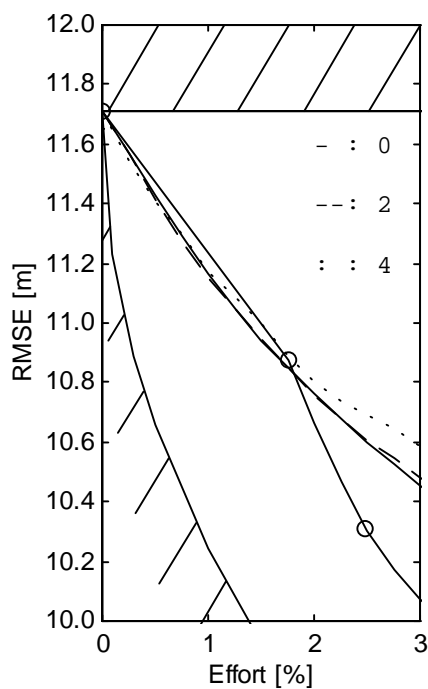


Figure 8 Evolution of the accuracy (measured by the RMSE in m) vs. the effort for the methods of Felicísimo (1994) (with the  $\ominus$  symbol) and López (1997). Results for  $w=8$ . Different lines correspond to different number of uncontrolled scores. Left plot 8a shows details of the right one 8b.

Under our assumptions both lines should meet at 0 and at 100 per cent. Even though both limits are hardly of practical interest (because it requires knowing the errors in advance) they give a better understanding of the process. Lines with the  $\ominus$  symbol are for the Felicísimo (1994) method while the others are for different controlled scores as of López (1997) method. Figure 8a has more detail in the low effort region, while figure 8b has been extended up to 15 per cent effort. It is clear that the Felicísimo's method outperforms the López's method in the long run, but at the low effort they are similar. This region is of primary importance for two reasons. Firstly because most users will not go too much further. End users



neither have extra data nor too much tools, so they will correct at most the worst errors. DEM producers might go back and make another measurement, but this might become a boring task if new values do not differ substantially from the old ones. Secondly, according to Torlegård *et al.* (1986) blunders typically account for less than 3 per cent of the dataset, 0.5 per cent being a median value. Thus pursuing the task over such limit might be misleading, because the methods have been designed for finding gross errors only.

It should be noticed that none of the methods shows at 0 per cent effort a slope comparable to the best possible method, which implies that the most important errors are not found in the early stages of the procedure.

We also tested some other options for the width parameter  $w$  which will not be presented here. Figure 9 compares the accuracy performance of the modified method vs. Felicísimo's one. The figure was obtained after subdividing the DEM in regions of width 72 rows, and building the strips taking every 9th row within the region. Thus the "strip" width is again 8. Notice that we skip nearly 10 rows, as suggested by the range of the variogram. The plot of the Felicísimo's method is again included for comparison. The most striking fact is the difference in the slope at 0 per cent effort which is markedly closer to the best one. This implies that larger errors are found earlier, leading to a faster decrease of the RMSE. However, once those important errors are removed, the remaining ones are difficult to locate, and the simpler Felicísimo's method is better if the effort exceeds 1.75 per cent.

The end user can calculate RMS values of the errors already found like those presented in figure 10. The x-coordinate is the effort defined as before, while the y-coordinate is the RMSE of the population already corrected. The 0 per cent value is not defined. Plots correspond to the Felicísimo (1994) approach and the modified method of López (1997). It is clear that the former finds larger errors in the "long" run (over 1.75 per cent effort) but the latter is fairly better for lower effort values. Three lines with different number of uncontrolled scores are shown, and it is clear that the one of 0 value is very similar to the one of 2, except very close to the 0 per cent effort.

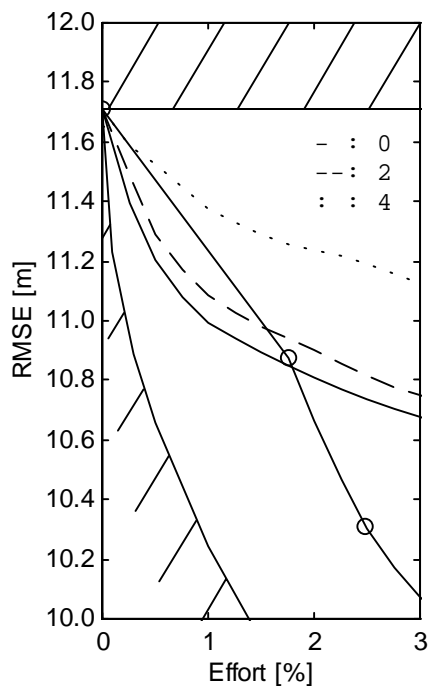


Figure 9 Evolution of the accuracy (measured by the RMSE in m) vs. the effort for the methods of Felicísimo (1994) (with the  $\ominus$  symbol) and the modified of López (1997). Results choosing every 9th row. Different lines correspond to different number of uncontrolled scores. Left plot 9a shows details of the right one 9b.

We also analyzed the spatial location of the errors found when a substantial amount of work has been done. Figure 11 shows the places where Felicísimo's method pointed out the errors up to the 3 per cent effort (in black), and up to 15 per cent (in gray). We noticed that most of them are concentrated along significant features of the DEM, namely breaklines where slope changes abruptly. In such points the second order polynomial is not a good approximation of the surface, so differences larger than expected arise. Once some values are corrected, such differences are even more evident, but since we do not allow any point to be corrected twice, its closer neighbors become candidates, explaining the "clear" image. Figure 12 correspond to the behaviour of the original method by López, 1997, while Figure 13 shows the pattern for the modified method. Both images look "noisy" with points randomly located, but it is clear that the modified method does not follow the features of the terrain, while the original method is more prone to do so.

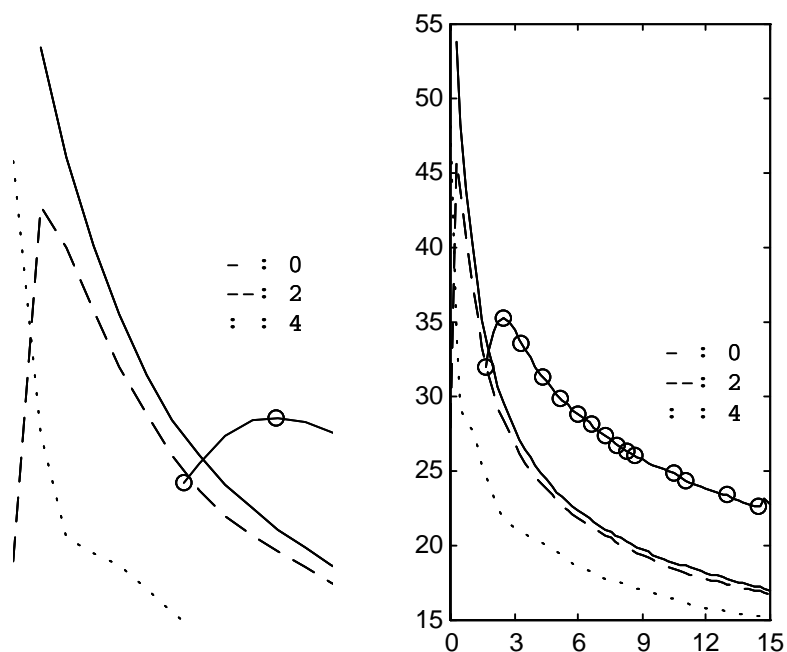


Figure 10 Evolution of the RMSE found of the cumulated errors up to a given effort vs. the effort, for the methods of Felicísimo (1994) (with the  $\circ$  symbol) and the modified of López (1997). Results choosing every 9th row, resulting strips of  $w=8$ . Left plot shows details of the right one

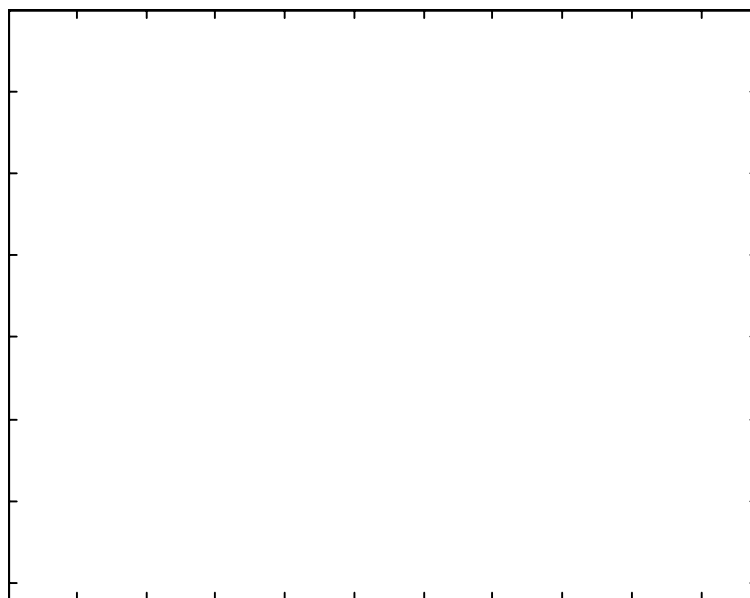


Figure 11 Binary map of the errors located up to the 15 per cent effort with the method of Felicísimo (1994). Black areas are for the suggested locations up to the 3 per cent effort; gray ones are obtained after 15 per cent effort

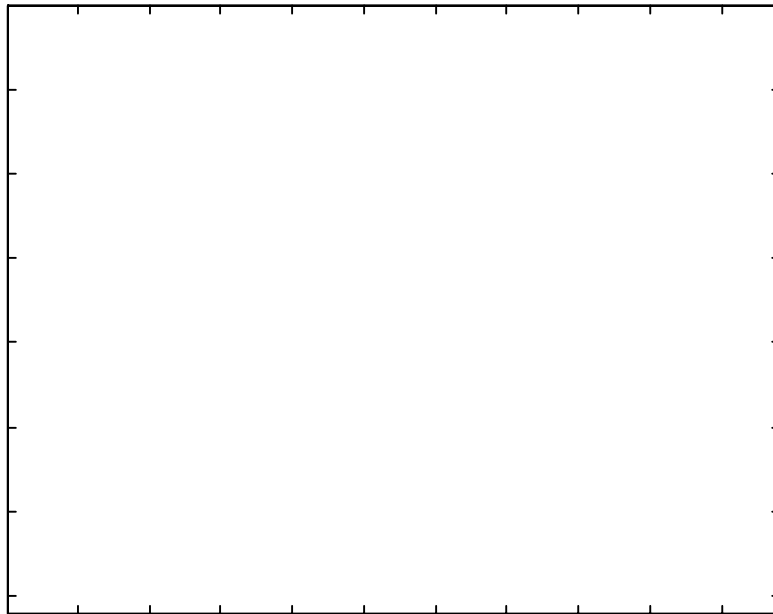


Figure 12 Binary map of the errors located up to the 15 per cent effort with the original method of López (1997). Black areas are for the suggested locations up to the 3 per cent effort; gray ones are obtained after 15 per cent effort

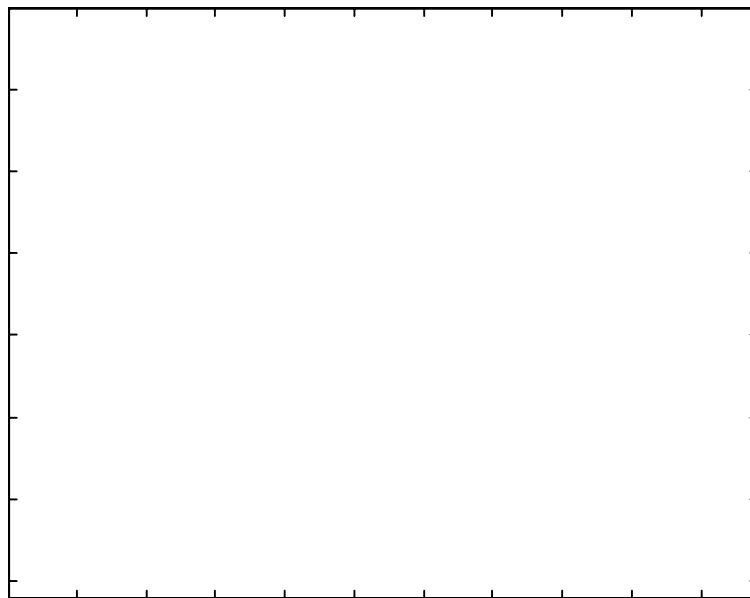


Figure 13 Binary map of the errors located up to the 15 per cent effort with the modified method of López (1997). Black areas are for the suggested locations up to the 3 per cent effort; gray ones are obtained after 15 per cent effort

## 5. Results using a "non perfect inspector" assumption

In figures 14 and 15 we show results obtained after removing the "perfect inspector" assumption. Under the new hypothesis, the inspector uses an interpolation procedure to correct those values which are believed to be wrong. The procedure uses the elevations in the neighborhood and finds a best fit using second order polynomials (as described for the Felicísimo 1994 method). After an elevation has been corrected, it will not be modified again. If errors are located in adjacent pixels, they will mutually interact masking each other and if the procedure detects one of them, the inspector will produce a modified but not correct interpolated surface. In this case there is no best and worst operation line, because the procedure might in principle introduce new and worse errors in the dataset. Both figures 14 and 15 were calculated as before using the high accuracy DEM as a reference

The Felicísimo method shows lower performance, mostly due to its preference to pinpoint clusters of elevations. Once an error is imputed using its neighbors it will become a "corrected" point; and if the neighbors are wrong, it will be so. Few improvements in any estimate of accuracy can arise in such case. Our proposed procedure also shows a degraded performance, but to a lesser extent. This can be easily explained analyzing the pattern of candidate locations (fig. 13) because our candidates are typically isolated; once imputed (maybe with also wrong values) the next candidate is usually not close to the previous one, and the error correction procedure is not trapped in the neighborhood of few candidates.

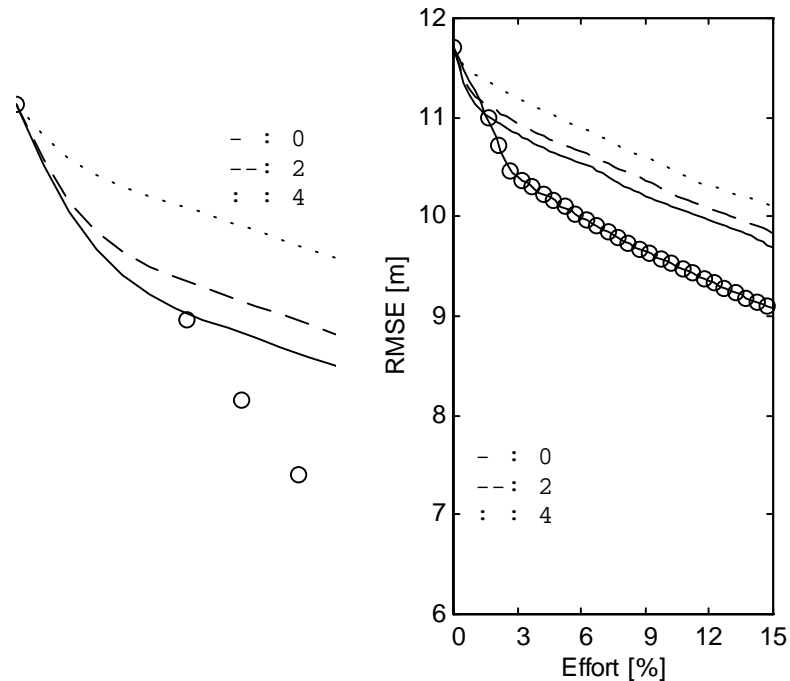


Figure 14 Evolution of the accuracy (measured by the RMSE in m) vs. the effort for the method of Felicísimo (1994) (with the  $\circ$  symbol) and the modified method. The perfect inspector hypothesis has been removed. Results are shown choosing every 9th row. Left plot shows details of the right one

## 6. Discussion

We have compared two published methods for locating errors (also named outliers or blunders) in raster datasets. We also suggested a modification for one of them, and we carried out a comparative test for all three methods using real data with known errors. The method suggested by Felicísimo (1994) is very simple, but no results using either synthetic or real errors were previously reported. One interesting fact is that this method is parameter free, so end users have nothing to adjust or choose. However it has been derived under some hypotheses that do not apply to the DEM used in this study. It relies on a low order polynomial interpolator using only nearest neighbors. We think that it will work better in smooth terrain. The use of low order polynomials tends to pinpoint locations which are close to each other, a situation which is more likely to occur with systematic errors. For further work we suggest considering the use of a local Universal Kriging interpolator (Samper *et al.* 1990) using more neighbors, which is in line with the findings of Giles *et al.* (1996) who also used a window with 11 by 11 elements. The Kriging approach also allows to model different spatial auto correlation scales.

The overall results show that *if* better elevation values can be derived using the same raw data, this approach leads to higher accuracy, provided that they there are no systematic errors.

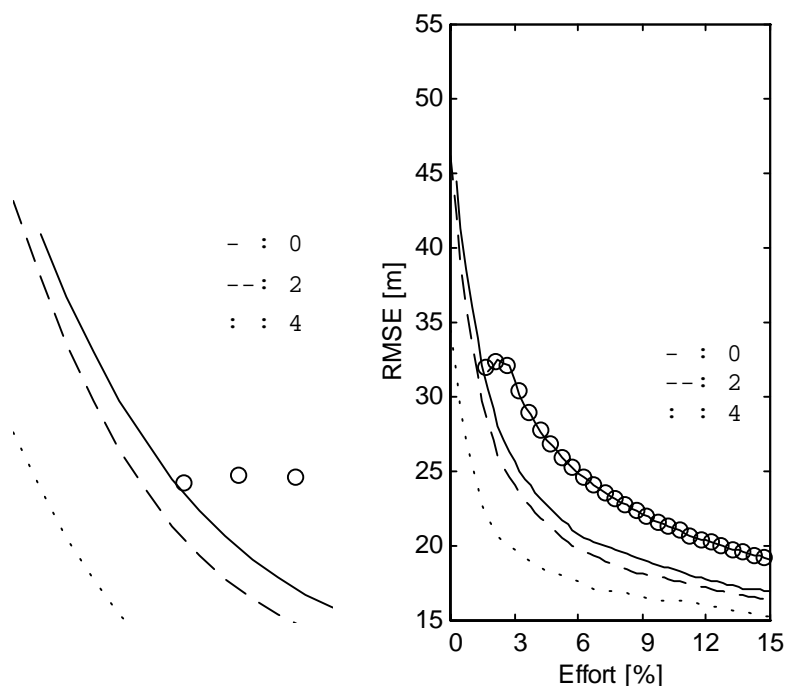


Figure 15 Evolution of the RMSE found of the cumulated errors up to a given effort vs. the effort, after removing the hypothesis of the perfect inspector. Plots are for the methods of Felicísimo (1994) (with the  $\circ$  symbol) and the modified method of López (1997). Results choosing every 9th row. Left plot shows details of the right one

The method outlined by López (1997) has been designed for and tested with synthetic errors with very low spatial auto correlation. For our case, where errors show heavy spatial auto correlation, it performs only slightly better than Felicísimo's for low effort, but it is outperformed in any other case. We consider the performance of López's method as poor.

In order to handle the spatial auto correlation of errors, we have proposed a modification of the method by López. We form the strips by subsampling the DEM at each  $k$ -th row. From a programming point of view this is a minor change. In real applications, the number  $k$  has to be fixed *a priori*. Östman (1987b) suggested that  $k$  is strongly connected with both the DEM and the acquisition method. We estimated the range from the sampled variogram. López (1997) describes a rule how to determine how many scores are considered describing the structure of the cloud. For this case, this rule suggest a value of 2. Slightly better results were obtained using 0. However, in a first approximation the rule is still valid.

All three methods have been used in an iterative fashion. Once some errors were removed, all the calculations have been carried out again, and new candidates appear. If this is not the case, some parameters are modified automatically (lowering confidence limits, for example) in order to continue the operation. We continue until 15 per cent of the DEM elevation values have been corrected or confirmed. According to Torlegård *et al.* (1986) gross errors account for less than 3 per cent of the population, so the 15 per cent limit is well within either the systematic (as defined by Thapa *et al.* 1992) or the random error set, provided the first 3 per cent were really gross errors.

We assumed that, once an error is located, it can be replaced by a "better" value. In real applications the procedure will be different depending on the user. In a DEM production environment, some action can be taken to check these identified isolated values. In photogrammetric measurements these checks can be done before removing the stereopair. The goal here is to improve the overall accuracy, while the effort is less crucial. On the other hand, the end user is left alone in most cases, because he may not be able to go to the original data sources. Therefore he will be interested in "evident" errors, i.e. those of relevant size (which are typically few).

All the results have been obtained using just one DEM corresponding to hilly terrain, derived using photogrammetric techniques. The relative weight of gross, systematic and random errors is different for different DEM production procedures, and to assess the impact of such differences in the methods presented is a topic of further research. The modified method attempts to filter out the systematic error, so it is likely to work well in either hilly or smooth terrain.

A comment about the computer time requirements: the method of Felicísimo (1994) is fairly cheap (of the order of  $m \cdot n$  operations, being  $(m, n)$  the size of the DEM), while the method of López (1997) and the modified procedure presented here involve, for each step, the computation of  $(m/w) \cdot (n/w)$  covariance matrices of size  $(w, w)$ , which takes  $[(n/w) \cdot O(n^2) + (m/w) \cdot O(m^2)] \cdot O(w^2)$  operations; to calculate the eigenvectors requires in turn  $[(n/w) + (m/w)] \cdot O(w^2)$  operations, and to project each strip to calculate the scores requires  $(m+n) \cdot w$  operations. Some other operations are required but depend linearly on  $m$  and  $n$ . In our example, a DEM of size  $m=360$  and  $n=216$ , for  $w=8$ , about 5 minutes per step are required using MATLAB in a SUN Sparc 20. The overall procedure is still considered cheap in terms of computer time.



## 7. Conclusions

Some methods to locate gross errors in quantitative raster data have been presented, and they were tested using a grid-based DEM with known errors. The DEM was produced by automatic stereo matchers from SPOT image pairs, and has elevations ranging from 181 to 1044 m. A more accurate DEM of the same area is available; it has been considered as the ground truth. The detection of gross errors is highly desirable, because they significantly affects all statistics commonly used to report quality. Since they are typically few, once located it might be practical to make some corrective actions to replace them with better quality estimates.

The usual hypothesis of errors uncorrelated in space (assuming no gross or systematic errors) seems to be wrong at least for this case, as well as the assumption of gaussian distribution for the residuals. This poses serious concern about the usefulness of some previously published algorithms (Felicísimo 1994; López 1997) and motivated this work.

The results suggest that Felicísimo (1994) method find mostly what is regarded as systematic errors, mainly due to the interpolation algorithm (biquadratic polynomial). López (1997) show similar results in terms of RMS of errors only in the early stages of the correction process.

In order to handle the significant spatial auto correlation observed, a modified version of the method of López (1997) has been designed. It was expected that it will ignore the systematic errors, and just bring up the gross and the random ones. Once tested with the same dataset, previous results were significantly improved and exceeded those of Felicísimo up to a certain level of effort, the effort being defined as the fraction of the DEM elevations corrected or revised. This effort level (1.75 per cent) is of the order of the number of gross errors typically found in photogrammetrically derived DEM as reported by Torlegård *et al.* (1986). Its location pattern looks sparse and random (as expected for gross errors), in opposition to the smooth pattern produced by Felicísimo's method, typical of systematic errors.

Different DEM production procedures might create particular combination of gross, systematic and random errors. The limited scope of this experiment does not allow us to generalize its conclusions to *any* DEM irrespective of its lineage; however, it is felt that gross errors will be easily detected if the systematic errors are negligible or they can be filtered out by some procedure.

The modified method has some free parameters; the most important is an estimate of the auto correlation lag (or the range of the variogram). It can be estimated from a limited number of independent control points; some authors claim that such value depends on the method for acquisition of the DEM and the DEM itself.

We assumed for most of the work that once an error is identified, it can be corrected without error. This is known as the *perfect inspector* hypothesis. In the case of using the error detection procedure in a semi-automatic production environment, the method warns the operator about possible errors before the stereopair is unmounted, enabling new measurement. In a fully digital production environment, some correlation thresholds are usually varied to minimize computer time. The method may be used to selectively strengthen the correlation thresholds in suspicious points. In case there is no possibility to verify the errors, e.g. for end users, the algorithm will help to locate the most unlikely values; we have simulated the situation under an *imperfect inspector* hypothesis. Unlikely values were imputed using a polynomial fit using nearest data, and the results were qualitatively similar to those obtained using ground truth data for imputation.

## 8. Acknowledgments

This work was started at the Department of Geodesy and Photogrammetry, Division of Geoinformatics, of the Royal Institute of Technology, Stockholm, Sweden, under the supervision of Assoc. Prof. Hans Hauska. Prof. Jan Peter Muller from the University College, London, have kindly supplied both DEM and related information. The author wishes to acknowledge BITS (Swedish Agency for International Technical and Economical Cooperation) and CONICYT (Consejo Nacional de Investigaciones Científicas y Técnicas, Uruguay) for partial funding. Computer resources have also been kindly supplied by the Centro de Cálculo of the School of Engineering, University of the Republic, Montevideo, Uruguay.

## References

- ABREU, E.; LIGHSTONE, M.; MITRA, S.K. AND ARAKAWA, K., 1996, A new efficient approach for the removal of impulse noise from highly corrupted images. *IEEE transactions on image processing*, **5**, 6, 1012-1025.
- ACKERMANN, F., 1995, Digitale Photogrammetrie - Ein Paradigma-Sprung. *Zeitschrift fur Photogrammetrie und Fernerkundung*, 3/95, 106-115
- ANONYMOUS, 1994, ESRI tops the charts again. *ESRI ARC News*, **16**, 3, 35-35

- BETHEL, J. S.; MIKHAIL, E. M., 1984, Terrain surface approximation and on-line quality assessment. In *Proceedings of the 15th. International Congress of the International Society for Photogrammetry and Remote Sensing, Rio de Janeiro. Commission III*, **25**(A3a), 23-32
- CRESSIE, N., 1993, Statistics for spatial data. John Wiley and Sons, ISBN 0-471-00255-0, 900 pp
- DAY, T.; MULLER, J.P., 1988, Quality assessment of Digital Elevation Models produced by automatic stereo matchers from SPOT image pairs. In *Proceedings of the 16th. International Congress of the International Society for Photogrammetry and Remote Sensing, Kyoto. Commission III*, 148-159.
- FELICÍSIMO, A., 1994, Parametric statistical method for error detection in digital elevation models. *ISPRS J. of Photogrammetry and Remote Sensing*, **49**(4): 29-33
- FÖRSTNER, W., 1983, On the morphological Quality of digital elevation models. In *International colloquium on Mathematical aspects of Digital Elevation Models, Stockholm*.
- GILES, P.T. AND FRANKLIN, S.E., 1996, Comparison of derivative topographic surfaces of a DEM generated from stereoscopic SPOT images with field measurements *PE&RS*. **62**, 10, 1165-1171
- HADI, A. S., 1992, Identifying Multiple Outliers in Multivariate Data. *J. Royal Statist. Soc. B* **54**,3,761-771
- HADI, A. S., 1994, A Modification of a Method for the detection of Outliers in Multivariate Samples. *J. Royal Statist. Soc. B* **56**, 2, 393-396
- HANNAH, M. J., 1981, Error detection and correction in Digital Terrain Models. *PE&RS* **47**, 1, 63-69
- HAWKINS D. M., 1974, The detection of errors in multivariate data, using Principal Components. *JASA*, **69**, 340-344.
- HAWKINS, D. M., 1993a, A feasible solution algorithm for the Minimum Volume Ellipsoid Estimator in Multivariate data. *Computational Statistics* **8**, 95-107
- HAWKINS, D. M., 1993b, The feasible set algorithm for least median of squares regression. *Computational Statistics & Data analysis* **16**, 81-101
- HUNTER, G. AND GOODCHILD, M., 1996, Communicating uncertainty in spatial databases. *Transactions in GIS* **1**, 1, 13-24
- JOHN, S. A., 1993, Data integration in a GIS - The question of data quality. *ASLIB Proceedings*, **45**, 4, 109-119.

- LEBART, L.; MORINEAU, A.; TABARD, N., 1977, Techniques de la description statistique: Methodes et logiciels pour l'analyse des grands tableaux. Ed. Dunod, Paris, 344 pp.
- LÓPEZ, C., 1997, Locating some types of random errors. *International Journal of Geographic Information Science*, 11, 7, 677-698
- ÖSTMAN, A., 1987a, Quality control of Photogrammetrically sampled Digital Elevation Models. *Photogrammetric Record*, **12**(69), 333-341.
- ÖSTMAN, A., 1987b, Accuracy estimation of Digital Elevation Data Banks. *Photogrammetric Engineering and Remote Sensing*, **53**, 4, 425-430.
- MITRA, S.K. AND YU, T.H., 1994, A new nonlinear algorithm for the removal of impulse noise from highly corrupted images. *IEEE 1994 International symposium on circuits & systems* **3**, 17-20
- NEBERT, D., 1995, Status of the National Geospatial Data Clearinghouse on the Internet. In *Proc. of the 15th. Annual ESRI User Conf., May 1995*. Also [www.fgdc.gov/clearinghouse/pubs/esri95/p196.html](http://www.fgdc.gov/clearinghouse/pubs/esri95/p196.html)
- NEBERT, D., 1996, Supporting Search for Spatial Data on the Internet: What it means to be a Clearinghouse node. In *Proc. of the 16th Annual ESRI User Conf.* Also [www.fgdc.gov/clearinghouse/pubs/esri96/revised.html](http://www.fgdc.gov/clearinghouse/pubs/esri96/revised.html)
- RICHMAN, M. B., 1986, Review article: Rotation of principal components. *J. of Climatology*, **6**, 293-335.
- SAMPER, F.J. AND CARRERA, J., 1990, Geoestadística: aplicaciones a la hidrología subterránea. ISBN 84-404-6045-7 480 pp (in spanish)
- STRANG, G., 1989, Introduction to applied mathematics. Wellesley-Cambridge Press, 510 pp
- THAPA, K. AND BOSSLER, J., 1992, Review article: Accuracy of spatial data used in Geographic Information Systems. *Photogrammetric Engineering & Remote Sensing*, **58**, 6, 835-841
- TORLEGÅRD, K; ÖSTMAN, A. AND LINDGREN, R., 1986, A comparative test of photogrammetrically sampled digital elevation models. *Photogrammetria*, **41**(1), 1-16