

This article was downloaded by:[Aguilar, F. J.]
[Aguilar, F. J.]

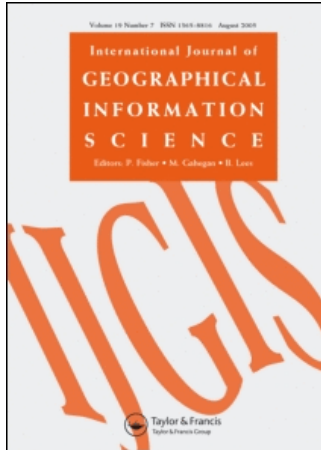
On: 14 May 2007

Access Details: [subscription number 778147506]

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Geographical Information Science

Publication details, including instructions for authors and subscription information:
<http://www.informaworld.com/smpp/title-content=t713599799>

Accuracy assessment of digital elevation models using a non-parametric approach

To cite this Article: , 'Accuracy assessment of digital elevation models using a non-parametric approach', International Journal of Geographical Information Science, 21:6, 667 - 686

To link to this article: DOI: 10.1080/13658810601079783

URL: <http://dx.doi.org/10.1080/13658810601079783>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007

Research Article

Accuracy assessment of digital elevation models using a non-parametric approach

F. J. AGUILAR*, M. A. AGUILAR and F. AGÜERA

Department of Agricultural Engineering, University of Almería, La Cañada de San Urbano, 04120 Almería, Spain

(Received 29 March 2006; in final form 20 October 2006)

This paper explores three theoretical approaches for estimating the degree of correctness to which the accuracy figures of a gridded Digital Elevation Model (DEM) have been estimated depending on the number of checkpoints involved in the assessment process. The widely used average-error statistic Mean Square Error (MSE) was selected for measuring the DEM accuracy. The work was focused on DEM uncertainty assessment using approximate confidence intervals. Those confidence intervals were constructed both from classical methods which assume a normal distribution of the error and from a new method based on a non-parametric approach. The first two approaches studied, called Chi-squared and Asymptotic Student t , consider a normal distribution of the residuals. That is especially true in the first case. The second case, due to the asymptotic properties of the t distribution, can perform reasonably well with even slightly non-normal residuals if the sample size is large enough. The third approach developed in this article is a new method based on the theory of estimating functions which could be considered much more general than the previous two cases. It is based on a non-parametric approach where no particular distribution is assumed. Thus, we can avoid the strong assumption of distribution normality accepted in previous work and in the majority of current standards of positional accuracy. The three approaches were tested using Monte Carlo simulation for several populations of residuals generated from originally sampled data. Those original grid DEMs, considered as ground data, were collected by means of digital photogrammetric methods from seven areas displaying differing morphology employing a 2 by 2 m sampling interval. The original grid DEMs were subsampled to generate new lower-resolution DEMs. Each of these new DEMs was then interpolated to retrieve its original resolution using two different procedures. Height differences between original and interpolated grid DEMs were calculated to obtain residual populations. One interpolation procedure resulted in slightly non-normal residual populations, whereas the other produced very non-normal residuals with frequent outliers. Monte Carlo simulations allow us to report that the estimating function approach was the most robust and general of those tested. In fact, the other two approaches, especially the Chi-squared method, were clearly affected by the degree of normality of the residual population distribution, producing less reliable results than the estimating functions approach. This last method shows good results when applied to the different datasets, even in the case of more leptokurtic populations. In the worst cases, no more than 64–128 checkpoints were required to construct an estimate of the global error of the DEM with 95% confidence. The approach therefore is an important step towards

*Corresponding author. Email: faguilar@ual.es

saving time and money in the evaluation of DEM accuracy using a single average-error statistic. Nevertheless, we must take into account that MSE is essentially a single global measure of deviations, and thus incapable of characterizing the spatial variations of errors over the interpolated surface.

Keywords: DEM; Accuracy assessment of spatial data; Estimating functions; Confidence intervals; Monte Carlo simulation method

1. Introduction

GIS users demand accurate and reliable information about the quality of their spatial data. For this reason, the extent to which the producer is able to guarantee the quality reported is important. We must take into account that going to the field and checking the positional accuracy of newly acquired geo-spatial data is very difficult and costly for the customer (Jakobsson 2002). In fact, it is probably not the user's task, and this kind of information should be presented by the producer as a standard part of the product. Therefore, measuring the positional error of geo-spatial data is one of the major research issues in the area of quality assessment of spatial data (Shi and Bedard 2004).

In practice, statistical tests are the only way to ensure that requirements are met at a moderate cost. This means following an inference process about certain parameters associated with the population of residuals. To carry out this inference, statistics computed over random samples are usually employed. Therefore, it should be adequate to establish general rules to guarantee an efficient (inexpensive) sampling.

With regard to sampling considerations, samples should be independent, representative of a very large population and truthful. If residuals at checkpoints present a significant spatial autocorrelation because they are located too close to each other (within the range of spatial autocorrelation), it would be mandatory to introduce some kind of correction for computing the equivalent independent sample size (Cressie 1993). Otherwise, the use of spatially autocorrelated residuals would lead to reporting a lower error than is correct.

With respect to sampling size, it is very difficult to determine the minimum number of checkpoints needed to guarantee a reliable assessment of DEM accuracy due to the high variability of the data that can be found. As a rule of thumb, a sample size of at least 30 checkpoints is employed in most situations (Griffith 1996). The widely quoted National Standards for Spatial Data Accuracy (NSSDA), elaborated by the Federal Geographic Data Committee (FGDC 1998), recommend the use of a minimum of 20 checkpoints to reflect the geographical area of interest and the error distribution in the dataset. Such standards assume a perfectly normal distribution of the residuals. Nevertheless, the central limit theorem only guarantees normality for a greater number of observations. As a result, it is likely that the positional accuracy assessment is overestimating the reliability of the DEMs. Ariza and Atkinson (2005) found recently that the use of the minimum sample size proposed by the NSSDA method leads to a variability of the horizontal accuracy of the order of 11%, which means that the confidence level of 89% is insufficient with respect to the theoretical value of 95%.

It is clear therefore that more than 20 checkpoints should be contemplated. But the question is, exactly how many? Some authors, such as Ley (1986), recommend a remarkably higher number of checkpoints (roughly 150) in order to guarantee an error assessment lower than around 10%. Ariza and Atkinson (2005) stated that

about 100 checkpoints should be sufficient for a 95% confidence level for the horizontal accuracy estimated from NSSDA.

Rigorous theoretical models designed to compute the sample size of residuals needed to guarantee a predetermined reliability are lacking. Thus, the model proposed by Li (1991) assumes a Gaussian distribution of residuals, but this is sometimes incompatible with the fact that vertical errors in a DEM often follow a non-normal distribution (López 1997, Maune *et al.* 2001b).

On the other hand, Aguilar *et al.* (2005b) proposed a theoretical model to compute the reliability of a DEM based on the Root Mean Square Error (RMSE). They measured the error of the accuracy as the coefficient of variation of the RMSE. This is the ratio of the expected standard deviation of vertical RMSE and the expected value of vertical RMSE. Either the numerator or denominator was expressed as a function of the number of checkpoints used for computing reliability and some parameters calculated from the residual distribution. In this case, residual data sets do not need to be constrained to the assumption of distribution normality.

A clear distinction can normally be established between the potential sources of DEM uncertainty ascribable to measuring the height of sample data points (sample data error) and the sum of the interpolation error and the error due purely to sampling the continuous terrain surface with a finite grid interval. The latter is sometimes referred to as information loss (Huang 2000). Of the two terms mentioned, sample data error is relatively easy to obtain from empirical observations. For example, the photographic scale is the main determinant of the accuracy of DEMs generated from photogrammetric methods. In this case, sample data error can be estimated with reasonable approximation from the flying height above mean terrain. Nonetheless, estimating information loss presents greater problems because terrain shape varies from one place to another. It is therefore impossible to find an analytical method to compute the standard deviation for all the height differences between terrain and DEM surface (Aguilar *et al.* 2006).

The main objective of the current paper is the development of a new approach to construct DEM vertical accuracy confidence intervals from a finite sample of residuals. Therefore, it could be considered as an extension of the work of Aguilar *et al.* (2005b). The purpose is not only to offer a simple measure of reliability to which the accuracy figures of a DEM have been estimated, but to compute an accurate confidence interval within which the global DEM accuracy measure should be located with a certain level of confidence.

For this purpose, residuals are calculated as height differences between 'true terrain elevation values' and corresponding elevation values extracted from the DEM. These true terrain elevation values or checkpoints should be obtained from an independent source of higher accuracy (Maune *et al.* 2001a). In this way, a global DEM accuracy measure, widely used and known as Mean Square Error (MSE), can be computed from the set of height differences at checkpoints.

The rest of this paper is organized as follows. Section 2 outlines a detailed theoretical development of the three approaches tested for constructing confidence intervals. The first approach, referred to here as the 'Chi-squared approach', assumes a normal distribution of the residuals and is developed in section 2.1. The second approach uses the asymptotic properties of the *t* distribution as an approximation to a hypothetical non-normal DEM residual population. It is examined in section 2.2. The third approach is the novel contribution of this paper and could be considered much more general than the previous cases because a non-parametric approach is used

where no particular distribution assumption is made. The confidence intervals are calculated using the standard theory of order statistics and estimating functions. It is developed in section 2.3. In section 3, data sets and the experimental method used for testing the performance of the theoretical models developed in section 2 are described. At the end of this section, the numerical procedure of the Monte Carlo method employed to simulate observed data is also explained. The results corresponding to the characteristics of the residual populations generated are discussed in section 4. The performance of the three models for constructing reliable MSE confidence intervals is discussed for samples taken from the populations generated in section 3. Conclusions are provided in the last section.

2. Theoretical approach to the assessment of DEM uncertainty

2.1 Chi-squared approach to the development of MSE confidence interval

Let us suppose that the difference in height between true terrain surface and interpolated DEM surface is a random variable, X , with a normal distribution. In this case, a finite sample of N height differences, i.e. $X = \{x_1, x_2, \dots, x_N\}$, could be considered as the residuals calculated at the N checkpoints used for computing DEM accuracy (Li 1988). In this way, a confidence interval can be constructed for estimating MSE from that sample. First of all, MSE is calculated according to equation (1):

$$\text{MSE} = \frac{\sum_{i=1}^N x_i^2}{N} = \frac{\sum_{i=1}^N e_i}{N} = \bar{e} \quad (1)$$

e_i being the square of each residual and \bar{e} the mean value. Note that the sum of the squares of N independent and random variables of mean 0 and variance 1 should be distributed as chi-squared (χ^2) with N degrees of freedom (Steel and Torrie 1980). Under this hypothesis, the following can be written:

$$\chi^2 = \sum_{i=1}^N y_i^2 = \sum_{i=1}^N \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \quad (2)$$

y_i being the standardized variables corresponding to each x_i residual of mean μ_i and variance σ_i^2 . In fact, variables y_i have approximately the normal distribution with mean 0 and variance 1 as long as the sample size is sufficiently large. However, equation (2) is more general than we need because all residuals x_i are supposedly extracted from the same population of mean μ and variance σ^2 . In this case, χ^2 is referred to $N - 1$ degrees of freedom and is given by:

$$\chi^2 = \frac{(N-1)\text{SD}_x^2}{\sigma^2} \quad (3)$$

SD_x being the standard deviation of sample X and σ^2 the population variance. Let us now consider the following expressions concerning variable χ^2 :

$$P(\chi_{0.025}^2 \leq \chi^2 \leq \chi_{0.975}^2) = 0.95 \quad (4)$$

with $P(\chi^2 \leq \chi_{0.025}^2) = 0.025$ and $P(\chi^2 \leq \chi_{0.975}^2) = 0.975$

where the tabulated values for χ^2 corresponding to 0.975 and 0.025 can be found, for

instance, in Thompson (1941). Obviously equation (4) means the probability that χ^2 will be within the expressed range with a confidence level $\alpha=0.05$. Thus, rearranging equations (3) and (4), the following expression can be obtained:

$$P\left(\chi_{0.025}^2 \leq \frac{(N-1)SD_x^2}{MSE - \mu^2} \leq \chi_{0.975}^2\right) = 0.95 \quad (5)$$

In equation (5), MSE denotes the MSE population, whereas μ is the residual mean population which can be estimated from the sample mean \bar{X} . Therefore, a 95% confidence interval for MSE could be computed by:

$$\begin{aligned} MSE_{upper} &= \frac{(N-1)SD_x^2}{\chi_{0.025}^2} + \bar{x}^2 \\ MSE_{lower} &= \frac{(N-1)SD_x^2}{\chi_{0.975}^2} + \bar{x}^2 \end{aligned} \quad (6)$$

Equation (6) shows that the expected value for the MSE (theoretically true MSE) falls within the upper and lower limits indicated with a probability of 95%. For instance, if the sample size was $N=10$ (i.e. 9 degrees of freedom), the tabulated values for $\chi_{0.975}^2$ and $\chi_{0.025}^2$ would be 19.02 and 2.7, respectively.

2.2 Asymptotic approach to the development of the MSE confidence interval

Let us suppose a situation equal to that contemplated in the last section. A confidence interval for the MSE estimation can be computed from a sample extracted from a residual population with an unknown variance. We must bear in mind that random variable \bar{e} (equation (1)) does not have to follow a normal distribution. However, from the Central Limit Theorem, we can say that when N becomes large ($N \rightarrow \infty$), the standardized variable Y tends to a Student t distribution with $N - 1$ degrees of freedom.

$$Y = \frac{\bar{e} - \mu_{\bar{e}}}{SD_{\bar{e}}}; \text{ with } SD_{\bar{e}} = \frac{SD_e}{\sqrt{N}} \quad (7)$$

In equation (7), Y is the Student t variable, and SD_e is the standard deviation of the variable square of residual ($e_i = x_i^2$). It is known that the t distribution is similar to the normal distribution as long as the sample size is large enough. So, a $1 - \alpha$ percent confidence interval for Y would be:

$$P(-t_{\alpha/2} \leq Y \leq t_{\alpha/2}) = P\left(-t_{\alpha/2} \frac{SD_e}{\sqrt{N}} \leq \bar{e} - \mu_{\bar{e}} \leq t_{\alpha/2} \frac{SD_e}{\sqrt{N}}\right) = 1 - \alpha \quad (8)$$

Where $t_{\alpha/2}$ is the two-tail critical value corresponding to confidence level α for $N - 1$ degrees of freedom. The first term means the probability that Y will be within the range from $-t_{\alpha/2}$ to $t_{\alpha/2}$. Therefore, rearranging equation 8, the following expression can be obtained:

$$\begin{aligned} \mu_{\bar{e} \text{ upper}} &= \bar{e} + t_{\alpha/2} \frac{SD_e}{\sqrt{N}} \\ \mu_{\bar{e} \text{ lower}} &= \bar{e} - t_{\alpha/2} \frac{SD_e}{\sqrt{N}} \end{aligned} \quad (9)$$

Equation (9) shows that the expected mean for the MSE (theoretically true MSE) falls within the upper and lower limits indicated with a probability of $1 - \alpha$. Let us observe how limits are symmetrically located around the MSE computed from the sample because the pivotal quantity is exactly the same. This pivotal quantity can be interpreted as the specific degree of accuracy for the MSE estimation (Li 1991). Furthermore, and following Chebyshev's Theorem, the probability that \bar{e} differs from $\mu_{\bar{e}}$ can be made arbitrarily small by choosing a sufficiently large N .

2.3 Estimating function approach applied to the assessment of the DEM uncertainty

Since the 1980s, the theory of estimating functions has been widely applied in generalized linear statistical models, sampling theory, and biostatistics. One of the major advantages of this theory is that it can be applied in a variety of contexts: parametric, semi-parametric, or even non-parametric (Chan and Ghosh 1998). A non-parametric approach is used in this paper.

As defined by Godambe (1960) in his pioneering work, an estimating equation for parameter θ is a $g(X, \theta)=0$ equation which has only one solution $\theta=\theta(X)$ for every possible X , where X is a random variable. In our case, X is a sequence of N independent residuals computed at corresponding checkpoints for evaluating DEM accuracy. The function g is named the estimating function, and an optimal estimation of θ can be achieved considering the class of estimating functions as having zero expectation given by the following linear combination (Godambe and Thompson 1989):

$$g = \alpha \cdot h_1 + \beta \cdot h_2 \quad (10)$$

where h_1 and h_2 are real functions which would be determined by the underlying nature of the statistical problem, and α and β are real functions that can be estimated by the locally optimal solution found by Godambe and Thompson (1989) and shown in equation (11). For this, h_1 and h_2 must be orthogonal. That is to say, $E(h_1 \cdot h_2)=0$, where E expresses the mathematical expectation of the operation in parentheses.

$$\alpha^* = \frac{E(\frac{\partial h_1}{\partial \theta})}{E(h_1^2)}; \beta^* = \frac{E(\frac{\partial h_2}{\partial \theta})}{E(h_2^2)} \quad (11)$$

After this brief background information, let us apply the estimating function theory to our scenario where parameter θ would be the mean or expected value for MSE ($\mu_{\bar{e}}$ in equation (1)). We must remember that our objective is focused on the estimation of a confidence interval for the variable MSE as an average DEM error measure. This implies the construction of a normally distributed pivotal quantity involving both the sample data and the parameter to be estimated ($\mu_{\bar{e}}$). The basic estimating functions, supposing the third and fourth central moments are known (skewness, δ_1 , and standardized kurtosis, δ_2 , respectively), would be (Godambe and Thompson 1989):

$$h_1 = \bar{e} - \mu_{\bar{e}}; h_2 = (\bar{e} - \mu_{\bar{e}})^2 - \sigma_{\bar{e}}^2 - \gamma_{1\bar{e}} \sigma_{\bar{e}} (\bar{e} - \mu_{\bar{e}}) \quad (12)$$

σ^2 being the second central moment or variance of variable \bar{e} . It can be easily proved that h_1 and h_2 are orthogonal functions, as stated before. Therefore, the following

functions can be deduced for the estimation of α and β by means of equation (11):

$$\alpha^* = -\frac{1}{\sigma_{\bar{e}}^2}; \beta^* = \frac{\gamma_{1\bar{e}}\sigma_{\bar{e}}}{\sigma_{\bar{e}}^4(\gamma_{2\bar{e}} + 2 - \gamma_{1\bar{e}}^2)} \tag{13}$$

It follows that the estimating function g for $\mu_{\bar{e}}$ would be estimated by $g^* = \alpha^*h_1 + \beta^*h_2$. Because the distribution of the statistic g^*/σ_{g^*} can be approximated by a standardized normal distribution ($N(0,1)$), the following expression can be written:

$$P\left(\left|\frac{g^*}{\sigma_{g^*}}\right| \leq Z_\alpha\right) = 1 - \alpha \tag{14}$$

where Z_α is the limit value within which the values of the mentioned statistics will fall with a probability of $1 - \alpha$, α being the confidence level for a standardized normal distribution. Hence, $Z_\alpha = 1.64$ for $\alpha = 0.05$.

Equation (15) can be obtained operating from equation (14) and solving a quadratic unequation for the variable $\bar{e} - \mu_{\bar{e}}$. The algebra from which this equation is derived is rather tedious and has not been included in this paper, but the reader is referred to Godambe and Thompson (1989).

$$\begin{aligned} \mu_{\bar{e} \text{ lower}} &= \bar{e} + \frac{\frac{\gamma_{2\bar{e}} + 2}{\gamma_{1\bar{e}}} - \sqrt{\left(\frac{\gamma_{2\bar{e}} + 2}{\gamma_{1\bar{e}}}\right)^2 + 4\left(\frac{Z_\alpha \sqrt{(\gamma_{2\bar{e}} + 2)(\gamma_{2\bar{e}} + 2 - \gamma_{1\bar{e}}^2)}}{|\gamma_{1\bar{e}}|} + 1\right)}}{2} \sigma_{\bar{e}} \\ \mu_{\bar{e} \text{ upper}} &= \bar{e} + \frac{\frac{\gamma_{2\bar{e}} + 2}{\gamma_{1\bar{e}}} + \sqrt{\left(\frac{\gamma_{2\bar{e}} + 2}{\gamma_{1\bar{e}}}\right)^2 + 4\left(\frac{Z_\alpha \sqrt{(\gamma_{2\bar{e}} + 2)(\gamma_{2\bar{e}} + 2 - \gamma_{1\bar{e}}^2)}}{|\gamma_{1\bar{e}}|} + 1\right)}}{2} \sigma_{\bar{e}} \end{aligned} \tag{15}$$

Note that the pivotal quantity is asymmetrical around the mean, and it is necessary for estimating several parameters which depend on the distribution of random variable \bar{e} (MSE). These parameters can be estimated from a finite sample of N residuals (x_i) by means of the following expressions:

$$\sigma_{\bar{e}} = \frac{\sigma_e}{\sqrt{N}} \approx \frac{SD_e}{\sqrt{N}} \text{ with } SD_e = \sqrt{\frac{\sum_{i=1}^N (e_i - \bar{e})^2}{N}} \tag{16}$$

$$\gamma_{1\bar{e}} = \frac{E(\bar{e} - \mu_{\bar{e}})^3}{\sigma_{\bar{e}}^3} = \frac{m_{3\bar{e}}}{\sigma_{\bar{e}}^3} = \frac{\frac{m_{3e}}{N^2}}{\frac{\sigma_e^3}{N^{3/2}}} = \frac{m_{3e}}{\sigma_e^3 \sqrt{N}} = \frac{\gamma_{1e}}{\sqrt{N}} \tag{17}$$

$$\gamma_{2\bar{e}} = \frac{E(\bar{e} - \mu_{\bar{e}})^4}{\sigma_{\bar{e}}^4} - 3 = \frac{m_{4\bar{e}}}{\sigma_{\bar{e}}^4} - 3 = \frac{\frac{m_{4e}}{N^3} + \frac{3(N-1)}{N^3} \sigma_e^4}{\sigma_e^4} - 3 = \frac{\gamma_{2e}}{N} \tag{18}$$

where SD_e is the standard deviation of the random variable e (squared residual; see equation (1)), m_3 and m_4 denote the third and fourth central moments of the variable indicated as subindex, and finally γ_1 and γ_2 are the population skewness and standardized kurtosis of the variable also indicated as subindex. The skewness and

standardized kurtosis from the squared residual population can be estimated from a residual sample X of size N ($X = \{x_1, x_2, \dots, x_N\}$) according to the following equations:

$$\gamma_{1e}^* = \left(\frac{N}{(N-1)(N-2)\text{SD}_e^3} \sum_{i=1}^N (x_i^2 - \bar{e})^3 \right) \quad (19)$$

$$\gamma_{2e}^* = \left(\frac{N(N+1)}{(N-1)(N-2)(N-3)\text{SD}_e^4} \sum_{i=1}^N (x_i^2 - \bar{e})^4 \right) - \frac{3(N-1)^2}{(N-2)(N-3)} \quad (20)$$

It must be stressed that we need to estimate σ from the standard deviation. The use of standard deviation instead of σ for computing Z values from normal tables is known to lead to results which are not sufficiently reliable, especially when working with small samples (in practice lower than 100 observations). It is more appropriate to use the Student t distribution and then calculate t_z values for $N - 1$ degrees of freedom.

3. Experimental validation of the proposed model

3.1 Study sites and original datasets

The two study areas are located in Almería, south-east Spain. The first is situated in what is known as the ‘Comarca del Mármol’, in the municipal area of Macael. It is a zone of marble quarries with a high level of extraction activity, which has formed a terraced and artificial relief with a predominance of steep slopes and even vertical walls. The second study area is situated in the ‘Comarca del Campo de Níjar’, bordering on the ‘Cabo de Gata’ Nature Reserve. This is an area with a smooth relief sculpted by natural agents. For this study, seven topographic surfaces were selected measuring 198×198 m, two situated in Macael and five in Níjar. The morphological characteristics of topographic surfaces can be observed in table 1. The Standard Deviation of Unitary Vectors perpendicular to the topographic surface (SDUV), calculated as explained in Aguilar *et al.* (2006), was used as a roughness descriptor.

The DEM of each topographic surface was obtained automatically by stereo-image matching followed by manual editing. A more detailed explanation about the methodology and procedures used to obtain the original DEMs can be found in Aguilar *et al.* (2006). In this way, seven DEMs with a grid spacing of 2 m were obtained. The grid points were the UTM map projection (zone 30 North; *European Datum* 1950), and elevation data were stored as orthometric heights.

3.2 Experimental design

The experimental design was devised for checking the accuracy of the three methods proposed in the last section for constructing confidence intervals. The generation of the residual population from each morphology tested was achieved by means of two different procedures called procedure 1 and procedure 2. In both, residual populations uniformly distributed with a grid spacing of 2 m were generated to reflect the geographical area of interest.

3.2.1 Procedure 1. The residual population was obtained using the schedule showed in figure 1. Seed datasets for each morphology, composed of the X , Y , and Z

Table 1. Characteristics of the topographic surfaces studied. All surfaces are grid DEMs composed of 10 000 points and with 2×2 m spacing^a.

Terrain descriptive statistics	Nijar					Macael	
	Flat	Rolling1	Rolling2	Slightly mountainous	Mountainous	Steep-rugged hillside	Very rugged
Average elevation (m)	166.54	176.94	195.42	178.96	215.16	762.29	
$Z_{\max} - Z_{\min}$ (m)	6.7	17.25	23.08	34.43	45.17	201.32	116.48
Elevation coefficient of variation (%)	0.97	2.28	2.2	3.79	3.98	6.52	4.28
Average slope (%)	3.3	9.27	10.01	19.42	31.18	82.14	65.12
Slope coefficient of variation (%)	48.02	45.66	69.59	58.62	38.27	30.91	77.01
SDUV (m)	0.03	0.09	0.1	0.19	0.31	0.35	0.64

^aThe standard deviation of unitary vectors perpendicular to topographic surface is denoted by SDUV.

coordinates of 128 ground points, were extracted from each original grid DEM by stratified random sampling (four-by-four sampling quadrants). This guarantees a homogenous distribution of the seed data over the whole working area. Based on these seed datasets, a DEM with a grid spacing of 2 m was generated for each morphology using the Inverse Distance Weighted (IDW) interpolation method from the six nearest neighbours. The IDW is a well-known interpolation method where weighting is assigned to data using the inverse separation distance to a power as a weighting function. A weighting power parameter of 2 was used in our case.

Since the interpolated and original grid DEMs are spatially coincident and equal in spacing, it is easy to compute the height differences between both DEMs at each node to obtain a set of 10 000 residuals which will be considered as the residual population from procedure 1.

3.2.2 Procedure 2. The residual population for each morphology was obtained by means of the algorithm shown in figure 2. The interpolation schedule, employing the six closest neighbours from adjacent columns, guarantees that the separation distance between sample points and interpolated points takes values from 2 to 2.82 m (figure 3).

The interpolation method, Multiquadric Radial Basis Function (MRBF), was applied because it is suitable for interpolating from high-resolution DEMs and performs very well with a local support. The selected value for the smoothing factor was always zero because it has been shown to give the best results (Aguilar *et al.* 2005a).

Note that in procedure 1, the separation distance between the interpolated points and the set of local support points will be greater and more variable than in procedure 2, and so the interpolation errors will be greater. At the same time, the interpolation method employed in procedure 1 is in more frequent use for DEM construction than the one used in procedure 2, and so procedure 2 is more artificial than procedure 1.

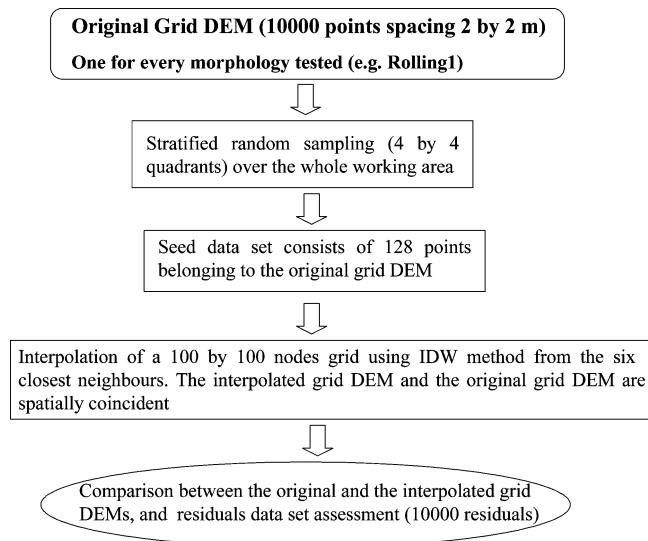


Figure 1. Flow chart of the scheme used to obtain the different residuals datasets in the procedure 1.

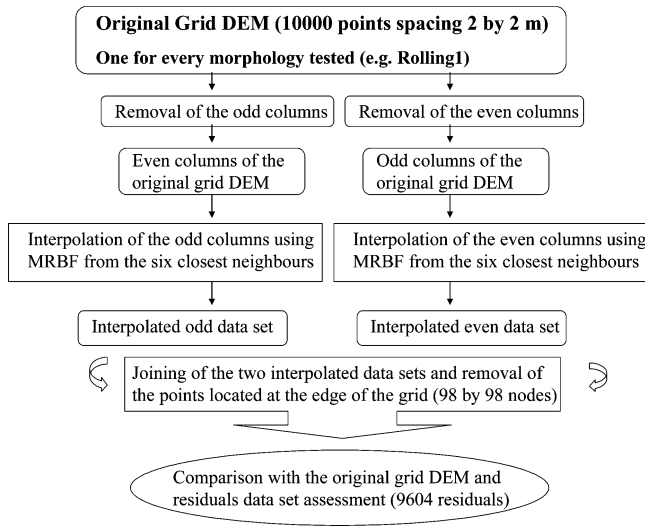


Figure 2. Flow chart of the scheme used to obtain the different residuals datasets in the procedure 2.

Tables 2 and 3 show some statistics of the residual population for each morphology after procedures 1 and 2 were applied. The non-normal nature of the datasets from procedure 2 is quite evident, whereas the datasets provided by procedure 1 seem to approximate a normal distribution, although they do not strictly follow it either.

3.2.3 Generating confidence intervals using the Monte Carlo simulation method. The Monte Carlo simulation method was applied to investigate the accuracy with which the proposed models describe the simulated data. This kind of numerical simulation is commonly used for investigating error propagation and

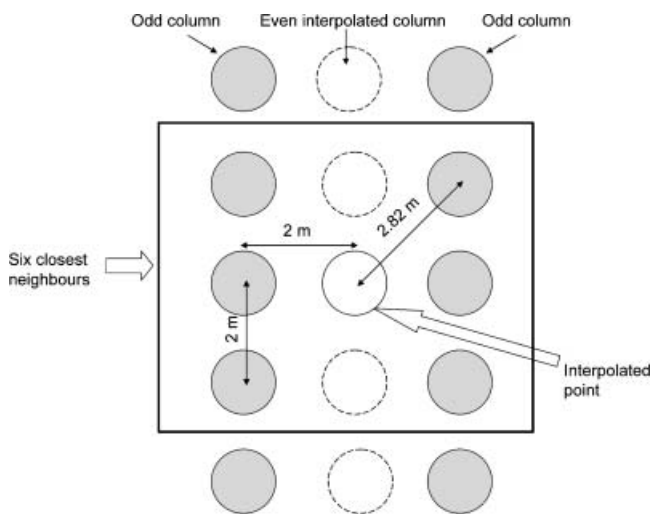


Figure 3. Scheme of the interpolation used in the procedure 2. Even columns have been interpolated from odd original grid DEM columns.

Table 2. Statistics of the residuals population for every morphology tested (Procedure 1: Inverse Distance Weighted interpolator)^a.

Morphology	μ (cm)	σ (cm)	Standardized		K-S statistic
			Skewness (γ_1)	Kurtosis (γ_2)	
Mountainous	-2.15	166.84	-0.49	1.67	0.064
Rolling 1	1.59	52.08	0.54	2.5	0.071
Flat	-1.65	16.75	-0.69	3	0.08
Steep rugged hillside	-68.44	448.56	-0.64	1.81	0.069
Very rugged	99.45	884.57	1.08	4.73	0.17
Slightly mountainous	-18.57	122.32	-0.49	2.9	0.107
Rolling 2	6.24	53.06	-0.44	4.93	0.081

^aK-S denotes the Kolmogorov–Smirnov goodness-of-fit statistic for a normal probability distribution (Royston 1982). The critical value of the K-S statistic at 95% confidence level for all populations is 0.014.

cartographic uncertainty, because it is easily implemented and generally applicable without using analytical equations (Heuvelink *et al.* 1989, Weir 2002, Ariza and Atkinson 2005).

The algorithm starts by extracting a sample X of size N from each different dataset of residual population (seven morphologies and two procedures) by means of random sampling, N taking the values 16, 32, 64, 128, 192, 288, 384, 576, and 960 checkpoints. Next the parameters related to each of the three models tested were computed from sample X and used for estimating the upper and lower limits of a 95% confidence interval for MSE (MSE_{upper} and MSE_{lower}). The population MSE (μ_e), for every morphology and procedure was calculated from all the residual values available in the dataset. Afterwards, it was necessary to confirm whether the corresponding MSE population fell within the limits of the calculated confidence intervals and whether the model had correctly predicted the expected uncertainty for the MSE estimation. This procedure was repeated up to 1000 times to compute the percentage of cases in which the MSE population was correctly situated within the theoretically established confidence interval bounds.

Finally, because we are extracting samples from finite populations, a correcting coefficient was applied to calculate the MSE variance in order to correct the reduced variability expected on consecutive random samplings extracted from the same finite

Table 3. Statistics of the residuals population for every morphology tested. Procedure 2: Multiquadric Radial Basis Function interpolator.

Morphology	μ (cm)	σ (cm)	Standardized		K-S statistic
			Skewness (γ_1)	Kurtosis (γ_2)	
Mountainous	0.02	11.16	0.39	12.18	0.158
Rolling 1	0.01	2.72	0.84	13.2	0.178
Flat	0.01	2.08	0.64	23.99	0.173
Steep rugged hillside	0.48	41.01	0.6	21.55	0.155
Very rugged	-0.87	135.02	0.12	31.95	0.224
Slightly mountainous	0.12	6.36	1.12	21.12	0.194
Rolling 2	-0.08	1.84	-0.37	29.66	0.196

^aK-S denotes the Kolmogorov–Smirnov goodness-of-fit statistic for a normal probability distribution (Royston 1982). The critical value of the K-S statistic at 95% confidence level for all populations is 0.014.

residual population (Steel and Torrie 1980). In equation (21), M represents the total number of residuals for each population (10 000 and 9604 for procedures 1 and 2, respectively), and N is the size of the random samplings.

$$\sigma_{\bar{e}}^2 = \frac{\sigma_e^2}{N} \left(\frac{M-N}{M} \right) \quad (21)$$

4. Results and discussion

4.1 Characteristics of the residual populations

Table 2 shows the statistics of the residual populations in the case of the procedure 1. The standard deviation (σ) can be seen to depend on the roughness of each morphology, lower values for flat terrain and higher values for very rugged ones.

The skewness values, in general, can be considered to be close to those of a normal distribution which could be accepted with absolute values lower than 0.5 (Daniel and Tennant 2001). At the same time, the mean values are very close to zero, except in the case of the very rugged terrain of Macael, indicating few systematic errors.

The values taken by the standardized kurtosis are greater than zero in all cases, indicating that the residual distribution is slightly leptokurtic for all morphologies. Leptokurtosis indicates that there are more occurrences far away from the mean than predicted by a standard normal distribution, probably because of the presence of outliers or gross errors. In other words, the probability distribution function peaks at the centre and has thicker tails. In fact, the values taken by the Kolmogorov–Smirnov statistics are higher than the critical value (Table 2), and thus, the hypothesis of the normal distribution of the underlying population should be rejected.

Table 3 shows some statistics of the residual population for the seven morphologies studied in the case of procedure 2. While the skewness is still similar to that observed in procedure 1, the standardized kurtosis is much greater, and therefore the statistical distribution of residual populations can be considered as clearly non-normal. This can also be checked observing the high values taken by the Kolmogorov–Smirnov statistics. Again, the presence of mean values lower than the standard deviations indicates the absence of systematic errors.

Note that the presence of outliers can corrupt the true statistical distribution of errors. From a statistical point of view, they cannot be considered as belonging to the same population as the other observations. Therefore, we have several datasets of clearly non-normal residual populations, especially in the case of the data generated from procedure 2. The testing of that data set is carried out in order to show the performance of the new proposed method based on the estimating function approach working with very non-normal residual distributions, because a theoretical model without practical restrictions is always more reliable and sound than a model subject to certain assumptions.

4.2 Model performance evaluation

Figures 4 and 5 present the results offered by the Monte Carlo numerical simulation. There is little doubt that the Chi-squared approach does not work as well as the other methods, above all when it was evaluated over datasets generated from procedure 2 with very non-normal residual populations. Thus, the percentage of

simulations which fall within the theoretical bounds ranged approximately between 80% for procedure 1 (figure 4), and only 35% to 55% for procedure 2 (figure 5). Let us remember that the statistically expected percentage should be 95%. This poor performance is probably due to the fact that the Chi-squared distribution is extremely sensitive to the normality assumption, and so its applicability to non-normal distributions is limited. However, its score is quite independent of the sample size, which is a useful behaviour.

As depicted in figures 4 and 5, the asymptotic approach works reasonably well on slightly non-normal residual populations like those from procedure 1, but worse results are reached in the case of procedure 2. In fact, even if the population is non-normal but bell-shaped, the t distribution could be a good approximation. Note also that the higher the number of checkpoints used, the greater the precision obtained for the calculation of the MSE confidence interval. Indeed, the distribution of sample means derived from a large size random sampling tended towards a normal distribution, although the main distribution was clearly non-Gaussian. Moreover, it has been necessary to estimate σ_e from a more or less reduced sample, which always implies a fair degree of uncertainty with regard to estimation. This is more evident when highly leptokurtic populations are sampled, as in the case of procedure 2 (figure 5) and morphologies such as Rolling 2, Very Rugged, or Steep-rugged Hillside. Over these kinds of data, many more checkpoints had to be taken in order to come close to the theoretical target value of 95%. In some cases, it was impossible to even reach that value for the range of the sample sizes tested.

Lastly, the third method based on the estimating function approach can be considered the most applicable in the majority of circumstances. Since it takes into account the first four central moments of the original population, it is free of any distributional assumption. However, the performance of the model could be highly influenced by the accuracy of skewness and standardized kurtosis estimations from small leptokurtic samples (Aguilar *et al.* 2005b). Therefore, all that is necessary is to correctly estimate the third and fourth central moments from equations (19) and (20) respectively, because the mean and variance are more robustly estimated from much smaller samples.

Figure 4 shows good results with regard to the performance of the model applied to datasets from procedure 1 (weak leptokurtosis). In all the morphologies, a low number of 32 checkpoints, and even of 16, was enough to construct an accurate 95% confidence interval. It should be noted that checkpoints should be approximately three times more accurate than the expected accuracy to be verified (FGCC 1984), and so it may be costly to obtain a large number of checkpoints for the DEM quality-assessment process (Li 1991). Thus, the proposed method could be of benefit to terrain analysts compared with the other two methods tested. Indeed, both the number of checkpoints needed to guarantee the goodness-of-error estimation and the uncertainty of that estimation can be decreased. In other words, producers could be very certain about the confidence level (for instance 95%) of the DEM error reported to the customer or DEM user, avoiding a probable overestimate of the reliability of their vertical accuracy assessment.

With regard to data from procedure 2 (high leptokurtosis and probably a more unrealistic scenario), figure 5 shows a worse behaviour on the part of the proposed model, requiring a greater number of checkpoints to come close to the stated confidence value of 95%. This feature is more pronounced in the case of terrains where the presence of outliers is to be expected, such as Rolling 2, Very rugged, or

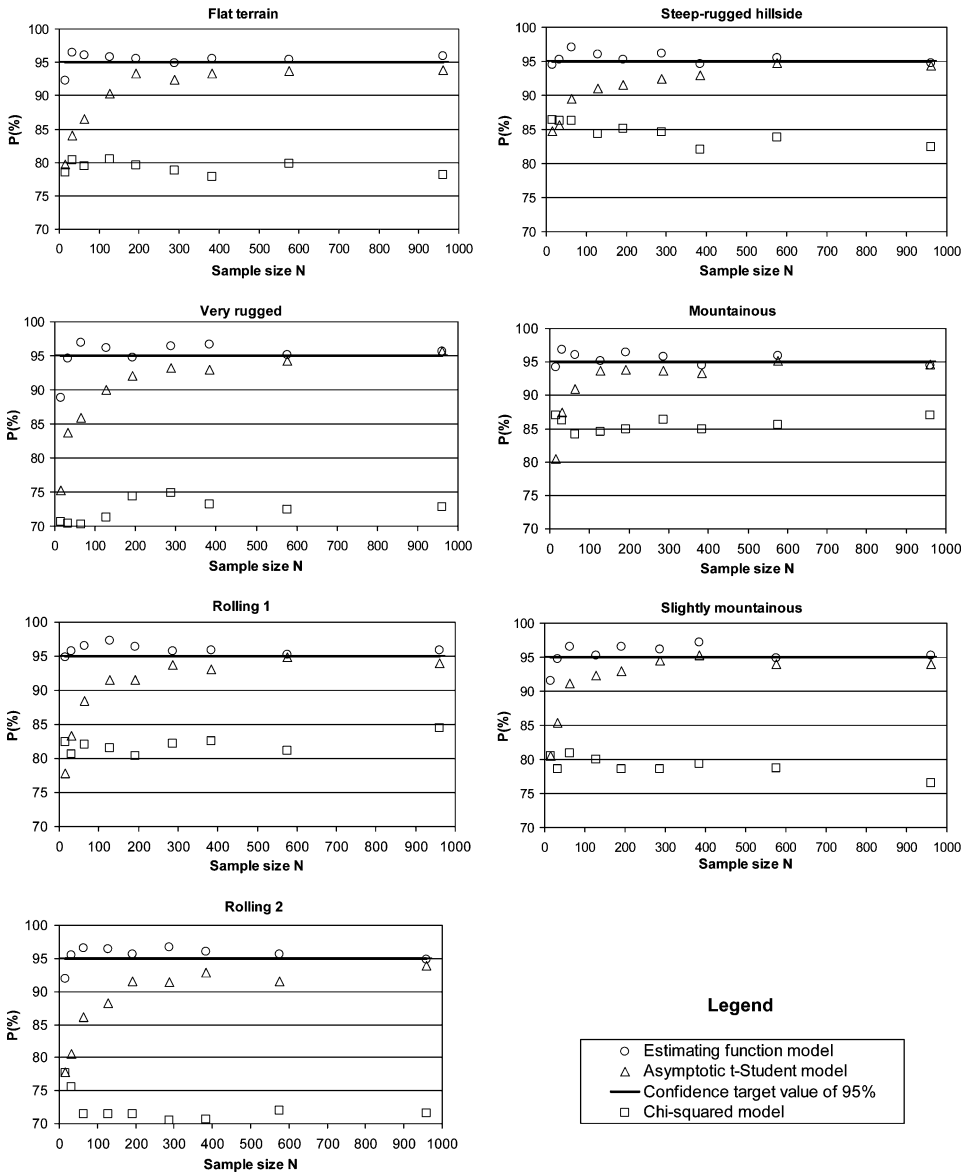


Figure 4. Comparison between the accuracy of the 95% confidence intervals constructed from Monte Carlo simulated data from procedure 1 using the Estimating function, Asymptotic Student t and Chi-squared approaches. P represents the percentage of simulated data registered within each calculated confidence interval.

Flat. With respect to this, the number of checkpoints which is sufficient for constructing the corresponding confidence interval successfully could be situated roughly between 64 and 128. Without doubt, this can be considered a reasonably low number if we take into consideration the artificial and non-normal nature of the original populations.

The main reason for these findings must be attributed to the fact that the estimating function approach presents wider confidence intervals than either the Chi-squared or the Student t approaches (figure 6). It is basically due to the

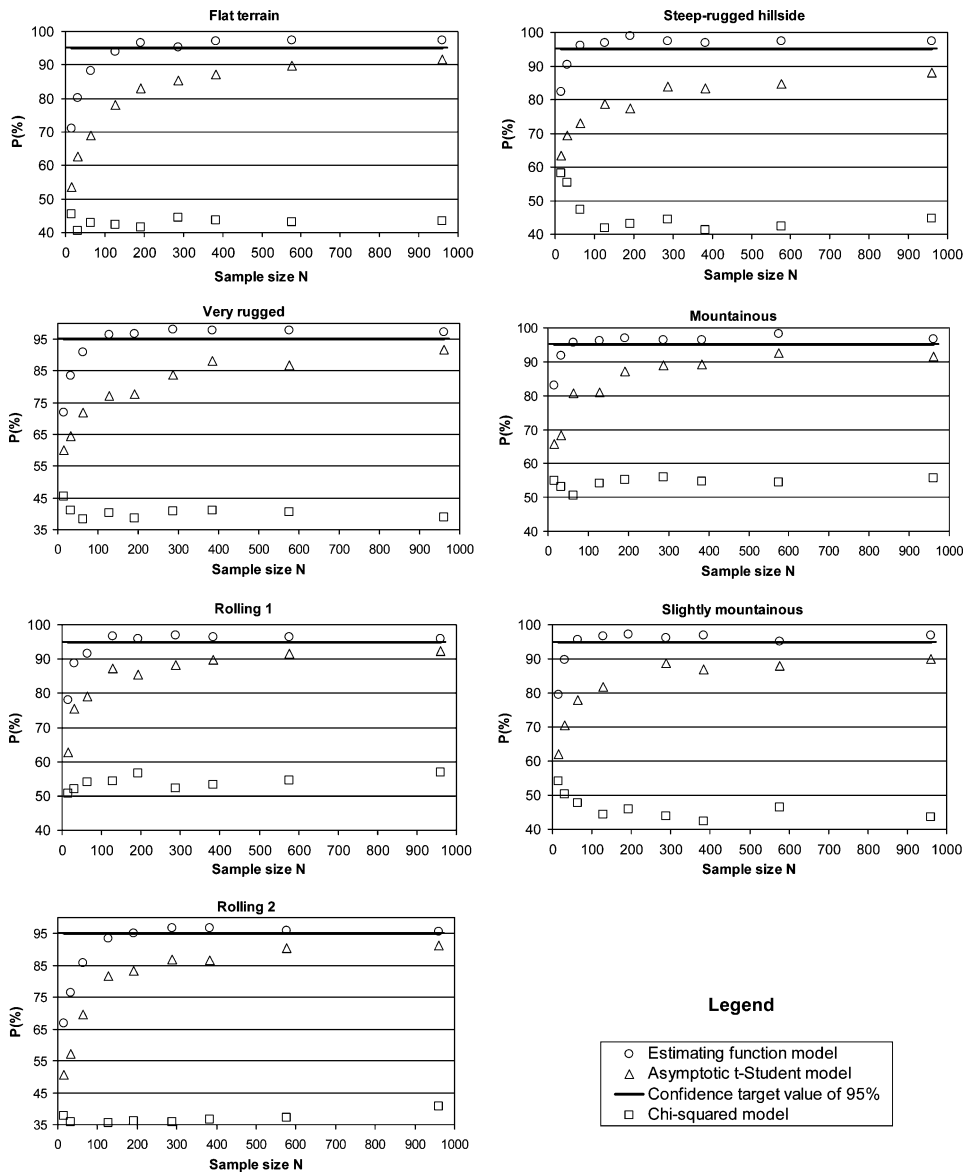


Figure 5. Comparison between the accuracy of the 95% confidence intervals constructed from Monte Carlo simulated data from procedure 2 using Estimating function, Asymptotic Student t and Chi-squared approaches. P represents the percentage of simulated data registered within each calculated confidence interval.

introduction of standardized kurtosis in the model, which described adequately the leptokurtosis scenario produced by the presence of anomalous values in the residual dataset. Conversely, the Chi-squared and Student t approaches, the latter to a lesser degree, assume the normal distribution of the residual population and so tend to overestimate the reliability of the uncertainty evaluation computing more narrow confidence intervals. In fact, a more important advantage of the proposed estimating function approach is its ability to detect outliers via standardized kurtosis (non-normal residual distributions) and so automatically expand the

confidence limits to guarantee that the reliability of the uncertainty evaluation always corresponds to the chosen confidence level. The new method is also sympathetic to the statistical nature of the residual distribution and so adapts its response to agree with the expected confidence level.

As we can see in figure 6, all the models tested present an asymptotic behaviour with regard to the sample size or number of checkpoints. That is to say, the higher the number of checkpoints used, the greater the accuracy that will be logically obtained for the calculation of the average error measure (MSE), and so the corresponding confidence interval will be narrower.

Finally, and from a theoretical point of view for the three models tested, we have supposed that residuals are absolutely independent, i.e. they are not correlated. However, spatial autocorrelation is a property of DEM error which has been reported by several authors (López 1997, Fisher 1998, Weng 2002). Note that if the error at the checkpoints is in fact not independent, the use of N spatially autocorrelated residuals would lead to a variability of the error lower than it should be, moving prediction intervals from a hypothetical leptokurtic profile to a platykurtic one. Thus, spatial autocorrelation may be defined as a measure of the true albeit masked information content in spatial data (Garrott 2004).

Since our residual populations presented a short-range spatial structure (Aguilar *et al.* 2005b), it is unlikely that the proposed models have been affected by this phenomenon because we are working with relatively small samples, and so the separation distance between the majority of checkpoints would be greater than the spatial structure range. Therefore, it would be possible to treat the observations as uncorrelated. On the other hand, if long-range residual autocorrelation exists, we could have difficulty in applying them without any type of correction. In a further study, we will explore, among other aspects, the performance of the estimating function approach with the sample size reduction proposed by Cressie (1993) from the following equation:

$$N' = \frac{N}{1 + 2\left(\frac{\rho}{1-\rho}\right)\left(\frac{N-1}{N}\right) - 2\left(\frac{\rho}{1-\rho}\right)^2(1-\rho^{N-1})\frac{1}{N}} \quad (22)$$

where ρ is the coefficient of the positive spatial autocorrelation, N is the original sample size and N' is the corrected original sample size ($N' \leq N$), which can be interpreted as the equivalent independent number of observations.

5. Conclusions

The theoretical approach to the estimation of confidence intervals for DEM average-error statistics (e.g. MSE) can be successfully achieved using the theory of estimating functions. It could be considered much more general than other classical approaches, such as the Chi-squared or Student t approximations, because it avoids the strong assumption of the normal distribution of the residuals. In fact, the estimating function approach is based on a non-parametric approximation where no particular distribution assumption is made.

The main reason for this finding must be attributed to the fact that the estimating function approach presents wider confidence intervals than either Chi-squared or Student t approaches. This is basically due to the introduction of the standardized kurtosis of the sum of residual squares in the model, which correctly describes the scenario of leptokurtosis produced by the presence of outliers in the residual dataset.

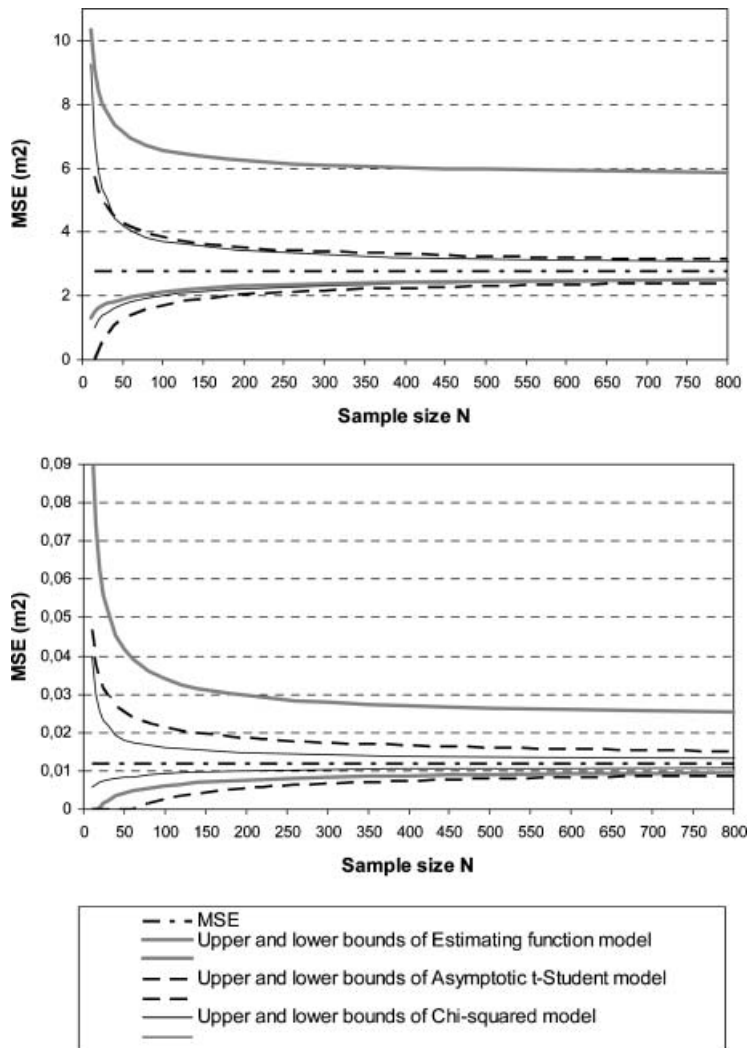


Figure 6. Graphical representation of the Mean Square Error (MSE) 95% confidence intervals constructed from the Estimating function model, Asymptotic Student t model, and Chi-squared model. The dashed and dotted horizontal line indicates the population MSE. All cases plotted concern Mountainous terrain. Top (data from procedure 1): $MSE=2.78\text{ m}^2$; $\sigma_e=5.35\text{ m}^2$; $\gamma_{1e}=3.86$; $\gamma_{2e}=19.28$. Bottom (data from procedure 2): $MSE=0.012\text{ m}^2$; $\sigma_e=0.047\text{ m}^2$; $\gamma_{1e}=9.13$; $\gamma_{2e}=112.58$.

Furthermore, the skewness and standard deviation of the sum of residual squares are also included. Note that skewness can be very useful to describe very biased populations with systematic errors.

It must be pointed out that the only uncertainty for the construction of confidence intervals associated with the estimating functions approach turns out to be the accurate estimation of those three statistics when they are calculated from small sample sizes. Thus, it has been proved that, even working with highly leptokurtic populations, the performance of the estimating function model is very good, and so the statistical inference process from samples is suitably realized. In fact, in the worst-case scenario 64–128 checkpoints are needed for constructing an MSE with an

accurate 95% confidence interval. Nevertheless, when populations tend to be normal, a good approximation can be reached with only 16–32 checkpoints.

From the perspective of DEM quality evaluation, the results obtained from the present work are very promising and can be considered highly applicable in current practice, especially taking into account the variability of tested morphologies and the extremely non-normal nature of residual populations generated from procedure 2.

Acknowledgements

The authors are grateful to the Consejería de Agricultura y Pesca of the Andalusia Government for financing part of this work through Project: ‘Basic cartography for the development of the network irrigation project in Campo de Níjar (Almería)’. Thanks are also due to the Centro Tecnológico de la Piedra de Macael for its contribution through project: ‘Production and compilation of digital cartography for planning and exploitation of mining resources’. The authors would like to thank three anonymous reviewers whose comments have notably improved the manuscript. Finally, the authors are particularly grateful to Professor Peter Fisher for his valuable and helpful comments on the original manuscript.

References

- AGUILAR, F.J., AGÜERA, F., AGUILAR, M.A. and CARVAJAL, F., 2005a, Effects of terrain morphology, sampling density and interpolation methods on grid DEM accuracy. *Photogrammetric Engineering & Remote Sensing*, **71**, pp. 805–816.
- AGUILAR, F.J., AGÜERA, F., AGUILAR, M.A. and CARVAJAL, F., 2005b, Modelling the effect of the number of checkpoints in the accuracy assessment of digital elevation models. In *Proceedings of the 21th International Cartographic Conference* (A Coruña: International Cartographic Association), CD-ROM Proceedings.
- AGUILAR, F.J., AGUILAR, M.A., AGÜERA, F. and SÁNCHEZ, J., 2006, The accuracy of grid digital elevation models linearly constructed from scattered sample data. *International Journal of Geographical Information Science*, **20**(2), pp. 169–192.
- ARIZA, F.J. and ATKINSON, A.D.J., 2005, Sample size and confidence interval when applying the NSSDA. In *Proceedings of the 21th International Cartographic Conference* (A Coruña: International Cartographic Association), CD-ROM Proceedings.
- CHAN, S. and GHOSH, M., 1998, Orthogonal projections and the geometry of estimating functions. *Journal of Statistical Planning and Inference*, **67**, pp. 227–245.
- CRESSIE, N., 1993, *Statistics for Spatial Data* (New York: Wiley).
- DANIEL, C. and TENNANT, K., 2001, DEM quality assessment. In *Digital Elevation Models and Applications: The DEM User's Manual*, D.F. Maune (Ed.), pp. 395–440 (Bethesda, MD: ASPRS).
- FGCC 1984, Standards and Specifications for Geodetic Control Networks. US Federal Geodetic Control Committee, Rockville, MD. Available online at: http://www.ngs.noaa.gov/FGCS/tech_pub/1984-stds-specs-geodetic-control-networks.htm (accessed 4 December 2006).
- FGDC 1998, Geospatial Positioning Accuracy Standards Part 3: National Standard for Spatial Data Accuracy (Reston, VA: US Federal Geographic Data Committee). Available online at: <http://www.fgdc.gov/FGDC-standards-projects/accuracy/part3/chapter3> (accessed 4 December 2006).
- FISHER, P.F., 1998, Improved modelling of elevation error with geostatistics. *GeoInformatica*, **2**(3), pp. 215–233.
- GARROTT, J., 2004, SAKWeb. Spatial autocorrelation and kriging Web. A W3 geocomputation perspective. PhD thesis, Universidad Nova de Lisboa, Portugal.
- GODAMBE, V.P., 1960, An optimum property of regular maximum likelihood equation. *Annals of Mathematical Statistics*, **31**, pp. 1208–1211.

- GODAMBE, V.P. and THOMPSON, M.E., 1989, An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference*, **22**, pp. 137–152.
- GRIFFITH, D., 1996, Some guidelines for specifying the geographical weights contained in spatial statistical models. In *Practical Handbook of Spatial Statistics*, S. Arlinghaus (Ed.), pp. 65–82 (Boca Raton, FL: CRC Press).
- HEUVELINK, G.B.M., BURROUGH, P.A. and STEIN, A., 1989, Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems*, **3**, pp. 303–322.
- HUANG, Y.D., 2000, Evaluation of information loss in digital elevation models with digital photogrammetric systems. *Photogrammetric Record*, **16**, pp. 781–791.
- JAKOBSSON, A., 2002, Data quality and quality management. Examples of quality evaluation procedures and quality management in European National Mapping Agencies. In *Spatial Data Quality*, W. Shi, P.F. Fisher and M.F. Goodchild (Eds), pp. 216–229 (London: Taylor & Francis).
- LEY, R.G., 1986, Accuracy assessment of digital terrain models. In *Proceedings of AutoCarto*, pp. 455–564 (London: International Cartographic Association).
- LI, Z., 1988, On the measure of digital terrain model accuracy. *Photogrammetric Record*, **12**, pp. 873–877.
- LI, Z., 1991, Effects of checkpoints on the reliability of DTM accuracy estimates obtained from experimental tests. *Photogrammetric Engineering & Remote Sensing*, **57**, pp. 1333–1340.
- LÓPEZ, C., 1997, Locating some types of random errors in digital terrain models. *International Journal of Geographical Information Science*, **11**, pp. 677–698.
- MAUNE, D.F., BINDER MAITRA, J. and MCKAY, E.J., 2001a, Accuracy standards. In *Digital Elevation Models and Applications: The DEM User's Manual*, D.F. Maune (Ed.), pp. 61–82 (Bethesda, MD: ASPRS).
- MAUNE, D.F., BLACK, T.A. and CONSTANCE, E.W., 2001b, DEM user requirements. In *Digital Elevation Models and Applications: The DEM User's Manual*, D.F. Maune (Ed.), pp. 441–460 (Bethesda, MD: ASPRS).
- ROYSTON, J.P., 1982, Expected normal order statistics (exact and approximate). *Applied Statistics*, **31**, pp. 161–165.
- SHI, W.Z. and BEDARD, Y., 2004, Theme issue: advanced techniques for analysis of geospatial data. *ISPRS Journal of Photogrammetry & Remote Sensing*, **59**, pp. 1–5.
- STEEL, R.G.D. and TORRIE, J.H., 1980, *Principles and Procedures of Statistics. a Biometrical Approach*, 2nd edition (New York: McGraw-Hill).
- THOMPSON, C.M., 1941, Table of percentage points of the χ^2 distribution. *Biometrika*, **32**, pp. 188–189.
- WEIR, M., 2002, Monte Carlo simulation of long-term spatial error propagation in forestry databases. In *Spatial Data Quality*, W. Shi, P.F. Fisher and M.F. Goodchild (Eds), pp. 294–303 (London: Taylor & Francis).
- WENG, Q., 2002, Quantifying uncertainty of digital elevation models derived from topographic maps. In *Advances in Spatial Data Handling*, D. Richardson and P. van Oosterom (Eds), pp. 403–418 (New York: Springer).