

## ACCURACY ASSESSMENT OF LIDAR-DERIVED DIGITAL ELEVATION MODELS

FERNANDO J. AGUILAR (faguilar@ual.es)  
*University of Almeria, Spain*

JON. P. MILLS (j.p.mills@newcastle.ac.uk)  
*Newcastle University*

(Based on a contribution to the Annual Conference of the Remote Sensing and Photogrammetry Society at Newcastle upon Tyne, 14th September 2007)

### *Abstract*

*Despite the relatively high cost of airborne lidar-derived digital elevation models (DEMs), such products are usually presented without a satisfactory associated estimate of accuracy. For the most part, DEM accuracy estimates are typically provided by comparing lidar heights against a finite sample of check point coordinates from an independent source of higher accuracy, supposing a normal distribution of the derived height differences or errors. This paper proposes a new methodology to assess the vertical accuracy of lidar DEMs using confidence intervals constructed from a finite sample of errors computed at check points. A non-parametric approach has been tested where no particular error distribution is assumed, making the proposed methodology especially applicable to non-normal error distributions of the type usually found in DEMs derived from lidar. The performance of the proposed model was experimentally validated using Monte Carlo simulation on 18 vertical error data-sets. Fifteen of these data-sets were computed from original lidar data provided by the International Society for Photogrammetry and Remote Sensing Working Group III/3, using their respective filtered reference data as ground truth. The three remaining data-sets were provided by the Natural Environment Research Council's Airborne Research and Survey Facility lidar system, together with check points acquired using high precision kinematic GPS. The results proved promising, the proposed models reproducing the statistical behaviour of vertical errors of lidar using a favourable number of check points, even in the cases of data-sets with non-normally distributed residuals. This research can therefore be considered as a potentially important step towards improving the quality control of lidar-derived DEMs.*

KEYWORDS: accuracy assessment, confidence intervals, DEM, lidar

### INTRODUCTION

ACCURATE DIGITAL ELEVATION MODELS (DEMs) of high spatial resolution from airborne lidar data are in increasing demand for a growing number of mapping and GIS tasks related to a wide variety of applications including forest management, urban planning, bird population

modelling, ice sheet mapping, flood control and road design (Lim et al., 2003). Lidar DEMs are also being increasingly used for new applications relating to change detection and geopositioning (James et al., 2006; Rodarmel et al., 2006; Miller et al., 2007). However, despite the growing range of applications, and in spite of the relatively high cost of this new type of digital product, such DEMs are usually presented without an appropriate associated estimate of their accuracy. Indeed, with a popular estimate being that 80% of all data used by managers and decision-makers is spatially referenced (van Oort and Bregt, 2005), many researchers have stressed the need to deal with issues of spatial quality. A recent, comprehensive review regarding the importance of assessing error in DEMs can be found in Fisher and Tate (2006).

In most cases, users of lidar-derived DEMs are content with some nominal specifications about raw data accuracy, as supplied by the instrument manufacturer or the data provider. Such statistics are often inappropriate given that the major component of the total error could well be the error introduced during the processing of the raw data (data filtering, gridding, segmentation and/or object reconstruction), rather than the error induced during the data capture process (which is related to vertical and horizontal error in the positioning of the laser platform, laser scan angle, surface reflectivity and terrain slope). In fact, gridding error can comprise a very important, and often neglected, source of error which should be taken into account (Smith et al., 2005; Aguilar et al., 2006). For example, it may be especially relevant for canopy height model estimation in forestry applications where there are fewer ground-return samples for effective DEM surface interpolation (Lim et al., 2003; Clark et al., 2004; Hopkinson et al., 2004; Su and Bork, 2006).

Although there have been a number of studies dealing with lidar-derived DEM error, most can be classified as empirical work in which the influence of different variables on DEM error was analysed (Hopkinson et al., 2004; Hyypä et al., 2005; Goodwin et al., 2006; Göpfert and Heipke, 2006; Su and Bork, 2006). In practice, statistical tests are the only way to ensure that requirements are met at a moderate cost by using an inference process based on sampling theory. At present, the great majority of DEM accuracy standards are based on computing the vertical accuracy of a finite sample data-set (check points) from the differences (residuals) between data-set heights and height values from an independent source of higher accuracy, usually obtained by differential GPS techniques and supposing those residuals follow a normal distribution. This is the case, for example, for the "National Standards for Spatial Data Accuracy" (NSSDA) document published by the US Federal Geographic Data Committee (FGDC, 1998). However, it should be recognised that vertical errors in lidar DEMs often follow a non-normal distribution when they are captured over non-open terrain (for instance, over vegetated or built-up areas) and, therefore, both filtering and gridding can introduce a large quantity of non-random noise (Flood, 2004; Hopkinson et al., 2004; Göpfert and Heipke, 2006), presenting high kurtosis and skewness due to the presence of systematic errors and outliers. The presence of systematic errors is relatively frequent in lidar-derived DEMs of forest areas where non-ground samples (such as low vegetation and logs) are included in the DEM interpolation process and the resulting mean error tends to be positive. In such cases a lidar-derived DEM tends to overestimate the reference elevation (Clark et al., 2004).

Concerned about the likely non-normal nature of lidar-derived DEM error under non-open terrain, the American Society for Photogrammetry and Remote Sensing (ASPRS), through the work of the ASPRS Lidar Committee, recently approved a new set of guidelines (Flood, 2004). The guidelines recommend that for open terrain, where error distribution is thought to be close to a normal distribution, the methodology proposed by the NSSDA should be adopted. That is, for assessing vertical accuracy at 95% confidence level:

$$\text{Vertical accuracy} = 1.96\text{rmse}_z \quad (1)$$

where rmse is the root mean square error. The resulting figure is known as the “fundamental” vertical accuracy. Away from open terrain, the fundamental vertical accuracy is replaced by the so-called “supplemental” vertical accuracy, tested using the 95th percentile method. The 95th percentile method may be used regardless of whether or not the errors follow a normal distribution and whether or not errors qualify as outliers, indicating that 95% of the errors in the data-set will have absolute values of equal or lesser value and 5% of the errors will be of larger value. Regarding the number of necessary check points, the ASPRS recommends a minimum of 20 check points (30 preferred) in each of the major land cover categories represented in the surveyed area. This gives rise to the question of exactly how many check points should be used in order to guarantee that a certain confidence level is achieved. Conversely, it may be necessary to derive the reliability of fundamental and supplemental vertical accuracy estimates when a particular number of check points is used. Deriving definitive values is not easy, and the ASPRS guidelines do not offer advice with respect to these issues.

Addressing the aforementioned issues, the main objective of this paper is the development of a theoretical methodology to assess lidar DEM vertical accuracy. Vertical accuracy is estimated by means of the construction of confidence intervals from a finite sample of residuals computed at check points evenly distributed across the whole survey area. One of the goals is to derive the minimum number of check points needed to guarantee a certain confidence level. In this way, a non-parametric approach based on the theory of estimating functions has been adopted and tested. This approach means that no particular distribution of residuals at check points is assumed and so it is applicable to non-normal error distributions which are usually found in lidar-derived DEMs.

#### DEVELOPMENT OF THE PROPOSED MODELS

The models tested in the work reported here concern two different scenarios:

- (1) Construction of confidence intervals relating to the distribution of lidar vertical errors from a finite sample of residuals coming from a population not necessarily expected to be normally distributed.
- (2) Construction of confidence intervals for the assessment of an average error statistic for which the meaning will easily be understood, both on the part of the user and the producer of lidar-derived DEMs. In this case rmse, as a widely used average error statistic, was chosen.

Both of the contemplated scenarios avoid the very common assumption of normally distributed errors, as adopted in the majority of current vertical accuracy standards.

#### *Model for Estimating Lidar Vertical Accuracies*

Suppose that the difference between a lidar-derived DEM and ground truth is a random variable  $X$ . This being the case, a sample of size  $N$  height differences may be used to estimate the mathematical expectation (mean value,  $\mu$ ) and dispersion (standard deviation,  $\sigma$ ) of the said height differences for the whole surface ( $X = \{x_1, x_2, \dots, x_N\}$ ). This is the procedure adopted by the NSSDA standards and the ASPRS Lidar Committee, as was previously discussed.

The first step is to obtain a confidence interval for  $\mu$  by means of statistical inference, taking into account the likely non-normal nature of the errors. It is therefore necessary to determine the confidence interval using a non-parametric approach where no particular distribution is assumed. This issue can be resolved using the theory of estimating functions, as demonstrated by Aguilar et al. (2007a). An outline of the main steps in deriving equations (2) and (3) can also be found in the Appendix located at the end of this paper.

$$\mu_{\text{upper}} = \bar{x} + \frac{\frac{\gamma_{2m}+2}{\gamma_{1m}} + \sqrt{\left(\frac{\gamma_{2m}+2}{\gamma_{1m}}\right)^2 + 4\left(\frac{t_\alpha\sqrt{(\gamma_{2m}+2)(\gamma_{2m}+2-\gamma_{1m}^2)}}{|\gamma_{1m}|} + 1\right)}}{2} \frac{\sigma}{\sqrt{N}} \quad (2)$$

$$\mu_{\text{lower}} = \bar{x} + \frac{\frac{\gamma_{2m}+2}{\gamma_{1m}} - \sqrt{\left(\frac{\gamma_{2m}+2}{\gamma_{1m}}\right)^2 + 4\left(\frac{t_\alpha\sqrt{(\gamma_{2m}+2)(\gamma_{2m}+2-\gamma_{1m}^2)}}{|\gamma_{1m}|} + 1\right)}}{2} \frac{\sigma}{\sqrt{N}} \quad (3)$$

where  $\bar{x}$  is the sample mean,  $N$  is the sample size,  $\sigma$  is the value of the population standard deviation (which will be estimated by sample standard deviation Sd), and finally  $\gamma_{1m}$  and  $\gamma_{2m}$  are the skewness and standardised kurtosis of the sampling distribution of the sample mean, respectively. Skewness and standardised kurtosis can be estimated from a finite sample by means of the following expressions:

$$\gamma_1^* = \frac{N \sum_{i=1}^N (x_i - \bar{x})^3}{(N-1)(N-2)\text{Sd}^3}; \gamma_{1m}^* = \frac{\gamma_1^*}{\sqrt{N}} \quad (4)$$

$$\gamma_2^* = \frac{N(N+1) \sum_{i=1}^N (x_i - \bar{x})^4}{(N-1)(N-2)(N-3)\text{Sd}^4} - \frac{3(N-1)^2}{(N-2)(N-3)}; \gamma_{2m}^* = \frac{\gamma_2^*}{N}. \quad (5)$$

The use of  $t_\alpha$ , a one-tailed critical value corresponding to a confidence level of  $1 - \alpha$ , instead of  $z_\alpha$  should be noted in equations (2) and (3). This change is recommended when estimating confidence intervals using the sample standard deviation instead of  $\sigma$ , especially when working with small samples (lower than 100 observations). In this case it is more appropriate to employ Student's  $t$ -distribution calculating  $t_\alpha$  values for  $N - 1$  degrees of freedom,  $N$  being the sample size (Steel and Torrie, 1980).

Once the confidence interval for the mean is determined, which therefore includes the disturbance effects of error bias, skewness and kurtosis, the second step implies treating data as if it came from a normal population. Under this hypothesis the following expressions can be written to estimate the maximum and minimum expected vertical error ( $x_{\text{upper}}$  and  $x_{\text{lower}}$ , respectively) for a  $1 - \alpha$  confidence level:

$$x_{\text{upper}} = \mu_{\text{upper}} + t_\alpha \text{Sd} \quad (6)$$

$$x_{\text{lower}} = \mu_{\text{lower}} - t_\alpha \text{Sd}. \quad (7)$$

Systematic errors are not assumed to have been removed and so the mean is not necessarily expected to be zero. This phenomenon is frequently encountered when working with lidar data in afforested areas where it is usual to find a positive bias due to dense low lying vegetation beneath the tree canopy (Kraus and Pfeifer, 1998; Goodwin et al., 2006; Göpfert and Heipke, 2006; Su and Bork, 2006).

*Model for Estimating Lidar-Derived DEM Accuracy Using Rmse*

In this case it is necessary to construct a confidence interval to estimate uncertainty when rmse is employed to evaluate lidar-derived DEM vertical accuracy, again supposing a non-normal error distribution. The approach proposed by Aguilar et al. (2007a), based on estimating functions theory, was again adopted to obtain the following equations:

$$rmse_{upper} = \sqrt{mse + \frac{\frac{\gamma_{2mse}+2}{\gamma_{1mse}} + \sqrt{\left(\frac{\gamma_{2mse}+2}{\gamma_{1mse}}\right)^2 + 4\left(\frac{t_x \sqrt{(\gamma_{2mse}+2)(\gamma_{2mse}+2-\gamma_{1mse}^2)}}{|\gamma_{1mse}|} + 1\right)}}{2}} \sigma_{mse} \quad (8)$$

$$rmse_{lower} = \sqrt{mse + \frac{\frac{\gamma_{2mse}+2}{\gamma_{1mse}} - \sqrt{\left(\frac{\gamma_{2mse}+2}{\gamma_{1mse}}\right)^2 + 4\left(\frac{t_x \sqrt{(\gamma_{2mse}+2)(\gamma_{2mse}+2-\gamma_{1mse}^2)}}{|\gamma_{1mse}|} + 1\right)}}{2}} \sigma_{mse} \quad (9)$$

where mse is the mean square error,  $\sigma_{mse}$  is the standard deviation of the mse sampling distribution, and finally  $\gamma_{1mse}$  and  $\gamma_{2mse}$  are mse skewness and standardised kurtosis, respectively. All described central moments for square errors may be estimated from a sample of size  $N$  check points by employing the following expressions:

$$mse = \frac{\sum_{i=1}^N x_i^2}{N} \quad (10)$$

$$Sd_{x^2} = \sqrt{\frac{\sum_{i=1}^N (x_i^2 - mse)^2}{N}}; \quad \sigma_{mse} = \frac{\sigma_{x^2}}{\sqrt{N}} \approx \frac{Sd_{x^2}}{\sqrt{N}} \quad (11)$$

$$\gamma_{1x^2}^* = \frac{N \sum_{i=1}^N (x_i^2 - mse)^3}{(N-1)(N-2)Sd_{x^2}^3}; \quad \gamma_{1mse}^* = \frac{\gamma_{1x^2}^*}{\sqrt{N}} \quad (12)$$

$$\gamma_{2x^2}^* = \frac{N(N+1) \sum_{i=1}^N (x_i^2 - mse)^4}{(N-1)(N-2)(N-3)Sd_{x^2}^4} - \frac{3(N-1)^2}{(N-2)(N-3)}; \quad \gamma_{2mse}^* = \frac{\gamma_{2x^2}^*}{N}. \quad (13)$$

METHODOLOGY

*Test Data*

The study sites concern lidar surveys of two different areas. The first area comprises the data-sets from the second phase of the European Organisation for Experimental Photogram-

metric Research (OEEPE—now European Spatial Data Research, EuroSDR) project on laser scanning. Data of the second area was acquired in a series of lidar surveys carried out by the Natural Environment Research Council’s Airborne Research and Survey Facility (NERC ARSF).

*Data-sets from the EuroSDR Project on Laser Scanning*

As part of the second phase of the EuroSDR project on laser scanning, different lidar data-sets, first and second pulse return, were acquired over the Vaihingen/Enz and Stuttgart city centre (Germany) test sites using an Optech ALTM laser scanner. International Society for Photogrammetry and Remote Sensing (ISPRS) Working Group III/3 has previously used these data-sets for testing numerous filtering algorithms (Sithole and Vosselman, 2004) and data has been made available through the working group’s website (<http://www.commission3.isprs.org/wg3>). A total of seven study sites (four urban and three rural) were selected for testing the aforementioned models, with sites chosen for their varied characteristics and diverse feature content (open fields, dense buildings, large buildings, quarries, bridges, dense vegetation and steep slopes, among others). The selected sites were appropriate to enable validation of the proposed models against a very wide range of operational conditions. Fifteen sub-sites were extracted from the original data (Table I).

For a better understanding of the nature of the data, a planimetric view of Site 2 and corresponding reference sub-sites are depicted in Fig. 1. Reference data (ground truth) was generated by manual filtering of the height data-set to segment bare earth data, the process assisted by knowledge of the landscape and aerial photographs (see Sithole and Vosselman (2003) for further details).

*Data-sets from the NERC ARSF*

Filey Bay is located approximately 11 kilometres south of the coastal town of Scarborough in North Yorkshire, UK. Three lidar data-sets of Filey Bay were acquired using

TABLE I. Characteristics of the selected EuroSDR data-sets.

Site	Point spacing	Reference data (sub-site)	Number of points	Characteristics
Site 1	1 m	Sample 11	17 626	Steep slopes, mixture of vegetation and buildings on hillside, buildings on hillside
	1 m	Sample 12	25 344	
Site 2	1 m	Sample 21	10 055	Large and irregularly shaped buildings, small tunnel, road and bridge
	1 m	Sample 22	21 570	
	1 m	Sample 23	11 025	
	1 m	Sample 24	3755	
Site 3	1 m	Sample 31	15 544	Vegetation between a high density of buildings, building with eccentric roof, open terrain with mixture of low and high features
Site 4	1 m	Sample 41	1653	Railway station with trains (low density of ground points)
	1 m	Sample 42	11 989	
Site 5	2 m	Sample 51	13 860	Steep vegetated slopes, quarry (sharp break lines), vegetated river bank
	2 m	Sample 52	17 678	
	2 m	Sample 53	25 113	
	2 m	Sample 54	3948	
Site 6	2 m	Sample 61	31 531	Large buildings, road including embankments
Site 7	2 m	Sample 71	12 786	Road, underpass, road including embankments

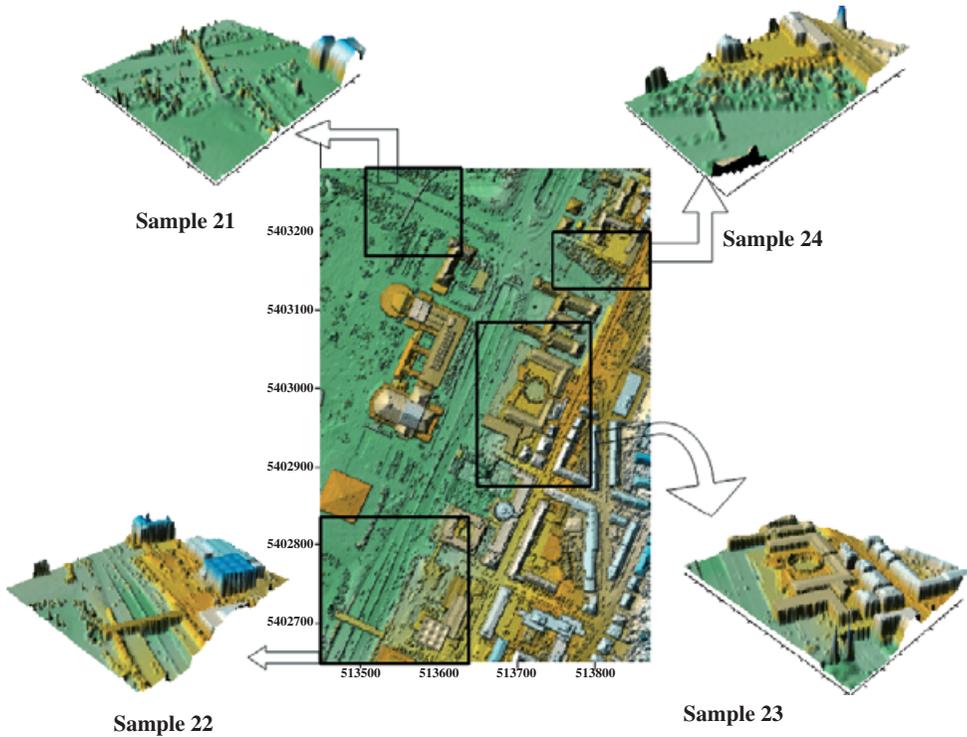


FIG. 1. Site 2 with corresponding reference sub-sites shown as perspective views.

an Optech ALTM 3033 scanner (operating at 1000 m flight altitude) in April 2005 (data-set 0405), August 2005 (0805) and May 2006 (0506); Miller et al. (2007) contains further details of the surveys. All issues related to georeferencing were carried out by the NERC ARSF. As independent quality control, delivered lidar data was compared against a testfield based on check points acquired during a previous survey in 2001 using kinematic GPS (anticipated vertical accuracy of around 1 to 2 cm). The test area consisted of a hard surface, an asphalt road circuit, located at Filey Brigg Country Park (Fig. 2) that enabled an open terrain vertical accuracy evaluation of the lidar data to be undertaken. In this case no error from filtering or gridding is expected, and so all error can be assumed to arise from the data capture process.

Table II gives statistical parameters associated to the vertical errors computed for the Filey Bay data-sets. It should be highlighted that error distribution generally fitted a normal distribution except in the case of the 0506 lidar data, where the “3-sigma” rule (Daniel and Tennant, 2001) had to be applied to remove three likely outliers. Notice that the Kolmogorov–Smirnov (K-S) test (Royston, 1982), with a confidence level of 95%, demonstrated that the error distribution for each data-set matched quite well a normal probability distribution because the K-S statistic was always lower than the 95% critical value. Despite the 0805 lidar data-set presenting an absolute value of skewness higher than the critical value of 0.5, proposed by some authors as a reasonable limit to consider data to come from a normal distribution (Daniel and Tennant, 2001), no outlier correction was applied because of the results yielded by the K-S test. Therefore, these data-sets were subsequently used for testing the proposed model to construct rmse confidence intervals from normally distributed lidar error data.

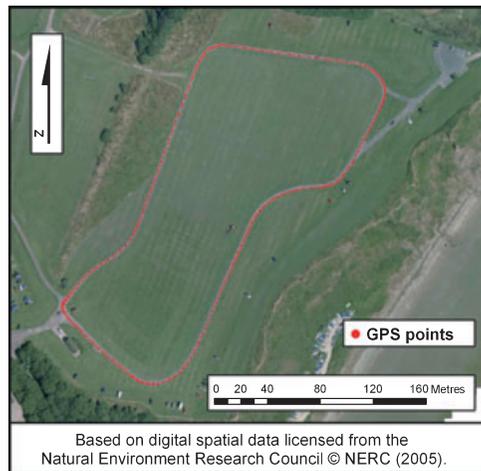


FIG. 2. GPS check point testfield located at Filey Brigg Country Park.

TABLE II. Statistical characteristics of Filey Bay data-sets.

	<i>Data-sets</i>			
	<i>0405 lidar</i>	<i>0805 lidar</i>	<i>0506 lidar (raw)</i>	<i>0506 lidar (3-sigma rule)</i>
Number of check points	840	833	834	831
Minimum error (m)	-0.16	-0.29	-0.03	-0.03
Maximum error (m)	0.03	0.23	1.31	0.20
Mean error (m)	-0.06	-0.16	0.09	0.08
Standard deviation (m)	0.03	0.05	0.06	0.04
Skewness	-0.04	1.24	10.75	0.19
Standardised kurtosis	0.12	9.07	179.45	0.00
K-S statistic	0.026	0.038	0.151	0.029
K-S critical value (95%)	0.047	0.047	0.047	0.047

### *Raw Data Processing of EuroSDR Data*

Last-return raw EuroSDR lidar data was filtered to segment the ground surface from vegetation, buildings and any gross errors embedded in the general point cloud (Axelsson, 1999). The filtering algorithm developed by Axelsson (2000) was used to segment ground points by means of a progressive triangular irregular network (TIN) densification method where the surface was allowed to fluctuate within certain values. Briefly, a sparse TIN is derived from neighbourhood minima, and then progressively densified through the laser point cloud. The algorithm has been implemented in TerraScan (integrated in Terrasolid, version 007.004), the software used to carry out the filtering process. As the first step in the filtering process, points regarded as too low or very high were removed as likely outliers. Subsequently, the ground point class was filtered and saved as points belonging to the DEM. Fig. 3 depicts an example of the results from the filtering process for Site 5. The interpolation method used to refill any gaps created in the filtered ground model (gaps resulting from the removal of points on buildings, vegetation and the like) was the

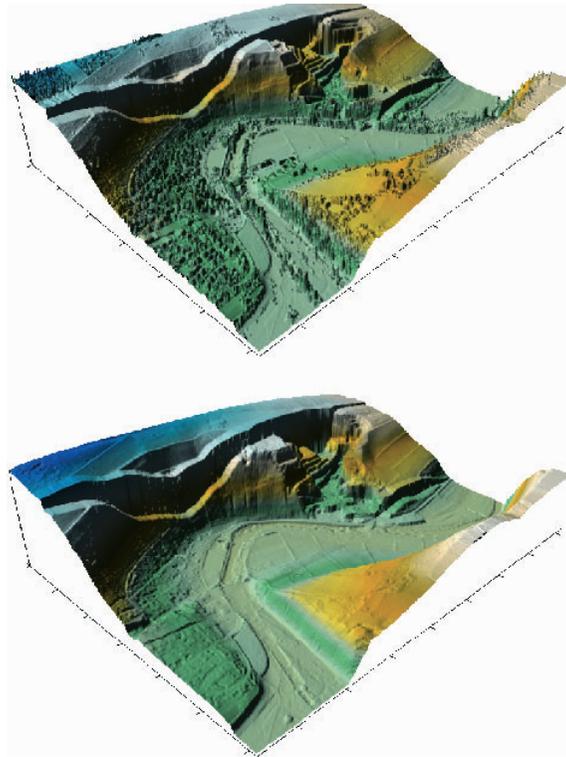


FIG. 3. Filtering process over Site 5. Top: last-return raw data. Bottom: last-return ground-filtered data.

Multiquadric Radial Basis Function working with the local support of the eight neighbours closest to the interpolated point, since this has previously been found generally to yield better results (Aguilar et al., 2005). The final grid spacing of the interpolated DEM was the same as presented by the raw lidar data.

Lidar error was computed as the difference between the derived ground DEM heights (obtained after applying Axelsson's filtering) and the heights from the manually filtered reference data for each of the 15 EuroSDR samples (as listed in Table I). It is important to note that large DEM data gaps caused by the filtering process (for example, the removal of points on large buildings) were not included in the error calculation in order to avoid the uncertainty in areas where no reference data (ground truth) was available. Hence the error data-sets obtained (Table III) can be considered very close to the errors found in real-world lidar surveys. Furthermore, the 3-sigma rule was applied to remove outliers which can corrupt the true statistical distribution of the errors. In fact, this technique is generally applied under operational conditions for removing outliers in the DEM quality assessment of lidar data (Daniel and Tennant, 2001). The percentage of data removed ranged between 0.5 and 3.5%, with a mean value of 1.8%. These results are fairly consistent with those published by Torlegård et al. (1986) referring to errors from photogrammetric DEMs. Nevertheless, despite applying the 3-sigma rule, corrected error data-sets still followed a non-normal distribution according to the K-S 95% confidence test.

TABLE III. Statistics for 3-sigma rule corrected error from EuroSDR data-sets.

Samples	Statistics							
	Points	Mean (m)	Sd (m)	$\gamma_1$	$\gamma_2$	K-S	Critical K-S	% data outliers
11	16 995	0.16	0.44	3.30	11.82	0.322	0.010	3.5
12	25 203	0.04	0.16	6.26	54.41	0.291	0.010	0.5
21	9742	0.02	0.05	1.83	4.77	0.153	0.014	3.1
22	21 193	0.04	0.13	5.12	32.39	0.272	0.009	1.7
23	10 871	0.05	0.20	5.06	31.26	0.299	0.013	1.4
24	3695	0.04	0.17	5.31	45.81	0.273	0.022	1.6
31	15 315	0.01	0.04	1.29	4.16	0.108	0.011	1.4
41	1626	0.25	1.11	5.05	24.41	0.424	0.034	1.6
42	11 743	0.02	0.07	2.98	13.22	0.201	0.013	2.0
51	13 701	0.00	0.06	0.26	5.32	0.107	0.012	1.1
52	17 368	0.08	0.30	2.52	8.41	0.270	0.010	1.7
53	24 702	0.16	0.71	6.64	57.08	0.360	0.009	1.6
54	3863	0.00	0.08	2.79	19.27	0.179	0.022	2.1
61	31 057	0.01	0.10	3.84	24.07	0.252	0.008	1.5
71	12 517	0.02	0.11	2.33	10.80	0.208	0.012	2.1

### Experimental Validation of the Proposed Models

The Monte Carlo method was employed for model validation. Monte Carlo numerical simulation is based on stochastic techniques, meaning it is based on the use of random numbers and probability statistics, to investigate complex problems without relying on analytical equations (Weir, 2002). For the first model evaluated, described through equations (2) to (7), the simulation started by extracting a sample of residuals (sample  $X$  of size  $N$ ) from each of the EuroSDR data-sets by means of random sampling. In this case  $N$  took the values of 20, 40, 60, 80 and 100 check points. The model parameters were then computed and the bounds of the interval,  $x_{\text{upper}}$  and  $x_{\text{lower}}$ , were obtained for a  $1 - \alpha$  confidence level. The number of errors from the whole data-set (population) which were within the bounds estimated by the model was then calculated. It must be noted that the percentage of errors included within the bounds should be close to 95% if, for instance, a significance level of  $\alpha = 0.05$  was chosen. This procedure was repeated 1000 times to estimate the mean value of the percentage of errors from the population falling within the bounds and its associated reliability measure (coefficient of variation). Over the same simulated samples, and purely for comparison purposes, fundamental and supplemental vertical accuracies were also computed according to the ASPRS guidelines for lidar data (Flood, 2004), their corresponding mean values and associated reliabilities also being obtained for 1000 runs.

For the evaluation of the second model, described through equations (8) to (13), the simulation procedure was quite similar to the first, with the exception that the estimated bounds were rmse values and the sample size increased in increments of 20, from 20 to 200 check points. The population rmse for every data-set was calculated from all residual values available in each. It was then necessary to determine whether the corresponding rmse population fell within the limits of the calculated confidence intervals and whether the model had correctly predicted the expected uncertainty for data-set estimation. In this case the procedure was run 2000 times because it was computationally faster than for the first model.

Note that because samples were extracted from finite populations, a correcting coefficient must be applied to calculate sample variance in order to correct the reduced variability expected on consecutive random samplings extracted from the same finite residual population (Steel and Torrie, 1980). Therefore, the following equation was applied:

$$Sd_m = \frac{Sd}{\sqrt{N}} \sqrt{\frac{M - N}{M}} \tag{14}$$

where  $M$  is the number of residuals,  $N$  is the sampling size,  $Sd$  and  $Sd_m$  are the standard deviation of the residuals sample and the mean sample, respectively.

RESULTS AND DISCUSSION

*Confidence Intervals for Lidar Vertical Accuracy Assessment*

Fig. 4 shows the performance of the proposed model compared against results achieved using the ASPRS guidelines (Flood, 2004). Only results for samples 11, 52 and 61 are given, although results are similar for all 15 samples.

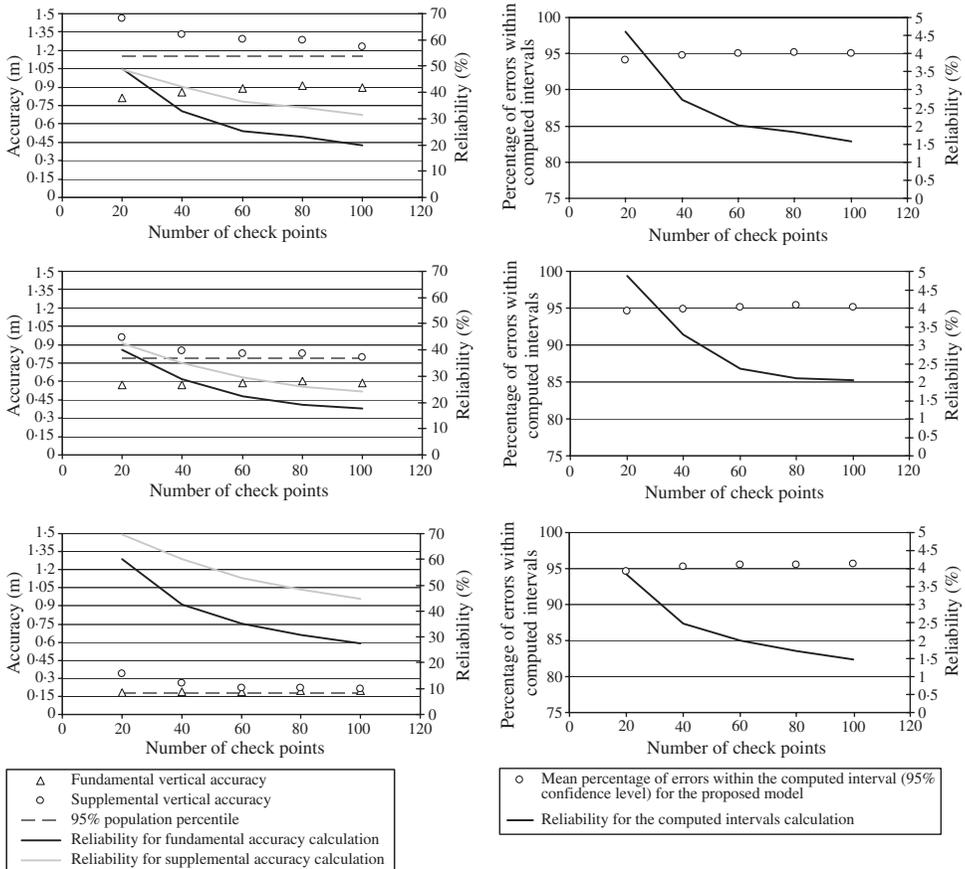


FIG. 4. Evaluation results for samples 11 (top graphs), 52 (middle) and 61 (bottom) from the EuroSDR data-set. Left column: accuracy following ASPRS guidelines. Right column: proposed model.

Results calculated using the fundamental accuracy assessment, as defined in equation (1) supposing a normal distribution of the error, were usually far from the true vertical accuracy computed for the error population. This is to be expected, taking into account the non-normally distributed nature of lidar error data-sets in non-open terrain, as previously confirmed from results yielded by the K-S test given in Table III. Therefore, following the recommendations of the ASPRS guidelines, the supplemental vertical accuracy test figures should be considered and computed using the 95th percentile method, as explained in the introduction. In this case supplemental accuracy always overestimated the true value, but closed in to the truth when the sample size increased.

The problem with both fundamental and supplemental vertical accuracy is that the variability of the estimates was very noticeable, especially in the case of the latter (always above  $\pm 20\%$  even when using up to 100 check points). Hence many more than 100 check points would have to be employed to ensure a confidence level near to the target value of 95%. However, the 95% confidence intervals computed by means of the proposed model, developed through equations (2) to (7), achieved a percentage of population errors within the bounds very close to the target value, even when working with only 20 check points. Better still, the variability of the estimates was very low, as can be seen in Fig. 4. Summing up, only around 60 check points would be needed to reach a confidence level of 95% with an estimate error below  $\pm 2.5\%$ .

Fig. 5 shows confidence interval limits (the mean of 1000 runs) computed from the proposed model for the three samples referred to in Fig. 4, varying randomly the initial finite sample of check points. The stability of the confidence interval length along the horizontal axis (sample size) should be highlighted. The confidence interval length basically depends on the sample standard deviation, skewness and standardised kurtosis, rather than the number of check points used. Therefore, precise and narrow confidence intervals can be calculated when working with as few as 60 check points. Notice that unlike the case for a normal distribution, the confidence intervals are not symmetrical around the mean value, but they are biased towards the direction of skewness.

The importance of the development of this simple and understandable model for accuracy estimation is supported by the fact that, traditionally, users are known to be failing to quantify risks due a lack of tools, theory and poor documentation on spatial data quality (van Oort and Bregt, 2005). The presentation of clear upper and lower limits rather than a single accuracy value may substantially improve the understanding of accuracy test results by users of lidar data and enhance communication with data providers. Furthermore, it is worth noting that the proposed model, unlike the ASPRS methodology, does not assume any previous lidar error distribution since all required information is extracted from the sample by means of a non-parametric approach. Thus, the so-called “strong assumption” of distribution normality accepted in many previous works, and in the majority of the current standards for positional accuracy, is avoided.

#### *Confidence Intervals for Rmse Estimate*

Figs. 6 and 7 show the performance of the model tested to compute rmse 95% confidence intervals as well as another model based on the Student's *t* asymptotic approach. The Student's *t* model has been formulated under the hypothesis of an almost-normal distribution of lidar vertical error and is given by the following expressions:

$$\text{rmse}_{\text{upper}} = \sqrt{\text{mse} + t_{\alpha/2} \sigma_{\text{mse}}} \quad (15)$$

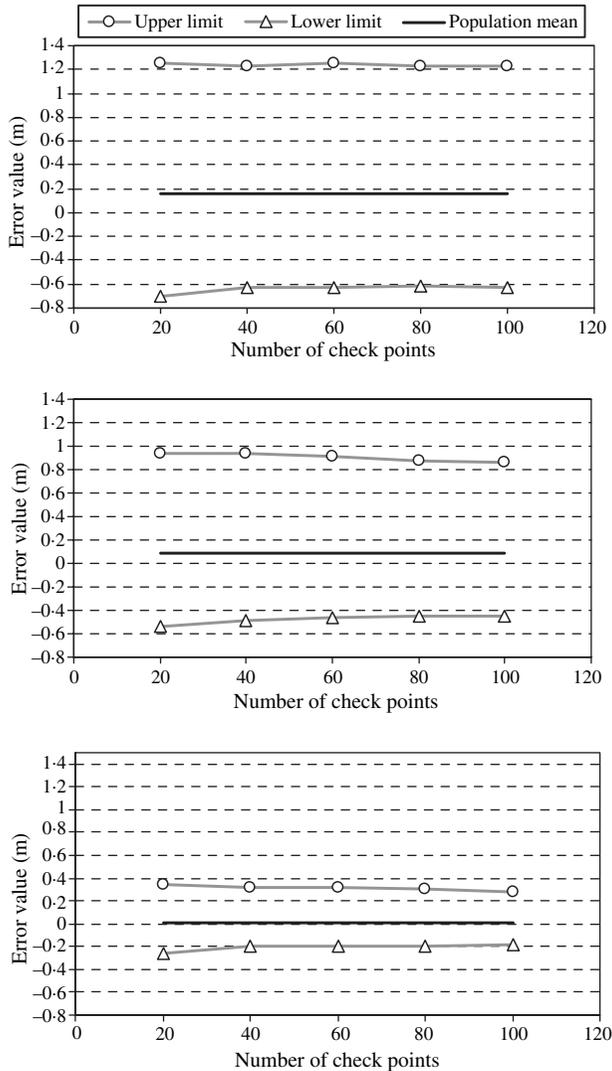


FIG. 5. Confidence intervals (95% confidence level) as a function of sample size computed, using the proposed model, for samples 11 (top graph), 52 (middle) and 61 (bottom) from the EuroSDR data-set.

$$\text{rmse}_{\text{lower}} = \sqrt{\text{mse} - t_{\alpha/2} \sigma_{\text{mse}}} \quad (16)$$

where mse is the mean square error given by equation (10),  $\sigma_{\text{mse}}$  is the standard deviation of the mse sampling distribution (equation (11)) and  $t_{\alpha/2}$  is the two tails critical value corresponding to confidence level  $\alpha$  for  $N - 1$  degrees of freedom,  $N$  being the sample size (see Aguilar et al. (2007a) for further details). In the case of error data coming from non-open terrain, such as those from the EuroSDR reference data-sets (Fig. 6), it seems to be clear

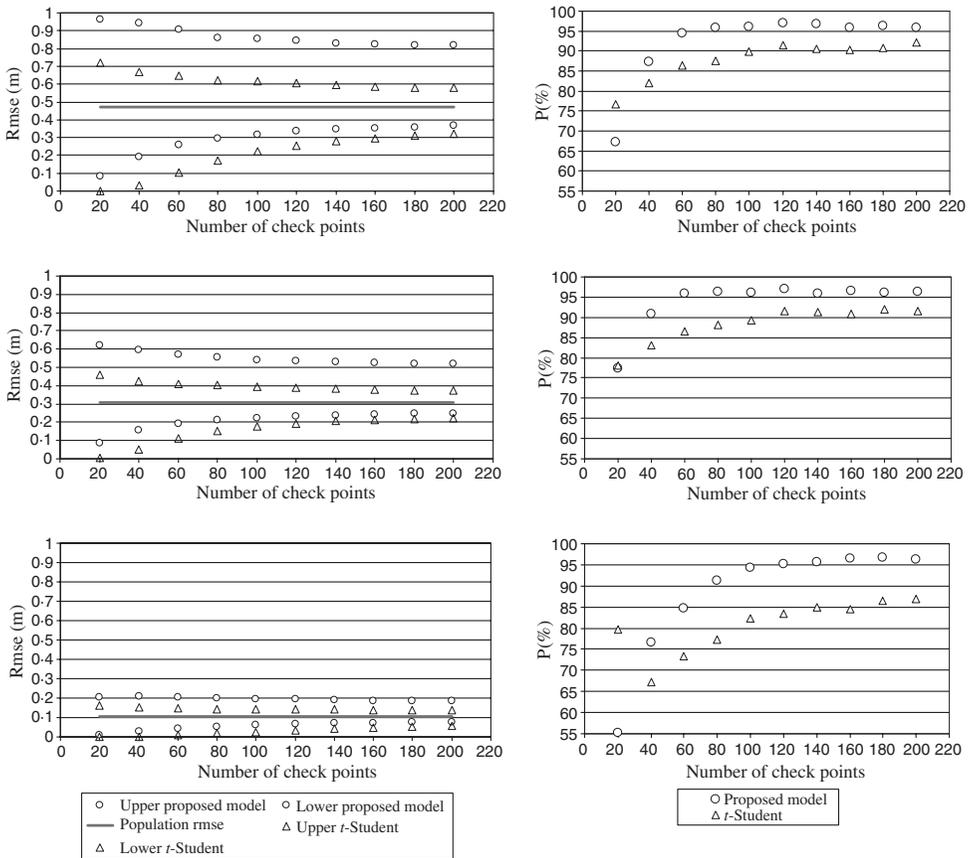


FIG. 6. Rmse evaluation results for samples 11 (top graphs), 52 (middle) and 61 (bottom) from the EuroSDR data-set. Left column: rmse confidence intervals (95% level) computed from the proposed model and the Student's *t* approach. Right column: percentage of times population rmse falls within intervals computed for the proposed model and the Student's *t* approach.

that the Student's *t* approach failed to reach the expected target value of 95% confidence, even when using sample sizes of up to 200 check points (Fig. 6, right). Conversely, the model tested here can be considered successful when using anything upwards from around 60 to 160 check points, 100 being the average for all 15 EuroSDR data-sets (see Fig. 9).

Notice that the length of the confidence intervals (Fig. 6, left) is narrower in the case of the Student's *t* model because the residual population has been assumed to be normal, and so standardised kurtosis and skewness have been set to zero (no outliers or systematic errors present). This is an unreal scenario which is overcome by means of the assumption-free non-parametric model. It is recognised that whilst non-normal, fairly symmetrical population shapes converge to normal looking sampling distributions with relatively small samples, strongly skewed distributions require much larger samples for the approximation to be good when a normal distribution is supposed. Thus the Student's *t* approach is expected to require a large number of check points when working with a population that exhibits high skewness.

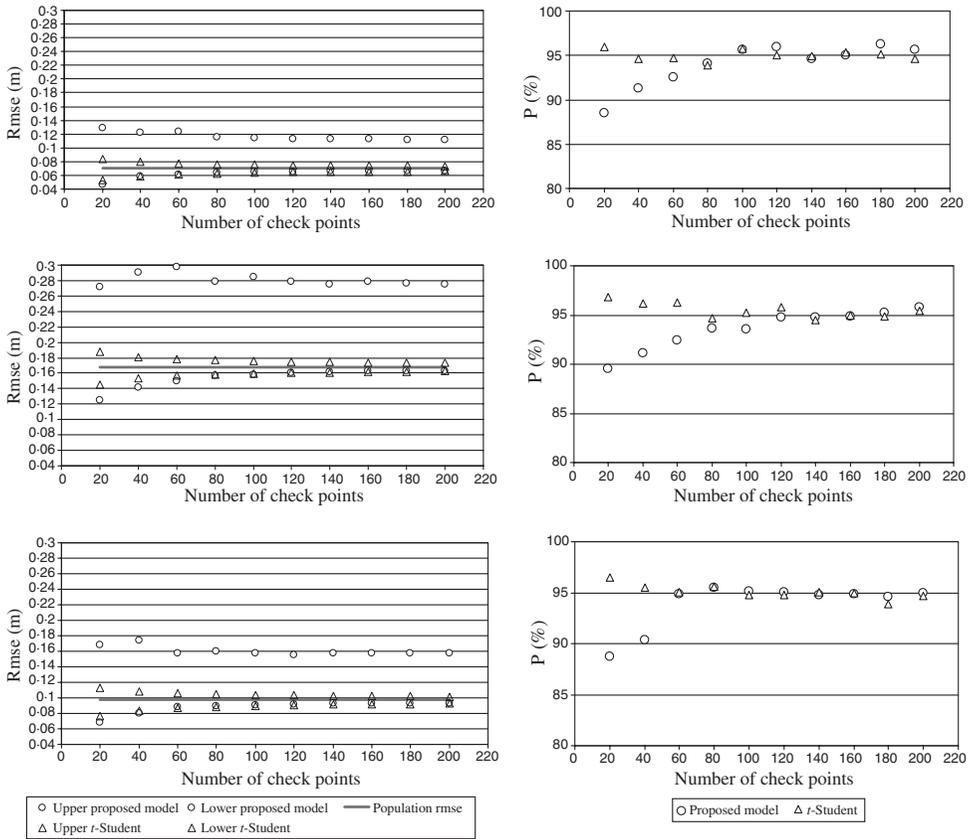


FIG. 7. Rmse evaluation results for 0405 lidar (top graphs), 0805 lidar (middle) and 3-sigma rule corrected 0506 lidar (bottom) from the NERC ARSF data-set. Left column: rmse confidence intervals (95% level) computed from the proposed model and Student's *t* approach. Right column: percentage of times population rmse falls within intervals computed for the proposed model and Student's *t* approach.

Results from the NERC ARSF test data (Fig. 7), however, demonstrated that the Student's *t* approach may be deemed appropriate when dealing with normal lidar vertical error distributions which are usually found over open terrain (no filtering error). It should be noted that the three NERC ARSF error data-sets clearly followed a normal distribution after being filtered using the 3-sigma rule, as can be observed from the K-S test results in Table II. In this case, results indicate that working with only a few check points, around 20 to 40, would be sufficient. In this instance the proposed model computed confidence intervals that were too wide and illustrated a somewhat erratic behaviour when working with a small number of check points. Indeed, despite the wide mean confidence interval that was calculated, the number of times the population rmse fell within the computed interval limits was found to be very poor until a sample size of between 60 and 80 check points was reached. This was mainly due to the high level of uncertainty in estimating the kurtosis value from a small sample size (Aguilar et al., 2007b). The Student's *t* model estimates only mean square error and the standard deviation of the mean square error sampling distribution, whose values are more reliable than kurtosis and skewness when estimated from small sample sizes.

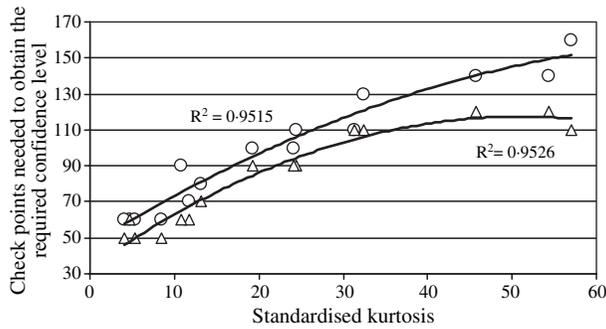


FIG. 8. Relationship between check points needed to reach a determined confidence level for the computed rmse bounds (circles 95%; triangles 90%) and the residual population kurtosis. Data from the 15 EuroSDR samples.

*Determining the Minimum Required Number of Check Points*

The nature of the relationship between the number of check points needed to reach a determined confidence level for the computed rmse bounds and the residual population kurtosis is depicted in Fig. 8. A second-order polynomial has been fitted to the experimental data to provide a better understanding of the non-linear increase in the number of required check points when errors come from a more leptokurtic data-set. The finding can be considered as logical since high kurtosis implies more errors are likely to be outliers, and so more check points are required for the construction of a reliable confidence interval. It should be stressed that when less stringent confidence levels are required (for instance, reducing confidence from 95 to 90%, as illustrated in Fig. 8), a smaller number of check points will be needed to construct a reliable interval according to the chosen confidence level. Kurtosis therefore seems to be the key to determining the required number of check points in order to achieve a certain confidence level in the uncertainty of the accuracy estimate. From these results, further work should be undertaken to determine some morphological terrain and landscape indicators (terrain roughness, landscape spatial distribution and arrangement of features such as buildings

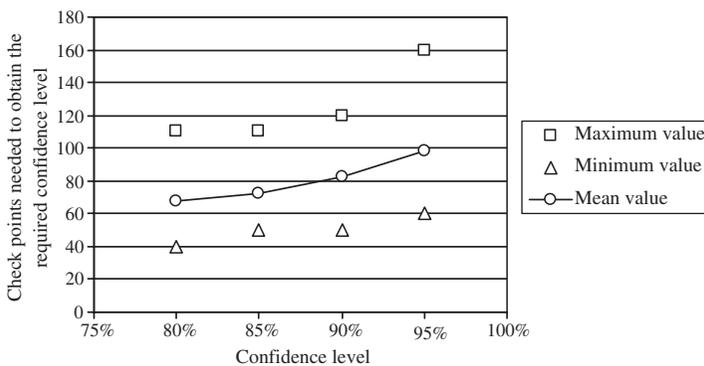


FIG. 9. Relationship between the number of check points needed to reach a determined confidence level and the desired confidence level. Maximum, minimum and mean values extracted from the 15 EuroSDR samples and for every confidence level tested.

and vegetation) to help predict the expected degree of kurtosis in the error budget after applying lidar data filtering and gridding processes.

Currently, the ASPRS guidelines specify that accuracies should be reported at the 95% confidence level. However, some researchers report applications related to geopositioning tasks which typically demand a 90% or even lower confidence level (Rodarmel et al., 2006). Therefore, rather than require 95%, the proposed method allows the user to determine the confidence level depending on the application of the data. Changing the confidence level is as simple as changing the significance level in the developed equations. Fig. 9 clearly shows that fewer check points are required when the degree of confidence is relaxed, although the relationship is not linear. In fact, in the worst case (maximum values in Fig. 9), the number of check points required for computing rmse bounds at the 95% confidence level would be 160, whilst it would be only 110 for an 80% confidence level.

### CONCLUSIONS

From the results obtained through this research the following conclusions can be drawn:

- (1) The model proposed to compute confidence intervals for lidar vertical error has proved to be advantageous compared to the methodology recommended by the ASPRS Lidar Committee for non-open terrain (non-normal error distribution). The proposed methodology works quite well under open and non-open terrain conditions because input parameters such as skewness and kurtosis allow the model to seek the statistical nature of the error data-set. Furthermore, reliable and narrow confidence intervals can be calculated when working from as few as 60 check points. This is because interval length depends basically on the sample standard deviation, skewness and standardised kurtosis, rather than the number of check points used.
- (2) The proposed methodology not only estimates confidence intervals for lidar vertical error, but also computes uncertainty in rmse estimates, which is usually better understood by DEM users and data providers. In this case the model developed by Aguilar et al. (2007a) is recommended when lidar error in non-open terrain is evaluated. Depending on the kurtosis of error population, between 60 and 160 check points (around 100 on average) could be required. However, lidar accuracy assessment via rmse in open terrain (normal error distribution) would be more successfully achieved by means of the Student's *t* approach because it needs fewer check points than the proposed model. In fact, only 20 to 40 check points would be required in this case.
- (3) Bearing in mind the variability of the data-sets tested, the results have proved to be very promising, the proposed models reproducing the statistical behaviour of lidar vertical errors using a reasonably low number of check points, even in the case of the more non-normally distributed residual data-sets. This research can therefore be considered as an important step towards improving and understanding the quality control of lidar-derived DEMs.

### ACKNOWLEDGMENTS

The authors are grateful to the Spanish Government for financing the first author's fellowship at Newcastle University through the Spanish Research Mobility Programme "Estancias de profesores e investigadores españoles en centros de enseñanza superior e investigación extranjeros". Thanks are due to the NERC ARSF and Pauline Miller, Newcastle

University, for assisting with the data from Filey Bay and also to ISPRS WG III/3 “3D Reconstruction from Airborne Laser Scanner and InSAR Data” for facilitating access to the EuroSDR lidar data.

REFERENCES

- AGUILAR, F. J., AGÜERA, F., AGUILAR, M. A. and CARVAJAL, F., 2005. Effects of terrain morphology, sampling density, and interpolation methods on grid DEM accuracy. *Photogrammetric Engineering & Remote Sensing*, 71(7): 805–816.
- 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): 169–192.
- AGUILAR, F. J., AGUILAR, M. A. and AGÜERA, F., 2007a. Accuracy assessment of digital elevation models using a non-parametric approach. *International Journal of Geographical Information Science*, 21(6): 667–686.
- AGUILAR, F. J., AGÜERA, F. and AGUILAR, M. A., 2007b. A theoretical approach to modeling the accuracy assessment of digital elevation models. *Photogrammetric Engineering & Remote Sensing*, 73(12): 1367–1380.
- AXELSSON, P., 1999. Processing of laser scanner data—algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2/3): 138–147.
- AXELSSON, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, 33(B4): 110–117.
- CLARK, M. L., CLARK, D. B. and ROBERTS, D. A., 2004. Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment*, 91(1): 68–89.
- DANIEL, C. and TENNANT, K., 2001. DEM quality assessment. *Digital Elevation Model Technologies and Applications: The DEM Users Manual* (Ed. D. F. Maune). American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland. 539 pages: 395–440.
- FGDC, 1998. Geospatial positioning accuracy standards. National standards for spatial data accuracy. <http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3> [Accessed: 23rd August 2007].
- FISHER, P. F. and TATE, N. J., 2006. Causes and consequences of error in digital elevation models. *Progress in Physical Geography*, 30(4): 467–489.
- FLOOD, M., 2004. ASPRS guidelines. Vertical accuracy reporting for lidar data. <http://www.asprs.org/society/divisions/ppd/standards/Lidar%20guidelines.pdf> [Accessed: 23rd August 2007].
- GODAMBE, V. P. (Ed.), 1991. *Estimating Functions*. Oxford University Press, Oxford. 356 pages.
- GODAMBE, V. P. and THOMPSON, M. E., 1989. An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference*, 22(2): 137–152.
- GOODWIN, N. R., COOPS, N. C. and CULVENOR, D. S., 2006. Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sensing of Environment*, 103(2): 140–152.
- GÖPFERT, J. and HEIPKE, C., 2006. Assessment of LiDAR DTM accuracy in coastal vegetated areas. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3): 79–85.
- HOPKINSON, C., CHASMER, L. E., ZSIGOVICS, G., CREED, I. F., SITAR, M., TREITZ, P. and MAHER, R. V., 2004. Errors in LiDAR ground elevation and wetland vegetation height estimates. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(8/W2): 108–113.
- HYYPÄ, H., YU, X., HYYPÄ, J., KAARTINEN, H., KAASALAINEN, S., HONKAVAARA, E. and RÖNNHOLM, P., 2005. Factors affecting the quality of DTM generation in forested areas. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3/W19): 85–90.
- JAMES, T. D., MURRAY, T., BARRAND, N. E. and BARR, S. L., 2006. Extracting photogrammetric ground control from lidar DEMs for change detection. *Photogrammetric Record*, 21(116): 312–328.
- KRAUS, K. and PFEIFER, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53(4): 193–203.
- LIM, K., TREITZ, P., WULDER, M., ST-ONGE, B. and FLOOD, M., 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27(1): 88–106.
- MILLER, P., MILLS, J., EDWARDS, S., BRYAN, P., MARSH, S., HOBBS, P. and MITCHELL, H., 2007. A robust surface matching technique for integrated monitoring of coastal geohazards. *Marine Geodesy*, 30(1/2): 109–123.
- OORT, P. A. J. VAN and BREGT, A. K., 2005. Do users ignore spatial data quality? A decision-theoretic perspective. *Risk Analysis*, 25(6): 1599–1610.

- RODARMEL, C., THEISS, H., JOHANESEN, T. and SAMBERG, A., 2006. A review of the ASPRS guidelines for the reporting of horizontal and vertical accuracy in lidar data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(1): 6 pages (on CD-ROM).
- ROYSTON, J. P., 1982. Expected normal order statistics (exact and approximate). *Applied Statistics*, 31(2): 161–165.
- SITHOLE, G. and VOSSELMAN, G., 2003. Report: ISPRS comparison of filters. <http://www.itc.nl/isprswgIII-3/filtertest/Report05082003.pdf> [Accessed: 23rd August 2007].
- SITHOLE, G. and VOSSELMAN, G., 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning points clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(1–2): 85–101.
- SMITH, S. L., HOLLAND, D. A. and LONGLEY, P. A., 2005. Quantifying interpolation errors in urban airborne laser scanning models. *Geographical Analysis*, 37(2): 200–224.
- STEEL, R. G. D. and TORRIE, J. H., 1980. *Principles and Procedures of Statistics: A Biometrical Approach*. Second edition. McGraw-Hill, New York. 631 pages.
- SU, J. and BORK, E., 2006. Influence of vegetation, slope, and lidar sampling angle on DEM accuracy. *Photogrammetric Engineering & Remote Sensing*, 72(11): 1265–1274.
- TORLEGÅRD, K., ÖSTMAN, A. and LINDGREN, R., 1986. A comparative test of photogrammetrically sampled digital elevation models. *Photogrammetria*, 41(1): 1–16.
- WEIR, M. J. C., 2002. Monte Carlo simulation of long-term spatial error propagation in forestry databases. *Spatial Data Quality* (Eds. W. Shi, P. F. Fisher and M. F. Goodchild). Taylor & Francis, London. 336 pages: 294–303.

APPENDIX

An estimating equation for parameter  $\theta$  is an equation  $g(X, \theta) = 0$  which has a unique solution  $\theta = \theta(X)$  for every value of  $X$ ,  $X$  being a random variable. In this case,  $X$  is a set of  $N$  independent residuals computed at corresponding check points for evaluating lidar-derived DEM accuracy. The function  $g$  is named the estimating function. An optimal estimation of  $\theta$  can be achieved considering the class of estimating functions as having zero expectation given by the following linear combination:

$$g = \alpha h_1 + \beta h_2 \tag{A1}$$

where  $h_1$  and  $h_2$  are real functions which would be computed depending on the underlying nature of the statistical problem, and  $\alpha$  and  $\beta$  are real functions that can be estimated by the locally optimal solution found by Godambe and Thompson (1989) and shown in equation (A2). For the correct application of this optimal solution,  $E(h_1 h_2)$  should be equal to zero ( $E$  expressing the mathematical expectation of the operation in parentheses). That is to say,  $h_1$  and  $h_2$  must be orthogonal. A more detailed discussion on the theory of estimating functions, together with numerous applications, can be found in Godambe (1991).

$$\alpha^* = \frac{E(\frac{\partial h_1}{\partial \theta})}{E(h_1^2)}; \quad \beta^* = \frac{E(\frac{\partial h_2}{\partial \theta})}{E(h_2^2)}. \tag{A2}$$

In this case, the parameter  $\theta$  would be the expected or mean value for the variable sample mean  $\bar{x}(\mu)$ . The basic estimating functions, supposing the third and fourth central moments are known (skewness and kurtosis, respectively), would be (Godambe and Thompson, 1989)

$$h_1 = \bar{x} - \mu; \quad h_2 = (\bar{x} - \mu)^2 - \sigma_m^2 - \gamma_{1m} \sigma_m (\bar{x} - \mu) \tag{A3}$$

where  $\sigma_m$  and  $\gamma_{1m}$  are the standard deviation and the skewness of the sampling distribution of the sample mean, respectively. It can be easily proved that  $h_1$  and  $h_2$  are orthogonal functions, as stated before. Therefore, the estimates of real functions  $\alpha$  and  $\beta$  can be obtained by means of equation (A2):

$$\alpha^* = -\frac{1}{\sigma_m^2}; \quad \beta^* = \frac{\gamma_{1m}\sigma_m}{\sigma_m^4(\gamma_{2m} + 2 - \gamma_{1m}^2)}; \quad \sigma_m = \frac{\sigma}{\sqrt{N}}. \quad (\text{A4})$$

Hence the estimating function  $g$  for  $\mu$  would be estimated by  $g^* = \alpha^* h_1 + \beta^* h_2$ . Given that the distribution of the statistic  $g^*/\sigma_{g^*}$  can be approximated by a standardised normal distribution, the following expression can be written:

$$P\left(\left|\frac{g^*}{\sigma_{g^*}}\right| \leq Z_\alpha\right) = 1 - \alpha \quad (\text{A5})$$

where  $Z_\alpha$  is the one-tailed critical value corresponding to a confidence level of  $1 - \alpha$ . Hence  $Z_\alpha = 1.64$  for  $\alpha = 0.05$ . Equations (2) and (3) in the main text can be obtained from equation (A5) and solving a quadratic equation for the variable  $\bar{x} - \mu$ . The reader is referred to Godambe and Thompson (1989) for further details.

### Résumé

*Malgré le coût relativement élevé des modèles numériques des altitudes (MNA) obtenus à partir de lidars aéroportés, ces produits sont généralement fournis sans qu'on les accompagne d'une estimation convenable de leur exactitude. Dans la plupart des cas, l'estimation de l'exactitude des MNA est fournie d'après une comparaison entre les altitudes obtenues au lidar et celles en un nombre limité de points de vérification dont les coordonnées proviennent d'une source indépendante et de précision supérieure; on suppose également que la distribution des erreurs et des écarts d'altitudes suit une loi normale. On propose dans cet article une nouvelle méthodologie pour évaluer l'exactitude altimétrique des MNA issus de lidars en recourant à des intervalles de confiance élaborés à partir d'un échantillonnage limité d'erreurs calculées aux points de vérification. On a essayé une solution non-paramétrique dans laquelle on ne suppose aucune distribution particulière des erreurs, ce qui permet d'appliquer la méthodologie proposée lorsque les distributions d'erreurs ne sont pas normales, ce qui est généralement le cas des MNA issus de lidars. On a testé la qualité de la modélisation proposée en utilisant une simulation expérimentale de type Monte Carlo sur 18 jeux de données d'erreurs altimétriques. On a établi 15 de ces jeux à partir des données-lidar originales fournies par le Groupe de Travail III/3 de la Société Internationale de Photogrammétrie et de Télédétection, en utilisant comme réalité de terrain leurs données de référence filtrées manuellement. Les trois derniers jeux de données ont été fournis par le système lidar du Natural Environment Research Council et son Airborne Research and Survey Facility, en déterminant les points de vérification par un système GPS cinématique de haute précision. Les résultats se sont montrés prometteurs, les modèles proposés reproduisant bien le comportement statistique des erreurs altimétriques du lidar pour un nombre convenable de points de vérification, y compris dans le cas où les jeux de données n'ont pas une distribution normale des résidus. On peut donc considérer que cette étude constitue une avancée importante pour l'amélioration du contrôle de la qualité des MNA issus de lidars.*

### Zusammenfassung

*Trotz der relativ hohen Kosten für Digitale Höhenmodelle (DEMs) aus flugzeuggestütztem Lidar werden diese Produkte ohne eine zufrieden stellende*

*Genauigkeitsabschätzung geliefert. Meist wird die Genauigkeit des Höhenmodells durch Vergleich einer Auswahl von Kontrollpunkthöhen aus unabhängigen Messungen höherer Genauigkeit mit den Höhen aus den Lidar Daten abgeleitet. Dazu wird eine Normalverteilung der abgeleiteten Höhendifferenzen, bzw. Höhenfehler angenommen. Dieser Beitrag schlägt eine neue Methode zur Schätzung der Höhengenaugigkeit eines Lidar Höhenmodells vor, das Konfidenzintervalle nutzt, die aus einer Stichprobe von Fehlern an Kontrollpunkten abgeleitet werden. Ein nicht-parametrischer Ansatz wurde getestet, bei dem keine besondere Fehlerverteilung angenommen wurde. Damit wird die vorgeschlagene Methode besonders auf die nicht-normalverteilten Fehler anwendbar, wie sie typischerweise in den Höhenmodellen aus Lidar auftreten. Das Verhalten des vorgeschlagenen Modells wurde empirisch mit einer Monte-Carlo Simulation an 18 Höhenfehlerdatensätzen untersucht. 15 dieser Datensätze wurden aus den originalen Lidar Daten der Arbeitsgruppe III/3 der Internationalen Gesellschaft für Photogrammetrie und Fernerkundung gerechnet, wobei die zugehörigen gefilterten Referenzdaten als Sollwerte angenommen wurden. Die restlichen drei Datensätze wurden durch das Lidar System des "Natural Environment Research Council's Airborne Research and Survey Facility" bereitgestellt zusammen mit Kontrollpunkten, die mit hochgenauem kinematischen GPS bestimmt worden sind. Die Ergebnisse waren viel versprechend. Die vorgeschlagenen Methode reproduzierte das statistische Verhalten der Höhenfehler der Lidar Daten mit einer günstigen Anzahl von Kontrollpunkten, sogar in den Datensätzen mit nicht-normalverteilten Verbesserungen. Diese Forschungsergebnisse können deshalb als wichtiger Schritt zur Verbesserung der Qualitätskontrolle von Höhenmodellen aus Lidar Daten betrachtet werden.*

#### Resumen

*Pese al relativamente elevado coste de los Modelos Digitales de Elevación (MDE) obtenidos mediante lidar aerotransportado, estos productos se proporcionan habitualmente sin una adecuada estimación de su precisión. En la mayoría de los casos la estimación de la exactitud de un MDE se hace comparando las elevaciones lidar con una muestra finita de puntos de comprobación procedentes de una fuente independiente de referencia de mayor exactitud. En estos casos se suele asumir que la distribución estadística de las diferencias de altura o error vertical sigue una distribución normal. Este artículo propone una nueva metodología para estimar la exactitud vertical de un MDE lidar mediante intervalos de confianza construidos a partir de una muestra finita de errores calculados en los puntos de comprobación. En este sentido, se ha adoptado una perspectiva no paramétrica, sin asumir ninguna distribución específica del error, de modo que la metodología propuesta es particularmente aplicable en el caso de distribuciones no normales como las que suelen encontrarse en MDE lidar. La bondad del modelo propuesto fue validada experimentalmente en 18 conjuntos de datos lidar mediante una simulación numérica basada en el método de Monte Carlo. 15 de las 18 poblaciones de errores se obtuvieron a partir de datos lidar brutos facilitados por el Grupo de Trabajo III/3 de International Society for Photogrammetry and Remote Sensing, usando sus respectivos datos de referencia, filtrados manualmente, como forma de validación. Las tres poblaciones de errores restantes fueron proporcionadas por el sistema lidar del Airborne Research and Survey Facility del Natural Environment Research Council, junto con puntos de comprobación adquiridos mediante GPS cinemático de*

*alta precisión. Los resultados obtenidos resultaron ser muy prometedores, observándose cómo los modelos propuestos reproducían acertadamente el comportamiento estadístico de los errores lidar verticales a partir de una muestra relativamente pequeña de puntos de comprobación, incluso en casos en que los datos tenían residuos con una distribución no normal. Por tanto, esta investigación puede considerarse como un avance muy importante en la mejora del proceso de control de calidad de los DEM lidar.*