International Journal of Applied Earth Observation and Geoinformation 34 (2015) 198–206

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Robust methods for assessing the accuracy of linear interpolated DEM

Bin Wang[a], Wenzhong Shi[a,*], Eryong Liu[b]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[b] School of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China

ARTICLE INFO

Article history:
Received 24 August 2013
Accepted 14 August 2014
Available online 6 September 2014

Keywords:
DEM accuracy
Interpolation residuals
Robust estimation
Confidence interval
Monte Carlo simulation

ABSTRACT

Methods for assessing the accuracy of a digital elevation model (DEM) with emphasis on robust methods have been studied in this paper. Based on the squared DEM residual population generated by the bi-linear interpolation method, three average-error statistics including (a) mean, (b) median, and (c) M-estimator are thoroughly investigated for measuring the interpolated DEM accuracy. Correspondingly, their confidence intervals are also constructed for each average error statistic to further evaluate the DEM quality. The first method mainly utilizes the student distribution while the second and third are derived from the robust theories. These innovative robust methods possess the capability of counteracting the outlier effects or even the skew distributed residuals in DEM accuracy assessment. Experimental studies using Monte Carlo simulation have commendably investigated the asymptotic convergence behavior of confidence intervals constructed by these three methods with the increase of sample size. It is demonstrated that the robust methods can produce more reliable DEM accuracy assessment results compared with those by the classical t-distribution-based method. Consequently, these proposed robust methods are strongly recommended for assessing DEM accuracy, particularly for those cases where the DEM residual population is evidently non-normal or heavily contaminated with outliers.

Introduction

The digital elevation model (DEM) has been widely used in GIS filed to digitally represent the variation of the Earth's surface. No matter appearing in the form of a rasterized DEMs with regular square-shaped cells or a vector-based triangulated irregular networks (TINs), an elevation surface is normally generated by interpolation of sampled terrain elevation points (Maune, 2007). The common approaches to capturing the terrain elevation values of a terrain surface include field surveying by total stations, photogrammetry or airborne LiDAR. These data acquisition methods are normally subject to measurement errors and the subsequent DEM interpolation models can propagate or further enlarge the errors (Fisher and Tate, 2006). A DEM, therefore subject to a certain level of errors, needs to be properly assessed and specified to the DEM users. Meanwhile, as the era of big data comes, the requirements of spatial data quality information increase significantly. Therefore, as pointed out by Shi et al. (2004), measurement of the positional error of geo-spatial data is a key research issue in the area of quality assessment of spatial data.

In the field of uncertainty modeling and spatial data quality analysis, DEM error modeling has occupied the core of considerable researches (Shi, 2010). Much attention has been paid to exploring the effects of all sorts of interpolation models upon various kinds DEM data with different spatial resolutions (Chaplot et al., 2006). It is well-known that the common indicator for assessing the DEM accuracy can be root mean squared error (RMSE), which has been employed together with trend analysis to evaluate the quality of the DEM produced from ASTER stereoscopy (de Oliveira and Paradella, 2009). Basically, the following three factors are mainly perceived to account for the interpolated DEM uncertainty: (1) source data errors from the spatial data acquisition stage; (2) the employed DEM interpolation models; (3) the complexity level of terrain variation (Shi et al., 2014). Based on the approximation theory, Hu et al. (2009) has discussed the mathematical relationship between DEM error and these three influencing factors. Normally, type (1) error is regarded as data-based error; hence it is highly related to the data capturing methods, while type (2) and (3) errors are perceived to be the model-based errors. These two latter type errors have aroused great attention of the scholars for investigating how close the interpolated DEM surface approximates the actual ground surface (Fisher and Tate, 2006). Furthermore, Chaplot et al. (2006) evaluated the performance of five common interpolation techniques: inverse distance weighting (IDW), ordinary/universal kriging, radial basis function (RBF), and the regularized spline

* Corresponding author. Tel.: +852 2766 5975; fax: +852 2330 2994.
E-mail addresses: wbcumt@163.com (B. Wang), lsgiwzshi@gmail.com (W. Shi), cumtley@hotmail.com (E. Liu).

http://dx.doi.org/10.1016/j.jag.2014.08.012
0303-2434/

with tension, upon natural landscapes with differing terrain morphologies and different geographical scales, pointing out that DEM interpolated accuracy is related to landform types, sampling data density and the spatial scale. Guo et al. (2010) investigated the effects of topographic variability and sampling density on the accuracy of LiDAR DEM generated by several interpolation methods at different spatial resolutions. Liu et al. (2012) presented as a new methodology to DEM accuracy assessment based on approximation theory and illustrated its application to DEMs created by linear interpolation using contour lines as the source data.

Actually, statistical testing is an alternative approach to ensure the spatial data quality requirements being met at a moderate cost. Existing spatial data accuracy standards, such as the National Standard for Spatial Data Accuracy (NSSDA) used in the United States, commonly assume that positional error of spatial data are normally distributed (Zandbergen, 2008). However, such straightforward assumption may become a bottleneck of these methods for coping with the interference of outliers in reality. For example, DEM source data from airborne LiDAR may contain both low and high outliers due to the presence of high-rise buildings or flying objects, like birds. As a result, it makes sense to develop robust methods for measuring DEM accuracy, thereby yielding more reliable and not biased assessment results, particularly for those cases with heavily contaminated outliers.

Robust approaches to DEM accuracy assessment have aroused much attention in recent studies. For instance, Aguilar et al. (2007) explored three approaches for DEM accuracy assessment, the third one of which was a non-parametric approach based on the theory of estimating functions without the assumption of normal distribution. Zandbergen (2008) stated that a non-normal distribution of positional errors in spatial data had implications for spatial data accuracy standards and error propagation modeling, hence promoted specific recommendations were then made for revising the NSSDA. Besides, Höhle and Höhle (2009) proposed alternative robust statistical measures such as median, normalized median absolute deviation, and sample quantiles for replacing the traditional error indicator – RMSE for the accuracy assessment of DEMs derived from laser scanning and automated photogrammetry. Meanwhile, they also discussed requirements regarding the DEM reference data and employed the bootstrap technique for constructing the confidence intervals of each robust statistical measure, as well as furthermore settling the question of how large a sample size is requisite so as to obtain sufficiently precise estimates of standard deviation and the sample quantiles, such treatment scheme thereby establishing a potential systematic framework for DEM accuracy assessment by robust statistical methods.

This study is devoted to further developing innovative robust statistical methods for the DEM accuracy assessment. Ideally, a DEM accuracy assessment result should not be affected distinctly by any possible outliers from the raw data with its source error of approximate normal assumption. This research not only presents the robust statistics for quantitatively accessing DEM accuracy, but also provides their corresponding confidence intervals to indicate the overall variation of the DEM accuracy. In this study, the interpolated residual population is firstly generated for each different sampling site. After that, three average error statistics are then examined upon these residuals for measuring the DEM accuracy. Corresponding confidence intervals for each statistic are then further constructed for characterizing the error variation of DEM quality. The first approach is the t-distribution-based method as proposed by Aguilar et al. (2007), while the other two approaches are proposed, mainly based on robust theories, aiming to counteract the outlier effects or even the skew distribution circumstance of the interpolated DEM residuals. Performance of these three DEM accuracy assessment methods has been verified using the ASTER GDEM (the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model) data from Shannxi Province of China typically due to the data's multiple morphological characteristics. Regarding the ASTER GDEM data, Slater et al. (2011) carried out empirical accuracy assessment upon a globally-distributed sample dataset and conducted statistical analyses by comparing the GDEM using known reference DEMs and ground control points (GCPs), with the results revealing a systematic bias in the ASTER GDEM elevations, higher average noise levels, and a lower effective ground resolution, as well as numerous topographic artifacts and anomalies. Besides, Chrysoulakis et al. (2011) pointed out that the ASTER GDEM production overall failed to meet its pre-production estimated vertical accuracy by investigating the whole area of Greece. However, such ASTER GDEM data remained widely used in terrain surface reconstruction, geomorphological modeling and landscape visualization. Upon such ASTER GDEM data, the following experimental studies based on Monte Carlo simulations indicate that robust methods produce more reliable assessment results compared with the classical t-distribution-based method.

The rest of this paper is organized as follows. Section 'Problem definition' gives the description of the research problem. In section 'Three "average error" statistics and their corresponding confidence intervals', three statistical methods, together with their corresponding confidence intervals, are proposed for conducting the DEM quality assessment. Their performances are investigated in section 'Numerical experiments' by numerical experiment upon the actual DEM data sets. Section 'Results and analysis' presents the follow-up discussions and findings from the experiments. Conclusions are provided in the final section 'Conclusions'.

Problem definition

This study aims to develop innovative robust statistical methods upon the population of interpolated residuals for the DEM accuracy assessment. The procedure consists of first partially sampling seed points from the original DEM data. Upon these points an interpolated DEM surface is then generated using the bi-linear interpolation method. Checkpoints are chosen from the interpolated DEM, with their elevations compared with the corresponding "true values" from the original DEM data, thereby yielding the desired interpolated residual population. As a matter of fact, the bi-linear interpolation method is not absolutely requisite for these robust statistical methods presented in this paper. These interpolation methods can be alternatively extended to any other interpolation methods or simply differencing any two DEM data sets for generating the residual population since our proposed robust models for assessing the DEM accuracy are totally based on the residual population and the fundamental starting point is to provide robust alternative error indicators for the traditional RMSE.

Suppose that the residuals population U is a random variable, from which finite samples, *i.e.* $U_1, U_2, \ldots, U_n$ are extracted (as shown in Fig. 1) for assessing the interpolated DEM accuracy. The squared residuals population can be denoted by $V = U^2$, resulting $V_1, V_2, \ldots, V_n (i = 1, 2, \ldots, n)$ to be the corresponding squared residual samples.

Specifically, the commonly employed Mean Squared Error (MSE) is a global error indicator, calculated by

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} V_i}{n} = \frac{\sum_{i=1}^{n} U_i^2}{n}. \quad (1)$$

The MSE is a quadratic scoring indicator which measures the average magnitude of error, weighing larger errors more heavily than smaller ones (Januchowski et al., 2010). In other words, the MSE indicator may behave well upon residuals of normal distribution or the circumstance with non-existent outliers. To cope with this drawback, three "average error" statistics (a) mean, (b) median, and (c) M-estimator, are fully explored for the DEM accuracy assessment based on the above squared residuals. Most noteworthy is

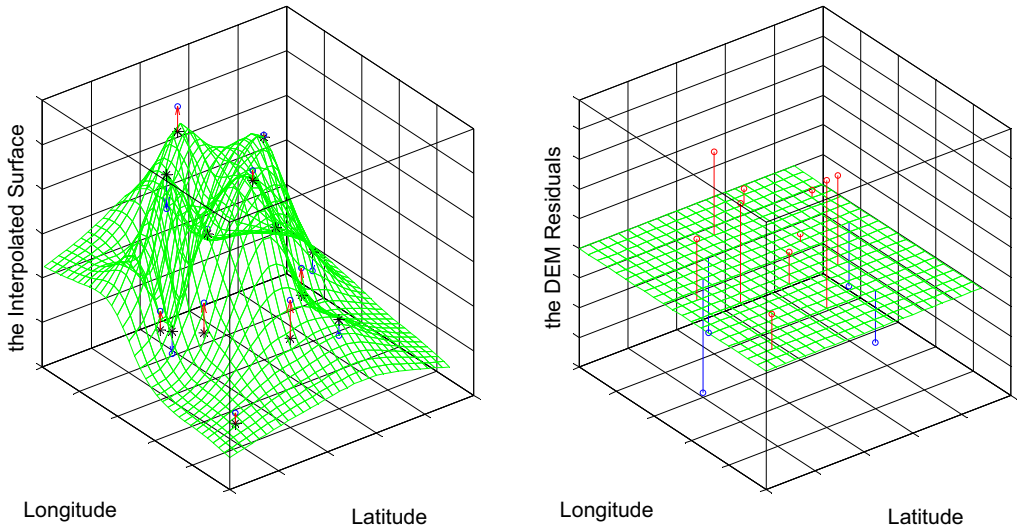

Fig. 1. A conceptual representation of DEM residuals between the original terrain surface and the interpolated one.

that no particular assumptions of distribution preference is necessary for the latter two robust statistical measures, median and M-estimator, for counteracting the outlier effects.

Three "average error" statistics and their corresponding confidence intervals

In this section, three "average error" statistics for squared interpolated residual population are presented, as well as their corresponding confidence intervals.

The t-distribution-based method

Firstly, by means of the Central Limit Theorem and the properties of Student t distribution, Aguilar et al. (2007) proposed an asymptotic approach to the development of the MSE confidence interval for the parameter μ (the expectation of population V) with a confidence level of $1-\alpha$ $(0<\alpha<1)$ as follows:

$$\left[\bar{V} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right], \tag{2}$$

where $\bar{V}$ is the sample mean of the population V and S is the sample standard deviation. The upper-tail value of the t-distribution is $t_{\alpha/2}(n-1)$ for n–1 degrees of freedom with a probabilistic area of $\alpha/2$ to its right. For instance, if the sample size is $n=25$ (*i.e.* 24 degrees of freedom), the tabulated value for $t_{0.025}$ corresponds to 2.0639. According to Eq. (2), it can be asserted that the true value of the expectation μ falls within the above interval with a confidence level of $100(1-\alpha)\%$. This interval defined in Eq. (2) can also be perceived as a characterization of the distribution uncertainty of the sample mean as it locates at the midpoint of that range. This model is hereafter named as the t-distribution-based method since the derivation of this confidence interval is mainly based on the student t distribution.

The Maritz–Jarrett median estimator method

As indicated above, the t-distribution-based method estimated the true MSE value under the normal distribution assumption of the squared population residuals. One concern is that this method may not be appropriate when the residuals contain significant outliers. Thus, it makes sense to develop more robust methods for coping with this issue as also raised by Aguilar et al. (2007). The following two sections are devoted to the introduction of two robust estimators, the median and the M-estimator, as well as their corresponding confidence intervals, for a squared residual population.

The explanation of the median estimator (Höhle and Höhle, 2009; Wilcox, 2012) is carried out for the squared residual population V in this section. When comparing two or more groups, the most common strategy is to use a single measure of location, the median (or 0.5th quantile) is usually an obvious choice, the corresponding sample median, M of which, is defined as

$$M = \begin{cases} V_{(n+1/2)} & \text{if } n \text{ is odd,} \\ \frac{1}{2}[V_{(n/2)} + V_{((n/2)+1)}] & \text{if } n \text{ is even.} \end{cases} \tag{3}$$

By means of the first-order statistic derived by Maritz and Jarrett (1978), the standard error of the sample median can be estimated as follows. Firstly, the function of beta probability density with positive integers a and b, is introduced, with the formula of

$$f(x) = \frac{(a+b+1)!}{a!b!} x^a (1-x)^b, \quad 0 \le x \le 1.$$

Suppose the random variable Y obeys a beta distribution (Johnson and Kotz, 1970) with $m = \lfloor (n+1)/2 \rfloor$, $a = m-1$ and $b = n-m$, where $\lfloor \cdot \rfloor$ is the floor function, rounding to the left closest integer on the axis. Further define $\omega_i = P\left(\frac{i-1}{n} \le Y \le \frac{i}{n}\right)$ and construct $C_k = \sum_{i=1}^{n} \omega_i V_{(i)}^k$ to estimate the k-th moment of sample median statistics, namely $E(M^k)$. Therefore, the standard error of the sample median, M, can be estimated with

$$\sigma(M) = \sqrt{E(M^2) - (E(M))^2} = \sqrt{C_2 - C_1^2}. \tag{4}$$

Consequently, it is reasonable to declare $Z = (M - \text{median})/\sigma(M)$ follows the standard normal distribution. An approximate $100(1-\alpha)\%$ confidence interval with median as its midpoint can then be established as

$$[M \pm Z_{\alpha/2}\sigma(M)]. \tag{5}$$

where $Z_{\alpha/2}$ is the upper-tail value of a standard normal distribution with an probabilistic area of $\alpha/2$ to its right. For instance, the tabulated value for $Z_{0.025}$ corresponds to 1.96. Formula (5) above indicates that the true median value of the squared DEM residuals population falls within such an interval consisting of its upper and lower limits with a confidence level of $100(1-\alpha)\%$.

The percentile bootstrapping M-estimator method

This section is devoted to the introduction of M-estimator (Wilcox, 2012) for the squared residuals population V, as well as its corresponding confidence interval constructed by the percentile bootstrap method. Actually the M-estimator is a maximum likelihood type of location measures including mean estimator as a special case. The following computation of the M-estimator, together with its confidence interval, appears to be more complicated than the foregoing two statistical methods.

Apparently, the expected value μ is the closest point to all possible squared residual values of V in terms of the expected squared distance, namely, $\mu = \frac{1}{n}\sum_{i=1}^{n} V_i = \underset{c}{\operatorname{argmin}} E(V-c)^2$. However, one practical concern is that $E(V-c)^2$ distributes inordinate weight amounts to the squared residual values in V which are far away from the aggregative center. Especially for a skewed distribution, extreme and relatively rare values can "pull" the value of μ into the tail of the distribution. In other words, μ is not a robust statistic for measuring the center of one population which has more general statistical distribution characteristics. This actually accords with the original motivation for the promotion of the M-estimator method (Wilcox, 2012).

In order to touch the more general case, let $\xi(V-\mu_m)$ be a function measuring the distance from μ_m, and ψ be its derivative with respect to μ_m. A location estimator, $\mu_m = \operatorname{argmin} E(\xi(V-\mu_m))$, is closest to all possible values of V as measured by its expected distance. Such μ_m is named the M-estimator, which satisfies the equation $E(\psi(V-\mu_m))=0$. It can be made scale equivariant by incorporating a measure of scale as follows;

$$E\left\{\psi\left(\frac{V-\mu_m}{\tau}\right)\right\}=0, \tag{6}$$

where τ is an appropriate adjustment factor measuring the scale of V. For the discrete squared residuals, the M-estimator, μ_m, is calculated with summation in Eq. (6). That is,

$$\sum_{i=1}^{n}\psi\left(\frac{V_i-\hat{\mu}_m}{\tau}\right)=0. \tag{7}$$

Three immediate followed concerns must be addressed before turning the M-estimator into practical applications, including the choice of: (a) ψ, (b) measure scale τ, and (c) a appropriate method for estimating $\hat{\mu}_m$ once ψ and τ have been determined. Martin and Zamar (1993) preferred the following Huber's ψ by which M-estimator demonstrated both quantitative and qualitative robustness,

$$\psi(x)=\max\{-K,\min(K,x)\}. \tag{8}$$

Here the 0.9 quantile of the standard normal distribution, i.e. $x=1.2816$ ($1/\sqrt{2\pi}\int_{-\infty}^{x} e^{-t^2/2}dt \approx 0.9$), is a common choice for the truncated boundary K (Huber and Ronchetti, 2009). The scale measure τ can be determined by $P(|V-V_{0.5}|\le\omega)=1/2$ (Wilcox, 2012), where $V_{0.5}$ is the population median. Suppose Z follows the standard normal distribution, implying $Z_{0.5}=0$. If τ satisfies $P(-\tau<Z<\tau)=0.5$, then τ is the 3/4 quantile of the standard normal distribution, approximately 0.6745, in that $1/\sqrt{2\pi}\int_{-\infty}^{0.6745} e^{-t^2/2}dt \approx 0.75$.

For measuring the scale τ, the absolute deviation statistic is employed upon the squared residuals by defining

$$\mathrm{MAD}=\mathrm{Median}\{|V_1-M|,|V_2-M|\ldots|V_n-M|\}, \tag{9}$$

where M is the sample median previously described in section 'The Maritz–Jarrett median estimator method'. In other words, MAD is the sample median of the n absolute deviation values $|V_1-M|,\ldots,|V_n-M|$ with its finite-sample breakdown point being 0.5 (Gather and Hilker, 1997).

Naturally a robust statistic should at least still be applicable in the simple normal distribution case, however, MAD does not estimate the standard deviation σ, but rather the $Z_{0.75}\sigma$, where $Z_{0.75}\approx 0.6745$ is the foregoing mentioned 3/4 quantile of the standard normal distribution. For this reason, the median absolute deviation typically needs to be normalized by dividing $Z_{0.75}$ so as to exactly estimate σ when the samples come from a normal distribution, namely,

$$\mathrm{MADN}=\frac{\mathrm{MAD}}{Z_{0.75}}\approx\frac{\mathrm{MAD}}{0.6745}. \tag{10}$$

For the squared residual samples, M-estimator of $\hat{\mu}_m$ then satisfies

$$\sum_{i=1}^{n}\psi\left(\frac{V_i-\hat{\mu}_m}{\mathrm{MADN}}\right)=0. \tag{11}$$

Solution to the above equation can be achieved through the following Newton–Raphson iterative algorithm.

Algorithm 1. Newton–Raphson Method for Computing the M-Estimator $\hat{\mu}_m$.

Step 0: Initialize $k=0$, $\hat{\mu}_k = M$ (the sample median), set $K=1.2816$ and calculate MADN according to Eqs. (9) and (10).

Step 1: Set $H=\sum_{i=1}^{n}\psi\left(\frac{V_i-\hat{\mu}_k}{\mathrm{MADN}}\right)$, Here ψ is given by Eq. (8).

Step 2: Set $DH=\sum_{i=1}^{n}\psi'\left(\frac{V_i-\hat{\mu}_k}{\mathrm{MADN}}\right)$, where $\psi'(x)=\begin{cases}1 & \text{if } -K\le x\le K\\ 0 & \text{otherwise}\end{cases}$.
Typically, DH counts the number of observations V_i satisfying $-K\le(V_i-\hat{\mu}_k)/\mathrm{MADN}\le K$.

Step 3: Update $\hat{\mu}_{k+1}=\hat{\mu}_k+\frac{\mathrm{MADN}\times H}{DH}$.

Step 4: If $|\hat{\mu}_{k+1}-\hat{\mu}_k|<\varepsilon=10^{-6}$, terminate iteration and obtain $\hat{\mu}_m=\hat{\mu}_{k+1}$. Otherwise, make increment k by 1 and repeat the steps 1–4 above.

In the following, confidence intervals corresponding to each "average error" statistics of the DEM squared residuals are constructed based on the Percentile bootstrap method (Wilcox, 2012) described below.

Bootstrap methods normally output new samples, $V_1^*, V_2^*, \ldots, V_n^*$, by randomly resampling with replacement from all the n squared residuals, $V_1, V_2, \ldots, V_n$. Apparently, different groups of bootstrap samples probably yield different bootstrap M-estimators. Suppose $\hat{\mu}_m^*$ be the M-estimator of $\hat{\mu}_m$ calculated by Algorithm 1 based on one time bootstrap resampling. A successive repetition of this process B times yield B bootstrap M-estimators: $\hat{\mu}_m^*1, \hat{\mu}_m^*2, \ldots, \hat{\mu}_m^*B$. If we denote $\bar{\mu}_m^* = 1/B\sum_{i=1}^{B}\hat{\mu}_m^*i$, the standard deviation of these M-estimators can then be calculated by

$$\hat{\sigma}_m=\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\hat{\mu}_m^*i-\bar{\mu}_m^*)^2}. \tag{12}$$

After a fairly large enough bootstrap number B (10,000 for instance) is configured, the new iteratively produced M-estimators can be obtained. Assume they all have been sorted in ascending order as $\hat{\mu}_m^*(1)\le\hat{\mu}_m^*(2)\le\cdots\le\hat{\mu}_m^*(B)$. Then, the confidence interval with an approximate $1-\alpha$ confidence level for the M-estimator $\hat{\mu}_m$ based on the squared residuals V can be defined as follows,

$$[\hat{\mu}_m^*(l),\quad \hat{\mu}_m^*(u)], \tag{13}$$

where $l=\lceil\alpha B/2\rceil$, $u=\lfloor(1-\alpha/2)B\rfloor$; $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are the floor and ceiling functions which map a real number to the largest previous or the smallest following integer, respectively. Formula (13) above indicates that true value of the M-estimator for the squared DEM

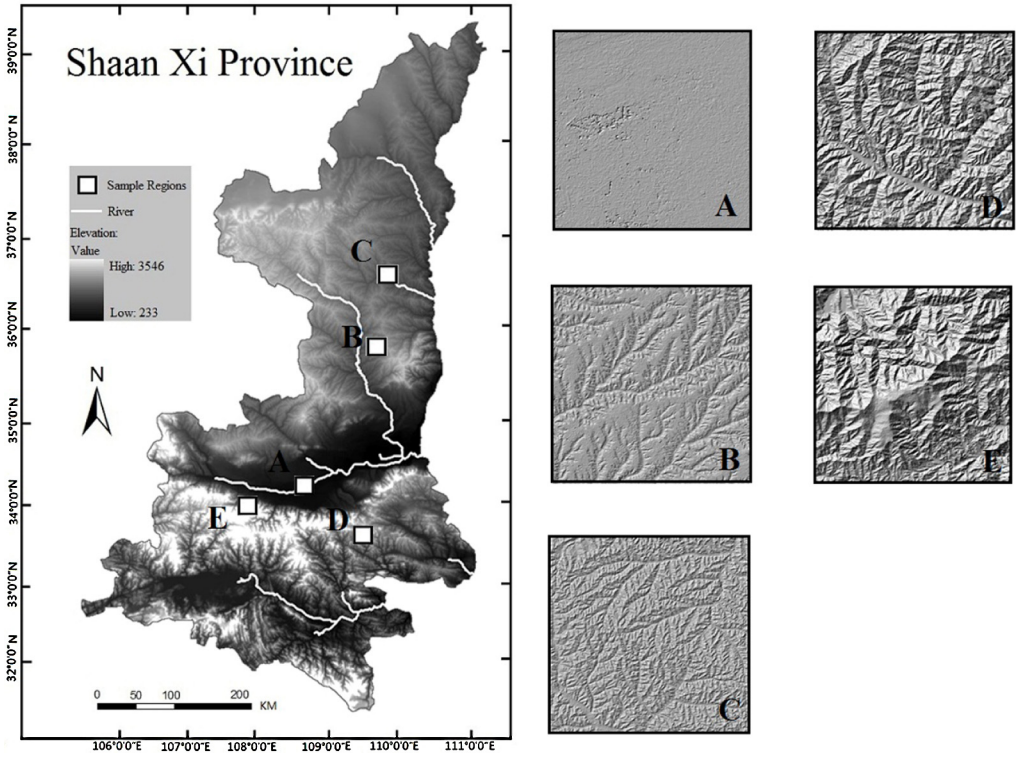

Fig. 2. The five sample regions from Shannxi province.

residuals falls into such an interval established using the upper and lower limits with the possibility of $100(1-\alpha)\%$.

Numerical experiments

The study area was chosen from Shannxi Province (left part of Fig. 2), locating in the hinterland of China, due to its multiple and diverse morphological characteristics, which are most suitable for conducting the performance examinations of here described robust statistical methods.

In this study, five sample regions were selected from this area with each sample region representing one typical topographic feature. Sample region A: between 34.06°–34.34° N and 108.56°–108.84° E is situated in the central Shaanxi plain, *i.e.* a 'plain' area type. Sample region B: between 35.59°–35.86° N and 109.40°–109.68° E, is located at the Loess Tableland area of Shaanxi, which is a transitional area between plain and hilly terrain. The sample region C: between 36.31°–36.59° N and 109.22°–109.50° E is located in the Loess hill and gully area of northern Shannxi, the topography of which is 'hilly'. Sample region D: between 33.44°–33.72° N and 109.27°–109.55° E, is located on a south-facing slope of the Qin Ling Mountains, a 'middle mountain' topography type. Sample region E: between 33.81°-34.09° N and 107.60°–107.88° E, is on the north slope of the Qin Ling Mountains, thus belonging to the 'high mountain' type. All these five sample regions are illustrated in Fig. 2, with their characteristics described in Table 1.

In this section, performance of three statistical methods for the DEM accuracy assessment presented in section 'Three "average error" statistics and their corresponding confidence intervals' is fully verified by the following numerical experiments, using data sets of ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) Global Digital Elevation Model, which is a joint operation between the United States National Aeronautics and Space Administration (NASA) and Japan's Ministry of Economy, Trade and Industry (METI). The ASTER GDEM data is released in GeoTIFF format with geographic lat/long coordinates and a 1 arc-second (30 m) grid of elevation postings. It is referenced by the WGS84/EGM96 geoid. Pre-production estimated accuracies for this global product were 20 meters at a 95% confidence level for vertical data and 30 meter at 95% confidence level for horizontal data (Pavelka, 2009). The selected ASTER DEM raw data has firstly been converted into the projection of UTM/WGS84 for conveniently extracting the terrain elevation values. Five topographic surfaces were gouged measuring 1000-by-1000 pixels from the original dataset as the test data for the following experiments.

The residual population for each morphology district was obtained following the operational flowchart drawn in Fig. 3. Seed data sets, composed of the X, Y, and Z coordinates of 2^{13} (the sampling fraction is approximately equal to 8.2‰) ground points, were extracted from the original grid DEM by the stratified random sampling (four-by-four sampling quadrants), which guarantees the sampled seed data with a homogenous distribution over the whole considered area. Based on these sampled seed data sets, a DEM with a grid resolution spacing same as the original grid DEM can

Table 1
Characteristics of the five sample regions.

Terrain descriptive statistics	Sample region				
	A	B	C	D	E
Min (m)	206	825	90	555	865
Max (m)	534	1424	1467	2000	3754
Average elevation (m)	389	1117	1196	1159	2184
Std dev. of elevation (m)	17.33	102	87	243	583
Min slope (°)	0	0	0	0	0
Max slope (°)	71.81	61.08	61.98	72.68	76.83
Average slope (°)	4.72	11.48	15.23	25.85	28.25
Std dev. of slope (°)	5.42	7.77	7.48	10.34	11.09

Note: All surfaces are grid DEMs composed of 10^6 points and with $30\,m \times 30\,m$ spacing.

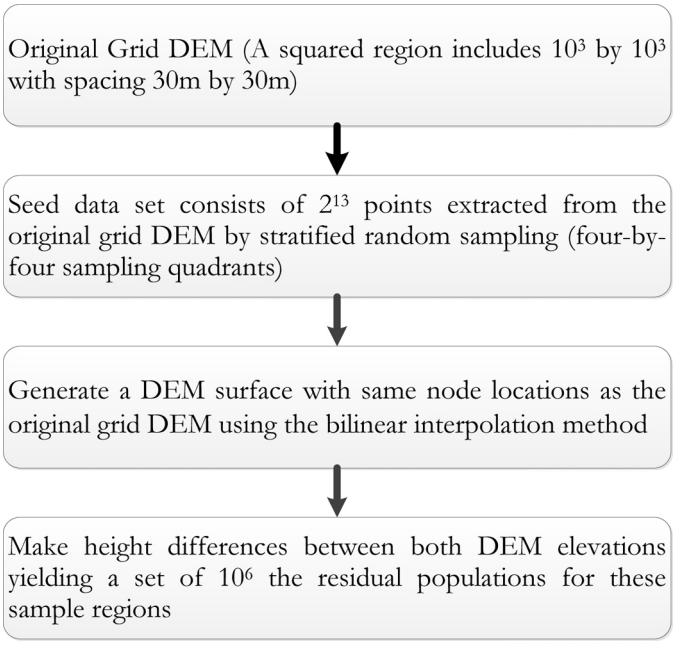

Fig. 3. Logic flow of generating the interpolated residuals for each sample region.

accordingly be generated for each terrain type using the bilinear interpolation method. As a result, the new interpolated and original DEM grids were spatially coincident and equal in resolution spacing. Height differences between these two DEM elevation surfaces at each node can then be made in each sampling region, thereby yielding a set of 10^6 sample residuals. For each different morphology district, the statistical characterizations were described in Table 2 and the intuitive visual inspections were illustrated in Fig. 4 for measuring the deviation of the interpolated residuals from standard normal distribution by the *Q*–*Q* plot and histogram distribution tools.

Table 3 tabulates the standard deviations of the three statistical indicators for the squared residual population from five sample regions. The standard error of the MSE statistic is computed using $S/\sqrt{n}$ in terms of Eq. (2), while the standard error of the median estimator is calculated according to Eq. (4) of the Maritz-Jarrett median estimate and the standard error of the M-estimator obtained in accordance with Eq. (12). The corresponding 95% confidence level intervals which character the distribution uncertainties of these squared residuals are also calculated for each of the five sample regions as presented in Table 4. The MSE confidence intervals are computed using formula (2). Confidence intervals for the median values of the squared residuals are calculated in line with formula (5), and the confidence intervals of the M-estimator for the squared residuals population are computed according to formula (13).

Monte Carlo simulations are employed here to explore the asymptotic convergence behaviors of the 95% confidence intervals in terms of the t-distribution-based method, the Maritz-Jarrett median estimator and the percentile bootstrapping M-estimator methods as discussed in section 'Three "average error" statistics and their corresponding confidence intervals' for the squared residual population generated from those five sample regions. In the first phase, using the simple random sampling approach, we pick out a series of samples X_i from the squared residual population in each of the five regions of sampling size N_i, which is assigned the values from 10 to 2000 with an increment interval of 10. For each group of the selected residuals X_i, the 95% confidence intervals corresponding to MSE, median and M-estimator of the samples can then computed same as previously described. And this procedure is repeated for 500 times so as to simulate the average distribution for each confidence interval. Finally, asymptotic curves are drawn by connecting the average end-points of all these 500 confidence intervals with a dash-dot line for MSE statistic, a solid line for the median estimator and a dashed line for the M-estimator in each of the five regions. Fig. 5 graphically illustrates with a comparative exhibition of all these asymptotic convergence characteristics.

Results and analysis

Table 2 provides the statistical characteristics of the residuals population for each of the five sample regions. The absolute mean values exert close relationship with the terrain complexity in each region. Apparently, these values appear smaller for a flat topography area and larger for rugged topography. Besides, it is found that systematic errors do obviously dominate and the standard deviations also reflect the roughness of the sampling regions. Meanwhile, the skewness statistic is also employed for measuring the asymmetry of the distribution, qualitatively negative skew indicating the left tail gathered, positive one corresponding to a right tail while values close to zero imply a relatively symmetric distribution. As can be seen in Table 2, the skewness value of region A is obviously deviating from zero compared with those of other regions. It appears that these residual populations are more likely to follow the normal distributions. However, the standardized

Table 2
Statistical characteristics of the interpolated residuals for the five sample regions.

Sample region	Mean (m)	Standard deviation (m)	Standardized		
			Skewness	Kurtosis	K–S statistic
A	0.2670	22.2284	−0.3942	11.9831	0.3866
B	0.7334	83.7219	0.0127	3.6343	0.4612
C	1.8055	109.1615	0.0884	3.9782	0.4693
D	3.1054	222.4256	0.0246	4.6869	0.4833
E	5.7406	699.2751	−0.1178	4.2724	0.4934

Table 3
The "average error" statistics of the squared residual population for the five sample regions (m^2).

Sample region	MSE	Median	M-estimator
A	15.9008	1.6827	2.3758
B	111.1450	64.0744	91.5513
C	218.8685	104.9526	142.2986
D	1046.6303	397.3977	547.7505
E	9253.7613	3024.5547	4727.0269

Table 4
The 95% confidence level intervals of the statistics of the squared residual population for the five sample regions (m^2).

Sample region	MSE	Median	M-estimator
A	$10^2 \times (4.267, 4.891)$	$10 \times (4.307, 4.967)$	$10 \times (6.757, 7.693)$
B	$10^3 \times (6.697, 7.133)$	$10^3 \times (1.650, 1.902)$	$10^3 \times (2.602, 2.960)$
C	$10^4 \times (1.179, 1.265)$	$10^3 \times (2.457, 2.869)$	$10^3 \times (3.896, 4.453)$
D	$10^4 \times (4.894, 5.309)$	$10^4 \times (1.039, 1.195)$	$10^4 \times (1.636, 1.849)$
E	$10^5 \times (4.910, 5.273)$	$10^4 \times (6.422, 7.608)$	$10^5 \times (1.070, 1.259)$

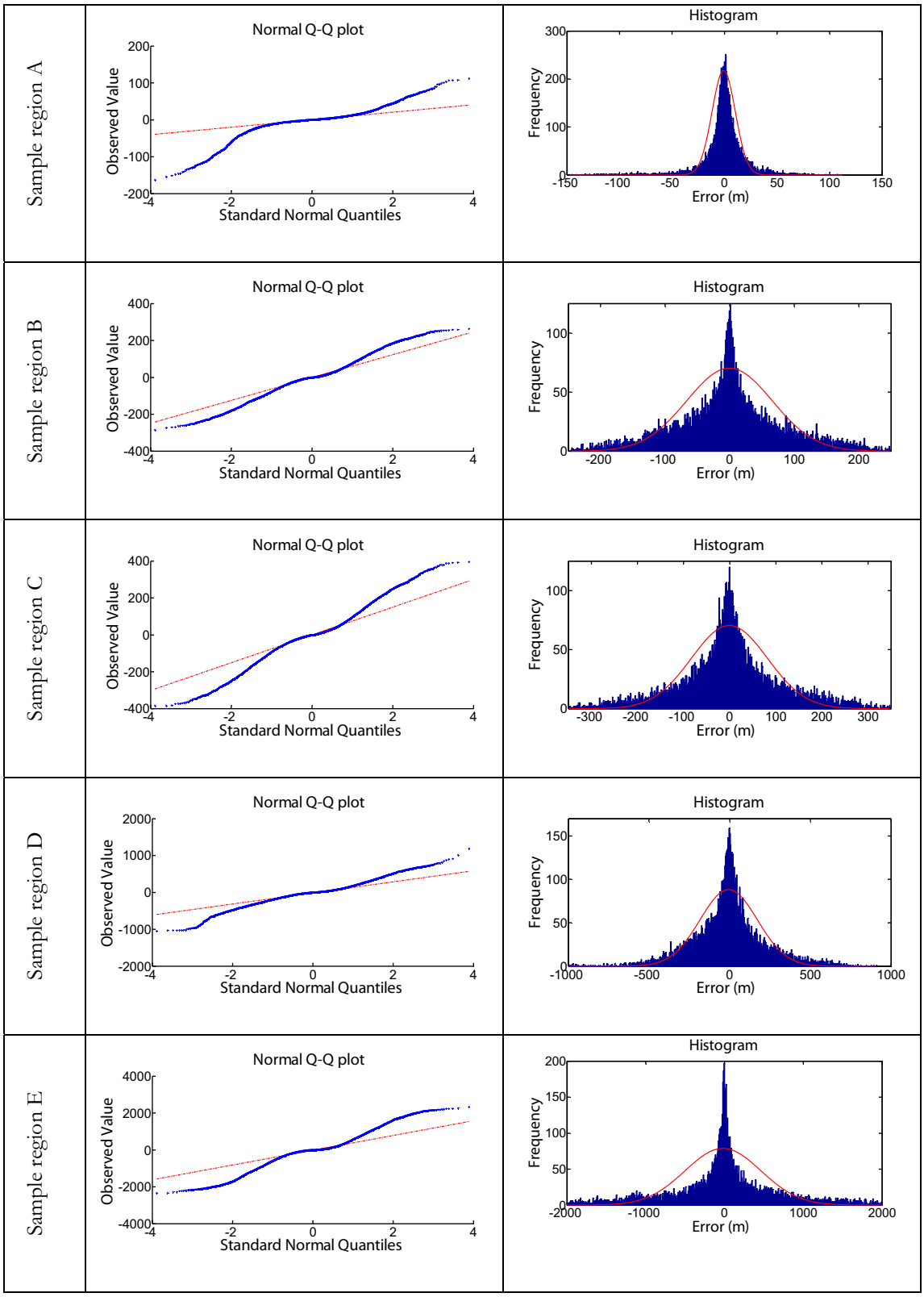

Fig. 4. Distributions of the interpolated residuals for the five sample regions.

kurtosis values are all positive for these five test regions, indicating that the residual population for each of the study region evidently obeys a leptokurtic distribution. This phenomenon is perceived to be caused by the presence of outliers. The K–S statistic (abbreviation of the Kolmogorov–Smirnov goodness-of-fit statistic) is also employed to examine the standardized sampling residuals for normality of the distribution, with its critical value at the 95% confidence level being 0.0136. However, the K–S statistic values shown in Table 2 are all beyond such critical value. Moreover, histograms illustrated in Fig. 4 obviously drag thicker tails on both sides and

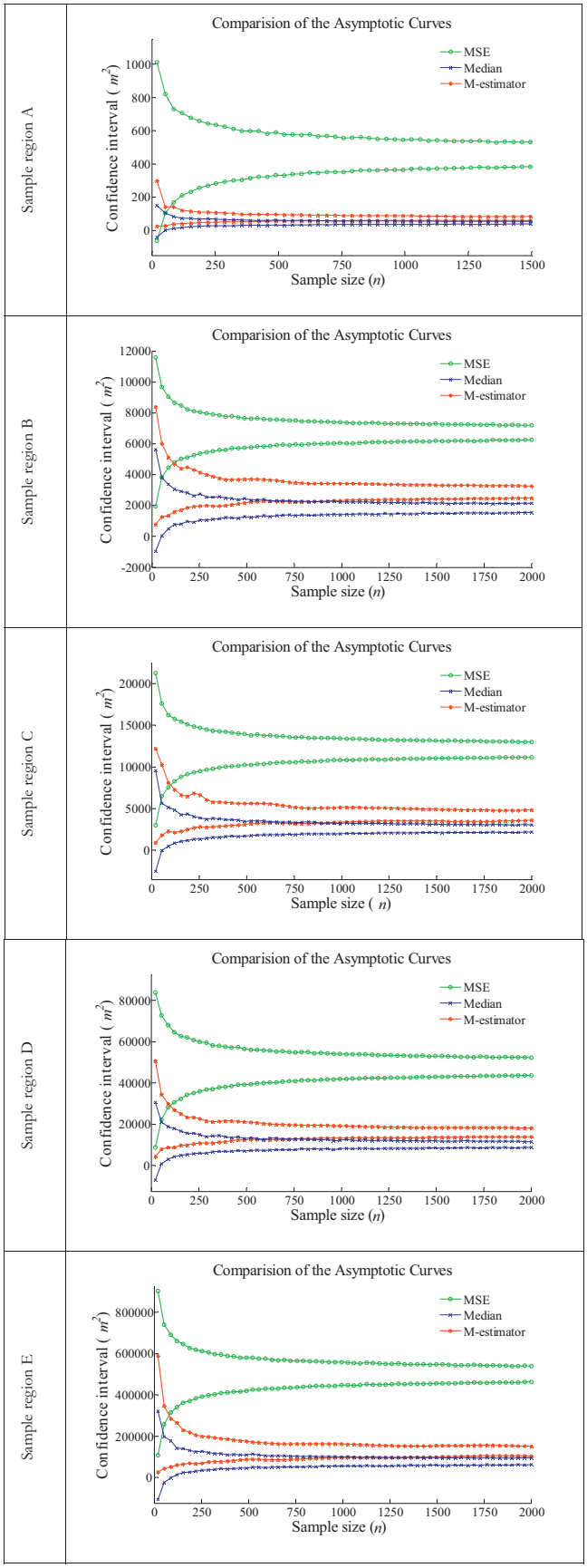

Fig. 5. Graphical representation of asymptotic convergence behaviors of the 95% confidence intervals resulting from the constructed Monte Carlo simulations using the MSE, median and M-estimator for the squared residual population.

have leptokurtic peaks at the center. If a residual population follows normal distribution, the plot should normally stretch to be a straight line. However, Fig. 4 illustrates that the *Q*–*Q* plots for the residual populations of the five sample regions deviate greatly from the indicated straight lines. Therefore, it can be concluded that these residual populations for all five sample regions fail to completely follow one normal distribution due to the contaminated outliers, thereby making the traditional error indicator – RMSE inappropriate for the DEM accuracy assessment.

Table 3 tabulates the "average error" statistics of the squared residuals population statistics for the five sample regions. It is noted that the median estimator deflects slightly, the M-estimator ranks second and MSE provides the largest deviations. The above experimental results demonstrate that the latter two robust estimators possess better capacity for counteracting the outlier effects or skew distributions from the interpolated residuals during the DEM accuracy assessment. A visual comparison exhibition of the confidence intervals corresponding to these three statistics is displaying in Fig. 5, which estimate the "average center" for the squared interpolated residuals of the five sample regions. The confidence interval widths for both the median estimator and the M-estimator are much narrower than that of the MSE estimator, which theoretically is more appropriately applicable to normal distribution case. The striking contrast shown in Fig. 5 fully indicates the superiority of proposed robust methods which provide better DEM accuracy assessment compared with that given by the traditional approach. Such superiority seems more noticeable when the residual population does not completely obey the normal distribution or contains certain amount of outliers.

It is further illustrated in Fig. 5 of the asymptotic convergence behavior of the 95% confidence intervals, based on the Monte Carlo simulations using the MSE, median and M-estimator for the squared interpolated residuals population of the five sample regions. The asymptotic behaviors of all the confidence intervals are demonstrated according to the increasing sample size, bringing about the corresponding progressively narrower confidence intervals until they gradually becoming stable. One reason may account for this finding should be the theory of large numbers since the sampling "average center" would converge in probability as the sample size grows toward infinity.

From Fig. 5, it can also be concluded that the MSE confidence interval is clearly wider than that from median and M-estimator for these five testing regions. This indicates that the median and M-estimator possess higher capability of accurately charactering the interpolated residual distribution when assessing the DEM accuracy than the *t*-distribution-based method derived under the assumption of normal distribution. The latter two methods are born with a greater tolerance (robustness) for non-normal distributions.

It should be pointed out that these three "average error" statistics discussed in this paper do not make a difference between positive and negative residual values. Other estimators, such as the mean error, median error, normalized median absolute derivation (NMAD) and sample quantiles, are thus recommended as auxiliary indicators for further DEM accuracy assessment, mainly because these estimators are single and global accuracy measures with the advantage of application convenience. However, a single statistic appears insufficient to suitably and thoroughly reflect as much of the information from DEM error distribution. In-depth, the DEM error can be deemed as a spatial variable, related to morphometric landform characteristics. As a result, appropriate regression methods can then be employed, with all order coefficients of the regression equation estimated by the proposed robust statistical methods. Mapping union of the overall parameters from the corresponding confidence intervals therefore generates the top and bottom fluctuating DEM error confidence surfaces, which promote

an intuitively better understanding of the accuracy assessment for DEM surface.

With regard to sample size, the National Standards for Spatial Data Accuracy (NSSDA) suggests that at least 20 checkpoints per terrain type should be put in place to estimate spatial data accuracy. NSSDA assumes that the data contains no systematic errors and that the errors follow a normal distribution, however, these prerequisites may not always be the case in practice. As illustrated in Fig. 5, at least 400 checkpoints are sampled for obtaining reliable intervals of the DEM accuracy assessment when the squared residual population fails to be normally distributed and even contains outliers. The underlying substance for such a big difference in the necessary sampling numbers from 10 to 400 or even more of constructing the confidence intervals corresponding to some error statistic should be attributed to the aforementioned inceptive normal distribution assumptions as it simplifies the circumstances in reality. The above discussions only broadly scratch the surface of issues in the development and application of innovative robust estimators for DEM accuracy assessment. Many specific concerns, however, may be involved in any concrete practice.

Conclusions

Three average error statistics: the mean, median and M-estimator for the squared interpolated DEM residuals have been thoroughly investigated regarding their robustness of the DEM accuracy assessment. Confidence intervals associated with certain confidence levels (95%) are also constructed for each of these statistics. Experimental results indicate that the median and M-estimator perform soundly well in counterstriking the outlier efforts from the interpolated DEM residuals compared with the case using mean (RMSE) to be the indicator for DEM accuracy assessment. In other words, median and M-estimator are more robust than the mean estimator when assessing interpolated DEM accuracy. Due to the common occurrence of non-normal distribution and outliers in actual interpolated DEM, the latter two robust methods are therefore strongly recommended for the henceforth DEM accuracy assessment.

As a matter of fact, many existing spatial data accuracy standards always assume that the interpolated residuals or DEM errors are normally distributed. It is, therefore, necessary to further renovate these accuracy assessment standards by supplementing robust statistical assessment methods being error indicators, which can alleviate the outlier effects and be applicable for the even skew distributed circumstance when assessing the DEM accuracy, without considering effects from the prerequisite normal distribution assumption of residual samples to check.

The bi-linear interpolation method, however, is not essential for these novel robust methods proposed in this paper. These robust statistical methods can be applicable to any other nonlinear interpolated DEM residuals or solely to compare any two groups of DEM data since these robust models totally operate upon only the generated DEM residuals. Besides, these three robust statistical methods are proposed based on statistical theories with assumption that the sampled DEM residuals are mutually independent. However, spatial autocorrelation seems to have varying analytical effects upon the error propagation determined by specific application. Thus, spatial autocorrelation should be further taken into considerations in the development of future innovative statistical methods for the DEM accuracy assessment.

Finally, from the experimental aspect in inspecting the robust estimators, ASTER DEM data can be replaced with DEM generated from LiDAR data for further exploit the performance of here proposed robust statistical methods of this study.

Acknowledgments

Mr. Bin WANG is the beneficiary of a doctoral grant from the **AXA Research Fund**. And Dr. Wenzhong SHI acknowledges the funding support from Ministry of Science and Technology of China (Project No.: 2012AA12A305, 2012BAJ15B04). Besides, the authors would also like to thank Teng ZHONG from the Hong Kong University for providing the ASTER GDEM data, Mrs. Elaine Anson of The Hong Kong Polytechnic University, for her careful language polishing work, and the handling editor's efforts, as well as several anonymous reviewers for their valuable comments.

References

Aguilar, F.J., Aguilar, M.A., Aguera, F., 2007. Accuracy assessment of digital elevation models using a non-parametric approach. Int. J. Geogr. Inf. Sci. 21, 667–686.
Chaplot, V., Darboux, F., Bourennane, H., Leguedois, S., Silvera, N., Phachomphon, K., 2006. Accuracy of interpolation techniques for the derivation of digital elevation models in relation to landform types and data density. Geomorphology 77, 126–141.
Chrysoulakis, N., Abrams, M., Kamarianakis, Y., Stanislawski, M., 2011. Validation of ASTER GDEM for the area of Greece. Photogramm. Eng. Remote Sens. 77, 157–165.
de Oliveira, C.G., Paradella, W.R., 2009. Evaluating the quality of the Digital Elevation Models produced from ASTER stereoscopy for topographic mapping in the Brazilian Amazon Region. An. Acad. Bras. Cienc. 81, 217–225.
Fisher, P.E., Tate, N.J., 2006. Causes and consequences of error in digital elevation models. Prog. Phys. Geogr. 30, 467–489.
Gather, U., Hilker, T., 1997. A note on Tyler's modification of the MAD for the Stahel–Donoho estimator. Ann. Stat. 25, 2024–2026.
Guo, Q.H., Li, W.K., Yu, H., Alvarez, O., 2010. Effects of topographic variability and lidar sampling density on several DEM interpolation methods. Photogramm. Eng. Remote Sens. 76, 701–712.
Höhle, J., Höhle, M., 2009. Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS J. Photogramm. Remote Sens. 64, 398–406.
Hu, P., Liu, X.H., Hu, H., 2009. Accuracy assessment of digital elevation models based on approximation theory. Photogramm. Eng. Remote Sens. 75, 49–56.
Huber, P.J., Ronchetti, E., 2009. Robust Statistics. John Wiley & Sons, Hoboken, NJ.
Januchowski, S.R., Pressey, R.L., VanDerWal, J., Edwards, A., 2010. Characterizing errors in digital elevation models and estimating the financial costs of accuracy. Int. J. Geogr. Inf. Sci. 24, 1327–1347.
Johnson, N.L., Kotz, S., 1970. Continuous Univariate Distributions. Hougton Mifflin, New York.
Liu, X.H., Hu, P., Hu, H., Sherba, J., 2012. Approximation theory applied to DEM vertical accuracy assessment. Trans. GIS 16, 397–410.
Maritz, J., Jarrett, R., 1978. A note on estimating the variance of the sample median. J. Am. Stat. Assoc. 73, 194–196.
Martin, R.D., Zamar, R.H., 1993. Efficiency-constrained bias-robust estimation of location. Ann. Stat. 21, 338–354.
Maune, D.F., 2007. Digital Elevation Model Technologies and Applications: The DEM Users Manual, 2nd ed. American Society for Photogrammetry and Remote, Bethesda.
Pavelka, K.,2009. Topographic and Thematic Mapping from Multi-Resolution Satellite Images. In: Proceedings of 24th ICA Conference. International Cartographic Association, Santiago, pp. 956–962.
Shi, W., 2010. Principles of Modeling Uncertainties in Spatial Data and Spatial Analyses. CRC Press/Taylor & Francis, Boca Raton.
Shi, W.Z.J., Li, Z.L., Bedard, Y., 2004. Theme issue: advanced techniques for analysis of geo-spatial data. ISPRS J. Photogramm. Remote Sens. 59, 1–5.
Shi, W., Wang, B., Tian, Y., 2014. Accuracy analysis of digital elevation model relating to spatial resolution and terrain slope by bilinear interpolation. Math. Geosci. 46, 1–37.
Slater, J.A., Heady, B., Kroenung, G., Curtis, W., Haase, J., Hoegemann, D., Shockley, C., Tracy, K., 2011. Global assessment of the new ASTER global digital elevation model. Photogramm. Eng. Remote Sens. 77, 335–349.
Wilcox, R.R., 2012. Introduction to Robust Estimation and Hypothesis Testing, 3rd ed. Elsevier Science & Technology, St. Louis, MO, USA.
Zandbergen, P.A., 2008. Positional accuracy of spatial data: non-normal distributions and a critique of the national standard for spatial data accuracy. Trans. GIS 12, 103–130.