

Research Article

Approximation Theory Applied to DEM Vertical Accuracy Assessment

XiaoHang Liu

*Department of Geography and
Human Environmental Studies
San Francisco State University*

Peng Hu

*School of Resource and
Environmental Engineering
AnHui University and
School of Resource and
Environmental Science
Wuhan University*

Hai Hu

*School of Resource and
Environmental Science
Wuhan University*

Jason Sherba

*Department of Geography and
Human Environmental Studies
San Francisco State University*

Abstract

Existing research on DEM vertical accuracy assessment uses mainly statistical methods, in particular variance and RMSE which are both based on the error propagation theory in statistics. This article demonstrates that error propagation theory is not applicable because the critical assumption behind it cannot be satisfied. In fact, the non-random, non-normal, and non-stationary nature of DEM error makes it very challenging to apply statistical methods. This article presents approximation theory as a new methodology and illustrates its application to DEMs created by linear interpolation using contour lines as the source data. Applying approximation theory, a DEM's accuracy is determined by the largest error of *any* point (not samples) in the entire study area. The error at a point is bounded by $\max(|\delta_{node}| + M_2 b^2/8)$ where $|\delta_{node}|$ is the error in the source data used to interpolate the point, M_2 is the maximum norm of the second-order derivative which can be interpreted as curvature, and b is the length of the line on which linear interpolation is conducted. The article explains how to compute each term and illustrates how this new methodology based on approximation theory effectively facilitates DEM accuracy assessment and quality control.

Address for correspondence: XiaoHang Liu, Department of Geography and Human Environmental Studies, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA. E-mail: xhliu@sfsu.edu

1 Introduction

Myriad applications count on digital elevation models (DEM) to infer topographical properties, thus DEM quality may have serious implications for geospatial activities. A pivotal indicator of DEM quality is its vertical accuracy which is the deviation of the estimated elevation (Z) from the ground truth (z), i.e. $|Z - z|$. In the literature, many methods have been proposed to assess DEM vertical accuracy but nearly all of them are statistics-based. Among them, variance and root mean square error (RMSE), which are based on the error propagation theory in statistics, have been most popular (Tempfli 1980, Li 1993, Huang 2000, Shi et al. 2005, Aguilar et al. 2010). This is exemplified by the U.S. National Standard on Spatial Data Assessment (NSSDA) which uses RMSE to assess DEM vertical accuracy (FGDC 1998). Despite the popularity of these methods, there have always been reservations on their effectiveness because the assumptions behind these methods contradict empirical observations (Fisher 1998, Wise 2000, Bonin and Rousseaux 2005, Oksanen and Sarjakoski 2006). Remedies have been proposed but they are still statistics-based, e.g. the 95th percentile method (NDEP 2004) and geostatistics (Kyriakidis and Goodchild 2006).

In previous research, we have presented approximation theory as a new viable methodology to assess DEM vertical accuracy (Hu et al. 2009a). Based on this methodology, a theoretical framework to assess DEMs generated by three interpolation methods was presented. The framework articulated for the first time the theoretical reasons behind several empirical observations, e.g. the correlation between DEM error, terrain complexity, and sampling density. However, it has not been shown how this framework can be applied to real-world analyses. This research fills this gap by explaining the framework in detail using a case study. Specifically, linear interpolation in 1D is applied to create a DEM using a topographic map as the source data. The vertical accuracy of the DEM is then assessed based on approximation theory. DEMs interpolated from topographic maps are currently used worldwide. It is a main type of DEM in the USGS's National Elevation Dataset (Gesch 2007). Until advanced technologies such as LiDAR and IFSAR become more affordable, DEMs generated from topographic maps are expected to remain popular. A case study of this type of DEM is thus valuable.

In the following sections, we first explain why statistical methodology is not applicable to vertical error assessment of DEMs generated by interpolation, then introduce approximation theory and its application to DEMs generated by linear interpolation in 1D. Results from a case study are then reported. The article ends by discussing the implications of the new methodology to DEM generation and quality control.

2 Challenges to Statistical Methodology

2.1 DEM Error Composition

To understand the fundamental challenges to statistical methodology, it is necessary first to clarify the nature and composition of DEM vertical error. Supposing T is a location, its vertical error ΔZ_T is the difference between the estimated elevation Z_T and the true elevation z_T , i.e. $\Delta Z_T = Z_T - z_T$. ΔZ_T is the sum of two components: interpolation error R_T and propagation error δ_T (Li 1993, Huang 2000, Hu et al. 2009a). Interpolation error is entirely due to the imperfection of the interpolation function and has nothing to do with

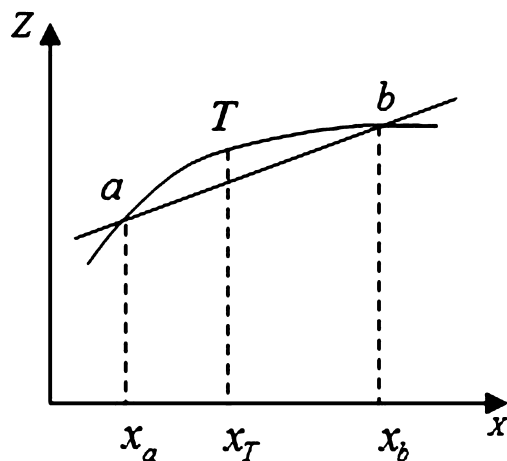


Figure 1 Linear interpolation in 1D. Note the convex flow line is approximated by a straight line, thus introducing a negative interpolation error

the source data. For example, when a flow line is approximated by a straight line (Figure 1) as in the case of linear interpolation in 1D, there is always an error even if the source data is perfect. If we use H_T to denote the elevation estimated using error-free source data, its deviation from the true elevation z_T defines interpolation error R_T , i.e. $R_T = H_T - z_T$. In reality, source data nearly always contain some errors such as random error, gross error, or measurement errors due to the confounding influences of canopy and ground vegetation, surface debris, buildings, and other objects, etc. The actual elevation recorded in a DEM is thus not H_T but Z_T . The difference between the two is the propagation error, i.e. $\delta_T = Z_T - H_T$. Propagation error δ_T describes how errors in the source data are propagated to T through interpolation. It depends on both the source data error and the interpolation function. Note δ_T should not be confused with Huevelink's (1998) propagation error which refers to the propagation of DEM error (i.e. ΔZ_T) to subsequent applications in environmental modeling and GIS.

2.2 Error Type of Each Component

According to error theory, there are three types of errors – random error, systematic error, and gross error. To understand the challenge to statistical methodology, the key is to articulate which type of error that propagation error and interpolation error belong to. In the literature, some studies have assumed both of them are random errors (Li 1993, Huang 2000, Aguilar et al. 2006) while the others suspected the existence of systematic error (Oksanen and Sarjakoski 2006). The research by Hu et al. (2009a) concludes that interpolation error is systematic error because its sign and magnitude are fixed. In the case of linear interpolation in 1D, interpolation error is always negative (i.e. underestimation) if the flow line in Figure 1 is convex and always positive (i.e. overestimation) if the flow line is concave. There is no randomness about the sign of interpolation error. Similarly, the magnitude of the interpolation error at a location is also fixed because its true elevation (z_T) and that estimated using error-free source data (H_T) are both determined values albeit unknown. Propagation error, on the other hand, is random error

under the best scenario that no systematic or gross error exists in the source data. Otherwise, it is not random error either but a mixture of random, systematic, and even gross error. Note, systematic and gross error in the source data may come from several sources. In that case, propagation error can be written as the sum of random error and each systematic error and gross error.

2.3 Challenges to Variance or RMSE Methodology

That interpolation error is systematic error determines that the vertical error at a location (ΔZ_T), which is the sum of interpolation and propagation error, cannot be random error. Rather, it is a mixture of random and systematic or even gross error. This directly challenges the applicability of error propagation theory which, as mentioned previously, is the rationale behind the most popular methods of using variance and RMSE to assess DEM vertical accuracy. Error propagation theory was first used by Greenwalt and Schultz (1962) to assess cartographic and geodetic data. According to the theory, if an error is the sum of several components (e.g. $\Delta Z_T = \delta_T + R_T$), each term can be quantified by its variance. Better yet, the variance of the total error is the sum of the variances of each component, e.g. $\sigma_{\Delta Z_T}^2 = \sigma_{R_T}^2 + \sigma_{\delta_T}^2$. This is the rationale behind the extensive use of variance and RMSE, which is equivalent to the square root of variance, in the literature.

However, error propagation theory hinges on a critical assumption that *all error components are random error and independent of each other*. Thus, in order to apply error propagation theory to DEM vertical accuracy, both propagation and interpolation error must be random error and independent of each other. Previous discussion has shown that interpolation error is not random error but systematic error, and propagation error may or may not be completely random error. Moreover, both interpolation and propagation error depend on the interpolation function, thus they are not necessarily independent. These reasons determined that error propagation theory is not applicable to DEM vertical error assessment. The validity of variance and RMSE are thus questionable.

In the literature, many studies have expressed concerns about RMSE based on their empirical observation that DEM errors do not seem to have a normal distribution (Fisher 1998, Wise 2000, Bonin and Rousseaux 2005, Oksanen and Sarjakoski 2006). It has to be pointed out that, while DEM errors are indeed not normally distributed, this is not the fundamental reason why RMSE is problematic. The fundamental reason is that DEM error is not random error but a mixture of random and systematic error or even gross error, consequently error propagation theory cannot be used to justify RMSE.

2.4 Challenges to Other Statistical Methodology

As a remedy to RMSE, other statistical methods have been proposed. One example is the modification to NSSDA proposed by the National Digital Elevation Program (NDEP 2004) by using the 95th percentile in areas where a normal-distribution error cannot be attained. A similar standard is used by FEMA to assess LiDAR-derived DEMs (FEMA 2010). Other statistics such as sample mean, range, maximum have also been proposed. While these methods, which do not require normal-distribution error, are an improvement, they do not solve the problem completely.

Central to much statistical inference is the assumption that observations are independent and have identical probability distributions. In the context of DEM error, this means the vertical errors from different locations must be from the same population. From the

discussion in Section 2.1, we know vertical error ΔZ_T consists of propagation error δ_T and interpolation error R_T . Under the simplest situation, that source data is free of systematic and gross error, propagation error is random error. Random error typically has a normal distribution with mean of 0, therefore there is $\delta_T \sim N(0, \sigma_T^2)$ where σ_T^2 is the variance of T 's propagation error. Interpolation error R_T is a constant, because its sign and magnitude are determined as explained previously. As the sum of these two errors, ΔZ_T also has a normal distribution, i.e. $\Delta Z_T \sim N(R_T, \sigma_T^2)$. However, the value of R_T and σ_T vary from location to location. The population of DEM vertical errors, denoted by $\{\Delta Z_T\}$ is thus made up by individuals from various normal distributions. From a statistical perspective, a population of such individuals is unlikely to have a normal distribution. This explains why so many studies have observed non-normal distributions of DEM vertical errors.

That DEM errors are unlikely to have a normal distribution explains why NSSDA's approach, which computes the 95% confidence interval of the vertical errors as $1.96 \cdot \text{RMSE}$, is problematic. The modification proposed by NDEP (2004) to use the 95th percentile is an improvement. The point of the 95th percentile is to inform that 95% of the errors are expected to be at or below this value. Apparently, the validity of the 95th percentile depends heavily on whether the samples truly represent the population. A small sample size of 30 is acceptable if individuals are from an identical population of normal distribution; otherwise, substantially more samples are necessary. Previous discussion has established that DEM errors are neither identically distributed nor from a normal distribution. Rather, it is expected to be smaller in flat terrain and larger in complex terrain. Hence unless all types of terrain are well sampled, the 95th percentile obtained from checkpoints will not describe the overall accuracy of a DEM. Existing standards only require a rather small sample size of 20 or more in each major land cover category. Whether these checkpoints can truly represent the entire terrain is thus a key challenge. In real-world applications, complex terrain is more difficult to access thus tends to be undersampled. This presents the risk that the 95th percentile obtained may overestimate DEM accuracy.

That vertical errors are not from an identical population also presents a challenge to geostatistical methodology. A central assumption in geostatistics is the stationarity of a process, meaning the mean and standard deviation of the process do not change in space. This does not always hold in the context of DEM error. As shown previously, each vertical error is from an individual normal distribution where the mean is the interpolation error and the variance is the variance of the propagation error, i.e. $\Delta Z_T \sim N(R_T, \sigma_T^2)$. Since R_T and σ_T vary from location to location, the stationarity assumption cannot be satisfied. This confirms the concern expressed by Oksanen and Sarjakoski's (2006) on using stationarity-based statistical methods to model DEM error.

In light of the reality that DEM error is not random error, not stationary, and not from an identical normal distribution, statistical methods including geostatistics are only appropriate if substantial numbers of checkpoints are used. Since this is rarely feasible in empirical analyses, non-statistical methods must be explored.

3 Overview of Approximation Theory

A viable non-statistical methodology is approximation theory which is routinely used in computational science to study how to approximate a complex function $z(x)$ using simpler functions $Z(x)$ and quantitatively characterize the errors introduced therein (Atkinson and Han 2004). For example, supposing function $z(x) = \sin(x)$ is to be approximated by linear

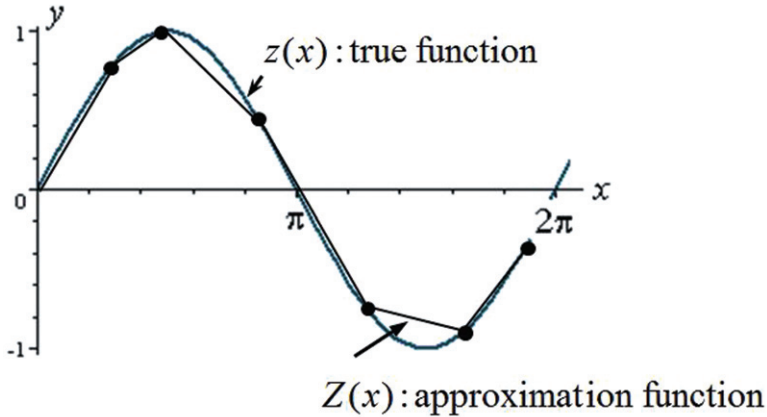


Figure 2 Approximation of a function using piecewise interpolation

polynomial $Z(x) = ax + b$ based on a set of reference points. The typical strategy is to use piecewise interpolation by dividing $z(x) = \sin(x)$ into segments, each of which is then approximated by a line (Figure 2). The accuracy of the approximation in a segment s is measured by the largest error at a point in this segment, i.e. $\max|z(x) - Z(x)|, x \in s$. The overall accuracy of the approximation is determined by the largest error of *any* point in the entire domain, i.e. $\max(|z(x) - Z(x)|)$. The rationale behind approximation theory is simple: If the largest error is acceptable, the error at *any* other point must be also acceptable, hence the accuracy of the overall approximation is guaranteed.

In the context of DEM generation, terrain is the complex function $z(x, y)$, DEM generation method is the approximation function $Z(x, y)$. $Z(x, y)$ is typically a piecewise function, i.e. it divides terrain into consecutive patches so that DEM generation can be conducted patch by patch. The vertical error at a location is $|Z(x, y) - z(x, y)|$. According to approximation theory, the overall accuracy of the DEM is controlled by the largest error at *any* location in *any* patch, i.e. $\max|Z(x, y) - z(x, y)|$. Note this maximum is not the statistical maximum but the maximum of the entire area covered by the DEM.

Based on approximation theory, Hu et al. (2009a) presented a framework on how to assess the vertical accuracy of DEMs created by different interpolation methods. Of relevance to this article is linear interpolation in 1D which is illustrated in Figure 1. Supposing a and b are two reference points whose elevation values are Z_a and Z_b respectively, the elevation at a location T is estimated by

$$Z_T = \omega_a Z_a + \omega_b Z_b \tag{1}$$

$$\omega_a + \omega_b = 1, \quad \omega_a, \omega_b \geq 0 \tag{2}$$

where ω_a (or ω_b) is the proportion of segment $x_b x_T$ (or $x_a x_T$) in $x_a x_b$. Linear interpolation in 1D is the optimal method of generating a DEM from contour intervals because it results in minimum interpolation error and does not amplify the errors in the source data (Hu et al. 2009a). Better yet, it preserves elevation orderliness proactively so that the resultant DEM has minimal artifacts and correct flow directions (Hu et al. 2009b).

Table 1 Vertical error in a DEM generated by linear interpolation in 1D

Vertical error at a point:	$\Delta Z_T = \delta_T + R_T$
δ_T : propagation error	$ \delta_T \leq \delta_{node} $
R_T : interpolation error	$ R_T \leq \frac{1}{8} M_2 h^2$
DEM vertical error:	$\max(R_T + \delta_T) \leq \max R_T + \delta_T $

As shown previously, the vertical error at a point T is the sum of propagation error (δ_T) and interpolation error (R_T). According to approximation theory, the overall accuracy of a DEM is determined by $\max|\delta_T + R_T|$ which is bounded by $\max|\delta_T| + \max|R_T|$. Using simple algebra and intermediate calculus, the results in Table 1 can be obtained (Hu et al. 2009a). $\max|\delta_T|$ is bounded by $\max|\delta_{node}|$ where δ_{node} is the error in the reference data used to interpolate T , e.g. a and b in Figure 1. $|R_T|$ is bounded by $\max\left|\frac{1}{8} M_2 h^2\right|$ where M_2 is the maximum norm of the second-order derivative in the flow path and h is the horizontal distance of the flow path. In the context of terrain, the second-order derivative describes how fast slope gradient changes, M_2 is thus equivalent to concavity or convexity. h describes the density of reference data. When sample elevations are abundant, the value of h is small. Table 1 articulated why DEM error is correlated with terrain complexity and sampling density, which has been observed by numerous studies (e.g. Oksanen and Sarjakoski 2006, Carlisle 2005). It also points out that it is concavity or convexity that is correlated with DEM error, not other terrain derivatives such as slope or aspect.

4 Approximation Theory Illustrated

4.1 Study Area and Data

To illustrate how Table 1 can be applied in empirical analyses, a case study is conducted to assess the vertical accuracy of a DEM generated from a topographic map. The spatial extent of the DEM is about 2.5 by 2.5 arc-minutes, covering the northern portion of the USGS 7.5-minute topographic quad of San Francisco South (Figure 3). The topographic map has a scale of 1:24,000 and contour interval of 25 feet. The elevation ranges from 0 to 925 feet. Various geomorphological types are found in the area including hills, valleys, lakes, and depressions, providing an excellent test of DEM generation and quality control. The social, economic, and ecological prominence of the area also means its DEM accuracy may have various implications.

Using the contour lines and spot elevations in the topographic map, a DEM was created using linear interpolation in 1D. During implementation, intermediate contour lines were generated recursively based on the concept of a weighted Voronoi diagram. The Voronoi diagram is a tool which partitions the space into regions such that all points in the region centered around a feature are closer to it than to any other features

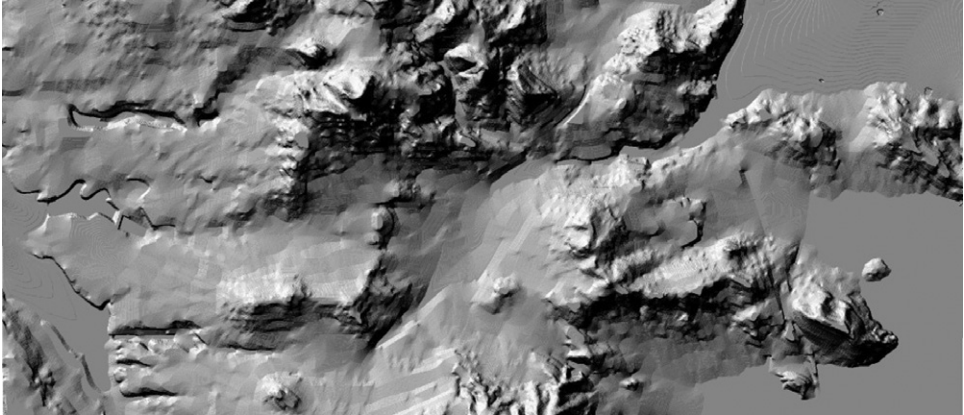


Figure 3 Terrain of the study area

(Aurenhammer 1991). The equidistant boundary between adjacent regions is called the Voronoi boundary. In the context of DEM generation, contour lines and spot elevations are the features. Their elevation values are the weights associated with them. Supposing there are two contour lines of 100 and 200 feet, their Voronoi boundary is then the 150-foot contour line. Similarly, the Voronoi boundary between a 100- and 150-foot contour line is a 125-foot contour line. The recursive process may continue until there is no further need of intermediate contour lines. Based on this idea, a raster-based algorithm similar to that in Cuisinaire and Macq (1999) was applied to generate the intermediate contour lines. The result is a DEM of 3.75 m spatial resolution. The choice of the spatial resolution is based on the 0.15 mm graphic resolution of the scanned contour lines. Given that points less than 3.75 m apart on the ground are not differentiable on the map, it is reasonable to assume that *every* point on the topographic map has been transferred to the 3.75 m DEM. The DEM is thus expected to have high accuracy and introduces minimal error when being used to estimate M_2 in Section 4.3.

In the next section, we illustrate how to assess the vertical accuracy of this DEM based on approximation theory, as presented in Table 1. Since the vertical accuracy is determined by the largest error at any location in the DEM, and the error at a location is the sum of propagation and interpolation error, the following section explains how to assess each error.

4.2 Propagation Error Assessment

The propagation error at a location T is bounded by the larger error in the source data used in its interpolation, e.g. the error in location a or b in Figure 1. The source data in this study is the 7.5-minute topographic map. The traditional assumption is that vertical errors in topographic maps are random error following a normal distribution, an assumption generally true for contours compiled by photogrammetry (Maune 2007). The mean of the normal distribution is expected to be 0. Its standard deviation can be inferred from the National Mapping Accuracy Standard (NMAS) (U.S. Bureau of the Budget 1947) which is used by USGS to control the quality of its topographic maps. According to NMAS, “no more than 10 percent of the elevations tested shall be in error

by more than one-half the contour interval.” Since a half interval is 12.5 ft in this study, it can be inferred that the standard deviation of the error in source data is 7.6 ft. Given that 99.95% samples in a normal distribution are smaller than 3.29 standard deviations, the bound of the error in the source data can be set as 25 ft, i.e. no error in the topographic map may exceed one contour interval. Note this error bound is a very conservative estimate; it is very likely to overestimate the largest error in the source data. A more accurate estimate is to use USGS’s method by selecting 20 or more well-defined checkpoints and then calculates the error by field visit using high precision devices such as GPS.

4.3 Interpolation Error

The interpolation error at a point T (R_T) is bounded by $\left| \frac{1}{8} M_2 b^2 \right|$ where M_2 is the maximum norm of the second-order derivative at a point and b is the horizontal length of the line segment used to interpolate T . This article introduces two methods to estimate $|R_T|$. The first one uses contour lines only, therefore is suitable for a pre-assessment before a DEM is generated. In Figure 4, assume a water drop flows downhill crossing a number of contour lines of Z_2, Z_1 , and Z_0 . Let T be a point on the flow path between Z_0 and Z_1 , and the contour interval is denoted by cl . M_2 can be approximated by the slope change rate at Z_1 , i.e.

$$M_2 \cong \frac{\frac{Z_2 - Z_1}{d_{12}} - \frac{Z_1 - Z_0}{d_{01}}}{d_{02}} = cl \frac{d_{01} - d_{12}}{d_{01}d_{12}d_{02}} \tag{3}$$

Considering that $d_{02} = d_{01} + d_{12}$ in Equation 3, the interpolation error at T (R_T) is bounded by:

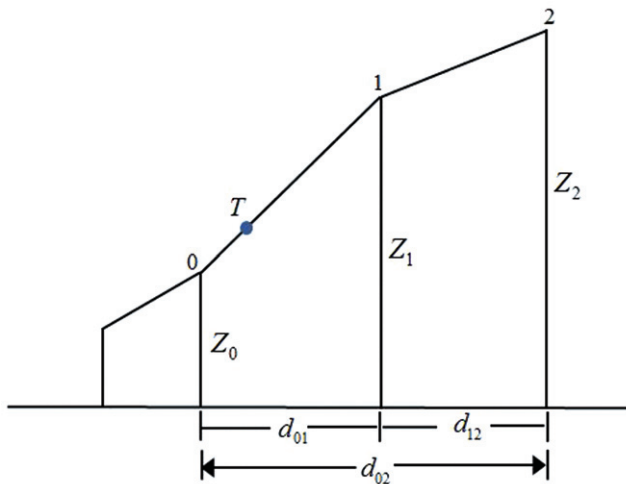


Figure 4 Estimation of M_2 using contour lines

Table 2 Estimation of interpolation error based on the horizontal distance ratio between adjacent contour lines

If d_{01}/d_{12} is less than	Interpolation error is bounded by
1.5	0.04 * cl
2	0.084*cl
3	0.18 * cl
4	0.30 * cl
5	0.417* cl
10	1.023* cl

$$R_T \leq \frac{1}{8} M_2 h^2 = \frac{1}{8} cl \cdot \frac{d_{01} - d_{12}}{d_{01} d_{12} d_{02}} \cdot d_{01}^2 = \frac{cl}{8} \frac{d_{01}}{d_{02}} \left(\frac{d_{01}}{d_{12}} - 1 \right) \tag{4}$$

The value of d_{01} and d_{12} are the horizontal distances between adjacent contour lines. Their ratio d_{01}/d_{12} describes how fast slope changes in the area, and indicates the density of contour lines. With simple mathematics, Equation 4 can be rearranged to obtain the results in Table 2. It can be seen that as d_{01}/d_{12} increases, R_T also increases but remains small. For example, as long as d_{01}/d_{12} is less than 4, R_T is bounded by one-third of the contour interval. Even when d_{01}/d_{12} reaches 10 which is extremely high, R_T is still only slightly more than one contour interval. In our study, we found that d_{01}/d_{12} is less than 1.5 in 72% of the cells calculated; less than 3 in 92% of the cells calculated; and less than 5 in 99% of the cells calculated. Extreme values such as 10, are very rare but do exist, as illustrated in Figure 5. These areas are where the largest interpolation errors are likely to occur, therefore they should be paid more attention during accuracy assessment. If one wishes to reduce the potentially high errors in these areas, an effective strategy is to collect additional reference elevations in between the contour lines and incorporate them in DEM generation.

The above method is valuable to obtain an assessment before DEM generation. If the DEM is already generated, R_T can be assessed by another method. In Figure 6, assume E is the cell under consideration. Its M_2 can be obtained by finding the maximum slope change rate (i.e. slope of slope which is equivalent to curvature) in its 3 by 3 neighborhood, i.e. $M_2 = \max\{|curv_{N-S}|, |curv_{W-E}|, |curv_{NW-SE}|, |curv_{NE-SW}|\}$.

The curvature in the north-south direction can be obtained by:

$$|curv_{N-S}| = \frac{|slope_1 - slope_2|}{2r} = \frac{\left| \frac{Z_E - Z_B}{r} - \frac{Z_H - Z_E}{r} \right|}{2r} \tag{5}$$

Similarly, the slope change rate in the northeast-southwest direction is:

$$|curv_{NE-SW}| = \frac{|slope_1 - slope_2|}{2\sqrt{2}r} = \frac{\left| \frac{Z_E - Z_C}{\sqrt{2}r} - \frac{Z_G - Z_E}{\sqrt{2}r} \right|}{2\sqrt{2}r} \tag{6}$$

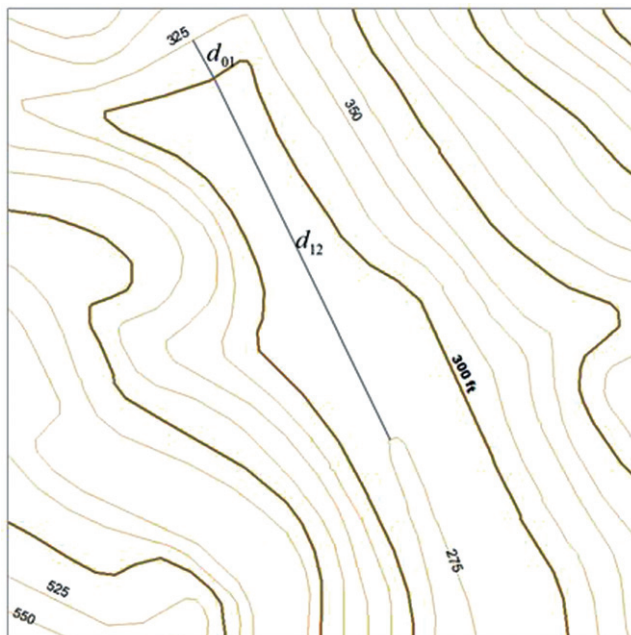


Figure 5 An example area where horizontal distance ratio between adjacent contour lines (d_{01}/d_{12}) exceeds 10

A	B	C
D	E	F
G	H	I

Figure 6 Estimation of M_2 using a DEM

The two other directions can be obtained in a similar manner. We applied this method to the 3.75 m DEM obtained by linear interpolation, and found M_2 values ranged between 0 and 2.5. Since h in a 3 x3 neighborhood is the horizontal distance in the direction with the maximum slope change rate, the interpolation error bound $\left| \frac{1}{8} M_2 h^2 \right|$ can be obtained for each pixel (Figure 7). In our study, it is found that 90% of the cells have an interpolation error bounded by 8.3 ft which is equivalent to one third of the contour interval; 6.25% cells have an interpolation error bounded between 8.3 ft and 12.5 ft, which is equivalent to one-third and one-half contour interval, respectively. Values larger than 12.5 ft, which are very few, are located in the boundary areas. These

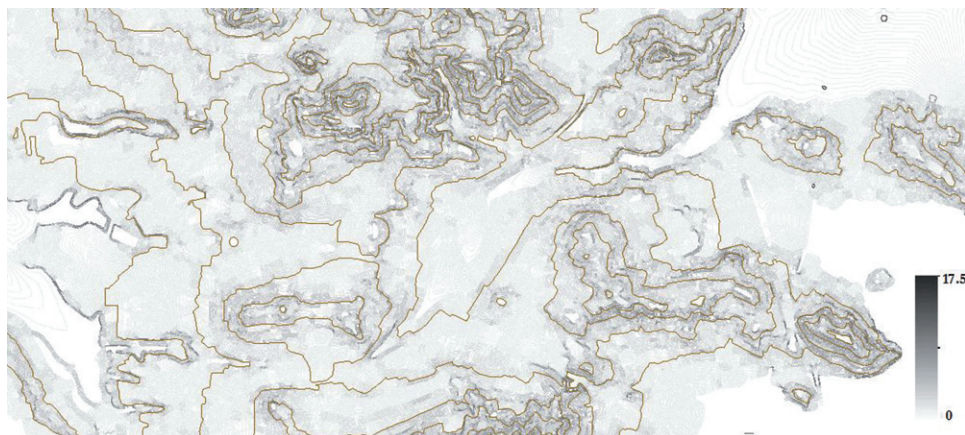


Figure 7 Interpolation error bound in the study area with index contour lines overlaid

values can be expected to drop if contour lines immediately outside the study area were included during DEM generation. In terms of spatial distribution, hilly area with greater slope change rate tends to have larger interpolation error. This confirms the results from the first method on using contour lines only.

4.4 Implications for DEM Quality Control

Vertical accuracy is an important aspect of DEM quality. That approximation theory is able to estimate propagation and interpolation error separately offers effective guidance to DEM production and quality control assuming a DEM user requires that the vertical error at any point should not exceed a threshold value. Since the total error bound at each point, which is the sum of the propagation and interpolation error, is available, the DEM producer can easily check whether the requirement is met. In areas with excessive errors, the producer can identify the cause by comparing propagation error with interpolation error. In the case that propagation error is dominant, the producer needs to improve the quality of the source data. On the other hand, if interpolation error is found to be the main cause, the producer can identify the location of large interpolation errors and reduce it by inserting additional samples in those areas before interpolation is conducted. Compared with RMSE, which is the existing method to control DEM vertical error, approximation theory is much more informative and effective. RMSE only offers a single summary statistic for the entire DEM based on a limited number of checkpoints. In the case that DEM quality is not satisfactory, the producer has no idea about the cause and how to improve it. In contrast, approximation theory not only provides an estimate of the error bound for each point in the entire study site, but also reveals where large errors are likely to occur; hence, pointing out how to effectively improve a DEM's vertical accuracy if one wishes.

5 Conclusions

As a pivotal indicator of DEM quality, DEM vertical accuracy has been studied by numerous researchers. Among the various methods developed, most are statistics

oriented despite concerns as to its effectiveness and validity. This article pointed out for the first time that it would be very challenging to continue the statistics paradigm because DEM errors are not random error, not normally distributed, not identically distributed, and not stationary. These characteristics determined that a very large number of check-points are necessary in order to assure the statistical validity. Unfortunately, this requirement is rarely met in real-world applications.

In this article, we presented approximation theory as a new methodology and illustrated how it can be applied to the accuracy assessment of a DEM generated by linear interpolation using contour lines as the source data. The results show that linear interpolation in 1D is an excellent interpolation method. Nearly all of the cells in our study area have an interpolation error less than 12.5 ft which is half the contour interval. Our research also points out that the terrain characteristic that directly determines interpolation error is the ratio between the horizontal distances of adjacent contour lines. Areas such as foothills are thus where large interpolation errors are likely to occur. It is impossible to change terrain characteristics, but interpolation error can always be effectively reduced by increasing sample density.

While the case study in this article is based on a DEM created by linear interpolation using topographic maps as the source data, the approximation theory methodology is applicable to other DEMs created by interpolation, e.g. LiDAR-derived DEMs. Hu et al. (2009a) has outlined the propagation and interpolation error bound for this type of DEMs generated by TIN interpolation, and the preliminary results from a case study are reported in Liu and Sherba (2012). Future research will expand the methodology to other interpolation methods such as bilinear interpolation in a rectangle which is often used in DEM resampling.

References

- Aguilar F, Aguilar A, Aguera F, and Sanchez J 2006 The accuracy of grid digital elevation models linearly constructed from scattered sample data. *International Journal of Geographical Information Science* 20: 169–92
- Aguilar F, Mills J, Delgado J, Aguilar M, Negreiros J G, and Pérez J 2010 Modelling vertical error in LiDAR-derived digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 103–10
- Atkinson K and Han W 2004 *Elementary Numerical Analysis*. Chichester, John Wiley and Sons
- Aurenhammer F 1991 Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Computing Surveys* 23: 345–405
- Bonin O and Rousseaux F 2005 Digital terrain model computation from contour lines: How to derive quality information from artifact analysis. *GeoInformatica* 9: 253–68
- Carlisle B 2005 Modelling the spatial distribution of DEM error. *Transactions in GIS* 9: 521–40
- Cuisenaire O and Macq B 1999 Fast Euclidean distance transformation by propagation using multiple neighborhoods. *Computer Vision and Image Understanding* 76: 163–72
- FEMA 2010 LIDAR Specifications for Flood Hazard Mapping. WWW document, http://www.fema.gov/plan/prevent/fhm/lidar_4b.shtm
- FGDC 1998 *Geospatial Positioning Accuracy Standards: Part 3, National Standards for Spatial Data Accuracy*. Washington, D.C., Federal Geographic Data Committee
- Fisher P F 1998 Improved modeling of elevation error with geostatistics. *GeoInformatica* 2: 215–33
- Gesch D 2007 The National elevation dataset. In Maune D (ed) *Digital Elevation Model Technologies and Applications: The DEM User's Manual* (Second Edition). Falls Church, VA, American Society of Photogrammetry and Remote Sensing: 449–73

- Greenwalt C and Schultz M 1962 *Principles of Error Theory and Cartographic Applications*. St. Louis, MO, U.S. Air Force, Aeronautical Chart and Information Center Technical Report No. 96
- Hu P, Liu X, and Hu H 2009a Accuracy assessment of digital elevation models based on approximation theory. *Photogrammetric Engineering and Remote Sensing* 75: 49–56
- Hu P, Liu X, and Hu H 2009b Isomorphism in digital elevation models and its implication to interpolation functions. *Photogrammetric Engineering and Remote Sensing* 75: 713–21
- Huang Y 2000 Evaluation of information loss in digital elevation models with digital photogrammetric systems. *Photogrammetric Record* 16(95): 781–91
- Huevelink G 1998 *Error Propagation in Environmental Modeling and GIS*. London, Taylor and Francis
- Kyriakidis P C and Goodchild M F 2006 On the prediction error variance of three common spatial interpolation schemes. *International Journal of Geographical Information Science* 20: 823–56
- Li Z 1993 Theoretical models of the accuracy of digital terrain models: An evaluation and some observations. *Photogrammetric Record* 14(82): 113–28
- Liu X and Sherba J 2012 Accuracy assessment of LiDAR-derived DEM based on approximation theory. In *Proceedings of the Annual Conference of American Society of Photogrammetry and Remote Sensing*, Sacramento, California
- Maune D 2007 DEM user requirements. In Maune D (ed) *Digital Elevation Model Technologies and Applications: The DEM User's Manual* (Second Edition). Falls Church, VA, American Society of Photogrammetry and Remote Sensing: 449–73
- NDEP 2004 Guidelines for Digital Elevation Data Version 1. WWW document, http://www.ndep.gov/NDEP_Elevation_Guidelines_Ver1_10May2004.pdf
- Oksanen J and Sarjakoski T 2006 Uncovering the statistical and spatial characteristics of fine toposcale DEM error. *International Journal of Geographical Information Science* 20: 345–69
- Shi W Z, Li Q Q, and Zhu C Q 2005 Estimating the propagation error of DEM from higher-order interpolation algorithms. *International Journal of Remote Sensing* 26: 3069–84
- Tempfli K 1980 Spectral analysis of terrain relief for the accuracy estimation of digital terrain models. *ITC Journal* 3: 478–510
- U.S. Bureau of the Budget 1947 *United States National Map Accuracy Standards*. Washington D.C., U.S. Bureau of the Budget.
- Wise S 2000 Assessing the quality for hydrological applications of digital elevation models derived from contours. *Hydrological Processes* 14: 1909–29

Copyright of Transactions in GIS is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.