

An Experiment Using a Circular Neighborhood to Calculate Slope Gradient from a DEM

Xun Shi, A-Xing Zhu, James Burt, Wes Choi, Rongxun Wang, Tao Pei, Baolin Li, and Chengzhi Qin

Abstract

The traditional 3×3 cell neighborhood used in a focal operation on a raster layer has a square shape that results in a dimensional neighborhood of which the orientation is eventually arbitrary to the physical features represented. This paper presents an experiment using a circular neighborhood to calculate slope gradient. Comparisons of the results from a circular neighborhood with the results from some traditional methods show that (a) for a smooth surface, the result from a circular neighborhood is more accurate than that from a square neighborhood, (b) a circular neighborhood is generally more sensitive to noise in the input DEM than a square neighborhood, and (c) in a validation using field measurements, the circular neighborhood performs better than the square neighborhood when the ratio of user-specified neighborhood size to cell size is high.

Introduction

The traditional 3×3 cell neighborhood used in a focal operation on a raster layer has two characteristics: its size is determined by the resolution of the input layer, and its shape is usually square. The resolution-determined size causes inconsistency in the terrain attribute values computed from gridded DEMs with different resolutions. This problem has been studied by many researchers (e.g., Chang and Tsai, 1991; Hodgson, 1995; Gao, 1997; Kienzle, 2004; Zhou and Liu, 2004). Conclusions drawn by these researchers do not totally agree with each other (mainly due to the different benchmarks or "true values" they used), but there is a consensus that the DEM resolution (eventually the size of the neighborhood determined by this resolution) has a strong effect on the accuracy of the terrain attributes derived from the DEM, and in turn on the outcome from the erosion or hydrological models based on these attributes. The resolution-determined neighborhood size also leads to the problem of mismatch between human-perceived and computer-calculated values of terrain attributes, people doing fieldwork (e.g., soil surveyors) always have their own measuring scales. This mismatch has become an important source of error in knowledge-based digital soil mapping. Ironically, this is becoming a more

serious problem since high-resolution DEMs are becoming more readily available. Burt and Zhu (2002) solved this mismatch problem by employing a user-defined neighborhood to replace the resolution-determined neighborhood. In this method, the user specifies the size of the neighborhood, and the computer selects the eight cells (not necessarily contiguous) that make up the square with the specified size to calculate the terrain attributes for the cell at the center of the neighborhood. This method allows the user to specify a neighborhood size that matches the scale in their mind, regardless of the resolution of the input DEM.

On the other hand, the square shape results in a dimensional neighborhood: the values in the diagonal directions are farther from the center of the neighborhood than the values in the cardinal directions. This dimensional property has a significant effect on derived terrain attributes (Zhou and Liu, 2004) while from a physical standpoint, the orientation of a DEM grid is arbitrary. To reduce this effect, some algorithms (e.g., Horn, 1981) assign different weights to the cells in different directions. The square shape, after all, is merely a convenient setting for the nine contiguous cells that define the neighborhood. Ideally, the neighborhood should be invariant under rotation, and an obvious choice would be to use a circular neighborhood.

To address the shape problem, this paper presents an experiment using a circular neighborhood in the calculation of slope gradient. The circular neighborhood is implemented for two gradient-calculation methods, the Evans method and a modified Zevenbergen-Thorne method. For these two methods, the results from the circular neighborhood are compared with the results from the traditional square neighborhood. Two other widely used methods, the Horn method and the original Zevenbergen-Thorne method, have also been included in the comparison as benchmarks. The Evans method (Evans, 1980) was chosen, because it is probably the most studied and has been argued to be the most optimal method under certain conditions (Jones, 1998a and 1998b; Florinsky, 1998). Also, this method is easily adapted to employ a circular neighborhood. The Zevenbergen-Thorne method (Zevenbergen and Thorne, 1987) was chosen as a representative of the methods based on the Lagrange polynomial. It was chosen also because Jones (1998a) concluded that the result from this method was more "accurate" than results from other methods, although we found this conclusion to be conditional (see the Results and Discussion Section). The original Zevenbergen-Thorne method only uses information from the four points in the cardinal directions, thus the shape of the neighborhood would not affect the

Xun Shi is with The Department of Geography, Dartmouth College, 6017 Fairchild, Hanover, NH 03755 (xun.shi@dartmouth.edu).

A-Xing Zhu, Tao Pei, Baolin Li, and Chengzhi Qin are with The State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Building 917, Datun Road, An Wai, Beijing 100101, China.

A-Xing Zhu, James Burt, Wes Choi, and Rongxun Wang are with The Department of Geography, University of Wisconsin-Madison, 550 North Park Street, Madison, WI 53706

Photogrammetric Engineering & Remote Sensing
Vol. 73, No. 2, February 2007, pp. 143–154.

0099-1112/07/7302-0143/\$3.00/0
© 2007 American Society for Photogrammetry
and Remote Sensing

result. To test the effect of the circular neighborhood, we implement a method that is based on the Zevenbergen-Thorne method but can take into account the information from both cardinal and diagonal directions. The Horn method (Horn, 1981) was chosen as a reference, because it is adopted by the SLOPE function of Arc/Info® and thereby might be the most popular method in practice.

The results from these chosen methods were compared in three ways: (a) the accuracies of different methods were compared by differencing their results using analytical values calculated from a synthetic surface; (b) sensitivities of different methods to error in the input DEM were compared using *noised* surfaces; and (c) correspondences of the results from different methods to human-perceived values were compared using a set of gradient values measured by soil scientists in a small watershed.

The implementation of the circular neighborhood in the two gradient-calculation methods are described in the next section, followed by two sections detailing how the results from the two kinds of neighborhoods were compared, one section explaining the methods, one describing the data, and then the comparisons are presented and discussed. The last section contains the conclusions drawn from this research.

Implementation of Circular Neighborhood

Definition of a Circular Neighborhood

A circular neighborhood is defined by eight points that are on a circle centered at a point for which a terrain attribute (e.g., slope gradient) is to be computed. The elevation at each of these eight points can be derived through a bilinear interpolation from the cells of the DEM. The definitions of the square neighborhood and the circular neighborhood are illustrated in Figure 1 for comparison. In this figure, w equals one half of the width of the neighborhood defined by the user; and z_1, z_2, \dots, z_8 are elevation values at those neighborhood-defining points.

Implementation of a Circular Neighborhood in Gradient-Calculation Methods

A number of methods have been proposed for calculating slope gradient. Although no consensus has been reached, it seems to be well accepted that the methods based on quadratic or Lagrange polynomials are better than other simpler methods (Jones, 1998a and 1998b; Florinsky, 1998).

Among the methods tested in this research, the Evans method and the Horn method are based on a quadratic polynomial and the Zevenbergen-Thorne method, and the modified Zevenbergen-Thorne method are based on a Lagrange polynomial. The quadratic polynomial can be given as follows (Shary, 1995):

$$Z = \frac{1}{2} rx^2 + \frac{1}{2} ty^2 + sxy + px + qy + u, \quad (1)$$

and the Lagrange polynomial can be given as follows (Florinsky, 1998):

$$Z = ax^2y^2 + bx^2y + cxy^2 + \frac{1}{2} rx^2 + \frac{1}{2} ty^2 + sxy + px + qy + u. \quad (2)$$

The values of the coefficients in these polynomials for the square and circular neighborhoods are listed in Tables 1 and 2. Details of the derivation can be found in the Appendix of this paper.

The coefficient values of the quadratic polynomial, as shown in Table 1, are quite different for the two kinds of neighborhoods. Since the gradient is calculated based on p and q , it can be expected that the Evans method will produce different gradient values under different neighborhood shapes. On the other hand, as shown by Table 2, among the nine coefficients of the Lagrange polynomial, $a, b, c,$ and s are different for the two kinds of neighborhoods, but $r, t, p, q,$ and u stay the same. Since the gradient is based on p and q , the shape of the neighborhood does not affect a calculation using the Lagrange polynomial. For testing the performance of the circular neighborhood under the Lagrange polynomial, a modified Zevenbergen-Thorne method is implemented. This new method first calculates the slope gradient of a given cell using the conventional Zevenbergen-Thorne method, and then rotates the kernel for 45 degrees and uses the four values in the diagonal directions to calculate the slope gradient again. The average of the two values is then assigned to the given cell as its overall slope gradient. This method is an integration of the Zevenbergen-Thorne method and the "Diagonal Ritter" method described by Jones (1998a) and allows incorporation of the information from the diagonal directions. In this paper, this new method is referred to as the Modified Z&T method.

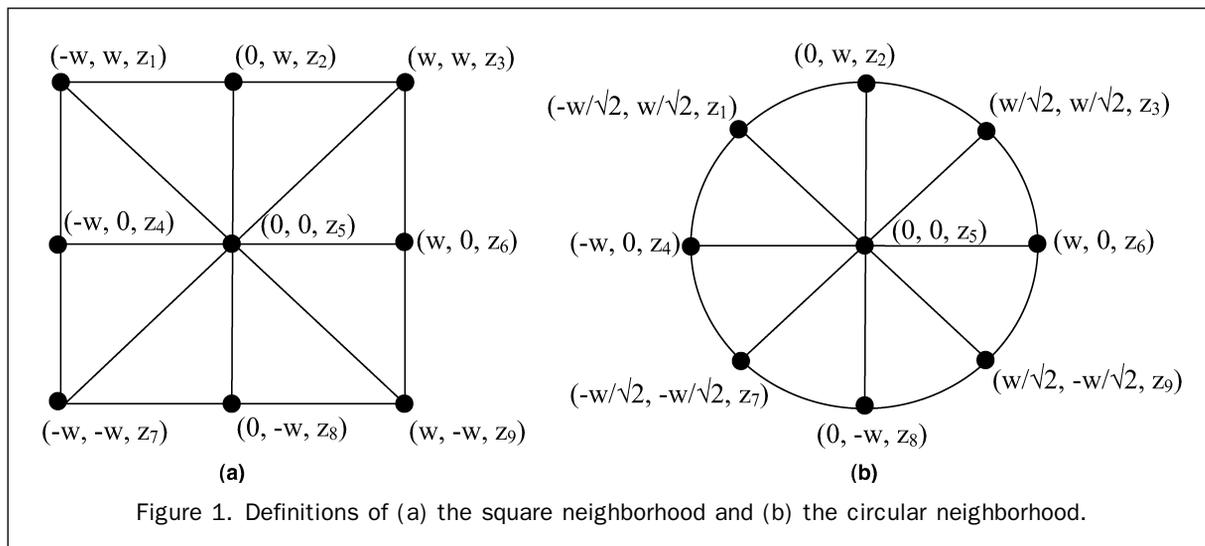


Figure 1. Definitions of (a) the square neighborhood and (b) the circular neighborhood.

TABLE 1. THE COEFFICIENT VALUES IN THE QUADRATIC POLYNOMIAL FOR THE TWO KINDS OF NEIGHBORHOODS

Coefficient	Square Neighborhood	Circular Neighborhood
<i>r</i>	$[(z_1 + z_3 + z_4 + z_6 + z_7 + z_9) - 2(z_2 + z_5 + z_8)]/3w^2$	$(z_1 + z_3 + 3z_4 + 3z_6 + z_7 + z_9 - z_2 - 8z_5 - z_8)/4w^2$
<i>t</i>	$[(z_1 + z_2 + z_3 + z_7 + z_8 + z_9) - 2(z_4 + z_5 + z_6)]/3w^2$	$(z_1 + z_3 + 3z_2 + 3z_8 + z_7 + z_9 - z_4 - 8z_5 - z_6)/4w^2$
<i>s</i>	$(z_3 + z_7 - z_1 - z_9)/4w^2$	$(z_3 + z_7 - z_1 - z_9)/2w^2$
<i>p</i>	$(z_3 + z_6 + z_9 - z_1 - z_4 - z_7)/6w$	$(\sqrt{2}z_3 + 2z_6 + \sqrt{2}z_9 - \sqrt{2}z_1 - 2z_4 - \sqrt{2}z_7)/8w$
<i>q</i>	$(z_1 + z_2 + z_3 - z_7 - z_8 - z_9)/6w$	$(\sqrt{2}z_1 + 2z_2 + \sqrt{2}z_3 - \sqrt{2}z_7 - 2z_8 - \sqrt{2}z_9)/8w$
<i>u</i>	$[5z_1 + 2(z_2 + z_4 + z_6 + z_8) - z_1 - z_3 - z_7 - z_9]/9$	z_5

TABLE 2. THE COEFFICIENT VALUES IN THE LAGRANGE POLYNOMIAL FOR THE TWO KINDS OF NEIGHBORHOODS

Coefficient	Square Neighborhood	Circular Neighborhood
<i>a</i>	$[(z_1 + z_3 + z_7 + z_9)/4 - (z_2 + z_4 + z_6 + z_8)/2 + z_5]/w^4$	$[(z_1 + z_3 + z_7 + z_9) - (z_2 + z_4 + z_6 + z_8)]/w^4$
<i>b</i>	$[(z_1 + z_3 - z_7 - z_9)/4 - (z_2 - z_8)/2]/w^3$	$[(z_1 + z_3 - z_7 - z_9) - \sqrt{2}(z_2 - z_8)]/w^3$
<i>c</i>	$[(-z_1 + z_3 - z_7 + z_9)/4 + (z_4 - z_6)/2]/w^3$	$[(-z_1 + z_3 - z_7 + z_9)/\sqrt{2} + (z_2 - z_8)]/w^3$
<i>r</i>	$[(z_4 + z_6)/2 - z_5]/2w^2$	$[(z_4 + z_6)/2 - z_5]/2w^2$
<i>t</i>	$[(z_2 + z_8)/2 - z_5]/2w^2$	$[(z_2 + z_8)/2 - z_5]/2w^2$
<i>s</i>	$(-z_1 + z_3 + z_7 - z_9)/4w^2$	$(-z_1 + z_3 + z_7 - z_9)/2w^2$
<i>p</i>	$(-z_4 + z_6)/2w$	$(-z_4 + z_6)/2w$
<i>q</i>	$(z_2 - z_8)/2w$	$(z_2 - z_8)/2w$
<i>u</i>	z_5	z_5

Methods for Evaluation of the Circular Neighborhood

Herein in this paper, the six methods tested in this research are referred to as EVANS-SQR, EVANS-CIR, M-Z&T-SQR, M-Z&T-CIR, HORN, and Z&T.

Accuracy for a Smooth Surface

Following Hodgson (1995) and Jones (1998a), Morrison's surface (Morrison, 1971 and 1974) was used to evaluate the accuracy of the results from different methods for a smooth surface. Morrison's surface is a synthetic smooth surface generated using a polynomial of 49 trigonometric terms. The gradient at any location on this surface can be calculated analytically. In this research, the gradient values so calculated were used as true values to validate the results from other methods. Rasterization of Morrison's surface was performed in order to calculate slope gradient using the six finite-difference methods tested in this research. The six methods were applied to the rasterized Morrison's surface at different resolutions, and the results were compared with the analytical values at the centers of cells. The statistics used for the comparison include the mean average error (MAE), root mean square error (RMSE), and R². These statistics were calculated between the gradient layer containing the analytical values and each of the layers generated using the finite-difference methods.

Sensitivity to Error

Hunter and Goodchild (1996) and Jones (1998b) used *noised* surfaces to evaluate the sensitivity of gradient-calculation methods to error in the input DEM. A *noised* surface is created by adding random noise to the original DEM surface. The same gradient-calculation method is then applied to both the *noised* surface and the original surface. If the difference between the results from the two surfaces is small, the method is considered to be insensitive to the noise. In this research, we used Morrison's surface and two USGS DEMs as the original surfaces.

Hunter and Goodchild (1996) especially discussed the influence of the spatial autocorrelation in the error on the slope gradient calculated from a DEM. In this research, spatially autocorrelated noise values were created through an Inverse Distance Weighted (IDW) interpolation: First, a given number of random points are generated, whose locations and values are randomly assigned. The values of these points vary

between (*DEM*_{min} - *DEM*_{max}) and (*DEM*_{max} - *DEM*_{min}), where *DEM*_{min} and *DEM*_{max} are the minimum and maximum values in the DEM. An IDW interpolation is then performed based on these random points to create a smooth *noise* surface. Then, a number of locations in this smooth *noise* surface are randomly selected, and their values are spatially shuffled to create a new *noise* surface with weaker spatial autocorrelation. As the percentage of shuffled locations increases, the spatial autocorrelation in the surface decreases. Figure 2 gives an example of *noise* surfaces with different degrees of spatial autocorrelation created in this way.

A *noised* surface is created by adding a *noise* surface to the original surface, which is represented by the equation below:

$$\text{Noised Surface} = \text{Original Surface} + c * \text{Noise Surface} \quad (3)$$

where *c* is a factor controlling the magnitude of noise in the resulting *noised* surface. Following Jones (1998b), RMSE was calculated between the gradient layer of the original surface and the gradient layer of its corresponding *noised* surface to evaluate the difference between the two.

Comparison Based on Ground Truth

A set of slope gradient values measured by soil scientists in the field were used as true values to validate the computer-calculated values. The MAE, RMSE, R², and a new index, agreement coefficient (AC), were used to compare the field values and the calculated values. AC is an index that measures how well the predicted values agree with the observed values (Zhu *et al.*, 1997). AC is defined as (Willmott, 1984)

$$AC = 1 - N * RMSE^2 / PE \quad (4)$$

where *N*, in this research, is the number of the field sites for which both the observed and calculated gradient values are available, and *PE* is the potential error variance. *PE* is defined as:

$$PE = \sum_{i=1}^N \left(|P_i - \bar{O}| + |O_i - \bar{O}| \right)^2 \quad (5)$$

where \bar{O} is the mean of the true values, and O_i and P_i are the observed and calculated gradient values for the *i*th site, respectively. AC varies between 0 and 1, where 1 indicates perfect agreement and 0 indicates complete disagreement.

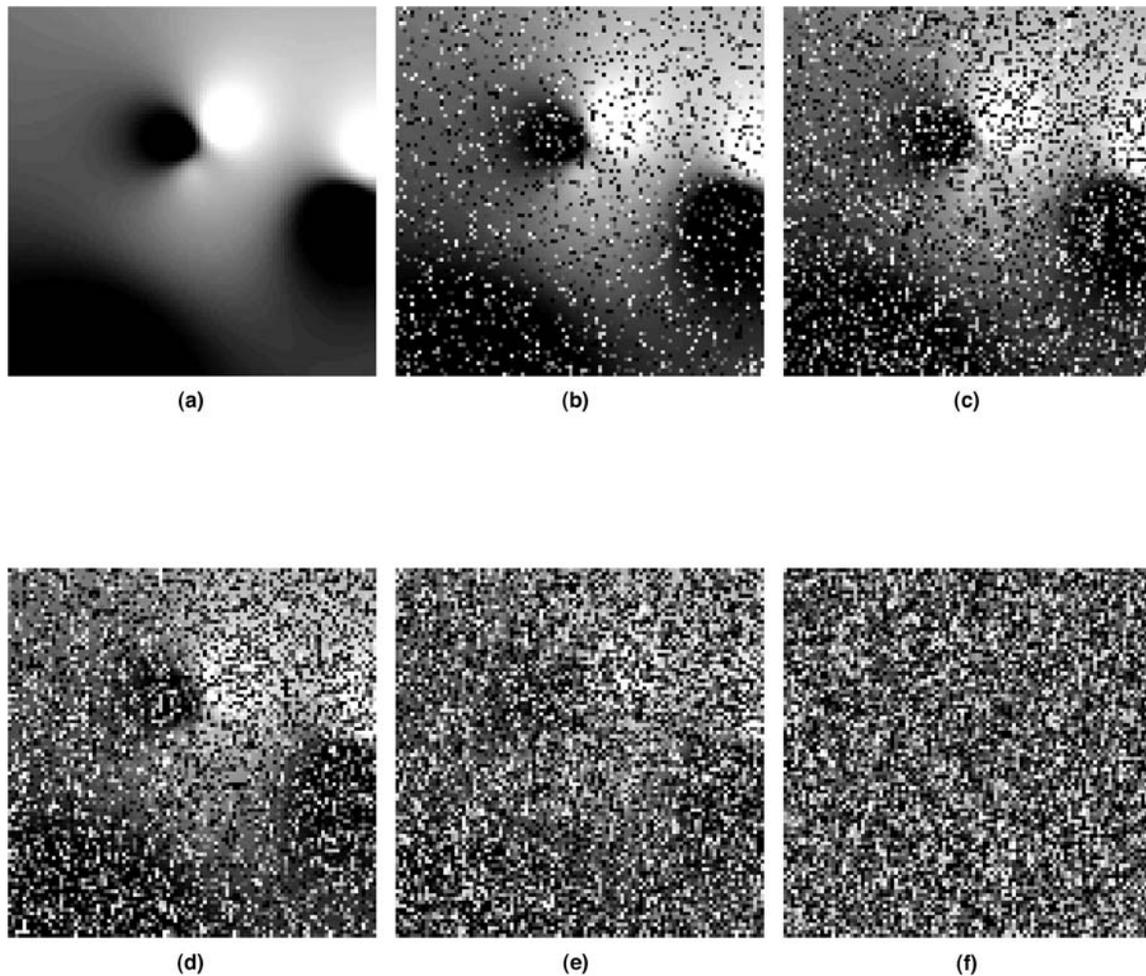


Figure 2. *Noise surfaces with different degrees of spatial autocorrelation: (a) shuffle rate = 0, (b) shuffle rate = 20 percent, (c) shuffle rate = 40 percent, (d) shuffle rate = 60 percent, (e) shuffle rate = 80 percent, and (f) shuffle rate = 100 percent.*

Data

Two series of surfaces were used in this research. The first series consists of rasterized Morrison's surfaces with different cell sizes, including 20, 10, 2, 1, and 0.5. The second series includes two USGS DEMs covering a small watershed called "Raffelson" in Wisconsin (Bangor Quadrangle). The two USGS DEMs cover the same area but one is at 30 m resolution and the other is at 10 m resolution. The Raffelson study area contains two broad ridges and a relatively open and flat floodplain. The terrain between the ridges and the floodplain is rugged, containing narrow alluvial valleys and steep side slopes. Elevation of this area varies between 250 m and 420 m; the mean elevation is about 320 m. Slope gradient of this area varies between 0 and about 70 percent. The mean slope gradient is about 25 percent. The contour lines in Figure 4 illustrate the two series of surfaces.

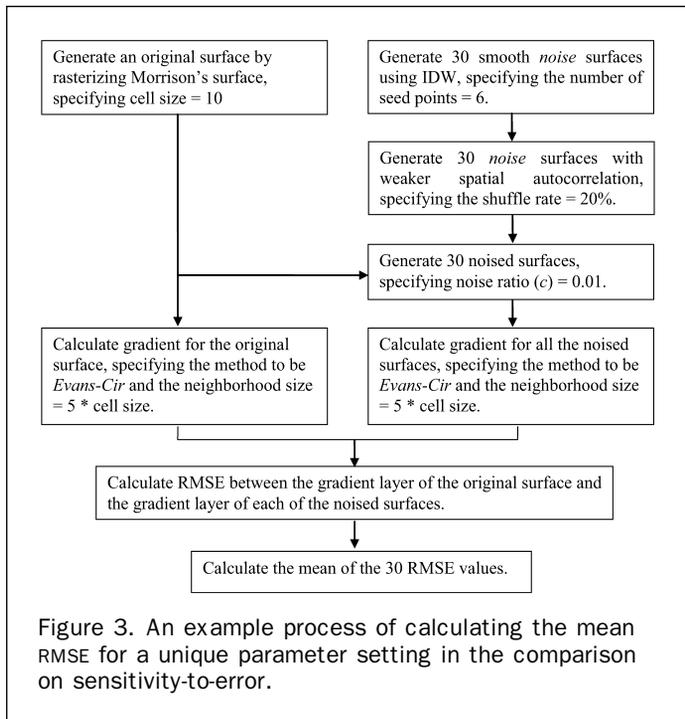
Rasterized Morrison's surfaces were used to evaluate the accuracy for a smooth surface. For each raster layer at a particular resolution, slope gradient was calculated using different neighborhood sizes, varying from one to ten times of that layer's cell size, with one cell size as the increment (i.e., a total of ten different neighborhood sizes were used for each layer).

Both series of surfaces were used in the comparison on sensitivity to noise. Neighborhood sizes tested included one to ten times of the cell size, with one cell size as the

increment. *Noised* surfaces were created using *noise* surfaces with different degrees of spatial autocorrelation and different noise factors (i.e., c in Equation 4). The percentages of shuffled locations included 0 percent, 20 percent, 40 percent, 60 percent, 80 percent, and 100 percent. The noise factors tested were 0.01, 0.02, 0.04, 0.08, and 0.16.

Under each unique combination of neighborhood size, shuffle rate, and noise factor, 30 different *noise* surfaces were created for an original surface, which in turn were used to create 30 *noised* surfaces. With the resulting 30 *noised* surfaces, 30 RMSE values were calculated between the *noised* surfaces and the original surface. The mean of these 30 RMSE values were used to compare results from different gradient-calculation methods and neighborhood sizes. Figure 3 describes the process of calculating the mean of RMSEs for an example setting. The process illustrated in Figure 3 was repeated for all parameter settings tested in this research.

In a soil survey project in the Raffelson area, soil scientists measured slope gradients at 90 sample locations in the field. These 90 field measurements were used to validate the results calculated by the six gradient-calculation methods from the two USGS DEMs (10 m and 30 m) using ten different neighborhood sizes (10 m to 100 m, with a 10 m increment). The locations of the 90 samples are shown in Figure 4b.



Results and Discussion

Difference of the Calculated Slope Gradient Values

The general spatial patterns of the calculated slope gradient values from all the six methods are very similar, but differences at a particular location can be significant. As an example, Figure 4 shows the images derived by subtracting the slope gradient layers calculated using EVANS-SQR from the layers calculated using EVANS-CIR. Figure 4a is based on a rasterized Morrison's surface (cell size = 10), and Figure 4b is based on the USGS 10 m DEM. The neighborhood sizes for deriving the two images are both five times of their corresponding cell sizes. Contour lines of elevation are superimposed on the images for checking the relationship between the difference values and landform.

The spatial pattern in the difference layer based on Morrison's surface is clear. Values from the circular neighborhood tend to be smaller than the values from the square neighborhood in relatively flat areas, including ridge tops, saddles, and valleys, but tend to be greater in relatively steep areas. In other words, for a smooth surface the circular neighborhood may present the flat areas to be "flatter" and the steep areas to be "steeper". When the slope gradient is measured in percentage (i.e., $100 \times \text{tangent of degree}$), at most locations the differences between the values from the two kinds of neighborhoods are smaller than 1. The values in Figure 4a range from -0.31 to 1.14 , with mean = 0.40 and standard deviation = 0.25 .

For the USGS 10 m DEM, which covers a rugged area, the pattern in the difference layer is not very simple. However, it seems that steep side slopes still tend to get higher values from the circular neighborhood. Figure 4b also shows that in a rugged area, the difference resulting from the two different neighborhoods can be more significant than that for a smooth surface. The values in Figure 4b range from -5.12 to 7.37 , with mean = 0.54 and standard deviation = 1.13 .

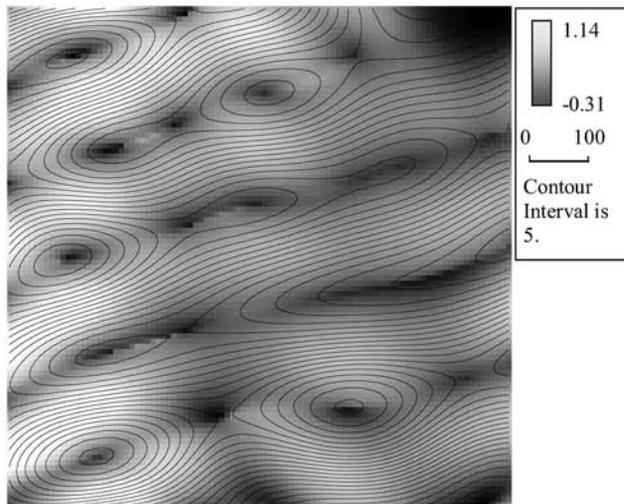
The difference between the values from the two different neighborhoods is eventually determined by the differences between the corresponding elevation values in the diagonal directions in the two neighborhoods, thus this difference is a

function of both the cell size of the DEM and the ratio of the user-specified neighborhood size to the cell size. When the cell size is fixed, increasing the ratio is likely to increase the difference, because a higher ratio results in a greater geographic distance between the corresponding diagonal points in the two neighborhoods, which may lead to a bigger change in elevation. When the ratio is fixed, a bigger cell size also leads to a greater geographic distance between the corresponding diagonal points in the two neighborhoods, and in turn a greater difference of elevation. These effects will be more significant in a rugged area than in a gentle or smooth area.

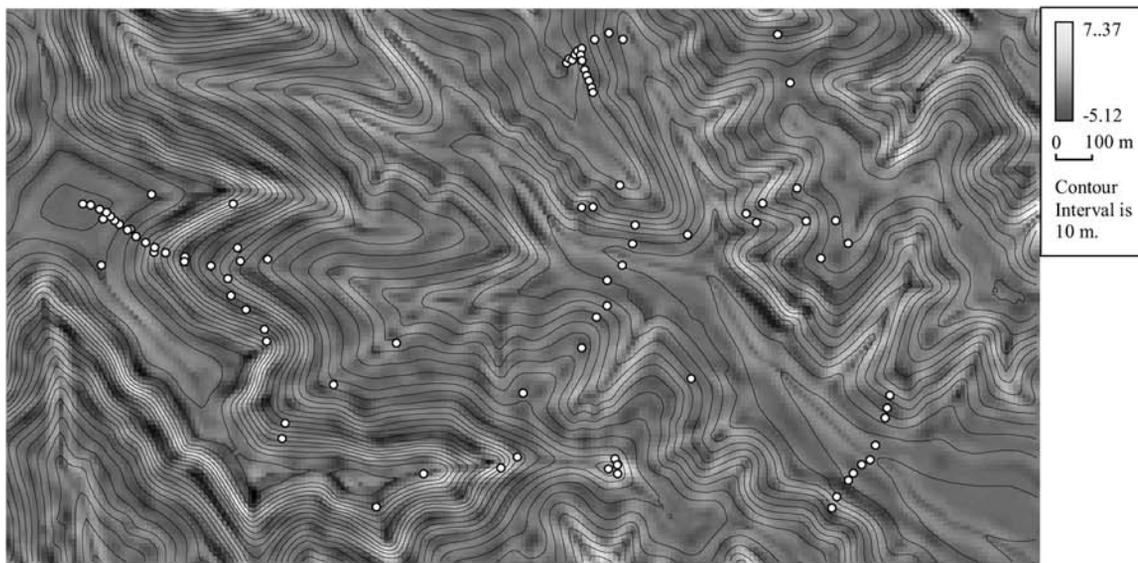
Accuracy for a Smooth Surface

The calculations using rasterized Morrison's surfaces at different resolutions show a common pattern. First, for all the methods, generally the smaller the neighborhood size, the more accurate the calculated slope gradient value (i.e., the smaller MAE, RMSE, and the higher R^2 values), no matter if the smaller neighborhood resulted from a smaller cell size or from the user-specification independent of cell size. The only exception is Z&T, whose accuracy is constant when the user-specified neighborhood is smaller than or equal to the traditional resolution-determined size (i.e., two times of the cell size). For all the other methods, specifying a neighborhood smaller than the traditional size leads to an accuracy higher than that from a calculation using the traditional size. Second, for all the tested cell sizes, when the neighborhood size is equal to or smaller than the traditional size, Z&T is the best among the six methods, which confirms Jones' finding (Jones, 1998a and 1998b). Z&T is followed by M-Z&T-CIR, EVANS-CIR, M-Z&T-SQR, HORN, and EVANS-SQR, but the differences among all the methods are very small. When the neighborhood extends beyond the traditional size, the two circular-neighborhood methods agree with Z&T and become significantly more accurate than the three square-neighborhood methods. Figure 5 illustrates the pattern using the result from one surface (cell size = 10) as an example. As shown by Figure 5, when the neighborhood is smaller than or equal to 20 (which is the traditional size for a surface with a cell size = 10), the differences among the six methods are so small that are even not visually discernable in this figure; as the neighborhood increases, the accuracies of all the methods decrease, but those of the circular-neighborhood methods and Z&T decrease slower than those of the square-neighborhood methods; once the neighborhood is large enough, the circular-neighborhood methods steadily surpass Z&T, in terms of MAE and RMSE (the actual data indicate that this threshold of neighborhood size is 5 times of the cell size or 2.5 times of the traditional neighborhood size).

This pattern is partly predictable. Morrison's surface is a smooth surface whose first-order derivative is continuous. Thus when using a finite-difference method it can be expected that the gradient value (derivative) calculated by a method using elevations of points closer to the center of the kernel will match the analytical result better. This explains the fact that smaller cell sizes and neighborhood sizes give better results in the comparison. This also explains why the circular-neighborhood methods perform better than the square-neighborhood methods: in the circular neighborhood, the values in the four diagonal locations are closer to the center of the kernel than their counterparts in the square neighborhood. One possible reason for that Z&T performs best when the neighborhood is small is that the Lagrange polynomial used by Z&T is an exact fit to the nine points in the neighborhood, and the difference calculated by it for a smooth surface will be closer to the analytical difference than the difference calculated from the least-squares fitted quadratic polynomial.



(a)



(b)

Figure 4. Difference between the results from the different neighborhoods (EVANS-CIR minus EVANS-SQR): (a) calculated from a rasterized Morrison's surface, cell size = 10. The contour lines show the elevation, and (b) calculated from 10 m USGS DEM covering the Raffelson study area. The contour lines show the elevation. The white dots show the locations where slope gradients were measured in the field.

After all, it is important to point out that the above pattern is only valid for smooth surfaces which are differentiable everywhere and the analytical values can be treated as true values. Also, this pattern has nothing to do with the degree of agreement between the computer-calculated values and human-measured values.

Sensitivity to Error

It is interesting to find that the sensitivity results from a synthetic smooth surface and the real DEMs covering a rugged area are very similar (Figures 6 and 7). Generally, the

circular-neighborhood methods are more sensitive to error added to the original surfaces than the square-neighborhood methods, but are less sensitive than Z&T. Figures 6 and 7 illustrate some details of the comparison results from one rasterized Morrison's surface (cell size = 10) and the 10 m USGS DEM. For all the tested noise factors (i.e., c in Equation 4) and shuffle rates, Figures 6 and 7 show only the results from some selected settings, which are sufficient for demonstrating the patterns. As illustrated by Figures 6 and 7, the overall trend is that the sensitivity decreases as the neighborhood size increases, no matter what method is used.

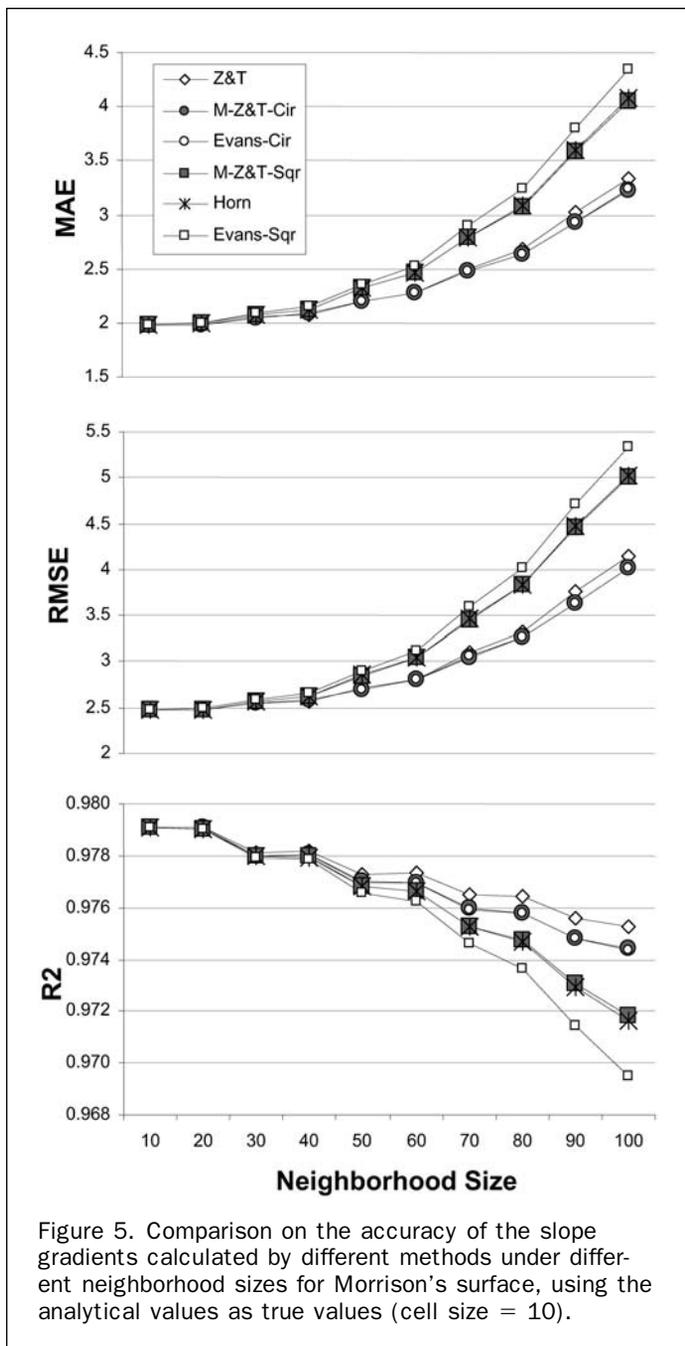


Figure 5. Comparison on the accuracy of the slope gradients calculated by different methods under different neighborhood sizes for Morrison's surface, using the analytical values as true values (cell size = 10).

However, different spatial autocorrelation levels of the noise (i.e., shuffle rates) may cause very different behaviors between the tested methods. When the autocorrelation is very strong (shuffle rate = 0), the RMSE is small, indicating that noise with strong spatial autocorrelation does not significantly affect the result of a gradient calculation. With a zero shuffle rate, the two circular-neighborhood methods group with Z&T and are consistently more sensitive than the three square-neighborhood methods. In addition, at a zero shuffle rate the difference between the two groups of methods increases as the neighborhood size increases. However, the absolute value of that difference remains very small. For example, even when the neighborhood size is ten times of the cell size and the noise factor (c) is as high as 0.16, the difference between EVANS-SQR and EVANS-CIR is

only about 0.1 for the Morrison's surface and about 0.04 for the USGS DEM.

Increasing the shuffle rate to 20 percent causes a jump in the magnitude of the RMSE, but that magnitude decreases dramatically as the neighborhood size increases. At this shuffle rate, especially when the neighborhood size is greater than three times of the cell size (i.e., 1.5 times of the traditional neighborhood size), the two circular-neighborhood methods deviate from Z&T and behave more like the three square-neighborhood methods. Occasionally they are even less sensitive than the square-neighborhood methods (e.g., when the neighborhood size is four times of the cell size). A totally random distribution of noise values (shuffle rate = 100 percent) results in a pattern similar to that from the 20 percent rate, albeit it noticeably increases the RMSE.

Comparison Based on Ground Truth

The results of validating the calculated gradient values using the 90 field-measured values are presented in Figures 8 and 9. Figure 8, which is for the result based on the 10 m DEM, shows that there indeed exists an optimal neighborhood size for matching the field measurements. For this particular study area and this particular DEM, this optimal neighborhood size is around 30 m, which is 1.5 times of the traditional resolution-determined size. Around 30 m, the results from different methods are fairly close to each other, except that Z&T is significantly worse than all the other methods in terms of R^2 . When the neighborhood size is smaller than 30 m, the square-neighborhood methods are slightly better than the circular-neighborhood methods. When the neighborhood size is greater than 30 m, especially when greater than 40 m, the circular-neighborhood methods perform better than all the other methods. The relatively gentle curves of the circular-neighborhood methods displayed in Figure 8 also indicate that the "optimal ranges" of these curves are wider than those of the other curves. An implication of this is that when the most optimal neighborhood size is unknown, using a circular neighborhood may have a higher probability of better matching the field measurements.

The result based on the 30 m DEM is somewhat different. First, all the four statistics indicate that the slope gradient values calculated from the 30 m DEM diverge more from the field measurements than the values from the 10 m DEM. Second, the "most optimal" neighborhood size is 10 m, which is the smallest user-specified size tested and is one sixth of the traditional resolution-determined size. This second finding indicates that using a user-specified neighborhood size may improve the quality of terrain attribute values derived from a low resolution DEM. Third, when the neighborhood is small, the best method for the 10 m DEM, EVANS-SQR, becomes the worst for the 30 m DEM, and the worst one for the 10 m DEM, Z&T, becomes the best for the 30 m DEM. Fourth, the accuracy of all methods drops sharply once the neighborhood extends beyond 60 m, but the accuracy of the two circular-neighborhood methods drops slower than all the other methods. Finally, for the 30 m DEM, the two circular-neighborhood methods always perform better than the three square-neighborhood methods.

Conclusion

In this research, a circular neighborhood was implemented for the Evans method and a modified Zevenbergen-Thorne method for calculating slope gradient from a gridded DEM. Three approaches were used to compare the results from the circular neighborhood with the results from the traditional

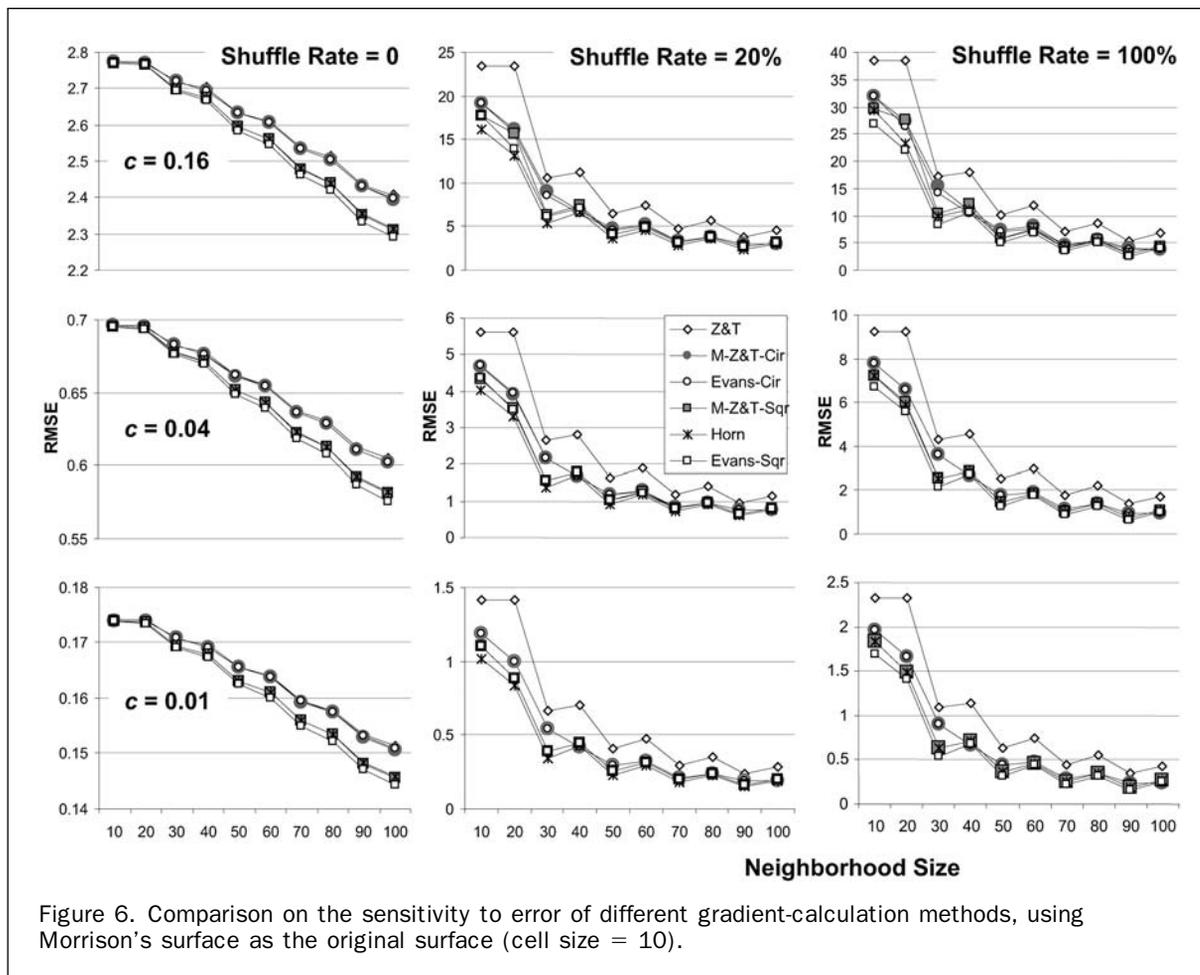


Figure 6. Comparison on the sensitivity to error of different gradient-calculation methods, using Morrison's surface as the original surface (cell size = 10).

square neighborhood. A comparison based on Morrison's surface reveals that, for a smooth surface, a circular-neighborhood method is more accurate than a square-neighborhood method. In addition, when the ratio of the user-specified neighborhood size to the cell size is large enough (the threshold is 5), a circular-neighborhood methods can be even more accurate than the Zevenbergen-Thorne method (as measured by MAE and RMSE).

Tests using *noised* surfaces show that a circular-neighborhood method is generally more sensitive to error in the DEM than a square-neighborhood method, but is generally less sensitive than the Zevenbergen-Thorne method. When the error has a strong spatial autocorrelation, the difference between a circular-neighborhood method and a square-neighborhood method is apparent and the difference increases as the neighborhood size increases. However, in this situation the RMSE between the values calculated from the original and *noised* surfaces itself is very small. When the spatial autocorrelation in the error is not so strong, a circular neighborhood method behaves more like a square-neighborhood method, especially when the neighborhood size is large, where occasionally a circular neighborhood method can be even less sensitive than a square-neighborhood method.

A validation based on field measurements shows that the circular neighborhood performs better in general than the square neighborhood when the neighborhood size is large and/or the resolution of the DEM is low. When both the neighborhood size and the cell size are small, the difference between the results from the two kinds of neighborhoods is

small and inconsistent. In any case, a circular neighborhood method has a wider optimal range of neighborhood size than other tested methods, indicating that when the most optimal neighborhood size is unknown, using a circular neighborhood may have a higher probability of better matching the field measurements.

Some general conclusions can be drawn from these findings. First, the circular-neighborhood methods tested in this research are a compromise between the Zevenbergen-Thorne method (using only the four points in the cardinal directions) and the methods using an eight-point square neighborhood. The Zevenbergen-Thorne method is more accurate for a smooth surface, but is more sensitive to noise in the DEM, whereas the square-neighborhood methods are less accurate for a smooth surface, but are less sensitive to noise. The circular-neighborhood methods perform between the two when the neighborhood size is small, and may surpass those traditional methods when the ratio of neighborhood size to cell size is large. If the sensitivity to error is the main concern when choosing a method, a circular neighborhood does not have advantage. However, a method insensitive to error may also be insensitive to real variation in the data. In terms of the overall performance, the circular neighborhood is a good alternative to the traditional square neighborhood. If a GIS package only implements one method for calculating slope gradient, the circular neighborhood should be considered.

Second, the circular neighborhood may be more advantageous when used together with a user-specified

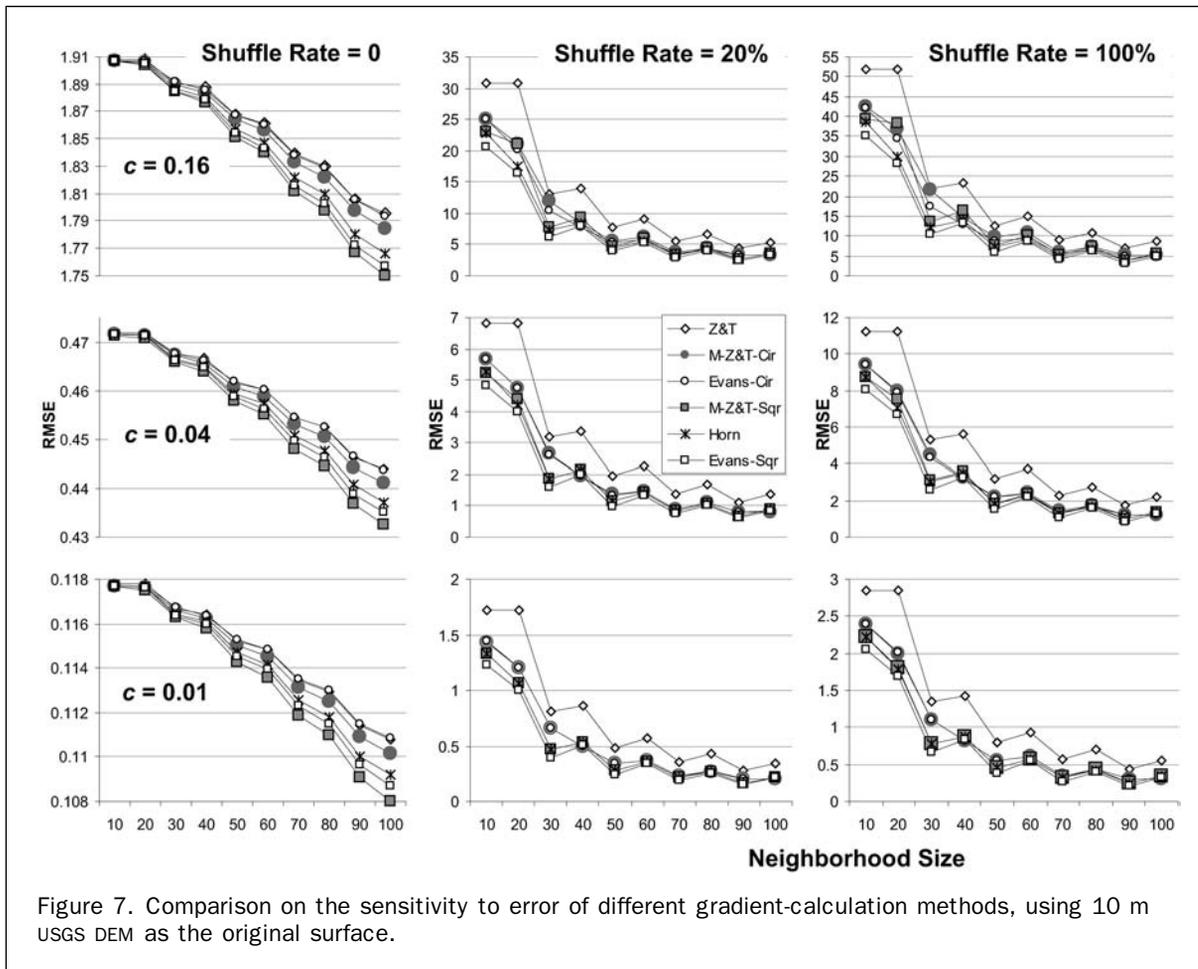


Figure 7. Comparison on the sensitivity to error of different gradient-calculation methods, using 10 m USGS DEM as the original surface.

neighborhood size on a high-resolution DEM. The case study presented in this paper demonstrates that the most optimal neighborhood size for matching field measurements may be different from the traditional resolution-determined size. The case study also shows that the circular neighborhood methods surpass all the other methods when the ratio of neighborhood size to cell size is large. More work is needed to find out if this is a general pattern. If it is, it may have significant implications for future terrain analyses in (human) knowledge-related applications (e.g., knowledge-based soil mapping), since high-resolution DEMs are becoming more readily available. For such DEMs, the user may want to specify a neighborhood size greater than the resolution-determined size for a better match between the calculated values and the field measurements.

Future work may include comparisons of the two kinds of neighborhoods using field measurements from different study areas, using DEMs at very high resolutions, and using higher-order terrain attributes, including various curvatures.

Acknowledgments

Support from the Walter and Constance Burke Award of Dartmouth College is gratefully acknowledged by Xun Shi. Support from the Chinese Academy of Science utilizing the "One-Hundred Talents" program to A-Xing Zhu is greatly appreciated.

A-Xing Zhu was also partially supported by the Chinese Academy of Sciences International Partnership Project "Human Activities and Ecosystem Changes," Project No. CXTD-Z2005-1.

Appendix

Derivation of Coefficient Values in the Quadratic and Lagrange Polynomials for the Circular Neighborhood

Refer to Equation 1 for the quadratic polynomial and Figure 1 for the positions of z_1, z_2, \dots, z_9 . Values for r, t, s, p, q , and u in Equation 1 can be obtained by fitting a second-order surface defined by Equation 1 to the nine points with elevation values z_1, z_2, \dots, z_9 . A least squares method to achieve this fit is to use the following equation (Pennock *et al.*, 1987):

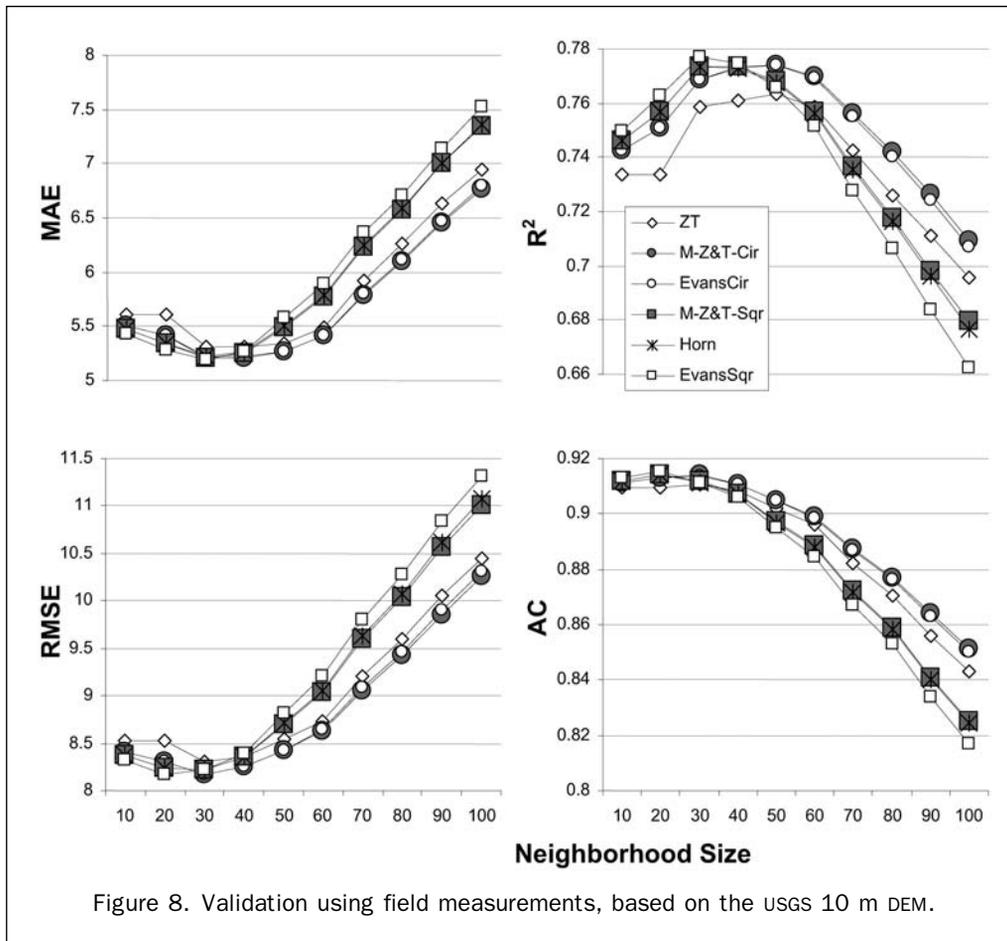
$$\mathbf{F}^T \mathbf{F} \boldsymbol{\beta} = \mathbf{F}^T \mathbf{z} \quad (\text{A1})$$

where \mathbf{z} is a 9×1 matrix composed of z_1, z_2, \dots, z_9 , $\boldsymbol{\beta}$ is a 6×1 matrix composed of the six coefficients in Equation 1 (Note that $\frac{1}{2}r$ and $\frac{1}{2}t$ should be used in $\boldsymbol{\beta}$). \mathbf{F} is a 9×6 matrix of the form:

$(w/\sqrt{2})^2$	$(w/\sqrt{2})^2$	$-(w/\sqrt{2})^2$	$-w/\sqrt{2}$	$w/\sqrt{2}$	1
0	w^2	0	0	w	1
$(w/\sqrt{2})^2$	$(w/\sqrt{2})^2$	$(w/\sqrt{2})^2$	$w/\sqrt{2}$	$w/\sqrt{2}$	1
w^2	0	0	$-w$	0	1
0	0	0	0	0	1
w^2	0	0	w	0	1
$(w/\sqrt{2})^2$	$(w/\sqrt{2})^2$	$(w/\sqrt{2})^2$	$-w/\sqrt{2}$	$-w/\sqrt{2}$	1
0	w^2	0	0	$-w$	1
$(w/\sqrt{2})^2$	$(w/\sqrt{2})^2$	$-(w/\sqrt{2})^2$	$w/\sqrt{2}$	$-w/\sqrt{2}$	1

and \mathbf{F}^T is the transpose matrix of \mathbf{F} . $\boldsymbol{\beta}$ is solved as

$$\boldsymbol{\beta} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{z} \quad (\text{A2})$$



$(\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T$ is constant and its resulting 6×9 matrix is as follows:

$1/(8w^2)$	$-1/(8w^2)$	$1/(8w^2)$	$3/(8w^2)$	$-1/w^2$	$3/8w^2$	$1/8w^2$	$-1/(8w^2)$	$1/(8w^2)$
$1/(8w^2)$	$3/(8w^2)$	$1/(8w^2)$	$-1/(8w^2)$	$-1/w^2$	$-1/(8w^2)$	$1/(8w^2)$	$3/(8w^2)$	$1/(8w^2)$
$-1/(2w^2)$	0	$1/(2w^2)$	0	0	0	$1/(2w^2)$	0	$-1/(2w^2)$
$-\sqrt{2}/(8w)$	0	$\sqrt{2}/(8w)$	$-1/(4w)$	0	$1/(4w)$	$-\sqrt{2}/(8w)$	0	$\sqrt{2}/(8w)$
$\sqrt{2}/(8w)$	$1/(4w)$	$\sqrt{2}/(8w)$	0	0	0	$-\sqrt{2}/(8w)$	$-1/(4w)$	$-\sqrt{2}/(8w)$
0	0	0	0	1	0	0	0	0

Therefore, the coefficients in β can be solved as:

$$r = 2[z_1/(8w^2) - z_2/(8w^2) + z_3/(8w^2) + 3z_4/(8w^2) - z_5/w^2 + 3z_6/8w^2 + z_7/8w^2 - z_8/(8w^2) + z_9/(8w^2)] \\ = (z_1 + z_3 + 3z_4 + 3z_6 + z_7 + z_9 - z_2 - 8z_5 - z_8)/4w^2$$

$$t = 2[z_1/(8w^2) + 3z_2/(8w^2) + z_3/(8w^2) - z_4/(8w^2) - z_5/w^2 - z_6/(8w^2) + z_7/8w^2 + 3z_8/8w^2 + z_9/(8w^2)] \\ = (z_1 + z_3 + 3z_2 + 3z_8 + z_7 + z_9 - z_4 - 8z_5 - z_6)/4w^2$$

$$s = -z_1/(2w^2) + z_3/(2w^2) + z_7/(2w^2) - z_9/(2w^2) \\ = (z_3 + z_7 - z_1 - z_9)/2w^2$$

$$p = -\sqrt{2}z_1/(8w) + \sqrt{2}z_3/(8w) - z_4/(4w) + z_6/(4w)$$

$$- \sqrt{2}z_7/(8w) + \sqrt{2}z_9/(8w) \\ = (\sqrt{2}z_3 + 2z_6 + \sqrt{2}z_9 - \sqrt{2}z_1 - 2z_4 - \sqrt{2}z_7)/8w$$

$$q = -\sqrt{2}z_1/(8w) + z_2/(4w) + \sqrt{2}z_3/(8w) - \sqrt{2}z_7/(8w) \\ - z_8/(4w) - \sqrt{2}z_9/(8w)$$

$$= (\sqrt{2}z_1 + 2z_2 + \sqrt{2}z_3 - \sqrt{2}z_7 - 2z_8 - \sqrt{2}z_9)/8w$$

$$u = z_5$$

Refer to Equation 2 for the Lagrange polynomial and Figure 1 for the positions of z_1, z_2, \dots, z_9 . Following Zevenbergen and Thorne (1987), the goal of the derivation is to exactly fit a Lagrange surface to the nine points in the neighborhood. Therefore, values for the coefficients in the polynomial can be obtained by solving the following equation group for the unknowns a, b, c, r, t, s, p, q , and u :

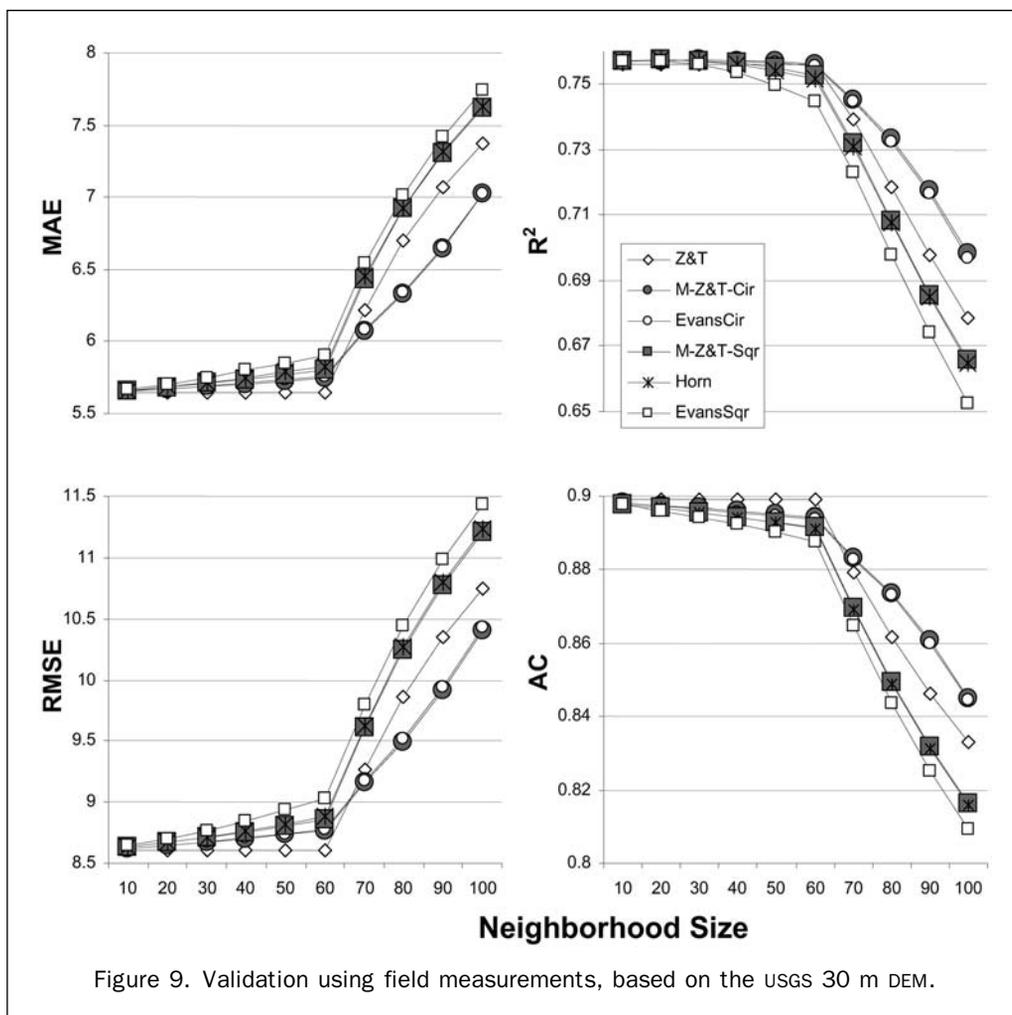


Figure 9. Validation using field measurements, based on the USGS 30 m DEM.

$$\begin{aligned}
 z_1 &= a(-w/\sqrt{2})^2(w/\sqrt{2})^2 + b(-w/\sqrt{2})^2(w/\sqrt{2}) \\
 &+ c(-w/\sqrt{2})(w/\sqrt{2})^2 + \frac{1}{2}r(-w/\sqrt{2})^2 + \frac{1}{2}t(w/\sqrt{2})^2 \\
 &+ s(-w/\sqrt{2})(w/\sqrt{2}) + p(-w/\sqrt{2}) + q(w/\sqrt{2}) + u \\
 z_2 &= \frac{1}{2}tw^2 + qw + u \\
 z_3 &= a(w/\sqrt{2})^2(w/\sqrt{2})^2 + b(w/\sqrt{2})^2(w/\sqrt{2}) + c(w/\sqrt{2})(w/\sqrt{2})^2 \\
 &+ \frac{1}{2}r(w/\sqrt{2})^2 + \frac{1}{2}t(w/\sqrt{2})^2 \\
 &+ s(w/\sqrt{2})(w/\sqrt{2}) + p(w/\sqrt{2}) + q(w/\sqrt{2}) + u \\
 z_4 &= \frac{1}{2}r(-w)^2 + p(-w) + u \\
 z_5 &= u \\
 z_6 &= \frac{1}{2}rw^2 + pw + u \\
 z_7 &= a(-w/\sqrt{2})^2(-w/\sqrt{2})^2 + b(-w/\sqrt{2})^2(-w/\sqrt{2}) \\
 &+ c(-w/\sqrt{2})(-w/\sqrt{2})^2 + \frac{1}{2}r(-w/\sqrt{2})^2 + \frac{1}{2}t(-w/\sqrt{2})^2 \\
 &+ s(-w/\sqrt{2})(-w/\sqrt{2}) + p(-w/\sqrt{2}) + q(-w/\sqrt{2}) + u \\
 z_8 &= \frac{1}{2}t(-w)^2 + q(-w) + u \\
 z_9 &= a(w/\sqrt{2})^2(-w/\sqrt{2})^2 + b(w/\sqrt{2})^2(-w/\sqrt{2}) \\
 &+ c(w/\sqrt{2})(-w/\sqrt{2})^2 + \frac{1}{2}r(w/\sqrt{2})^2 + \frac{1}{2}t(-w/\sqrt{2})^2 \\
 &+ s(w/\sqrt{2})(-w/\sqrt{2}) + p(w/\sqrt{2}) + q(w/\sqrt{2}) + u.
 \end{aligned}$$

References

- Burt, J.E., and A.-X. Zhu, 2002. *The Help Document of 3dMapper*, Terrain Analytics, LLC, Madison, Wisconsin, URL: <http://www.terrainanalytics.com> (last date accessed: 03 November 2006).
- Chang, K., and B. Tsai, 1991. The effect of data precision on the calculation of slope and aspect using gridded DEMs, *Cartographica*, 29:22–34.
- Evans, I.S., 1980. An integrated system of terrain analysis and slope mapping, *Zeitschrift für Geomorphologie*, N.F., Supplementband, 36:274–295.
- Florinsky, I., 1998. Accuracy of local topographic variables derived from digital elevation models, *International Journal of Geographical Information Science*, 12:47–61.
- Gao, J., 1997. Resolution and accuracy of terrain representation by grid DEMs at a micro-scale, *International Journal of Geographical Information Science*, 11:199–212.

- Hodgson, M.E., 1995. What cell size does the computed slope/aspect angle represent?, *Photogrammetric Engineering & Remote Sensing*, 61:513–517.
- Horn, B.K.P., 1981. Hill shading and the reflectance map, *Proceedings of the IEEE*, 69:14–47.
- Hunter, G., and M. Goodchild, 1997. Modeling the uncertainty of slope and aspect estimates derived from spatial databases, *Geographical Analysis*, 29:35–49.
- Jones, K.H., 1998a. A comparison of algorithms used to compute hill slope as a property of the DEM, *Computer & Geosciences*, 24:315–323.
- Jones, K.H., 1998b. A comparison of two approaches to ranking algorithms used to compute hill slopes, *Geoinformatica*, 2: 235–256.
- Kienzle, S., 2004. The effect of DEM raster resolution on first order, second order and compound terrain derivatives, *Transactions in GIS*, 8:83–111.
- Morrison, J.L., 1971. *Method-produced Error in Isarithmic Mapping*, American Congress on Surveying and Mapping, Washington, D.C., Technical Monograph No. CA-5, 24.
- Morrison, J.L., 1974. Observed statistical trends in various interpolation algorithms useful for first stage interpolation, *The Canadian Cartographer*, 11:142–159.
- Pennock, D.J., B.J. Zebarth, and E. de Jong, 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada, *Geodema*, 40:297–315.
- Shary, P.A., 1995. Land surface in gravity points classification by complete system of curvatures, *Mathematical Geology*, 27: 373–390.
- Travis, M.R., G.H. Elsner, W.D. Iverson, and C.G. Johnson, 1975, VIEWIT computation of seen areas, slope and aspect for land-use planning, *U.S. Dept of Agriculture Forest Service General Technical Report PSW 11/1975*, Pacific Southwest Forest and Range Experimental Station, Berkeley, California.
- Willmott, C.J., 1984. On the evaluation of model performance in physical geography, *Spatial Statistics and Models* (G.L. Gaile and C.J. Willmott, editors), D. Reidel Publishing Company, Dordrecht, The Netherlands, pp. 443–440.
- Zevenbergen, L.W., and C.R. Thorne, 1987. Quantitative analysis of land surface topography, *Earth Surface Processes and Landforms*, 12:47–56.
- Zhou, Q., and X. Liu, 2004. Analysis of errors of derived slope and aspect related to DEM data properties, *Computers & Geosciences*, 30:369–378.
- Zhu, A.X., L.E. Band, R. Vertessy, and B. Dutton, 1997. Derivation of soil properties using a soil land inference model (SoLIM), *Soil Science Society of America Journal*, 61: 523–533.

(Received 03 August 2005; accepted 27 September 2005; revised 10 October 2005)