October 2004

The Discontinuous Nature of Kriging Interpolation for Digital Terrain Modeling

Thomas H. Meyer
University of Connecticut, thomas.meyer@uconn.edu

The Discontinuous Nature of Kriging Interpolation for Digital Terrain Modeling

Thomas H. Meyer

ABSTRACT: Kriging is a widely employed method for interpolating and estimating elevations from digital elevation data. Its place of prominence is due to its elegant theoretical foundation and its convenient practical implementation. From an interpolation point of view, kriging is equivalent to a thin-plate spline and is one species among the many in the genus of weighted inverse distance methods, albeit with attractive properties. However, from a statistical point of view, kriging is a best linear unbiased estimator and, consequently, has a place of distinction among all spatial estimators because any other linear estimator that performs as well as kriging (in the least squares sense) must be equivalent to kriging, assuming that the parameters of the semivariogram are known. Therefore, kriging is often held to be the gold standard of digital terrain model elevation estimation. However, I prove that, when used with local support, kriging creates discontinuous digital terrain models, which is to say, surfaces with "rips" and "tears" throughout them. This result is general; it is true for ordinary kriging, kriging with a trend, and other forms. A U.S. Geological Survey (USGS) digital elevation model was analyzed to characterize the distribution of the discontinuities. I show that the magnitude of the discontinuity does not depend on surface gradient but is strongly dependent on the size of the kriging neighborhood.

KEYWORDS: Kriging, digital terrain model, discontinuity

Introduction

Kriging (Matheron 1963) is a popular technique for interpolating and estimating elevation values from digital terrain data. General references on the subject include David (1977), Isaaks and Srivastava (1989), Goovaerts (1997), and Journal and Huijbregts (1997). Bailey (1994, p. 32) asserts that "there is an argument for kriging to be adopted as a basic method of surface interpolation in geographic information systems (GIS) as opposed to the standard deterministic tessellation techniques that currently prevail and which can produce artificially smooth surfaces." This argument was supported by Laslett (1994) whose study gives an example of a data set for which splines are "too smooth," and kriging results in more precise estimations. Regarding the "spline vs. kriging" debate, it has been shown (Kimeldorf and Wahba 1971; Wahba 1990) that kriging is mathematically equivalent to thin-plate splines, and Almansa et al. (2002) established its place in a more general class of functions, namely, the absolutely minimizing Lipschitz extension.

While kriging is not without its critics (Philip and Watson 1986) there is no question that its use is widespread. The properties of any mathematical surface being used as a terrain model define the properties imbued to the model. The onus is on the modeler to choose the model wisely so that its properties match the desired traits of the terrain. Continuity properties are of paramount importance. Discontinuous surfaces have "holes" or "tears" in them. Continuous surfaces might not be smooth, meaning that the surface might have one or more "creases" in it.

The computational geometry literature is replete with discussions about the continuity of patches within themselves and with their neighbors (de Boor 1978; Lancaster and Šalkauskas 1986; Dierckx 1993; Farin 1993), but there is relatively little attention given to surface discontinuity properties in the GIS literature. Although Meyer (1999) gives a review of the continuity properties of digital terrain models categorizing them according to their topological and continuity characteristics, Lam (1983, p.134) set the stage for the current research when she reported concerning kriging that, "Choice of neighborhood will also affect the continuity properties of the estimates…If the change of data points from one neighborhood to the next is too abrupt there may be discontinuities even though the actual phenomenon is continuous." The purpose of this paper is to carefully document this undesirable characteristic of kriging, namely, that kriging as it is typically used for digital ter-

Thomas H. Meyer, Department of Natural Resources Management and Engineering, University of Connecticut, Storrs, Connecticut, W.B. Young Building, rm 308, 1376 Storrs Road, Storrs, CT 06256-4087. Tel: (860) 486-0145; Fax: (860) 486-5408. E-mail: <thomas.meyer@uconn.edu>.

rain modeling produces a piece-wise discontinuous surface.

This paper is organized as follows. A brief explanation of the notation and the theoretical setting of digital terrain modeling in the context of this paper is presented. This is followed by a proof that kriging can produce discontinuous surfaces along with a discussion that generalizes the results. This is followed by a section presenting an analysis of the discontinuities found in a USGS 7.5' digital elevation model, which provides some idea of the magnitude and distribution of the worst discontinuities encountered in this real-world data set. The final section of the paper provides a discussion, summary, and conclusions.

Theoretical Setting

Define a *surface* to be a real-valued function of two variables, $z = f(x,y)$. Let f be defined over an open, bounded, simply connected, two-dimensional region R that is a subset of three-dimensional Euclidean real space. A digital terrain model (DTM) is often constructed from a set of samples, s, taken from the area of interest, such that these samples are intended to capture the essence of the terrain's shape. Digital terrain modeling includes the problem of interpolating s to derive heights at places in R for which no sample was taken. Some surfaces are defined upon all the sample points in s. Examples include Lagrange polynomials, Fourier transformations, and kriging. Such functions are said to have *global support*, meaning that every point in s contributes to the formulation of f.

Global support is generally not desirable for digital terrain modeling for several reasons. It imposes heavy computational burdens for large data sets. Also, it has the counter-intuitive property that, for certain techniques such as the Lagrange polynomials, making a small change in any particular sample can produce large changes over the entire surface. This runs contrary to Tobler's Law of Geography, "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970, p. 236). Also, polynomial global support surface models which interpolate all the points in s must be of an order equal to the cardinality of s, or greater. This can produce unwanted behaviors such as extreme surface departures and unrealistic undulations. From a statistical point of view, global support leads to over-fitting and yields poor validation.

Although kriging is defined with global support (Siška et al.1997), in practice it is not typically used that way for terrain modeling. Instead, s is subdivided into *neighborhoods*, which are subsets of s with relatively few elements that roughly (or strictly) partition s. Then, kriging is used to interpolate over the neighborhoods in the following way. Suppose I want to interpolate a surface value at the point $p = (x,y)$ and p is not an element of s. Let n denote a neighborhood surrounding p. Then, the height estimate at p is a weighted sum of the heights of n. The weights are related inversely to the distance from the sample to p in a way that the variance of the estimate is minimized. The surface produced over a neighborhood is called a *patch*.

There are many heuristics for choosing the neighborhood (Isaaks and Srivastava, 1989, p.338). Some of the heuristics include using:

- All samples within some circle or ellipse enclosing (x,y);
- All samples within some convex polygon enclosing (x,y),
- A limited number of samples from the four quadrants enclosing (x,y), or
- Voronoi nearest neighbors.

These heuristics divide the region R into subregions, or *patches*, such that all points in a patch have the same interpolation neighborhood. In the case of Voronoi nearest neighbors, the patches are Voronoi polygons. The equivalence of kriging and thin-plate splines guarantees that, within a patch, $f(x,y)$ is smooth and, therefore, continuous. This is a consequence of splines being polynomials and polynomials being infinitely differentiable. However, it has not been documented that $f(x,y)$ can be discontinuous on the borders between the patches. This shall be established in the next section.

Discontinuity Proof

I claim that a surface created by ordinary kriging can be patch-wise discontinuous. It suffices to produce a single example to establish the claim. The proof proceeds as follows.

1. Choose a data set.
2. Find the patches defined over the data set. Each patch constitutes a neighborhood.
3. Select a border between two contiguous patches.
4. Interpolate the border twice, once for each neighborhood of the two patches.
5. If the two interpolated borders differ in elevation, then $f(x,y)$ is discontinuous along the border.
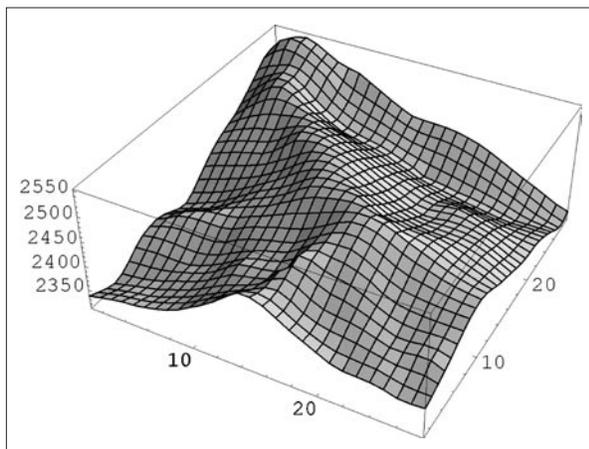
The Data Set and Its Covariance Function

The topographic data set comes from the USGS 7.5' digital elevation model (DEM) *Sandia Crest*, a
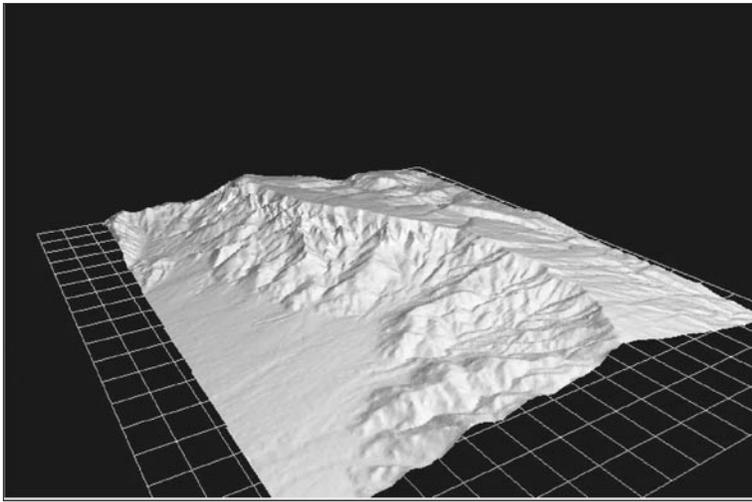
Figure 1. The USGS DEM of Sandia Crest, New Mexico. The mountains run south to north, south being in the foreground.
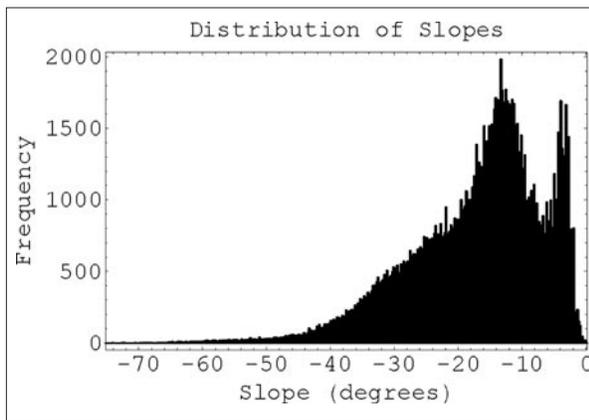
Figure 2. The surface gradients at each posting of the USGS Sandia Crest DEM.

Figure 3. The topographic data set for the proof.

mountainous region in north-central New Mexico (Figure 1). This area was chosen because it contains a diverse set of topographic surface types, including alluvial fans, talus slopes, cliffs, and inclined planes. This DEM has a minimum and maximum elevation of 1,770 and 3,248 m, respectively, for a vertical relief of 1,478 m and has a wide variety of gradients (Figure 2). The city of Albuquerque covers the alluvial fans seen in the foreground of the figure. Parts of this region are very rugged, but such terrain is often encountered in the common practice of surveying and mapping.

For the proof, I will use a small subset of the DEM taken from the foothills, see Figure 3. The site is located in UTM (NAD83) zone 13 with corner coordinates (364,140 E, 3,900,690 N) and (364,980 E, 3,901,590 N) and measures 840 meters east to west and 810 meters north to south.

The empirical omni-directional semivariogram of the data set was computed using custom programs written in *Mathematica* v.5.0 (Wolfram 1999). The semivariogram is the set of points shown in Figure 4. There is no discernable anisotropy for distances less than two hundred meters, a distance greater than the largest nearest neighbor distance. Therefore, the omni-directional variogram was judged to be an adequate model, and no directional variograms were computed. Throughout this paper I select analytical models of empirical semivariograms by fitting linear, spherical, exponential, and Gaussian functions to the data (Isaaks and Srivastava 1989, pp. 372-375). I chose the function with the largest correlation coefficient. The semivariogram in Figure 4 is best fit by a Gaussian function and, thus, is of the form:

$$\gamma(h) = 1 - e^{-3h^2/a^2}$$

where h = lag distance; and
a = the practical range. In this case, a was chosen to be 500 meters. A least-squares fit yielded the model:

$$\gamma(h) = 3219.67(1 - e^{-3h^2/250000})$$

The model and the data are depicted in Figure 4. By assuming stationarity, the covariance function is related to the variogram by:
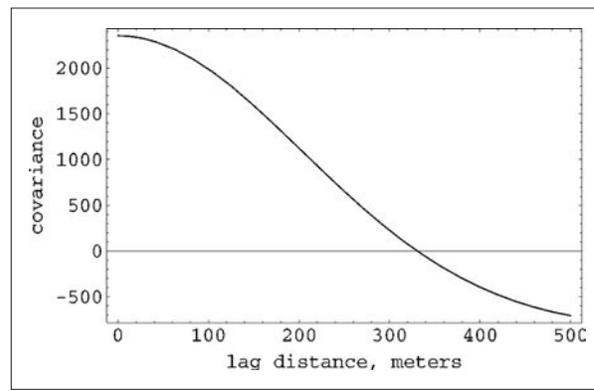
$$\text{cov}(h) = \sigma^2 - \gamma(h)$$

where σ^2 is the variance of the data set and has a value of 2,352.829. Figure 5 shows the graph of the covariance function.
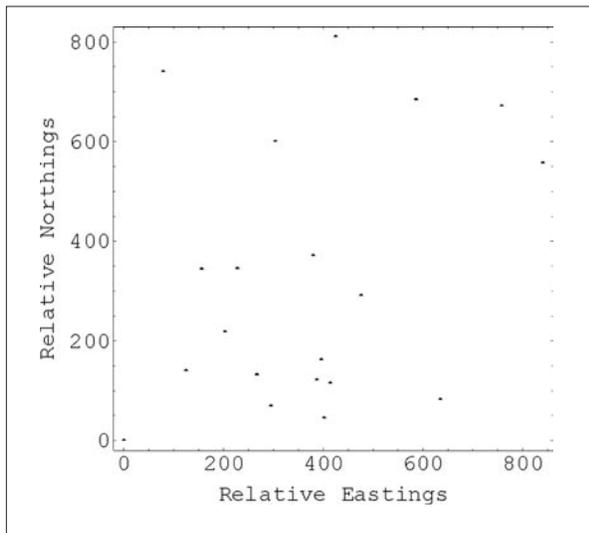
Patches
Having created a covariance function with which to form the kriging weights matrix, the next task is to tessellate the region into patches. It was
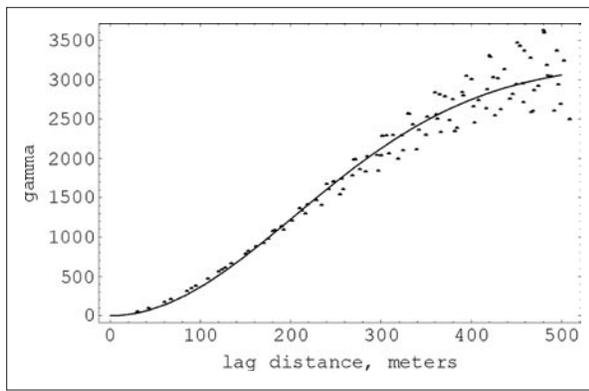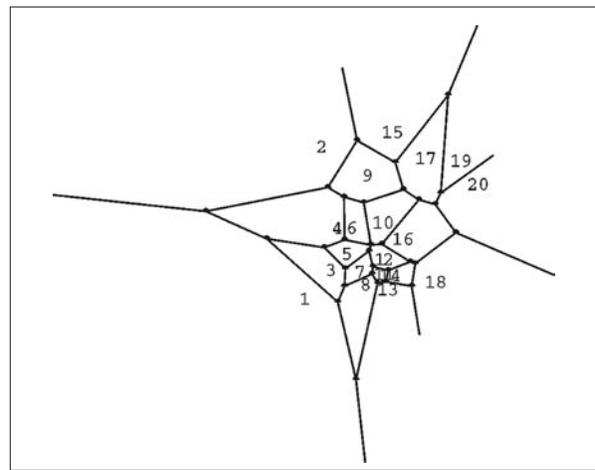
Figure 4. The omni-directional variogram and its least squares Gaussian model.

Figure 5. The covariance function.

Figure 6. The points used to create a Voronoi diagram for the study area are indicated by the integers.

Figure 7. The Voronoi diagram for the study area. Individual sample points are indicated by numbers. The polygons are subregions of R such that every point in a polygon corresponds to the same interpolation neighborhood. Compare with Figure 6.

decided to use Voronoi nearest neighbors and Voronoi polygons as interpolating neighborhoods and patches, respectively. Twenty (easting, northing) pairs were generated randomly from a uniform distribution. These pairs were constrained to fall within the study area and are shown in Figure 6. Figure 7 shows the Voronoi diagram of the pairs. The numbers in Figure 7 correspond to the points in Figure 6.

Proof by Example

Consider the border between the polygons generated by points 6 and 10. The nearest neighbors of point 6 are {10, 9, 4, 5, 12}. The nearest neighbors of point 10 are {6, 12, 16, 17, 9}. The two Voronoi vertices defining the common border between points 6 and 10 are (285.4, 466.7) and (320.5, 261.3). The border between the polygons was enumerated by kriging at 100 points distributed evenly between these two Voronoi vertices.

Kriging was done using custom programs written in *Mathematica* v5.0 by constructing the matrices and computing their inverses. The results are shown in Figures 8 and 9. The visual similarity of the two borders confirms that the interpolation is working correctly; they should be very similar. However, the difference of the interpolated values shown in Figure 10 clearly depicts the discontinuity. The discontinuity ranges in value from 3.08 meters to -1.72 meters. This completes the proof.

Generalization of Results

The fundamental source of the discontinuity is that the border is being interpolated twice according to two different covariance matrices. A kriging estimate is formed by the product of a matrix and a vector. The matrix contains values related to the value of a covariance function (Figure 5) evaluated at the lag distances between the point of interest
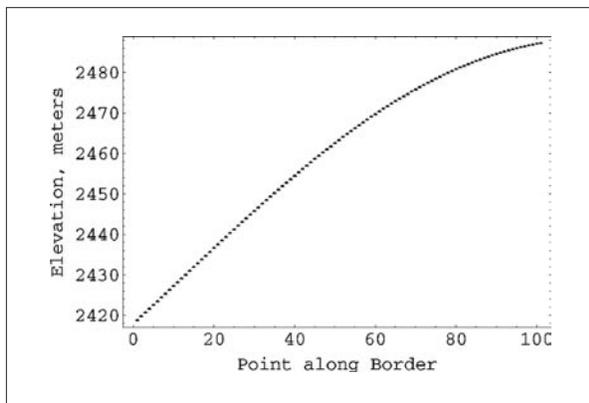
Figure 8. The border as interpolated with the nearest neighbors of point 6.

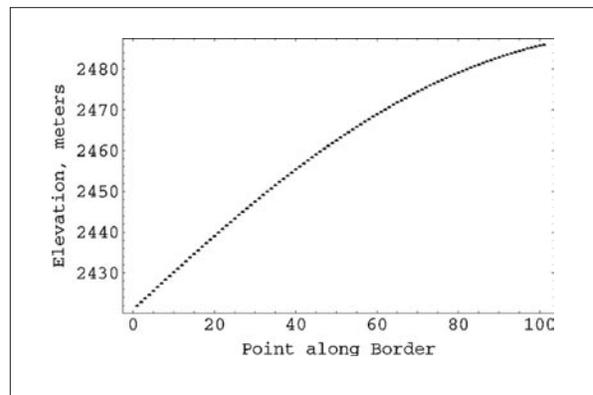**Figure 9.** The border as interpolated with the nearest neighbors of point 10.

and the neighborhood points. The vector contains the elevations of the neighborhood points. In general, if two kriging neighborhoods are different, meaning they do not consist of the same points, then the covariance matrices and vectors formed from those points will also be different, and this typically leads to different estimations. Even so, it is possible for the two estimates to agree, and Figure 10 shows exactly one place where the two estimated borders cross near the 50th point; there is a place where they had exactly the same value. Even so, the different covariance matrices lead to different estimations, in general. Furthermore, all forms of kriging (universal, co-kriging, block kriging) form estimates by this same basic procedure. Therefore, the discontinuity property is general. It is not a consequence of this data set or of the choice of using ordinary kriging.

Experimentation

Having proven that kriging can produce discontinuities at its patch borders, I wanted to investigate just how large of a discontinuity might be encountered in practice and undertook to characterize causes that might exacerbate or mitigate the size of the discontinuities.

The experiment was to repeat the analysis detailed in the proof over the entire Sandia Crest DEM. I felt that the entire DEM was too inhomogeneous to be modeled by a single semivariogram. Therefore, the DEM was subdivided into 900 disjoint regions of essentially equal size (square with the same number of postings), and an empirical semivariogram was computed for each region and an analytical model fit to each semivariogram.

The postings of a USGS DEM form a square lattice, and I had observed that *Mathematica's* Delaunay triangulation algorithm occasionally produces an

incorrect tessellation for square grids. Therefore, the Delaunay triangulation and Voronoi diagram for each region were computed after displacing each posting by a small (max ±1 cm) random amount. This small random perturbation produced a distribution of border lengths. I measured the lengths of all the borders and observed that over 95 percent of them were between 20 and 40 meters. For uniformity and to exclude borders that extended beyond the convex hull, I did not consider borders shorter than 20 m or longer than 40 m. A total of 163,017 borders were tested. Ten positions were computed along each border, and elevations were estimated at each position twice, once for the neighborhood of postings defined by each Voronoi polygon sharing that border. I speculated that slope and neighborhood size might influence the results, so slope was computed at every DEM posting using the method found in Meyer et al. (2001). I then repeated the experiment using neighborhoods of the first extended nearest neighbors, meaning the set of postings that are the nearest neighbors of the point-of-interest's nearest neighbors. The extended neighborhoods always overlapped each other. As each border was tested, I kept track of summary statistics for the maximum discontinuity (absolute value) and the variability of the discontinuities at each point.

Results

Figure 11 shows two histograms of the worst discontinuities for the 900 regions. On the left is a histogram for the no-overlapping neighborhood case, which has a maximum discontinuity of 13.2 m. On the right is the histogram for the overlapping neighborhood case, which has a maximum discontinuity of 3.0 m. Note the dissimilar ordinate scales. In 100 percent of the cases, the overlapping neighborhoods produced smaller discontinuities.

Figure 12 shows a histogram of the improvements afforded by overlapping the neighborhoods. The ordinate gives the ratio of the size of a border's discontinuity with overlapping neighborhoods to the size of a border's discontinuity without over-lapping neighborhoods. Thus, an ordinate value of 1.0 indicates no improvement, and a value of 0.1 indicates the overlapping discontinuity is 10 times smaller. The average value was 0.2, or a five-time improvement.

I tested to see if there was a correlation with slope and found none. Figure 13 shows a "typical" scatter plot of slope vs. discontinuity. In all cases, there was no significant correlation.

To illustrate the idea further, Figure 14 shows a portion from the center of the DEM interpolated
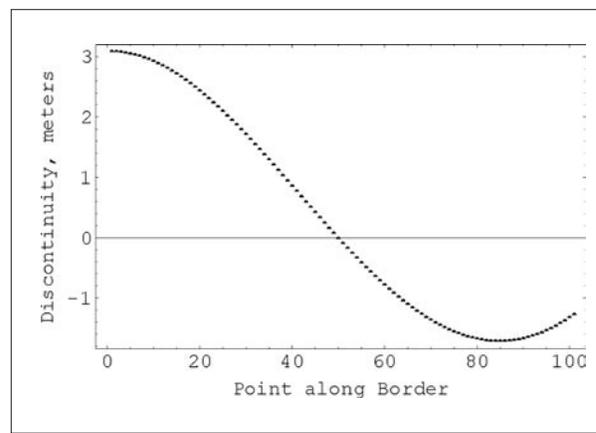
Figure 10. The difference of the two interpolated borders. Any value not equal to zero indicates a discontinuity.
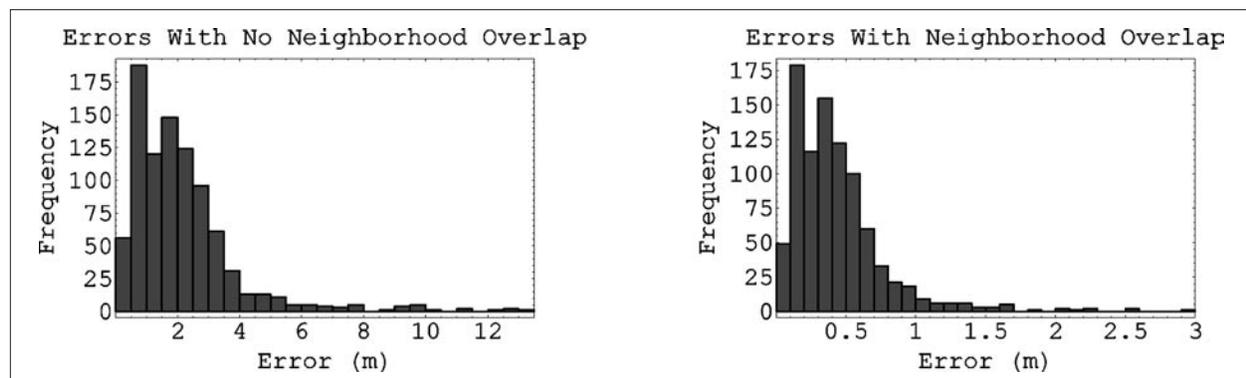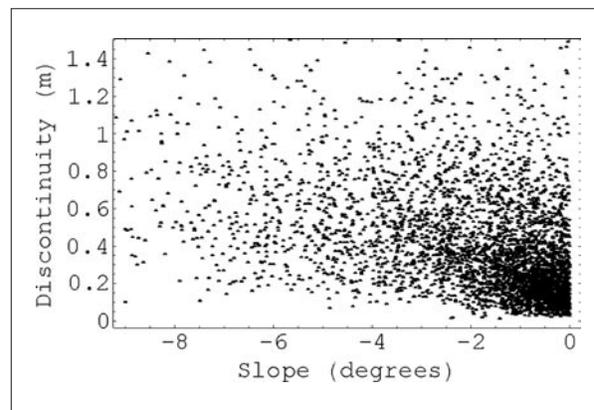

Figure 11. Histograms of the worst discontinuities among the 900 regions. The histogram on the left shows discontinuities for the disjoint kriging neighborhoods. The histogram on the right shows discontinuities for the overlapping kriging neighborhoods.
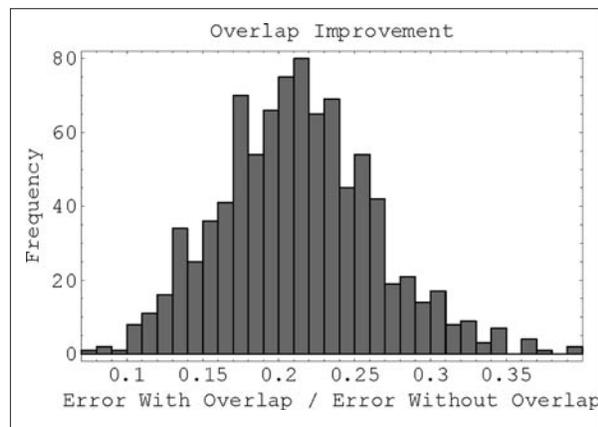
Figure 12. Histograms of the improvement realized by having larger, overlapping neighborhoods. The value for the ordinate is a ratio of discontinuity magnitudes for the two neighborhood sizes:(disjoint/overlapping). The average value of 0.2 indicates a five-times improvement.

Figure 13. Scatter plot of slope vs. magnitude of discontinuity for one region.

twice. The two left-most pictures were created using non-overlapping neighborhoods, whereas the two right-most pictures were created using neighbor-hoods that overlap by one. The eye easily sees the discontinuities in the top left image, and its contour plot in the bottom left exposes them clearly, as well. The discontinuities are still visible in the top right picture, although they are less pronounced than its counterpart. There is far less evidence of the discontinuities in its contour plot in the bottom right.

Discussion

In every case kriging performed best when the estimate resulted from overlapping neighborhoods. This is not surprising because, if the neighborhoods share only one point, then their common border is on the convex hull of the Voronoi polygon defined by that neighborhood. Consequently, estimates along such a border essentially constitute extrapolation.

An examination of the worst discontinuities revealed that the two neighborhoods causing these discontinuities were very dissimilar in shape. The topography for one neighborhood was typically essentially planar and the other was very rough. Because the two neighborhoods have only one point in common, they share almost no information. Therefore, any extrapolation done from the planar neighborhood carries that assumption with it: the estimates produced by the planar region are planar, meaning they assume that the terrain continues on in a plane forever. Similarly, the rough shape of the other neighborhood is carried outwards in its idea of what the terrain is like. Therefore, where the two met, there was a large discontinuity between the linear extrapolation and the rough extrapolation.

In contrast, for the neighborhoods that overlapped by more than one point, both neighborhoods had much more common information about their overlapping area and the estimations, being well within the convex hulls, were interpolated instead of extrapolated. Consequently, the discontinuities were smaller, sometimes ten times smaller. Even so, although mitigated, all borders had discontinuities, if only small ones.

It is natural to wonder whether certain types of terrain are more problematic than others; perhaps one should expect flat lands to exhibit smaller discontinuities than alluvial fans or cliffs. This is certainly true in the sense that no discontinuity can exceed the size of local relief. However, as shown in the scatter plot (Figure 13), many borders in steeply sloped regions had small discontinuities compared to borders in relatively flat regions, and vice versa. Slope is thus not a factor in the sense that there was no correlation between slope and the maximum discontinuity.

These results suggest that roughness plays a part but, in itself, does not cause large discontinuities. Rather, the large discontinuities seem to occur at the junction between two terrain types of dissimilar roughness: for example, where rugged mountain ridges intersected relatively flat talus slopes. This is a discontinuity of the second derivative of the theoretical terrain surface, an interpretation that fits well into kriging's mathematical framework because kriging is smooth. Fitting a smooth surface to a discontinuous surface ought not to work well at the discontinuity, and this is what we observed, that the large discontinuities occurred as the apparent junction of disparate terrain types. This also fits well with the quote from Lam (1983) given above.

The source of the large discontinuities can, however, be seen in a different light—as an under-sampling issue. Had the terrain been sampled adaptively (Makarovic 1973; 1977; 1979; 1984; Carter 1988), the transition probably would have not been abrupt and the large discontinuities would not have occurred. I emphasize that the discontinuities will always be present in general, but adaptive sampling should reduce the occurrence of large discontinuities.

Conclusions and Summary

This study has documented that kriging produces a piece-wise continuous surface and that a kriged surface has zero-order continuity along its patch borders. This means that a kriged surface has "rips" or "tears" as shown in Figure 14. The discontinuities are a consequence of the mathematical formulation of kriging and are not a consequence of the choice of the data set; the result is general. It is common knowledge among those with a background in computational geometry that higher-order surface patches will not, in general, enjoy along-patch continuity unless they have been specifically formulated to do so. Kriging was not developed to be used piece-wise and, as a result, is lacking the mathematical constraints to enforce continuity along the borders.

The ramifications of this discontinuity depend largely upon the needs of the user. Discontinuities such as the ones presented in Figure 14 will be clearly visible on high-accuracy topographic maps. Construction maps frequently have one-foot contour lines, and discontinuities measured in meters will be significant. In contrast, the data for the Sandia Crest DEM were created from a topographic map with 40-foot contour lines. The overlapping neighborhood discontinuities shown in Figures 11 and 14 would generally not compromise such a map's conformance with the National Map Accuracy Standards and, therefore, could be ignored in its compilation.
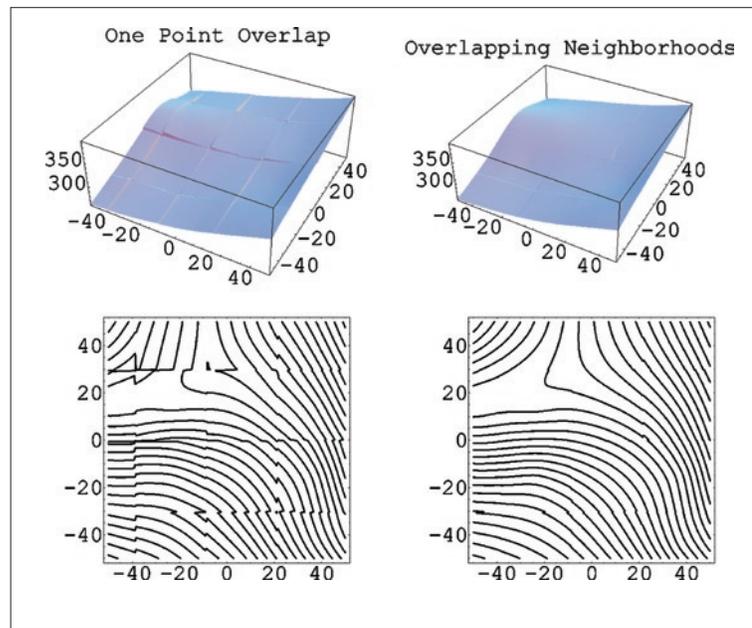
The conclusions of this study are thus two-fold. First, the experimentation documents the discontinuous property of kriging surfaces that has not been studied in the literature. Second, it illustrates a potential pitfall for practitioners who might have otherwise unknowingly assumed that kriging produces globally smooth surfaces, as are common for the production of contour lines.

REFERENCE LIST

Almansa, A., F. Cao, Y. Gousseau, and B. Rouge. 2002. Interpolation of digital elevation models using AMLE and related methods. *IEEE Transactions on Geoscience and Remote Sensing* 40 (2): 314-25.

Bailey, T.C. 1994. A review of statistical spatial analysis in geographical information systems. In: S. Fotheringham and P. Rogerson (eds), *Spatial analysis and GIS*. London, U.K.: Taylor & Francis. pp. 13-44.

Carter, J.R. 1988. Digital representations of topographic surfaces. *Photogrammetric Engineering and Remote Sensing* 54 (11): 1577-1580.

David, M. 1977. *Geostatistical ore reserve estimation*. Amsterdam, the Netherlands: Elsevier Scientific Publishing Company. 384p.

de Boor, C. 1978. *A practical guide to splines*. New York, New York: Springer-Verlag. 392p.

Dierckx, P. 1993. *Curve and surface fitting with splines*. Oxford, U.K.: Clarendon Press. 285p.

Farin, G. 1993. *Curves and surfaces for CAGD: A practical guide*, 3rd ed. Boston, Massachusetts: Academic Press, Inc. 473p.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*. New York, New York: Oxford University Press. 483p.

Isaaks, E.H., and R. M. Srivastava. 1989. An introduction to applied geostatistics. New York, New York: Oxford University Press. 561p.

Journal, A.G., and C. J. Huijbregts. 1997. *Mining geostatistics*. London, U.K.: Academic Press. 600p.

Kimeldorf, G., and G. Wahba. 1971. Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis & Applications* 33: 82-95.

Lam, N.S.-N. 1983. *Spatial interpolation methods: A review*. The American Cartographer 10 (2): 129-49.

Lancaster, P., and K. Šalkauskas. 1986. *Curve and surface fitting: An introduction*. New York, New York: Academic Press. 280p.

Laslett, G.M. 1994. Kriging and splines: An empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association* 89 (426): 391-409.

Figure 14. The same region of the DEM interpolated twice with different neighborhoods. The two pictures on the left result from neighborhoods that share a single point, whereas the two pictures on the right result from first extended nearest neighborhoods.

Makarovic, B. 1973. Progressive sampling for digital terrain models. *ITC Journal* 3: 397-416.

Makarovic, B. 1977. Composite sampling for digital terrain models. *ITC Journal* 3: 406-33.

Makarovic, B. 1979. From progressive sampling to composite sampling for digital terrain models. *Geo-Processing* 1: 145-66.

Makarovic, B. 1984. Structures for geo-information and their application in selective sampling for digital terrain models. *ITC Journal* 4: 285-95.

Matheron, G. 1963. *Principles of geostatistics*. Economic Geology 58: 1246-66.

Meyer, T.H., M. Eriksson, and R.C. Maggio. 2001. Gradient estimation from irregularly spaced data sets. *Mathematical Geology* 33 (6): 693-717.

Meyer, T.H. 1999. A conceptual framework for digital terrain modeling. Ph.D. dissertation, Texas A&M University, 110 pp.

Philip, G.M., and D.F. Watson. 1986. Matheronian geostatistics—Quo vadis? *Mathematical Geology* 18 (1): 93-117.

Siška, P., M. Eriksson, and R. Maggio. 1997. Ordinary kriging and the construction of a digital elevation model. *Geografický Časopis* 49: 91-112.

Tobler, W.R. 1970. Computer model simulating urban growth in the Detroit region. *Economic Geography* 46 (2): 234-40.

Wahba, G. 1990. *Spline models for observational data*. Philadelphia, Pennsylvania: SIAM. 169p.

Wolfram, S. 1999. *The Mathematica book*, 4 ed. Cambridge, UK: Cambridge University Press. 1470p.