

The Discontinuous Nature of Kriging Interpolation for Digital Terrain Modeling

Thomas H. Meyer
Department of Natural Resources Management and Engineering
University of Connecticut
Storrs, Connecticut
`thomas.meyer@uconn.edu`

January 21, 2003

Abstract

Kriging is a widely accepted method for interpolating and estimating elevations from digital elevation data. Its place of prominence is due to its elegant theoretical and practical properties. From an interpolation point of view, kriging is equivalent to a thin-plate spline and is just 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 estimator that performs as well as kriging (in the least squares sense) must be equivalent to kriging. Therefore, kriging is held by many to be the gold standard of digital terrain model (DTM) elevation interpolation. Even so, as used for DTM interpolation, kriging has an undesirable property that has not been previously documented. Kriging produces discontinuous surfaces along the boundaries of interpolation patches. This paper documents the situation, explains its source, provides bounds on the discontinuity, and provides real-world examples from a digital elevation model of mountainous terrain in central New Mexico.

Keywords: kriging, discontinuity, digital terrain modeling, spatial interpolation.

INTRODUCTION

Kriging (Matheron, 1963) is a popular technique for interpolating and estimating elevation values from digital terrain data. General references on the subject are (Journal and Huijbregts, 1978; David, 1977; Lam, 1983; Isaaks and Srivastava, 1989; Goovaerts, 1997). Bailey (1994) asserts that “there is an argument for kriging to be adopted as a basic method of surface interpolation in GIS as opposed to the standard deterministic tessellation techniques

which 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 certain splines are “too smooth,” whereas kriging produced more precise estimations. Declercq (1996) compared polynomials, splines, linear triangulation, proximation, distance weighting, and kriging to test their efficacy to visualize spatial patterns in addition to their performance in predicting unvisited sample locations. Kriging performed among the best in both categories. Katzil and Doytsher (2000) found third-order polynomials to perform comparably with kriging but they did not have either the need to compute a variogram or kriging’s complicated matrix computations. Yang and Hodler (2000) found that kriging outperformed four other interpolation techniques in preserving the visual character of digital terrain models. 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. Almansa and others (2002) established kriging’s place in a more general class of functions, namely, the absolutely minimizing Lipschitz extension. Although 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 surface model wisely to properly match the properties of the surface with the desired traits of the terrain. Continuity properties are of paramount importance. Discontinuous surfaces have “holes” or “tears” in them, so to speak. A continuous surfaces might not be smooth, meaning that the surface might have “creases” in it, such as with a triangulated irregular network (TIN). It is important to catalogue surface model continuity properties and this paper establishes these properties for kriging interpolation.

Surface Models

In their seminal work, Miller and LaFlamme (1958) suggested, “The digital terrain model (DTM) is simply a statistical representation of the continuous surface of the ground by a large number of selected points with known xyz coordinates in an arbitrary coordinate field.” A DTM was envisioned to be a set of height samples, that is to say, a set of points. We will denote this set of sample locations by $s = \{(x_i, y_i)\}$ and denote the set of heights measured at those locations by $h = \{z(x_i, y_i) \mid (x_i, y_i) \in s\}$. Miller goes on to say, “Just as the engineer must interpolate on the topographic map, the computer will have to interpolate with the DTM.” The mathematical scheme used for this interpolation is what we call a **surface model**. We formally define a surface model to be a real-valued function of two variables, $f : \mathbf{p} \in \mathcal{R}^2 \rightarrow \mathcal{R}$, or $z = f(x, y)$, where \mathcal{R} denotes the set of reals.

Some surface models 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. Global support imposes relatively heavy computational burdens, especially for large data sets. Also, it has the counter-intuitive property that, for certain techniques such as 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, that everything is related to everything, but closer things are more related. Also, polynomial global support surface models that 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.

Although kriging is defined with global support, in practice, it is not typically used that way for terrain modeling. Instead, s is subdivided into **neighborhoods**, being subsets of s with relatively few elements that roughly (or strictly) partition s . Then, kriging is used to interpolate over these neighborhoods in the following way. Suppose we want to interpolate a surface value at the point $\mathbf{p} = (x, y)$ and $\mathbf{p} \notin s$. Let n_j denote a neighborhood containing \mathbf{p} . Then, the height estimate at \mathbf{p} is a weighted sum of the heights of the points in n_j . The weights are related inversely to the distance from the sample to \mathbf{p} in such a way that minimizes the variance of the estimate. The surface produced over a neighborhood is called a **patch**.

There are several heuristics for choosing neighborhoods for a point of interest (e.g., Isaaks and Srivastava, 1989, p. 338). Some of the heuristics include using all samples within some circle or ellipse enclosing \mathbf{p} , using all samples within some convex polygon enclosing \mathbf{p} , using a limited number of samples from the four quadrants enclosing \mathbf{p} , or using Voronoi nearest neighbors.

Continuity

Although the physical surface of the Earth is arguably not mathematically continuous (Meyer, 1999), it is generally modeled as if it were. The Earth's surface is seen as being piece-wise continuous, meaning that, on the whole, the surface is continuous but it is possible that there are local discontinuities such as knick points and cliffs. Consequently, piece-wise continuity is commonly considered a desirable property in terrain surface models.

Any surface model that is based on polynomials (i.e., not a fractal) will be smooth within a patch. This is a direct consequence of the infinite differentiability of polynomials. However, the piece-wise continuity of surface patches cannot be taken for granted. There is a huge body of literature in the Computer Aided Geometric Design (CAGD) community devoted to creating piece-wise continuous surfaces of various orders with first and second order being the most common (e.g., see Lancaster and Šalkauskas (1986); Farin (1993, 1995); Dierckx (1995)). These patches are smooth in their interiors

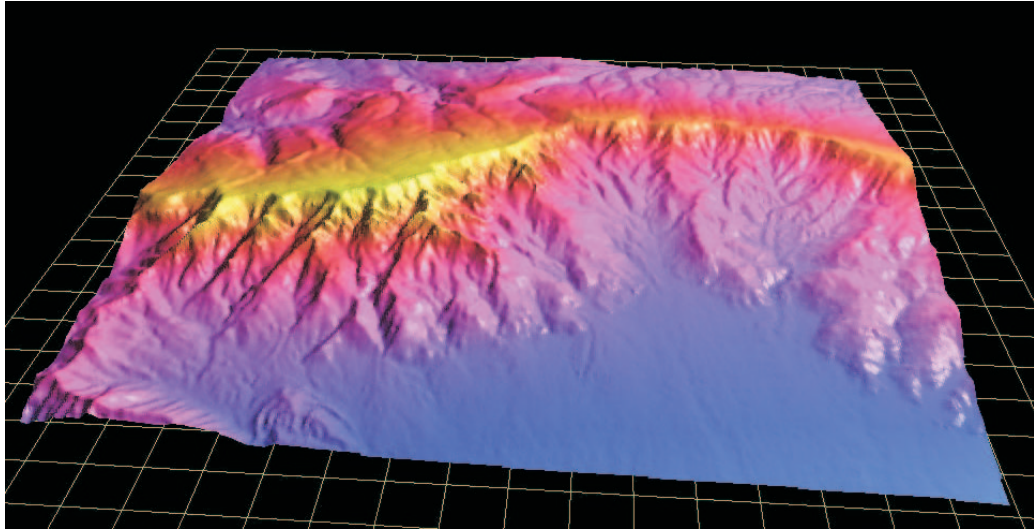


Figure 1: DEM of the Sandia Mountains in north-central New Mexico.

and form, say, a first-order continuous surface on the whole. The remainder of the paper will document that kriging creates a zero-order, piecewise continuous surface model.

Kriging Continuity

The equivalence of kriging and thin-plate splines guarantees that, within a patch, kriging produces a continuous surface. The question then becomes what continuity can be expected along the border between two neighborhoods?

Claim: When used with local support, kriging is in the class of piecewise, zero-order continuous surfaces.

Proof: To establish the claim, it suffices to produce a single example. The proof proceeds as follows.

1. Choose a sample height data set.
2. Subdivide the data set into neighborhoods.
3. Select a border between two abutting neighborhoods.
4. Interpolate the border twice, once for each neighborhood.
5. If the two interpolated borders differ in elevation, then kriging is discontinuous along the border.

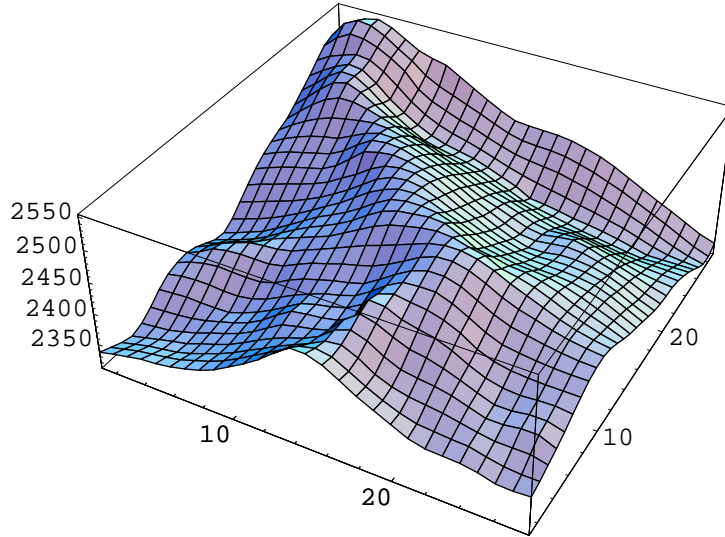


Figure 2: The topographic data set.

The data set and its covariance function

The topographic data set comes from the the Sandia Mountains of north-central New Mexico, see Fig. 1. This DEM was chosen because of the wide variety of topography within a single DEM, from flat tails of alluvial planes to sheer cliffs. These mountains, although quite rugged, are not remote in the sense that Albuquerque, New Mexico, is built up to and around them. Therefore, surveyors can encounter terrain of this type in their work.

The data for this proof comes from the foothills of these mountains, see Fig. 2. The site is located in UTM zone 13 with corner coordinates (e364140, n3900690) and (e364980, n3901590) and measures 840 meters east to west and 810 meters north to south. The corners and extent of the site were chosen to coincide with elevations available in the USGS Sandia Crest 7½ minute digital elevation model of the area.

The omni-directional variogram of the data set was computed using relational database queries as described in (Maggio and others, 1997). The result is shown in Fig. 3. 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. The variogram was deemed to be Gaussian and, thus, be of the form

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

where h is the lag distance and a is the practical range. In this case, a was

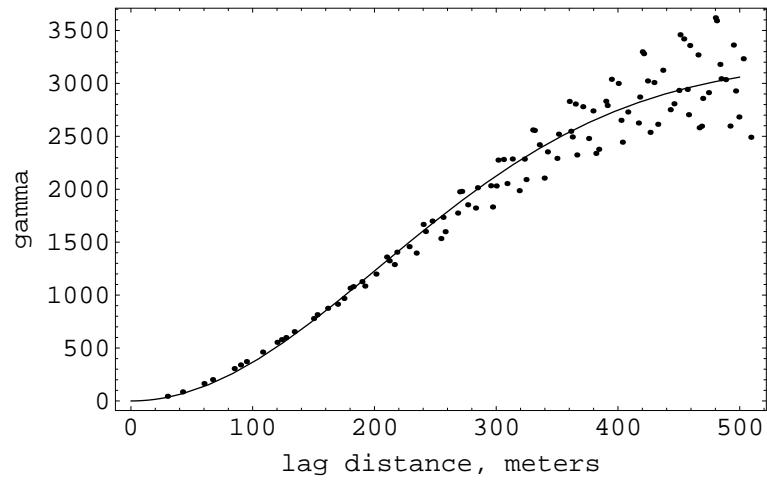


Figure 3: The omni-directional variogram and its least squares model.

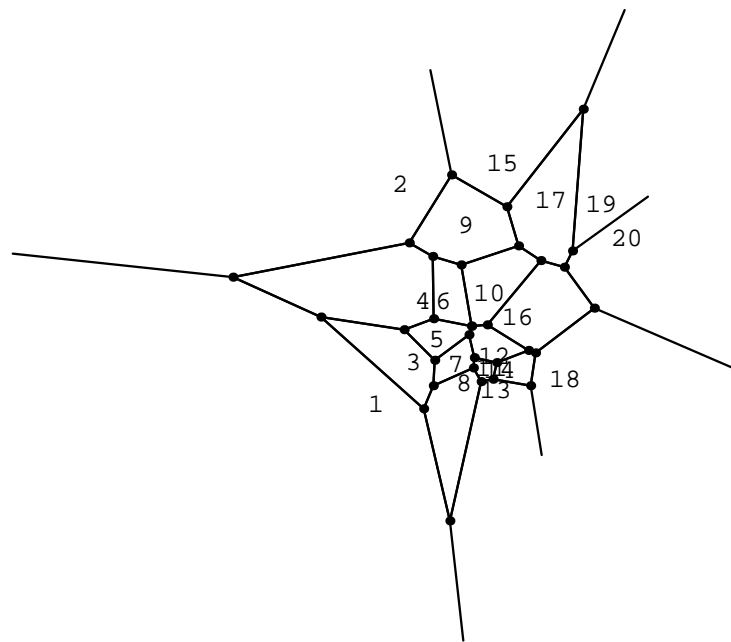


Figure 4: The Voronoi diagram for the study area. Individual sample points are indicated by numbers. The polygons are subregions of R such that every point in a polygon corresponds to the same interpolation neighborhood.

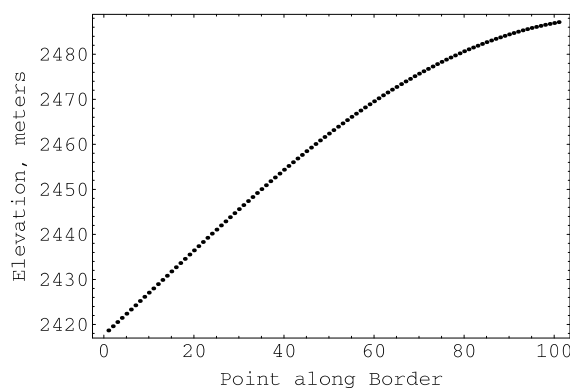


Figure 5: The border as interpolated with the nearest neighbors of point 6.

chosen to be 500 meters. A least squares fit yielded the model

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

The model and the data are depicted in Fig. 3.

By assuming stationarity, the covariance function is related to the variogram by

$$\text{cov}(h) = \sigma^2 - \gamma(h), \quad (1)$$

where σ^2 is the variance of the data set and has a value of 2352.829 m².

Patches

Having created a covariance function with which to kriging, the next task was to tessellate the region into neighborhoods. It was decided to create neighborhoods using Voronoi nearest neighbors. This choice is arbitrary; any other neighborhood scheme would confirm the claim. Twenty *easting*, *northing* pairs were generated randomly from a uniform distribution. These pairs were constrained to fall within the study area. Figure 4 shows the points and the Voronoi diagram of the points.

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 defined by kriging the surface at 100 points distributed evenly between these two Voronoi vertices. The ordinary kriging was done using Mathematica v3.0 by constructing the matrices and

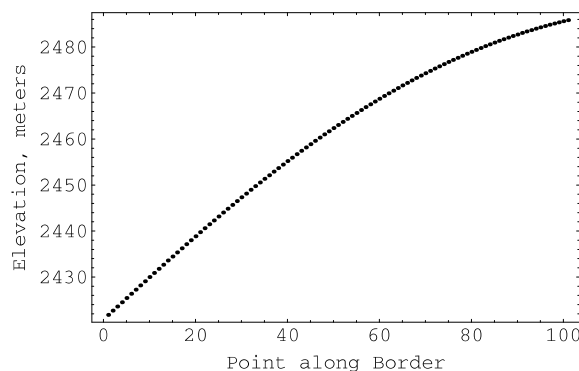


Figure 6: The border as interpolated with the nearest neighbors of point 10.

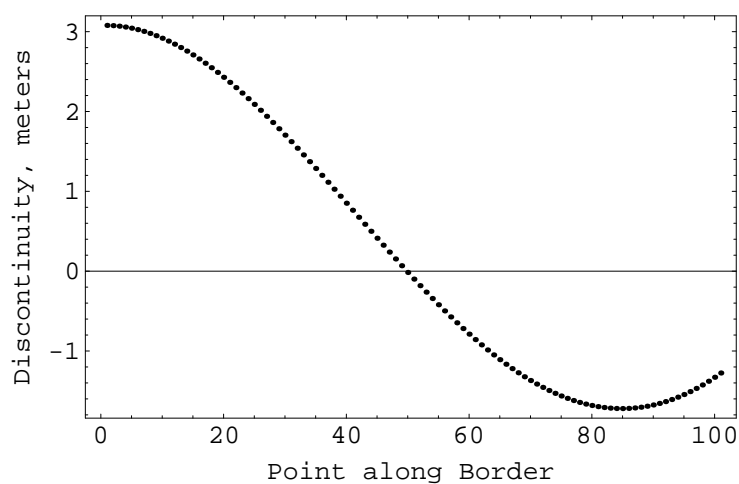


Figure 7: The difference of the two interpolated borders. Any value not equal to zero indicates a discontinuity.

computing their inverses. However, the other types (universal kriging, kriging with a trend, block kriging, and cokriging) have the same basic structure and the argument applies equally to them, as well.

The results are shown in Fig. 5 and 6. The visual similarity of the two borders confirms that the interpolation is working correctly; one could expect them to be very similar. However, the difference of the interpolated values shown in Fig. 7 clearly depicts the discontinuity. The discontinuity ranges in value from 3.08 meters to -1.72 meters. This completes the proof.

Causes

The source of the discontinuity comes from the observation that the border is being interpolated via two different covariance matrices. The matrices

Gap (m)	Count
0-1	206939
1-2	102100
2-3	13383
3-4	3449
4-5	1047
5-6	450
6-7	228
7-8	132
8-9	85
9-10	49
10-11	40
11-12	20
12-13	15
13-14	17
14-15	5
15-16	8
16-17	2
17-18	1
18-19	1
19-20	1

Table 1: Distribution of Discontinuity Magnitude (m).

are different because the interpolation neighborhoods are different. If the samples are not reflective symmetric, then there is no reason to expect the patches to match on their borders. Therefore, this is a completely general result; it is not specific to the chosen data set.

This property of polynomial surfaces is essentially why CAGD exists. Global support has long been known to be undesirable for reasons listed above. These drawbacks lead to using piecewise curves and surfaces that have local support. However, it is usually desirable to have surface-wide continuity of first or second order. If the piecewise curves are not constructed to be continuous, in general, they will not be (de Boor, 1978). The central theme of CAGD is to develop piecewise curves and surfaces that enjoy continuity across their borders, in addition to other properties as well.

Examination of a DEM

To explore the range of possible discontinuities further, the entire Sandia Crest DEM was analyzed in the manner described above. That is to say, random points were generated to cover the entire DEM and the borders between all adjacent neighborhoods were interpolated. This resulted in the analysis of 327972 borders. 94.2% of the borders had a maximum absolute discontinuity (gap) of one meter or less. The maximum gap observed was 19.48 meters. The average maximum gap was 0.54 meters with a standard deviation of 0.62 meters. See Table 1.

We observed a correlation between slope and increasing discontinuity size so we grouped the tests by slope and created histograms of the discontinuity size. Slope was computed using a technique specifically designed for irregularly spaced data (Meyer and others, 2001). The gaps were grouped by slope. Table 2 shows the distributions of the gaps for four particular slope values. For 0° , meaning flat ground for both neighborhoods, 86% of the gaps are between zero and one meter in magnitude, with the overall shape of the distribution appearing to be a decreasing exponential. Interestingly, there was a single instance in which flat ground produced a nine meter gap. We next show the results for the next slope group, those less than 3.75° . Again, the distribution appearing to be a decreasing exponential. We skip to the results for 49.7° . Notice that the distribution now appears to be lognormal with a mean around 3 meters. Finally, the distribution for the steepest sloped neighborhoods has very few samples but appears to be roughly uniform.

Conclusions and Discussion

We note several trends in the gap distributions. Overall, the distributions of the gaps appear to be lognormal. Next, the average gap size steadily increases with slope. Third, the variability of gap size increases with slope.

(m)	0°	3.75°	49.7°	52.7°
0	1901	38765	1	
1	288	3644	7	
2	9	89	18	2
3	5	12	13	1
4		4	14	2
5		2	12	4
6			7	1
7			7	
8			3	
9	1		4	
10			3	
11			0	
12			2	
13				
14				
15			3	
16				
17			1	
18				
19				
20			1	

Table 2: Slope (degrees) vs. Discontinuity Magnitude (m).

This suggests that kriging with a trend might perform better than the ordinary kriging used in this study because kriging with a trend removes trends (slope) in the data, which can arguably be violating the stationarity assumption upon which kriging is based. This is a subject for future research.

The ramifications of this discontinuity depend largely upon the needs of the user. Discontinuities such as the one presently above will be clearly visible on high-accuracy topographic maps. Survey maps frequently have one-foot contour lines and discrepancies of three meters will be significant. In contrast, these data were taken from a digital elevation model developed from a map with 40 foot contour lines. These discontinuities do not compromise such a map's conformance with the National Map Accuracy Standards and, therefore, could be ignored in its compilation.

Acknowledgements

This work was funded by the Connecticut Institute of Water Resources in cooperation with the State Water Resources Research Institute Program of the U.S. Geological Survey under USGS award number 01HQGR0078, as authorized by the Water Resources Research Act of 1984 (P.L. 98-242), as amended.

References

- Almansa, Andrés, Cao, Frédéric, Gousseau, Yann, and Rougé, Bernard, 2002, Interpolation of Digital Elevation Models Using AMLE and Related Methods: *IEEE Transactions on Geoscience and Remote Sensing*, v. 40, no. 2, pp. 314–325.
- Bailey, Trevor C., 1994, A review of statistical spatial analysis in geographical information systems: *in* Fotheringham, Stewart and Rogerson, Peter (eds.), *Spatial Analysis and GIS*, Taylor & Francis, London, pp. 13–44.
- David, M., 1977, *Geostatistical Ore Reserve Estimation*: Elsevier Scientific Publishing Company, Amsterdam.
- de Boor, Carl, 1978, *A Practical Guide to Splines*: Springer-Verlag, New York.
- Declercq, Franky Albert Noël, 1996, Interpolation Methods For Scattered Sample Data: Accuracy, Spatial Patterns, Processing Time: *Cartography and Geographic Information Systems*, v. 23, no. 3, pp. 128–144.
- Dierckx, Paul, 1995, *Curve and Surface Fitting with Splines*: Oxford Science Publications, Oxford.
- Farin, Gerald, 1993, *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide*, 3rd ed.: Academic Press, New York.

- Farin, Gerald, 1995, *NURB Curves and Surfaces: From Projective Geometry to Practical Use*: A K Peters, Wellesley, MA.
- Goovaerts, Pierre, 1997, *Geostatistics for Natural Resources Evaluation*: Oxford University Press, New York.
- Isaaks, Edward H and Srivastava, R. Mohan, 1989, *An Introduction to Applied Geostatistics*: Oxford University Press, New York.
- Journal, A. G. and Huijbregts, C. J., 1978, *Mining Geostatistics*: Academic Press, London.
- Katzil, Y. and Doytsher, Y., 2000, Height Estimation Methods for Filling Gaps in Gridded DTM: *Journal of Surveying Engineering*, v. 126, no. 4, pp. 145–162.
- Kimeldorf, G. and Wahba, G., 1971, Some results on Tchebycheffian spline functions: *Journal of Mathematical Analysis & Applications*, v. 33, pp. 82–95.
- Lam, Nina S., 1983, Spatial Interpolation Methods: A Review: *The American Cartographer*, v. 10, no. 2, pp. 129–149.
- Lancaster, Peter and Šalkauskas, Kęstutis, 1986, *Curve and Surface Fitting: An Introduction*: Academic Press, New York.
- Laslett, G. M., 1994, Kriging and Splines: An emperical comparison of their predictive performance in some applications: *Journal of the American Statistical Association*, v. 89, no. 426, pp. 391–409.
- Maggio, Robert C., Meyer, Thomas H., and Siska, Peter P., 1997, Spatial education using relational databases: *in Proceedings of GIS/LIS '97*, Cincinnati, Ohio, pp. 432–440.
- Matheron, G., 1963, Principles of geostatistics: *Economic Geology*, v. 58, pp. 1246–1266.
- Meyer, Thomas H., Eriksson, Marian, and Maggio, Robert C., 2001, Gradient Estimation from Irregularly Spaced Data Sets: *Mathematical Geology*, v. 33, no. 6, pp. 693–717.
- Meyer, Thomas Henry, 1999, *A Conceptual Framework for Digital Terrain Modeling*: Ph.D. thesis, Texas A&M University, Department of Forest Science, College Station, Texas.
- Miller, C. L. and LaFlamme, R. A., 1958, The Digital Terrain Model - Theory and Application: *Photogrammetric Engineering*, v. 24, no. 7, pp. 433–442.
- Philip, G. M. and Watson, D. F., 1986, Matheronian Geostatistics - Quo Vadis?: *Mathematical Geology*, v. 18, no. 1, pp. 93–117.

Wahba, G., 1990, *Spline Models for Observational Data*: Society for Industrial and Applied Mathematics, Philadelphia.

Yang, X. and Hodler, T., 2000, Visual and Statistical Comparisons of Surface Modeling Techniques for Point-based Environmental Data: *Cartography and Geographic Information Science*, v. 27, no. 2, pp. 165–176.