Earth Surface Processes and Landforms
Earth Surf. Process. Landforms **30**, 601–612 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/esp.1164

Predicting ordinary kriging errors caused by surface roughness and dissectivity

Peter P. Siska[1]*, Pierre Goovaerts[2], I-Kuai Hung[3] and Vaughn M. Bryant[4]
[1] *School of Agriculture and Geosciences, PO Box 4418, Austin Peay State University, Clarksville, TN 37044, USA*
[2] *Biomedware, 516 North State Street, Ann Arbor, MI 48104, USA*
[3] *College of Forestry, Stephen F. Austin State University, Nacogdoches, TX 75962-6109, USA*
[4] *Palynology Laboratory, Department of Anthropology (TAMU 4352), Texas A&M University, College Station, Texas 77843-4352, USA*

**Correspondence to: P. Siska, School of Agriculture and Geosciences, PO Box 4418, Austin Peay State University, Clarksville, TN 37044, USA. E-mail: siskap@apsu.edu*

Abstract

The magnitude of kriging errors varies in accordance with the surface properties. The purpose of this paper is to determine the association of ordinary kriging (OK) estimated errors with the local variability of surface roughness, and to analyse the suitability of probabilistic models for predicting the magnitude of OK errors from surface parameters. This task includes determining the terrain parameters in order to explain the variation in the magnitude of OK errors. The results of this research indicate that the higher order regression models, with complex interaction terms, were able to explain 95 per cent of the variation in the OK error magnitude using the least number of predictors. In addition, the results underscore the importance of the role of the local diversity of relief properties in increasing or decreasing the magnitude of interpolation errors. The newly developed dissectivity parameters provide useful information for terrain analysis. Our study also provides constructive guides to understanding the local variation of interpolation errors and their dependence on surface dissectivity. Copyright © 2005 John Wiley & Sons, Ltd.

Received 22 July 2003;
Revised 29 June 2004;
Accepted 7 September 2004

Keywords: ordinary kriging; surface roughness and dissectivity; surface parameters; multiple regression

Introduction

Geostatistics provides a set of probabilistic methods designed for modelling spatial patterns and estimating values in unsampled locations. The ordinary kriging interpolator is one of the most frequently used geostatistical estimators currently available in several software packages including geographic information systems (GIS). The construction of continuous surfaces including the digital elevation and terrain models (DEM, DTM) can be rapidly performed in GIS using the ordinary kriging interpolator. Due to the frequent variation of local means in the recording of Earth science data, the assumption of the second order stationary is often too restrictive. However, ordinary kriging allows one to account for such local variations of the mean by limiting the domain of the mean stationarity to the local neighbourhood $N(u)$, which is centred on the location 'u' that is being estimated (Goovaerts, 1997). In addition, ordinary kriging yields fairly accurate results when compared to other geostatistical interpolators or non-linear methods (Lloyd and Atkinson, 2001; Moyeed and Papritz, 2002). The purpose of this study is to investigate the dependency of elevation interpolation errors on terrain properties. This relationship is examined under 'ordinary' conditions (i.e., the data input is represented by a set of widely accessible internet elevation data that may not always comply with the assumptions of independence), identity, and the random distribution of samples. An expected change in the magnitude of the ordinary kriging estimation errors is modelled as a function of explanatory variables describing the surface dissectivity in planar (low diversity) and mountain (high diversity) surfaces. The key term used throughout this paper is the 'error' term. It should be noted here that various error terms are used in spatial analysis and their usage often becomes confusing for readers. For example, Lloyd and Atkinson (2001) use the terms 'observed standard error, the actual errors of estimation, the estimation errors, the standard errors of ordinary kriging (OK) estimates, and the kriging standard error' to describe basically only the two types of errors: (a) the model inherent error, and (b) the estimation errors. Isaaks and Srivastava (1989) use 'error variance' in the place of 'kriging variance', and additional terms are used such as ‘actual error variance, global error variance, predicted error variance, observed error variance, absolute kriging error, global and local estimation errors, kriging standard error, and so forth’, which are either synonyms of the two errors mentioned earlier or are overlapping terms that thus confuse the reader. For the purpose of consistency, in this paper we focus on two major error groups:

(a) an estimation (prediction) error that arises through the use of interpolation procedures, defined as a difference between the actual (true) value and the estimated (predicted) value;
(b) the kriging variance or kriging standard error (the square root of kriging variance) that is inherent to the kriging system and is available for every interpolated value. Its mathematical form is:

$$\sigma^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) - \mu(\mathbf{u}) \tag{1}$$

where $C(0)$ corresponds to the covariance value when the lag distance (sample separation) is zero, λ represents the kriging weights, $C(\mathbf{u}_\alpha - \mathbf{u})$ is the covariance between the sample values and the point being predicted, and $\mu(\mathbf{u})$ is a Lagrange multiplier that results from minimizing the estimation variance subject to the unbiasedness constraint on the estimator (Goovaerts, 1997).

For practical reasons the square root of the kriging variance, the kriging standard error, will be used in this paper. Two measures of the first type of error (i.e., estimation or prediction error) are used in this study:

$$E_1 = \sqrt{\frac{\sum_{i=1}^{N} (e_i - \bar{e})^2}{N - 1}} \tag{2}$$

where e_i is an interpolation error at location i, $\bar{e}$ is the mean error for a moving window, and N is the number of observations in each moving window.

$$E_2 = \frac{1}{N}\sum_{i=1}^{N} |o_i - \hat{o}_i| = \frac{1}{N}\sum_{i=1}^{N} |e_i|, \tag{3}$$

where o_i stands for the observed elevation at location i, $\hat{o}_i$ is an interpolated value at the same location, and N is the number of sample points in the moving window.

In the last two decades the errors, their magnitude, and distribution in spatial digital databases have become a major concern for scientists and practitioners (Goodchild, 1994). Some researchers investigated and compared interpolation errors produced by different algorithms and then made recommendations regarding the uncertainty of estimates (Goovaerts, 2000, 2001; Lloyd and Atkinson, 2001; Siska and Hung, 2001; Kastens and Staggenborg, 2002).

A special category of interpolation errors is associated with the Earth’s surface modelling. The digital elevation (terrain) models have been one of the most common representations of the Earth’s surface and interpolation methods have been heavily used in their production (Miller and La Flamme, 1958; Cooper *et al*., 1987; Skidmore, 1989; Giles *et al*., 1994; Eklundh and Martensson, 1995; Siska *et al*., 1997). Therefore, the objective of our paper is to address the fundamental questions of interpolation accuracy with regard to the variation of terrain properties. For example, the increase in surface roughness and dissectivity is accompanied by a greater magnitude of changes in morphometric parameters and landscape processes (Krcho, 1991). Therefore, we believe it is appropriate to investigate the influence of the Earth’s relief roughness–dissectivity on the magnitude of interpolation errors and then search for a pattern in this relationship that could be modelled and predicted. In an abstract sense and for application purposes, surfaces can be generated for any of a large variety of mapping applications including such aspects as geological formations, topography, demography, financial investments, education resources, growth and yield estimates, the probability of toxic metals, etc. Because there is no perfect algorithm, the goal of spatial analysts is to create representations of continuous phenomena with minimum errors and to evaluate the uncertainty of the interpolation. Thus, the understanding of errors, especially in regard to their magnitude, local variability, and dependence on surrounding surface parameters, is an important aspect of surface analysis. For this paper we developed several multiple regression models to predict the magnitude of OK estimated errors and then to evaluate the relationship between surface parameters and the local variation of errors. One important objective of this study is to determine how well the magnitude of OK estimated errors is associated with the local variability of terrain properties. For example, an expected change in the

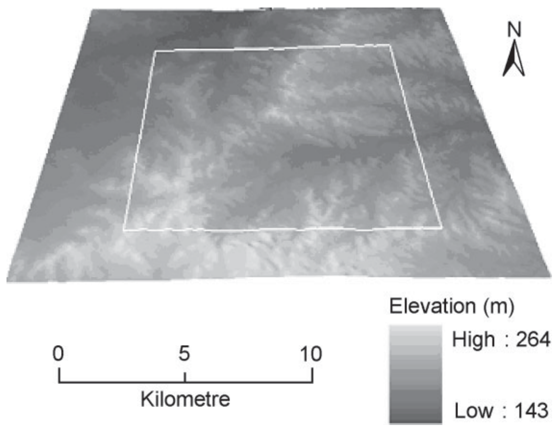

Figure 1. Low (planar) surface.

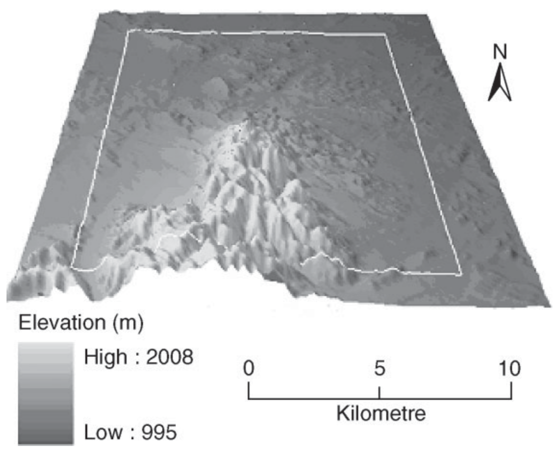

Figure 2. High (mountain) surface.

magnitude of OK estimated errors is modelled as a function of explanatory variables that describe surface roughness–dissectivity in planar (Figure 1) and mountain surface (Figure 2) types. Another objective of this study is to point out the need to develop new roughness–dissectivity parameters explaining the variation in the magnitude of errors using the least variance inflation factor. This problem will be addressed in a later research article in which we also plan to examine the problem of multicollinearity. This phenomenon commonly occurs in the regression of Earth science data (Brower and Merriam, 2001).

Definitions, Data and Methods

The exterior of the Earth's barren crust is frequently called 'relief' and consists of the fundamental properties of roughness and dissectivity. Both terms are closely related reflecting the essential characteristics of the Earth's surface. The roughness usually refers to the irregular shape of the Earth's surface while dissectivity emphasizes the particular relief 'cut' at a specific location. Dissectivity is derived from the Latin words *dissectum* and *dissecare* meaning cut or cutting apart. Analogously, the terrain roughness and dissectivity originated through the work of geological processes and fluvial erosion that cut apart the Earth's surface into a variety of geomorphic entities whose principal components are slopes and gradients. Throughout the course of time the encounter between endogenetic (plate tectonic, folding, volcanic activity) and exogenetic (physical and chemical weathering) forces has resulted in an infinite variety of landforms that are too intricate to be expressed using only one united classification system. Serious attempts have been made to describe the Earth's relief quantitatively by Chorley (1972), Mark (1975), Phillips and Watson (1986), Krcho (1991), Ritter *et al*. (2001), and Schmidt *et al*. (2003).

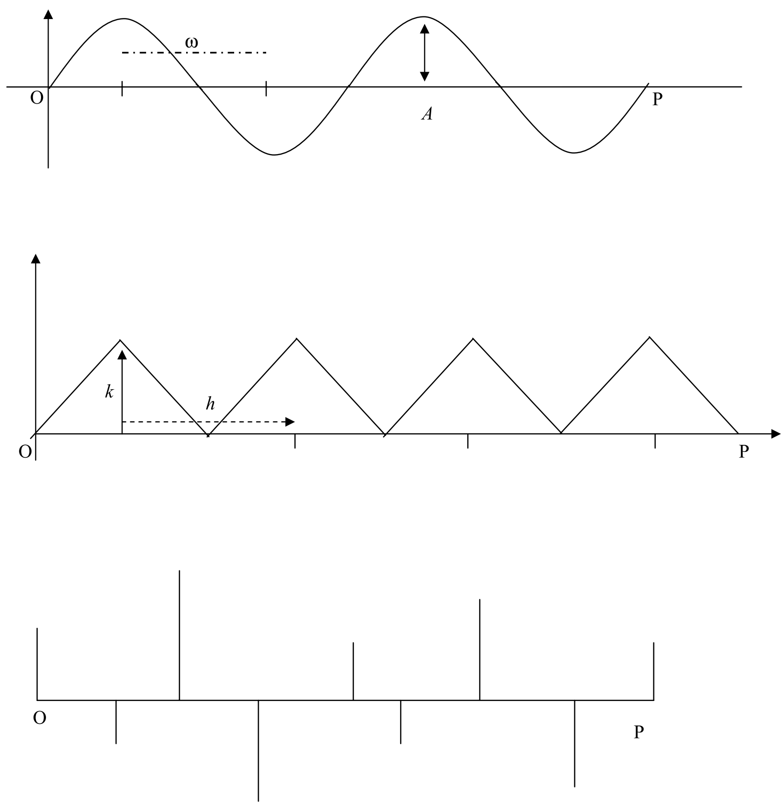

Figure 3. Theoretical and practical concepts of dissectivity. Top figure: roughness and dissectivity as a functional relationship, $f(u) = A \sin \omega u$; middle figure: $f(u) = (u - h) + k$; bottom figure: computation of *Vh* parameter as a ratio between the sum of all vertical differences and horizontal distances among the smple points.

The concept of surface, or terrain, dissectivity builds on the idea of topographic wavelength. Similarly, the terms 'texture' and 'grain' are defined as the significant wavelengths of topography. Indeed, the transcendental functions, such as the sinusoidal functions, could ideally represent the terrain dissectivity. The crests and valleys would correspond to the upper and lower segments of periodically repeated sine or cosine functions $f(u) = A \sin \omega u$, where A is the amplitude of a vertical stretch or flattening and ω is the horizontal stretch or compression. The dissectivity parameter is then defined as the ratio between the length of a sinusoidal function between the two points O and P on the surface, and the shortest horizontal distance between points O and P. The relationship between dissectivity and amplitude is positive (i.e., the dissectivity value increases with the vertical stretch or amplitude). In contrast, it decreases with the increasing ω value. This function can be modified to a more realistic form: $f(u) = f(u - h) + k$, where h is the horizontal shift and k represents the vertical shift (Figure 3, middle part).

Roughness–dissectivity parameters

The roughness is approximated using summary statistics and the roughness index (Phillips and Watson, 1986) in the moving windows environment. The dissectivity is evaluated using three new measures proposed and tested here. These are: dissectivity mean (Equation 4), vertical horizontal ratio (Equation 5), and the dissectivity index (Equation 6). The dissectivity mean (*Dm*) is the sum of the partial slopes inside each moving window divided by the number of non-repeating pairs:

$$Dm = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} |z_i - z_j| d_{ij}^{-1}}{N} \qquad (4)$$

where z_i and z_j are attribute values (elevations) measured at locations i and j, d_{ij} is the horizontal distance between them, m is the number of point data in the moving window and N is the number of non-repeating pairs used to calculate the slope:

$$N = \binom{m}{2} = \frac{m!}{2!(m-2)!}$$

The Vh parameter represents the ratio between the sum of the non-repeatable vertical differences and the sum of the horizontal distances between them:

$$Vh = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} |z_i - z_j|}{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} d_{ij}} \times 100 \tag{5}$$

The dissectivity index is the simplest parameter computed in each window and is the ratio of the difference between the maximum and minimum values in that window over the horizontal distance between them:

$$DI = \frac{|z_{\max} - z_{\min}|}{d_{\max,\min}} \times 100 \tag{6}$$

Data sets

The data consisted of irregularly spaced spot elevation points that were digitized from 7.5′ USGS topographic maps. The irregularly spaced point samples guaranteed that no errors would be introduced into the original data sets prior to the interpolation. Regular grids, such as the DEM, already contain an interpolation bias and therefore were excluded from this analysis. Thus, to accentuate the influence of surface roughness and dissectivity on the magnitude of OK errors, two contrasting landform settings were selected: (1) a low landform with small terrain variability [planar], and (2) a high and diverse landform [mountain]. The elevation data were extracted from target quadrangles, halves of the four directly adjacent quadrangles and quarters of the four diagonally adjacent quadrangles. The target quadrangle for the planar surface was the Piloncillo USGS 7.5′ quadrangle and the mountain data originated from the Gomez Peak 7.5′ USGS quadrangle. The entire area covers more than *c*. 600 km^2. The summary statistics in Table I indicate that there is a much higher dissectivity of the mountain surfaces relative to the planar surface. The actual computation of the parameters was performed in the adjustable moving windows system using the AML program, which was developed for this project.

The optimal size of the moving windows for analysis that were deemed appropriate was a window of 4000 × 4000 m with a 1000 m overlap (Figure 4). These dimensions guaranteed the presence of at least five samples for computing

Table I. Difference in surface roughness–dissectivity parameters: planar (low) and mountain (high) surfaces

Surface parameter (mean value per moving window)	**Planar (low)**	**Mountain (high)**
Elevation (m)	197	1331
Variance of elevations (m^2)	109	6630
Vh (Equation 5)	1	3
Dm (Equation 4)	1	7
DI (Equation 6)	1	4
Roughness mean index	16	980
Standard deviation of roughness	12	863
Range of roughness	43	2670
Variance of roughness	194	217 067
E1 (Equation 2)	5	46
E2 (Equation 3)	4	35

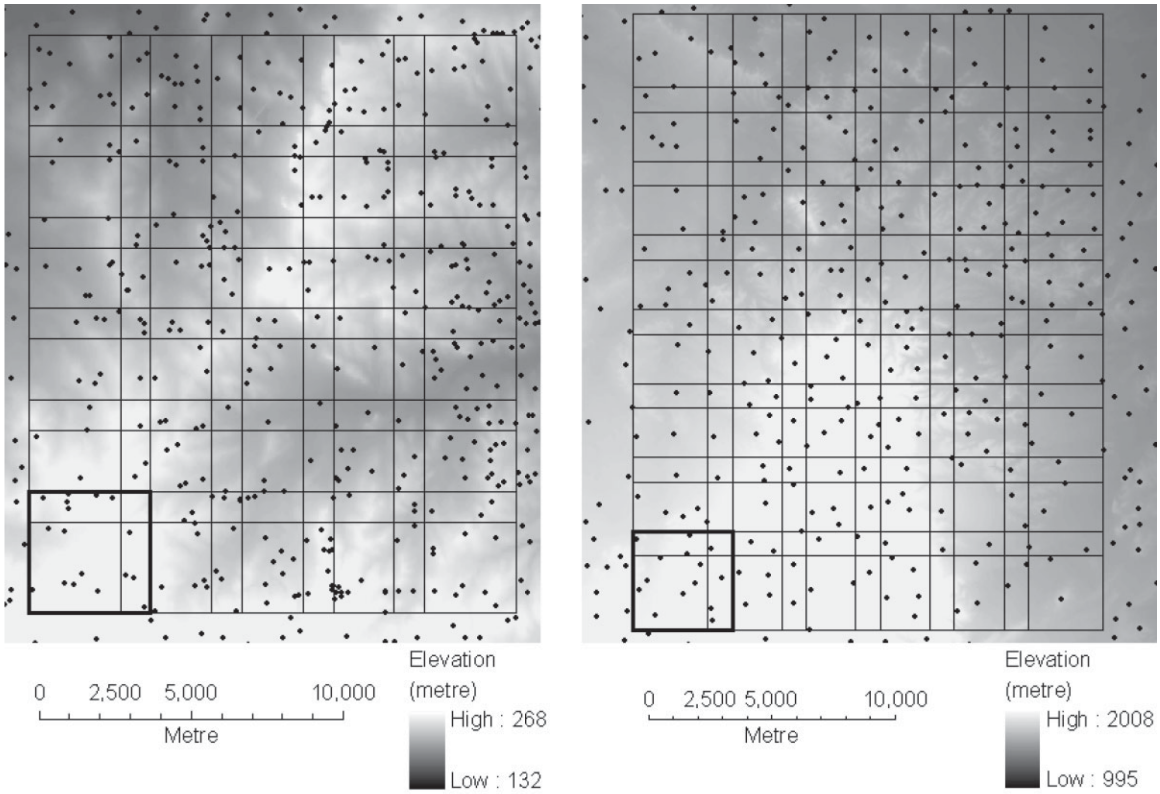

Figure 4. The adjustable moving windows system superimposed on two surface types represented by irregularly spaced sample points (DEM raster in background).

statistics inside the moving windows. A larger overlap, on the other hand, would have extensively smoothed out the local and regional differences between the moving windows.

Geostatistical methods

In order to get the spatial interpolation by ordinary kriging, as well as the computation of predictor x_7 (Equation 7), we had to first model the spatial variability of elevation data. In geostatistics the variability in space is quantified by the semivariogram, which is denoted $\gamma(\mathbf{h})$, and measures the average dissimilarity between observations as a function of separating distance and direction. The experimental variograms were computed and a model was fitted visually using Variowin software (Pannatier, 1996). The spatial structure was considered isotropic (i.e., direction-independent) for planar surfaces and the following spherical model with a nugget effect was fitted:

$$\gamma(\mathbf{h}) = 17{\cdot}33\delta(\mathbf{h}) + 681{\cdot}33\left(\left(\frac{3\mathbf{h}}{2a}\right) - 0{\cdot}5\left(\frac{\mathbf{h}}{a}\right)^3\right)$$

where $a = 27\,321$ m is the range of influence of elevation data, and the nugget effect is 17·33. The Dirac function $\delta(\mathbf{h})$ equals 0 for zero distance (i.e., $|\mathbf{h}| = 0$, and 1 otherwise). For mountain surfaces the variability changes as a function of direction (i.e., anisotropy) and was modelled using the following combination of a nugget effect and the Gaussian variogram model:

$$\gamma(\mathbf{h}) = 1500\delta(\mathbf{h}) + 19\,033\left(1 - e^{\left(\frac{\mathbf{h}^2}{a^2}\right)}\right)$$

where $a = 3056$ m is the range in the direction of maximum continuity (major range) and 2261 m is the minor range in the perpendicular direction. A small nugget value was added to the Gaussian model in order to stabilize the kriging system. The variogram parameters were used in the ordinary kriging system to compute the weights assigned to observations z_i in the spatial estimation:

$$\hat{z}_0 = \sum_{i=1}^{N} \lambda_i z_i$$

where $\hat{z}_0$ represents the estimated value at location $\boldsymbol{u}_0$, λ_i are the weights for the sample values participating in the estimation of $\hat{z}_0$, and z_i are the sample values. The unbiasedness constraint requires that the kriging weights must sum to one. Hence, kriging is a form of weighted average estimator. The weights are assigned on the basis of a model fitted to a function such as the variogram that depicts the spatial continuity of the variable in question. More detailed information concerning variogram modelling and kriging can be found in the articles and textbooks by Journel and Huijbregts (1978), Isaaks and Srivastava (1989), Cressie (1985, 1991), Goovaerts (1997), Gorsich and Genton (2000), and Ortiz and Deutsch (2002).

Model Development

Brower and Merriam (2001) used stepwise multiple regression models for predicting stratigraphic units in the Kansas area. In this study, a set of surface dissectivity parameters, roughness indexes, and statistical measures describing the local characteristics of planar and mountain surfaces were computed for each moving window and the stepwise backward regression procedure was applied to ascertain the relationship between the error magnitude and the surface properties. This procedure operated in two steps: (a) the initial model was first determined from the full regression model (Equation 7) using the backward stepwise procedure; and (b) interactive and polynomial models were then tested and compared to the previous model to determine which functional relationship was more effective in explaining the variations in the magnitude of OK estimated errors. The full regression model is:

$$\hat{m}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{k-1} x_{(k-1)i} + \beta_k x_{ki} + \varepsilon_i \qquad (7)$$

where $\hat{m}$ is the dependent variable, the magnitude of estimation errors for each moving window i, β_j are the regression parameters, x_{ji} are the explanatory variables: x_1, elevation mean; x_2, elevation variance; x_3, minimum elevation; x_4, maximum elevation; x_5, sum of elevations; x_6, range of elevations; x_7, mean kriging standard error (square root of kriging variance); x_8, *Dm* parameter (Equation 4); x_9, *Vh* parameter (Equation 5); x_{10}, *DI* parameter (Equation 6); x_{11}, maximum slope; x_{12}, vertical sum; x_{13}, base slope sum; x_{14}, mean roughness index; x_{15}, x_{16}, minimum and maximum roughness values; x_{17}, range of roughness; x_{18}, variance of roughness; x_{19}, sum of roughness values; x_{20}, number of samples; x_{21}, easting (UTM coordinate system); x_{22}, northing (UTM coordinate system); and ε_i are the residuals with the condition $E(\varepsilon) = 0$. The slope values were computed using the following formula:

$$slope = \sqrt{\left(\frac{\partial_z}{\partial_x}\right)^2 + \left(\frac{\partial_z}{\partial_y}\right)^2}$$

The vertical sum (x_{12} parameter) was computed by summarizing the triangle heights

$$\sum_{i=1}^{N} \left(\frac{\partial_z}{\partial_y}\right)$$

and the base slope sum using

$$\sum_{i=1}^{N} \left(\frac{\partial_z}{\partial_x}\right)$$

for each moving window. The tested hypothesis is:

$$H_0: \quad \beta_0 = \beta_1 = \beta_2 = \beta_k = 0,$$

$$H_1: \quad \text{at least one of } \beta_j,\ j = 1, 2, \ldots k, \text{ is not equal to } 0.$$

The backward elimination consisted of evaluating each t-ratio $\hat{\beta}_j/s_{\hat{\beta}_j}(j = 1, \ldots 22)$ appropriate for testing H_0: $\beta_j = 0$ versus $\beta_j \neq 0$. If the t-ratio with the smallest absolute value is less than a pre specified constant t, which is:

$$\min_{j=1,\ldots 22} \left| \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}} \right| < t$$

then the predictor corresponding to the smallest ratio is eliminated from the model (Devore, 1995). The backward elimination represented the initial step in developing the best model for the magnitude of OK estimated errors.

Planar surface

As the name indicates, the majority of the elevation points are located on a plane within a 50 m altitude range. The central part is moderately elevated and stretches in a NE–SW direction. The backward-stepwise elimination using all 22 parameters resulted in the following model:

$$y_1 = -234{\cdot}5 + 0{\cdot}18x_1 - 0{\cdot}015x_2 - 0{\cdot}008x_5 - 2{\cdot}7x_7 + 5{\cdot}5x_9 + 0{\cdot}0007x_{12} + 0{\cdot}02x_{18} + 1{\cdot}5x_{21} + 5{\cdot}9 \times 10^{-5}x_{22} + e$$

where y_1 indicates the regression of interpolation errors as represented in Equation 2. The F-test for regression, 8·05 at $p < 0{\cdot}0001$, leads one to reject the null hypothesis, which entails that at least one of the coefficients is different from zero. The t-tests for regression parameters shows that the elevation mean (x_1), the elevation variance (x_2), and the mean kriging standard error (x_7) contribute the most to the model ($p = 0{\cdot}001$). The smallest contribution was measured for the northing parameter (x_{22}, $p > |t| = 0{\cdot}08$). Taken together, this model, with nine explanatory variables, explained 78·4 per cent ($r^2 = 0{\cdot}7836$) of the y_1 variance. The same elimination process using the second type of error (Equation 3) resulted in the following model:

$$y_2 = -51{\cdot}4 + 0{\cdot}2x_1 - 0{\cdot}01x_5 + 2{\cdot}45x_7 - 0{\cdot}66x_{10} + 0{\cdot}38x_{11} + 0{\cdot}0007x_{12} + 0{\cdot}04x_{17} - 0{\cdot}052x_{18} + 1{\cdot}83x_{20} + e$$

where y_2 indicates the regression of interpolation errors as represented in Equation 3. The F-test for regression, 15·05 at $p < 0{\cdot}0001$, and r-squared value 0·87 indicate that the model better explains the variation in y_2. The structure of predictors is similar to some extent to the previous model; on the other hand, new parameters began to play an important role. The contribution of the square root of the kriging variance (OK standard error) to both models is noteworthy. Its function in spatial analysis is very limited due to the second-order stationary assumptions, and therefore it was deemed useful only as a parameter describing the sampling configuration or the sampling density (Oliver and Webster, 1991; Goovaerts, 1997; Lloyd and Atkinson, 2001).

Interactive models

In some cases, it is possible to construct other models by using predictors that are mathematical functions of the other explanatory variable x_j's. The intersection between Vh (x_9), Dm (x_8), maximum slope (x_{11}), number of samples (x_{20}), and the base slope sums (x_{12}) appears to be an effective predictor. The interaction term used was determined from a thorough analysis of partial regression coefficient tests in numerous regression models by trial and error methods. Less significant explanatory variables from the previous initial models were dropped from the full set of predictors and the backward procedure was used with the addition of the intersection terms: x_8 (Dm), x_9 (Vh), x_{11} (maximum slope), and x_{12} (vertical sum). The resulting model:

$$y_1 = -47{\cdot}8 + 0{\cdot}21x_1 - 0{\cdot}01x_5 + 1{\cdot}9x_7 + 8{\cdot}4 \times 10^{-8}x_8x_9x_{11}x_{12} + e$$

provides an improved fit in terms of the following parameters. The F value increased to 12·5 at $p < 0{\cdot}0001$, leading to a clear rejection of the H_0 hypothesis. In addition, the t-statistics for each parameter exhibited a higher level of significance ($p = 0{\cdot}001$). The r-squared value decreased slightly to 0·723 due to a smaller number of explanatory variables present in the model and the residuals were evenly distributed around a zero mean. The results underlined the importance of using the elevation mean and the OK standard error as predictors of the magnitude of estimation errors in the planar surface.

The regression of y_2 values with all 22 predictors resulted in the model:

$$y_2 = -51{\cdot}4 + 0{\cdot}2x_1 - 0{\cdot}01x_5 + 2{\cdot}5x_7 + 0{\cdot}38x_{11} + 0{\cdot}7 \times 10^{-3}x_{12} - 0{\cdot}66x_{10} + 0{\cdot}037x_{18} - 5{\cdot}2 \times 10^{-3}x_{19} + 1{\cdot}8x_{21} + e$$

The elevation mean (x_1), the sum of elevations (x_5), the sum of vertical distances (x_{12}), the OK standard error (x_7), and the number of samples (x_{20}), contributed considerably to the model with p being <0·0001. The new variables, *DI* ($p = 0{\cdot}03$), range ($p = 0{\cdot}02$), and the variance of roughness ($p = 0{\cdot}016$), also were found to be useful explanatory variables. The highest r-squared value (0·87) indicates that 87 per cent of variation in the OK estimation errors could be explained using nine explanatory variables. Consequently, this model had the highest F-test statistics. Note that due to the unequal sample size in each moving window the sums and means of elevation values are not duplicate parameters, and their correlation is especially low for a planar surface ($r = 0{\cdot}25$). Therefore, each parameter contributed in its own way to the regression model.

High (mountain) surface

If the magnitude of the kriging estimation errors is independent from surface properties, then its relationship with the roughness–dissectivity parameters would remain approximately constant in a plane or mountain topography. Our results from the model fitting indicated that the general structure of predictors in the regression models rapidly changed for the mountain surface in comparison to the low (planar) surface. For example, the following predictors were eliminated from the model: the sum of elevations (x_5), the OK standard error (x_7), and the number of samples (x_{20}). On the other hand, the importance of all dissectivity parameters rapidly increased along with the coefficients of roughness. Again, the best model (i.e. highest $r^2 = 0{\cdot}95$) was achieved using two interaction terms: (a) the interaction between x_9, x_{12}, (*Vh* and vertical sum), and (b) x_2, x_9, x_{10}, x_{19} (elevation variance, *Vh*, *DI*, variance of roughness). The following model:

$$y_1 = 56 + 0{\cdot}0061x_2 - 0{\cdot}0421x_3 + 3{\cdot}5x_{10} + 0{\cdot}1x_{16} + 2{\cdot}2 \times 10^{-5}x_{19} - 0{\cdot}0042x_9x_{12} - 122 \times 10^{-16}\,x_2(x_9)^3x_{10}x_{19} + e$$

yielded the highest F-test value (105·5, $p < 0{\cdot}0001$), and the p-values for the parameters' t-tests were consistently below 0·0001 for all predictors except the minimum elevation parameter with $p = 0{\cdot}0003$ (see Table II).

The backward stepwise regression for y_2 was significant at the initial stage:

$$y_2 = -196{\cdot}95 + 0{\cdot}07x_1 - 0{\cdot}08x_4 + 7{\cdot}06x_8 - 11{\cdot}3x_9 + 2{\cdot}63x_{10} + 0{\cdot}007x_{12} - 7{\cdot}3x_{10} - 7{\cdot}3 \times 10^{-5}x_{14} + 0{\cdot}05x_{15} - 0{\cdot}004x_{20} - 0{\cdot}003x_{18} + 0{\cdot}0004x_{22} + e$$

The F-statistic 57·34 at $p < 0{\cdot}0001$ led to the rejection of the H_0 hypothesis (i.e. at least one of the coefficients was not zero). The model also explained 95 per cent of the variation in the magnitude of the OK estimated errors. The

Table II. Correlations of surface parameters with the magnitude of OK errors

	E1						E2					
	Planar (low)			Mountain (high)			Planar (low)			Mountain (high)		
Parameter	r	r²	t	r	r²	t	r	r²	t	r	r²	t
Dm	0·39	0·15	2·06	0·87	0·76	5·85	0·5	0·25	2·63	0·86	0·74	5·79
Vh	0·37	0·14	1·95	0·86	0·74	5·79	0·48	0·23	2·53	0·86	0·74	5·79
DI	0·44	0·19	2·32	0·88	0·77	5·92	0·48	0·23	2·53	0·86	0·74	5·79
MS	0·11	0·01	0·58	0·83	0·69	5·59	0·18	0·03	0·95	0·82	0·67	5·52
RM	0·17	0·03	0·90	0·82	0·67	5·52	0·27	0·07	1·43	0·81	0·66	5·45
EVA	0·34	0·11	1·80	0·84	0·71	5·65	0·43	0·18	2·27	0·83	0·69	5·59
MEL	0·24	0·06	1·27	0·72	0·52	4·86	0·18	0·03	0·95	0·69	0·47	4·66

Dm, Dissectivity mean; *Vh*, vertical horizontal ratio; *DI*, dissectivity index; MS, maximum slope; RM, mean of roughness index; EVA, elevation variance; MEL, mean of elevations.

dissectivity metrics were found to be important parameters for predicting the magnitude of the OK estimated errors and they were associated with the lowest variance inflation value.

Discussion

Results indicate that the functional relationship between the independent variable (magnitude of OK estimation errors) and the roughness–dissectivity is stronger in the mountain topography where steeper slopes and more rapid changes in elevation dominated the surface. Table II contains elements of the correlation matrix and the percentage of variance in the y parameters explained by independent variables. The t-test for correlation significance was computed using the formula:

$$t = r\sqrt{\frac{k-2}{1-r^2}}$$

where k is the total number of moving windows and r is the correlation coefficient. The t-test revealed that the strength of this relationship, especially in the mountain surface, is clearly above the critical value of 1·69 at alpha 0·05.

Regression analysis results demonstrated that the dissectivity parameters, the roughness index, and the measurements of central tendency (mean, variance, range etc.) can be used as valuable predictors of local variations in the magnitude of the OK estimated errors, and their explanatory potential increases with roughness–dissectivity. Thus, the regression models can explain up to 95 per cent of the variation in the magnitude of OK estimated errors in more dissected landform settings. The presence or absence of a particular surface parameter in the regression model is also a function of the roughness–dissectivity of the interpolated surface. For example, the square root of the kriging variance (labelled here as the OK standard error), the sum of elevations, and the number of observations played an important role as predictors only in the regression of a low (planar) surface. The kriging standard error, which usually has a limited use mostly as an indicator of data density and sample configuration, seems to be a useful parameter notably contributing to the planar regression model. In contrast, the roughness metric only played an important role as a predictor in the mountainous surface. The most effective predictor appeared to be the Vh parameter (Equation 5), which was included in every regression model either as a stand alone independent explanatory variable or as a combination with other parameters in the interactive regression model. In contrast, the maximum slope, the easting and northing values, and the maximum elevation made a very limited contribution to the regression models as predictors. The elevation range parameter (x_6) was not useful in regression even though it is considered an important measure of local relief (Mark, 1975). Overall, the increase in z-values (elevation) has only a small effect on the increase in the magnitude of kriging errors for a planar surface, which is expected because the increase is small in magnitude.

The regression coefficients of the independent variables can be confounded by multicollinearity as indicated by a high variance inflation factor for a few explanatory variables. The Earth science data are typically autocorrelated; therefore, multicollinearity is an expected phenomenon in the regression process. Likewise, it is not a modelling error, rather it is a condition of deficient data. From the t-values (Table II) it follows that most of the regressors are important in the model and the deletion of one predictor would have a significant effect on the explanatory power of the model (predictors cannot serve as a proxy for one another). Therefore, the linear regression models do not suffer from a severe case of multicollinearity (Chatterjee and Price, 1977). Nevertheless, the results of this work indicate that there is a need to develop and test new orthogonal predictors in the future to eliminate, or at least decrease, the multicollinearity to a minimum for surface dissectivity analysis. Lognormal transformation and consequent regression using logarithmically transformed data did not significantly improve the model. On the other hand, the relationship between the error magnitude and surface variability, hence the model structure, depends on the definition of error magnitude. In order to understand more profoundly this relationship, two definitions of error magnitude were used (Equations 2 and 3) in the modelling process.

Conclusion

This paper has presented multiple regression models and regression coefficients that relate the magnitude of OK estimated errors to a set of dissectivity and roughness parameters. The magnitude of estimation errors increases or decreases depending on the surface type (low or high) and the local variability of the surface. The research identified several key issues regarding the magnitude of the OK estimated errors and their relationship to the variability of local surface parameters.

(a) The increase of terrain roughness–dissectivity increases the magnitude of the OK estimation errors. In addition, the variation in the magnitude of the OK estimation errors is explained more fully in a mountainous type than in a planar surface.
(b) The structure of predictors in the regression models and their explanatory power changes with roughness–dissectivity.
(c) The most important predictors in the mountain surface were dissectivity (*Dm*, *Vh*, *DI*) parameters, while the kriging variance, the number of samples, and the mean of elevations were more important in the planar surface.
(d) The effect of the square root of the kriging variance (kriging standard error) on the magnitude of the OK estimation errors (the expected change in the OK estimated error when increased by one unit of kriging variance), was several times stronger than the rest of the parameters present in the model. Moreover, the kriging standard error indicated only a small inflation of the model variance (the multicollinearity aspect).
(e) The most significant contribution to the regression models was made by the *Vh* parameter (Equation 5) while the effect of the mean elevation and the elevation variance was insignificant.
(f) The interactive regression models yielded the largest model's *r*-squared value and the largest *t*-test statistics (the *p*-values less than 0·0001) for parameter estimates. The interactive term, which is composed of the elevation variance, the dissectivity index (*DI*), the variance of roughness, and the *Vh* parameter, produced the smallest *p*-value (less than 0·0001) and a relatively small variance inflation value.
(g) Results indicate that the magnitude of the OK errors varies locally according to the surface characteristics and can be modelled using multiple regression models and surface parameters as predictors.

Acknowledgments

This work was financially supported by the College of Forestry at Stephen F. Austin State and the USDA Cooperative State Forestry Research (McIntire-Stennis grant no. 414424). The authors are especially thankful for the detailed comments, suggestions and corrections of two reviewers that played an instrumental role in the revisions of the original manuscript. Their recommendations are gratefully acknowledged.

References

Brower JC, Merriam DF. 2001. Thematic map analysis using multiple regression. *Mathematical Geology* **33**(3): 353–368.
Chatterjee S, Price B. 1977. *Regression Analysis by Example*. John Wiley & Sons: New York.
Chorley RJ. 1972. *Spatial Analysis in Geomorphology*. Harper and Row: New York.
Cooper PR, Friedman DE, Wood SA. 1987. The Automatic generation of digital terrain models from satellite images by stereo. *Acta Astronautica* **15**: 171–180.
Cressie N. 1985. Fitting variogram models by weighted least squares. *Mathematical Geology* **17**(5): 563–584.
Cressie N. 1991. *Statistics for Spatial Data*. John Wiley & Sons: New York.
Devore JL. 1995. *Probability and Statistics for Engineering and the Sciences*. Duxbury Press: Belmont.
Eklundh L, Martensson U. 1995. Rapid generation of digital elevation models from topographic maps. *International Journal of Geographical Information Systems* **9**(3): 320–340.
Giles PT, Chapman MA, Franklin SA. 1994. Incorporation of a digital elevation model derived from stereoscopic satellite imagery in automated terrain analysis. *Computer and Geosciences* **20**: 441–450.
Goodchild MF. 1994. *The Accuracy of Spatial Databases*. Taylor & Francis: London.
Goovaerts P. 1997. *Geostatistics for Natural Resource Evaluation*. Oxford University Press: New York.
Goovaerts P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* **228**: 113–129.
Goovaerts P. 2001. Geostatistical modeling of uncertainty in soil science. *Geoderma* **103**: 3–26.
Gorsich JD, Genton GM. 2000. Variogram model selection via nonparametric derivative estimation. *Mathematical Geology* **32**(3): 249–269.
Isaaks ED, Srivastava RM. 1989. *Introduction to Applied Geostatistics*. Oxford University Press: New York.
Journel AG, Huijbregts Ch J. 1978. *Mining Geostatistics*. Academic Press: London.
Kastens T, Staggenborg S. 2002. Spatial interpolation accuracy. *Kansas State Precision Conference*, Great Bend. www.agmanager.info/crops/prodecon/precision/default.asp.
Krcho J. 1991. Georelief as a subsystem of landscape and the influence of morphometric parameters of georelief on spatial differentiation of landscape-ecological processes. *Ecology* (*Ekologia*) **10**: 115–157.
Lloyd CD, Atkinson PM. 2001. Assessing uncertainty in estimates with ordinary and indicator kriging. *Computers & Geosciences* **27**: 929–937.
Mark MD. 1975. Geomorphometric parameters: a review and evaluation. *Geografiska Annalen* **57**(3/4): 165–176.
Miller CL, La Flamme AR. 1958. The digital theory application. *Photogrammetric Engineering and Remote Sensing* **7**: 433–442.
Moyeed RA, Papritz A. 2002. An empirical comparison of kriging methods for nonlinear spatial point prediction. *Mathematical Geology* **34**(4): 365–386.

Oliver MA, Webster R. 1991. How geostatistics can help you? *Soil Use and Management* **7**: 206–217.
Ortiz CJ, Deutsch VC. 2002. Calculating of uncertainty in the variogram. *Mathematical Geology* **34**(2): 169–183.
Pannatier Y. 1996. *VARIOWIN Software for spatial data analysis in 2D*. Springer-Verlag: New York.
Phillips GM, Watson DF. 1986. A method for assessing local variation among scattered measurements. *Mathematical Geology* **18**(8): 759–764.
Ritter DF, Kochel CR, Miller JR. 2001. *Process Geomorphology*. McGraw-Hill: New York, fourth edition.
Schmidt J, Evans IS, Brinkmann J. 2003. Comparison of polynomial models for land surface curvature calculation. *International Journal of Geographical Information Science* **17**(8): 797–814.
Siska P, Eriksson M, Maggio RC. 1997. Ordinary kriging and construction of a digital elevation model. *Geografický Časopis* **49**(2): 91–113.
Siska PP, Hung I-K. 2001. Propagation of errors in spatial analysis. *Papers and Proceedings of the Applied Geography Conferences, University of North Texas* **24**: 284–290.
Skidmore AK. 1989. A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. *International Journal of Geographical Information Systems* **3**(4): 323–334.

