GEOGRAFICKÝ ČASOPIS

49 1997 2

Peter Šiška, Marian Eriksson**, Robert C. Maggio**

ORDINARY KRIGING AND THE CONSTRUCTION OF A DIGITAL ELEVATION MODEL

Peter Šiška, Marian Eriksson, Robert Maggio: Ordinary Kriging and the Construction of a Digital Elevation Model. Geografický časopis, *49*, 2, 1997, 8 figs., 1 tab., 16 refs.

Implementation of geostatistical methods into natural resource management studies marked the end of the 1980s and the beginning of the 1990s. Areas of application included soil science, entomology, ecology, forestry, and geography, among others. In particular, ordinary kriging has seen widespread use in recent years. Its use can be expected to increase as understanding of the procedure and availability of geostatistical software increases. One objective for this paper is to introduce kriging to a new audience. During this period there has also been a dramatic increase in Geographic Information Systems (GIS) both in terms of capabilities and the breadth of applications. Many GIS applications make use of digital elevation models (DEMs), most of which have been generated from elevation point data using deterministic models. We present results of research into the use geostatistical methods to develop and test a stochastic (probabalistic) DEM. Both GIS and spatial statistical software were used during the process of generating and analyzing the data and displaying the resulting DEM. The use of stochastic methods was motivated by the observation that many earth-related attributes are generated by stochastic processes. Kriging was the spatial statistical tool selected for model development. Results corraborated the hypothesis that kriging can produce highly accurate DEMs. The density and distribution of sample locations determines the absolute accuracy.

Key words: geostatistics, variogram, interpolation, geographic information system

1 INTRODUCTION

The term "digital elevation model" (DEM) refers to the numerical representation of the earth's surface. Increasing implementation of computerized technology in the

* Mapping Sciences Laboratory, Centeq Plaza Suite 220, 1500 Research Parkway, College Station, TX 77843-2120, USA

** Department of Forest Science, Texas A&M Univesity, College Station, TX 77843-2135, USA

natural resource and geo-sciences is accompanied by growing interest in three dimensional surface modeling. DEMs are being applied in a wide variety of projects such as soil erosion calculations, hydrologic modeling, vegetation mapping, landuse development, planning, architecture etc. Geographic Information Systems (GIS) widely support terrain representations and provide creative environments for further applications and extensions.

DEMs can be generated from a variety of sources. Giles et al. (1994) derived DEMs from satellite imagery using multispectral stereoscopy, while Cooper et al. (1987) used simulated satellite data to generate their DEMs. Recently, DEMs are increasingly generated by direct measurements from digital aerial photographs. However, DEMs can also be developed entirely by spatial interpolation of point data or contour lines, including second- or higher-order polynomials. Examples can be found in Miller La Flamme (1958), Birkhoff and Mansfield (1974), and Nelson and Jones (1995). In contrast to these deterministic interpolation models, a variety of spatial statistical tools, such as kriging, can be utilized to develop interpolation models. The purpose of this article is to introduce kriging and to develop the statistical model and resulting estimation procedures for ordinary kriging. A kriging-based DEM for an area located in western Texas, USA, is developed as an example. Kriging procedures, however, can be applied to a wide variety of spatial data in the natural resource and geo-sciences that can be assumed to have been generated by random processes.

2 THE KRIGING MODEL

Kriging is based on probabilistic principles. The decision to use kriging instead of other interpolation methods for developing our DEM was supported by the fact that kriging is a "best linear" spatial interpolation procedure. It is "best" in the sense that, given the model assumptions, it is unbiased and it has the smallest mean squared prediction error. That is, the smallest prediction variance. It is linear in the sense that the predicted value at *any* location within the study domain is taken to be a linear combination of data values actually observed at other locations within the domain. In particular, let $Z(x)$ be a random variable taking on values, $z(x)$, at locations x in the area (domain, D) of study. At any particular x, $Z(x)$ is a random variable. In the literature, the set $Z(x) = \{Z(x): x \in D\}$ is called a *random function*.

In our study the $Z(x)$ is the random process that generated elevation at x and the $z(x)$ are elevation values. Kriging models are of the form

$$Z(x) = \mu(x) + \varepsilon(x), \tag{1}$$

where $\mu(x)$ is a mean function and $\varepsilon(x)$ is a random error process with $E(\varepsilon(x)) = 0$. This implies that $\mu(x) = E(Z(x))$. Actually, there are a number of different kriging models. Most notable are *simple kriging*, in which $\mu(x)$ is assumed to be known, *ordinary kriging*, for which $\mu(x)$ is assumed to be an unknown constant: $\mu(x) = \mu$ for all x in the domain, *universal kriging*, where $\mu(x)$ is assumed to be an unknown function to be estimated, much like a regression model, and *nonlinear kriging*, in which case the mean function is a nonlinear function of the data. Because ordinary kriging is, by far, the most commonly used kriging technique, and because an understanding of the various types of kriging are derived from an understanding of ordinary kriging, only ordinary kriging will be considered herein.

The ordinary kriging estimator for the elevation at *any*, sampled or unsampled, location x_0 in the domain is linear in the data. That is, it has the form:

$$\hat{z}(x_0) = \sum_{i=1}^{n} w_{0_i} z(x_i), \tag{2}$$

where $z(x_i)$, $i = 1,2,...,n$, are known elevation data available for model development. This implies the random variable

$$\hat{Z}(x_0) = \sum_{i=1}^{n} w_{0_i} Z(x_i). \tag{3}$$

The w_{0_i} are called the kriging weights for location x_0. They are truly weights in the sense that $\sum_{i=1}^{n} w_{0_i} = 1$. This is true for all prediction locations, x_0. The kriging weights are unknown constants that must be estimated as a part of model construction. The condition that weights must sum to one is actually imposed as a constraint in the development of the kriging estimator. It ensures unbiasedness.

Since the weights and μ are all constants (not stochastic), upon taking the expectation of $\hat{Z}(x_0)$ in equation (3) we see that

$$E(\hat{Z}(x_0)) = E\left(\sum_{i=1}^{n} w_{0_i} Z(x_i)\right) = \sum_{i=1}^{n} w_{0_i} E(Z(x_i)) = \sum_{i=1}^{n} w_{0_i} E(\mu + \varepsilon(x_i)) =$$
$$= \sum w_{0_i}\mu + \sum w_{0_i} E(\varepsilon(x_i)) = \mu \sum w_{0_i} + \sum w_{0_i} 0 = \mu \sum w_{0_i} .$$

This indicates that, in order for $\hat{Z}(x_0)$ to be unbiased for μ, we must have the constraint that the kriging weights for location x_0 sum to one.

As noted earlier, the kriging estimator is "best" in the sense that it is unbiased and that it minimizes the mean squared prediction error for models having the form of the of equation (1) and estimators of the form of (2). The mean squared prediction error, also called the kriging variance or the prediction variance, for location x_0 is given by $E(\hat{Z}(x_0) - Z(x_0))^2$. We want to minimize this quantity subject to the constraint that $\sum_{i=1}^{n} w_{0_i} = 1$. This can be accomplished by introducing the Lagrangian multiplier, λ, and setting to zero the derivatives of the Lagrangian function,

$$L(w_{0_i}, \lambda : i = 1,2,...,n) = E(\hat{Z}(x_0) - Z(x_0))^2 + 2\lambda\left(\sum_{i=1}^{n} w_{0_i} - 1\right). \tag{4}$$

In Appendix 1 we show that

$$L(w_{0_i},\lambda:i=1,2,...,n)=\sum_{i=1}^{n}\left(w_{0_i}^2\sigma_{ii}+2w_{0_i}\sum_{j\neq i}^{n}w_{0_j}\sigma_{ij}\right)-2\sum_{i=1}^{n}w_{0_i}\sigma_{0_i}+\sigma_{00}+2\lambda\left(\sum_{i=1}^{n}w_{0_i}-1\right),$$

where:

$$\sigma_{ij}=Cov(Z(x_i),Z(x_j)),$$
$$\sigma_{0_i}=Cov(Z(x_0),Z(x_i)),\text{ and}$$
$$\sigma_{ii}=Cov(Z(x_{i)},Z(x_i))=Var(Z(x_i)),$$

for $i,j=0,1,...,n$.

From introductory calculus, we know that setting the derivatives of L, with respect to the n+1 parameters, λ, w_{0_i}, $i=1,2,...,n$, equal to zero minimizes L and, consequently, minimizes the mean squared prediction error subject to the constraint that the weights sum to one.

Since there are n+1 parameters, there are n+1 derivatives; the first of which is

$$\frac{\partial L}{\partial\lambda}=2\left(\sum_{i=1}^{n}w_{0_i}-1\right).$$

The remaining n equations take the form

$$\frac{\partial L}{\partial w_{0_i}}=2\sum_{j=1}^{n}w_{0_j}\sigma_{ij}-2\sigma_{0_i}+2\lambda,\quad i=1,2,...n.$$

Setting all n+1 derivatives equal to zero results in $\sum_{i-1}^{n}w_{0_i}=1$, our constraint, and

$$\sum_{j=1}^{n}w_{0_j}\sigma_{ij}+\lambda=\sigma_{0_i},\quad i=1,2,...,n.$$

Writing these n+1 equations out in all their glory we obtain:

$$\begin{aligned}
&\sigma_{11}w_{01}+\sigma_{12}w_{02}+\ldots+\sigma_{1n}w_{0n}+\lambda=\sigma_{01}\\
&\sigma_{21}w_{01}+\sigma_{22}w_{02}+\ldots+\sigma_{2n}w_{0n}+\lambda=\sigma_{02}\\
&\quad\vdots\qquad\qquad\ddots\qquad\qquad\vdots\\
&\sigma_{n_1}w_{01}+\sigma_{n_2}w_{02}+\ldots+\sigma_{nn}w_{0n}+\lambda=\sigma_{0n}\\
&\qquad w_{01}+w_{02}+\ldots+w_{0n}=1
\end{aligned}\qquad(5)$$

which can be simply expressed in the matrix form: $\dot{\mathbf{C}}\dot{\mathbf{w}}_0=\dot{\mathbf{c}}_0$. The dots indicate partitioned matrices. The "sub-0" subscript is a reminder that the kriging equations are for the prediction at a particular location, x_0. Components of those matrices are

$$\mathbf{C} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{n2} \\ \cdot & \cdot & \dots & \cdot \\ \sigma_{1n} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix}, \quad \mathbf{c}_0 = \begin{bmatrix} \sigma_{01} \\ \sigma_{02} \\ \cdot \\ \sigma_{0n} \end{bmatrix}, \quad \mathbf{w}_0 = \begin{bmatrix} w01 \\ w02 \\ \cdot \\ w0n \end{bmatrix}, \quad \text{and} \quad \mathbf{j} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{bmatrix}.$$

The partitioned matrices are

$$\dot{\mathbf{C}} = \begin{bmatrix} \mathbf{C} & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix}, \quad \dot{\mathbf{w}}_0 = \begin{bmatrix} \mathbf{w}_0 \\ \lambda \end{bmatrix}, \quad \text{and} \quad \dot{\mathbf{c}}_0 = \begin{bmatrix} \mathbf{c}_0 \\ 1 \end{bmatrix}.$$

The weights and the Lagrangian multiplier are determined by $\dot{\mathbf{w}} = \dot{\mathbf{C}}^{-1}\dot{\mathbf{c}}_0$. Expressions for the kriging (prediction) variance are provided in Appendix 3.

3 FUNCTIONS OF DISTANCE AND THEIR IMPLEMENTATION IN THE KRIGING SYSTEM

There are two problems with the system (5) of equations which extend to the corresponding matrix representation. First, it involves unknown population covariances that must be estimated before the system can be solved. Secondly, it involves as many as $n(n + 1)/2$ distinct covariances in $\mathbf{C}$, plus another n covariances in $\mathbf{c}_0$. It would be impossible to estimate all $(n^2 + 3n)/2$ of these covariances from our n data points without some simplifying assumptions. The most fundamental of these assumptions is that the covariance of elevation between points is a function of only the distance between those points. Elevations, or other attributes, at locations relatively close to each other will tend to be similar and those of locations distant from one another will tend to be more different. The further the distance apart, the more dissimilar the data will tend to be. That is, correlation between data values tends to decrease with distance. Like correlation, covariance decreases with increasing distance. By modeling the covariance structure as a function of distance between points, and knowing the distance between all locations x_i, x_j, and x_0, in (5), we will be able to estimate the covariances needed to solve for the kriging weights.

Covariance, expressed as a function of distance, is called both the *covariance function* and the *covariogram*. Specifically, if h is the horizontal distance between two points separated by a vector $\boldsymbol{h}$, then $C(h) = C(\|\boldsymbol{h}\|) = \text{Cov}(Z(x), Z(x + \boldsymbol{h}))$. In the case of the covariances in system (5):

$$\sigma_{ii} = \text{Cov}(Z(x_i)), Z(x_i)) = C(0) = \sigma^2,$$

$$\sigma_{ij} = \text{Cov}(Z(x_i)), Z(x_j)) = C(\| x_i - x_j \|), \text{ and}$$

$$\sigma_{0i} = \text{Cov}(Z(x_0)), Z(x_i)) = C(\| x_0 - x_i \|).$$

It turns out that, rather than modeling the covariogram directly, applied researchers often prefer to model a closely related function, the *semivariogram*, $\gamma(h) = C(0) - C(h)$. The variogram, itself, is defined as $2\gamma(h)$. It is easy to show that, for locations $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, $2\gamma(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|) = E(Z(\boldsymbol{x}_i) - Z(\boldsymbol{x}_j))^2$.

(Some authors, e.g. Pannatier 1996, call $\gamma(h)$ the variogram). Estimated variogram functions tend to be better behaved than estimated covariogram functions (Cressie 1994). If $C(0)$ exists, it is a constant and, since $C(h)$ is a decreasing function of h, it is clear that $\gamma(h)$ is an increasing function of distance. This simple relationship between the covariance function and the variogram indicates that, when $C(0)$ exists, the kriging weights can be equally well expressed by estimated covariance or variogram values. Indeed, rather than expressing the system (5) of equations and the resulting matrix equation in terms of covariances, many authors choose to express it in terms of variogram values. The correspondence is presented in Appendix 2. Because the covariogram and the associated variogram are assumed to be functions only of distance, they are sometimes (e.g., Isaaks and Srivastava 1989) referred to as models of spatial continuity.

Reference was made in the preceding paragraph to the existence of $C(0)$. Two fundamental model assumptions are that $E(Z(\boldsymbol{x})) = \mu$ for all $\boldsymbol{x}$ in the domain of study and that the $E(Z(\boldsymbol{x})-Z(\boldsymbol{x}+\boldsymbol{h}))^2 = 2\gamma(\|\boldsymbol{h}\|)$ is a function only of the distance between the points. Together, these two assumptions constitute the assumption that $Z(\boldsymbol{x})$ is *intrinsically stationary*. If, in addition, the covariance between points is a function only of the distance between the points, ie. that $\mathrm{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{h})) = C(\|\boldsymbol{h}\|)$ for all points in the domain, the process is said to be *second-order stationary*. Second-order stationarity implies the existence of $C(0)$. The assumption of second-order stationarity is stronger than the intrinsic assumption and allows a wider class of inferences. For models that are intrinsically, but not second-order, stationary, the variogram exists but $C(0)$ does not. As a result, the variogram can be modeled but not the covariogram. This is one of the reasons that many applied researchers prefer to work with the variogram rather than the covariogram for modeling spatial continuity.

3.1 MODELING SPATIAL CONTINUITY

The idea behind modeling the variogram is quite simple:

Step 1. Choose a sequence of distances, say $h_1, h_2, \ldots, h_m$. The difference, $h_{k+1} - h_k$, between successive values is often constant and called the lag-distance.

Step 2. For each chosen distance, h_k, select from the set $\{\boldsymbol{x}_i: i = 1,2,\ldots,n\}$, those points that are separated, or approximately separated, by h_k. Estimate the variogram value, $2\tilde{\gamma}(h_k)$, between those points. If $N(h_k)$ is the number of distinct pairs of locations that are separated by h_k the usual estimator is $2\tilde{\gamma}(h_k) = \sum_{N(h_k)} (z(\boldsymbol{x}_i) - z(\boldsymbol{x}_j))^2 / N(h_k)$.

Step 3. Fit a smooth function, $2\gamma(h)$, to the computed $2\tilde{\gamma}(h_k)$s, $k = 1,2,\ldots,m$.

Step 4. Estimate $\gamma(\ \|\ \boldsymbol{x}_i - \boldsymbol{x}_j\ \|\ )$ and $(\ \|\ \boldsymbol{x}_0 - \boldsymbol{x}_i\ \|\ )$, i and j ranging from 1 to n, required to find the kriging weights. These values are substituted into Γ and γ_0 in Appendix 2 in order to solve the kriging equations.

The algorithm, as stated above, is a bit of a simplification. First, it is often the case that the (co)variogram structure is not the same in all directions. In the case of elevation, it is reasonable to expect that, if we have a mountain range oriented along an axis, then the correlation (covariance) between pairs of points in the direction of that axis may be different than the correlation between point pairs perpendicular to

the axis. If variograms fitted in two directions are the same, they are said to be *isotropic* to one another. If variograms fitted in all directions are the same, the variogram is said to be *omnidirectional* or *universal*. Conversely, if variograms fitted from data oriented in two directions are different, they are *anisotropic* to one another. If anisotropy is present, then Step 2 of the algorithm must be modified to first select a direction for computing the variogram, and to only select pairs of points that are separated by h_j, in that direction. The resulting variogram is called a *directional variogram*. Omnidirectional variograms are ideal because all of the available data, not just pairs of data oriented along a given axis, can be used to estimate variogram values, $\tilde{\gamma}(\boldsymbol{h}_j)$; giving rise to more stable estimates. Also, if anisotropy is present, then variograms in at least two directions must be fitted and then combined to give the final variogram estimator: $\gamma(\boldsymbol{h}) = f(\gamma_1(\|\boldsymbol{h}_1\|),\gamma_2(\|\boldsymbol{h}_2\|),...)$, where f is some combining function. Such combining of variograms is beyond the scope of this paper. A follow-up paper will consider the question of anisotropy in some detail. We simply note that it is important to always compute and to graphically compare directional variograms in order to assess whether the omnidirectional variogram is justified.

Variograms can be constructed from regularly or irregularly spaced data. If data are irregularly spaced, as were our elevation data, it is unlikely that any two locations will be exactly separated by the distances chosen in Step 1. As a result, distance tolerances are specified and those points separated by $h_i \pm$ tolerance are selected in Step 2. For directional variograms, tolerances on the direction of separation are also imposed.

As for Step 3, there are potentially a very large number of functional forms suitable for modeling the variogram. Most spatial modelers restrict their attention to a small number of models that have proven quite useful and flexible in most situations. Many variograms are well-fitted by the *spherical model*. This model is good for defining some of the terms related to variogram modeling. The spherical model is specified as:

$$\gamma_S(h) = \begin{cases} 0, & h = 0, \\ c_0 + c(3h/a - (h/a)^3)/2, & 0 < h \le a \\ c_0 + c & h > a \end{cases} \qquad (6)$$

A generic spherical variogram is shown in Figure 1a. There is a discontinuity in this function. It is equal to zero at h=0, but it approaches c_0 as h approaches zero from the right. The discrepancy, c_0 = 20, is called the *nugget effect*. It is usually attributed to measurement error; specifically to the fact that we can only measure elevation at distances finitely close to any given location. The other discontinuity is at $h = a = 35$; a is called the *range*. Beyond the range, the variogram is constant at a value of $c_0 + c = 100$. This value is called the *sill*.

An exponential variogram is shown in Figure 1b. Exponential models take the form

$$\gamma_E(h) = \begin{cases} 0, & h = 0, \\ c_0 + c(1 - \exp(-h/a)), & 0 < h \le a. \end{cases} \qquad (7)$$

The nugget effect is again represented by c_0 and, as $h \to \infty$, $\gamma_E(h)$ approaches the sill of $c_0 + c$. However, since $\gamma_E(h)$ never actually reaches the sill, the concept of a

range is more vague. The *practical range* is taken as $h = 3a$. This is the distance at which the variogram reaches approximately 95% of its sill.

The spherical and exponential models both have sills. Models of this type are sometimes called *transition models*. Figure 1c shows three different power variogram models. A power model is of the form:

$$\gamma_P(h) = \begin{cases} 0, & h = 0, \\ c_0 + c.h^a, & 0 < a < 2. \end{cases} \tag{8}$$

These variograms increase without bound as h increases. That is, they do not have sills. It turns out that variograms of second-order stationary random functions, $Z(x)$, have sills while those of only intrinsically stationary random functions do not.

Figure 1d depicts a fourth variogram model, the Gaussian model, that is sometimes used. The Gaussian model takes the form: $\gamma_G(h) = c_0 + c(1-\exp(-h^2/a^2)$; $\gamma_G(0) = 0$. It has a practical range of $a\sqrt{3}$. It turns out that valid variogram models must induce *positive definite* variogram matrices ($\dot{\Gamma}$ in Appendix 2). This ensures that $\dot{\Gamma}^{-1}$in (A2.4) will exit. Models that are guaranteed to produce positive definite matrices, regardless of the separation distance, are themselves said to be positive definite. While it is possible to specify any functional form for a variogram and then test for positive definiteness (eg., Cressie 1991, section 2.5.2), the spherical, exponential, power, and Gaussian models are known to be positive definite and have been found to be flexible enough to cover most modeling situations.

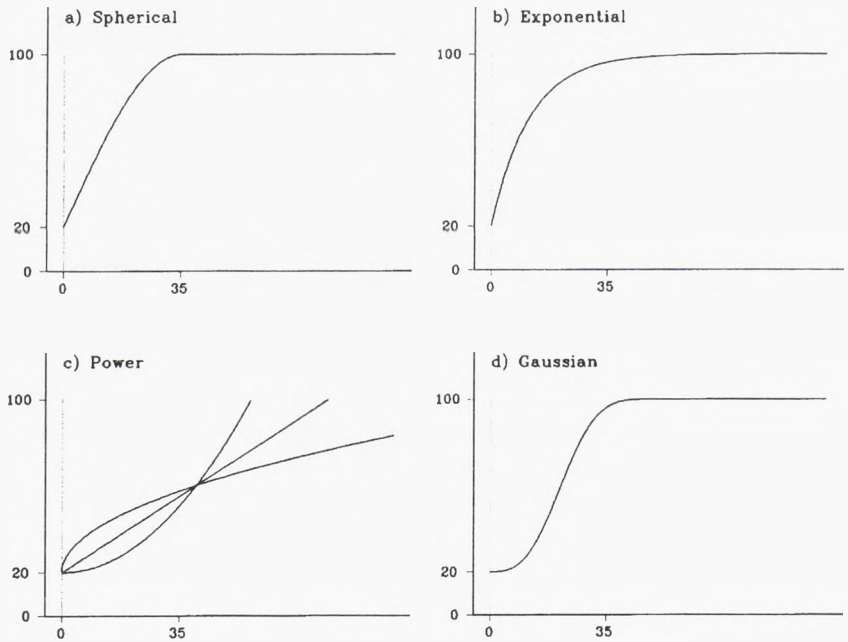

Fig. 1. Four positive definite variogram models. All models have a nugget effect of 20. The spherical, exponential, and Gaussian models each have a range of 35 and a sill of 100. Three power models, corresponding to a = 0.5, a = 1, and a = 2 are shown in panel c.

A variety of techniques are available to actually fit the covariogram model, $\gamma(h)$, to the estimated covariogram values $\{\tilde{\gamma}(h_j):, j = 1,2,...,m\}$. Pannatier (1996) implements a trial-and-error approach in the program VARIOWIN, whereby the user chooses from one of the three model forms specified above and then tweaks the model parameters, c_0, c, and a, until a good fit is achieved. His technique is largely visual, although a "goodness-of-fit" value is computed. Alternatively, regression or likelihood-based techniques can be used. In the case of regression, since the covariogram models are nonlinear in their parameters, nonlinear regression is indicated. Also, since some of the $N(h_k)$ are likely to be quite different than others, weighted regressions are often preferred over unweighted regressions. The visual, trial-and-error approach is quite effective, but the regression approach guarantees repeatability.

Finally, we note that rather than predicting the elevation (or other attribute), $\hat{z}(x_0)$, using all n data points, often only a subset of points nearest the point, x_0, in question is used. This is done partly for efficiency reasons and partly for practicality. Suppose there are $n = 1000$ data points available for kriging. Then both $\mathbf{C}$ and Γ are 1000×1000 matrices. One of these matrices must be inverted in order to estimate the kriging weights. But the inversion of such large matrices is quite time consuming, even on the most powerful of today's workstations. From a more practical point of view, points distant from x_0 will have small to negligible kriging weights. They are assumed to be zero. Consequently, only those n' points nearest x_0 are used to determine the weights, necessitating the inversion of only an $n' \times n'$ matrix. Different computer programs implement different windowing systems around x_0. Sometimes the window is of a user specified physical dimensions (rectangle or circle) centered on x_0; sometimes the user specifies n', and sometimes a combination of these windowing techniques is employed. The range is sometimes useful in determining window dimensions.

4 DATA AND MODELING PROCEDURES

As was stated at the beginning of this paper, the objectives of this article where two-fold: to introduce the kriging procedure to a new audience and to apply kriging to surface modeling. The data were $n = 806$ coordinate locations, x_i, within an area in western Texas, USA, and the elevations, z, at those points. Point locations were digitized from 7.5 minute (1/24,000 scale) US. Geological Survey maps using a GIS. The primary study area consisted of the Piloncillo quadrangle (B5-28099). This particular quadrangle was selected with the help of a program that calculated parameters of relief diversity in a 60 mile wide zone which is adjacent to the Rio Grande River. The Piloncillo and adjacent quadrangles were selected as representative of areas having low relief diversity. Benchmarked elevations were read directly from the maps. The majority of the elevation data were not benchmarked but estimated from the contour lines. The Lambert projection was used for coordinate data and the elevation data are in feet (1 inch = 2.54 cm; 1 ft = 12 inches). Elevations within the clipped quads ranges from approximately 480 feet to about 855 feet. To avoid border effects within the primary study area a buffer around the Pilo quadrangle was established that was equal to 1/2 of the area covered by a 7.5 minute quadrangle. This resulted in the 806 points total in the GIS database.

Table 1 exhibits the first seven data records. Each record consists of a point identification number (ID-Number), the Lambert easting, Lambert northing, and elevation of the point, and an elevation category. The ID-numbers originated internal to

the GIS database during digitization process. The ID-numbers and elevation categories are inconsequential to the variogram modelling and subsequent kriging. The data were extracted from the GIS database and transferred to Geostatistical Software Library (Deutsch and Journel 1992) for further analysis and processing.

Tab. 1. Example of raw data for digital elevation modeling

ID Number	X-Coordinate	Y-Coordinate	Elevation (feet)	Elevation category
1	1050107.875	662767.500	691	5
2	1050443.000	663189.812	691	5
3	1052474.250	663108.812	622	6
4	1053226.250	663276.812	624	6
5	1054871.500	663374.188	643	8
19	1054888.000	662187.188	707	6
20	1053893.500	661948.812	625	2

4.1 SELECTION OF A VARIOGRAM MODEL

Over one hundred variograms were generated in order to determine the "best" functional form, and to assess variogram specifications. For example, large lag-distances increased the "jaggedness" of the overall variogram shape and small lag distances were used to study local changes in variogram shape. On the other hand, with very small lag distances, the number of data points used to calculate some variogram values is low and, as a result the variogram values were not reliable measures of spatial continuity. The final variogram was generated with the lag distance of 500 meters. This lag distance maintains the smooth overall shape of the variogram while preserving sufficient detail. The distance tolerance was half the lag-distance, or ±250 feet.

The purpose of analyzing directional variograms is to determine the presence of anisotropy. Eight directional variograms were calculated in the following degrees north of east: 0, 30, 45, 60, 90, 120, 135 and 150 with directional tolerances of 15 degrees. For example, the 30 degree directional variogram actually considered all points within 15 to 45 degrees from a point. Figure 2a shows the 0 (East-West, squares) and 90 (North-South, triangles) degree directional variograms together with the omnidirectional variogram (circles). The other directional variograms took positions between these two (NS vs. EW) extremes. While both the East-West and the North-South variograms are clearly transitional (have sills), the East-West variogram appears to have a lower sill, indicating a weak presence of anisotropy.

While the directional variograms indicated a small amount of anisotropy, it was not significant enough to justify the building of a complex variogram model. Complex (nested) variogram models do not guarantee significant improvement in the accuracy of kriging interpolation and consequently the resulting DEM. Indeed, for distances less than about 7000 meters, all of the directional variograms were very close to one another. Because maximum window size (end of last section) used for kriging was considerably smaller than 7000 meters, the principle of parsimony in variogram modeling prevailed and a universal variogram was assumed.

The shapes of the experimental (directional and universal) variograms clearly indicated a transitional model that is quite linear for smaller distances, h. The directional variograms were calculated up to a distance of 15,000 meters. This is approximately half of the total North-South distance in the clipped maps and over half the

East-West total distance. Experimental variograms often become erratic at distances much greater than half the maximum distance between points on the map. At 15,000 meters the universal variogram has not yet reached a sill. The Gaussian model has considerably more curvature at small distances than indicated by our experimental variograms and was dismissed.

Using the trial-and-error approach to variogram model fitting implemented in VARIOWIN (Pannatier 1996), a Spherical(0, 4932, 17000), Exponential(0, 7600, 15000), and a Power(0, 0.947, 0.906) model were each fitted to the universal variogram. The numbers within parentheses are the fitted values of the c_0, c and a parameters in models (6) - (8). These models, fitted to the universal variogram, are shown in Figure 2b. Figures 2c and 2d show the fitted models on the East-West and the North-South directional variograms. The universal variogram is very well fitted by both the spherical and the exponential models and is well fitted by the power model at distances less than about 10,000 meters. Likewise, the North-South variogram is well fitted by all three models at all distances and the East-West variogram is well fitted by the three models for distances less than 7,000 meters. Since all three models are essentially the same for distances less than 7,000 meters and since the windowing technique used for kriging only considered locations less than 7,000 meters, all three fitted variogram models will produce nearly identical kriging interpolations. Therefore, from a practical point of view, it is inconsequential which variogram model is selected. From a theoretical point of view, the power model was dismissed since it does not reach a sill and all directional variograms appeared to be transitional. Recall that a transitional variogram indicates a second-order stationary process which is desired.

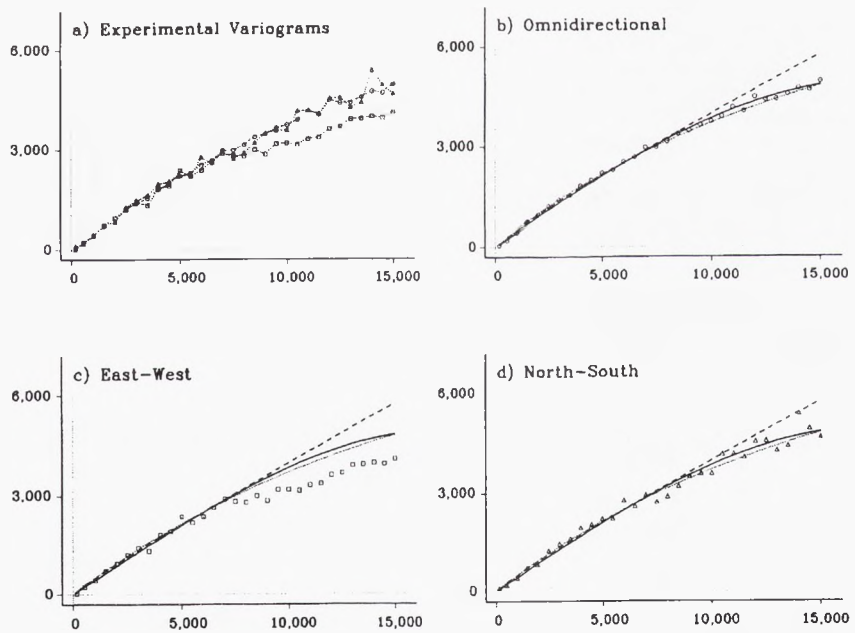

Fig. 2. Experimental and fitted variograms. The symbols for the omnidirectional, East-West, and North-South experimental variograms are circles, squares, and triangles, respectively. Solid lines = fitted spherical model, dotted lines = fitted exponential model, and dashed lines = fitted power model.

As fitted, the spherical model appears "better" than the exponential model in the sense that it reaches a sill equal to the variance of the random process ($\sigma^2 = C(0) =$ 0+4932). The exponential model approaches a sill of 7,600. Since $\gamma(h) = C(0) - C(h)$, a sill greater than $C(0)$ indicates that the covariance, and hence the correlation, between elevation values at distances larger than h = -15000 ln(1 - (4932 - 0) / 7600) is negative. In the case of elevation, negative correlations at larger distances may not be unreasonable due to undulating topography. In other situations, negative correlations may be unreasonable due to the physical characteristics of the Z-process under study. In considering the fitted spherical model, however, we see that since that sill is reached at a distance h = 17,000 meters - greater than the maximum distance, 15,000 meters, considered for modeling. As a result, there are many spherical models that will fit the experimental variograms well at distances less that 7,000 meters. An example is shown in Figure 3 where both Spherical(0, 4932, 17000) and Spherical(0, 8000, 28500) are plotted. The two variograms are essentially identical at distances less than 7,000 meters.

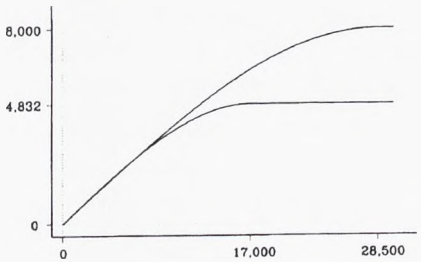

Fig. 3. Spherical (0, 17000, 4932) and Spherical (0, 8000, 28500) models.

The considerations above argue that there are many variogram models that will fit the experimental variograms equally well at distances less than 7,000 meters. Since the windowing process used in kriging will never select points this far apart, any one of these models will produce kriging interpolations that are essentially the same. We elected to use the model

$$\gamma(h) = 50 + 7{,}600(1 - \exp(-h / 15{,}000),$$

for finding the kriging weights. This model has no nugget effect, a sill of 7,600 ft^2, and the effective range of 15,000*3 meters.

4.2 KRIGING THE ELEVATION DATA

The OKB2D program in the Geostatistical Software Libarary (Deutsch and Journel 1992) was used for kriging. This is a widely available library and was chosen largely out of cost considerations. Ancillary analysis and checking was done using custom programs written in GAUSS (Aptech Systems, Inc.).

The earlier development of the kriging procedure was presented as if the elevation, $Z(x_0)$, of a single location is to be predicted. In practice, we are usually interested in predicting for a large number of locations. Indeed, for the creation of a DEM, we want to predict the elevations on a grid of locations. This presents no problem: we simply repeat the prediction process separately for each point on the grid. The grid should be fine enough to capture the topographical phenomena of interest; but at the

same time, a grid that is too fine unduly increases processing time and file sizes without providing additional information (due to the reuse of data).

Figure 4 shows the distribution of the digitized data on the *xy*-plane. The distribution is generally even. The extent in the East-West direction is approximately 24,560 meters and in the North-South direction approximately 27,740 meters. There were just over 800 digitized points. This indicates that each point nominally covers about 920 by 920 meters and that the map dimensions are nominally about 27 points in the East-West direction by 30 points in the North-South direction. In choosing the grid dimensions, we should also consider the window dimensions used for kriging each grid location. Three was chosen as the minimum number of digitized points to use for predicting the elevation at a grid location. The minimum window radius to capture at least three neighboring points was found to lie between 2,000 and 2,500 meters.

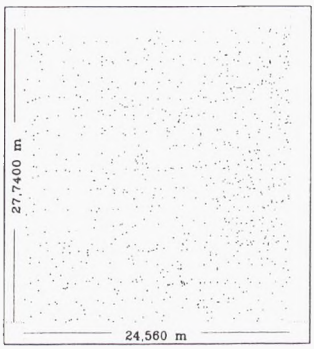

Fig. 4. Spatial distribution of sample data.

In trial runs, it was observed that for all grid locations, regardless of the grid and window sizes, the weights of only three or four points within the window were noticeably different from zero. These were usually the digitized sample locations nearest the grid location being predicted. Occasionally, one of the four nearest locations had a small weight while another, slightly further away, had a larger weight. This situation typically arises when the nearer-smaller-weighted point is redundant with other nearby points. In no case did using more than the eight sample (digitized) locations nearest a grid location significantly change the kriging weights. There were almost always eight points within the maximum window radius of 2,500 meters. Indeed, with a nominal spacing of 920 by 920 meters, we could expect between 20 and 25 sample points within a circular window of radius 2,500 meters. Consequently, we selected three as the minimum and eight as the maximum number of sample points with which to krig. These values were achieved with a kriging window radius of at most 2,500 meters.

Given the nominal spacing of 920 by 920 meters and the window size, it was decided that a grid of 50 by 50 meters was sufficiently fine for illustration purposes, and would guarantee that most adjacent grid locations were kriged with overlapping, but different, sets of sample data. The grid of 50 by 50 implies, for the Piloncillo dataset, that each cell is about 490 meters by 555 meters. This may be too coarse for some natural resource or geophysical studies using DEM results. A detailed analysis was not attempted, but is unlikely that the Piloncillo data would support a kriging

grid as fine as 100 by 100 without digitizing more sample data. Figure 6 shows the results, $\hat{z}(G)$, of kriging the Piloncillo data on a 50 by 50 grid; $G = \{x_0 : x_0$ is on the grid$\}$.

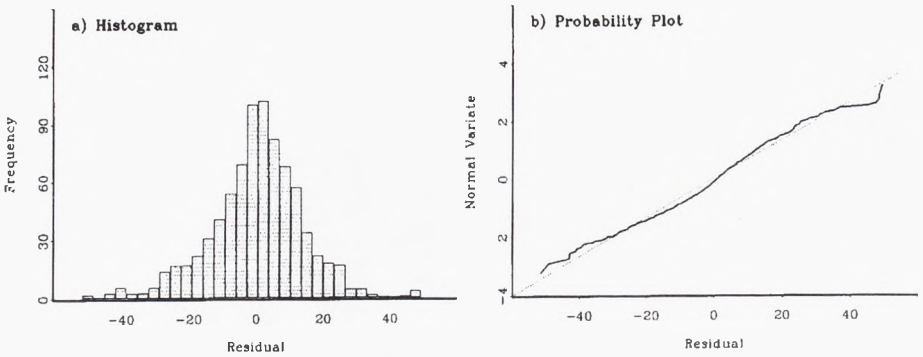

Fig. 5. Histogram and normal probability plot of the cross-validation residuals, $\hat{z}(x_i) - z(x_i)$, $i = 1, 2, ..., n$.

4.3 MODEL ACCURACY AND TESTING

One of the primary advantages of kriging over other interpolation methods is that kriging allows assessments of the goodness-of-fit of the interpolation. This is because a variance can be estimated at each grid point. Cross-validation is a leave-one-out type of analysis that is sometimes used with kriging. Under this scenario, each of the digitized sample points is left out of the data set, one at a time, and the elevation at that point is predicted using the kriging equation from the remaining 805 sample points. Since we know the elevation at each of these points and we have a prediction of the elevation at each point, we can use the differences, $\hat{z}(x_i) - z(x_i)$ for i=1,2,...,806, to evaluate model performance. Figure 5a shows a histogram of these differences and Figure 5b a normal probability plot of the differences. Data following a normal distribution will fall close to the dotted line in Figure 5b. These two plots indicate that the residuals are approximately normally distributed. The minimum residual was -50 feet and the maximum was about +50 feet; the residual mean and standard

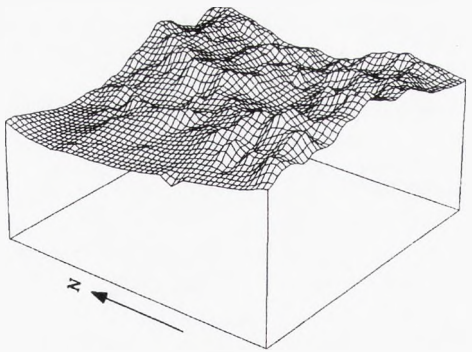

Fig. 6. Predicted elevations, $\hat{z}(G)$, for grid locations $G = \{x_0 : x_0$ on the grid$\}$.

deviation were 0.2 feet and 13.9 feet, respectively, implying a confidence interval half width of about 27 feet.

Since the residuals are approximately normally distributed, we can also compute confidence intervals for each point on our grid using $\hat{z}(x_0) \pm 1.96\hat{\sigma}_K(z(x_0))$ where $\hat{\sigma}^2_K(z(x_0))$ is a sample- and variogram model-based estimate of the kriging variance (Appendix 3). Figure 7 shows the confidence-interval-half-width surface: $1.96\hat{\sigma}_K(\hat{z}(G))$ and Figure 8 shows the result of adding the half width to the predictions: $\hat{z}(G) + 1.96\hat{\sigma}_K(\hat{z}(G))$. Given the scaling of the graphics, Figure 8 seems to indicate that the kriging variances are within acceptable limits. The range of the confidence interval half widths is 5 to 56 feet, with a mean half width of about 32 feet. The widest intervals were near the boundary of the clipped region. Within the Piloncillo quadrangle itself, the maximum was about 48 feet. This amount of prediction error may seem undue, but for grid cells of 490 by 555 meters, it is reasonable. The magnitudes of the variances and resulting confidence intervals was also consistent with the cross-validation results.

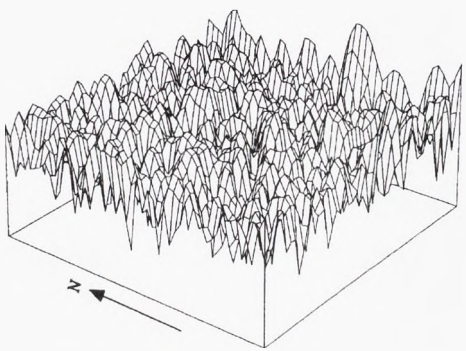

Fig. 7. Confidence interval half-widths, $1.96\hat{\sigma}_K(\hat{z}(G))$, for grid locations $G = \{x_0 : x_0 \text{ on the grid}\}$.

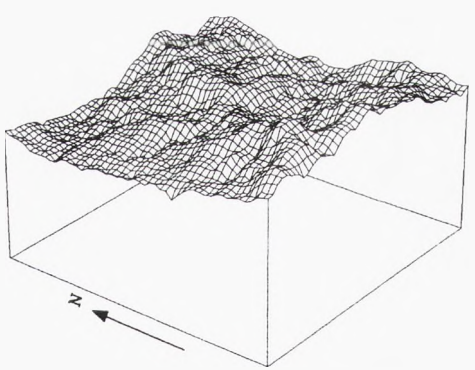

Fig. 8. Predicted plus confidence half-widths, $\hat{z}(G) + 1.96\hat{\sigma}_K(\hat{z}(G))$, for grid locations $G = \{x_0 : x_0 \text{ on the grid}\}$.

5 DISCUSSION AND CONCLUSION

Elevations on the clipped Piloncillo quadrangle were considered as random va-

riables. As such, elevation was studied and predicted by stochastic rather than deterministic models. In fact, a particular elevation is a result of a number of endogenetic and exogenetic forces such as plate tectonics, stratigraphy, chemical and physical weathering which are not fully understood. Hence, each elevation occurs on a particular place with certain probability. Stochastic models such as kriging appear to be highly suitable for modeling of many types of geo-physical and natural resource data. Kriging predicts the elevations with minimum mean square error of prediction and can take into account directionally (anisotropic) spatial dependence of the random variable. The digitized Piloncillo data were irregularly spaced. There is some indication that, for irregularly spaced data, kriging performs better than deterministic algorithms (Laslett 1994). For the density of digitized sample points, our confidence intervals indicated acceptable predictions. Better predictions could be obtained by increasing the sample size.

There is no easy answer to the question of how to construct the most accurate digital elevation models or which approach is the most efficient (Lam 1983). However, increased knowledge of geostatistical methods provides significant insights not only into our understanding of spatial distribution of elevations but also into the spatial modeling of any natural resource phenomena. Geographic Information Systems store large quantities of spatial data and have functions to manipulate spatial database. Hence, the link of spatially oriented scientific research with GIS significantly improves the efficiency of digital elevation modeling and enhances the power of spatial analysis.

Appendix 1: The Lagrange Function Expressed in Terms of Covariances of the Random Function

By expanding $E(\hat{Z}(x_0) - Z(x_0))^2$, in this Appendix we will show that the Lagrange function, (4), can be expressed in terms of covariances such as $\mathrm{Cov}(Z(x_i), Z(x_j))$. We begin by adding and subtracting $\mu = E(Z(x_0)) = E(\hat{Z}(x_0))$ and using equation (3) to subsitute for $\hat{Z}(x_0)$. For now, ignore the expectation operator:

$$(\hat{Z}(x_0) - Z(x_0))^2 = ((\hat{Z}(x_0) - \mu) - (Z(x_0) - \mu))^2 =$$

$$= \left(\left(\left(\sum_{i=1}^{n} w_{0i} Z(x_i) \right) - \mu \right) - (Z(x_0) - \mu) \right)^2 \qquad \text{(A1.1)}$$

Now, expanding the square,

$$(\hat{Z}(x_0) - Z(x_0))^2 = \left(\left(\sum_{i=1}^{n} w_{0i} Z(x_i) \right) - \mu \right)^2 -$$

$$- 2 \left(\left(\left(\sum_{i=1}^{n} w_{0i} Z(x_i) \right) - \mu \right) (Z(x_0) - \mu) \right) + E(Z(x_0) - \mu))^2. \qquad \text{(A1.2)}$$

Since $\sum w_{0i} = 1$, we can write $\mu = \mu \sum w_{0i} = \sum w_{0i} \mu$. Substituting this into (A1.2) we have

$$(\hat{Z}(x_0) - Z(x_0))^2 = \left(\left(\sum_{i=1}^{n} w_{0i} Z(x_i)\right) - \left(\sum_{i=1}^{n} w_{0i}\mu\right)\right)^2 -$$

$$- 2\left(\left(\left(\sum_{i=1}^{n} w_{0i} Z(x_i)\right) - \left(\sum_{i-1}^{n} w_{0i}\mu\right)\right)(Z(x_0) - \mu))\right) + E(Z(x_0) - \mu))^2.$$

In both the first and second terms on the right-hand side, we can now consolidate the summations. Also, since $(Z(x_0) - \mu)$ in the second term does not depend on i, we can move it inside the summation:

$$(\hat{Z}(x_0) - Z(x_0))^2 = \left(\sum_{i=1}^{n} w_{0i}(Zx_i) - \mu)\right)^2 -$$

$$- 2\left(\sum_{i=1}^{n} w_{0i}(Z(x_i) - \mu)(Z(x_0) - \mu))\right) + E(Z(x_0) - \mu))^2. \qquad \text{(A1.3)}$$

The first term in (A1.3) expands to

$$\left(\sum_{i=1}^{n} w_{0i}\,(Z(x_{0i}) - \mu)\right)^2 = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{0i}\, w_{0j}\,(Z(x_{0i}) - \mu)(Z(x_{0j}) - \mu)\right) =$$

$$= \left(\sum_{i=1}^{n}\left(w_{0i}^2\,(Z(x_{0i}) - \mu)^2 + 2w_{0i}\sum_{j\neq i}^{n} w_{0j}\,(Z(x_{0i}) - \mu)(Z(x_{0j}) - \mu)\right)\right).$$

Substituting this back into (A1.3), we arrive at

$$(\hat{Z}(x_0) - Z(x_0))^2 = \sum_{i=1}^{n}\left(w_{0i}(Z(x_i) - \mu)^2 + 2w_{0i}\sum_{j\neq i}^{n} w_{0j}((Z(x_i) - \mu(Z(x_j) - \mu))\right) -$$

$$- 2\sum_{i=1}^{n} w_{0i}((Z(x_i) - \mu)(Z(x_0) - \mu)) + (Z(x_0) - \mu)^2.$$

Finally, taking expectations

$$E(\hat{Z}(x_0) - Z(x_0))^2 = \sum_{i=1}^{n}\left(w_{0i}\,E(Z(x_i) - \mu)^2 + 2w_{0i}\sum_{j\neq i}^{n} w_{0j}E((Z(x_i) - \mu(Z(x_j) - \mu))\right) -$$

$$- 2\sum_{i=1}^{n} w_{0i}E((Z(x_i) - \mu)(Z(x_0) - \mu)) + E(Z(x_0) - \mu)^2. \qquad \text{(A1.4)}$$

But $E((Z(x_i) - \mu)(Z(x_j) - \mu))$ is the definition of the covariance, σ_{ij}, between $Z(x_i)$ and $Z(x_j)$, and $E((Z(x_i) - \mu)(Z(x_0) - \mu)) = \sigma_{i0}$ is the covariance between $Z(x_i)$ and $Z(x_0)$. Also, $E(Z(x_0) - \mu)^2 = \sigma_0^2$ is the variance at location x_0. In order to maintain a consistent notation, lets write this as a covariance: $E(Z(x_0) - \mu)^2 = E(Z(x_0) - \mu)(Z(x_0) - \mu) = \sigma_{00}$. Similarly, $E(Z(x_i) - \mu)^2 = \sigma_i^2 = \sigma_{ii}$ for $i = 1, 2, ..., n$. The variance is assumed to be the same at all locations in the domain, so we will also use the notation $\sigma^2 = \sigma_0^2 = \sigma_i^2$, $i = 1, 2, ..., n$.

Substituting these covariance expressions into (A1.4)

$$E(\hat{Z}(x_0) - Z(x_0))^2 = \sum_{i=1}^{n} \left(w_{0i}^2 \sigma_{ii} + 2w_{0i} \sum_{j \neq i}^{n} w_{0j} \sigma_{ij} \right) - 2 \sum_{i=1}^{n} w_{0i} \sigma_{0i} + \sigma_{00} .$$

So that the Lagrangian function is:

$$L(w_{0i}, \lambda : i = 1,2,...,n) = E(\hat{Z}(x_0) - Z(x_0))^2 =$$

$$= \sum_{i=1}^{n} \left(w_{0i}^2 \sigma_{ii} + 2w_{0i} \sum_{j \neq i}^{n} w_{0j} \sigma_{ij} \right) - 2 \sum_{i=1}^{n} w_{0i} \sigma_{0i} + \sigma_{00} + 2\lambda \left(\sum_{i=1}^{n} w_{0i} - 1 \right).$$

Appendix 2: Kriging Equations Expressed in Terms of Variogram Values

The definitions of the covariance, variogram, and variance of the $Z(x)$'s allow us to write:

$$\sigma_{ij} = E(Z(x_i) - \mu)(Z(x_j) - \mu) = E(Z(x_i)\, Z(x_j)) - \mu^2 \qquad \text{(A2.1)}$$

$$2\gamma_{ij} = E(Z(x_i) - Z(x_j)^2 = 2E(Z(x_i)\, Z(x_i)) - 2E(Z(x_i)\, Z(x_j)) \qquad \text{(A2.2)}$$

$$\sigma^2 = C(0) = \sigma_{ii} = E(Z(x_i - \mu)(Z(x_i) - \mu) = E(Z(x_i)Z(x_i)) - \mu^2 \qquad \text{(A2.3)}$$

From (A2.3), $E(Z(x_i)Z(x_i)) = \sigma^2 + \mu^2$. Substitute this into (A2.2) to obtain: $\gamma_{ij} = \sigma^2 + \mu^2 - E(Z(x_i)Z(x_j))$. But from (A2.1), $E(Z(x_i)Z(x_j)) = \sigma_{ij} + \mu^2$. So

$$\gamma_{ij} = \sigma^2 + \mu^2 - \sigma_{ij} - \mu^2 = \sigma^2 - \sigma_{ij},$$

or: $\sigma_{ij} = \sigma^2 - \gamma_{ij}$. Substituting each of these into the matrix **C**, we see that

$$\mathbf{C} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix} = \begin{bmatrix} \sigma^2 - \gamma_{11} & \sigma^2 - \gamma_{12} & \dots & \sigma^2 - \gamma_{1n} \\ \sigma^2 - \gamma_{21} & \sigma^2 - \gamma_{22} & \dots & \sigma^2 - \gamma_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \sigma^2 - \gamma_{n1} & \sigma^2 - \gamma_{n2} & \dots & \sigma^2 - \gamma_{nn} \end{bmatrix} =$$

$$= \sigma^2 \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 1 & 1 & \dots & 1 \end{bmatrix} - \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{bmatrix} = \sigma^2 \mathbf{J} - \Gamma$$

Similarly,

$$\mathbf{c}_0 = \begin{bmatrix} \sigma_{01} \\ \sigma_{02} \\ \cdot \\ \cdot \\ \cdot \\ \sigma_{0n} \end{bmatrix} = \begin{bmatrix} \sigma^2 - \gamma_{01} \\ \sigma^2 - \gamma_{02} \\ \cdot \\ \cdot \\ \cdot \\ \sigma^2 - \gamma_{0n} \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} - \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{0n} \end{bmatrix} = \sigma^2 \mathbf{j} - \gamma \,.$$

Substituting these expressions into the kriging equations, $\dot{\mathbf{C}}\dot{\mathbf{w}} = \dot{\mathbf{c}}_0$:

$$\begin{bmatrix} \mathbf{C} & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2 \mathbf{J} - \Gamma & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \lambda \end{bmatrix} = \begin{bmatrix} \sigma^2 \mathbf{j} - \gamma \\ 1 \end{bmatrix}$$

This allows us to write:

$$\left(\begin{bmatrix} \sigma^2 \mathbf{J} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} - \begin{bmatrix} \Gamma & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{w} \\ -\lambda \end{bmatrix} = \begin{bmatrix} \sigma^2 \mathbf{j} \\ 0 \end{bmatrix} - \begin{bmatrix} \gamma \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2 \mathbf{J} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ -\lambda \end{bmatrix} - \begin{bmatrix} \Gamma & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ -\lambda \end{bmatrix} = \begin{bmatrix} \sigma^2 \mathbf{j} \\ 0 \end{bmatrix} - \begin{bmatrix} \gamma \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2 \mathbf{J} \mathbf{w} \\ 0 \end{bmatrix} - \begin{bmatrix} \Gamma & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ -\lambda \end{bmatrix} = \begin{bmatrix} \sigma^2 \mathbf{j} \\ 0 \end{bmatrix} - \begin{bmatrix} \gamma \\ 1 \end{bmatrix}.$$

But each row of **Jw** evaluates to $\sum_{i=1}^{n} w_{0i} = 1$. Therefore, **Jw = j** and

$$\begin{bmatrix} \sigma^2 \mathbf{j} \\ 0 \end{bmatrix} - \begin{bmatrix} \Gamma & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ -\lambda \end{bmatrix} = \begin{bmatrix} \sigma^2 & \mathbf{j} \\ 0 \end{bmatrix} - \begin{bmatrix} \gamma \\ 1 \end{bmatrix}.$$

This allows us to write the kriging equations in matrix form as $\dot{\Gamma}\dot{\mathbf{w}} = \dot{\gamma}$ where

$$\dot{\Gamma} = \begin{bmatrix} \Gamma & \mathbf{j} \\ \mathbf{j}^T & 0 \end{bmatrix}, \qquad \dot{\mathbf{w}} = \begin{bmatrix} \mathbf{w} \\ -\lambda \end{bmatrix}, \qquad \text{and} \quad \dot{\gamma} = \begin{bmatrix} \gamma \\ 1 \end{bmatrix}. \tag{A2.4}$$

The similarity with the original, covariance-based, formulation of the kriging equations is obvious.

Appendix 3: The Kriging Variance

The kriging (or prediction) variance is simply the mean squared prediction error that was minimized in order to determine the kriging weights. That is,

$$\sigma_K^2(Z(x_0)) = E(\hat{Z}(x_0) - Z(x_0))^2 .$$

There are a number of different ways that the kriging variance can be expressed and estimated. We showed in Appendix 1 that it can be expressed as

$$\sigma_K^2(Z(x_0)) = \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{0i}\, w_{0j}\, \sigma_{ij}) - 2\sum_{i=1}^{n} w_{0i}\, \sigma_{0i} + \sigma_{00} .$$

Remembering that, if it exists, $\sigma_{00} = \sigma^2$ is the variance of the Z-process: $\sigma^2 = E(Z(x_0) - \mu)^2$, we can re-write this in matrix form as:

$$\sigma_K^2(Z(x_0)) = \mathbf{w}_0^T\, \mathbf{C}\mathbf{w}_0 - 2\mathbf{w}_0^T\, \mathbf{c}_0 + \sigma^2 \tag{A3.1}$$

Another expression can be found by noting, from the covariance-based formulation of the kriging equations (equations (5) and the matrix formulation immediately following), that $\mathbf{C}\mathbf{w}_0 + \mathbf{j}\lambda = \mathbf{c}_0$. This allows us to write

$$\sigma_K^2(Z(x_0)) = \mathbf{w}_0^T\, (\mathbf{c}_0 - \mathbf{j}\lambda) - 2\mathbf{w}_0^T\, \mathbf{c}_0 + \sigma^2 = \sigma^2 - (\mathbf{w}_0^T \mathbf{c}_0 + \lambda) .$$

Here we used the fact that $\mathbf{w}_0^T\, \mathbf{j} = 1$. Now noting that $\dot{\mathbf{w}}_0^T\, \mathbf{C}\dot{\mathbf{c}}_0 = [\mathbf{w}_0 \;\; \lambda]^T \begin{bmatrix} \mathbf{c}_0 \\ 1 \end{bmatrix} = \mathbf{w}_0^T\, \mathbf{c}_0 + \lambda$, we have that

$$\sigma_K^2(Z(x_0)) = \sigma^2 - \dot{\mathbf{w}}_0^T\, \dot{\mathbf{c}}_0 . \tag{A3.2}$$

Equations (A3.1) and (A3.2) are equivalent ways to express the variance when working with the covariogram formulation of kriging. In Appendix 2 we showed that $\mathbf{C} = \sigma^2\mathbf{J} - \Gamma$ and that $\mathbf{c}_0 = \sigma^2\mathbf{j} - \gamma_0$. If we substitute these terms into (A3.1) and (A3.2) we obtain expressions that are useful when working the variogram formulation of

kriging. In particular,

$$\begin{aligned}\sigma_K^2(Z(x_0)) &= \mathbf{w}_0^T\,\mathbf{C}\mathbf{w}_0 - 2\mathbf{w}_0^T\,\mathbf{c}_0 = \sigma^2 = \\ &= \mathbf{w}_0^T\,(\sigma^2\,\mathbf{J} - \Gamma)\,\mathbf{w}_0 - 2\mathbf{w}_0^T\,(\sigma^2\,\mathbf{j} - \gamma_0) + \sigma^2 = \\ &= 2\mathbf{w}_0^T\,\gamma_0 - \mathbf{w}_0^T\,\Gamma\,\mathbf{w}_0\,,\end{aligned} \tag{A3.3}$$

and

$$\begin{aligned}\sigma_K^2(Z(x_0)) &= \sigma^2 - \dot{\mathbf{w}}_0^T\dot{\mathbf{c}}_0 = \\ &= \sigma^2 - (\mathbf{w}_0^T\mathbf{c}_0 + \lambda) = \\ &= \sigma^2 - \mathbf{w}_0^T\,(\sigma^2\,\mathbf{j} - \gamma_0) - \lambda = \\ &= \mathbf{w}_0^T\gamma_0 - \lambda = \\ &= [\mathbf{w}_0\ \ \lambda]^T\begin{bmatrix}\gamma_0\\ 1\end{bmatrix} = \dot{\mathbf{w}}_0^T\,\dot{\gamma}_0\,.\end{aligned} \tag{A3.4}$$

Because it can be confusing when comparing the variance formulations used by various authors, expecially when one is first trying to understand kriging, we specifically note that the sign on the Lagrange multiplier, in our case λ, depends on how the Lagrangian function is specified for the development of the kriging equations. We added the term $2\lambda(\sum w_{0i} - 1)$. Some authors subtract this term. Whether the term is added or subtracted is inconsequential in the sense that the quantity within the parentheses is zero. Nor does it really matter whether λ is multiplied by 2. However, λ itself is estimated as part of the kriging procedure and shows up in the variance equations. An advantage of using the (A3.2) or (A3.4) expressions of the kriging variance, is that the sign of the Lagrange multiplier is automatically accounted for. Also, in Appendix 2 we saw that the Lagrange multiplier for the variogram and the covariogram formulations of the kriging procedure have opposite sign.

Finally, we should note that, as expressed, equations (A3.1) - (A3.4) are population variances. In order to use these equations, sample- and model-based estimates must be substituted for the population values. For example, to use (A3.1), we estimate the population variance, σ^2, with the sample-based $\hat{\sigma}^2 = \sum_{i=1}^{n} (z_i - \bar{z})^2 / (n-1)$, and we use the kriging equations to estimate the weights (elements of $\mathbf{w}_0$). Knowing the distances between the $\boldsymbol{x}_i$'s, i=1,...,n, and between x_0 and each of the $\boldsymbol{x}_i$'s, elements of Γ and γ_0 are estimated from the variogram model. Similarly, if we have a distance-based model of covariance, elements of $\mathbf{C}$ and $\mathbf{c}_0$ can be estimated from that model.

REFERENCES

APTECH SYSTEMS, Inc. (1992). *The GAUSS System Version 3.0.* Maple Valley, WA (Aptech Systems).

BIRKHOFF, G. and MANSFIELD, L. (1974). Compatible triangular finite elements. *Journal of Mathematical Analysis and Applications*, 47, 531.

COOPER, P.R., FRIEDMAN, D.E. and WOOD, S. A. (1987). The automatic generation of digitalterrain models from satellite images by stereo. *Acta Astronautica*, 15, 171-180.

CRESSIE, N., (1994). *Statistic for Spatial Data.* New York (Wiley).

CRESSIE, N., HAWKINS, D.M. (1980). Robust estimation of the variogram. *Mathematical Geology,* 12, 127.

DEUTSCH, C., JOURNEL, A.G. (1992). *Geostatistical Software Library.* New York (Oxford University Press).

GILES, P. T., ChAPMAN, M. A., FRANKLIN, S. E. (1994). Incorporation of a digital elevation model derived from stereoscopic satellite imagery in automated terrain analysis. *Computers and Geosciences*, 20, 441-450.

ISAAKS, E.D., SHRIVASTAVA, R.M. (1989). *Applied Geostatistics.* New York (Oxford University Press).

JOURNEL, A. G., HUIJBREGTS, C.H.J. (1978). *Mining Geostatistics*. New York (Academic Press).

LAM, N. S. (1983). Spatial interpolation methods: a review. *American Cartographer,* 10, 129 - 220.

LASLETT, G. M. (1994). Kriging and Splines: An empirical comparison of their predictive performance in some applications. *Journal of the American Statistical Association,* 89, 391-421.

MATHERON, G. (1963). Principles of Geostatistics, *Economic Geology.* 58, 1246 -1260.

MILLER, C. L., LA FLAMME, A. R. (1958). The digital theory and application. *Photogrammetric Engineering and Remote Sensing*, 7, 433-442.

NELSON, E. J., JONES, N. L. (1995). Reducing elevation roundoff errors in digital elevation models. *Journal of Hydrology*, 169, 36-49.

PANNATIER, Y. (1996). *VARIOWIN Software for Spatial Data Analysis in 2D.* New York (Springer).

PAVLÍK, Z., KUEHN, K. (1981). *Úvod do kvantitativních metod pro geografy.* Praha (Univerzita Karlova, Státní pedagogické nakladatelství).

Peter Šiška, Marian Eriksson, Robert Maggio

TVORBA DIGITÁLNEHO MODELU RELIÉFU POMOCOU KRIGINGOVEJ INTERPOLAČNEJ METÓDY

Digitálne výškové modely sú aproximáciou trojrozmerného zemského povrchu. Ich presnosť je sčasti ovplyvnená disekciou reliéfu. Rovinné geomorfologické útvary sú relatívne presnejšie interpretovateľné digitálnymi výškovými modelmi. Na druhej strane značne členité časti zemského povrchu, ako napríklad vysokohorské štruktúry, zlomové línie, erózne brázdy, kaňony, členité vrásové štruktúry, atď., sú príčinou nepresnosti pri vytváraní digitálnych terénnych modelov. Predložená práca podrobne vysvetľuje a zároveň aj uplatňuje princípy krigingovej (optimálnej) interpolačnej metódy pri konštrukcii digitálneho výškového modelu rovinného typu. Kriging je stochastickou metódou a vyznačuje sa minimálnou strednou kvadratickou chybou predikcie.

Zákonitosti v prírodnej sfére, vrátane procesov, ktoré sa dotýkajú dynamiky zemskej kôry, sa často nemôžu plne opísať a vyjadriť deterministickými modelmi. Predložená práca dôkladne analyzuje priestorovú interpolačnú metódu, ktorá je podobná skupine tzv. BLUE estimátorov v lineárnej regresii. Kriging má dobré predpoklady kvantitatívne vyjadriť zákonitosti priestorového rozloženia geografických javov v prírodnej sfére a zároveň poukazuje na možnosti využitia digitálnych modelov v rámci geografických informačných systémov, ktoré podporujú výskumné úlohy praktického zamerania.

Obr. 1 Štyri pozitívne definitné modely rozptylogramov (variogramov). Všetky modely majú zhodne chybu merania 20 jednotiek. Sférický, exponenciálny a gaussov model majú rozsah 35 a asymptotu 100. Tri mocninové modely, ktoré zodpovedajú exponentom a = 0,5, a = 1, a a = 2 sú zobrazené na paneli c.

Obr. 2. Experimentálne a fitované rozptylogramy. Grafické symboly pre multismerový a jednotlivé smerové experimentálne rozptylogramy (východ-západ a sever-juh) sú vyjadrené v podobe kruhov, štvorcov a trojuholníkov v uvedenom poradí. Plné čiary reprezentujú sférický

model, bodkovaná čiara reprezentuje exponenciálny model a prerušovaná čiara aproximovaný mocninový model.

Obr. 3. Sférické modely (0, 17000, 4932 a 0, 8000, 28500).

Obr. 4. Priestorové rozloženie vzoriek.

Obr. 5. Histogram a normálna pravdepodobnostná sieť rezíduí $\hat{z}(x_i) - z(x_i)$, i = 1,2, ..., n.

Obr. 6. Modelom predpovedané výškové hodnoty $\hat{z}(G)$ pre polohy buniek $G = \{x_0 : x_0$ leží na bunkovej sieti$\}$.

Obr. 7. Polovičná šírka intervalu spoľahlivosti $1.96\hat{\sigma}_K(\hat{z}(G))$ pre polohy buniek $G = \{x_0 : x_0$ leží na bunkovej sieti$\}$.

Obr. 8. Hodnoty predpovedané modelom a polovičná šírka ich intervalu spoľahlivosti $\hat{z}(G) + 1.96\hat{\sigma}_K(\hat{z}(G))$ pre polohy buniek G = $\{x_0 : x_0$ leží na bunkovej sieti$\}$.

Translated by Peter Šiška