THE FRONTIERS OF
STATISTICAL SCIENTIFIC THEORY & INDUSTRIAL APPLICATIONS
(Volume II of the Proceedings of ICOSCO-I,
The First International Conference on Statistical Computing)

ON VARIOGRAM ESTIMATION

Donald E. Myers
Department of Mathematics
University of Arizona
Tucson, Arizona 85721
U.S.A.

SYMPTOTIC ABSTRACT

Geostatistics is concerned with the analysis of spatially distributed and correlated data such as arises in mining, hydrology, soil physics, geotechnics, and environmental monitoring and assessment. All of the various forms of kriging utilize the variogram to determine the weight vector. Because the distribution of the sample variogram is known only under very restrictive assumptions and because the (negative of the) variogram must be conditionally positive definite, adequate methods of statistical inference are as yet lacking. Common practice incorporates both subjective judgement by the user and some form of cross-validation. The efficacy of the cross-validation is in turn dependent on the robustness of the kriging estimator with respect to the variogram.

The process of variogram estimation is examined beginning with the determination of sampling schemes. Properties of the sample variogram are compared with various proposed robust estimators including the effects of non-uniform sampling patterns. The difficulties resulting from the presence of drift or anisotropies are examined.

Key words: variogram, cross-validation, positive definiteness, radial basis functions.

1. INTRODUCTION

Geostatistics is a branch of stochastic processes and statistics primarily concerned with estimation problems in hydrology, mining engineering, soil physics, geosciences and more recently environmental monitoring. In particular it is concerned with the use of spatially correlated data. The spatial correlation is quantified by the semi-variogram, hereafter simply called the variogram. The original terminology, semi-variogram, was utilized by Matheron because of a one-half factor, i.e., half of a variance. Because only one of the two quantities (semi-variogram vs variogram) is actually needed a number of authors have used variogram in lieu of semi-variogram for simplicity and that practice is followed here. While the practice is not universal, it is quite common particularly in the English literature. The entire process of variogram estimation/modelling is reviewed in subsequent sections, including the problem of the lack of adequate statistical tests.

Let $Z(x)$ be a random function defined in 1, 2 or 3 space. $Z(x)$ might represent ore grade in a mineral deposit, percent sulfur in coal, hydraulic conductivity or depth to water table. Given data $Z(x_1), Z(x_2), \ldots, Z(x_n)$; one objective may be to estimate

$$Z_V = \frac{1}{V}\int_V Z(x)dx \tag{1}$$

where V is a line segment, area or volume. Assuming $Z(x)$ satisfies the Intrinsic Hypothesis (see (5), (6)) Matheron (1965) defined the variogram as follows:

$$\gamma(h) = \frac{1}{2}\mathrm{Var}[Z(x+h) - Z(x)] \tag{2}$$

Knowledge of the variogram is then sufficient to determine the coefficients in the Ordinary kriging estimator

$$Z_V^* = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{3}$$

so that Z_V^* is unbiased and has minimum error variance.

If $Z(x)$ is second order stationary then

$$\gamma(h) = \sigma^2 - \sigma(h) \tag{4}$$

where $\sigma(h)$ is the (auto)covariance. Note that the variogram may exist when the covariance does not. In particular if the variogram is unbounded the covariance does not exist. For example Brownian motion has a variogram but not a stationary autocovariance. In general estimation/modelling of the variogram is simpler than for the corresponding covariance function. In contrast to some techniques which allow the use of a sample covariance, the above estimator requires the use of a theoretical model and this is the source of part of the difficulties in variogram estimation. The theoretical variogram must satisfy two conditions

i. Conditional Positive Definiteness

$$-\iint d\mu(x)\gamma(x-y)d\lambda(y) > 0$$

for any non-zero measures λ, μ with finite support and

$$\int d\mu(x) = 0, \int d\lambda(y) = 0$$

ii.

$$\lim_{|h|\to\infty} \frac{\gamma(h)}{|h|^2} = 0$$

Let $\mathcal{A}$ denote the set of valid variograms. $\mathcal{A}$ is closed under addition and multiplication by positive constants. A number of known valid models are listed in Appendix A. Matheron (1973) has given a Bochner type representation theorem for variograms. The Radial Basis Functions as described by Micchelli (1986) are in fact variograms or generalized covariances.

2. THE KRIGING ESTIMATOR

The cross-validation process utilizes properties of the kriging estimator and we include a brief review of its properties. $Z(x)$ is assumed to satisfy the Intrinsic Hypothesis (Matheron, 1971)

$$i. \quad E[Z(x+h) - Z(x)] = 0 \tag{5}$$

for all x, h

$$ii. \quad \gamma(h) = \frac{1}{2}\mathrm{Var}[Z(x+h) - Z(x)] \tag{6}$$

depends only on h

The kriging equations are

$$\sum_{i=1}^{n} \gamma(x_i - x_j)\lambda_i + \mu = \gamma(x_0 - x_j) \tag{7}$$

$$j = 1, \ldots, n$$

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{8}$$

The kriging estimator may also be written in the dual form

$$Z^* = \sum_{i=1}^{n} b_i \gamma(x_0 - x_i) + a \tag{9}$$

where

$$\sum_{i=1}^{n} b_i \gamma(x_i - x_j) + a = Z(x_j) \tag{10}$$

$$j = 1, \ldots, n$$

$$\sum_{i=1}^{n} b_i = 0 \tag{11}$$

The kriging estimator is exact, that is, if $Z(x_i)$ is a data value then $Z^*(x_i) = Z(x_i)$ if $Z(x_i)$ is retained in the data set.

3. VARIOGRAM ESTIMATORS

A. The Sample Variogram

If $Z(x)$ satisfies the Intrinsic Hypothesis then

$$\gamma(h) = \frac{1}{2}E[Z(x+h) - Z(x)]^2 \tag{12}$$

and hence a logical estimator would be

$$\gamma^*(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x_i+h) - Z(x_i)]^2 \tag{13}$$

where $N(h)$ is the number of pairs $Z(x_i+h), Z(x_i)$. Because $\gamma^*(h)$ is essentially a sample mean it has all the disadvantages commonly associated with the sample mean, in particular it is non-robust. Since in general no assumptions are made concerning the distribution of $Z(x)$ and hence the sampling distribution of $\gamma^*(h)$ will not be known. Davis and Borgman (1978, 1982) have shown under reasonable conditions that a central limit property applies

$$\frac{\gamma^*(h) - \gamma(h)}{\sigma_{\gamma(h)}} \to N(0,1) \tag{14}$$

as $N(h) \to \infty$. In addition they have tabulated the sample distribution for $\gamma^*(h)$ for certain variogram models assuming $Z(x)$ is multivariate normal by using Fourier Transforms. It is easy to see that under a Normality assumption, $\gamma^*(h)$ is approximately chi-square. It is however, not sufficient to obtain confidence intervals for each h although they may be informative.

Aside from the non-robust nature of $\gamma^*(h)$ a number of other difficulties can arise in practice. In 2 or 3 dimensional space, h must be treated as a vector; i.e. $\gamma^*(h)$ is a function of distance r and direction θ. In order to identify possible anisotropies $\gamma^*(h)$ must be computed for a range of discussed later. In most instances a preliminary visual fit coupled with a form of cross-validation is the preferred method.

When distance classes and angle windows are used it is often necessary to experiment with the width of the classes and windows to obtain a reasonably smooth plot. This use also complicates the problem of designing a sample pattern to optimize the sample variogram. For regular or near-regular grids, the number of pairs is small for short distances, largest for intermediate distances and decreases as the distance exceeds half the diameter of the region of interest. Warrick and Myers (1987) have shown that it is possible to force the number of pairs for each distance/angle class to approximate a prescribed distribution. It is desirable to have a larger number of pairs for short distances since it is that portion of the variogram that is most critical. It should be noted that an optimal sampling pattern for variogram estimation will be quite different from that of an optimal pattern for the subsequent kriging with a known variogram.

By definition $\gamma(0) = 0$ whereas it is often found that

$$\lim_{|h| \to 0} \gamma^*(h) > 0 \tag{19}$$

This discontinuity is incorporated into the variogram model and is known as the Nugget affect. It may reflect several causes. Since there is always a shortest intersample distance, γ^* is not known for shorter distances and hence γ is not directly estimated in this interval. In terms of the kriging variance, modelling this uncertainty as a Nugget effect is the conservative approach. When interpreted in terms of the covariance as in eq. (4), the Nugget effect is seen to include data uncertainty, for example measurement errors.

If $Z(x)$ does not satisfy the Intrinsic Hypothesis, a weaker formulation would be given by

$$Z(x) = Y(x) + m(x) \tag{20}$$

$$E[Z(x)] = m(x) \tag{21}$$

where $Y(x)$ satisfies the Intrinsic Hypothesis. In this, known as universal kriging, the estimator remains unchanged but equations (7, 8) become

$$\sum_{i=1}^{n} \lambda_i \gamma(x_i - x_j) + \sum_{k=0}^{p} \mu_k f_k(x_j) = \gamma(x_0 - x_j) \qquad (7')$$

$$j = 1, \ldots, n$$

$$\sum_{\ell=1}^{n} \lambda_i f_k(x_i) = f_k(x_0) \qquad (8')$$

$$k = 0, \ldots, p$$

The variance of the error of estimation is then larger to reflect the uncertainty about $m(x)$. In the above equations, it is assumed that

$$m(x) = \sum_{k=0}^{p} a_k f_k(x)$$

where $f_0, \ldots, f_p$ are linearly independent, usually taken to be polynomials. These functions are assumed known but $a_0, \ldots, a_p$ are unknown. Unfortunately new difficulties arise in estimating/modelling the variogram for $Z(x)$. If $m(x)$ is a first degree polynomial in the coordinates of x then the sample variogram for Z includes a quadratic term, i.e.

$$\gamma_Z^*(h) = \gamma_Y^*(h) + \frac{1}{2N(h)} \sum [Y(x_i + hy) - Y(x_i)][m(x_i + h) - m(x_i)] + \frac{1}{2N(h)} \sum [m(x_i + h) - m(x_i)]^2 \qquad (22)$$

Since a valid variogram grows less rapidly than a quadratic, quadratic or faster growth in the sample variogram is usually taken as evidence of nonstationarity. If $m(x)$ is fitted by least squares and the sample variogram is computed from the residuals there will be a bias, as was noted by Matheron (1971) and Sabourin (1976). Neuman and Jacobson (1983) proposed an iterative method for determining the order of $m(x)$ and at the same time estimating the variogram. Cressie (1986a) has proposed the use of

Median Polish for estimating the drift and then modeling the variogram on the residuals followed by kriging of the residuals and subsequent addition back of the drift. While this method does not require a complete grid it does depend on a grid structure for the sample locations. Inasmuch as the Neuman and Jacobson technique does not completely remove the bias resulting from least squares fitting of the drift and Cressie's method requires a grid structure the problem of variogram fitting in the presence of drift is not fully resolved. Matheron's theory of intrinsic random functions (1973) provides another approach by using generalized covariances. The problem of estimation of generalized covariances is quite different from that of estimating variograms. The basic idea is that increments will filter out polynomials. Let $\lambda_1, \ldots, \lambda_n$ satisfy the conditions given in Eq. (8′) then

$$Z(\lambda) = \sum \lambda_i Z(x_i)$$

is a generalized increment. The error of estimation of the kriging estimator is such a generalized increment. $K(h)$ is called a Generalized Covariance if

$$\mathrm{Var}\, Z(\lambda) = \sum \lambda_i \lambda_j K(x_i - x_j)$$

for generalized increments. Generalized covariances provide an alternative way of dealing with non-stationarity. Variograms are zero-order generalized covariances. The software package BLUEPACK, (Delfiner, 1976) incorporates an automatique structure recognition routine for generalized covariances. Unfortunately in some instances, anisotropies and non-stationarity may be difficult to distinguish from each other.

Some erratic or unusual features of the sample variogram may be explainable by noting unusual characteristics of the sample location pattern or the histogram of the data values. If the location pattern is long and narrow or irregularly shaped, anisotropies may appear to be present. If the histogram of the data is strongly bi-modal there may be jumps in the variogram.

All of the models listed in Appendix A are characterized by a few parameters and many of these are geometrically interpretable. For example, the Spherical Model is determined by the Nugget C_0, the sill $C_0 + C_1$ and the range a, each of which can be visually estimated from the graph of the sample variogram.

Kitanidis (1983) has proposed estimating the variogram parameters by restricted maximum likelihood estimation, this requires a multi-variate Normal assumption. Samper (1986) utilized an alternative approach to maximum likelihood estimation by coupling it with cross-validation and assuming the residuals from cross-validation are multivariate normal.

B. Other Estimators

Because of the sensitivity of the sample variogram to outliers, several other estimators have been proposed. Armstrong and Delfiner (1980) used median and quartile estimators. Cressie and Hawkins (1980) proposed the square root of the fourth power of first order differences. Although this estimator is biased, they computed the bias adjustment under a Normality assumption. Essentially all of the methods proposed thus far still utilize first order differences and hence many of the difficulties that arise in connection with the use of the sample variogram are present with alternative estimators.

C. Transformations

It is common in many forms of statistical analysis to apply a non-linear transformation, for example square root or logarithmic for Analysis of Variance. While such transformations may result in better behaved sample variograms, with the exception of the logarithmic in the case of a log-Normal distribution, it is generally difficult to relate the variogram of the transformed data to the variogram of the original data or to utilize the variogram of the transformed data for the kriging process. Cressie (1985b) has shown that in the case of second order stationarity with known mean the δ method can be used to approximate the variogram of the transformed variable in terms of the known mean and the variogram of the original. The approximation assumes that the transformation has a continuous second derivative and is not applicable under the weaker Intrinsic Hypothesis.

Under the assumption of multivariate log-normality the variogram for $Z(x)$ is given by

$$\gamma(h) = M^2[e^{\sigma^2}(1 - e^{-\gamma_Y(h)})] \tag{23}$$

where

$$M = E[Z(x)], \sigma^2 = \text{Var}[Y(x)] \tag{24}$$

In the case of a log-normal distribution or for hydrologic parameters such as hydraulic conductivity, a logarithmic transformation may help to avoid the effect of the skewed distribution. Two approaches may be followed: (1) transform the data, model the variogram and then re-transform the variogram (2) transform the data, model the variogram, krige and then re-transform the krige values. In the latter case a bias adjustment is necessary. For further details, see Rendu (1979), Journel (1980) or Dowd (1982). The basic problem is as follows: Let

$$Y(x) = lnZ(x) \tag{25}$$

$$Y^*(x_0) = \sum_{i=1}^{n} \lambda_i Y(x_i) \tag{26}$$

where the weights $\lambda_1, \ldots, \lambda_n$ are found using the variogram for $Y(x)$. Unfortunately $\exp[Y^*(x_0)]$ is a biased estimator of $Z(x_0)$. To compute the bias correction factor one considers four cases; Simple and Ordinary, Punctual and Block kriging as shown in Journel (1980).

In the case of non-stationarity the use of non-linear transformations will complicate the problem of simultaneous estimation/modelling of the drift and the variogram. In particular in the case of a logarithmic transformation the drift becomes multiplicative.

D. Regularized Variograms

For many geostatistical variates, the value represents an average over a small volume rather than a point value. In turn the objective may be to estimate a spatial average rather than to interpolate a point function. Non-point support for the data leads to two problems. The first concerns the estimation of the variogram and the second changes in the kriging equations. The latter requires only a minor change in the kriging equations (7, 8), the right hand side term $\gamma(x_0 - x_j)$ is replaced by $\bar{\gamma}(V, x_j)$ where

$$\bar{\gamma}(V, x_j) = \frac{1}{V} \int_V \gamma(x - x_j) dx \tag{27}$$

The same substitution is made in the Dual form of the kriging estimator. For the former let v_x be an elemental volume "centered" at x, v_{x+h} the translate by the vector h. Denote by

$$Z_{v_x} = \frac{1}{v} \int_{v_x} Z(y) dy \tag{28}$$

and

$$Z_{v_{x+h}} = \frac{1}{v} \int_{v_{x+h}} Z(y) dy = \frac{1}{v} \int_{v_x} Z(y + h) dy \tag{29}$$

and

$$\gamma_v(h) = \frac{1}{2} \text{Var}[Z_{v_{x+h}} - Z_{v_x}] \tag{30}$$

if $Z(x)$ satisfies the Intrinsic Hypothesis then

$$\gamma_v(h) = \bar{\gamma}(v, v_h) - \bar{\gamma}(v, v) \tag{31}$$

where $\bar{\gamma}(u, v) = \frac{1}{uv} \int_u \int_v \gamma(x - y) dx dy$. The most obvious change from the point to the non-point variogram is the decrease in the sill, this is comparable to the reduction in variance with an increase in sample size. The models listed in Appendix A are all point models, hence if the data used to estimate the variogram is non-point then one must exercise care in fitting to a standard model. This complication also arises in the use of cross-validation.

As noted above, the variogram value must be replaced by an average value when kriging a spatial average. Dunn and Alldredge (1982) have proposed combining the steps of estimation of the variogram and (numerically) integrating to obtain the average variogram. Unfortunately the linear/polygonal approximation process does not ensure the use of a valid variogram model.

4. ROBUSTNESS/CONTINUITY

It is not sufficient to approximate the values of the variogram at a finite number of points, i.e. the positive definiteness condition must be satisfied, hence it is necessary to approximate the variogram by valid models. See for example Dunn (1983) and Myers (1984). Since the principal use of the variogram is to determine the weights in the kriging estimator one way to quantify the proximity of two variograms is by the change in the weight vector. Diamond and Armstrong (1984) defined a neighborhood for isotropic variograms

$$N_\gamma(\delta) = \{g | g \varepsilon \mathcal{A}, \sup_{0 \le r} |\frac{g(r)}{\gamma(r)} - 1| < \delta\} \tag{32}$$

and then obtained bounds for $\| X \|$ where $X = [\lambda_1, \ldots, \lambda_n, \mu]^T$, the solution vector of the kriging system, in the case of an ℓ_1, ℓ_2, or ℓ_∞ norm. Myers (1985, 1986) generalized the neighborhood definition and considered two additional definitions obtaining bounds for $[\lambda_1, \ldots, \lambda_n]^T$. The generalized neighborhood is given by

$$N_\gamma(\delta, r) = \{g | g \epsilon \mathcal{A}, \sup_{0 \le |h| \le r} |\frac{g(h)}{\gamma(h)} - 1| < \delta\} \tag{33}$$

The second is given by

$$M_\gamma(\delta, \gamma_1) = \{g | g(h) = \gamma(h) + \varepsilon \gamma_1(h), 0 \le \varepsilon < \delta; \gamma_1 \varepsilon \mathcal{A}\} \tag{34}$$

and the third by

$$NM_\gamma(\epsilon, \delta) = \{g | g \epsilon \mathcal{A}, \sup_{0 \le |h| \le \epsilon} |\gamma(h) - g(h)| < \delta\} \tag{35}$$

The first and third correspond to continuity of the weight vector with respect to a change in the variogram. The second corresponds to Frechet differentiability of the weight vector. The kriged value and the kriging variance are both determined by the weight vector. Note that the weight vector is not affected by changes in the data values whereas the kriged value is, there is a distinction then between the continuity of the weight vector and subsequently the kriged value with respect to the variogram and the robustness of the kriged value with respect to the data values.

5. CROSS-VALIDATION

Since the kriging estimator is exact it suggests the following procedure: sequentially, one at a time, delete a data value and krig the value for that location using the remaining data. If the variogram model adequately reflects the spatial correlation implicit in the data set then the kriged values should be close to the observed values. This "closeness" can be characterized in a number of ways, the following statistics are usually computed:

(a) $\frac{1}{n}\sum_{i=1}^{n}[Z(x_i) - Z^*(x_i)]$

(b) $\frac{1}{n}\sum_{i=1}^{n}[Z(x_i) - Z^*(x_i)]^2$

(c) $\frac{1}{n}\sum_{i=1}^{n}[\frac{Z(x_i)-Z^*}{\sigma_i}(x_i)]^2$

(d) Sample correlation of $Z(x), Z^*(x)$

(e) Sample correlation of $Z^*(x), (Z(x) - Z^*(x))/\sigma_i$.

Theoretically the expected values of; (a) should be zero, (b) should be small, (c) should be one, (d) should be close to one (this depends on the Lagrange multipliers μ_k, $k = 0, \cdots, p$ as in (7')), (e) should be close to zero (this also depends on the Lagrange multipliers). In addition the histogram of the Normalized errors is plotted and a list compiled for the sample locations with large normalized errors. The latter is frequently useful for identifying outliers, suspicious data or abnormalities of some other kind. Ideally all of the above conditions should be satisfied in practice an improvement in one statistic may degrade another. Temporary suppression of the data locations with large normalized errors prior to recomputation of the sample variogram may produce significant improvements.

It was noted earlier that weighted least squares estimation of the variogram is not optimal, this is because the loss function is not directly related to the continuity of the kriging estimator. Any valid variogram model will result in a minimal variance estimator since the minimized variance is computed using that variogram. The least squares estimator will work best for estimating those parameters appearing linearly in the model whereas identification of the model type(s) appearing in the nested structure is more important. For example if a linear model is used then neither parameter has any effect on the weight vector hence "optimal" estimation of the parameters is of little importance. Moreover in many cases the geometry of the sample location pattern may be more important than the variogram model in determining the weight vector.

By using the Dual form of the kriging estimator another form of cross-validation is possible. Recall that the Dual form is

$$Z^*(x_0) = \sum b_i \gamma(x_0 - x_i) + a \tag{36}$$

and $E(a) = E[Z(x)]$. By analogy with regression methods consider the following statistics

$$\frac{1}{n}\sum_{i=1}^{n}(Z(x_i) - a)^2$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{[Z(x_i) - a]^2}{\mathrm{Var}(a)}$$

The latter should be close to 1. The advantage of these statistics over those given earlier is that all the data set can be used, i.e. no jack-knifing is necessary.

6. THE MULTIVARIATE PROBLEM

If the random function $Z(x)$ is replaced by a random vector function $\bar{Z}(x) = [Z_1(x), \ldots, Z_m(x)]$ then as shown in Myers (1982, 1984) the variogram is replaced by a variogram matrix $\bar{\gamma}(h)$ where the diagonal entries are the variograms of the separate components and the off-diagonals are cross-variograms, that is,

$$\bar{\gamma}(h) = [\gamma_{ij}(h)] \tag{37}$$

$$\gamma_{ii}(h) = \frac{1}{2}\mathrm{Var}[Z_i(x+h) - Z_i(x)] \tag{38}$$

$$\gamma_{ij}(h) = \frac{1}{2}\mathrm{Cov}(Z_i(x+h) - Z_i(x), Z_j(x+h) - Z_j(x)) \tag{39}$$

the positive definite condition for the variogram matrix is given by the following

$$-\mathrm{Trace} \sum_{i=1}^{n} \sum_{j=1}^{n} \Gamma_i^T \bar{\gamma}(x_i - x_j) \Gamma_j \geq 0 \tag{40}$$

for any points $x_1, \ldots, x_n$ and matrices $\Gamma_1, \ldots, \Gamma_n$ with

$$\sum_{i=1}^{n} \Gamma_i = 0 \tag{41}$$

One consequence of this as seen in Myers (1984) is that cross-variograms can only be validated in conjunction with the corresponding variograms. While the sample cross-variogram

$$\gamma_{ij}^*(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [Z_i(x_k+h) - Z_i(x_k)][Z_j(x_k+h) - Z_j(x_k)] \tag{42}$$

estimates the cross-variogram $\gamma_{ij}(h)$ the modelling process is more complicated It is easily seen that

$$\gamma_{ij}(h) = \frac{1}{2}[\gamma_{ij}^+(h) - \gamma_{ii}(h) - \gamma_{jj}(h)] \tag{43}$$

where $\gamma^{+}_{ij}(h)$ is the variogram for $Z_i(x) + Z_j(x)$. This suggests a procedure for modelling the cross-variogram. The variograms for $Z_i(x), Z_j(x), Z_i(x) + Z_j(x)$ are separately estimated and modelled including cross-validation then using Eq. (38), $\gamma_{ij}(h)$ is obtained. A further adjustment may be necessary to ensure that the Cauchy-Schwartz condition is satisfied, i.e.

$$|\gamma_{ij}(h)| \leq [\gamma_{ii}(h)\gamma_{jj}(h)]^{\frac{1}{2}} \tag{44}$$

Finally by using Co-kriging (Carr and Myers 1985), the variogram matrix can be cross-validated.

7. FINAL COMMENTS

As yet without strong distributional assumptions statistical tests for evaluating variogram estimators are still lacking. In light of the coincidence between the dual form of the kriging estimator and interpolation by Radial Basis Functions it may be desirable to search for a deterministic characterization of the efficacy of variogram estimators. Diamond and Armstrong (1984) have obtained bounds for the change in the solution vector of the system (7), (8) [or (7'), (8')] for changes in the variogram characterized by a particular choice of a neighborhood as well as for changes in the order of the drift or the sampling pattern. Myers (1985), (1986) extended these results to exclude the Lagrange Multiplier(s) since the latter do not explicitly appear in the estimator and gave results using a generalization of the neighborhood used by Diamond and Armstrong as well as two new neighborhood definitions. In none of these three papers were the results explicitly extended to a characterization of the efficacy of variogram estimators.

NOTICE

Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through Cooperative Agreement 811938-02 to University of Arizona, it has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

References

Armstrong, M. and Delfiner, P., (1980). Towards a more Robust Variogram. N-671, Centre de Geostatistique, Fontainebleau.

Armstrong, M. and Jabin, R., (1981). Variogram models must be positive definite. Mathematical Geology, **13**, No. 5, 455-460.

Armstrong, M., (1984). Commoner Problems seen in Variograms. Mathematical Geology, **16**, No. 3, 305-313.

Carr, J., Myers, D. E. and Glass, C., (1985). Cokriging - A computer program. Computers and Geosciences, **11**, No. 2, 111-127.

Cressie, N. and Hawkins, D., (1980). Robust Estimation of the Variogram. Mathematical Geology, **12**, No. 2, 112-125.

Cressie, N. (1985a). Fitting Variogram models by weighted least squares. Mathematical Geology, **17**, No. 5, 563-586.

Creassie, N., (1985b). When are relative variograms useful in geostatistics. Mathematical Geology, **17**, 7, 693-702.

Cressie, N. (1986). Kriging Non-Stationary data. J.Amer.Stat.Ass., **81**, No. 395, 625-634.

Davis, B. M. and Borgman, L. E., (1978). Some Exact sampling distributions for variogram estimators. Mathematical Geology, **11**, No. 6, 643-653.

Davis, B. M. and Borgman, L. E. (1982). A note on the Asymptotic distribution of the sample variogram. Mathematical Geology, **14**, No. 2, 189-194.

Delfiner, P., (1976). Linear Estimation of Non-Stationary phenomena. in Advanced Geostatistics for the Mining Industry. Guarascio et al (eds), D. Reidel Publishing, Dordrecht, 49-68.

Diamond, P. and Armstrong, M. (1984). Robustness of Variogram and Conditioning of Kriging matrices. Mathematical Geology, **16**, No. 8, 809-822.

Dowd, P. (1982). Log Normal Kriging - The General Case. Mathematical Geology, **14**, No. 5, 475-500.

Dunn, M. R. and Alldredge, J. R., (1982). General Polygonal Variogram Functions: Evaluation of Estimation Variance Integrals. Mathematical Geology, **14**, No. 1, 77-85.

Dunn, M. R. (1983). A Simple Sufficient Condition for a Variogram model to yield positive variances under restrictions. Mathematical Geology, **15**, No. 4, 553-564.

Journel, A. (1980). The Log-Normal Approach to predicting Local Distributions of Selective Mining Unit Grades., Mathematical Geology, **12**, No. 4, 285-303.

Kitandis, P. (1983). Statistical Estimation of polynomial generalized covariance functions and hydrologic applications. Water Resources Research, **19**, No. 4, 901-921.

Kitanidis, P. (1985). Minimum-Variance Unbiased quadratic estimation of covariances of regionalized variables. Mathematical Geology, **17**, No. 2, 195-208.

Matheron, G. (1965). Les Variables regionaliseés et leur estimation. Masson, Paris 305 p.

Matheron, G. (1973). The Intrinsic Random Functions and their Applications. Advances in Applied Probability, **5**, 439-468.

Matheron, G., (1971). The Theory of Regionalized Variables and its Applications. Centre de Geostatistique, Fontainebleau, France, 212 p.

Micchelli, C., (1986). Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions. Constructive Approximation, **2**, 11-22.

Myers, D. E. (1982). Co-Kriging: The Matrix Form. Mathematical Geology, **14**, No. 3, 249-257.

Myers, D. E. (1984a). Co-Kriging: New Developments. in Geostatistics for Characterization of Natural Resources, G. Verly et al (eds), D. Reidel Pub., Dordrecht, 295-306.

Myers, D. E. (1984b). Letter to the Editor: Conditions for a Variogram model to yield positive variances under restrictions: Comment. Mathematical Geology, **16**, No. 4, 431-432.

Myers, D. E. (1985). Some Aspects of Robustness., Sciences de la Terre, No. 24, 63-79.

Myers, D. E. (1986). The Robustness and Continuity of Kriging. (unpublished) Joint statistics Meeting, Chicago, August 21, 1986.

Neuman, S. and Jacobson, E. (1984). Analysis of NonIntrinsic Spatial Variability by residual kriging with application to regional ground water levels. Mathematical Geology, **16**, No. 5, 499-521.

Rendu, J.-M. (1979). Normal and Lognormal estimation. Mathematical Geology, **11,** No. 4, 407-422.

Sabourin, R., (1976). Application of Two Methods for the interpretation of the underlying variogram. in Advanced Geostatistics for the Mining Industry, Guarascio et al (eds), D. Reidel Pub., Dordrecht, 101-112.

Samper, J. (1986). Statistical Methods of Analyzing Hydrochemical, Isotropic, and Hydrologic data from Regional aquifers. Unpublished doctoral dissertation, University of Arizona, Tucson.

Warrick, A. and Myers, D. E. (1987). Optimization of Sampling locations for variogram calculations, Water Resources Research, **23**, 3, 496-500.

Appendix A

Valid, Istropic Variogram Models

1. Spherical

$$\gamma(r) = \begin{cases} 0\ , & r = 0 \\ C_0 + C_1[\frac{3}{2}(\frac{r}{a}) - \frac{1}{2}(\frac{r}{a})^3], & 0 < r < a \\ C_0 + C_1, & a \leq r \end{cases}$$

$$C_0 \geq 0, \quad C_1 \geq 0,$$

2. Power

$$\gamma(r) = \begin{cases} 0\ , & r = 0 \\ a + br^{\alpha}, & 0 < r \end{cases}$$

$$a \leq 0, \quad b \geq 0, \quad 0 \leq \alpha < 2$$

3. Gaussian

$$\gamma(r) = \begin{cases} 0\ , & r = 0 \\ C_0 + C_1[1 - \exp(\frac{-r^2}{a^2})], & 0 < r \end{cases}$$

$$C_0 \geq 0, \quad C_1 \geq 0, \quad a > 0$$

4. Exponential

$$\gamma(r) = \begin{cases} 0\ , & r = 0 \\ C_0 + C_1[1 - \exp(\frac{-r}{a})], & 0 < r \end{cases}$$

$$C_0 \geq 0, \quad C_1 \geq 0, \quad a > 0$$

magnitudes and also a range of angles. Unfortunately $N(r,\theta)$ may be very small for any particular choice of r,θ. This is true even for regular grids. Most software packages incorporate the use of distance classes and angle windows for computing γ^*. The simple form of the estimator (13) is replaced by

$$\gamma^*(\bar{r},\bar{\theta}) = \frac{\iint G(|x-y|, x-y, u)[Z(x)-Z(y)]^2 dxdy}{2\iint G(|x-y|, x-y, u)dxdy} \tag{15}$$

where

$$G(|x-y|, x-y, u) = [1_{h_1}(|x-y|) - 1_{h_2}(|x-y|)][1_{\theta_1}(\theta) - 1_{\theta_2}(\theta)] \tag{16}$$

$$\theta = \sin^{-1}\frac{< x-y, u >}{|x-y|} \tag{17}$$

and u is a unit vector in the direction $\bar{\theta}$. $\gamma^*(\bar{r},\bar{\theta})$ is a sample variogram but all pairs $Z(x), Z(y)$ with $h_1 < |x-y| < h_2$, $\theta_1 < \theta < \theta_2$ are used to compute $\gamma^*(\bar{r},\bar{\theta})$. We assume $\theta_1, \theta_2, h, h_2$ are chosen so that $h_1 < \bar{r} < h_2$, $\theta_1 < \bar{\theta} < \theta_2$. This has the advantage of increasing the number of pairs used to compute γ^* for each choice of $\bar{r},\bar{\theta}$. One disadvantage though is that $\gamma^*(\bar{r},\bar{\theta})$ is then not an unbiased estimator of $\gamma(\bar{r},\bar{\theta})$. In general

$$E[\gamma^*(\bar{r},\bar{\theta})] = \frac{\iint G(|x-y|, x-y, u)\gamma(|x-y|,\theta)dxdy}{\iint G(|x-y|, x-y, u)dxdy} \tag{18}$$

The bias is in general not uniform with respect to $\bar{r},\bar{\theta}$ and is very dependent on both the sample patterns and the "true" variogram model. The bias will be least where γ is smoothest and greatest when γ is non-differentiable.

Some have proposed (Cressie, 1985) and others have used weighted least-squares fitting of a valid model to the sample variogram. The relationship to the robustness/continuity of the kriging estimator has not been established. There are a number of reasons why this is not an optimal method and the results may even be mis-leading. These will be