MULTIVARIABLE GEOSTATISTICAL ANALYSIS FOR ENVIRONMENTAL MONITORING

Donald E. MYERS*

CONTENTS

* Department of Mathematics, University of Arizona, TUCSON, Arizona 85721 (USA)

Geomathematics and Geostatistics Analysis Applied to Space and Time Dependent Data, in Sci. de la Terre, Sér. Inf., Nancy, 1988, 27, 411-427.

ABSTRACT

The application of statistics to environmental monitoring incorporates many aspects, some of which are the same as those utilized in applying geostatistics to problems in mining and hydrology. By its very nature the problem is multivariate either because there are multiple pollutants that are of interest or because the laboratory analysis produces data requiring a multivariate approach.

Classical multivariate techniques do not explicitly incorporate the locations of the samples nor the spatial correlation and also do not reflect differences in the support of the samples or the support of the region for which an appraisal is desired. Univariate geostatistics does have these properties but it does not incorporate intervariable correlation. Tools are needed then which incorporate the important features of both approaches. Multivariable geostatistics based on cokriging provides such tools. A brief review is given of cokriging in its general form as developed by Myers (1982, 1983, 1984, 1986) together with new results characterizing positive definiteness. This is followed by a discussion of the practice of cokriging including software development. This is placed in the context of applications to environmental monitoring.

RESUME

L'utilisation des méthodes statistiques pour le contrôle de l'environnement est, par bien des égards, semblable aux applications dans l'industrie minière ou en hydrologie. Le problème est, par sa nature même, multivariable soit à cause du nombre de pollutants d'interet, soit parce que la méthode d'analyse fournit des données exigeant une approche multivariable.

Les techniques multivariables classiques ne tiennent compte ni de la localisation des points de mesure ni de la corrélation spatiale. Le support des échantillons et la taille de la région étudiée n'entrent pas en jeu non plus. La géostatistique monovariable a l'avantage d'en tenir compte, sans, pour autant,prendre en compte les correlations entre variables. Il faut, donc, des outils conçus pour les données qui sont à la fois multivariables et spatiales; d'où l'origine du cokrigeage. Après un bref resumé de la technique dans la forme developpée par Myers, de nouveaux résultats au sujet de la positivité et des variogrammes croisés sont presentés. On fait quelques remarques sur des considérations pratiques de l'implémentation du cokrigeage (ainsi que des logicels correspondants) dans le contexte des études d'environnement.

A – INTRODUCTION

Environmental monitoring is concerned with the determination of the presence or absence as well as the identification of one or more pollutants in a region, all of the conclusions drawn or decisions made to be based on information resulting from the analysis of samples taken from the environment. The sampling and analytical process which is often cyclical might be thought of as having five components:

i. Decisions about where to sample, what to sample for, kinds of samples
ii. Actual sampling process including possible in-situ measurements
iii. Laboratory analysis of samples
iv. Statistical or other analysis of data from analyses
v. Conclusions or Decisions based on results of step iv

Of course these are all related and inter-dependent. While we shall concentrate on the fourth component we shall see that the importance of a multivariable approach is in part a consequence of problems arising out of the other components.

Of course any well designed statistical analysis requires prior consideration of the sampling plan but environmental problems like problems arising out of mining applications require consideration of where to sample and the supports of the samples. The formulation of an optimal sampling plan requires identifying appropriate loss function(s), sample size alone is not an adequate characterization. Note that there is a clear distinction between the sampling requirements and subsequent handling of the samples with respect to the ultimate decision making process and that pertaining to the analytical process. Even for the analytical chemist the analysis of the samples is no longer simply a laboratory process but rather is often one utilizing computer controlled experiments with subsequent statistical analysis to make the final determinations. This intermingling of mathematics and statistics with analytical chemistry has spawned a new field known as Chemometrics (see for example the summaries in the Fundamental Reviews issues of Analytical Chemistry, 1980,1982,1984 and 1986). While this paper is not primarily concerned with the problems pertaining to the laboratory analysis suffice it to say that the single most distinguishing characteristic of Chemometrics is its emphasis on multivariable analysis both in the sense of the methods of data analysis and also in the sense of compounds or elements analyzed. This increased complexity has occurred concurrently with an escalation in costs both as a cause and as a response. In many applications a multivariable approach should be seen as one incorporating coroborating information or proxy information used to replace missing or unobtainable information.

The simplest problem is simply to determine the presence or absence at the sampled locations in which case one must decide which pollutants are of interest and little if any data analysis is required unless one wishes to draw an inference about non-sampled locations. In this case one may wish to estimate the probability of the presence, estimate a spatial mean for a given region or estimate the probability distribution for the concentrations in a given region. These are exactly the kinds of problems addressed by geostatistics in its various forms. The simplest of

these being linear kriging and the tool used to formulate the other methods such as disjunctive kriging or probability kriging. As will be seen, cokriging incorporates univariate kriging as a special case and hence all the advantages of kriging as an estimator are retained. In some of the extensions of linear geostatistics to non-linear techniques cokriging provides the tool to reduce these to linear problems. It was originally assumed that cokriging would require much greater computations and hence to be avoided where possible, however programs such as those given in Carr, Myers and Glass (1985) show that this need not be a deterent.

B - WHY LINEAR COMBINATIONS?

Univariate geostatistics uses a linear estimator, i.e., a linear regression as an approximation to the conditional expectation. This is motivated both by the special case of the multivariate gaussian distribution wherein the conditional expectation is a linear function of the data but also because the linear estimator only requires knowledge of the the covariance function. These reasons are still relevant in the case of several variables moreover nearly all classical (as opposed to geostatistical) multivariate statistical techniques are linear, for example principal components and the various related approaches such as cluster analysis, factor analysis and discriminant analysis. Suppose $Z_1(x),\ldots,Z_m(x)$ are random functions representing the variables of interest, x denoting the position in 1,2 or 3 space. If $x_1,\ldots,x_n$ are the sample locations then denote the data as $Z_1(x_1),\ldots,Z_1(x_n),\ldots..Z_m(x_n)$. In general of course these samples are not punctual but will have an associated support. We shall see that the non-punctual support problem can be resolved in the context of multivariable geostatistics in exactly the same way as in univariate geostatistics. It is also seen that consideration of linear estimators and linear combinations leads quite naturally to questions pertaining to positive definiteness. Finally we shall see that linear geostatistics particularly multivariate geostatistics provides a way to incorporate non-linear techniques.

C - COKRIGING

1 - THE ESTIMATOR AND THE ASSUMPTIONS

Let $Z_1(x),....Z_m(x)$ be random functions representing the variables of interest as above. Let $x_1,...x_n$ be the sample locations with data $Z_j(x_i)$; i=1,....,n and j=1,....,m. For the moment it is assumed that all variables are sampled at all locations, later this restriction will be removed. Write

$$\bar{Z}(x) = [Z_1(x),....,Z_m(x)] \qquad (1)$$

$\bar{Z}(x)$ is assumed to satisfy the Intrinsic Hypothesis (analagous to the Intrinsic Hypothesis for univariate geostatistics), namely

. $E\ (\bar{Z}(x+h)-\bar{Z}(x))= [0,\ldots,0]$ for all x, h (2)

. $\bar{\gamma}(h)=0.5\ E(\bar{Z}(x+h)-\bar{Z}(x))^T(\bar{Z}(x+h)-\bar{Z}(x))$ exists and depends only on h

$\bar{\gamma}(h)$ is the variogram matrix, the entries in $\bar{\gamma}(h)$ are the covariances, $\gamma_{jk}(h)= 0.5\ Cov(Z_j(x+h)-Z_j(x),Z_j(x+h)-Z_j(x))$. Note that the matrix of covariances for the components of $\bar{Z}(x)$ may not be symmetric whereas $\bar{\gamma}(h)$ is, however covariances are in general bounded whereas variograms and cross-variograms need not be.

The cokriging estimator as given by Myers (1982) is

$$\bar{Z}^*(x_0)= \Sigma\bar{Z}(x_i)\Gamma_i = [Z_1^*(x_0),\ldots,Z_m^*(x_0)] \tag{3}$$

and the weight matrices $\Gamma_1,\ldots,\Gamma_n$ are chosen so that

.. $E\ (\bar{Z}^*(x_0)-\bar{Z}(x_0))= [0,\ldots,0]$ (4)

.. $\Sigma a_j^2 Var(Z^*_j(x_0)-Z_j(x_0))$ is minimised for fixed positive numbers $a_1,\ldots a_m$

The entries in the weight matrices have an immeadiate interpretation, if λ_{jk}^i is an entry in Γ_i then it is the weight assigned to $Z_j(x_i)$ in estimating $Z_k(x_0)$. The conditions given in (2) constitute a weak form of stationarity, these conditions can be further weakened to allow for Universal cokriging. It will also be seen that the estimator can be re-written in dual form.

2 – THE COKRIGING EQUATIONS

The first condition in (4) is the unbiasedness condition and the second is a weighted sum of the estimation variances. The original derivation of the equations given by Myers (1982) used an equally weighted sum which is a special case. When the Intrinsic Hypothesis (2) is satisfied and if

$$\Sigma\ \Gamma_i= I \tag{5}$$

which is sufficient to ensure that the unbiasedness condition is satisfied then the weighted sum of the estimation variances can be written as

$$Tr\ \Sigma\Sigma B_j^T E[\bar{Z}(x_j)-\bar{Z}(x_0)]^T[\bar{Z}(x_i)-\bar{Z}(x_0)]B_i \tag{6}$$

where $B_i = \Gamma_i A$ and A is a diagonal matrix with the a_i's as the diagonal entries. Tr denotes the trace of the matrix, that is, the sum of the diagonal entries. We then replace (5) by the equivalent condition (5')

$$\Sigma B_i = A \tag{5'}$$

To obtain the B_i's or rather the Γ_i's the scalar valued expression in (6) must be minimised subject to the conditions given in (5'). The reason for allowing unequal weighting in (4) or (6)

is two fold, the first is that we may consider some variables less important (for example in the case of environmental sampling some variables may only be pathfinders or indicators rather than pollutants themselves) and the second is simply one of generality. After introducing Lagrange multipliers because of the constraints we obtain the cokriging equations

$$\Sigma\bar{\gamma}(x_i-x_j)B_i + \mu_A = \bar{\gamma}(x_0-x_j);\ j=1,\ldots n \quad : \quad \Sigma B_i = A \tag{7}$$

and the minimised value of (6) is given by

$$\sigma_{CK}^2 = Tr[\Sigma\ \bar{\gamma}(x_0-x_j)B_j + \mu_A] = Tr[(\Sigma\ \bar{\gamma}(x_0-x_j)\Gamma_j + \mu)A] \tag{8}$$

where $\mu = \mu_A(A^{-1})$. Note that the individual variances may be extracted as the diagonal entries. In fact it is seen that the matrix A is irrelevant and affects neither the estimator nor the separate minimised estimation variances.

3 – FULL OR UNDERSAMPLED?

All of the early applications and a number of examples more recently in print, for example Vauclin et al (1983), Aboufirassi and Marino (1984), represent the "undersampled" problem. That is, where one variable is of principal interest and the other variables are used to enhance the estimation of the first. Moreover in most of these examples, the supplementary variable is used to compensate for a lack of data for the primary variable. There were and are two apparent reasons for this emphasis. First, in most mining applications one metal is of principal interest and decisions are made based only on grades for that metal. Secondly it was assumed that this made the system of equations simpler hence reducing computer time. It was shown in Myers (1984) that the undersampled problem is a special case of the general formulation and there is a simple algorithm for reducing the system. This was implemented in the program given in Carr,Myers and Glass (1985). In many deposits for example some low grade copper deposits, there may be precious metals such as gold,silver,molybdenum, etc. which are insufficient to justify sole extraction but when mined with the primary metal contribute significantly to the profitability of the mine. If this information is not used in the ore selection process, there may be significant variance between potential and realized profit. Hence even in mining applications the first reason for concentrating on the undersampled problem may not be valid, in the context of environmental monitoring it is even less valid. The second reason is not really a valid one in that the system is not appreciably simplified.

First we briefly review the algorthim given in Myers (1984). Write the system given by (7) in matrix form

$$\begin{bmatrix} \bar{\gamma}(x_1-x_1) & \cdots & \bar{\gamma}(x_1-x_n) & I \\ \bar{\gamma}(x_2-x_1) & \cdots & \bar{\gamma}(x_2-x_n) & I \\ \cdots & \cdots & \cdots & \cdot \\ \cdots & \cdots & \cdots & \cdot \\ \bar{\gamma}(x_n-x_1) & \cdots & \bar{\gamma}(x_n-x_n) & I \\ I & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \cdot \\ \cdot \\ \Gamma_n \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_0-x_1) \\ \gamma(x_0-x_2) \\ \cdot \\ \cdot \\ \bar{\gamma}(x_0-x_n) \\ A \end{bmatrix} \tag{7'}$$

It is immeadiately obvious that this is exactly like a uni-variate kriging system except that all entries are replaced by matrices and hence the form of the system is independent of the number of variables as well as the number of sample locations. We see that each of the first n rows of all three "big" matrices (the coefficient matrix on the left, the vector of unknowns on the left and the vector on the right hand side) corresponds to a sample location. If there were no data available at location x_i, then that row would be deleted. Likewise each of the first n columns of the coefficient matrix corresponds to a sample location. But each of these rows or columns has as entries matrices, the columns or rows within these matrices correspond to variables. Hence if data for Z_2 is missing at x_4 we first find column (respectively row) 4 and within that column (row) of matrices we delete column 2 (row 2). In the case of rows we do the same to all three "big" matrices but in the case of columns only the coefficient matrix is altered. As was shown in Carr,Myers and Glass (1985) the software can alter the general system to fit the particular undersampled problem. Subsequent versions of the program has made the process more efficient by never incorporating the rows and columns that would subsequently be deleted. To put the relationship of the undersampled problem to the general formulation in another light consider the former in the usual way.

We wish to estimate $Z_1(x_0)$ using data for Z_1, Z_2 at x_1, x_2, x_3. In non-matrix form we have

$$Z_1^*(x_0) = \Sigma\, \lambda_{11}{}^i Z_1(x_i) + \Sigma\, \lambda_{21}{}^i Z_2(x_i) \qquad (9)$$

and the system of equations

$$\begin{aligned}
&\Sigma\lambda_{11}{}^i\gamma_{11}(x_i-x_j)+\Sigma\lambda_{12}{}^i\gamma_{12}(x_i-x_j)+\ \mu_{11} \ = \ \gamma_{11}(x_0-x_j) && j=1,\ldots,n\\
&\Sigma\lambda_{11}{}^i\gamma_{21}(x_i-x_j)+\Sigma\lambda_{12}{}^i\gamma(x^i-x_j)+\ \mu_{21} \ = \ \gamma_{21}(x_0-x_j) && j=1,\ldots,n \qquad (10)\\
&\Sigma\lambda_{11}{}^i = 1 \quad , \qquad \Sigma\lambda_{12}{}^i = 0
\end{aligned}$$

is obtained by minimizing $\mathrm{Var}[Z_1^*(x_0)-Z_1(x_0)]$ subject to $E[Z_1^*(x_0)-Z_1(x_0)]= 0$. However if we re-arrange this system of equations then it can be written in the form

$$\begin{bmatrix} \bar{\gamma}(x_1-x_1)\ldots\bar{\gamma}(x_1-x_n) & I \\ \bar{\gamma}(x_n-x_1)\ldots\bar{\gamma}(x_n-x_n) & I \\ I \quad \ldots \quad I & 0 \end{bmatrix} \begin{bmatrix} \lambda_{11}{}^1 \\ \lambda_{12}{}^1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_{11}{}^n \\ \lambda_{21}{}^n \\ \mu_{11} \\ \mu_{21} \end{bmatrix} = \begin{bmatrix} \gamma_{11}(x_0-x_1) \\ \gamma_{21}(x_0-x_1) \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{11}(x_0-x_n) \\ \gamma_{21}(x_0-x_n) \\ 1 \\ 0 \end{bmatrix} \qquad (7")$$

which is exactly like (7') except that the column is missing for the λ's for estimating the other variable as well as the second column on the right hand side but these have no effect on the coefficient matrix and it is the inversion of that matrix that requires all the computer time. Hence we can see that solving the full system is essentially no different than solving the single variable system. However the algorithm for producing the undersampled system from the full sampled system is much simpler and much more obvious when the matrix formulation is used.

4 - LINEAR COMBINATIONS

In a mining application one might try to avoid a multivariable approach by forming a linear combination, for example each variable is weighted by its value (for example market price normalized for any differences in the cost of extraction). This new variable then is used for all estimation purposes, variograms are computed etc. This would appear to provide a substantial reduction in time and effort but one must ask what is the price for this simplification and is it always an adequate replacement. In an environmental application, it may not be so simple to assign relative weights since the "value" of a pollutant may not be easily determined and more importantly our perception of the danger (i.e.,cost) associated with a pollutant may change over time.

The basic relationship between kriging linear combinations and forming linear combinations of co-kriged values is given in Myers (1983) and will be briefly reviewed along with an elaboration of the relationship to more classical multivariate techniques. Using Z(x) as above let

$$W(x) = \bar{Z}(x)A, \qquad A^T = [a_1, \ldots, a_m] \tag{11}$$

then
$$\gamma_W(h) = A^T \bar{\gamma}_Z(h) A \tag{12}$$

If $W^*(x0) = \sum \lambda_i W(x_i)$ then the first question is, is $W^*(x_0) = \bar{Z}^*(x_0)A$? The kriging equations for the estimator given in (11) are

$$\sum \lambda_i \gamma_W(x_i - x_j) + \mu_W = \gamma_W(x_0 - x_j) \quad , \quad \sum \lambda_i = 1 \tag{13}$$

If (12) is substituted into (13) then it seen that in order for $W^*(x_0)$ to coincide with $\bar{Z}^*(x_0)A$ we must have $\lambda_i A = \Gamma_i A$ for i=1,...,n. Since A is a column vector and hence is not invertible this limits the choice of the Γ_i's and hence the optimum is in general not acheived. This means that the variance associated with W^* is in general greater than that with $\bar{Z}^*A$.

In a more classical approach one generates new variables, i.e. factors, as linear combinations of the original variables with several objectives in mind. For example if the variables are correlated then the number of factors required to adequately explain the original data set is less than the number of original variables. Several authors, Myers and Carr (1984), Myers (1984), Aboufirassi and Marino (1984), Davis and Greenes (1983) have attempted to merge classical methods such as Principal Components with kriging. In each case Principal Components is used to reduce the

number of variables to a smaller number of "uncorrelated variables", variograms are computed for each new variable and the original variable is obtained as a linear combination of the new variables. One should be aware that orthogonal factors in the sense of Principal Components is not the same as having cross-variograms that are zero. Davis and Greenes did check the sample cross-variograms of the factors and found that it was reasonable to conclude that the factors were uncorrelated in the sense of geostatistics. To put it in a more general context let

$$\bar{Y}(x) = \bar{Z}(x)A, \quad A \text{ an } m \times p \text{ matrix} \quad \text{or} \quad \bar{\gamma}_Y(h) = A^T \bar{\gamma}_Z(h) A \tag{14}$$

each component of $\bar{Y}(x)$ is a linear combination of the components of $\bar{Z}(x)$. If the components of $\bar{Y}(x)$ are uncorrelated in the sense of geostatistics then $\bar{\gamma}_Y(h)$ must be a diagonal matrix. This is very nearly the same as requiring that the sample correlation matrix be diagonalized as in Principal Components. In general however one would expect that the matrix A would depend on h. This perspective is also relevant to the problem of models for variogram matrices and we shall return to this in a later section.

5 – PUNCTUAL vs BLOCK COKRIGING

In the uni-variate form of geostatistics, specifically the kriging estimator, if one wishes to estimate a spatial average such as

$$\bar{Z}_V = (1/V) \int_V \bar{Z}(x) dx \tag{15}$$

where V is an area or volume then it is only necesary to alter the right hand side of the system of kriging equations. Specifically one replaces the variogram matrix by an average value of the variogram matrix. As shown in Myers (1984) the same simple replacement is sufficient for cokriging. Although non-punctual support for the data causes the same problems for the estimation of cross-variograms as for variograms, the use of non-punctual data in the estimator causes no new problems.

6 – NON-STATIONARITY

Although the Intrinsic Hypothesis (2) is itself a form of non-stationarity this assumption can be weakened further. Suppose that

$$\bar{Z}(x) = \bar{Y}(x) + M(x) \tag{16}$$

where $M(x) = F(x)B$. $F(x) = [f_0(x), \ldots, f_p(x)]$ is a vector of known linearly independent functions (usually taken to be polynomials in the position coordinates) and B is a p x m matrix of constants. Y(x) is assumed to satisfy the Intrinsic Hypothesis. The system of cokriging equations (7) or (7') must be modified by replacing each identity matrix by a row (repectively column) of

matrices and the single matrix of Lagrange multipliers is replaced by several such matrices. Write $F_\ell(x_j) = f_\ell(x_j)I$ then the new system is

$$\Sigma\bar{\gamma}(x_i-x_j)\Gamma_i + \Sigma F_\ell(x_i)\mu_\ell = \bar{\gamma}(x_j-x_0)$$

$$\Sigma F_\ell(x_i)\Gamma_i = F_\ell(x_0) \quad ; \; j=1,\ldots,n \text{ and } \ell=0,\ldots,p \qquad (17)$$

Further details may be found in Myers (1982). As in the simplest form of cokriging using the intrinsic hypothesis, Universal cokriging is completey analogous to (uni-variate)Universal kriging in that each entry in the system of equations is replaced by a matrix. The algorithm for obtaining the undersampled version of the system works equally well for Universal cokriging.

7 - TEMPORAL DEPENDENCE

In many applications of geostatistics such as in mining, the time scale for deposition or formation is exceedingly long in comparison to the time span for the period of analysis and hence one may reasonably ignore any temporal dependence. In many environmental problems however the time scale for change is relatively short, for example the dispersion of a plume in an aquifer or the spread of sulfur dioxide in the air. Moreover in many instances the tendency is to sample at multiple time points for a small number of sample locations. We are concerned with estimation/ interpolation problems for a single random function $\bar{Z}(x,t)$ or vector function $Z(x,t)$. There are several possible ways for reducing this problem to a simpler one and at least one leads to the use of cokriging.

a. Assume that $Z(x,t)$ (or $\bar{Z}(x,t)$) satisfies the Intrinsic Hypothesis with respect to x for each t and that moreover the variogram depends on t only in the parameters. This formulation is not too different from an assumption of local stationarity, in some such instances the relative variogram is a useful way to filter out the non-stationarity. In practice only a finite number of t's will be considered.

b. In some instances it is the time dependence that is of primary importance and the formulation in a. might be used with x,t interchanged.

c. If only one variable is of interest and data is available at a finite number of locations and times, re-write the problem as a vector function. That is, $\bar{Z}(x)= [Z(x,t_1),\ldots,Z(x,t_m)]$. This approach has the advantage that data for one time point can be used in the estimation for another time point. If $\bar{Z}(x,t)$ is a vector then the same approach may be used except that the dimension will increase more rapidly.

8 - DUALITY

It is well-known that the uni-variate kriging estimator can be written in a dual form,

$$Z^*(x_0)= \Sigma\lambda_i Z(x_i) = \Sigma b_i\gamma(x_0-x_i) + a \qquad (18)$$

where the λ_i's and b_i'satisfy respectively

$$\Sigma\lambda_i\gamma(x_i-x_j)+\mu = \gamma(x_0-x_j) \; ; \; j=1,\ldots,n \; : \; \Sigma\lambda_i = 1 \tag{19}$$

$$\Sigma b_i\gamma(x_i-x_j) + a = Z(x_j) \; ; \; j=1,\ldots,n \; : \; \Sigma b_i = 0 \tag{20}$$

We see however that the coefficient matrix is exactly the same in (19) and (20). Matheron (1980) and others have shown that this duality implies that thin plate splines are a special case of kriging and smoothing splines are a special case of cokriging. Myers (1986) has shown that this duality can be extended to cokriging and in fact programs to solve the cokriging system can provide the dual solution as well. This duality is not so readily apparent if the cokriging system is written in the "one variable" form as given by equations (9). First we need some notation:

$$K = \begin{bmatrix} \bar{\gamma}(x_1-x_1) \ldots\ldots & \bar{\gamma}(x_1-x_n) \\ \cdot & \cdot \\ \bar{\gamma}(x_n-x_1) \ldots\ldots & \bar{\gamma}(x_n-x_n) \end{bmatrix} \; ; \; F = \begin{bmatrix} I \\ \cdot \\ I \end{bmatrix}$$

$$\hat{\Gamma} = \begin{bmatrix} \Gamma_1 \\ \cdot \\ \Gamma_n \end{bmatrix} \; ; \; K_0 = \begin{bmatrix} \bar{\gamma}(x_1-x_0) \\ \cdot \\ \bar{\gamma}(x_n-x_0) \end{bmatrix} \; ; \; \hat{Z} = [\bar{Z}(x_1),\ldots\bar{Z}(x_n)] \tag{21}$$

then the analogues of (19) and (20) are

$$\begin{bmatrix} K & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \Gamma \\ \mu \end{bmatrix} = \begin{bmatrix} K_0 \\ I \end{bmatrix} \tag{22}$$

$$\begin{bmatrix} K & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} B \\ M \end{bmatrix} = \begin{bmatrix} \hat{Z}^T \\ 0 \end{bmatrix} \tag{23}$$

and the cokriging estimator can be written in either of the two forms

$$Z^*(x_0) = \hat{Z}\,\hat{\Gamma} = [K_0^T \quad I][B^T \quad M^T]^T \tag{24}$$

Further properties of the Dual form of the cokriging estimator and in particular of the matrices B, M are found in Myers (1986)

D – POSITIVE DEFINITENESS

1 – EQUIVALENT DEFINITIONS

It is well-known as shown in Matheron (1973) that in order for a function to be a variogram it must satisfy a (conditional) positive definiteness condition as well as a growth

condition. Although a number of different estimators have been proposed and used for the variogram, in all cases one of the steps involves fitting the estimated variogram to a theoretical model which is usually taken to be a linear combination of known valid models such as the spherical, exponential, power, etc. In particular this ensures that the estimation variance is non-negative and hence has a minimum value. There are several ways to extend these concepts to the multivariable version.

a. As was noted above it is reasonable to consider linear combinations and since the linear combination is uni-variate we might define positive definiteness for matrix valued functions by transforming to a univariate problem. We shall consider only the multi-variate version of variograms but the extension to generalized covariances is immeadiate.

If $\bar{\gamma}(h)$ is a m x m matrix valued function with $\bar{\gamma}(0)= 0$ then $-\bar{\gamma}$ is conditionally positive definite if for all vectors $A^T= [a_1,\ldots,a_m]$; $-A^T\bar{\gamma}(h)A$ is conditionally positive definite. That is , $\gamma(h)$ is a valid variogram matrix if $A^T\bar{\gamma}(h)A$ is a valid variogram. In particular this means that

$$-\Sigma\Sigma\Sigma\Sigma a_r a_s \lambda_i \lambda_j \gamma_{rs}(x_i-x_j) > 0 \qquad (25)$$

for any choice of A, any choice of the $x_1,\ldots x_n$ and any choice of the $\lambda_1,\ldots,\lambda_n$ with $\Sigma\lambda_i = 0$. Note several important special cases. Suppose that only two the a_i's are non-zero and A^T is either [1 1] or [1 -1] then we have that $\gamma_{ii}+\gamma_{jj}+2\gamma_{ij}$, $\gamma_{ii}+\gamma_{jj}-2\gamma_{ij}$ must both be valid models. This observation provides further insight into the form of valid models.

b. Recalling the form of the estimator used when data on several variables is used to estimate a single variable we see that another possible formulation might be that

$$-\Sigma\Sigma\Sigma\Sigma \lambda^i{}_r \lambda^j{}_s \gamma_{rs}(x_i-x_j) > 0 \qquad (26)$$

for any choice of the $x_1,\ldots,x_n$ and any choice of the λ's with $\Sigma\lambda_i = 0$ for all r. Clearly this definition is equivalent to the first one.

c. The obvious definition that corresponds to the general form of the cokriging equations was given by Myers (1984) as

$$-Tr\Sigma\Sigma\Gamma_i{}^T\gamma(x_i-x_j)\Gamma_j > 0 \qquad (27)$$

when $\Sigma\Gamma_i = 0$. Moreover in that same paper it was shown that this is equivalent to (26) although the later was not formulated as a definition at that time.

2 - MODELS FOR VARIOGRAM MATRICES

Irrespective of which estimator is used for the variogram at some point in the process one uses two important properties of valid variograms; the set of valid variograms is closed under positive linear combinations and secondly it contains certain known models such as the spherical, exponential, power, cubic and gaussian. Moreover all of these are characterized by parameters that are interpretable on the graph for an isotropic model. By analogy we might attempt to characterize

valid variogram matrices by requiring that the entries be valid models. For the diagonal entries this is appropriate and each corresponds to a single component of the vector random function, i.e., each diagonal entry is a variogram. For the off-diagonal entries the problem is not so simple. One may utilize a sample cross-variogram to estimate the off-diagonal entries but now the question is, what does a valid cross-variogram look like? In a number of instances, authors have used variogram models for cross-variograms, in general this will lead to an invalid model in view of the first definition of positive definite given above. Moreover that approach would asume that the correlation between the two components is always positive, that is, while variograms are always positive cross-variograms could be negative. Moreover variograms except possibly for hole-effect models are non-decreasing functions but clearly the cross-variogram need not be either non-decreasing or non-increasing.

From the first definition of positive definiteness given above we see that there are three equivalent general forms of the cross-variogram

$$\gamma_{ij}(h)= 0.5[\gamma_{ij}^{+}(h)-\gamma_{ii}(h)-\gamma_{jj}(h)] = 0.5[\gamma_{ii}(h)+\gamma_{jj}(h)-\gamma_{ij}^{-}(h)]$$

$$= 0.25[\gamma_{ij}^{+}(h)-\gamma_{ij}^{-}(h)] \qquad (28)$$

where γ_{ij}^{+}, γ_{ij}^{-} respectively are the variograms for $Z_i(x)+Z_j(h)$, $Z_i(h)-Z_j(h)$. In general a cross-variogram does not look like a variogram but rather like the difference of two variograms.

In Myers (1982) it was suggested that cross-variograms be modelled by modelling γ_{ij}^{+}, γ_{ii}, γ_{jj} then using the first form given in (28). The dis-advantage of this is that the three variograms are modelled independently and one is still not assured of having the correct model. If all three relations given in (28) are used then this problem is overcome. This process does not avoid the fact that for m components there are m variograms and m(m-1)/2 cross-variograms, this is not an impossible task as is seen in Myers and Carr (1984) where eleven variables were considered. If the relations given in (28) are combined with the integral representation formumla given by Matheron (1973) then an integral representation is obtained for the cross-variogram but this is of more theoretical interest than practical. The key point about models for cross-variograms is that they must always be considered in connection with the variograms for the two components.

3 - THE LINEAR MODEL

The construction that has received perhaps the most attention is the linear model. Suppose that $Y_1(x),\ldots,Y_p(x)$ are uncorrelated random functions and that $Z_j(x)= \sum a_{ij}Y_i(x)$ or $\bar{Z}(x) = \bar{Y}(x)A$. Then $\bar{\gamma}_Z(h)=A^T\bar{\gamma}_Y(h)A$ with $\bar{\gamma}_Y$ a diagonal matrix of variograms. This suggests a way to construct general variogram matrices. By using Version 1 of the definition of positive definiteness we see that for any choice of A $\bar{\gamma}_Z$ is a valid model. The problem of course is to determine the Y's or rather their variograms and the entries in A when we begin with $\bar{Z}(x)$. In some instances the decomposition has a physical interpretation such as for geochemical variables, Wackernagel (1985) has exploited this possibility together with the use of Principal Components to model both A and $\bar{\gamma}_Y$. Note that such a representation is equivalent to diagonalizing $\bar{\gamma}_Z$, this is analogous to

the use of Principal Components on a correlation matrix except that the factors are required to be the same for each choice of h. In the case of a multivariate gaussian this decomposition is reasonable since it corresponds to a rotation but in general one would not expect such a decomposition. In all of the early applications only the linear model was used which is very restrictive.

E - PRACTICAL ASPECTS

1 - PROBLEMS AND DIFFICULTIES

Most of the problems encountered in the practice of cokriging are the same as those encountered in the practice of kriging but perhaps magnified by the number of variables incorporated. With respect to the variograms there are no new problems except there are multiple variograms to model.

The problems are essentially of two types both relating to the estimation and modelling of cross-variograms or cross-covariances

a. No matter which estimator is used nor how the cross-variograms are modelled, it is necessary to have a sufficient number of sample locations where both variables are sampled (for each pair of variables). At least in theory the supports should be the same as well. This condition is frequently not satisfied if the form of the analysis is decided after the sampling is complete. Where the condition is not satisfied one sometimes uses an ad-hoc process of kriging one variable at the sample locations of the other variable and estimating the cross-variograms using the kriged data. This should always be done in connection with some form of cross-validation using cokriging.

b. In the case of variograms or covariances it is easy to ensure that the positive definiteness condition is satisfied by using nested models, i.e., positive linear combinations of valid models moreover the parameters of the models are at least partially identifiable from the graph of the sample variogram. However positive definiteness is not a property of a cross-variogram rather it is a property of the triple of two variograms and the associated cross-variogram. As shown above the cross-variogram has a more complex structure and one can not easily identify the components or the parameters from the graph of the sample cross-variogram. One should follow several approaches i. compute and plot the sample cross-variogram (or cross-covariance) ii. model the variograms of each variable separately and cross-validate iii. model the variograms for the sum and difference of each pair of variables then compute and model the cross-variograms using the relations in (28), this later step will likely require sequential adjustments to get a match between the three representations of the cross-variogram. It should be noted that in the case of one variable the variogram generalizes the covariance but in the case of cross-variograms where the cross-covariance is not symmetric then they do not necessarily coincide in the case of second order stationarity. A sufficient condition is given in Myers (1982), in particular it may be preferable to use cross-covariances in some cases since the variograms,cross-variograms are always symmetric.

While cross-validation does not lead to tests of hypotheses in a strong statistical sense it does provide a mechanicism for ascertaining whether the variogram model reproduces the characteristics of the data, in the case of cokriging however the problem is more complex. It is possible to use cokriging to cross-validate in exactly the same way the univariate kriging is used but the question of which statistics to use to select the better model has not been resolved as yet.

2 - SOFTWARE

Proto-type general cokriging and co-coditional simulation programs are given in Carr, Myers and Glass (1985) and in Carr, Myers (1985). Sample variogram and sample cross-variogram programs are given in Journel and Huijbregts (1978) and are generally available in a number of commercially available packages or the public domain STATPAC available from the USGS. At the University of Arizona newer versions of the cokriging programs including dual cokriging and cross-validation using cokriging have been written in FORTRAN 77 and implemented on PC compatibles, the same source codes are useable on VAXs. A program for testing the Cauchy-Schwartz inequality has also been written and is being altered to incorporate graphics. The availability of high resolution color graphics and version 4.1 of the MicroSoft FORTRAN compiler which allows much larger source codes and data sets makes the micro quite adequate both for computation and for display of results. There is a considerable in savings in capital investment in comparison to the use of the mainframe.

3 - EXTENSIONS

Because of the weaknesses associated with interpolating or estimating non-linear functions such as probability distributions by linear combinations, Matheron introduced Disjunctive Kriging. Myers (1986) has shown that the multivariate form of Disjunctive Kriging is easily obtained using the general form of cokriging, this extension is in the process of being implemented by software. Disjunctive Kriging incorporates a strong gaussian distribution assumption and Journel has proposed the use of Indicator or Probability Kriging which uses a non-linear transformation. In its most general form the weights are obtained as the solution to a cokriging system. It was noted above that univariate kriging incorporates thin plate splines as a special case via the dual form of the kriging estimator, dual cokriging provides a way to define multivariable splines and this connection is currently be developed.

In that kriging allows the computation of the variance without knowledge of the data values, a number of authors have dealt with the possibility of the design of an optimal sampling plan where the function to be minimized in some combination of kriging variances or is the kriging variance associated with the estimation for a region. In the case of environmental monitoring similar problems arise but the loss function will now be much more complicated including components such as the overhead on the sampling process, costs for individual sites, costs

associated with particular types of samples, type of laboratory analysis to be used (i.e., pollutants or indicators to be sampled for), costs (health, social, etc) associated with Type I, Type II errors. It is important to recall that sampling designs in this sense refer to the collection of data for cokriging (in one of its various forms) not for the purpose of estimation of variograms or cross-variograms. General software for the former is under development and unlike previous examples in the univariate case does not select new sample locations from a pre-specified grid, this requires a different search/optimization technique.

F – ACKNOWLEDGEMENT

Portions of this article were written while the author was visiting at the Centre de Géostatistique, Fontainebleau.

G – NOTICE

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

REFERENCES

Aboufirassi,M. and Marino,M. ,1984, Co-Kriging of Aquifer Transmitivity from Field Measurements of Transmitivity and Specific Capacity: Math. Geology, 16, p. 11

Carr,J. and Myers,D.E. ,1984, Applications of the Theory of Regionalized Variables to the Spatial Analysis of Landsat Data: in IEEE 1984 Proceedings of the PECORA 9 Spatial Information Technologies for Remote Sensing Today and Tomorrow, IEEE Computer Science Press

Carr,J.,Myers,D.E.,and Glass,C.1985,Co-Kriging-A Computer Program: Computers and Geosciences,11,p. 111

Carr,J. and Myers,D.E., 1985, COSIM:A FORTRAN IV Program for Co-Conditional Simulation: Computers and Geosciences, p.675

Chauvet,P. ,1977, Exemple d'Application de Geostatistique Non-Stationnaire,Le Co-Krigeéage: in the Proceedings of the Pribram Mining Congress, Prague

Davis,B. and Greenes,K.,1983, Estimation Using Spatially Distributed Multivariate Data;An Example with Coal Quality Data: Math. Geology, 15, p. 287

Francois-Bongarcon,D., 1981, Le Co-Regionalizations: C-86, Centre de Géostatistique, Fontainebleau, 27 p

Kirsch,Ch. and Pawlowsky,U., 1985, Modelisation de variogrammes croises: comment satisfaire à l'inegalite de Cauchy-Schwartz: Sciences de la Terre, p. 53

Matheron,G.,1971, The Theory of Regionalized Variables and its Application: Centre de Geostatistique,Fontainebleau, 211 p

Matheron,G.,1973, Intrinsic Random Functions and their Applications: Advances in Applied Probability, vol 5, p.439

Matheron,G.,1979,Recherche de Simplification dans un Probleme de Cokrigéage: N-628, Centre de Géostatistique, Fontainebleau

Matheron,G.,1980, Splines et Krigeage;leur equivalence formelle: N-667, Centre de Géostatistique, Fontainebleau , 26p

McBratney,A. and Webster,R., 1983, V. CoRegionalization and Multiple Sampling Strategy: J. Soil Science, 34, p. 137

Myers,D.E., 1982, Matrix Formulation of Co-Kriging: Math. Geology, 14, p. 249

Myers,D.E.,1983, Estimation of Linear Combinations and Co-Kriging: Math. Geology,15,p. 633

Myers,D.E., 1984, Co-Kriging-New Developments: in Geostatistics for Natural Resource Characterization, G. Verly et al (eds), D. Reidel Publishing Co, Dordrecht

Myers,D.E. and Carr, J., 1984, Co-Kriging and Principal Components;Bentonite Data Re-visited: Sciences de la Terre, 21,p. 65

Myers, D.E., 1985, Co-Kriging,Methods and Alternatives: in The Role of Data in Scientific Progress, P. Glaeser (ed), Elsevier Publishing Co.

Myers,D.E.,1986, Some New Aspects of Multivariate Analysis: NATO A.S.I. on Statistical Treatment for Estimation of Mineral and Energy Resources , 22 June-4 July 1986 , IL Ciocco(Lucca), Italy

Myers,D.E. 1986, New Multivariate Techniques: EPA Conference on Progress in Chemometrics, August 1986, Las Vegas

Vauclin,M., Vierra,S., Vachaud,G., and Nielsen, D.R.,1983, The Use of Co-Kriging with Limited Field Soils Observations: Soil Science Society J., 47, p. 175

Wackernagel,H.,1985, L'ajustement multivariable d'un modele linéaire: N-945, Centre de Géostatistique, Fontainebleau

Wackernagel,H.,1985, Etude des correlations de variables géochimique en modele linéaire des coregionalisations: N-954, Centre de Géostatistique, Fontainebleau

Wackernagel,H.,1985, The Inference of the Linear Model of the Coregionalization in the case of a Geochemical Data Set: Sciences de la Terre,

Wackernagel, H.,1986, Geostatistical Techniques for Interpreting Multivariate Spatial Information: NATO A.S.I. on Statistical Treatment for Estimation of Mineral and energy Resources, 22 June-4 July 1986, Il Ciocco(Lucca), Italy