

Matrix Formulation of Co-Kriging¹

Donald E. Myers²

The matrix form of the general co-kriging problem is presented. Matrix solutions are given for SRFs with covariance functions, for IRFs of order zero using variograms and for universal co-kriging. General methods for obtaining cross-covariance or cross-variogram models are given. The relationship of the general co-kriging problem to the problem of one under sampled variable is presented.

KEY WORDS: co-kriging, cross-variance, cross-variogram, joint estimation, estimation variance, matrix form, linear model, trace.

INTRODUCTION

Although the basic principles and theory of co-kriging are well known it is a tool that is not often used. There are several reasons why this may be so. These include the notational and computational complexities, difficulties in modeling cross variance or cross variograms and a primary interest in problems which do not require joint estimation, in particular an emphasis on problems where spatial correlation is more important than intervariable correlation. Although Journel and Huijbrechts (1978) note that in mining applications co-kriging is used mainly when one variable is undersampled there are applications both in mining, meteorology, soils, geochemical dispersion patterns, hydrology where co-kriging is relevant if other difficulties are surmountable. Matheron (1979) has noted that in general, kriging a linear combination of dependent variables is not the same as the linear combination of co-kriged variables and has given an alternative solution when certain assumptions are justifiable. The use of co-kriging may be determined by whether the principal objective is the reduction of the estimation variance for one variable or the need to jointly estimate several variables. By formulating the problem in matrix form we see that the problem of one variable being undersampled is not the general co-kriging problem, and an emphasis on that problem obscures the question of the appropriate formulation of the esti-

¹Manuscript received 1 October 1981.

²Department of Mathematics, University of Arizona, Tucson, Arizona 85721.

mation variance. There is a natural interpretation for the estimation variance but it occurs in perhaps an unexpected manner which still retains the similarity with the single variable form. For simplicity then we begin with the problem of jointly estimating point values for several dependent random functions.

THE STATIONARY CASE

As in the case of single variable kriging, it is easiest to begin with the case of stationary random functions with finite variance and then generalize.

Let Z_1, \dots, Z_m denote random functions representing the variables of concern, for example, lead-zinc-silver, clay-silt-bulk density, sodium content-cover loam thickness-stoniness, bariometric pressure-temperature-humidity, to name a few possibilities. $S = \{x_1, x_2, \dots, x_n\}$ is the set of sample locations. If we write

$$\bar{Z}(x) = [Z_1(x), \dots, Z_m(x)]$$

then our objective is to estimate the vector $\bar{Z}(x)$ given the data $\bar{Z}(x_1), \dots, \bar{Z}(x_n)$. To place the problem in the same setting as with single variable kriging, the vectors $\bar{Z}(x), \bar{Z}(x_1), \dots, \bar{Z}(x_m)$ must be viewed as elements of a vector space with an inner product so that the estimate of $\bar{Z}(x)$ is obtained as the projection onto a certain subspace determined by the data vectors. The inner product must be scalar valued, though, so that the estimation variance will be scalar valued and hence that the possibility of a minimum is provided. Rather than presenting the problem in the full context of Hilbert spaces we use the more familiar probabilistic description; that is, we want to construct a linear combination of the data vectors such that the estimator is unbiased and the "estimation variance" is minimized. We can write the linear estimator in the form

$$\bar{Z}^*(x) = \sum_{k=1}^n \bar{Z}(x_k) \Gamma_k \quad (1)$$

where each Γ_k is an $m \times m$ matrix. If we let λ_{ij}^k denote the element of the i th row, j th column then this element represents the contribution of the i th variable at location x_k , to the estimate of the j th variable. In nonmatrix form this would be

$$Z_j^*(x) = \sum_k \sum_i Z_i(x_k) \lambda_{ij}^k$$

Since expectation is linear

$$E[\bar{Z}^*(x)] = \sum_k E[\bar{Z}(x_k)] \Gamma_k$$

Since the Z_i are stationary, a sufficient condition for \bar{Z}^* to be unbiased is that

$$\sum_{k=1}^n \Gamma_k = I \quad (2)$$

Kriging each variable separately corresponds to replacing (2) by the stronger conditions

$$\begin{aligned}\lambda_{ij}^k &= 0, & \text{for } i \neq j \\ \sum_k \lambda_{ii}^k &= 1, & \text{for each } i\end{aligned}\quad (2a)$$

In the language of projections, separate kriging results in projecting into a smaller subspace and hence in general the estimation variance will be larger. If one or more of the variables is undersampled it results in adding additional constraints to those implicit in (2), namely

$$\lambda_{i_o j}^{k_o} = 0, \quad \text{for } j = 1, \dots, m \quad (2b)$$

where x_{k_o} is an unsampled location for the i_o variable.

It is seen then that expectation can be taken to be a vector, that is, the vector of the separate expectations, but variance can not be treated this way. There are at least two ways to define an estimation variance and

$$\max_{1 \leq i \leq m} \{ \text{Var} [Z_i(x) - Z_i^*(x)] \}$$

and

$$\sum_{i=1}^m \text{Var} [Z_i(x) - Z_i^*(x)] \quad (3)$$

A_f is seen later; this is equivalent to considering only $\text{Var} [Z_i(x) - Z_i^*(x)]$ if one is interested only in estimating Z_i and it will be shown that the estimation variances attributable to the separate components may be selected out.

The second form is more natural in the context of Hilbert spaces and computationally more tractable so it will be used. In terms of the metrics induced by these two possibilities, they are equivalent. When (2) is satisfied, (3) may be written in the form

$$E[\bar{Z}(x) - \bar{Z}^*(x)] [\bar{Z}(x) - \bar{Z}^*(x)]^T \quad (4)$$

Of course, to determine the Γ_k so that (2) will be satisfied and (4) will be minimized we must express (4) in terms of the covariances. Let $E[Z_i(x) Z_j(y)] = C_{ij}(x - y)$ and then

$$\bar{C}(x - y) = [C_{ij}] = E[\bar{Z}(x)^T \bar{Z}(y)] \quad (5)$$

where T denotes the transpose of the matrix. $\bar{C}(x - y)$ is an $m \times m$ matrix and in general is nonsymmetric, that is $\bar{C}(x - y) \neq \bar{C}(y - x)$ and $\bar{C}(x - y) \neq \bar{C}(x - y)^T$ but it always satisfies $\bar{C}(x - y)^T = \bar{C}(y - x)$. Our objective now is to express (4) in terms of $\bar{C}(x_i - x_j)$ and Γ_k . This is accomplished by noting that

$$\text{Tr } E[\bar{Z}(x) - \bar{Z}^*(x)]^T [\bar{Z}(x) - \bar{Z}^*(x)] = E[\bar{Z}(x) - \bar{Z}^*(x)] [\bar{Z}(x) - \bar{Z}^*(x)]^T \quad (6)$$

We also note that (2) can be written in the form (Tr denotes the trace)

$$\sum_{k=1}^n \lambda_{ij}^k = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (7)$$

Let $\bar{\mu}$ be an $m \times m$ matrix whose elements are denoted by μ_{ij} , these are the Lagrange multipliers. Set

$$\phi(\Gamma_1, \dots, \Gamma_n, \bar{\mu}) = \sum_i \sum_j \mu_{ij} \left[\sum_k \lambda_{ij}^k - \delta_{ij} \right] \quad (8)$$

$$\Phi(\Gamma_1, \dots, \Gamma_n, \bar{\mu}) = \text{Tr } E[\bar{Z}(x) - \bar{Z}^*(x)]^T [\bar{Z}(x) - \bar{Z}^*(x)] + 2\phi(\Gamma_1, \dots, \Gamma_n, \bar{\mu}) \quad (9)$$

To minimize Φ we must solve the system of equations

$$\begin{aligned} \partial\Phi/\partial\lambda_{ij}^k &= 0, & k = 1, \dots, n; \quad i, j = 1, \dots, m \\ \partial\Phi/\partial\mu_{ij} &= 0 \end{aligned} \quad (10)$$

After computing the derivatives and simplifications we obtain

$$\begin{aligned} \sum_{j=1}^n \bar{C}(x_i - x_j) \Gamma_j + \bar{\mu} &= \bar{C}(x_i - x), \quad i = 1, \dots, m \\ \sum_{j=1}^n \Gamma_j &= I \end{aligned} \quad (11)$$

or simply

$$\begin{bmatrix} \bar{C}(x_1 - x_1) & \cdots & \bar{C}(x_1 - x_n) & I \\ & \cdots & & \\ \bar{C}(x_n - x_1) & \cdots & \bar{C}(x_n - x_n) & I \\ I & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_n \\ \bar{\mu} \end{bmatrix} = \begin{bmatrix} \bar{C}(x_1 - x) \\ \vdots \\ \bar{C}(x_n - x) \\ I \end{bmatrix} \quad (12)$$

The form of (12) is exactly the same as for simple kriging of one variable except that the entries are matrices instead of scalars.

The estimation variance may be written in several forms, one of which is

$$\sigma_K^2 = \text{Tr} [\bar{C}(0)] - \text{Tr} \left[\sum_{j=1}^n \bar{C}(x - x_j) \Gamma_j \right] - \text{Tr } \bar{\mu} \quad (13)$$

Note that the order of multiplication for matrices must be preserved and also the nonsymmetry of \bar{C} must be respected. Although σ_K^2 is a cumulative estimation variance the variance attributable to the sequence components is easily de-

terminated. If we write V as

$$V = \bar{C}(0) - \sum_{i=1}^n \bar{C}(x - x_i) \Gamma_i - \bar{\mu} = [v_{jk}], \quad \text{then}$$

$$\sigma_K^2 = \text{Tr } V = \sum_{j=1}^m v_{jj}, \quad \text{and} \quad (14)$$

$$\sigma_{K,j}^2 = v_{jj} = C_{jj}(0) - \sum_i \sum_k C_{ji}(x - x_k) \lambda_{ij}^k - \mu_{jj}$$

We can now make several useful observations about the matrix eq. (12). First, in spite of the nonsymmetry of $\bar{C}(x - y)$, the coefficient matrix in (12) is symmetric, and second, except for the zero matrix in the lower right-hand corner, all other entries are invertible so that the coefficient matrix could be reduced to lower (or upper) triangular form by operating on matrices and hence simplify the computations. As noted (12) has the same form independent of m and is clearly the appropriate generalization of one variable kriging. If the \bar{C} s are diagonal matrices, that is, the components are uncorrelated, and (2) and (2a) are equivalent, but if the \bar{C} s are not diagonal then (2a) is stronger and the system of equations is not the same as (12) and in general the estimation variance is larger than that given by (13). Likewise (2b) is stronger than (2) and the system of equations is not the same as (12).

INTRINSIC RANDOM FUNCTIONS

As with simple kriging of one variable, it is simpler to utilize cross variograms than to use cross covariances but this is not always possible. Subject to a symmetry condition, however, it is possible to rewrite the system of equations using cross variograms. Instead of assuming that the Z s are stationary random functions we assume

- (i) $E[Z_i(x+h) - Z_i(x)] = 0, \quad \text{for } i = 1, \dots, m, \quad \text{that is}$
- $$E[\bar{Z}(x+h) - \bar{Z}(x)] = [0, 0, \dots, 0] = 0$$
- (ii) $\text{CoVar } [Z_i(x+h) - Z_i(x), Z_j(x+h) - Z_j(x)] = 2\gamma_{ij}(h)$

exists and is dependent only on h for $i, j = 1, \dots, m$

If (i) is satisfied then (ii) may be written as

$$\frac{1}{2} E[\bar{Z}(x+h) - \bar{Z}(x)]^T [\bar{Z}(x+h) - \bar{Z}(x)] = \bar{\gamma}(h) = [\gamma_{ij}(h)] \quad (15)$$

To express the estimation variance in terms of the $\bar{\gamma}(x_i - x_j)$ and the Γ_k , it is necessary to also assume or require that

- (iii) $E[\bar{Z}(x+h) - \bar{Z}(0)]^T [\bar{Z}(x) - \bar{Z}(0)] = E[\bar{Z}(x) - \bar{Z}(0)]^T [\bar{Z}(x+h) - \bar{Z}(0)],$
- for every x, h ; that is

$$E[Z_i(x+h) - Z_i(0)] [Z_j(x) - Z_j(0)] = E[Z_i(x) - Z_i(0)] [Z_j(x+h) - Z_j(0)],$$

for $i, j = 1, \dots, m$ and all x, h

For stationary random functions with finite variances this condition is equivalent to

$$(iii)' \quad C_{ij}(h) - C_{ij}(-x) - C_{ij}(x+h) = C_{ij}(-h) - C_{ij}(x) - C_{ij}(-x-h)$$

which is obviously satisfied if $C_{ij}(y) = C_{ij}(-y)$ for all i, j, y . (iii) is necessary to derive the following identity

$$E[\bar{Z}(x+h) - \bar{Z}(0)]^T [\bar{Z}(x) - \bar{Z}(0)] = \bar{\gamma}(x+h) + \bar{\gamma}(x) - \bar{\gamma}(h) \quad (16)$$

which is needed to express the estimation variance in terms of the variogram matrices. The estimator \bar{Z}^* has the same form as before given by (1), and (2) is a sufficient condition for the unbiased property which then allows the estimation variance to be again written in the form given by (6). Utilizing the identity, an analogous expression to (9) is formal and after differentiation we obtain the system of matrix equations

$$\sum_{j=1}^n \bar{\gamma}(x_i - x_j) \Gamma_j + \bar{\mu} = \bar{\gamma}(x_i - x); \quad i = 1, \dots, n, \quad (17)$$

or

$$\sum_{j=1}^n \Gamma_j = I$$

$$\begin{bmatrix} \bar{\gamma}(x_1 - x_1) & \cdots & \bar{\gamma}(x_1 - x_n) & I \\ & \cdots & & \\ \bar{\gamma}(x_n - x_1) & \cdots & \bar{\gamma}(x_n - x_n) & I \\ I & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_n \\ \bar{\mu} \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_1 - x) \\ \vdots \\ \bar{\gamma}(x_n - x) \\ I \end{bmatrix} \quad (18)$$

The estimation variance then may be written in the following form

$$\sigma_K^2 = \text{Tr} \left[\sum_{i=1}^n \gamma(x - x_i) \Gamma_i \right] + \text{Tr} \bar{\mu} \quad (19)$$

and as in the stationary case the estimation variance attributable to the separate components may be selected out.

UNIVERSAL KRIGING

If the components of the random function $\bar{Z}(x)$ are not assumed to be stationary but have expected values that are locally represented by linear combinations of known functions then the above results are easily extended in a manner analogous to universal kriging for one variable.

Let $\bar{F}(x) = [f_1(x), \dots, f_p(x)]$ be a $1 \times p$ vector of known functions, linearly independent over the support x_1, \dots, x_n . $E[\bar{Z}(x)]$ is then assumed to be of the form

$$\begin{aligned}\bar{M}(x) &= \bar{F}(x) B \\ B &= [b_{ij}], \quad \text{a } p \times m \text{ matrix}\end{aligned}\tag{20}$$

For \bar{Z}^* to be unbiased it is sufficient that

$$\bar{F}(x) B = \sum_{j=1}^n \bar{F}(x_j) B \Gamma_j \tag{21}$$

whatever the matrix B . If we write $F_l(x_j) = f_l(x_j) I$, I the identity matrix then (21) becomes

$$\sum F_l(x_j) \Gamma_j = F_l(x), \quad \text{for } l = 1, \dots, p \tag{22}$$

and the universal kriging system is

$$\begin{aligned}\sum_{j=1}^n \bar{C}(x_i - x_j) \Gamma_j + \sum_{l=1}^p F_l(x_i) \bar{\mu}_l &= \bar{C}(x_i - x) \\ \sum_{j=1}^n F_l(x_j) \Gamma_j &= F_l(x), \quad l = 1, \dots, p;\end{aligned}\tag{23}$$

$i = 1, \dots, n$ or if

$$\begin{bmatrix} \bar{C}(x_1 - x_1) \cdots \bar{C}(x_1 - x_n) & F_1(x_1) \cdots F_p(x_1) \\ \bar{C}(x_n - x_1) \cdots \bar{C}(x_n - x_n) & F_1(x_n) \cdots F_p(x_n) \\ F_1(x_1) & \cdots F_1(x_n) \\ F_p(x_1) & \cdots F_p(x_n) \end{bmatrix} \begin{matrix} \\ \\ \bigcirc \end{matrix} = W$$

$$\begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_n \\ \bar{\mu}_1 \\ \vdots \\ \bar{\mu}_p \end{bmatrix} = X, \quad \begin{bmatrix} \bar{C}(x_1 - x) \\ \vdots \\ \bar{C}(x_n - x) \\ F_1(x) \\ \vdots \\ F_p(x) \end{bmatrix} = L$$

The system is

$$WX = L \tag{24}$$

The kriging variance is

$$\sigma_K^2 = \text{Tr } \bar{C}(0) - \left[\text{Tr } \sum_{i=1}^n \bar{C}(x - x_i) \Gamma_i \right] \quad (25)$$

$$\left[- \text{Tr } \sum_{l=1}^p F_l(x) \bar{\mu}_l \right]$$

and the variance attributable to various components may be selected out.

MODELS FOR $\bar{C}(h)$ AND $\bar{\gamma}(h)$

As noted earlier, one of the difficulties occurring in the use of co-kriging is modeling the cross covariances or cross variograms. One model which is used, but which does not lead to a reduction in the estimation variance when compared to separate variable kriging, is the strict linear model. That is, it is assumed that each $Z_j(x)$ has a representation

$$Z_j(x) = \sum_{k=1}^s a_{jk} Y_k(x) \quad (26)$$

where the Y_k s are stationary, independent with separate covariances $K_k(h)$. Then

$$C_{ij}(h) = \sum_{k=1}^s a_{ik} a_{kj} K_k(h) \quad (27)$$

An elementary observation about variances provides a much wider choice of models. For any pair $Z_i(x)$, $Z_j(x)$ let $U_{ij}(x) = Z_i(x) + Z_j(x)$ then

$$E[U_{ij}(x+h) U_{ij}(x)] = C_{ii}(h) + C_{jj}(h) + 2C_{ij}(h) \quad (28)$$

that is, cross-covariance models can be obtained as linear combinations of covariances for the separate components and covariance for the sum. For intrinsic random functions this becomes

$$\gamma_{ij}(h) = \frac{1}{2} [\gamma_{i,j}^+(h) - \gamma_i(h) - \gamma_j(h)] \quad (29)$$

we see that spherical, exponential, polynomial, gaussian models may be used to fit $\gamma_{i,j}^+$, γ_i , γ_j and thus $\gamma_{ij}(h)$. In general it will be necessary to verify that $|\gamma_{ij}(h)| \leq [\gamma_i(h) \gamma_j(h)]^{1/2}$ after the separate modeling of γ_i , γ_j , $\gamma_{i,j}^+$. This suggests that rather than computing and plotting sample cross covariograms, sample variograms for paired sums should be computed and plotted instead.

An examination of the integral representation theorem for generalized covariances, Matheron (1973), shows that the above analysis also leads to a representation for generalized cross covariances.

CONCLUSIONS

Matrix formulation of the co-kriging problem leads to a clear identification of appropriate conditions and suggests computational simplifications and methods for modeling cross covariances or cross covariograms.

ACKNOWLEDGMENTS

Portions of this research were completed while on sabbatical from the University of Arizona and visiting at the Centre de Géostatistique, Fontainebleau.

REFERENCES

- Francois-Bongarcon, D., 1981, Les coregionalisations, le cokrigeage; Centre de Géostatistique, Fontainebleau, C-86.
- Journel, A. and Huijbrechts, Ch. J., 1978, Mining Geostatistics: Academic Press.
- Matheron, G., 1973, The intrinsic random functions and their applications: Adv. Appl. Prob. vol. 5, p. 437-468.
- Matheron G., 1979 Recherche de simplification dans un probleme de cokrigeage: Centre de Géostatistique, Fontainebleau, N-698.
- Myers, D. E., 1981, Joint estimation of random functions: the matrix form: Research report, Department of Mathematics, University of Arizona.