Estimation of the Spatial Distribution of Soil Chemicals Using Pseudo-Cross-Variograms

R. Zhang,*D. E. Myers, and A. W. Warrick

ABSTRACT

In using cokriging to study soil spatial variability, a key step is to determine cross-variograms. A recently developed approach was utilized to compute pseudo-cross-variograms, from which cross-variograms can be formulated. The approach does not require a large number of common locations where data are available for all variables used in the cokriging modeling and estimation processes. In this study, with only one-thirteenth of the original data for NO_3 and Ca, valid cross-variograms, each with the electrical conductivity (EC), were obtained by using pseudo-cross-variograms. Based on the cross-variograms, cokriging with EC improved the estimation of NO_3 and Ca significantly. Cokriging yielded a smaller mean squared error (MSE) and kriging variance, and a higher correlation between estimates and measurements. Using 20 points of NO_3 and 130 points of EC, cokriging provided a similar distribution pattern for NO_3 as that generated with 100 points of NO_3. Cokriging with EC reduced MSE and the mean kriging variance of the estimated Ca up to 78 and 85%, respectively, compared with kriging.

SOIL SPATIAL VARIABILITY can be a frustrating issue in the field-scale description of soil physical, chemical, and hydrological parameters. Geostatistics is a helpful tool to study the spatial distribution of soil properties (Yost et al., 1982a,b; Burgess et al., 1981; Hatfield et al., 1984; Morkoc et al., 1985; Warrick et al., 1986). Most geostatistical research in soil science has been based on determinations of the spatial correlation function, i.e., the variogram, and a variety of linear estimators such as kriging.

Cokriging is an extended technique of kriging, which incorporates both spatial and intervariable correlation. Cokriging can estimate all or some variables using the same coefficient matrix. Several applications of cokriging for describing soil-water phenomena have been reported during the past 10 yr. For example, Vauclin et al. (1983) used kriging and cokriging to predict available water content and sand content. McBratney and Webster (1983) applied cokriging to interpret the topsoil silt content, using subsoil silt or subsoil sand as joint variables. Carr and Myers (1984) studied the application of cokriging to the analysis of satellite data. Yates and Warrick (1987) and Mulla (1988) estimated soil water content using a cokriging procedure in which the bare soil surface temperature and the sand content were used as auxiliary functions. Using soil-map delineations, Stein et al. (1988) pointed out that the use of cokriging resulted in an average increase in precision of about 10% in 30-yr average moisture deficit maps. Zhang et al. (1992) improved the estimation of soil texture by using cokriging with spectral properties.

The key step in cokriging is to formulate cross-variograms or cross-covariances of two random variables (or more). Standard approaches for modeling cross-variograms are usually based on sums or differences of the two random variables, which are measured at the same locations. One disadvantage to the standard approaches is that only common locations can be used. Estimating the cross-variograms requires a large number of locations where data is collected for both variables, a condition that is frequently not satisfied in practice. For cases in which this type of information is not available, cokriging cannot be used to improve estimation. Clark et al. (1989) presented a variation of cokriging, using pseudo-cross-variograms, which did not require that both variables be measured at the same locations. Our objective was to apply pseudo-cross-variograms and cokriging to improve estimations of the spatial distribution of soil chemicals over relatively large areas of land.

R.Zhang, U.S. Salinity Lab. USDA-ARS, 4500 Glenwood Dr., Riverside, CA 92501; D.E. Myers, Mathematics Dep., and A.W. Warrick, Dep. of Soil and Water Science, Univ. of Arizona, Tuscon, AZ 85721. Received 2 Aug. 1991. *Corresponding author.

Published in Soil Sci. Soc. Am. J. 56:1444–1452 (1992).

THEORY

Pseudo-Cross-Variograms

Cokriging is a method for estimating one or more variables of interest using data from several variables by incorporating not only spatial correlation but also intervariable correlation. Let $Z_1(x)$, ..., $Z_m(x)$ denote the value of the variables Z_1, ..., Z_m at location x, and $\mathbf{Z}(x) = [Z_1(x), ..., Z_m(x)]$. If x_1, ..., x_n are sample locations with data $\mathbf{Z}(x_1)$, ..., $\mathbf{Z}(x_n)$, then the cokriging estimator can be written in the form

$$\mathbf{Z}^*(x) = \sum_{i=1}^{n} \mathbf{Z}(x_i)\mathbf{\Gamma}_i \qquad [1]$$

where $\mathbf{\Gamma}_i$s are weight matrices. In many applications, the data may not include values for all variables at all locations. This is called the undersampled problem. The estimator adapts to this case by inserting zeros in the appropriate places of the weight matrices. It is shown in Myers (1984) that the undersampled case is a special one both in terms of the estimator and the set of equations used to determine the weight matrices.

To determine the weight matrices, the variables (Z_j (j = 1, 2, ..., m) at x, $Z_j(x)$, are considered to be random functions and statistical conditions are imposed on the cokriging estimator. The estimator is required to be unbiased and the sum of the variances of the estimator errors (of the respective variables) is minimized. As shown in Myers (1982) this leads to a linear system of matrix equations. To quantify the spatial correlation of each variable and also the intervariable correlation, either variograms and cross-variograms, or covariances and cross-covariances are used. These statistical techniques are used to compute the variances of the errors of estimations. Covariances and cross-covariances can be used when the random functions are second-order stationary, i.e., when

(i) $E\{Z_j(x)\} = m_j$, a constant depending on j,
(ii) $\text{Cov}\{Z_j(x + h), Z_j(x)\} = (\gamma_{jj}(h)$ exists and depends only on h and j,

where E denotes the expected value and Cov denotes the convariance. These conditions are sufficient for the cross-covariances to exist

$$\text{Cov}\{Z_j(x + h), Z_k(x)\} = C_{jk}(h) \quad [2]$$

Alternatively, the weaker assumption called the Intrinsic Hypothesis could be used

(i') $E\{Z_j(x)\} = m_j(x)$, a function of j and x,
(ii') $0.5\text{Var}\{Z_j(x + h) - Z_j(x)\} = \gamma_{jj}(h)$ exists and depends only on h and j,

where Var denotes the variogram. These conditions are sufficient for the cross-variogram to exist:

$$0.5\text{Cov}\{Z_j(x + h) - Z_j(x), Z_k(x + h) - Z_k(x)\} = \gamma_{jk}(h) \quad [3]$$

Whichever of these combinations is used, it is necessary to estimate and model correlation functions using the data. The advantage of using variograms and cross-variograms is that it is not necessary to separately estimate m. To estimate $\gamma_{jk}(h)$, it is necessary to find many pairs of sample locations with a separation h (or nearly so) such that there are data for both Z_j and Z_k at the locations. However, in the undersampled case, there may not be enough data pairs at sample locations. In an extreme case there may not be any such pairs. This problem can be alleviated by the use of the pseudo-cross-variogram, since it can be estimated even without such pairs. When using variograms and pseudo-cross-variograms, only minor changes are needed in the cokriging equations. In some cases the pseudo-cross-variograms can be used to model the cross-variograms.

Clark et al. (1989) and Myers (1991) defined the pseudo-cross-variogram of Z_j and Z_k as

$$\hat{g_{jk}}(h) = 0.5E[Z_j(x) - Z_k(x + h)]^2 \quad [4]$$

It is assumed that this function depends only on the separation h. The pseudo-cross-variogram is not necessarily symmetric, may not be zero at $h = 0$, and is not a variogam or a covariance. When $j = k$ and the function satisfies the Intrinsic Hypothesis, Eq. [4] becomes the variogram. Since in general the two random functions Z_j and Z_k do not have the same means, Eq. [4] is not the same as one-half the variance of the difference, even if both random functions separately satisfy the Intrinsic Hypothesis. A more general definition of the pseudo-cross-variogram is given as follows:

$$g_{jk}(h) = g_{kj}(-h) = 0.5\text{Var}[Z_j(x) - Z_k(x + h)] \quad [5]$$

Since the translation of a second-order stationary random function is also second-order stationary, Eq. [5] is the variance of a random function and depends only on h. However, the equation is not the covariance or the variogram of a random function unless $j = k$. If $Z(x) = Y(x) + m(x)$, where $Y(x)$ is second-order stationary and $m(x)$ is the mean of $Z(x)$, and if the pseudo-cross-variograms g_{jk}, g_{kj} are symmetric, Eq. [5] may be rewritten as (Myers, 1991)

$$g_{jk}(h) = 0.5(\sigma_j - 2\sigma_{jk} + \sigma_k) + \gamma_{jk}(h) \quad [6]$$

where σ_j and σ_k are the variances of variables j and k, respectively, σ_{jk} is the covariance between variables j and k at $h = 0$, and $\gamma_{jk}(h)$ is the cross-variogram. Equation [6] demonstrates that the pseudo-cross-variogram differs from the cross-variogram by a positive constant.

Cokriging Equations

The cokriging equations are normally obtained and expressed in terms of the variograms and cross-variograms or, alternatively, the covariances and cross-covariances. The following section provides a brief overview of how to adapt the cokriging equations to the pseudo-cross-variograms and details are given in Myers (1991).

If each component of $\mathbf{Z}(x)$ satisfies the Intrinsic Hypothesis, then Eq. [1] is unbiased if

$$\sum \boldsymbol{\Gamma}_i = [1,0,\ldots,0]^T \quad [7]$$

where T indicates a transpose (in Eq. [7] and in the following discussion, all single sums are taken from $i = 1$ to $i = n$ and the double sums are from $i = 1$ to $i = n$ and $j = 1$ to $j = n$). The variance of the error for variable Z_1 can be written in the form

$$\text{Var}[Z_1^*(x_0) - Z_1(x_0)] = \sum\sum \boldsymbol{\Gamma}_i^T E[W_i^T W_j]\boldsymbol{\Gamma}_j \quad [8]$$

where

$$W_i = Z_1(x_i) - Z_1(x_0), \ldots, Z_m(x_i) - Z_1(x_0) \quad [9]$$

By use of

$$E\{[Z_j(x) - Z_p(u)][Z_k(y) - Z_p(u)]\} = \hat{g_{jp}}(u - x) + \hat{g_{kp}}(u - y) - \hat{g_{jk}}(y - x) \quad [10]$$

and

$$\text{Cov}\{[Z_j(x) - Z_p(u)][Z_k(y) - Z_p(u)]\} = g_{jp}(u - x) + g_{kp}(u - y) - g_{jk}(y - x) \quad [11]$$

the "covariance" matrix in Eq. [8] can be written in the form

$$E[W_i^T W_j] = \hat{\mathbf{G}_p}(x_0 - x_i) + [\hat{\mathbf{G}_p}(x_0 - x_j)]^T - \hat{\mathbf{G}_p}(x_i - x_j) \quad [12]$$

where

$$\hat{\mathbf{G}_p}(x - y) = \begin{vmatrix} \hat{g_{1p}}(x - y) & \cdots & \hat{g_{1p}}(x - y) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \hat{g_{mp}}(x - y) & \cdots & \hat{g_{mp}}(x - y) \end{vmatrix} \quad [13]$$

and

$$\hat{\mathbf{G}}(x - y) = \begin{vmatrix} \hat{g_{11}}(x - y) & \cdots & \hat{g_{1m}}(x - y) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \hat{g_{m1}}(x - y) & \cdots & \hat{g_{mm}}(x - y) \end{vmatrix} \quad [14]$$

Using Eq. [7], the right-hand side of Eq. [8] can be written as

$$\sum\sum \boldsymbol{\Gamma}_i^T E[W_i^T W_j]\boldsymbol{\Gamma}_j = 2\sum[\hat{g_{11}}(x_0 - x_j), \ldots, \hat{g_{1m}}(x_0 - x_j)]\boldsymbol{\Gamma}_j - \sum\sum \boldsymbol{\Gamma}_i^T \hat{\mathbf{G}}(x_i - x_j)\boldsymbol{\Gamma}_j \quad [15]$$

To minimize the variance of the error of an unbiased estimate, m Lagrange multipliers $\mu_1, \ldots, \mu_m$ are introduced.

The cokriging equations are then

$$\sum \hat{\mathbf{G}}(x_i - x_j)\boldsymbol{\Gamma}_j + [\mu_1,\ldots,\mu_m]^T = [\hat{g_{11}}(x_0 - x_i),\ldots,\hat{g_{1m}}(x_0 - x_i)]^T \quad [16]$$

for $i = 1, \ldots, n$. Unlike the matrix variogram function used in Myers (1982), $\hat{\mathbf{G}}(h)$ does not need to be a symmetric matrix. For the special case where $\hat{g_{jk}}(h) = g_{jk}(h) + a_{jk}$ with $a_{jj} = 0$, the above equations are the same as the usual system of co-kriging equations in terms of variograms and cross-variograms.

METHODS

The sample pseudo-cross-variograms for Variable 1 (Z_1) and Variable 2 (Z_2) are computed by means of

$$g^*_{12}(h) = \frac{1}{2N}\sum_{i=1}^{N}[Z_1(x_i) - Z_2(x_i + h)]^2 \quad [17]$$

and

$$g^*_{21}(h) = \frac{1}{2N}\sum_{i=1}^{N}[Z_2(x_i) - Z_1(x_i + h)]^2 \quad [18]$$

where N is the number of sample pairs for lag h. If $g_{12}{}^*(h)$ and $g_{21}{}^*(h)$ are symmetrical, i.e., $g_{12}(h) = g_{21}(h)$, we may obtain a cross-variogram from the pseudo-cross-variogram by utilizing the relationship between g_{12} and γ_{12} (Eq. [6]). The nugget of the cross-variogram is determined by cross-validation. The cross-validation procedure involves taking the pseudo-cross-variogram with a smaller nugget value as a cross-variogram, and producing estimates corresponding to measured locations using the cokriging technique. In this procedure, every known point is estimated by using a neighborhood around it, but not itself. Based on the statistical analysis of the estimates and measurements, an appropriate nugget of the cross-variogram can be chosen. The choice of a nugget of the cross-variogram should result in a near-zero value of mean error, and near unity for the reduced kriging variance.

The following methods are used to compare results estimated with kriging and cokriging. The MSE is calculated with

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}[Z(x_i) - Z^*(x_i)]^2 \quad [19]$$

where $Z(x_i)$ and $Z^*(x_i)$ are the measured and estimated values at x_i, respectively. Relative improvement, or relative reduction, of MSE is defined by

$$100(\text{MSE}_k - \text{MSE}_{ck})/\text{MSE}_k \quad [20]$$

where MSE_k and MSE_{ck} are the mean squared error for kriging and cokriging, respectively. Relative reduction of kriging variance is defined in the same way by replacing MSE_k and MSE_{ck} in Eq. [20] with kriging variances of kriging and cokriging, respectively.

APPLICATION

The estimation procedure using pseudo-cross-variograms was applied to a large data set of soil chemical properties as described by Warrick et al. (1991). The data were collected in May of 1989 near Marana in Arizona. The experiment involved 130 measurement points over an area of 1000 by 1000 m. Two soil series were present, the Mohave (fine-loamy, thermic Typic Haplargid) and the Tubac (fine, mixed, thermic Typic Paleargid). Both series are deep, well-drained soils. Soil texture ranged from clay to sandy loam. One hundred sampling points were obtained from a 100 by 100 m grid, whereas an additional 30 locations were optimally distributed. Optimization was with respect to fitting a predefined distribution of sample separation (Warrick and Myers, 1987). Figure 1a shows the locations of the 130 points. Samples were air dried and sieved, and extracts were made from a 1:1 mixture of dry soil to water by weight. Correlations (r^2) between several physical and chemical parameters of the 130 soil samples are listed in Table 1. The variables EC, NO_3, and Ca were used for the following analysis. The EC was scaled to the same magnitude of NO_3 by multiplying by 100. The scaling does not affect accuracy of kriging or co-kriging results, although kriging variances may become larger. The descriptive statistics for the nontransformed and transformed data are given in Table 2 and include the mean, variance, minimum, maximum, skewness, and kurtosis of the data set. The results suggest that the soil chemical properties are more closely log-normally distributed. Therefore, a log transformation was performed on the variables. In the following discussion, all results were based on the log-transformed data.

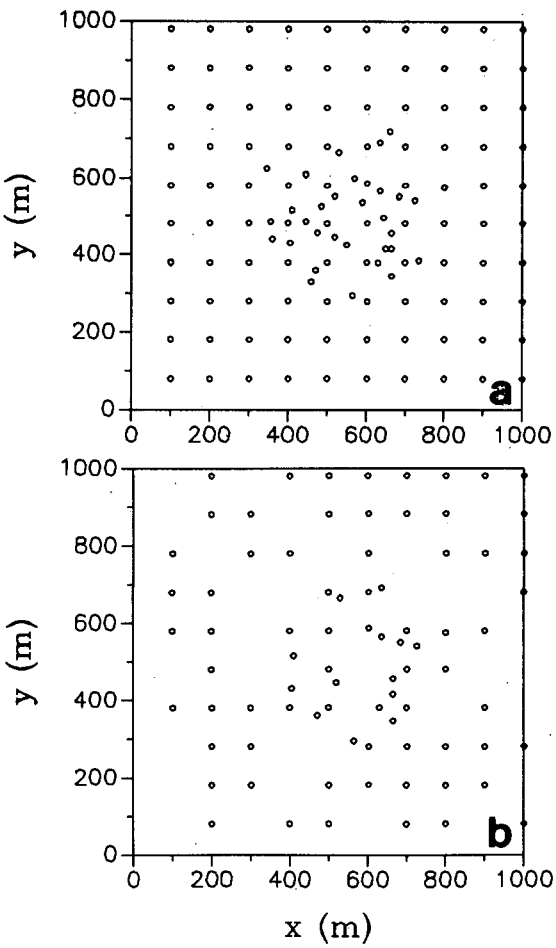

Fig. 1. Locations of the (a) 130 soil samples and (b) 80 soil samples at the Marana Farm.

Table 1. Correlations (r^2) of physical and chemical parameters of 130 soil samples from the Marana Farm.

	Gravel	Sand	Silt	Clay	pH	EC†	Cl	NO_3	SO_4	Na	Ca	Mg
Gravel	1	0.612	0.359	0.545	0.028	0.077	0.064	0.014	0.134	0.114	0.053	0.059
Sand		1	0.613	0.871	0.068	0.106	0.133	0.017	0.272	0.166	0.076	0.077
Silt			1	0.258	0.074	0.055	0.048	0.043	0.040	0.036	0.064	0.042
Clay				1	0.041	0.099	0.143	0.004	0.210	0.207	0.055	0.071
pH					1	0.166	0.061	0.197	0.106	0.106	0.159	0.183
EC†						1	0.662	0.689	0.524	0.785	0.949	0.959
Cl							1	0.317	0.441	0.699	0.561	0.558
NO_3								1	0.096	0.333	0.750	0.693
SO_4									1	0.773	0.416	0.492
Na										1	0.615	0.679
Ca											1	0.956
Mg												1

† Electrical conductivity.

Cokriging between Nitrates and Electrical Conductivity

As shown in Table 1, NO_3 has a relatively high correlation with EC ($r^2 = 0.689$), and EC is much more easily measured than NO_3. It seems reasonable to try to improve the estimation of NO_3 by using information of EC with relatively few measurements of NO_3. We assumed that 130 EC and 80 NO_3 data points were available and used them to compute the pseudo-cross-variogram and variograms. The locations of the 80 points chosen randomly from the original 130 points are shown in Fig. 1b. Plotted in Fig. 2 are the sample variograms for EC and NO_3, along with the fitted variograms, which are determined with cross-validation. The fitted variograms are linear models of the forms

$$\gamma_1(h) = 0.08 + 1.8(10)^{-4}h \quad [21]$$

$$\gamma_2(h) = 0.23 + 4.7(10)^{-4}h \quad [22]$$

for EC and NO_3, respectively. The sample pseudo-cross-variograms for Variable 1 (EC) and Variable 2 (NO_3), $g_{12}^*(h)$ and $g_{21}^*(h)$, are plotted in Fig. 2c. The sample pseudo-cross-variograms are quite symmetrical, as illustrated by the fact that the two sets of results fall on nearly the same line in Fig. 2c. The fitted pseudo-cross-variogram is also a linear model,

$$g_{12}(h) = 0.15 + 3.0(10)^{-4}h \quad [23]$$

The nugget of the cross-variogram of EC and NO_3 was determined by cross-validation, while its slope was taken from Eq. [23]. Table 3 presents results of cross-validation, using the data sets of 80 NO_3 and 130 EC points, the variograms of Eq. [21] and [22], and a cross-variogram of the form

$$\gamma_{12}(h) = 0.055 + 3.0(10)^{-4}h \quad [24]$$

For comparison, the results of cross-validation for the variogram of NO_3 (Eq. [22]) are also shown in Table 3. The entries in Columns 2 and 4 indicate that supplemental information of EC improves the estimation of NO_3, resulting in a smaller mean error (27% reduction), a smaller mean sum of square error (43% reduction), and a lower mean kriging variance (25% reduction), as well as a higher correlation (r^2) between the estimates and the actual measurements.

Using the same approach with only 20 NO_3 data points, as shown in Fig. 3a, and again all 130 EC measurements, we obtained a pseudo-cross-variogram

$$g_{12}(h) = 0.075 + 3.25(10)^{-4}h \quad [25]$$

which is shown in Fig. 3b with the sample pseudo-cross-variograms. The cross-validation results for this data set and the variogram models of Eq. [21] and [22] are summarized in Table 4. The cross-variogram model in this case was

$$\gamma_{12}(h) = 0.05 + 3.25(10)^{-4}h \quad [26]$$

based on the pseudo-cross-variogram (Eq. [25]).

Using the 20 NO_3 data and the 130 EC points, and the variograms and cross-variogram, we cokriged and kriged 100 NO_3 data points on the 100 by 100 m grid where measured data were available. The MSE between predicted and measured values for cokriging and kriging were 0.297 and 0.463, respectively. Thus, cokriging reduced the MSE by 36%.

The estimated errors and kriging variance using kriging and cokriging are illustrated in Fig. 4a and 4b for the 87 estimates (the 13 data points out of 100 estimates that were estimated exactly with both techniques were not plotted). The results clearly show that kriging has a larger estimated error, and a larger kriging variance than cokriging. The contour maps of NO_3 shown in Fig. 5a, 5b, and 5c were generated using information from the original 100 measurements, the 100 cokriging estimates, and the 100 kriging estimates, respectively. Cokriging gives a similar pattern

Table 2. Descriptive statistics for the electrical conductivity (EC), and NO_3 and Ca contents (130 soil samples).

Variable	Mean	Variance	Min.	Max.	Skewness	Kurtosis
			Nontransformed data			
EC	0.694	0.126	0.280	3.45	3.92	29.6
NO_3	96.60	8720	13.1	844	4.53	33.5
Ca	57.0	1510	18.7	380	4.88	38.7
			Transformed data			
EC	4.15	0.161	3.33	5.84	0.564	4.24
NO_3	4.31	0.474	2.57	6.74	0.402	3.47
Ca	3.91	0.223	2.93	5.94	0.791	4.65

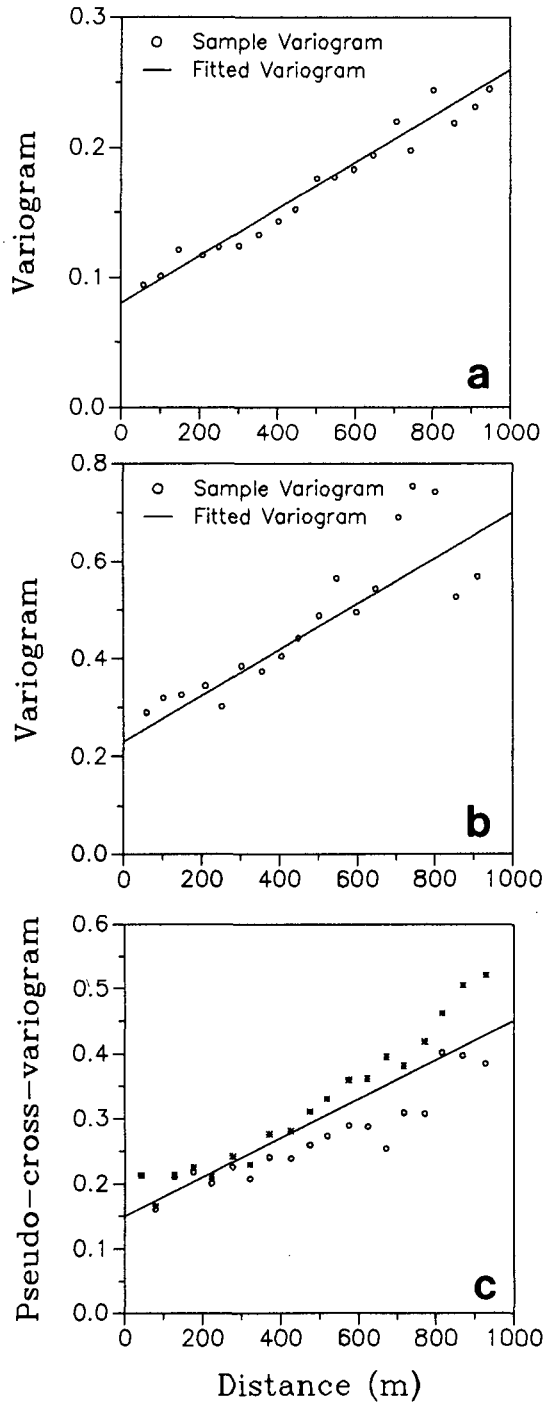

Fig. 2. Variograms of (a) electrical conductivity (EC) using 130 data points, (b) NO_3 using 80 data points, and (c) the pseudo-cross-variograms of NO_3 and EC using 80 NO_3 and 130 EC data. The asterisks and open circles in (c) represent the sample pseudo-cross-variogram g_{12}* and g_{21}*, respectively, whereas the solid line represents the fitted pseudo-cross-variogram.

Table 3. Summary statistics of cross-validation for the cross-variogram (80 NO_3 and 130 electrical conductivity [EC] data points) and variogram of NO_3.

	Cross-Variogram		Variogram
	NO_3	EC	NO_3
Number of validation points	80	130	80
Mean error	−0.024	0.003	−0.033
Mean square error	0.217	0.097	0.378
Mean kriging variance	0.229	0.088	0.307
Reduced kriging variance	0.948	1.032	1.213
Correlation of estimates and error	−0.164	0.050	0.048
Correlation of estimates and measurements	0.708	0.630	0.426

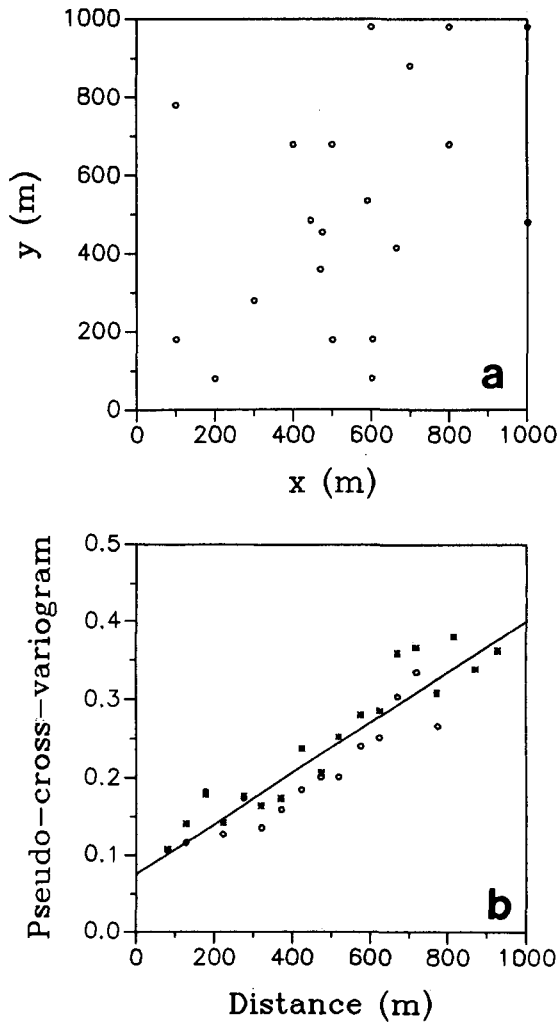

Fig. 3. (a) Locations of the 20 soil samples and (b) pseudo-cross-variograms of NO_3 and electrical conductivity (EC) assuming 20 NO_3 and 130 EC data. The asterisks and open circles in (b) represent the sample pseudo-cross-variograms g_{12}* and g_{21}*, respectively, whereas the solid line represents the fitted pseudo-cross-variogram.

of the NO_3 distribution as the original data. On the other hand, kriging smoothes the estimates and gives a quite different pattern of the NO_3 distribution.

It should be noted that, with the 20 pairs of NO_3 and EC as shown in Fig. 3a, the cross-variogram is not well defined if standard methods are used to compute their cross-variogram. The sample variograms of the sums and differences for the two variables are shown in Fig. 6a and 6b, respectively, with a maximum sample pair of 17. The irregular behavior and deficient sample pairs make it impossible to fit theoretical models. Therefore, cokriging could not be per-

Table 4. Summary statistics of cross-validation (20 NO_3 and 130 electrical conductivity [EC] data points).

	NO_3	EC
Number of validation points	20	130
Mean error	−0.069	0.004
Mean square error	0.255	0.123
Mean kriging variance	0.280	0.101
Reduced kriging variance	0.934	1.194
Correlation of estimates and error	0.114	0.138
Correlation of estimates and measurements	0.426	0.495

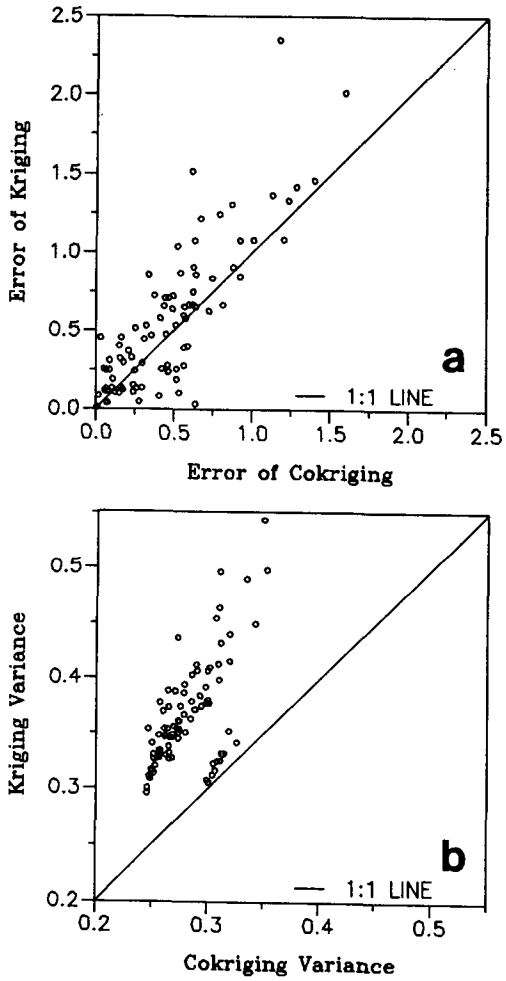

Fig. 4. Plots of (a) estimated error, and (b) kriging variance with kriging and cokriging.

formed. In contrast to the standard methods of computing cross-variograms, pseudo-cross-variograms can be computed with even fewer data pairs at the same sample locations. Figure 6c presents sample pseudo-cross-variograms computed with 10 NO_3 and 130 EC points (only 10 pairs of NO_3 and EC at common locations), and a fitted model.

Cokriging between Calcium and Electrical Conductivity

This second example shows that estimates of Ca are improved significantly by cokriging using pseudo-cross-variograms of Ca and EC. The following analyses

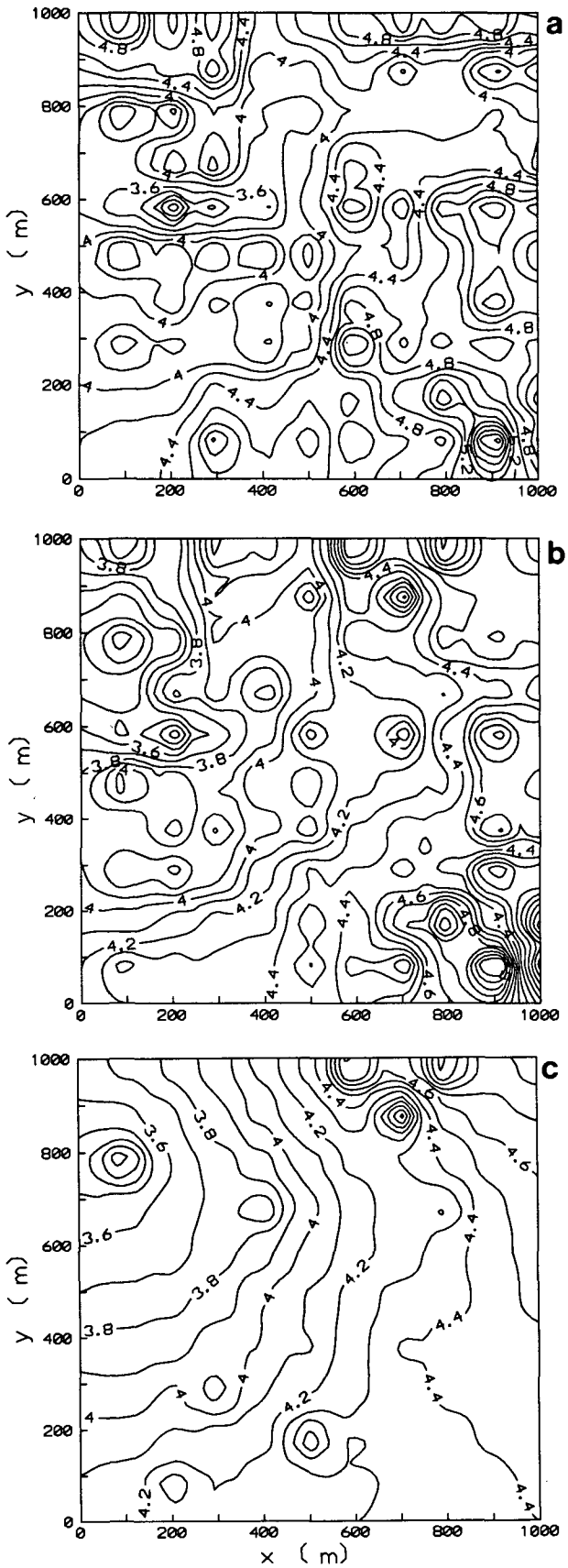

Fig. 5. Contour maps of NO_3 calculated from (a) 100 data points, (b) 100 estimates by cokriging with 20 NO_3 and 130 electrical conductivity (EC) data points, and (c) 100 estimates by kriging with the same 20 NO_3 points as cokriging.

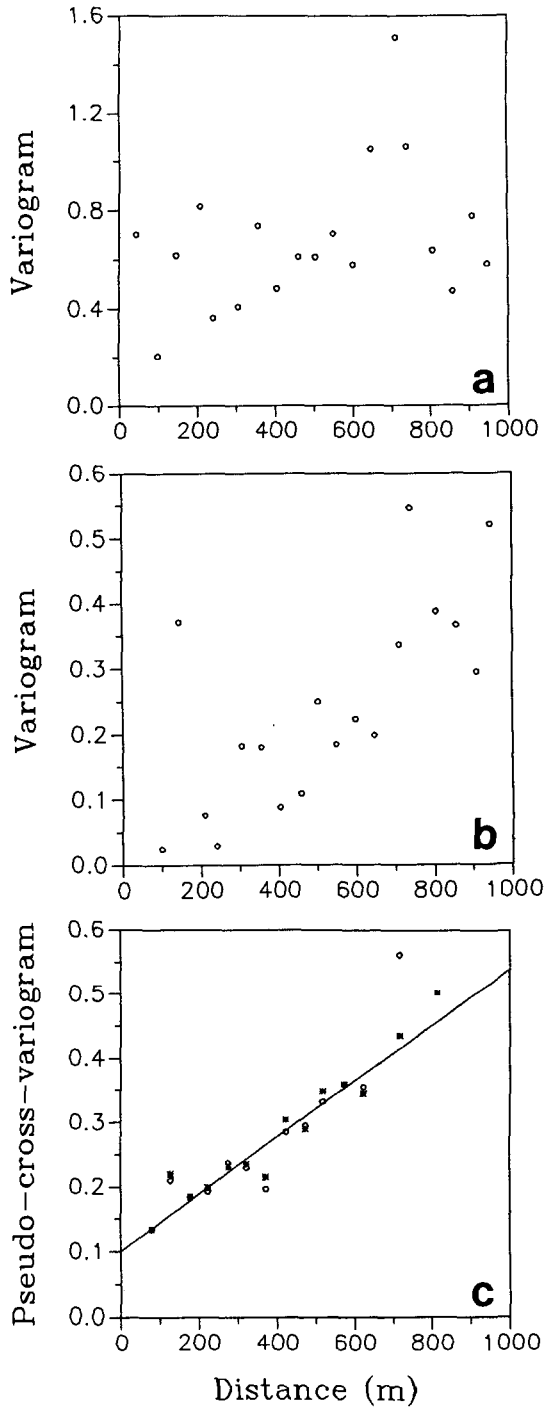

Fig. 6. The sample variograms of (a) the sum of NO_3 and electrical conductivity (EC), (b) the difference between NO_3 and EC using 20 pairs of NO_3 and EC data, and (c) pseudo-cross-variograms of NO_3 and EC using 10 NO_3 and 130 EC data. The asterisks and open circles in (c) represent the sample pseudo-cross-variograms g_{12}* and g_{21}*, respectively, whereas the solid line represents the fitted pseudo-cross-variogram.

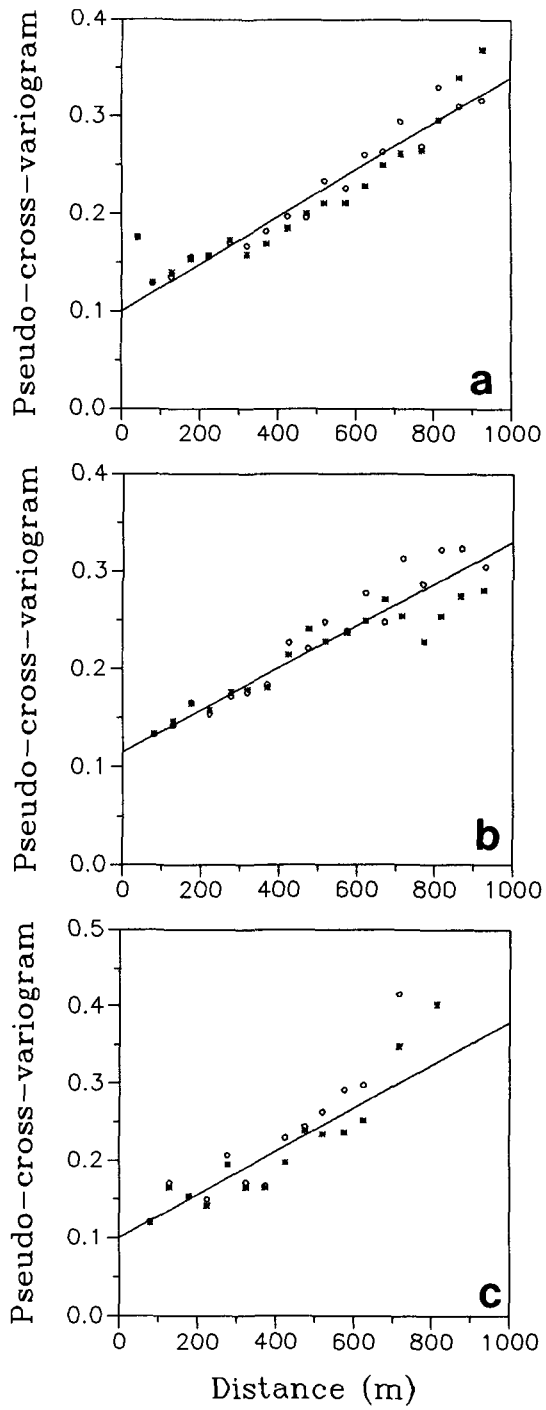

Fig. 7. Pseudo-cross-variograms of Ca and electrical conductivity (EC), based on (a) 80 Ca and 130 EC data, (b) 50 Ca and 130 EC data, and (c) 10 Ca and 130 EC data. The asterisks and open circles represent the sample pseudo-cross-variograms g_{12}* and g_{21}*, respectively, whereas the solid lines represent the fitted pseudo-cross-variograms.

were based on a linear variogram of Ca, having a nugget of 0.06 and slope of 4.8×10^{-4}, which was computed using 50 data points randomly chosen from the original 130 points. The choice of the variogram was considered in two aspects. First, computation of the sample variogram only requires a small fraction of the total number of available data (in terms of lower sampling cost). Second, the kriging interpolation processes are not very sensitive to the model parameters in terms of accuracy (Warrick et al., 1988). By assuming different "known" data points, seven subsets of Ca were randomly selected from the original

Table 5. Summary statistics of cross-validation for Ca; the cross-variograms were selected based on the pseudo-cross-variograms computed from 130 electrical conductivity (EC) data points and different numbers of data points of Ca.

Statistic	Number of validation points						
	80	60	50	40	30	20	10
Mean error	−0.007	0.003	0.002	0.001	0.008	−0.082	−0.062
Mean square error	0.074	0.028	0.040	0.093	0.026	0.124	0.038
mean kriging variance	0.072	0.031	0.039	0.089	0.027	0.112	0.042
Reduced kriging variance	1.022	1.012	1.057	1.226	1.236	1.183	1.090
Correlation of estimates and error	−0.391	−0.354	−0.316	−0.178	−0.216	−0.067	0.318
Correlation of estimates and measurements	0.867	0.943	0.931	0.795	0.934	0.594	0.799

data set. These data sets consisted of 80, 60, 50, 40, 30, 20, and 10 data points. Using these subsets and 130 EC measurements, we computed sample pseudo-cross-variograms. All of the sample pseudo-cross-variograms' g_{12}^* and g_{21}^* were symmetrical and linear. Results of the pseudo-cross-variograms are shown in Fig. 7a, 7b, and 7c for 80, 50, and 10 known Ca data points, respectively. The cross-variograms were modeled based on the pseudo-cross-variograms. Nuggets of cross-variograms of the data sets were determined by cross-validation. Table 5 presents the summary statistics of cross-validation for Ca. The results again show that a valid cross-variogram can be obtained with a dramatically reduced data set. Table 6 summarizes the intercepts (a) and the slopes (b) of the pseudo-cross-variograms and the nuggets of the cross-variograms. For instance, the pseudo-cross-variogram between Ca (80 data points) and EC (130 data points) is

$$g_{12}(h) = 0.10 + 2.4(10)^{-4}h \quad [27]$$

while the cross-variogram is

$$\gamma_{12}(h) = 0.04 + 2.4(10)^{-4}h \quad [28]$$

The nuggets of the linear cross-variogram range from 0.03 to 0.07, and the slopes range from 2 to 3. Using the variograms of Ca and EC and their cross-variogram, as well as each subset, kriging and cokriging were carried out to estimate the 100 points on the 100 by 100 m grid. The MSE, the correlation between estimates and measurements (r^2), and the mean kriging variance were calculated from the predicted and measured values. Cokriging always gave a smaller MSE, a lower mean kriging variance, and a higher correlation between estimates and measurements than kriging. These results are presented in Table 6. For cokriging, the kriging variance is not fully related to the number of Ca points because EC plays an important role in the estimation of Ca. The improvement obtained by cokriging may be quantified by the difference in the kriging variances for kriging and cokriging. As shown in Table 6, the relative reductions in MSE and in mean kriging variance by cokriging are up to 78 and 85%, respectively, compared with kriging.

We also estimated 625 points on a 50 by 50 m grid by kriging with 30 Ca data points, and by cokriging assuming the same 30 Ca and 130 EC data points. Cokriging reduces the kriging variance by up to 40%, compared with kriging. For most of the estimates, the reduction is about 20%.

For this example, the correlation between Ca and EC is high; therefore, simple linear regression may give satisfactory estimates of Ca by using information on EC. Based on the 30 Ca and 130 EC data points, linear regression gives an MSE of 0.040 and a correlation between estimates and measurements of 0.917 for 100 Ca estimates on the 100 by 100 m grid. The results are comparable to those in Table 6 (MSE = 0.033 and r^2 = 0.928). In general, however, much more accurate results are obtained by cokriging than by linear regression if the correlation of two variables is not so high. Additional advantages to using cokriging include an estimation that can be carried out anywhere in a studied domain, and a kriging variance that is provided in the estimation process.

Table 6. Estimated parameters of pseudo-cross-variograms of Ca and electrical conductivity (EC) (130 points), mean squared error (MSE), correlation between estimates and measurements, and mean kriging variance of 100 estimates of Ca by kriging and cokriging.

Data points of Ca	Pseudo-cross-variogram		MSE		Correlation of estimates and measurements		Mean kriging variance	
	a†	$b \times 10^{-4}$‡	kriging	cokriging	kriging	cokriging	kriging	cokriging
80	0.10 (0.04§)	2.4	0.055	0.026 (53¶)	0.875	0.942	0.0059	0.0029 (51#)
60	0.125 (0.068)	2.05	0.104	0.033 (68)	0.743	0.928	0.0082	0.0015 (82)
50	0.115 (0.065)	2.15	0.202	0.046 (77)	0.635	0.902	0.0100	0.0019 (81)
40	0.15 (0.04)	2.2	0.153	0.054 (65)	0.613	0.891	0.0136	0.0065 (52)
30	0.11 (0.07)	2.5	0.151	0.033 (78)	0.616	0.928	0.0200	0.0030 (85)
20	0.145 (0.03)	3.0	0.246	0.085 (65)	0.405	0.847	0.0227	0.0112 (51)
10	0.10 (0.07)	2.8	0.195	0.079 (59)	0.573	0.850	0.0509	0.0135 (73)

† Intercepts of the linear pseudo-cross-variograms.
‡ Slopes of the linear pseudo-cross-variograms.
§ Values in parentheses in this column are nuggets of the linear cross-variograms of Ca and EC; their slopes are the same as those of pseudo-cross-variograms.
¶ Values in parentheses in this column indicate relative reductions in MSE.
# Values in parentheses in this column indicate relative reductions in the mean kriging variance.

CONCLUSIONS

In this study, pseudo-cross-variograms were used to compute the spatial distribution of NO_3 and Ca across a 1000 by 1000 m field. The results suggest that the approach using pseudo-cross-variograms overcomes several of the shortcomings in the standard methods for computing cross-variograms. The main advantage of the pseudo-cross-variogram approach is that the computation of sample cross-variograms does not require that all variables be measured at the same locations. Therefore, the method maximizes the use of available information.

For our examples, valid cross-variograms of NO_3 and EC, as well as Ca and EC, were obtained based on their pseudo-cross-variograms. The pseudo-cross-variograms were computed from one-thirteenth (10 vs. 130) of the original NO_3 or Ca with the EC data. Using the reduced NO_3 data set, cokriging gave more accurate predicted results for NO_3 than kriging, as shown by the smaller MSE and kriging variance. Estimation of NO_3 from 20 data points using cokriging correctly showed the measured pattern of the chemical distribution contoured from 100 data points, whereas kriging produced a much smoother pattern. For different Ca subsets, cokriging always provided more accurate estimates for Ca than kriging. In general, as the number of data points increased, MSE and kriging variance decreased, and the correlation between estimates and measurements increased. The relative reductions in MSE and mean kriging variance were as much as 78 and 85% when cokriging was used.

Cokriging with the help of pseudo-cross-variograms may be performed in a similar way as discussed above to predict other soil chemicals, such as Cl, SO_4, Na, and Mg in the data set, using the information of EC. Cokriging can also be employed to estimate NO_3, using other chemicals besides EC as auxiliary variables to compute pseudo-cross-variograms.

When sample pseudo-cross-variograms are not symmetrical, the representation for pseudo-cross-variograms is more complicated than for cross-variograms. The conditions for verifying positive-definiteness are also more complicated. Further research is being carried out in such situations.

REFERENCES

Burgess, T.M., R. Webster, and A.B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. J. Soil Sci. 31:643–659.

Carr, J.R., and D.E. Myers. 1984. Applications of the theory of regionalized random variables to the spatial analysis of Landsat data. p. 55–61. *In* IEEE Computer Soc. Proc. Pecora 9 Spatial Techno. Remote Sensing Today and Tomorrow. IEEE Computer Soc. Press, Silver Spring, MD.

Clark, I., K.L. Basinger, and W.V. Harper. 1989. MUCK — A novel approach to co-kriging. p. 473–494. *In* B.E. Buxton (ed.) Proc. Conf. geostatistical, sensitivity and uncertainty methods for ground-water flow and radionuclide transport modeling. Batelle Press, Columbus, OH.

Hatfield, J.L., M. Vauclin, S.R. Vieira, and R. Barnard. 1984. Surface temperature variability patterns within irrigated fields. Agric. Water Manage. 8:429–437.

McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Corregionalization and multiple sampling strategy. J. Soil Sci. 34:137–162.

Morkoc, F., J.W. Biggar, D.R. Nielsen, D.E. Rolston. 1985. Analysis of soil water contents and temperature using state-space approach. Soil Sci. Soc. Am. J. 49:798–803.

Mulla, D.J. 1988. Estimating spatial patterns in water content, matric suction, and hydraulic conductivity. Soil Sci. Soc. Am. J. 52:1547–1553.

Myers, D.E. 1982. Matrix formation of co-kriging. Math. Geol. 14:249–257.

Myers, D.E. 1984. Co-kriging — New developments. p. 205–305. *In* G. Verly et al. (ed.) Geostatistics for natural resources characterization. Part 1. D. Reidel Publ. Co., Hingham, MA.

Myers, D.E. 1991. Pseudo-cross variograms, positive-definiteness, and cokriging. Math. Geol. 23:805–816.

Stein, A., M. Hoogerwerf, and J. Bouma. 1988. Use of soil-map delineations to improve (co)-kriging of point data on moisture deficits. Geoderma 43:163–177.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175–184.

Warrick, A.W., S.A. Musil, J.F. Artiola, D.M. Hendricks, and D.E. Myers. 1991. Sampling strategies for hydrological properties and chemical constituents in the upper vadose zone. USDI Water Resour. Res. Grant Program Final Tech. Rep. BP91-234724, Nat. Tech. Inform. Serv., Springfield, VA.

Warrick, A.W., D.E. Myers. 1987. Optimization of sampling locations for variogram calculations. Water Resour. Res. 23:496–500.

Warrick, A.W., D.E. Myers, and D.R. Nielsen. 1986. Geostatistical methods applied to soil science. p. 53–82. *In* A. Klute (ed.) Methods of soil analysis. Part 1. 2nd ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI.

Warrick, A.W., R. Zhang, M.K. El-Haris, and D.E. Myers. 1988. Direct comparisons between kriging and other interpolators. p. 505–510. *In* P.J. Wierenga and D. Bachelet (ed.) Int. Conf. Worksh. validation of flow and transport models for the unsaturated zone. New Mexico State Univ., Las Cruces.

Yates, S.R., and A.W. Warrick. 1987. Estimating soil water content using cokriging. Soil Sci. Soc. Am. J. 51:23–30.

Yost, R.S., G. Uehara, and R.L. Fox. 1982a. Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms. Soil Sci. Soc. Am. J. 46:1028–1032.

Yost, R.S., G. Uehara, and R.L. Fox. 1982b. Geostatistical analysis of soil chemical properties of large land areas. II. Kriging. Soil Sci. Soc. Am. J. 46:1033–1037.

Zhang, R., A.W. Warrick, and D.E. Myers. 1992. Improvement of soil textural estimates using spectral properties. Geoderma 52:223–234.