

Use of Pseudo-Crossvariograms and Cokriging to Improve Estimates of Soil Solute Concentrations

R. Zhang,* P. Shouse, and S. Yates

ABSTRACT

Estimating mass and distributions of chemicals in soils is one of the key steps to study chemical transport in the vadose zone and groundwater systems. In this study, solute mass and distributions were computed in an 800 by 800 by 1.8 m soil volume using kriging and cokriging with nonsymmetric pseudo-crossvariograms. Among the measured chemicals of Cl^- , SO_4^{2-} , Ca^{2+} , Mg^{2+} , Na^+ , P_2O_5 , K^+ , and NO_3^- in the three-dimensional system, Cl^- was used as the model to show the estimation process and results. Using pseudo-crossvariograms maximized the use of available information at different soil depths and improved solute estimation. All sample pseudo-crossvariograms between depths were modeled successfully with common variogram functions such as spherical and linear. Therefore, it was relatively easier to test the positive definiteness of the cokriging coefficient matrix adapted to pseudo-crossvariograms. Cokriging allows easily obtained information at shallower depths to be used to improve solute estimations at deeper depths. Compared with kriging, cokriging reduced the mean squared errors of estimations between 30 and 60% at different depths, and reduced the mean kriging variances between 35 and 58%. In the total mass estimation of Cl^- in the soil, cokriging with nonsymmetric pseudo-crossvariograms used less than half the data; potentially it could reduce more than half the sampling cost of kriging estimation. Meanwhile, cokriging reduced the estimation error by about 18%, when compared with kriging estimates using all observations. Using the same data at each layer, cokriging would reduce the estimation error 40% more than kriging because cokriging efficiently incorporated information at upper layers without increasing the sample requirement. Cokriging with nonsymmetric pseudo-crossvariograms is an accurate and economical way to calculate solute distributions and total mass in a large field.

ACCURATELY ESTIMATING SOLUTE MASS and distributions in the vadose zone and groundwater systems is critical in many environmental studies, such as characterizing hazardous waste fields, designing best management practices of irrigation systems, and modeling chemical movement and fate in soils. Field soil sampling and analysis for chemicals is a direct way to measure solute distributions; however, it is often too expensive to carry out detailed sampling in a large field. With limited data, geostatistical analyses provide an alternative means to characterize the spatial distribution of soil properties (Warrick et al., 1986), and therefore offer an economic way to estimate solute mass and distributions in soils.

Geostatistics can use interrelationships between two or more spatially dependent variables to improve the estimation of the variables. The interrelationships can be among different variables or among different depths for the same variable. Cokriging is an extension of the kriging method and incorporates both spatial and intervariable correlations into the estimation process. Cokriging has been applied to study various spatial variables. For example, Yates and Warrick (1987) estimated soil water content using a cokriging procedure in which the bare soil surface temperature and the sand content

were used to supply additional information. Stein et al. (1988) used cokriging to increase computation precision in moisture deficit maps. Zhang et al. (1992b) improved the estimates of soil texture by including associated spectral properties in a cokriging procedure. Zhang et al. (1995) estimated concentrations of trace elements in soils and plants using kriging and cokriging.

An essential part of any geostatistical investigation using cokriging is the characterization of spatially dependent variables or the formulation of crossvariograms among the variables. A major disadvantage of standard approaches for modeling crossvariograms is that only values for the properties of interest having common sample locations can be used to estimate the cross-correlation functions. Therefore, estimating crossvariograms requires a large number of locations where data are collected for each variable, a condition that is frequently not satisfied in practice. To alleviate this difficulty, Clark et al. (1989) described a variation of cokriging that does not require measurements of variables at the same locations. The approach involves the development of pseudo-crossvariograms and their use in a cokriging procedure. Zhang et al. (1992a) have successfully used cokriging with symmetric pseudo-crossvariograms to estimate the spatial distribution of soil chemical concentrations.

In this study, our first objective was to study relationships between nonsymmetric sample pseudo-crossvariograms and valid models of pseudo-crossvariograms to estimate solute concentrations at different soil depths. The second objective was to compare the accuracy and advantages of using pseudo-crossvariograms and cokriging to estimate solute distributions and mass in the soil with kriging when limited data is available at deeper locations.

THEORY

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

$$\mathbf{Z}^*(x) = \sum_{i=1}^n \mathbf{Z}(x_i) \Gamma_i \quad [1]$$

where Γ_i s are weight matrices.

To determine the weight matrices, the variables Z_j ($j = 1, 2, \dots, m$) 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 estimation variances is minimized. As shown in Myers (1982) this leads to a linear system of matrix equations. Variograms and crossvariograms are used to quantify the spatial correlation of each variable and also the intervariable correlation. These statistical techniques are utilized to compute the variances of the estimation errors.

R. Zhang, Dep. of Plant, Soil and Insect Sciences, Univ. of Wyoming, Laramie, WY 82071-3354; P. Shouse and S. Yates, U.S. Salinity Lab., USDA-ARS, Riverside, CA 92507-4617. Received 2 May 1996. *Corresponding author (renduo@uwyo.edu).

To maximize the use of available data, a pseudo-crossvariogram of Z_j and Z_k is defined as (Myers, 1991)

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

It is assumed that this function depends only on the separation h . Note that it is not required that Z_j and Z_k are sampled at the same locations. The pseudo-crossvariogram is not necessarily symmetric, may not be zero at $h = 0$, and is not a variogram or a covariance. A more general definition of the pseudo-crossvariogram is given by Myers (1991) as

$$g_{jk}(h) = g_{kj}(-h) = 0.5\text{Var}[Z_j(x) - Z_k(x + h)] \quad [3]$$

Since the translation of a second-order stationary random function is also second-order stationary, Eq. [3] is the variance of a random function and depends only on h .

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

$$g_{12}^*(h) = \frac{1}{2N} \sum_{i=1}^N [Z_1(x_i) - Z_2(x_i + h)]^2 \quad [4]$$

and

$$g_{21}^*(h) = \frac{1}{2N} \sum_{i=1}^N [Z_2(x_i) - Z_1(x_i + h)]^2 \quad [5]$$

where N is the number of sample pairs for lag h . If $g_{12}^*(h)$ and $g_{21}^*(h)$ are symmetrical, we may obtain a crossvariogram γ_{12} from the pseudo-crossvariogram g_{12} (Myers, 1991; Zhang et al., 1992a). If $g_{12}^*(h)$ and $g_{21}^*(h)$ are nonsymmetrical, we assumed in this study that they can be modeled by combining the common variogram functions, such as exponential, Gaussian, spherical, and linear models. Parameters of the pseudo-crossvariogram are determined by cross validation. The cross-validation procedure produces estimates corresponding to measured locations using the cokriging technique to be discussed below. 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 pseudo-crossvariogram can be chosen. Similar to the selections of variograms and crossvariograms through cross validation (Yost et al., 1982a,b), the choice of a pseudo-crossvariogram should result in a near-zero value of mean error, and near unity for the reduced kriging variance.

The cokriging equations are normally obtained and expressed in terms of the variograms and crossvariograms. The general cokriging equations adapted to the pseudo-crossvariograms are given in Myers (1991) as

$$\sum_{j=1}^n G(x_i - x_j) \Gamma_j + [\mu_1, \dots, \mu_m]^T = [\gamma_1(x_0 - x_i), \dots, g_{1m}(x_0 - x_i)]^T \quad [6]$$

for $i = 1, \dots, n$. In the cokriging equation,

$$\sum_{i=1}^n \Gamma_i = [1, 0, \dots, 0]^T \quad [7]$$

where T indicates a transpose, and μ_1, \dots, μ_m are Lagrange multipliers. The cokriging coefficient matrix \mathbf{G} is in terms of pseudo-crossvariograms in a form of

$$\mathbf{G}(h) = \begin{bmatrix} \gamma_1(h) & \dots & g_{1m}(h) \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ g_{m1}(h) & \dots & \gamma_m(h) \end{bmatrix} \quad [8]$$

in which γ is the variogram and g is the pseudo-crossvariogram. Unlike the cokriging coefficient matrix used in Myers (1982, 1984), $\mathbf{G}(h)$ does not need to be a symmetric matrix.

For the following discussion consider two random variables $Z_1(x)$ and $Z_2(x)$, which are the physical variables collected at two depths. The cokriging coefficient matrix then takes the form

$$\mathbf{G}(h) = \begin{bmatrix} \gamma_1(h) & g_{12}(h) \\ g_{21}(h) & \gamma_2(h) \end{bmatrix} \quad [9]$$

It should be verified that the variograms and pseudo-crossvariograms in the matrix satisfied the Schwarz's inequality given by (Myers, 1982)

$$\gamma_1(h)\gamma_2(h) > g_{12}(h)g_{21}(h) \quad [10]$$

This condition guarantees the cokriging coefficient matrix to be positive definite and guarantees that the variance of the estimated variables is positive.

MATERIALS AND METHODS

To study spatial distributions of several important chemicals, a field experiment was conducted at an irrigation farm in the western part of Fresno County near the town of Mendota, CA, in the fall of 1991. Soil samples were collected in an 800 by 800 m field at depths of 0.15, 0.30, 0.45, 0.60, 0.90, 1.20, 1.50, and 1.80 m. The number of samples varied at different depths, ranging from 114 data points at 0.15 m to 82 data points at 1.80 m as shown in Fig. 1. The field site is a commercial farm that has a varying cropping history. Furrow irrigation using California Aqueduct water and well water was the primary irrigation practice. The soil type on the site is the Panoche silty clay loam (fine-loamy, mixed (calcareous), thermic Typic Torriorthent). Soil samples were collected at the speci-

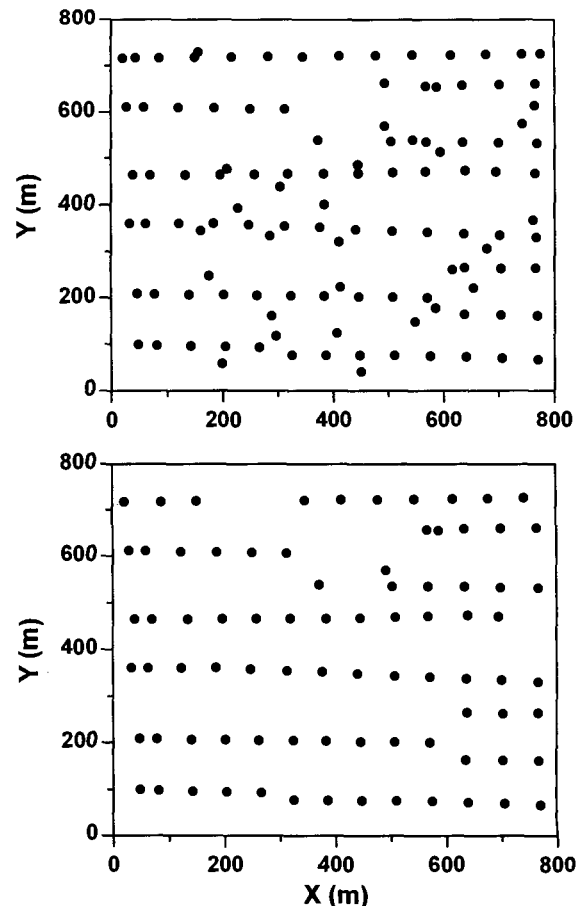


Fig. 1. Horizontal spacing of the sampling scheme at (top) 0.15 and (bottom) 1.80 m.

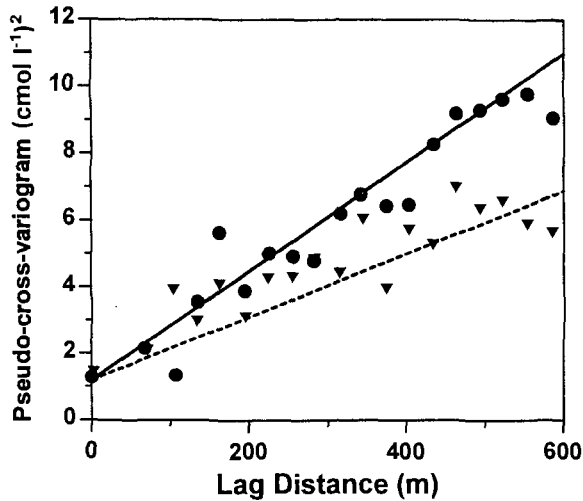


Fig. 3. Pseudo-crossvariograms of Cl⁻, g_{12} and g_{21} , between the depths of 0.90 and 1.20 m. The circles and triangles represent the sample pseudo-crossvariograms, g_{12}^s and g_{21}^s , respectively, whereas the solid and dashed lines represent the fitted pseudo-crossvariograms.

2. Model-specific variograms and pseudo-crossvariograms were chosen based on the sample variograms and sample pseudo-crossvariograms calculated at Step 1 with cross-validation.
3. Cokriging and the data points at 0.60 and 0.90 m were used to estimate Cl⁻ at 0.90 m at the same 112 horizontal locations as 0.60 m.
4. The 112 estimates at 0.90 m were used as “data” with cross-validation to recalculate the variogram for the depth (updated variogram model).
5. This procedure was repeated using the 112 estimates at 0.90 m and 45 data at 1.20 m to obtain a variogram at 1.20 m and pseudo-crossvariograms of the two depths.
6. The same procedure was repeated to obtain variograms and pseudo-crossvariograms for depths 1.50 and 1.80 m.

Note that the equations for computation of pseudo-

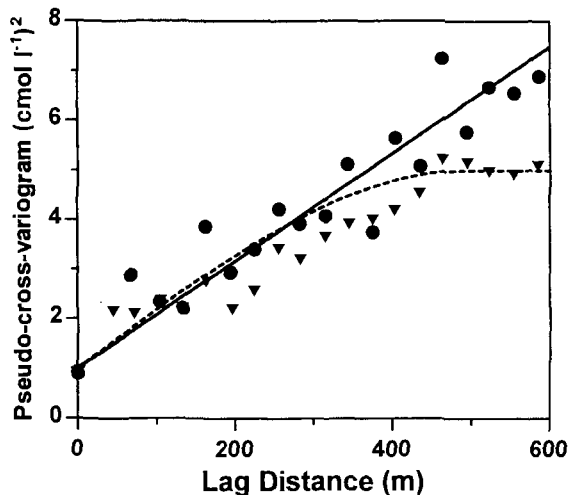


Fig. 4. Pseudo-crossvariograms of Cl⁻, g_{12} and g_{21} , between the depths of 1.20 and 1.50 m. The circles and triangles represent the sample pseudo-crossvariograms, g_{12}^s and g_{21}^s , respectively, whereas the solid and dashed lines represent the fitted pseudo-crossvariograms.

Table 2. Variogram models for Cl⁻ concentrations at different depths.

Depth	Variogram	
m	(cmol L ⁻¹) ²	
0.60	$\gamma(h) = 4 + 0.0183h$	
0.90	$\gamma(h) = 3 + 0.0175h$	$\gamma(h) = 1 + 0.015h$ (updated)
1.20	$\gamma(h) = 2 + 0.0133h$	$\gamma(h) = 1.5 + 0.0167h$ (updated)
1.50	$\gamma(h) = 3 + 0.0108h$	$\gamma(h) = 1.8 + 0.01h$ (updated)
1.80	$\gamma(h) = 1.8 + 0.00667h$	

crossvariograms (Eq. [4] and [5]) originally are defined for two variables (Z_1 and Z_2) collected in the same field. In our study, we used the equations to compute pseudo-crossvariograms for the same variable collected at two depths (Z_1 and Z_2). Comparing the distance between the depths (0.15–0.30 m) with the horizontal scale (800 m), the vertical variability of the solutes should be smaller than the horizontal variability.

Figures 2A and 2B show the sample variograms and models for the depths 0.60 and 1.80 m, respectively. The pseudo-crossvariograms between the depths 0.90 and 1.20 m are presented in Fig. 3. Both g_{12} and g_{21} are linear models; however, the slope of g_{21} is about two times that of g_{12} . As shown in Fig. 4, one of the pseudo-crossvariograms between 1.20 and 1.50 m is a linear model and another is a spherical model. Tables 2 and 3 list the models of the variograms (including the updated ones) and pseudo-crossvariograms, respectively. All variograms are linear models and their slopes decrease with depth from 0.018 to 0.0067. The use of common variogram functions, linear and spherical models, to fit all of the sample pseudo-crossvariograms enabled us to test the positive definite condition of the cokriging coefficient matrix. The variograms and pseudo-crossvariograms in the tables satisfied Schwarz's inequality given in Eq. [10] within the interest domain between Depths 1 and 2. For example, between the depths 0.90 and 1.20 m, we have

$$\begin{aligned} \gamma_1(h)\gamma_2(h) &= (1 + 0.015h)(1.5 + 0.0167h) \\ g_{12}(h)g_{21}(h) &= (1.2 + 0.008h)(1.2 + 0.0163h) \end{aligned}$$

Therefore,

$$\begin{aligned} \gamma_1(h)\gamma_2(h) - g_{12}(h)g_{21}(h) &= \\ 0.06 + 0.01h + 0.00012h^2 &> 0 \end{aligned}$$

Using the variograms and pseudo-crossvariograms in

Table 3. Pseudo-cross-variogram models for Cl⁻ concentrations between depths.

Depths	Pseudo-cross-variogram
m	(cmol L ⁻¹) ²
0.60 and 0.90	$g_{12}(h) = 1 + 0.0142$ $g_{21}(h) = 1 + 8[1.5(h/500) - 0.5(h/500)^3]$ ($h \leq 500$) $g_{21}(h) = 9$ ($h > 500$)
0.90 and 1.20	$g_{12}(h) = 1.2 + 0.008h$ $g_{21}(h) = 1.2 + 0.0163h$
1.20 and 1.50	$g_{12}(h) = 1 + 0.0108h$ $g_{21}(h) = 1 + 4[1.5(h/500) - 0.5(h/500)^3]$ ($h \leq 500$) $g_{21} = 5$ ($h > 500$)
1.50 and 1.80	$g_{12}(h) = 1 + 0.00783h$ $g_{21}(h) = 1 + 0.0108h$

Table 4. Summary statistics of concentration estimation using cokriging and kriging.

Depth 1	Depth 2	Estimates at Depth 2	Cokriging			Kriging			RMSE
			MSE	MKV†	r‡	MSE	MKV	r	
m		no.	(cmol L ⁻¹) ²			(cmol L ⁻¹) ²			%
0.60	0.90	103	1.76	3.37	0.901	3.76	5.17	0.774	53
0.90	1.20	90	1.62	2.22	0.867	2.22	3.61	0.814	27
1.20	1.50	85	1.31	2.46	0.893	2.11	4.66	0.811	38
1.50	1.80	82	0.58	1.14	0.940	1.44	2.71	0.835	60

† Mean kriging variance.

‡ Correlation coefficient between measurements and estimates.

Tables 2 and 3, we estimated Cl⁻ concentration at the sample locations at depths of 0.90, 1.20, 1.50, and 1.80 m using kriging and cokriging techniques. The reduced data set was used for the estimation, i.e., the available 112 data at 0.60 m and the common 45 data points at each depth below 0.60 m. At each of the data points, the concentration was estimated by using a neighborhood around it, but not itself. The results are summarized in Table 4. Compared with kriging, cokriging reduced the MSE between 30 and 60%, reduced the mean kriging variances between 35 and 58%, and significantly increased the correlation coefficients between the measurements and estimates.

To estimate solute distributions and total mass in the 800 by 800 by 1.8 m soil volume, kriging and cokriging were used to calculate 441 values of the solute concentration on a grid of 40 by 40 m at each soil layer. For soil layers above 0.90 m, all data and kriging were utilized for the concentration estimation. Three methods were applied to compute solute distributions below 0.60 m: (A) kriging with all data at each layer; (B) kriging using the 45 data points at each layer; and (C) cokriging with the same 45 data points at each layer and the 112 estimates at the upper layer. In terms of sampling cost, the cokriging with pseudo-crossvariograms (Method C) used less than half the data of Method A. In terms of estimation accuracy, cokriging produced better results than kriging using all of the data. For the concentration estimates at 0.90 m, kriging variance values by kriging using all data were higher than those by cokriging with less than half the data at 0.90 m and incorporating the data at 0.60 m.

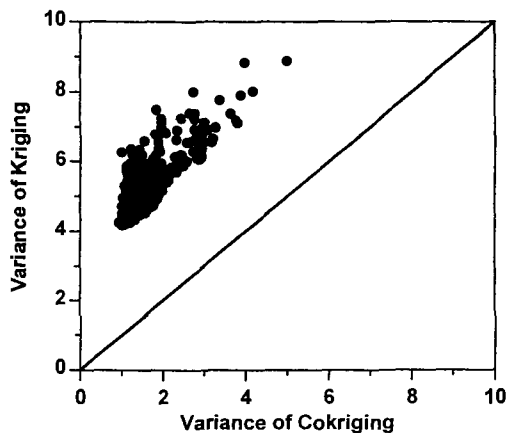


Fig. 5. Comparison of kriging variances [(cmol L⁻¹)²] for Cl⁻ concentration estimation at 0.90 m between kriging using 45 data points at 0.90 m and cokriging using the same data at 0.90 m and other information at 0.60 m.

Cokriging produced much better results than kriging using the same 45 data points at 0.90 m (Fig. 5). The kriging variance values are more than two times the cokriging variance at the estimated points. Figures 6A,

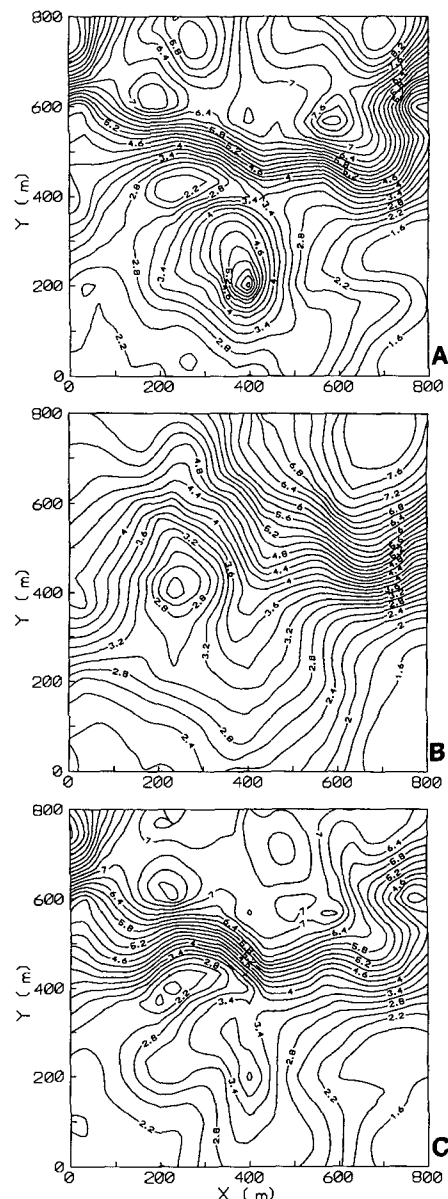


Fig. 6. Chloride concentration (cmol L⁻¹) contour maps at 0.90 m generated by (A) kriging with 103 data at 0.90 m; (B) kriging using 45 data points at 0.90 m; (C) cokriging with the pseudo-crossvariograms using the same 45 data points at 0.90 m as well as 112 data points at 0.60 m.

6B, and 6C show the concentration contour maps generated by Methods A, B, and C, respectively.

The total mass in the soil volume was calculated based on the concentration distributions using a depth average and the spatial distributions of the gravimetric water content and bulk density. Since cokriging with the nonsymmetric pseudo-crossvariograms used less than half of the data, this approach could reduce the sampling costs considerably compared with kriging. In addition, cokriging reduced the estimation error about 18%. If the reduced data set was used, cokriging reduced the estimation error up to 40% compared with kriging.

SUMMARY

It is critical to estimate solute distributions and total mass in soils for many agricultural and environmental problems. A large-scale field experiment was conducted to measure Cl^- , SO_4^{2-} , Ca^{2+} , Mg^{2+} , Na^+ , P_2O_5 , K^+ , and NO_3^- at depths of 0.15, 0.30, 0.45, 0.60, 0.90, 1.20, 1.50, and 1.80 m in an area of 800 by 800 m. Assuming only 45 data points to be available at each layer below 0.60 m, we computed pseudo-crossvariograms between depths and carried out cokriging for studying sampling and estimation strategies. Sample pseudo-crossvariograms, $g_{12}^*(h)$ and $g_{21}^*(h)$, were nonsymmetrical for all depths and were modeled with the common variograms, such as spherical and linear. The variograms and the pseudo-crossvariograms used in the cokriging coefficient matrix were shown to satisfy Schwarz's inequality. This condition guaranteed the matrix to be positive definite, which was a guarantee that the variance of the estimated variables was positive.

Cokriging with nonsymmetric pseudo-crossvariograms was successfully used to calculate solute distributions and the total mass of Cl^- in the soil. Kriging and cokriging were utilized to estimate the solute concentration at the measured locations based on less than half of the available data. It was shown that cokriging can be used to increase estimation accuracy and reduce sampling requirements. Through cokriging, much more easily sampled information at shallower depths was used to improve estimations at deeper soil layers. Compared

with kriging, cokriging reduced the mean squared errors between 30 and 60%, reduced the mean kriging variances between 35 and 58%, and significantly increased the correlation coefficients between the measurements and estimates. Cokriging with the nonsymmetric pseudo-crossvariograms used less than half data for the estimation of chemical distributions, thus it potentially could reduce more than half the sampling cost compared with kriging estimation. In addition, compared with kriging using all data, cokriging reduced the estimation error about 20%. If the same information was used for kriging and cokriging estimations, cokriging reduced the estimation error 40% more than kriging.

REFERENCES

- 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.
- 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.
- 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.
- 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. 1992b. Improvement of soil textural estimates using spectral properties. *Geoderma* 52:223–234.
- Zhang, R., S. Rahman, G.F. Vance, and L.C. Munn. 1995. Geostatistical analyses of trace elements in soils and plants. *Soil Sci.* 159:383–390.
- Zhang, R., D.E. Myers and A.W. Warrick. 1992a. Estimation of the spatial distribution for soil chemicals using pseudo-cross-variograms. *Soil Sci. Soc. Am. J.* 56:1444–1452.