Mathematical Geology, Vol. 30, No. 7, 1998

The Variance-Based Cross-Variogram: You Can Add Apples and Oranges[1]

Noel Cressie[2] and Christopher K. Wikle[3]

The variance-based cross-variogram between two spatial processes, $Z_1(\cdot)$ and $Z_2(\cdot)$, is var $(Z_1(\mathbf{u}) - Z_2(\mathbf{v}))$, expressed generally as a bivariate function of spatial locations $\mathbf{u}$ *and* $\mathbf{v}$*. It characterizes the cross-spatial dependence between $Z_1(\cdot)$ and $Z_2(\cdot)$ and can be used to obtain optimal multivariable predictors (cokriging). It has also been called the pseudo cross-variogram; here we compare its properties to that of the traditional (covariance-based) cross-variogram, cov $(Z_1(\mathbf{u}) - Z_1(\mathbf{v}), Z_2(\mathbf{u}) - Z_2(\mathbf{v}))$. One concern with the variance-based cross-variogram has been that $Z_1(\cdot)$ and $Z_2(\cdot)$ might be measured in different units ("apples" and "oranges"). In this note, we show that the cokriging predictor based on variance-based cross-variograms can handle any units used for $Z_1(\cdot)$ and $Z_2(\cdot)$; recommendations are given for an appropriate choice of units. We review the differences between the variance-based cross-variogram and the covariance-based cross-variogram and conclude that the former is more appropriate for cokriging. In practice, one often assumes that variograms and cross-variograms are functions of* $\mathbf{u}$ *and* $\mathbf{v}$ *only through the difference* $\mathbf{u} - \mathbf{v}$*. This restricts the types of models that might be fitted to measures of cross-spatial dependence.*

KEY WORDS: cokriging, equivariance, pseudo cross-variogram.

INTRODUCTION

There has been considerable discussion recently on the merits of two types of cross-variogram for cokriging (Cressie, 1991, p. 140, 141; Myers, 1991; Papritz, Kunsch, and Webster, 1993; Ver Hoef and Cressie, 1993; Wackernagel, 1995; Ver Hoef and Barry, 1998). Both types measure cross-dependence between two spatial variables and both yield the variogram as a special case when the two variables are the same (see Eqs. 6 and 7 below).

Let

$$\{(Z_1(\mathbf{s}), Z_2(\mathbf{s})): \mathbf{s} \in D\} \tag{1}$$

[1]Received 20 August 1997; accepted 5 January 1998.

[2]Department of Statistics, Iowa State University, Ames, Iowa 50011-1210. e-mail: ncressie@iastate.edu

[3]National Center for Atmospheric Research, P.O. Box 3000, Boulder, Colorado 80307-3000. e-mail: wikle@ucar.edu

0882-8121/98/1000-0789$15.00/1

be a bivariate spatial process defined over a domain of interest $D \subset \mathbb{R}^d$, the d-dimensional Euclidean space. We assume henceforth that $\text{var}(Z_1(\mathbf{s})) < \infty$ and $\text{var}(Z_2(\mathbf{s})) < \infty$, for all $\mathbf{s} \in D$, a regularity condition that guarantees the existence of all mean-squared prediction errors. We also assume here, for ease of presentation, that $E(Z_1(\mathbf{s})) = \mu_1$ and $E(Z_2(\mathbf{s})) = \mu_2$, for all $\mathbf{s} \in D$.

Kriging of any one variable, say $Z_1(\cdot)$, exploits the spatial dependence within $Z_1(\cdot)$ to obtain an optimal linear predictor of $Z_1(\mathbf{s}_0)$ based on data,

$$\mathbf{Z}_1 \equiv (Z_1(\mathbf{u}_1), \ldots, Z_1(\mathbf{u}_{n_1}))' \tag{2}$$

Cokriging of the variable $Z_1(\cdot)$, based on the data $\mathbf{Z}_1$ and

$$\mathbf{Z}_2 \equiv (Z_2(\mathbf{v}_1), \ldots, Z_2(\mathbf{v}_{n_2}))' \tag{3}$$

exploits the spatial dependencies within $Z_1(\cdot)$ and $Z_2(\cdot)$, *as well as* the cross-spatial dependencies *between* $Z_1(\cdot)$ and $Z_2(\cdot)$, to yield an optimal linear predictor,

$$p_1(\mathbf{Z}; \mathbf{s}_0) \equiv \boldsymbol{\lambda}_1' \mathbf{Z}_1 + \boldsymbol{\lambda}_2' \mathbf{Z}_2 \tag{4}$$

of $Z_1(\mathbf{s}_0)$, where $\mathbf{Z}' \equiv (\mathbf{Z}_1', \mathbf{Z}_2')$. Notice that the observation locations in (2) and (3) do not have to be the same. Clearly, because the cokriging predictor of $Z_1(\mathbf{s}_0)$ uses more information (namely observations on $Z_2(\cdot)$), it will have smaller mean-squared prediction error than the kriging predictor of $Z_1(\mathbf{s}_0)$. Equally, one can obtain an analogous optimal linear predictor, $p_2(\mathbf{Z}; \mathbf{s}_0)$, of $Z_2(\mathbf{s}_0)$ based on $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

Kriging equations that yield the optimal coefficients of $\mathbf{Z}_1$ in the kriging predictor of $Z_1(\mathbf{s}_0)$, can be written either in terms of spatial covariance functions or in terms of spatial variograms (e.g., Journel and Huijbregts, 1978, p. 306; Cressie, 1991, p. 142, 143). It is easy to show (e.g., Journel and Huijbregts, 1978, p. 325) that cokriging equations to obtain optimal coefficients $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ in the cokriging predictor (4) can be obtained in terms of the spatial covariance functions,

$$C_{ij}(\mathbf{u}, \mathbf{v}) \equiv \text{cov}\,(Z_i(\mathbf{u}), Z_j(\mathbf{v})); \qquad i, j = 1, 2 \tag{5}$$

where $\mathbf{u}, \mathbf{v} \in D$. However, what has proved controversial is the analogous result in terms of spatial cross-variograms. In particular, the controversy has been around the *choice* of cross-variogram that appropriately generalizes the variogram. Two types of cross-variogram have been proposed and they have been discussed extensively in the literature cited above. In historical order of appearance, they are:

$$2\nu_{ij}(\mathbf{u})\,\mathbf{v}) \equiv \text{cov}\,(Z_i(\mathbf{u}) - Z_i(\mathbf{v}), Z_j(\mathbf{u}) - Z_j(\mathbf{v})); \qquad i, j = 1, 2 \tag{6}$$

$$2\gamma_{ij}(\mathbf{u}, \mathbf{v}) \equiv \text{var}\,(Z_i(\mathbf{u}) - Z_j(\mathbf{v})); \qquad i, j = 1, 2 \tag{7}$$

where $\mathbf{u}, \mathbf{v} \in D$. That is, the quantity defined in (6) is a covariance of direct increments and the quantity defined in (7) is a variance of cross-increments. Notice that we have chosen to write (5), (6), and (7) in their most general forms, as functions of $\mathbf{u}$ and $\mathbf{v}$, and not in their stationary forms, as functions of $\mathbf{u} - \mathbf{v}$. This is because there is nothing in kriging and cokriging methodology that *requires* stationarity. A nonstationary covariance or variogram model may result from a spatial partial differential equation with a few unknown parameters. Estimation of the nonstationary covariance function is easily obtained by plugging in estimates of those few parameters (e.g., ML or REML estimates such as described in Cressie, 1991, p. 91–93). Then (co)kriging based on the nonstationary, "plug-in" covariance functions proceeds as described in the next section.

Myers (1982) and Wackernagel (1988) have been early proponents of ν_{ij} given by (6). Clark, Basinger, and Harper (1989) proposed a version of (7) without mean correction and Cressie (1991, p. 140, 141) showed how cokriging equations for λ_1 and λ_2 in (4) could be written in terms of γ_{ij} given by (7). Myers (1991) gave the same result and obtained algebraic relationships between (5), (6), and (7); he coined the term "pseudo cross-variogram" for (7).

It has long been known (e.g., Journel and Huijbregts, 1978, p. 326; Myers, 1982) that cokriging in terms of ν_{ij} in (6) requires a rather strong symmetry condition: $C_{12}(\mathbf{u}, \mathbf{v}) = C_{12}(\mathbf{v}, \mathbf{u})$. However, as is illustrated in the concluding section, it is not hard to construct models that do *not* satisfy this condition (see also Ver Hoef and Cressie, 1993; Ver Hoef and Barry, 1998). Through matrix algebra and a small example involving asymmetric C_{12}, Ver Hoef and Cressie (1993) show that use of the ν_{ij} (given by Eq. 6) in cokriging equations yields nonoptimal linear (in $\mathbf{Z}_1$ and $\mathbf{Z}_2$) predictors, sometimes badly so. In contrast, the use of γ_{ij} given by (7) always yields the optimal cokriging predictor. This is the same optimal cokriging predictor one would obtain by using cross-covariances C_{ij} given by (5) (see Ver Hoef and Cressie, 1993).

For reasons given above, the adjective "pseudo," for the cross-variograms $2\gamma_{ij}$, is a misnomer. In what is to follow, when we want to make a distinction, we shall refer to covariance-based cross-variograms $2\nu_{ij}$, and to variance-based cross-variograms $2\gamma_{ij}$, but, when no distinction is necessary, $2\gamma_{ij}$ will henceforth be referred to simply as the cross-variograms. Finally, we do not claim that the $2\gamma_{ij}$ given by (7) are unique in the sense that they are the only cross-variograms that will yield cokriging predictors (see Kunsch, Papritz, and Bassi, 1997, where this nonuniqueness is apparent). We shall discuss this further in the concluding section.

A practical problem in the use of γ_{ij} for cokriging has been noted. If $Z_1(\cdot)$ and $Z_2(\cdot)$ are measured in different units, what can be made of $\text{var}(Z_1(\mathbf{u}) - Z_2(\mathbf{v}))$? Can we add (or, here, subtract) "apples" and "oranges?" In practice, $Z_1(\cdot)$ and $Z_2(\cdot)$ are typically standardized in the usual way. Consider the stan-

dardized datasets $\{(Z_1(\mathbf{u}_i) - \bar{Z}_1)/S_1: i = 1, \ldots, n_1\}$, where $\bar{Z}_1 \equiv \Sigma Z_1(\mathbf{u}_i)/n_i$ and $S_1^2 \equiv \Sigma(Z_1(\mathbf{u}_i) - \bar{Z}_1)^2/(n_1 - 1)$, and $\{(Z_2(\mathbf{v}_i) - \bar{Z}_2)/S_2: i = 1, \ldots, n_2\}$, where $\bar{Z}_2$ and S_2^2 are likewise defined. Then the standardized variables are unitless, can be subtracted, and incidentally have zero mean (important for estimation of γ_{ij}) under symmetry assumptions.

No justification of this common-sense, but *ad hoc* standardization has ever been given, and that is the purpose of this note. The next section shows that the cokriging predictor obtained using cross-variograms $2\gamma_{ij}$ is equivariant to mean-parameter and scale-parameter changes in the variables $Z_1(\cdot)$ and $Z_2(\cdot)$. Although the results in the next section is rather technical, its consequences for cokriging are important. It says that no matter what units $Z_1(\cdot)$ and $Z_2(\cdot)$ are defined in, the cokriging predictor $p_1(\mathbf{Z}; \mathbf{s}_0)$, based on cross-variograms $2\gamma_{ij}$, is always in the same units as $Z_1(\cdot)$. These and other related issues are discussed in the concluding section.

EQUIVARIANCE OF COKRIGING TO MEAN AND SCALE CHANGE

Equivariance is best explained through an example. Consider a scientific study of sea-surface temperature (SST) over the tropical Pacific. A statistical methodology is equivariant if the resulting scientific inferences are identical, regardless of the units that the SST data come in (e.g., °F, °C, °K). For example, if $Z(\cdot)$ is SST in °F and $Y(\cdot)$ is SST in °C, then

$$Y(\mathbf{s}) = (Z(\mathbf{s}) - 32)/(9/5); \qquad \mathbf{s} \in D \tag{8}$$

Let $p^{(Z)}(\mathbf{Z}; \mathbf{s}_0)$ denote a predictor of $Z(\mathbf{s}_0)$ based on data $\mathbf{Z}$ and an assumed statistical model for the Z-process. Further, let $p^{(Y)}(\mathbf{Y}; \mathbf{s}_0)$ denote a predictor of $Y(\mathbf{s}_0)$ depending on data $\mathbf{Y}$ and the Y-process model, in the same way that $p^{(Z)}(\mathbf{Z}; \mathbf{s}_0)$ depends on $\mathbf{Z}$ and the Z-process model. Then the statistical methodology, that specifies how the data and the process model define the predictor, is *equivariant* to whether the study is conducted in °F or °C if

$$p^{(Y)}(\mathbf{Y}; \mathbf{s}_0) = (p^{(Z)}(\mathbf{Z}; \mathbf{s}_0) - 32)/(9/5) \tag{9}$$

or, equally, if

$$p^{(Z)}(\mathbf{Z}; \mathbf{s}_0) = (9/5)p^{(Y)}(\mathbf{Y}; \mathbf{s}_0) + 32 \tag{10}$$

More generally, if

$$Y(\mathbf{s}) = (Z(\mathbf{s}) - \eta)/\tau; \qquad \eta \in \mathbb{R}, \tau > 0 \tag{11}$$

then the statistical methodology that yields the predictor is mean and scale *equivariant* if

$$p^{(Z)}(\mathbf{Z};\, \mathbf{s}_0) = \tau p^{(Y)}(\mathbf{Y};\, \mathbf{s}_0) + \eta \tag{12}$$

for all mean parameters $\eta \in \mathbb{R}$ and all scale parameters $\tau > 0$. Clearly, equivariance is desirable in any scientific study since then the units of analysis do not affect the substantive conclusions.

In the context of cokriging, equivariance can be formulated as follows. Define

$$Y_i(\mathbf{s}) \equiv (Z_i(\mathbf{s}) - \eta_i)/\tau_i; \qquad i = 1, 2 \tag{13}$$

For example, consider the case of SST and sea-level pressure (SLP). In this case, $Y_1(\cdot)$ could refer to SST in °C and $Z_1(\cdot)$ to SST in °F (as in Eq. 8). Furthermore, $Y_2(\cdot)$ could refer to SLP in Pascals and $Z_2(\cdot)$ to SLP in inches of mercury. As is often the case in multivariate analysis, we are analyzing data and cokriging with variables that measure quite different properties ("apples" and "oranges," if you will).

Recall from (2) and (3) that data $\mathbf{Z}' = (\mathbf{Z}_1', \mathbf{Z}_2')$ are observed. These data are transformed to $\mathbf{Y}' = (\mathbf{Y}_1', \mathbf{Y}_2')$ according to (13). Let $p_1^{(Z)}(\mathbf{Z};\, \mathbf{s}_0)$ and $p_1^{(Y)}(\mathbf{Y};\, \mathbf{s}_0)$ denote the cokriging predictors of $Z_1(\mathbf{s}_0)$ and $Y_1(\mathbf{s}_0)$, given in terms of cross-variograms $2\gamma_{ij}^{(Z)}$ and $2\gamma_{ij}^{(Y)}$, respectively, where,

$$2\gamma_{ij}^{(Z)}(\mathbf{u}, \mathbf{v}) \equiv \text{var}\,(Z_i(\mathbf{u}) - Z_j(\mathbf{v}))$$

$$2\gamma_{ij}^{(Y)}(\mathbf{u}, \mathbf{v}) \equiv \text{var}\,(Y_i(\mathbf{u}) - Y_j(\mathbf{v})); \qquad \mathbf{u}, \mathbf{v} \in D$$

Notice that

$$2\gamma_{ij}^{(Z)}(\mathbf{u}, \mathbf{v}) = \text{var}\,(Z_i(\mathbf{u})) + \text{var}\,(Z_j(\mathbf{v})) - 2\,\text{cov}\,(Z_i(\mathbf{u}), Z_j(\mathbf{v})) \tag{14}$$

and likewise for $\gamma_{ij}^{(Y)}(\mathbf{u}, \mathbf{v})$. Then cokriging in terms of variance-based cross-variograms is equivariant if

$$p_1^{(Z)}(\mathbf{Z};\, \mathbf{s}_0) = \tau_1 p_1^{(Y)}(\mathbf{Y};\, \mathbf{s}_0) + \eta_1 \tag{15}$$

where

$$p_1^{(Y)}(\mathbf{Y};\, \mathbf{s}_0) = (\boldsymbol{\lambda}_1^{(Y)})'\mathbf{Y}_1 + (\boldsymbol{\lambda}_2^{(Y)})'\mathbf{Y}_2 \equiv (\boldsymbol{\lambda}^{(Y)})'\mathbf{Y} \tag{16}$$

$$(\boldsymbol{\lambda}_1^{(Y)})'\mathbf{1} = 1,\ (\boldsymbol{\lambda}_2^{(Y)})'\mathbf{1} = 0 \tag{17}$$

are unbiasedness conditions, and $\boldsymbol{\lambda}_1^{(Y)}$, $\boldsymbol{\lambda}_2^{(Y)}$ are given by cokriging equations in terms of $\gamma_{ij}^{(Y)}$ (see Eq. 19). The quantity $p_1^{(Z)}$ in (15) is defined in exactly the same way as in (16) and (17), but in terms of $\mathbf{Z}' \equiv (\mathbf{Z}_1', \mathbf{Z}_2')$ and $\gamma_{ij}^{(Z)}$.

The cokriging predictor is optimal in that it minimizes the mean squared prediction error among all linear unbiased predictors. Cressie (1991, p. 141) shows that the linear predictor (16), subject to unbiasedness conditions (17),

has a mean squared prediction error that can be written in terms of variance-based cross-variograms $2\gamma_{ij}^{(Y)}$:

$$\begin{aligned} & E(Y_1(\mathbf{s}_0) - p_1(\mathbf{Y}; \mathbf{s}_0))^2 \\ &= -\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\lambda_{1i}\lambda_{ij}\gamma_{11}^{(Y)}(\mathbf{u}_i, \mathbf{u}_j) + 2\sum_{i=1}^{n_1}\lambda_{1i}\gamma_{11}^{(Y)}(\mathbf{s}_0, \mathbf{u}_i) \\ &\quad + 2\sum_{k=1}^{n_2}\lambda_{2k}\gamma_{12}^{(Y)}(\mathbf{s}_0, \mathbf{v}_k) - 2\sum_{i=1}^{n_1}\sum_{k=1}^{n_2}\lambda_{1i}\lambda_{2k}\gamma_{12}^{(Y)}(\mathbf{u}_i, \mathbf{v}_k) \\ &\quad - \sum_{k=1}^{n_2}\sum_{l=1}^{n_2}\lambda_{2k}\lambda_{2l}\gamma_{22}^{(Y)}(\mathbf{v}_k, \mathbf{v}_l) \end{aligned} \tag{18}$$

Upon minimizing (18) with respect to $\boldsymbol{\lambda}' \equiv (\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2')$, one obtains cokriging equations, which are given in matrix form by Myers (1991). Ver Hoef and Cressie (1993, p. 225) solve these equations, giving an explicit formula for the cokriging predictor as,

$$\begin{aligned} p_1^{(Y)}(\mathbf{Y}; \mathbf{s}_0) &= (\boldsymbol{\lambda}^{(Y)})'\mathbf{Y} \\ &= (\boldsymbol{\gamma}_1^{(Y)})'(\boldsymbol{\Gamma}^{(Y)})^{-1}\mathbf{Y} \\ &\quad + \{\mathbf{w} - \mathbf{X}'(\boldsymbol{\Gamma}^{(Y)})^{-1}\boldsymbol{\gamma}_1^{(Y)}\}'(\mathbf{X}'(\boldsymbol{\Gamma}^{(Y)})^{-1}\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\Gamma}^{(Y)})^{-1}\mathbf{Y} \end{aligned} \tag{19}$$

where $\boldsymbol{\Gamma}^{(Y)}$, assumed to be invertible, is given by

$$\boldsymbol{\Gamma}^{(Y)} = \begin{bmatrix} \boldsymbol{\Gamma}_{11}^{(Y)} & \boldsymbol{\Gamma}_{12}^{(Y)} \\ \boldsymbol{\Gamma}_{21}^{(Y)} & \boldsymbol{\Gamma}_{22}^{(Y)} \end{bmatrix}_{(n_1+n_2)\times(n_1+n_2)}$$

$$\boldsymbol{\Gamma}_{12}^{(Y)} = (\gamma_{12}^{(Y)}(\mathbf{u}_i, \mathbf{v}_j))_{n_1\times n_2}$$

and $\boldsymbol{\Gamma}_{11}^{(Y)}$, $\boldsymbol{\Gamma}_{22}^{(Y)}$, $\boldsymbol{\Gamma}_{21}^{(Y)}$ are likewise defined. Also,

$$\boldsymbol{\gamma}_1^{(Y)} = \begin{bmatrix} \boldsymbol{\gamma}_{11}^{(Y)} \\ \boldsymbol{\gamma}_{21}^{(Y)} \end{bmatrix}_{(n_1+n_2)\times 1}$$

$$\boldsymbol{\gamma}_{11}^{(Y)} = (\gamma_{11}^{(Y)}(\mathbf{u}_i, \mathbf{s}_0))_{n_1\times 1}, \qquad \boldsymbol{\gamma}_{21}^{(Y)} = (\gamma_{21}^{(Y)}(\mathbf{v}_i, \mathbf{s}_0))_{n_2\times 1}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \end{bmatrix}_{(n_1+n_2)\times 2}, \qquad \mathbf{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{2\times 1}$$

and $\mathbf{1}_p$ is the p-dimensional column vector of 1s ($p = 1, 2, \ldots$). Myers (1992) discusses the consequences of $\boldsymbol{\Gamma}^{(Y)}$ not being invertible.

Now, from Equation (14),

$$2\gamma_{12}^{(Y)}(\mathbf{u}, \mathbf{v}) = (\tau_1\tau_2)^{-1}2\gamma_{12}^{(Z)}(\mathbf{u}, \mathbf{v}) + (\tau_1^{-1} - \tau_2^{-1}) \cdot \{\tau_1^{-1} \text{ var } (Z_1(\mathbf{u})) - \tau_2^{-1} \text{ var } (Z_2(\mathbf{v}))\} \tag{20}$$

Substituting (20) into (18) and using the unbiasedness conditions (17), we find that the variance terms in (20) cancel and the scales τ_1, τ_2 enter multiplicatively. That is,

$$\begin{aligned} &E(Y_1(\mathbf{s}_0) - p_1(\mathbf{Y}; \mathbf{s}_0))^2 \\ &= -\tau_1^{-2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \lambda_{1i}\lambda_{1j}\gamma_{11}^{(Z)}(\mathbf{u}_i, \mathbf{u}_j) + 2\tau_1^{-2} \sum_{i=1}^{n_1} \lambda_{1i}\gamma_{11}^{(Z)}(\mathbf{s}_0, \mathbf{u}_i) \\ &\quad + 2(\tau_1\tau_2)^{-1} \sum_{k=1}^{n_2} \lambda_{2k}\gamma_{12}^{(Z)}(\mathbf{s}_0, \mathbf{v}_k) - 2(\tau_1\tau_2)^{-1} \sum_{i=1}^{n_1} \sum_{k=1}^{n_2} \\ &\quad \lambda_{1i}\lambda_{2k}\gamma_{12}^{(Z)}(\mathbf{u}_i, \mathbf{v}_k) - \tau_2^{-2} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} \lambda_{2k}\lambda_{2l}\gamma_{22}^{(Z)}(\mathbf{v}_k, \mathbf{v}_l) \end{aligned} \tag{21}$$

Define the matrices

$$\mathbf{G} = \begin{bmatrix} \tau_1^{-1}\mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \tau_2^{-1}\mathbf{I}_{n_2} \end{bmatrix}_{(n_1+n_2)\times(n_1+n_2)}$$

$$\mathbf{H} = \begin{bmatrix} \tau_1^{-2}\mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & (\tau_1\tau_2)^{-1}\mathbf{I}_{n_2} \end{bmatrix}_{(n_1+n_2)\times(n_1+n_2)}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix; $p = 1, 2, \ldots$. Now create cross-variogram matrices

$$\boldsymbol{\Gamma}^* = \mathbf{G}\boldsymbol{\Gamma}^{(Z)}\mathbf{G}$$

$$\boldsymbol{\gamma}_1^* = \mathbf{H}\boldsymbol{\gamma}_1^{(Z)}$$

and notice from (19) and (21) that,

$$(\boldsymbol{\lambda}^{(Y)})' = (\boldsymbol{\gamma}_1^*)'(\boldsymbol{\Gamma}^*)^{-1} + \{\mathbf{w} - \mathbf{X}'(\boldsymbol{\Gamma}^*)^{-1}\boldsymbol{\gamma}_1^*\}' \cdot (\mathbf{X}'(\boldsymbol{\Gamma}^*)^{-1}\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\Gamma}^*)^{-1} \tag{22}$$

Further, $\mathbf{Y}$ and $\mathbf{Z}$ are connected by

$$\mathbf{Y} = \mathbf{G}(\mathbf{Z} - X\boldsymbol{\eta}) \tag{23}$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2)'$. Thus,

$$(\boldsymbol{\lambda}^{(Y)})'\mathbf{G}\mathbf{X}\boldsymbol{\eta} = (\boldsymbol{\lambda}^{(Y)})'\mathbf{G}\mathbf{Z} - (\boldsymbol{\lambda}^{(Y)})'\mathbf{Y}$$

which, after considering the unbiasedness constraints (17), becomes

$$\tau_1(\boldsymbol{\lambda}^{(Y)})'\mathbf{Y} + \eta_1 = \tau_1(\boldsymbol{\lambda}^{(Y)})'\mathbf{G}\mathbf{Z} \tag{24}$$

Substituting (22) into the right-hand side of (24) and noting that

$$\mathbf{X}'\mathbf{G}^{-1} = \begin{bmatrix} \tau_1\mathbf{1}' & \mathbf{0}' \\ \mathbf{0}' & \tau_2\mathbf{1}' \end{bmatrix}_{2\times(n_1+n_2)}$$

and $\tau_1\mathbf{H}\mathbf{G}^{-1} = \mathbf{I}_{n_1+n_2}$, we obtain

$$\begin{aligned} &\tau_1(\boldsymbol{\lambda}^{(Y)})'\mathbf{Y} + \eta_1 \\ &\quad = (\boldsymbol{\gamma}_1^{(Z)})'(\boldsymbol{\Gamma}^{(Z)-1}\mathbf{Z} + \{\mathbf{w} - \mathbf{X}'(\boldsymbol{\Gamma}^{(Z)})^{-1}\boldsymbol{\gamma}_1^{(Z)}\}' \\ &\qquad \cdot (\mathbf{X}'(\boldsymbol{\Gamma}^{(Z)})^{-1}\mathbf{X})^{-1}\mathbf{X}'(\boldsymbol{\Gamma}^{(Z)})^{-1}\mathbf{Z} \\ &\quad = (\boldsymbol{\lambda}^{(Z)})'\mathbf{Z} \end{aligned}$$

The latter equality is obtained by putting $\eta_1 = \eta_2 = 0$ and $\tau_1 = \tau_2 = 1$ in (19). That is, from (15), cokriging using variograms $2\gamma_{ij}$ is mean and scale equivariant.

DISCUSSION AND CONCLUSIONS

The implications of the equivariance result for cokriging, proved in the previous section, are worth emphasizing. The user can choose linear transformations $Y_1(\cdot)$, $Y_2(\cdot)$ in *any* units (including those of the original $Z_1(\cdot)$, $Z_2(\cdot)$) and, provided $\gamma_{ij}^{(Y)}$ is known, the optimal cokriging predictor $p_1^{(Y)}(\mathbf{Y}; \mathbf{s}_0)$ is equivariant.

In practice, $\gamma_{ij}^{(Y)}$ is seldom known and has to be estimated from the data $\mathbf{Y}$. If units are chosen for $Y_1(\cdot)$ and $Y_2(\cdot)$ so that the variability of $Y_1(\cdot)$ (as measured by the sample standard deviation $S_1^{(Y)}$, say) is very different from the variability of Y_2 (as measured by $S_2^{(Y)}$), then the cross-dependence expressed in $\gamma_{ij}^{(Y)}$ runs the risk of being swamped by estimation error. For example, if var $(Y_i(\mathbf{s})) = (\sigma_i^{(Y)})^2$; $i = 1, 2$, and $\sigma_1^{(Y)} >> \sigma_2^{(Y)}$, then from (14),

$$2\gamma_{12}^{(Y)}(\mathbf{s}, \mathbf{s} + \mathbf{h}) \simeq (\sigma_1^{(Y)})^2$$

and any method-of-moments estimator $2\hat{\gamma}_{12}^{(Y)}$ of $2\gamma_{12}^{(Y)}$ would behave as,

$$2\hat{\gamma}_{12}^{(Y)}(\mathbf{s}, \mathbf{s} + \mathbf{h}) \simeq (S_1^{(Y)})^2$$

regardless of the strength of cross-spatial dependence. By contrast, if $\sigma_1^{(Y)} = \sigma_2^{(Y)} = \sigma^{(Y)}$, then

$$2\gamma_{12}^{(Y)}(\mathbf{s}, \mathbf{s} + \mathbf{h}) = 2(\sigma^{(Y)})^2(1 - \text{corr}\,(Y_1(\mathbf{s}), Y_2(\mathbf{s} + \mathbf{h})))$$

so that the form of the cross-dependence is governed directly by the correlation function (corr) as $\mathbf{s}$ and $\mathbf{h}$ vary.

Consequently, we recommend that units for $Y_1(\cdot)$ and $Y_2(\cdot)$ be chosen so that they have comparable variability. The easiest way to do this empirically, is to define

$$Y_i(\mathbf{s}) \equiv (Z_i(\mathbf{s}) - \bar{Z}_i)/S_i^{(Z)}; \qquad i = 1, 2, \mathbf{s} \in D \tag{25}$$

Not only can we now claim that var $(Y_1(\mathbf{s})) \simeq$ var $(Y_2(\mathbf{s})) \simeq 1$, but also that $E(Y_1(\mathbf{s})) \simeq E(Y_2(\mathbf{s})) \simeq 0$. Thus, $2\gamma_{ij}^{(Y)}(\mathbf{u}, \mathbf{v}) =$ var $(Y_i(\mathbf{u}) - Y_j(\mathbf{v})) \simeq E(Y_i(\mathbf{u}) - Y_j(\mathbf{v}))^2$, for which a method-of-moments estimator is natural (replace $E(\cdot)$ with ave $(\cdot)$).

It is obvious that the technical result established in the previous section generalizes to more than two variables. That is, the cokriging predictor of $Z_1(\mathbf{s}_0)$ is equivariant under mean and scale changes to $Z_1(\cdot)$, $Z_2(\cdot)$, . . . , $Z_k(\cdot)$, for any integer $k \geq 2$.

Although it is not relevant to the central result and conclusions of this paper, a few comments on stationarity assumptions can be made. To obtain good estimates of $\gamma_{ij}^{(Y)}$, a stationarity assumption like, $2\gamma_{ij}^{(Y)}(\mathbf{u}, \mathbf{v})$ depends only on $\mathbf{u} - \mathbf{v}$, is often made. For example, Ver Hoef and Cressie (1993), Cook and others (1994), and Majure and Cressie (1997) consider,

$$Y_1(\mathbf{s}) = W(\mathbf{s}) + \epsilon_1(\mathbf{s})$$

$$Y_2(\mathbf{s}) = W(\mathbf{s} + \Delta) + \epsilon_2(\mathbf{s}); \qquad \Delta \neq \mathbf{0}$$

where W is a second-order stationary process with covariance $C_W(\mathbf{h})$; ϵ_1 and ϵ_2 are zero-mean white noise processes with variances σ_1^2 and σ_2^2, respectively; and W, ϵ_1 and ϵ_2 are mutually independent. Notice the asymmetry caused by Δ in this shifted spatial factor model. Straightforward calculations yield,

$$2\gamma_{11}(\mathbf{u}, \mathbf{v}) = 2\{C_W(\mathbf{0}) + \sigma_1^2 - C_W(\mathbf{u} - \mathbf{v})\}$$

$$2\gamma_{22}(\mathbf{u}, \mathbf{v}) = 2\{C_W(\mathbf{0}) + \sigma_2^2 - C_W(\mathbf{u} - \mathbf{v})\}$$

$$2\gamma_{12}(\mathbf{u}, \mathbf{v}) = 2\{(C_W(\mathbf{0}) + \tfrac{1}{2}(\sigma_1^2 + \sigma_2^2) - C_W(\mathbf{u} - \mathbf{v} - \Delta))$$

$$2\gamma_{21}(\mathbf{u}, \mathbf{v}) = 2\{C_W(\mathbf{u}) + \tfrac{1}{2}(\sigma_1^2 + \sigma_2^2) - C_W(\mathbf{u} - \mathbf{v} + \Delta)\}$$

all of which are functions of $u - \mathbf{v}$.

It is worth noting that when estimating cross-variograms nonparametrically, $2\gamma_{12}(\mathbf{h})$ can be estimated even when $Z_1(\cdot)$ is observed at far fewer and different sites than those at which $Z_2(\cdot)$ is observed (Clark, Basinger, and Harper, 1989). By contrast, $2\nu_{12}(\mathbf{h})$ can only be estimated from pairs of locations at which *both* $Z_1(\cdot)$ and $Z_2(\cdot)$ are observed.

Kunsch, Papritz, and Bassi (1997) have shown that there exist generalized cross-covariance functions ξ_{ij} that yield cokriging predictors, although they do not have a closed form expression for the ξ_{ij}. Further, they give models for which $\xi_{ij}(\mathbf{u}, \mathbf{v})$ depends only on the difference $\mathbf{u} - \mathbf{v}$, but for which γ_{ij} and C_{ij} do not; for example, simply make $W(\cdot)$ intrinsically stationary in the example given above. Nevertheless, our equivariance result in the previous section is true regardless of whether γ_{ij} depends on $(\mathbf{u}, \mathbf{v})$ or on $\mathbf{u} - \mathbf{v}$.

In conclusion, we have shown that cokriging with variance-based cross-variograms $2\gamma_{ij}^{(Z)}$ given by (7) are equivariant to mean and scale changes in either $Z_1(\cdot)$ or $Z_2(\cdot)$. Thus, a cokriging analysis in any set of units yields consistent predictors. In practice, it makes most sense to transform the processes so that their variabilities match (e.g., Eq. 25). Nevertheless, if they are not matched exactly, our result shows cokriging to be insensitive to mismatches.

ACKNOWLEDGMENTS

This research was supported by the U.S. Environmental Protection Agency under cooperative agreement CR822919-01-0. This article has not been subjected to review by the USEPA and thus no official endorsement should be inferred. The authors would like to thank Hans Wackernagel and an anonymous referee for their comments that led to improvements in the article.

REFERENCES

Clark, I., Basinger, K. L., and Harper, W. V., 1989, MUCK: A novel approach to co-kriging, *in* Buxton, B. E., ed., Proceedings of the Conference on Geostatistical, Sensitivity, and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling: Battelle Press, Columbus, OH, p. 473–493.

Cook, D., Cressie, N., Majure, J., and Symanzik, J., 1994, Some dynamic graphics for spatial data (with multiple attributes) in a GIS, *in* Dutter, R., and Grossman, W., eds., Compstat, Proceedings in Computational Statistics: Physica-Verlag, Heidelberg, p. 105–119.

Cressie, N., 1991, Statistics for spatial data: Wiley, New York, 900 p.

Journel, A. G, and Huijbregts, C. J., 1978, Mining geostatistics: Academic Press, New York, 600 p.

Kunsch, H. R., Papritz, A., and Bassi, F., 1997, Generalized cross-covariances and their estimation: Math. Geology, v. 29, no. 6, p. 779–799.

Majure, J. J., and Cressie, N., 1997, Dynamic graphics for exploring spatial dependence in multivariate spatial data: Geographical Systems, v. 4, no. 2, p. 131–158.

Myers, D., 1982, Matrix formulation of co-kriging: Math. Geology, v. 14, no. 3, p. 249–257.

Myers, D. E., 1991, Pseudo cross-variograms, positive-definiteness, and cokriging: Math. Geology, v. 23, no. 6, p. 805–816.

Myers, D. E., 1992, Kriging, cokriging, radial basis functions and the role of positive definiteness: Computers and Mathematics with Applications, v. 24, no. 12, p. 139–148.

Papritz, A., Kunsch, H. R., and Webster, R., 1993, On the pseudo cross-variogram: Math. Geology, v. 25, no. 8, p. 1015–1026.

Ver Hoef, J. M., and Barry, R. P., 1998, Constructing and fitting models for cokriging and multivariable spatial prediction: Journal of Statistical Planning and Inference, forthcoming.

Ver Hoef, J. M., and Cressie, N., 1993, Multivariable spatial prediction: Math. Geology, v. 25, no. 2, p. 219–239.

Wackernagel, H., 1988, Geostatistical techniques for interpreting multivariate spatial information, *in* Chung, C. F., Fabbri, A. G., and Sinding-Larsen, R., eds., Quantitative analysis of mineral and energy resources: Reidel, Dordrecht, p. 393–409.

Wackernagel, H., 1995, Multivariate geostatistics: Springer-Verlag, Berlin, 256 p.