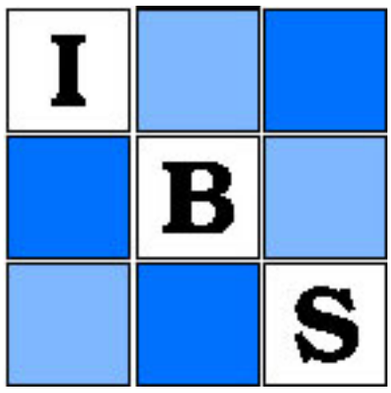

BIOMETRICS 47, 575–587
June 1991

Universal Kriging and Cokriging as a Regression Procedure

A. Stein

Soil Science and Geology Department, Agricultural University,
Duivendaal 10, 6701 AR Wageningen, The Netherlands

and

L. C. A. Corsten*

Department of Mathematics, Agricultural University, Wageningen, The Netherlands

SUMMARY

Prediction of a property on the basis of a set of point measurements in a region is required if a map of this property for the region is to be made. Of the spatial interpolation and prediction techniques, kriging is optimal among all linear procedures, as it is unbiased and has minimal variance of the prediction error. In cokriging, which has this same attractive property, additional observations of one or more covariables are used, which may lead to increased precision of the predictions. Both techniques are often applicable in different fields such as soil science, meteorology, medicine, agriculture, biology, public health, and environmental sciences (e.g., atmospheric or soil pollution). In this study we try to remove the cloud of obscurity covering the notions of kriging and cokriging by embedding them into regression procedures. This leads to a straightforward formulation of the two techniques. It turns out that kriging and cokriging differ only slightly from each other. The procedures are illustrated by two numerical examples, one to demonstrate the methodology, and one practical problem encountered in a soil study. Cokriging is found to be most valuable when a highly correlated covariable is sampled intensely.

1. Introduction

Traditional map production is increasingly being replaced by computer-based statistical prediction. An interesting well-known prediction technique has been developed in France since 1963 by Matheron and his coworkers (cf. Matheron, 1963, 1971; Delfiner, 1976) and is generally known as kriging, named after the South African D. Krige, the first practitioner of the method in the 1950s (cf. Krige, 1951). This technique performs well in predicting the value of a possible but actually not taken observation of a spatially distributed variable such as a mine grade (Krige, 1966), a soil characteristic (Webster, 1985), rainfall (Witter, unpublished thesis, Agricultural University, Wageningen, 1984), or gene frequency (Piazza, Menozzi, and Cavalli-Sforza, 1981). After predicting a spatially distributed variable at many locations in a region of interest (e.g., in the nodes of a grid) from a few point observations, one can produce a map of this variable (e.g., by ''contours'' or isopleths, differently shaded areas, or graphical techniques for surfaces in three-dimensional space). This activity may be termed interpolation. The predictor in any location will be a linear combination of the observations with positive or negative weights. This technique provides not only a prediction, but also an estimate of the variance of the prediction error. So a general indication of the quality of a map can be given, determined by the data and by the configuration of the observations as a whole. Among all linear predictors kriging is the one without systematic prediction error and with smallest prediction variance.

*Current address: Ritzemabosweg 20, 6703 AX Wageningen, The Netherlands.

Key words: Cokriging; Covariance function; Cross-covariance function; Interpolation; Kriging; Map quality; Prediction; Regression; Spatial phenomena.

Kriging is not the only spatial prediction method; other nonstatistical methods include tessellation and splines. Kriging, however, is the only method with a sound statistical base and is to be applied when uncertainties exist concerning measurement error and when the variation is a function of the distance between measurements (e.g., Webster, 1985). Kriging is compared with splines in Dubrule (1983).

Although kriging is best in the above sense, more information is often available to a scientist or a surveyor than what is contained in just one variable. So quite naturally an extension to kriging, known as cokriging, emerged (Matheron, 1979; Journel and Huijbregts, 1978). This technique, with attractive properties similar to those of kriging (e.g., unbiasedness, minimum prediction error variance), utilises observations of more than one variable in predicting one variable. It has found many applications, e.g., in soil science: the prediction of water stored in soil at $\frac{1}{3}$ bar pressure (Vauclin et al., 1983), reducing the costs of soil surveys (Stein et al., 1988), mapping top-soil silt content (McBratney and Webster, 1983). It has also been used in other disciplines: e.g., predicting radar-measured rainfall data (Krajewski, 1987) and jointly estimating peak values of earthquake response spectra (Carr and McCallister, 1985). In all these situations, the different variables are combined to give optimal predictions of one variable.

Cokriging, however, has had up to now two serious disadvantages. First, it is quite difficult to grasp, due to the complexity of notation (cf. Journel and Huijbregts, 1978, p. 171). Second, in contrast to kriging (Cressie, 1986), it could not fully be used in the presence of a trend in the variables due to the fact that the theory of estimation of spatial cross-covariance and its limitations is not completely available. Therefore we shall restrict ourselves to the situation where covariances are known.

Often Lagrange multipliers are used to derive the kriging and cokriging formulae, and they are used in the presentation of the final formulae as well (see, for example, Journel and Huijbregts, 1978). We believe that the complexity of notation is largely due to their use in the derivation and in the presentation of the results. Further, they are not necessary, as we have to minimize quadratic functions only.

It is known how kriging can be understood in the context of linear regression applied to dependent observations (cf. Corsten, 1985, 1989). The present article shows how cokriging can be seen in a similar fashion; kriging emerges as a special case. There are links with other fields in statistics, e.g., query response with best linear unbiased predictors (Gianola and Goffinet, 1982), current topics in experimental designs such as Bayesian analysis (Steinberg, 1988) and nearest neighbour models (Gleeson and Cullis, 1987), decomposition of the prediction error (Harville, 1985), and kernel smoothing and splines (Friedman and Silverman, 1989). In all the topics mentioned here, the variable is decomposed into three constituent parts: a fixed effect, locally dependent errors, and a purely random effect. As is shown below, this applies to universal kriging and cokriging as well.

One of the important aspects to be stressed here is that we aim to estimate a stochastic variable (i.e., to *predict* a possible future observation) instead of to estimate a parameter (i.e., to *estimate the expectation* of an observation). In ordinary linear regression theory with uncorrelated disturbances, all with the same variance, the two approaches differ only in their standard deviations. The best prediction and the estimate of its expectation for a new point with regressor values contained in the vector $\mathbf{x}_0$ will, admittedly, take the same value in this case. But the standard deviation of the estimator of the expectation is the residual standard deviation, multiplied by $\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$, while for the predictor the last factor must be replaced by $\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$. Here $\mathbf{X}$ is the matrix where each of the k columns is an n-vector containing the observations of a regressor in the n observation points. The difference between these multiplication factors can be impressively large. Apart from numerical differences, prediction of a random variable and estimation of its expectation are fundamentally different. From now on we deal mainly with the estimator of a stochastic effect, i.e., with prediction.

2. Universal Cokriging as a Regression Procedure

The term *regionalized variable* is commonly used to indicate that a variable takes values in a specific one-, two-, or three-dimensional space. The regionalized variable for which predictions are to be provided is called the target variable, or the *predictand*. It is assumed that the predictand has been observed in a given pattern of m points. In addition, n observations of another covariable in possibly different points are available. We restrict ourselves here to only one covariable, the more general case of $k - 1$ covariables being treated in Appendix B.

The prediction problem can be stated in the following way. We require a random predictor t for the random value y_0 of the predictand at a point z_0 on the basis of the m observations at the points $z_1, \ldots, z_m$ of the predictand, which are contained in the stochastic m-vector $\mathbf{y}_1$. In addition, observations for the covariable at n completely or partially different points $z_{m+1}, \ldots, z_{m+n}$ are available and are contained in the stochastic n-vector $\mathbf{y}_2$. Although in practice there will be shared measurement locations, which are in fact necessary to estimate the joint spatial structure function, there is no need to have those in the present prediction problem.

The expectation of each element of $\mathbf{y}_1$ and $\mathbf{y}_2$ is supposed to be a polynomial in the one, two, or three coordinates of the observation points z_i $(i = 1, \ldots, m)$ and z_j $(j = m + 1, \ldots, m + n)$, respectively, with respect to an arbitrary coordinate system, and so the expectations obey the structure

$$\mathrm{E}\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1, \tag{2.1}$$

$$\mathrm{E}\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2. \tag{2.2}$$

Both $\mathrm{E}\mathbf{y}_1$ and $\mathrm{E}\mathbf{y}_2$ are linear combinations of the regressor vectors with as yet unknown parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. The matrix $\mathbf{X}_1$ consists of m rows and p columns, where p is dependent on the degree (d) of the polynomial expectation or *trend* of the variable and the dimension (dim) of the region in which the observations are taken. It is easily shown that

$$p = \frac{(dim + d)!}{dim!\,d!} = \binom{dim + d}{dim}. \tag{2.3}$$

In a two-dimensional region and without trend, $\mathbf{X}_1$ is merely the vector $\mathbf{1}_m$, consisting of m ones only. If the degree of the trend is equal to 1, each row of $\mathbf{X}_1$ consists of $1, \xi_1, \xi_2$, where ξ_1 and ξ_2 represent the coordinates of an observation point of the predictand. If the trend is quadratic, each row of $\mathbf{X}_1$ consists of $1, \xi_1, \xi_2, \xi_1^2, \xi_1\xi_2, \xi_2^2$. Similarly, $\mathbf{X}_2$ is an $n \times p$ matrix where each row consists of p monomial values of the coordinates of an observation point of the covariable. The term *universal* for kriging and cokriging indicates the presence of trend terms.

Obviously now $\mathrm{E}y_0 = \mathbf{x}_0'\boldsymbol{\beta}_1$, in which the vector $\mathbf{x}_0'$ consists of one row of p monomial values of the coordinates of the point z_0 where a prediction is required, similarly to the p regressors in $\mathbf{X}_1$.

In contrast to trend surface analysis (Watson, 1971), where observations are supposed to be independent, the observations in kriging and cokriging have a special dependence structure. This is modelled by means of the covariance between (actual and hypothetical) observation points or their disturbances with respect to expected values.

The dependence structure of the vector $(\mathbf{y}_1', \mathbf{y}_2', y_0)'$ is assumed to be given by the symmetric covariance matrix $\mathbf{C}^*$, of order $m + n + 1$. This matrix may be partitioned as

$$\mathbf{C}^* = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{c}_{01} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{c}_{02} \\ \mathbf{c}_{01}' & \mathbf{c}_{02}' & c_{00} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{c}_0 \\ \mathbf{c}_0' & c_{00} \end{bmatrix}, \tag{2.4}$$

where $\mathbf{C}_{11}$ is the covariance matrix of the elements of $\mathbf{y}_1$, $\mathbf{C}_{22}$ of the elements of $\mathbf{y}_2$, $\mathbf{C}_{12}$ $(= \mathbf{C}'_{21})$ between the elements $\mathbf{y}_1$ and those of $\mathbf{y}_2$, and $\mathbf{c}_{01}$, $\mathbf{c}_{02}$, and c_{00} between y_0 and the elements of $\mathbf{y}_1$, $\mathbf{y}_2$, and y_0, respectively, while

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}.$$

The matrix $\mathbf{C}^*$ must be positive-definite in order that the variance of any linear predictor be nonnegative. Each element of $\mathbf{C}_{11}$ and of $\mathbf{c}_{01}$ as well as c_{00} is assumed to be known as an isotropic function $c_1(r)$, only of the distance r between the pair of observation points concerned, the so-called covariance function for the predictand. Likewise, each element of $\mathbf{C}_{22}$ is assumed to be known as an isotropic function $c_2(r)$ of the distance r between the pair of observation points concerned, the covariance function for the covariable; and each element of $\mathbf{C}_{12}$ and of $\mathbf{c}_{02}$ is assumed to be known as an isotropic function $c_{12}(r)$ of the distance r between the pair of observation points concerned, the so-called cross-covariance function between predictand and covariable. Further assumptions about the covariance functions are not required. To avoid complications we do not consider here the conditions that $c_1(r)$, $c_2(r)$, $c_{12}(r)$ must satisfy for $\mathbf{C}^*$ to be positive-definite. Neither are problems of estimating covariance functions addressed in this paper (see, e.g., Ripley, 1981). Nor do we deal with generalized covariance functions, which are to be applied only to linear combinations of possible observations $(\mathbf{y}'_1, \mathbf{y}'_2)'$ whose coefficient vector must obey certain linear restrictions (see, e.g., Delfiner, 1976; Dowd, 1989). The semivariogram, which is a special case of the opposite of such a generalized covariance function particularly in the absence of a trend, is not considered in this expository context, once more in order to keep the material as simple as possible (Corsten, 1989).

We have now the following linear model for actual observations $\mathbf{y}$ and hypothetical observation of the predictand y_0:

$$\begin{bmatrix} \mathbf{y} \\ y_0 \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{x}_0^{*\prime} \end{bmatrix} \boldsymbol{\beta} + \mathbf{e}; \quad \mathrm{E}\mathbf{e} = \mathbf{0}; \quad \mathrm{cov}(\mathbf{e}) = \mathbf{C}^*; \tag{2.5}$$

in which we have introduced the abbreviations

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix}, \quad \mathbf{x}_0^* = \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}. \tag{2.6}$$

We require the best random predictor $t = \boldsymbol{\lambda}'\mathbf{y}$, linear in the observations $\mathbf{y}$, to satisfy the conditions that

$$\mathrm{E}(\boldsymbol{\lambda}'\mathbf{y} - y_0) = 0 \tag{2.7}$$

and

$$\mathrm{var}(\boldsymbol{\lambda}'\mathbf{y} - y_0) \text{ is minimal}. \tag{2.8}$$

With $\mathbf{V} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}$, Appendix A proves that

$$t = \mathbf{x}_0^{*\prime}\hat{\boldsymbol{\beta}} + \mathbf{c}'_0\mathbf{C}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{2.9}$$

where $\hat{\boldsymbol{\beta}} = \mathbf{V}\mathbf{X}'\mathbf{C}^{-1}\mathbf{y}$, the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$.

The predictor t in (2.9) is the sum of the estimated local expectation of the predictand only (that is, $\mathbf{x}_0^{*\prime}\hat{\boldsymbol{\beta}} = \mathbf{x}'_0\hat{\boldsymbol{\beta}}_1$) and a linear combination of the observed residuals contained in $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, with the best linear multivariate approximation of the predictand by all $m + n$ observations as coefficients. Because $t - \mathbf{x}_0^{*\prime}\hat{\boldsymbol{\beta}} = \mathbf{c}'_0\mathbf{C}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, the procedure can be interpreted as regression of the residual of t with respect to $\mathbf{x}_0^{*\prime}\hat{\boldsymbol{\beta}}$ on the residuals of $\mathbf{y}$ with respect to $\mathbf{X}\hat{\boldsymbol{\beta}}$.

The variance of the prediction error equals

$$\mathrm{var}(t - y_0) = c_{00} - \mathbf{c}'_0\mathbf{C}^{-1}\mathbf{c}_0 + \mathbf{x}'_a\mathbf{V}\mathbf{x}_a, \tag{2.10}$$

where $\mathbf{x}_a = \mathbf{x}_0^* - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c}_0$.

This variance of the prediction error (2.10), free of Lagrange multipliers, can be interpreted as follows:

c_{00} is the variance of the variable under study;
$\mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0$ is the reduction of that variance due to the best linear approximation by the other observations considered as $m + n$ additional multivariate characteristics;
$\mathbf{x}_a'\mathbf{V}\mathbf{x}_a = \text{var}(\mathbf{x}_a'\hat{\boldsymbol{\beta}})$ is the variance of the sum of estimated expectations of predictand and covariable, however, not in the prediction location with regressors contained in $\mathbf{x}_0^*$, but in a different but related point with regressors contained in $\mathbf{x}_a$.

In fact, the prediction error variance reflects the orthogonal decomposition of the prediction error into $y_0 - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}$ and $\mathbf{x}_a'\hat{\boldsymbol{\beta}}$, respectively.

The predictor t is an exact predictor: if a prediction is carried out at an observation point, the vector $\mathbf{c}_0$ is equal to the corresponding column of $\mathbf{C}$, leading to the observation itself as the best prediction. Then, it turns out that the variance of the prediction error will vanish.

Repeated application of t at different locations requires revision of $\mathbf{x}_0$ and $\mathbf{c}_0$ only, whereas the costly inversion of the matrix $\mathbf{C}$ is required only once.

A confidence interval for the predictand can be given if $\mathbf{y}$ has a multivariate Gaussian distribution. It follows that with probability $1 - \alpha$ the interval $[t - u(\alpha/2)\text{var}(t - y_0)^{1/2}, t + u(\alpha/2)\text{var}(t - y_0)^{1/2}]$ will cover y_0, where $u(\alpha/2)$ is the value such that the probability that $u < u(\alpha/2)$ equals $1 - \frac{1}{2}\alpha$, u following the standard normal distribution.

3. Particular Cases

3.1 *Universal Kriging*

The universal kriging equations are obtained when observations are present for the predictand only—that is, with n equal to 0. All submatrices with subscript 2 will now vanish. As a result, we obtain the same equations for the predictor and the variance of the prediction error as equations (2.9) and (2.10), but where $\mathbf{y}$ now reduces to $\mathbf{y}_1$, $\mathbf{X}$ to $\mathbf{X}_1$, $\mathbf{C}$ to $\mathbf{C}_{11}$, $\mathbf{c}_0$ to $\mathbf{c}_{01}$, $\hat{\boldsymbol{\beta}}$ to $\hat{\boldsymbol{\beta}}_1$, and $\mathbf{V}$ and $\mathbf{x}_a$ to the corresponding submatrix and subvector, while $\mathbf{x}_0^*$ reduces to $\mathbf{x}_0$.

3.2 *The Nugget Effect*

In many studies, a so-called "nugget effect", representing measurement error variance plus variation at very small distance, is encountered, often in combination with spatial dependence. We then have for $c_1(r)$ that $c_1(r) = B\delta(r) + f_1(r)$, where $\delta(r) = 1$ if $r = 0$ and $\delta(r) = 0$ if $r \neq 0$, the coefficient B is the size of the nugget effect, and $f_1(r)$ models the spatial dependence. Similar structures apply for $c_2(r)$ and $c_{12}(r)$.

A feature sometimes encountered in regionalized studies is the so-called pure nugget effect, i.e., correlation zero for nonzero distance. Then the covariance functions have the structure $c_1(r) = \sigma_1^2\delta(r)$, $c_2(r) = \sigma_2^2\delta(r)$, and $c_{12}(r) = \sigma_{12}\delta(r)$.

Without nugget effect the interpolation surface will be entirely smooth, passing through the observations, while in the presence of a nugget effect the observations will appear as spikes above or below the smooth prediction surface. With pure nugget effect this prediction surface will coincide with the estimated trend surface.

4. Numerical Examples

Example 1 To illustrate different aspects of universal kriging and cokriging, a simple example was constructed. Along a straight line two observations of the predictand, with values equal to 21 and 23, were situated at locations z_1 and z_2 with coordinates 1 and 3, respectively. Also, three observations of the covariable with values 5, 6, and 6 were situated on locations z_3, z_4, and z_5 with coordinates 1, 2, and 3, respectively. The covariance functions $c_1(r)$ and $c_2(r)$ as

Table 1
Values for the coefficients of the covariance functions in Example 1

	B	A	b
$c_1(r)$	1	3	.5
$c_2(r)$	.3	1.7	.5
$c_{12}(r)$	.4	1.9	.5

well as the cross-covariance function $c_{12}(r)$ were assumed to be known, and to obey an exponential model with nugget effect (cf. §3.2): $c(r) = B\delta(r) + A\exp(-r/b)$, where $\delta(r) = 1$ if $r = 0$ and $\delta(r) = 0$ if $r \neq 0$. The values of B, A, and b for each of the three covariance functions are presented in Table 1. The values are such that the correlation coefficient between the variables equals $+.73$. Attention was focused on the difference between kriging and cokriging, with and without a linear trend. Some intermediate steps are shown in Table 2. For kriging without trend the sole value of $\boldsymbol{\beta}$ equals the average of the two observations. For (universal) kriging with a linear trend the values of $\boldsymbol{\beta}$ are those for a line through the points. However, the values are modified by the introduction of the covariable, and become less clearly interpretable.

Predictions were carried out for all points with coordinates between 0 and 4 (Figure 1); standard deviations of prediction error are given in Figure 2. Numerical attention is focused on the locations P_1, P_2, and P_3 with coordinates 2, 2.5, and 3.5, respectively (Table 3). Point P_1 is of interest as there is an observation of the covariable, but no observation of the predictand; point P_2 is an interior point for which typically predictions are needed, e.g., if a map is to be constructed; and with point P_3 extrapolation can be illustrated.

The prediction in P_1 increases by .58 if a covariable is used, because of the relatively large value of the observation of the covariable in the prediction location. An increase of the prediction in P_2 by approximately .25 is observed, due both to trend and to the use of a covariable. The prediction in P_3 is barely influenced by the presence of a covariable, but the presence of a trend has major influence. Finally, it is seen from Figure 1 how predictions are in fact modifications with respect to the trend surface; also noted are the discontinuities at the observation points of the predictand, due to the nugget effect.

The strongest reduction in the standard deviation of the prediction error when the covariable is included takes place in P_1, due to the large value of $\mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0$ caused by the strong cross-correlation between the two variables. Further, a major part of the variance of the prediction error is due to the fact that the parameters in $\boldsymbol{\beta}$ have to be estimated, implying that $\mathbf{x}_a'\mathbf{V}\mathbf{x}_a$ is far from negligible (Table 3).

In summary, this example shows how kriging and cokriging may perform differently, in particular on a small set of data.

Example 2 The second example stems from a soil study in the Limagne area in central France (Stein, Bouma, Mulders, and Weterings, 1989). The central problem of this example is to characterize soil variability within an area without clear physiographic features and to map a crucial soil parameter for crop growth simulation modelling over the area. As a rule, physiographic features such as faults, relief, brooks, erosion forms, etc., are useful tools to study soil spatial variability in an area and to delineate soil units, each with a "representative profile."

The study area is located on one of the most recent terraces of the river Allier, some 30 kilometers downstream of the city of Clermont-Ferrand. The area of 200 ha is naturally bounded at one side by the Allier, and at the other side by a somewhat higher-situated and older terrace. Moisture availability is one of the most important land qualities for agriculture as depth to groundwater is greater than 1.2m. On a generalized 1:100,000 soil map, which is the only one available for this region (INRA, 1965) no soil delineations are presented, which implies that this region is considered to be fairly homogeneous on this scale. When more detailed investigations

Table 2
Intermediate results for Example 1

	Kriging		Cokriging	
	No trend	Linear trend	No Trend	Linear trend
y	21 23	21 23	21 23 6 6 5	21 23 6 6 5
X	1 1	1 1 1 3	1 0 1 0 0 1 0 1 0 1	1 1 0 0 1 3 0 0 0 0 1 2 0 0 1 3 0 0 1 1
C	4.000 .055 .055 4.000		4.000 .055 .257 .035 2.300 .055 4.000 .257 2.300 .035 .257 .257 2.000 .230 .230 .035 2.300 .230 2.000 .031 2.300 .035 .230 .031 2.000	
V	2.027	9.918 −3.945 −3.945 1.973	1.703 .888 .888 .774	9.593 −3.945 5.418 −2.265 −3.945 1.973 −2.265 1.133 5.418 −2.265 4.712 −1.969 −2.265 1.133 −1.969 .984
$\hat{\beta}$	22.000	20.000 1.000	22.178 5.654	20.178 1.000 4.654 .500

	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3
$\mathbf{c}_0$	.406 .406	1.104 .149	1.104 .020	.406 .406	1.104 .149	1.104 .020	.406 .406 2.300 .257 .257	1.104 .149 .699 .699 .095	1.104 .020 .699 .095 .013	.406 .406 2.300 .257 .257	1.104 .149 .699 .699 .095	1.104 .020 .699 .095 .013
$\mathbf{x}_0^*$	1.00	1.00	1.00	1.00 2.00	1.00 2.50	1.00 3.50	1.00 .00	1.00 .00	1.00 .00	1.00 2.00 .00 .00	1.00 2.50 .00 .00	1.00 3.50 .00 .00
$\mathbf{x}_a$	.80	.69	.72	.80 1.60	.69 1.64	.72 2.67	.83 −.94	.75 −.34	.78 −.10	.83 1.65 −.94 −1.89	.75 1.80 −.34 −.77	.78 2.83 −.10 −.30

are carried out on a larger scale (e.g., 1:20,000), clay, sand, and gravel layers exhibit irregular patterns that are difficult to map, due again to lack of clearly visible physiographic features (Stein et al., 1989). A regular sampling scheme was therefore used to investigate the spatial variability (Figure 3).

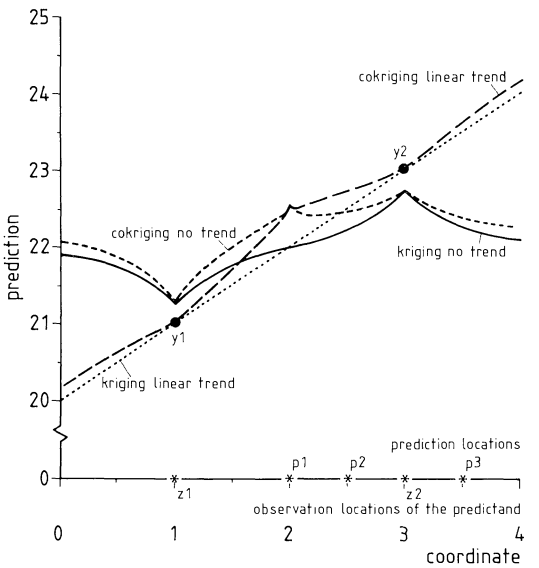

Figure 1. Predictions carried out with kriging and with cokriging, without a trend and with a linear trend.

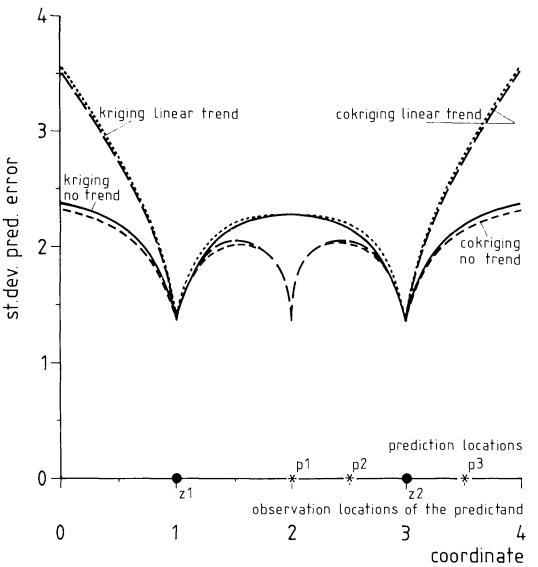

Figure 2. Standard deviation of the prediction error for kriging and for cokriging, without a trend and with a linear trend.

Four transects were planned with observation points 100m apart. Two transects were located more or less parallel to the Allier River, and two transects were chosen perpendicular to the river, yielding a total of 62 observation points. To investigate variability of soil survey data at short distances, an additional sampling scheme was used (Figure 3) with observations at distances 1m, 2m, 4m, 8m, 16m, and 32m, in two directions from every fourth observation point of the main transect perpendicular to the river. From any of those points one additional observation was taken parallel to the river at a distance 1m, 2m, 4m, 8m, 16m, and 32m, respectively. In every observation point a number of basic soil survey measurements were made

Table 3
Predictions and negative or positive contributions to the prediction error variance by $\mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0$ *and* $\mathbf{x}_a'\mathbf{V}\mathbf{x}_a$ *in Example* 1. *The value of* $\mathbf{c}_{00}$ *is equal to* 4.0 *in all cases.*

		Trend	Prediction	$\mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0$	$\mathbf{x}_a'\mathbf{V}\mathbf{x}_a$	var($\mathbf{t} - \mathbf{y}_0$)
P_1	Kriging	None	22.0	.081	1.297	5.215
		Linear	22.0	.081	1.297	5.215
	Cokriging	None	22.58	2.666	.468	1.803
		Linear	22.58	2.666	.468	1.803
P_2	Kriging	None	22.24	.309	.968	4.659
		Linear	22.5	.309	1.099	4.790
	Cokriging	None	22.47	.507	.588	4.081
		Linear	22.74	.507	.722	4.215
P_3	Kriging	None	22.28	.305	1.059	4.755
		Linear	23.5	.305	4.021	7.717
	Cokriging	None	22.39	.311	.898	4.587
		Linear	23.63	.311	3.863	7.552

Table 4
Values for the coefficients of the covariance functions in Example 2

		B	A	b
AVW	$c_1(r)$	1,310	2,580	.28
DG	$c_2(r)$	306	522	.27
AVW × DG	$c_{12}(r)$	342	1,100	.25

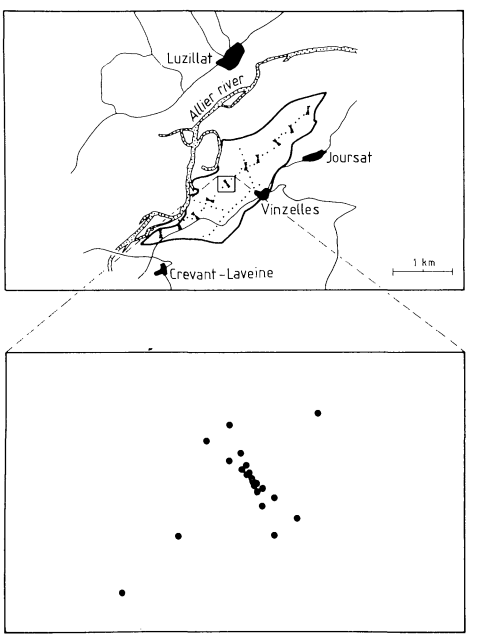

Figure 3. The sampling scheme on the river terrace in the Limagne area in France.

by means of a boring to 1.2m, or to an impenetrable gravel layer, whichever appeared first. Gravel layers within 1.2m were encountered in 192 of the 304 observation points, sometimes occurring close to the surface. Soil physical properties were measured on the main transects, yielding 62 calculations of "available water" (AVW), which is defined as the volume of water

between field capacity and 15 bar. Available water is directly related to crop yield (e.g., FAO, 1976).

Prediction of the land quality AVW for unvisited locations throughout the terrace was carried out with kriging. Since no trend was expected, covariance and cross-covariance functions were estimated by existing standard methods (Journel and Huijbregts, 1978). Estimated coefficients are given in Table 4. Because soil physical properties require relatively expensive and time-consuming measurement procedures, cokriging is attractive to be used as well. The choice for "depth to gravel" (DG) as a covariable is self-evident, since it is easy to measure and highly correlated with AVW (correlation coefficient .88).

The AVW map presented in Figure 4 shows optimal predictions for unvisited locations. It is noted how the gravel bank in the center of the area gives rise to restricted water availability.

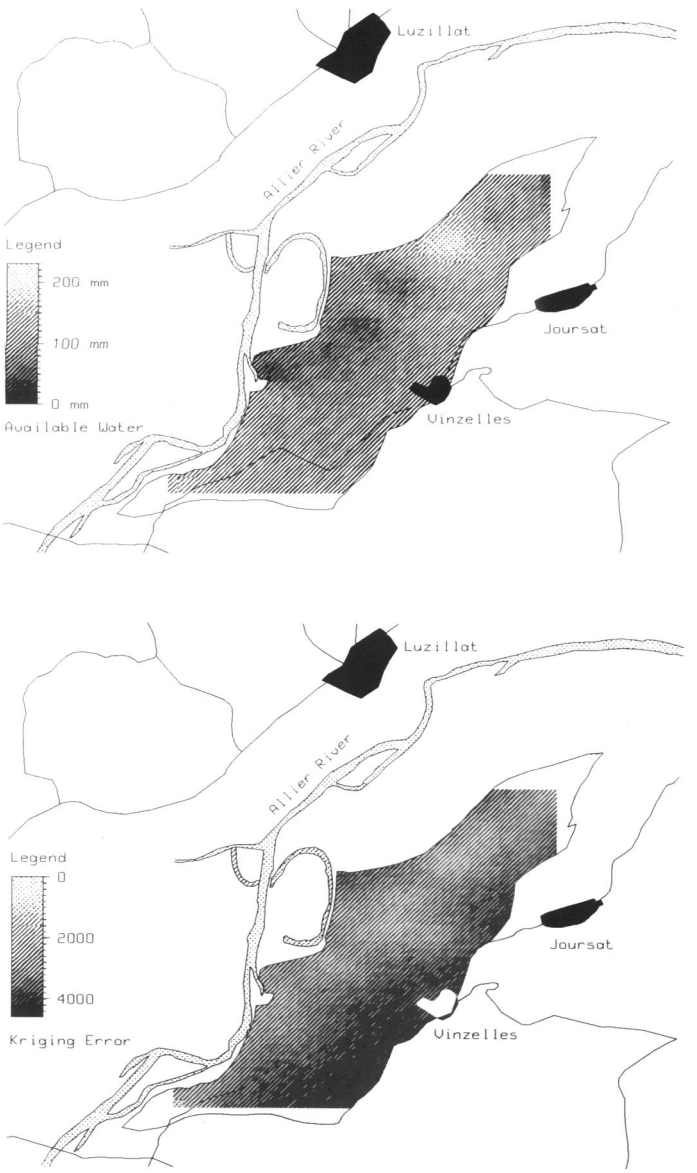

Figure 4. The cokriged AVW map (top) and the map with the prediction error variances (bottom).

Table 5
Minimum, maximum, and mean prediction error variances in Example 2, for 1,200 random points, for two subsets thereof, and for points with an observation of the covariable but not of the predictand

Test set	Number of points	Kriging			Cokriging		
		Min	Max	Mean	Min	Max	Mean
All random data	1,200	1,955	3,985	3,099	1,591	3,917	2,917
Subset 1	833	1,955	3,907	2,909	1,591	3,800	2,694
Subset 2	205	1,955	3,198	2,541	1,591	2,809	2,162
All observations with DG < 1.2m and without observation of AVW	159	1,878	3,133	2,252	1,005	1,206	1,080

To compare the gain in precision relative to kriging, the prediction error variance was determined in 1,200 randomly located points within the level river terrace, for both kriging and cokriging. Minimum, maximum, and average values are presented in Table 5. In order to investigate the relation between the prediction error variance and the distance to measurement locations, the prediction error variances were determined for subset 1, containing the points within 280m of observations of the predictand and for subset 2, containing those points within 100m of observations of the predictand. Finally, the prediction error variance was determined in the 159 data points, all within 64m of an AVW observation where a DG measurement was present, but no AVW observation.

The soil variability patterns were checked by multispectral aerial photographs, made with a near-infrared (NIR) filter and with a green (GR) filter. The (NIR − GR)/(NIR + GR) ratio is used as a vegetation index. For 37 of the 62 locations of the 100m transects containing maize, the greytones were classified according to a standard sensitivity guide. As AVW correlates highly with DG, it is likely that differences in greytones were caused by stress situations, due to shortage of available water in the profile. The vegetation index yielded a significant regression equation (at $\alpha = .05$ level) with AVW as explanatory variable and the vegetation index as dependent variable. This is considered a validation of the AVW map obtained by cokriging.

Conclusions from this example are:

1. The prediction error variance is a markedly increasing function of the distance to observation points. This implies a need for dense sampling of the predictand, which is, however, usually prevented by the associated costs.
2. A gain in precision with respect to kriging of 50% in terms of the mean variance of the prediction error is observed if a highly correlated covariable is included. This covariable has to be sampled on all grid points where a prediction is carried out to produce a predictive soil map. The gain in precision for other points is structural, but small, ranging in the average from 6% to 12%, depending on the test set.
3. Evidence on aerial photographs supports the AVW values obtained in this study.

Acknowledgements

The authors thank Dr A. C. van Eijnsbergen from the Mathematics Department, Prof. J. Bouma from the Soil Science and Geology Department, and Dr J. J. de Gruijter from the Agricultural Mathematics Group, Wageningen, The Netherlands, for their helpful suggestions.

Résumé

La prédiction d'une caractéristique à partir d'un ensemble de mesures ponctuelles dans une région est nécessaire pour dresser la carte de cette région. Parmi toutes les techniques d'interpolation et de prédiction d'une variable spatiale, le krigeage est la meilleure des procédures linéaires car elle est sans biais et

minimise la variance de l'erreur. Dans le cokrigeage, qui posséde les mêmes propriétés, on utilise les mesures complémentaires d'une ou plusieurs covariables, ce qui amène une meilleure précision des valeurs prédites. Ces deux techniques sont souvent applicables dans des domaines variés tels que la pédologie, la météorologie, la médecine, l'agriculture, la biologie, la santé publique, l'étude de l'environnement (comme la pollution de l'atmosphère et des sols). Dans cette étude nous essayons de dissiper l'obscurité qui couvre les notions de krigeage et de cokrigeage en les assimilant à des procédures de régression. Il en découle une formulation très simple des deux techniques, qui montre que celles-cis diffèrent très peu l'une de l'autre. Les procédures sont illustrées par deux exemples numériques, l'une explicitant la méthodologie, l'autre concernant un problème pratique rencontré en pédologie. Le cokrigeage semble être plus intéressant lorsque l'on utilise une covariable fortement corrélée mesurée un grand nombre de fois.

References

Carr, J. R. and McCallister, P. G. (1985). An application of cokriging for estimation of tripartite earthquake response spectra. *Mathematical Geology* **17**, 527–545.

Corsten, L. C. A. (1985). Current statistical issues in agricultural research. *Statistica Neerlandica* **39**, 159–168.

Corsten, L. C. A. (1989). Interpolation and optimal linear prediction. *Statistica Neerlandica* **43**, 69–84.

Cressie, N. A. C. (1986). Kriging nonstationary data. *Journal of the American Statistical Association* **81**, 625–634.

Delfiner, P. (1976). Linear estimation of nonstationary spatial phenomena. In *Advanced Geostatistics in the Mining Industry*, M. Guarascio, M. David, and C. Huijbregts (eds), 49–68. Dordrecht: Reidel.

Dowd, P. A. (1989). Generalized cross-covariances. In *Geostatistics*, M. Armstrong (ed.), 151–162. Dordrecht: Kluwer.

Dubrule, O. (1983). Two methods with different objectives: Splines and kriging. *Mathematical Geology* **15**, 245–257.

FAO (1976). A framework for land evaluation. *FAO Soils Bulletin* **32**. Rome: Food and Agriculture Organization of the United Nations.

Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modelling. *Technometrics* **31**, 3–21.

Gianola, D. and Goffinet, B. (1982). Query Response: Sire evaluation with best linear unbiased predictors. *Biometrics* **38**, 1085–1088.

Gleeson, A. C. and Cullis, B. R. (1987). Residual maximum likelihood estimation of a neighbour model for field experiments. *Biometrics* **43**, 277–288.

Harville, D. (1985). Decomposition of prediction error. *Journal of the American Statistical Association* **80**, 132–138.

Institute Nationale de la Recherche Agronomique (1965). Carte pédologique du val d'Allier. Feuille sud: Clermont-Ferrand. Montpellier, France: Service d'Etudes des Sols.

Journel, A. G. and Huijbregts, Chr. J. (1978). Mining geostatistics. London: Academic Press.

Krajewski, W. F. (1987). Cokriging radar-rainfall and rain gauge data. *Journal of Geophysical Research* **92**, 9571–9580.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* **52**, 119–138.

Krige, D. G. (1966). Two-dimensional weighted moving average trend surfaces for ore evaluation. In *Proceedings of Symposium on Mathematical Statistics and Computer Applications in Ore Valuation*. Special edition of *Journal of South African Institute of Mining and Metallurgy*, Johannesburg, 13–38.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246–1266.

Matheron, G. (1971). The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathématique, No. 5. Paris: Ecole de Mines de Paris.

Matheron, G. (1979). Recherche de simplification dans un problème de cokrigeage. Fontainebleau: Centre de Géostatistique.

McBratney, A. B. and Webster, R. (1983). Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. *Journal of Soil Science* **34**, 137–162.

Piazza, A., Menozzi, P., and Cavalli-Sforza, L. (1981). The making and testing of geographic gene-frequency maps. *Biometrics* **37**, 635–659.

Ripley, B. (1981). *Spatial Statistics*. New York: Wiley.

Stein, A., Bouma, J., Van Dooremolen, W., and Bregt, A. K. (1988). Cokriging point data on moisture deficits. *Soil Science Society of America Journal* **52**, 1418–1423.

Stein, A., Bouma, J., Mulders, M. A., and Weterings, M. H. W. (1989). Using spatial variability studies to estimate physical land qualities of a level river terrace. *Soil Technology* **2**, 385–402.

Stein, A., Bouma, J., Kroonenberg, S. B., and Cobben, S. (1989). Sequential sampling to measure the infiltration rate within relatively homogeneous land units. *Catena* **16**, 91–100.

Steinberg, D. M. (1988). Factorial experiments with time trends. *Technometrics* **30**, 259–269.
Vauclin, M., Vieira, S. R., Vachaud, G., and Nielsen, D. R. (1983). The use of cokriging with limited field soil observations. *Soil Science Society of America Journal* **47**, 175–184.
Watson, G. S. (1971). Trend surface analysis. *Mathematical Geology* **3**, 215–226.
Webster, R. (1985). Quantitative spatial analysis of soil in the field. In *Advances in Soil Science* **3**, B. A. Stewart (ed.), 1–70. New York: Springer-Verlag.

Received October 1988; *revised July* 1989 *and February* 1990; *accepted March* 1990.

Appendix A
Derivation of Predictor and Prediction Error Variance

This appendix gives the derivation of the formulae (2.9) and (2.10). We start by making some preliminary observations and consider a number, say k, of stochastic characteristics y_i, on the basis of which the best linear predictor for y_0 is to be formulated [that is, the predictor u, not necessarily unbiased, with minimal $\mathrm{var}(u - y_0)$], while $\mathbf{C}$ and $\mathbf{c}_0$, the covariances among the y_i and those between the y_i and y_0, are known. Then one has to determine in $\langle y_1, \ldots, y_k \rangle$, the space spanned by the y_i's, the vector $\boldsymbol{\mu}'\mathbf{y}$ such that $(y_0 - \boldsymbol{\mu}'\mathbf{y})$ has minimum norm or length, where the inner product is defined as the covariance between two vectors. This is equivalent to finding the orthogonal projection of y_0 on $\langle y_1, \ldots, y_k \rangle$, or $\mathrm{cov}(y_0 - \boldsymbol{\mu}'\mathbf{y}, y_i) = 0$ for $i = 1, \ldots, k$, or $\mathbf{C}\boldsymbol{\mu} = \mathbf{c}_0$, the normal equations, leading to $u = \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}$. The vector $y_0 - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}$ is orthogonal (uncorrelated) to any linear combination of $y_1, \ldots, y_k$ and $\mathrm{var}(y_0 - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}) = \mathrm{var}(y_0) - \mathrm{var}(\mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}) = c_{00} - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0$, c_{00} being the variance of y_0.

Now with $\mathrm{E}\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{E}\,y_0 = \mathbf{x}_0^{*\prime}\boldsymbol{\beta}$, according to (2.5), we wish to minimize $\mathrm{var}(t - y_0)$ under the condition that $\mathrm{E}(t - y_0) = \mathbf{0}$. In order to accomplish this, the prediction error is split into

$$t - y_0 = \left(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}\right) - \left(y_0 - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}\right), \tag{A.1}$$

the difference of two orthogonal terms, and $\mathrm{var}(t - y_0) = \mathrm{var}(\boldsymbol{\lambda}'\mathbf{y} - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}) + c_{00} - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0$.

Now, for $\boldsymbol{\nu}'\mathbf{y}$, with $\boldsymbol{\nu}'$ equal to $\boldsymbol{\lambda}' - \mathbf{c}_0'\mathbf{C}^{-1}$, we require that $\mathrm{E}(\boldsymbol{\nu}'\mathbf{y}) = \mathbf{x}_a'\boldsymbol{\beta}$, with $\mathbf{x}_a' = (\mathbf{x}_0^{*\prime} - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{X})$, because of unbiasedness. Minimizing $\mathrm{var}(t - y_0)$ is equivalent to minimizing $\mathrm{var}(\boldsymbol{\nu}'\mathbf{y})$ while $\mathrm{E}(\boldsymbol{\nu}'\mathbf{y}) = \mathbf{x}_a'\boldsymbol{\beta}$. This minimum is attained by the GLS estimator $\hat{\boldsymbol{\beta}} = \mathbf{V}\mathbf{X}'\mathbf{C}^{-1}\mathbf{y}$, where $\mathbf{V} = (\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}$. Hence $\boldsymbol{\nu}'\mathbf{y} = \mathbf{x}_a'\hat{\boldsymbol{\beta}}$ and thus $t = \boldsymbol{\nu}'\mathbf{y} + \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y}$ will be

$$t = \mathbf{x}_a'\hat{\boldsymbol{\beta}} + \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{y} = \mathbf{x}_0^{*\prime}\hat{\boldsymbol{\beta}} + \mathbf{c}_0'\mathbf{C}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right), \tag{A.2}$$

as mentioned in (2.9). Further, the minimum of $\mathrm{var}(t - y_0)$ is according to decomposition (A.1) equal to $\mathrm{var}(\mathbf{x}_a'\hat{\boldsymbol{\beta}}) + c_{00} - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0$. Replacement of $\mathrm{var}(\mathbf{x}_a'\hat{\boldsymbol{\beta}})$ by $\mathbf{x}_a'\mathbf{V}\mathbf{x}_a$ yields (2.10).

Appendix B
Extension to k Variables

Extension of the preceding method to k variables, i.e., prediction of a variable with $k - 1$ covariates, is straightforward. Again, $\mathbf{y}_1$ is the vector containing observations of the predictand, whereas $\mathbf{y}_2, \ldots, \mathbf{y}_k$ are the vectors containing observations of the covariables. In a linear model we have

$$\mathrm{E}\begin{bmatrix}\mathbf{y}\\ y_0\end{bmatrix} = \begin{bmatrix}\mathbf{X}\\ \mathbf{x}_0^{*\prime}\end{bmatrix}\boldsymbol{\beta}, \tag{B.1}$$

where $\mathbf{X} = \mathrm{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_k)$, a block-diagonal matrix with the matrix $\mathbf{X}_i$ as ith block and the vectors $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_k', y_0)'$ and $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_k')$, $\mathbf{x}_0^*$ containing a string of $k-1$ null-vectors in addition to the regressors for the predictand. With $\mathbf{C}$ the covariance matrix of actual observations (of predictand and covariables), $\mathbf{c}_0$ the vector of covariances between the predictand at the unvisited location with all observations contained in $\mathbf{y}$, and c_{00} the variance of the predictand at the unvisited spot, we obtain

$$t = \mathbf{x}_0^{*\prime}\hat{\boldsymbol{\beta}} + \mathbf{c}_0'\mathbf{C}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right), \tag{B.2}$$

$$\mathrm{var}\left(t - y_0\right) = c_{00} - \mathbf{c}_0'\mathbf{C}^{-1}\mathbf{c}_0 + \mathbf{x}_a'\mathbf{V}\mathbf{x}_a, \tag{B.3}$$

where now $\hat{\boldsymbol{\beta}}' = (\hat{\boldsymbol{\beta}}_1', \ldots, \hat{\boldsymbol{\beta}}_k')$, and $\mathbf{V}$ and $\mathbf{x}_a$ are completely analogous to the expressions defined above. The formulae (B.2) and (B.3) are similar to (2.9) and (2.10), respectively.