Mathematical Geology, Vol. 23, No. 5, 1991

Indicator Principal Component Kriging[1]

V. Suro-Pérez[2] and A. G. Journel[2]

An alternative to multiple indicator kriging is proposed which approximates the full coindicator kriging system by kriging the principal components of the original indicator variables. This transformation is studied in detail for the biGaussian model. It is shown that the cross-correlations between principal components are either insignificant or exactly zero. This result allows derivation of the conditional cumulative density function (cdf) by kriging principal components and then applying a linear back transform. A performance comparison based on a real data set (Walker Lake) is presented which suggests that the proposed method achieves approximation of the conditional cdf equivalent to indicator cokriging but with substantially less variogram modeling effort and at smaller computational cost.

KEY WORDS: indicator kriging, principal component analysis, biGaussian model, indicator covariance matrix, orthogonalization.

INTRODUCTION

Most Earth Sciences data feature patterns of spatial continuity which can be used to model and assess the uncertainty prevailing at unsampled locations. Models of spatial continuity allow going beyond the actual data toward an assessment of the uncertainty specific to each unsampled location. Characterization of uncertainty and spatial interpolation are a primary goal of any geostatistical approach.

The uncertainty associated with an unsampled value $z(\mathbf{x})$ at location $\mathbf{x}$ can be modeled by the probability distribution of a random function (RF) $Z(\mathbf{x})$. This distribution is made conditional on the surrounding information. The degree to which the distribution of $Z(\mathbf{x})$ is influenced by the surrounding data is dictated by the prior model of spatial continuity or dependence between the $(n + 1)$ RVs $Z(\mathbf{x})$, $Z(\mathbf{x}_\alpha)$, $\alpha = 1, \ldots, n$. Consider then the conditional cumulative distribution function (cdf) of $Z(\mathbf{x})$:

$$F(\mathbf{x}; z|\{n\}) = P\{Z(\mathbf{x}) \leq z|\{Z(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha), \quad \alpha = 1, \ldots n\}\}$$

[1]Manuscript received 17 July 1990; accepted 11 January 1991.

[2]Applied Earth Sciences Department, Stanford University, Stanford, CA 94305-2225.

0882-8121/91/0700-0759$06.50/1

For any set $\{n\}$ of data values, there are as many conditional cdf's of type $F(\mathbf{x}; z|\{n\})$ as there are models of spatial continuity which relate the unknown value $z(\mathbf{x})$ to the data values. All such conditional cdf's $F_l(\mathbf{x}; z|\{n\})$, $l = 1, \ldots, L$, including those provided by all forms of indicator kriging, should be seen as alternative models of the uncertainty prevailing at the unsampled location $\mathbf{x}$, rather than different estimates of an elusive "true" cdf $F(\mathbf{x}; z|\{n\})$.

Prior to selection of any estimated value $z^*(\mathbf{x})$, access to a conditional cdf model $F_l(\mathbf{x}; z|\{n\})$ allows determination of probability intervals and probabilities of exceedence such as:

$$P\{Z(\mathbf{x}) \in (z_1, z_2]|\{n\}\} = F_l(\mathbf{x}; z_2|\{n\}) - F_l(\mathbf{x}; z_1|\{n\})$$

$$P\{Z(\mathbf{x}) > z_1|\{n\}\} = 1 - F_l(\mathbf{x}; z_1|\{n\})$$

The concept of a loss function (Journel, 1984) allows deriving from the model $F_l(\mathbf{x}; z|\{n\})$ not just one but as many optimal estimates for $z(\mathbf{x})$ as there are different criteria for optimality. This usually involves the minimization of the expected value of some loss function associated with the estimation error $z^*(\mathbf{x}) - z(\mathbf{x})$.

Disjunctive Kriging (DK) (Matheron, 1976; Rivoirard, 1989), Indicator Kriging (IK) (Journel, 1983), Multigaussian Kriging (MG) (Verly, 1983), Probability Kriging (PK) (Sullivan, 1984), Uniform Conditioning (Guibal and Remacre, 1984), and BiGaussian Kriging (Marcotte and David, 1985) are all techniques providing models for such conditional cdf's. Except for IK and PK, all these algorithms are parametric in the sense that they call for a prior bivariate or multivariate distribution model for the random function $Z(\mathbf{x})$; the parameters of the conditional cdf $F_l(\mathbf{x}; z|\{n\})$ are then derived by some form of kriging. In the case of IK and PK, the bivariate distribution of any pair of RFs $Z(\mathbf{x})$ and $Z(\mathbf{x}')$ is inferred directly from data rather than derived from prior models.

Principal Component Analysis (PCA) (Anderson, 1984; Borgman and Frahme, 1976; Davis and Greenes, 1983) and the indicator formalism where any attribute value $z(\mathbf{x})$ is coded into an indicator vector of 0's and 1's, are used in this paper to model the conditional cdf. The approach presented here attempts to fill the gap between the IK approach which uses indicator autocovariances to model $F_l(\mathbf{x}; z|\{n\})$ and the theoretically better Indicator Cokriging (CoIK) which makes use of both indicator autocovariances and cross-covariances. The proposed approach provides a model of the conditional cdf by kriging the indicator principal components instead of the indicators themselves.

Presentation of the principal components approach is accomplished first by considering a parametric point of view. The biGaussian model is studied in detail and analytical expressions for the corresponding indicator crosscovariances are derived. It is shown that these indicator crosscovariances are not negligible with respect to indicator autocovariances. CoIK would account for such

cross-information, however, it is difficult to implement. PCA amounts to summarizing the indicator auto- and cross-covariances information into the principal component (pc) autocovariances. The resulting pc crosscovariances are almost null.

The performance of this latter approach, hereafter denoted IPCK (Indicator Principal Component Kriging), is analyzed using a real data set (Walker Lake). Prediction intervals are obtained using CoIK, IPCK, and IK. The theoretically better CoIK algorithm results are compared to those provided by the nonparametric technique IPCK. It is shown that IPCK provides a good approximation to the CoIK reference model, yet requires much less modeling and computational effort.

THE INDICATOR APPROACH

The indicator approach calls for the coding of the stationary random function $Z(\mathbf{x})$ into a series of indicators (binary random variables) defined for different cutoffs z_k:

$$I(\mathbf{x}; z_k) = \begin{cases} 1 & \text{if } Z(\mathbf{x}) \leq z_k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

This coding defines a stationary indicator vector:

$$I(\mathbf{x}; \mathbf{z}) = [I(\mathbf{x}; z_1) \ldots I(\mathbf{x}; z_K)]^T \tag{2}$$

where K is the total number of cutoffs. Note the indicator vector above does not reproduce exactly the original value $Z(\mathbf{x})$. If the indicator vector is such that $i(\mathbf{x}; z_k) = 0$ and $i(\mathbf{x}; z_{k+1}) = 1$, then the attribute $z(\mathbf{x})$ recovered from the indicator vector is only known to be in the class $z(\mathbf{x}) \in (z_k, z_{k+1}]$. The indicator coding (1) allows an immediate relationship between the indicator expected value and the original Z univariate distribution:

$$E\{I(\mathbf{x}; z_k)\} = P\{Z(\mathbf{x}) \leq z_k\} = F(z_k) \tag{3}$$

Similarly, the bivariate cdf of any pair of random variables $Z(\mathbf{x})$, $Z(\mathbf{x} + \mathbf{h})$ can be expressed as the noncentered indicator covariance $K_I(\mathbf{h}; z_k z_{k'})$:

$$\begin{aligned} K_I(\mathbf{h}; z_k, z_{k'}) &= E\{I(\mathbf{x}; z_k) I(\mathbf{x} + \mathbf{h}; z_{k'})\} \\ &= P\{Z(\mathbf{x}) \leq z_k, Z(\mathbf{x} + \mathbf{h}) \leq z_{k'}\} \qquad k, k' = 1, \ldots, K \end{aligned} \tag{4}$$

Note that strict stationarity is assumed in Eq. (4) for expressing the noncentered covariance as function of only the distance $\mathbf{h}$. Furthermore, the indicator approach amounts to discretizing the Z-bivariate distribution for each lag $\mathbf{h}$ by K^2 indicator covariance values.

CoIndicator Kriging

The CoIndicator kriging (CoIK) algorithm makes use of the full bivariate distribution as discretized by K cutoff values as in Eq. (4), and thus requires the modeling of K^2 indicator auto- and cross-covariances of the type:

$$C_I(\mathbf{h}; z_k, z_{k'}) = K_I(\mathbf{h}; z_k, z_{k'}) - F(z_k)F(z_{k'})$$

The CoIK model for the conditional cdf is written:

$$F_{\text{CoIK}}(\mathbf{x}; z_{k_0}|\{n\}) = \sum_{k'=1}^{K} \sum_{\alpha=1}^{n} \lambda_{k_0 k'\alpha} I(\mathbf{x}_\alpha; z_{k'}) \tag{5}$$

The weights $\lambda_{k_0 k'\alpha}$ are obtained by minimizing the estimation variance:

$$E[I(\mathbf{x}; z_{k_0}) - F_{\text{CoIK}}(\mathbf{x}; z_{k_0}|\{n\})]^2 \tag{6}$$

which entails the following system of constrained normal equations (Myers, 1982):

$$\sum_{k'=1}^{K} \sum_{\beta=1}^{n} \lambda_{k_0 k'\beta} C_I(\mathbf{x}_\alpha - \mathbf{x}_\beta; z_k, z_{k'}) + \mu_k = C_I(\mathbf{x} - \mathbf{x}_\alpha; z_{k_0}, z_k)$$

$$k = 1, \ldots, K; \alpha = 1, \ldots n$$

$$\sum_{\beta=1}^{n} \lambda_{k_0 k\beta} = \delta_{kk_0}, \qquad k = 1, \ldots, K \tag{7}$$

where the μ_k's are Lagrange multipliers resulting from the minimization of (6). δ_{kk_0} is the Kronecker delta:

$$\delta_{kk_0} = \begin{cases} 1 & \text{if } k = k_0 \\ 0 & \text{otherwise} \end{cases}$$

Remarks

- For each cutoff z_{k_0}, the CoIK estimator requires solution of a system of $K(n + 1)$ equations.
- Spatial clustering and preferential sampling in high or low values can render the inference of $O(K^2)$ indicator autocovariances and cross-covariances difficult. Even if there are enough data and the sampling is unbiased, the number of indicator autocovariances and cross-covariances required to achieve enough resolution (K large enough) may be prohibitive. This is precisely why CoIK is rarely, if ever, used in practice.

- Disjunctive Kriging (DK) allows cutting the previous inference burden to the single autocovariance of a specific transform of the original variable $Z(\mathbf{x})$. But this is possible only by calling for a specific bivariate distribution model for any pair of RVs $Z(\mathbf{x})$, $Z(\mathbf{x} + \mathbf{h})$, which may or may not be appropriate. The major difference between DK and CoIK is the decision in the latter case to model the bivariate distribution from data.
- A model of conditional cdf that is fully conditional on all n data values $z(\mathbf{x}_\alpha)$ taken altogether would call for the $(n + 1)$-variate distribution of the $(n + 1)$ RVs $Z(\mathbf{x})$, $Z(\mathbf{x}_\alpha)$, $\alpha = 1, \ldots n$ (Journel, 1977). Clearly, such high-order multivariate distribution cannot be inferred from the sparse data available, and would necessarily come from a parametric model—for example, from the multiGaussian model (Verly, 1983). The unique parameter (actually a function) of this Gaussian model that is inferred from actual data is the correlogram of either $Z(\mathbf{x})$ or its normal score transform.

Indicator Kriging

The IK model can be seen as an approximation of the CoIK model (5), where the cross-covariances are ignored:

$$F_{\mathrm{IK}}(\mathbf{x};\, z_{k_0} | \{n\}) = \sum_{\alpha=1}^{n} \lambda_{k_0\alpha} I(\mathbf{x}_\alpha;\, z_{k_0}) \tag{8}$$

The weights $\lambda_{k_0\alpha}$ are obtained by minimizing the corresponding estimation variance, leading to the following system of constrained normal equations (Journel, 1983):

$$\sum_{\beta=1}^{n} \lambda_{k_0\beta} C_I(\mathbf{x}_\beta - \mathbf{x}_\alpha;\, z_{k_0}) + \mu_{k_0} = C_I(\mathbf{x} - \mathbf{x}_\alpha;\, z_{k_0}), \qquad \alpha = 1, \ldots, n$$

$$\sum_{\beta=1}^{n} \lambda_{k_0\beta} = 1 \tag{9}$$

The required information is now limited to K autocovariances corresponding to the K cutoffs z_k.

Remarks

- The IK model $F_{\mathrm{IK}}(\mathbf{x};\, z_{k_0})$ requires solution of a system of $(n + 1)$ equations for each cutoff z_{k_0}.
- Although IK ignores the cross-covariances, the resulting model values $F_{\mathrm{IK}}(\mathbf{x};\, z_{k'} | \{n\})$ and $F_{\mathrm{IK}}(\mathbf{x};\, z_k | \{n\})$ are not independent since the con-

ditioning information $I(\mathbf{x}_\alpha; z_{k'})$ and $I(\mathbf{x}_\alpha; z_{k''})$ is correlated across cutoff values.

- By retaining only the indicator of nonexceedence $i(\mathbf{x}_\alpha; z_{k_0})$, for modeling the conditional cdf at z_{k_0}, IK does not recognize how large or small is each original sample value $z(\mathbf{x}_\alpha)$. CoIK recognizes within which class falls each original data by considering the full indicator data vector $I(\mathbf{x}_\alpha; \mathbf{z})$.

PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is an algebraic technique involving a linear transform of one vector into another. In two dimensions, PCA can be seen as a rotation of orthogonal axes, with an angle chosen such that the spread of the first transformed variable along the first axis is maximum and the spread of the second transformed variable along the second axis is minimum (Anderson, 1984).

PCA is an orthogonalization procedure which does not require any prior statistical hypothesis about the data. However, statistical interpretation of this orthogonalization shows that the cross-covariance between the transformed variables, say $Y_i(\mathbf{x})$, is null:

$$E\{(Y_i(\mathbf{x}) - \mu_{Y_i})(Y_j(\mathbf{x}) - \mu_{Y_j})\} = 0, \qquad \forall\, j \neq i$$

μ_{Y_i} being the mean of component $Y_i(\mathbf{x})$. Additionally, the first transformed variable $Y_1(\mathbf{x})$, or first principal component, has maximum variance, the second principal component, $Y_2(\mathbf{x})$, the second largest variance, and so on.

These properties have made PCA a popular way to reduce the dimension of the data by selecting a reduced number of principal components, those which contribute most to the variance. For example, in geochemistry where the number of variables is usually large, PCA has been used extensively to reduce the dimension of the problem (Howarth, 1983). Such a decision is based on the assumption that the variance is the most important aspect of variability and that retaining those variables with the largest contribution to that variance would provide a concise yet satisfactory explanation of the source of that variability. However, the choice of the variance as the sole criterion to rank sources of variability is debatable.

Transforming the Indicators

The indicator vector at location $\mathbf{x}$ is defined by expressions (1) and (2). The corresponding indicator covariance matrix $\Sigma_I(\mathbf{h}_1)$ for any specific distance

vector $\mathbf{h}_1$ is:

$$\Sigma_I(\mathbf{h}_1) = \begin{bmatrix} C_I(\mathbf{h}_1; z_1, z_1) & \cdots & C_I(\mathbf{h}_1; z_1, z_K) \\ C_I(\mathbf{h}_1; z_2, z_1) & \cdots & C_I(\mathbf{h}_1; z_2, z_K) \\ \vdots & \cdots & \vdots \\ C_I(\mathbf{h}_1; z_K, z_1) & \cdots & C_I(\mathbf{h}_1; z_K, z_K) \end{bmatrix} \tag{10}$$

where:

$$C_I(\mathbf{h}_1; z_k, z_{k'}) = \text{cov}\ \{I(\mathbf{x}; z_k), I(\mathbf{x} + \mathbf{h}_1; z_{k'})\}$$

One way to obtain the corresponding principal components is to consider the spectral decomposition of $\Sigma_I(\mathbf{h}_1)$ (Anderson, 1984) defined as:

$$\Sigma_I(\mathbf{h}_1) = \mathbf{A}\Lambda\mathbf{A}^T \tag{11}$$

where $\mathbf{A}$ is an orthonormal matrix and Λ is a diagonal matrix. Both matrices are defined specifically for the vector $\mathbf{h}_1$. By virtue of this decomposition, the columns of matrix $\mathbf{A}$ are the eigenvectors of $\Sigma_I(\mathbf{h}_1)$ and the elements of the diagonal matrix Λ are the eigenvalues of $\Sigma_I(\mathbf{h}_1)$ ordered from largest to smallest ($\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$) (Wilkinson, 1965).

Once the matrix $\mathbf{A}$ is calculated from (11) for some specific lag $\mathbf{h}_1$, the indicator principal component vector is obtained by simple matrix multiplication:

$$\mathbf{Y}(\mathbf{x}) = \mathbf{A}^T\mathbf{I}(\mathbf{x}; \mathbf{z}) \tag{12}$$

Each element of $\mathbf{Y}(\mathbf{x})$ is thus written:

$$Y_k(\mathbf{x}) = \sum_{k'=1}^{K} a_{k',k} I(\mathbf{x}; z_{k'}) \tag{13}$$

where $a_{k',k}$ and $I(\mathbf{x}; z_{k'})$ are the elements of matrix $\mathbf{A}$ and vector $\mathbf{I}(\mathbf{x}; \mathbf{z})$, respectively. The new variables Y_k are linear combinations of the original indicators with the property that their cross-covariances for $\mathbf{h} = \mathbf{h}_1$ are exactly zero, $C_Y(\mathbf{h}_1; k, k') = 0$ for all $k \neq k'$. The variance of Y_k is equal to the k^{th} eigenvalue (ω_k) of $\Sigma_I(\mathbf{h}_1)$:

$$C_Y(0; k, k) = \omega_k, \qquad \forall\ k = 1, \ldots, K \tag{14}$$

The orthogonalization (12) does *not* ensure that the cross-covariances $C_Y(\mathbf{h}; k, k')$ are zero for $\mathbf{h} \neq \mathbf{h}_1$. If that were the case, the cokriging of $\mathbf{Y}(\mathbf{x})$ would reduce to the separate kriging of each of its elements $Y_k(\mathbf{x})$. However, since $Y_k(\mathbf{x})$ and $Y_{k'}(\mathbf{x} + \mathbf{h})$ are uncorrelated by construction at $\mathbf{h}_1$, it is conjectured and hereafter checked, that their level of correlation at any larger $\mathbf{h}$ will be negligible.

Kriging of Indicator Principal Components

A set of n data vectors $\mathbf{Y}(\mathbf{x}_\alpha)$ are obtained from the transformation (13) and a cokriging-type estimator of the principal component $Y_{k_0}(\mathbf{x})$ is implemented:

$$Y^*_{k_0}(\mathbf{x}) = \sum_{k=1}^{K} \sum_{\alpha=1}^{n} \lambda_{k_0 k\alpha} Y_k(\mathbf{x}_\alpha) \tag{15}$$

Assuming that orthogonality holds true for all $\mathbf{h}$:

$$C_Y(\mathbf{h}; k, k') \approx 0, \qquad \forall\, k \neq k' \tag{16}$$

allows reducing the estimator (15) to its simpler ordinary kriging version:

$$Y^*_{k_0}(\mathbf{x}) = \sum_{\alpha=1}^{n} \lambda_{k_0\alpha} Y_{k_0}(\mathbf{x}_\alpha) \qquad k_0 = 1, \ldots, K \tag{17}$$

with the weights $\lambda_{k_0\alpha}$ being derived from a constrained normal system of type (9):

$$\sum_{\beta=1}^{n} \lambda_{k_0\beta} C_Y(\mathbf{x}_\beta - \mathbf{x}_\alpha; k_0) + \mu_{k_0} = C_Y(\mathbf{x} - \mathbf{x}_\alpha; k_0), \qquad \alpha = 1, \ldots, n$$

$$\sum_{\beta=1}^{n} \lambda_{k_0\beta} = 1$$

with:

$$C_Y(\mathbf{x}_\alpha - \mathbf{x}_\beta; k_0) = \text{cov}\,(Y_{k_0}(\mathbf{x}_\alpha), Y_{k_0}(\mathbf{x}_\beta))$$

The IPCK Model

The inverse of transform (12) provides a model for the conditional cdf of $Z(\mathbf{x})$:

$$\mathbf{F}_{\text{IPCK}}(\mathbf{x}; \mathbf{z}|\{n\}) = \mathbf{A}\mathbf{Y}^*(\mathbf{x}) \tag{18}$$

where the vector $\mathbf{F}_{\text{IPCK}}(\mathbf{x}; \mathbf{z}|\{n\})$ is defined as:

$$\mathbf{F}_{\text{IPCK}}(\mathbf{x}; \mathbf{z}|\{n\}) = [F_{\text{IPCK}}(\mathbf{x}; z_1|\{n\}) \ldots F_{\text{IPCK}}(\mathbf{x}; z_K|\{n\})]^T$$

and the vector $\mathbf{Y}^*(\mathbf{x})$ is:

$$\mathbf{Y}^*(\mathbf{x}) = [Y^*_1(\mathbf{x}) \ldots Y^*_K(\mathbf{x})]^T \tag{19}$$

Note that the estimator (17) and the model (18) appear as linear combinations of all indicator data associated with all K cutoffs, as is the CoIK model (5).

THE BIGAUSSIAN CASE

Consider a stationary random function $Z(\mathbf{x})$, such that any pair $Z(\mathbf{x})$, $Z(\mathbf{x} + \mathbf{h})$ has a standard bivariate normal distribution with zero mean and covariance matrix:

$$\Sigma_Z(\mathbf{h}) = \begin{bmatrix} 1 & \rho(\mathbf{h}) \\ \rho(\mathbf{h}) & 1 \end{bmatrix}, \quad \text{with } |\rho(\mathbf{h})| < 1 \tag{20}$$

With these parameters, the bivariate normal distribution is expressed as (Anderson, 1984):

$$f(z, z'; \rho(\mathbf{h})) = \frac{1}{2\pi\sqrt{1 - \rho^2(\mathbf{h})}} \exp\left[- \frac{z^2 + z'^2 - 2\rho(\mathbf{h})zz'}{2(1 - \rho^2(\mathbf{h}))} \right] \tag{21}$$

where $\rho(\mathbf{h})$ is the z-correlogram defined as:

$$\rho(\mathbf{h}) = \frac{\text{cov } (Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h}))}{\text{var } (Z(\mathbf{x}))} = E[Z(\mathbf{x})Z(\mathbf{x} + \mathbf{h})]$$

Using this standard bigaussian model, the corresponding indicator covariance can be expressed as (Suro-Perez, 1988):

$$C_I(\mathbf{h}; z, z') = \frac{1}{2\pi} \int_0^{\arcsin \rho(\mathbf{h})} \exp\left[- \frac{z^2 + z'^2 - 2zz' \sin\theta}{2 \cos^2\theta} \right] d\theta \tag{22}$$

The advantage of expression (22) against the traditional expression:

$$C_I(\mathbf{h}; z, z') = \int_{-\infty}^{z} \int_{-\infty}^{z'} f(x, y; \rho(\mathbf{h}))\, dx\, dy - F(z)F(z')$$

is the reduction of a double integral to a single integral.

Indicator Covariance Matrix $\Sigma_I(\mathbf{h})$

From expression (21), two properties of symmetry for the Gaussian indicator cross-covariances can be derived:

$$C_I(\mathbf{h}; z, z') = C_I(\mathbf{h}; z', z) \tag{23}$$

and

$$C_I(\mathbf{h}; z, z') = C_I(\mathbf{h}; -z, -z') \tag{24}$$

Expression (23) entails the symmetry of the covariance matrix $\Sigma_I(\mathbf{h})$ with respect to the main diagonal. The combination of expressions (23) and (24) entails symmetry with regard to the other diagonal if the K cutoff values are

chosen such that:

$$z_k = -z_{K-k+1} \qquad \forall\, k = 1, \ldots, K \tag{25}$$

This property is called persymmetry (Golub and Van Loan, 1983). The particular choice of cutoff values (25) implies that the indicator cross-covariances satisfy the relation:

$$C_I(\mathbf{h};\, z_k,\, z_{k'}) = C_I(\mathbf{h};\, z_{K-k'+1},\, z_{K-k+1}) \qquad \forall\, k,\, k' \tag{26}$$

Thus, for a proper choice of cutoff values, the covariance matrix $\Sigma_I(\mathbf{h})$ for a standard biGaussian distribution is not only symmetric:

$$\Sigma_I(\mathbf{h}) = \Sigma_I^T(\mathbf{h}) \tag{27}$$

it also satisfies the relationship of persymmetry:

$$\Sigma_I(\mathbf{h}) = \mathbf{E}\Sigma_I(\mathbf{h})\mathbf{E} \tag{28}$$

with:

$$\mathbf{E} = [\mathbf{e}_n \ldots \mathbf{e}_1]$$

and the vector e_i being defined as the i^{th}-column of a $K \times K$ identity matrix. Note that results (27) and (28) are valid for all $\mathbf{h}$.

Eigenvectors of $\Sigma_I(\mathbf{h}_1)$

Symmetry and persymmetry of $\Sigma_I(\mathbf{h}_1)$ yields specific properties for the corresponding eigenvectors. The eigenvectors associated with the odd order eigenvalues are written:

$$\mathbf{a}_l = [a_{1,l} \ldots a_{m-1,l} \quad a_{m,l} \quad a_{m-1,l} \ldots a_{1,l}]^T \qquad l = 2k-1,\ k = 1, \ldots, \frac{K+1}{2} \tag{29}$$

while the eigenvectors associated with the even-order eigenvalues are:

$$\mathbf{a}_u = [a_{1,u} \ldots a_{m-1,u} \quad 0 \quad -a_{m-1,u} \ldots -a_{1,u}]^T \qquad u = 2k,\ k = 1, \ldots, \frac{K-1}{2} \tag{30}$$

where K is an odd number of cutoff values and m is:

$$m = \frac{K+1}{2}$$

As will be seen in the following section, these specific expressions have important consequences for the computation of the principal component covariance matrix.

Computation of the Principal Component Cross-Covariances

By definition, indicator principal components are linear combinations of the original indicators, with weights corresponding to the $\Sigma_I(\mathbf{h}_1)$-eigenvectors. If $Y_l(\mathbf{x})$ is the l^{th}-principal component and $Y_u(\mathbf{x} + \mathbf{h})$ is the u^{th}-principal component, the cross-covariance is written as:

$$C_Y(\mathbf{h};\, l,\, u) = \text{cov}\,(\mathbf{a}_l^T\mathbf{I}(\mathbf{x};\, \mathbf{z}),\, \mathbf{a}_u^T\mathbf{I}(\mathbf{x} + \mathbf{h};\, \mathbf{z})) = \mathbf{a}_l^T\Sigma_{\mathbf{I}}(\mathbf{h})\mathbf{a}_u \qquad (31)$$

where $\mathbf{a}_l$ and $\mathbf{a}_u$ are the previously defined eigenvectors associated to the matrix $\Sigma_I(\mathbf{h}_1)$. The latter product (31) is zero for all $\mathbf{h}$ under the conditions of the following theorem:

Theorem. The crosscovariance of the principal components y_l and y_u derived from the indicator variable $\mathbf{I}(\mathbf{x};\, \mathbf{z})$ are zero:

$$C_Y(\mathbf{h};\, l,\, u) = 0 \;\forall\; \mathbf{h}$$

if $\Sigma_I(\mathbf{h})$ is symmetric and persymmetric, and the indexes l and u are odd and even, respectively, or vice versa.

Proof. Assuming l odd and u even and the expressions (29) and (30) for the corresponding eigenvectors, the product $\mathbf{a}_l^T\Sigma_I(\mathbf{h})$ is written:

$$\mathbf{a}_l^T\Sigma_I(\mathbf{h}) = [d_1 \;\ldots\; d_{m-1} \;\; d_m \;\; d_{m-1} \;\ldots\; d_1]$$

Thus, the final product is:

$$\mathbf{a}_l^T\Sigma_I(\mathbf{h})\mathbf{a}_k = 0, \qquad \forall\; \mathbf{h}$$

This result implies that some principal component cross-covariances are zero exactly for all $\mathbf{h}$ and, consequently, that the joint estimation of the elements of $\mathbf{Y}(\mathbf{x})$ can be achieved by the solution of a sparse cokriging system. That sparse cokriging system can be further approximated by kriging each Y_k separately. This last approximation assumes that all the cross-covariances are zero when in fact only some are so exactly.

THE PRACTICE OF IPCK

The seven following steps represent a typical sequence for application of IPCK:

Declustering the Univariate CDF

For the choice $|\mathbf{h}| = 0$, all elements of the indicator covariance matrix $\Sigma_I(\mathbf{h})$ are sole functions of the univariate cdf $F(z)$. From data an experimental distribution $F^*(z)$ can be inferred. In the presence of preferential sampling,

these data need to be declustered in order to correct the bias introduced by their preferential location. This problem is discussed in Journel (1983) and Isaaks and Srivastava (1989).

Selection of Cutoff Values

The choice of cutoff values conditions the structure of the indicator covariance matrix and the corresponding orthogonal matrix **A**. Symmetric quantiles for in Gaussian case implies that some principal component crosscovariances are exactly zero for all **h**. The decision about the number of cutoffs depends on each specific application. If the main interest focuses on the high quantiles, then this part of the cdf should be discretized more. The number of cutoffs in this approach is not a serious problem since the burden of covariance modeling is alleviated by the fact that the higher the order of the indicator principal component, in general, the lesser its autocorrelation (Suro-Perez, 1988). The last indicator principal component autocovariances are often quasi-pure nugget effect.

Choice of h for the Spectral Decomposition of $\Sigma_I(\mathbf{h})$

Orthogonalization of $\Sigma_I(\mathbf{h})$ can be done for any **h**. Ideally, the specific distance $\mathbf{h}_1$ at which orthogonalization is to be performed should be such that after transformation (12), the indicator principal component cross-covariances can be considered negligible. A reasonable first choice is the smallest lag distance of the experimental indicator covariances/variograms, because it is expected that for most bivariate distribution models the indicator principal component cross-covariances are decreasing functions of **h**. Another choice would be either $\mathbf{h} = 0$ or $\mathbf{h}_1 \to 0$.

Checking for Zero Cross-Correlation

Since the IPCK goal is to approximate CoIK, the relevant check consists in evaluating the relative magnitude of the principal component cross-covariances with respect to the corresponding autocovariances. If that relative magnitude is small, then IPCK can be applied safely, no matter the bivariate distribution, whether biGaussian or not. If that relative magnitude is large, IPCK is not recommended.

Order Relations Problems

The resulting model $F_{\text{IPCK}}(\mathbf{x};\, z_k|\{n\})$ does not necessarily satisfy order relations for cdf's that are:

$$F_{\text{IPCK}}(\mathbf{x};\, z_k|\{n\}) \in [0,\, 1]$$

$$F_{\text{IPCK}}(\mathbf{x};\, z_k|\{n\}) \leq F_{\text{IPCK}}(\mathbf{x};\, z_{k'}|\{n\}), \qquad \forall\, z_{k'} > z_k$$

The first condition may not be satisfied because kriging-type estimates are nonconvex linear combinations of the conditioning data. Indeed, the kriging weights can be negative and the estimate can lie outside the limits defined by the maximum and minimum data values. The second type of order relations problems is due to the fact that the model value $F_{\text{IPCK}}(\mathbf{x}; z_k|\{n\})$ is not constrained by $F_{\text{IPCK}}(\mathbf{x}; z_{k+1}|\{n\})$; indeed, the two respective kriging systems do not impose any such constraint. However, experience has shown (Sullivan, 1984; Suro-Perez and Journel, 1990) that although order relations are numerous, they are all of small magnitude and can be corrected easily to obtain a model for the conditional distribution which satisfies the requirements of a cdf.

Optimal Estimates

Various estimates for the original unsampled value $z(\mathbf{x})$ can be derived from the conditional cdf $F_{\text{IPCK}}(\cdot)$; they are all optimal but for different criteria of optimality established from a loss function concept (Journel, 1989). For a given loss function $L(\cdot)$ of the random error $z^*(\mathbf{x}) - Z(\mathbf{x})$, an estimate of the expected loss is:

$$E\{L(z^*(\mathbf{x}) - Z(\mathbf{x}))|\{n\}\} = \int_0^1 L(z^*(\mathbf{x}) - z)\, dF_{\text{IPCK}}(\mathbf{x}; z|\{n\}) \quad (32)$$

Minimization of expression (32) yields the corresponding L-optimal estimate $z_L^*(\mathbf{x})$. Numerical evaluation can be obtained through repetitive numerical integration for a series of values $z^*(\mathbf{x})$.

Probability Intervals

Probability intervals can be computed directly from the conditional cdf model:

$$P^*\{Z(\mathbf{x}) \in (a, b)|\{n\}\} = F_{\text{IPCK}}(\mathbf{x}; b|\{n\}) - F_{\text{IPCK}}(\mathbf{x}; a|\{n\}) \quad (33)$$

The probability of exceedence of any given threshold c is:

$$P\{Z(\mathbf{x}) \geq c|\{n\}\} = 1 - F_{\text{IPCK}}(\mathbf{x}; c|\{n\};) \quad (34)$$

Note that probability intervals and probability of exceedence are independent of the choice of the particular optimal estimate $z^*(\mathbf{x})$. These measures of uncertainty are intrinsic to the information $\{n\}$ retained and the model of spatial correlation used to build the conditional cdf model.

WALKER LAKE DATA SET

This data is a subset of the larger Walker Lake data set presented by Isaaks and Srivastava (1989). This subset was defined and used by the US-EPA (United States Environmental Protection Agency) to evaluate a "variance of geostatis-

Table I. Walker Lake Statistics[a]

No. of data	Mean	Variance	$q_{0.25}$	$q_{0.75}$	min	max	cv
19,800	194.134	127,570.2	4.324	229.66	0.0	5505.924	1.84

[a]cv is the coefficient of variation, q_p is the p-quantile, min is the minimum value, and max is the maximum value.

ticians'' (Englund, 1990). The variable is a local elevation roughness index associated to the variance of 25 contiguous topographic evaluation data. Table I presents some univariate statistics of the data set. Note the large coefficient of variation (1.8) indicating a highly skewed distribution. Figure 1 shows a gray scale map based on all reference 19800 data. A NE–SW structure is observed, which is captured in the Z-correlogram map shown in Fig. 2.

Considering only the 126 sample data retained by the EPA study, 660 conditional cdf's are modeled at locations regularly spaced and shown in Fig. 3. Table II shows the statistics of the 126 samples considered for this study. The kriging plan consists of retaining a minimum of 3 and a maximum of 16 sample data, with a maximum of 4 samples per quadrant. The search neighborhood is an ellipse oriented in the direction of maximum continuity (N68°E), with a major axis equal to 25 units and a minor axis equal to 13 units.

CoIK, IK, and IPCK are the three algorithms employed to model the conditional distributions. Nine decile cutoffs corresponding to the exhaustive information are considered and are shown in Table III. For all three algorithms the exhaustive information is used to evaluate all indicator auto- and cross-covariances and the indicator principal component autocovariances.

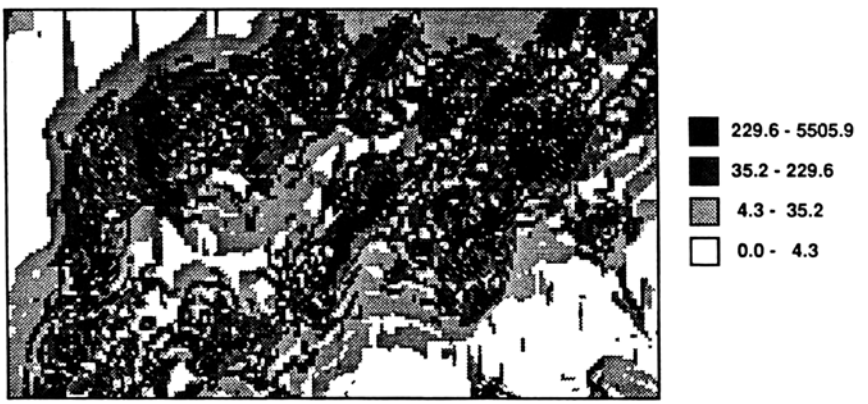


Fig. 1. Gray scale map of the Walker Lake data set.

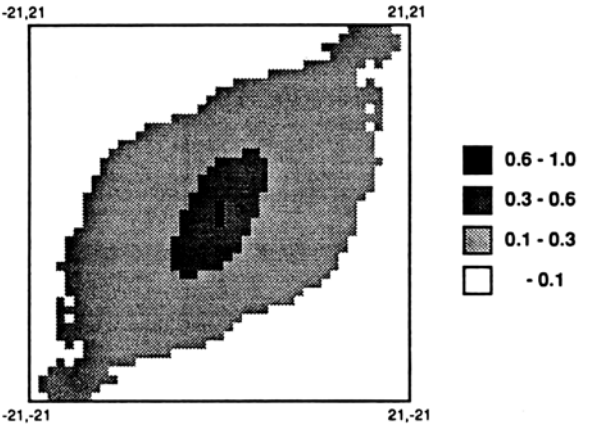


Fig. 2. Z-correlogram map for Walker Lake.

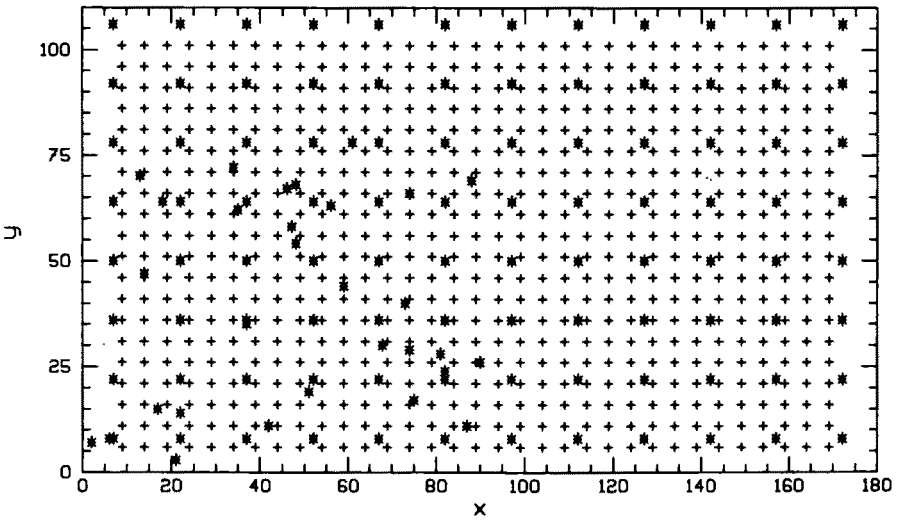


Fig. 3. Configuration of sample data and locations where conditional cdf models are built.

Table II. Walker Lake Statistics: 126 Samples[a]

No. of data	Mean	Variance	$q_{0.25}$	$q_{0.75}$	min	max	cv
126	208.535	119,263.1	9.600	295.595	0.0	1653.66	1.66

[a]These statistics have been obtained considering only the 126 data whose locations are shown in Fig. 3.

Table III. Nine Deciles (Walker Lake Data)[a]

q_p	p
0.057	0.1
1.636	0.2
7.957	0.3
17.898	0.4
35.225	0.5
74.396	0.6
156.072	0.7
323.342	0.8
593.247	0.9

[a] q_p corresponds to the p-quantile of the Walker Lake data, and p is the proportion of values below q_p.

Conditional Distributions for Walker Lake

The 126 sample locations represent a realistic situation. Clustering and undersampling are features frequently encountered in practice. Since these data were selected independently for the EPA study (Englund, 1990), they should not favor any of the three algorithms retained for testing.

Although statistical inference of representative covariance models is crucial to any geostatistical approach, this problem has not been addressed here. Instead, we allow CoIK, IPCK, and IK to use covariance models deduced from the exhaustive (19800 data) information. Indeed, the main goal of this section is to compare the performance of these three algorithms, under perfect inference.

Indicator Cokriging and Indicator Kriging

Forty-five indicator auto-crosscovariances were computed from the 19800 data constituting the exhaustive information (Fig. 1). In all the cases a nonergodic estimator (Isaaks and Srivastava, 1988) was considered. The indicator cross-covariances were assumed symmetric with respect to the cutoffs:

$$C_I(\mathbf{h}; k, k') = C_I(\mathbf{h}; k', k)$$

The symmetry assumption allows reduction of the $9^2 = 81$ indicator covariances and cross-covariances required by CoIK to only 45.

Figure 4a and b shows the indicator autocovariances for the first decile cutoff and for the median, respectively. Notice how anisotropy is more prevalent at the median cutoff. Figure 5a and b shows the indicator cross-covariances between the first and second cutoff and the median and sixth cutoff. In both cases, the magnitude of these cross-covariances are significant when compared to the magnitudes of the corresponding indicator autocovariances. CoIK does

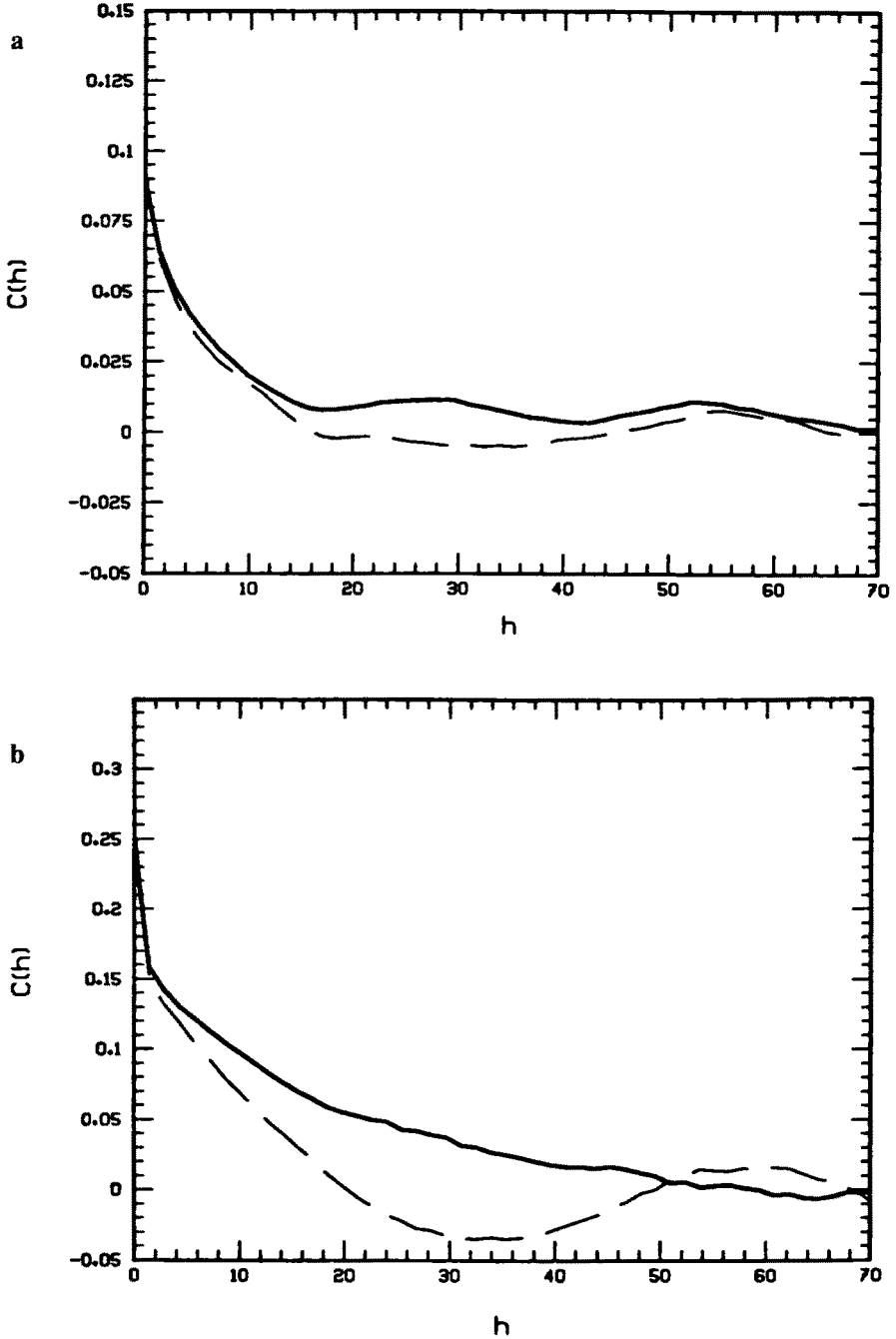


Fig. 4. Indicator covariances in the N45°E (solid line) and N45°W (dashed line) directions. This convention will be maintained hereafter. (a) First decile; (b) median.

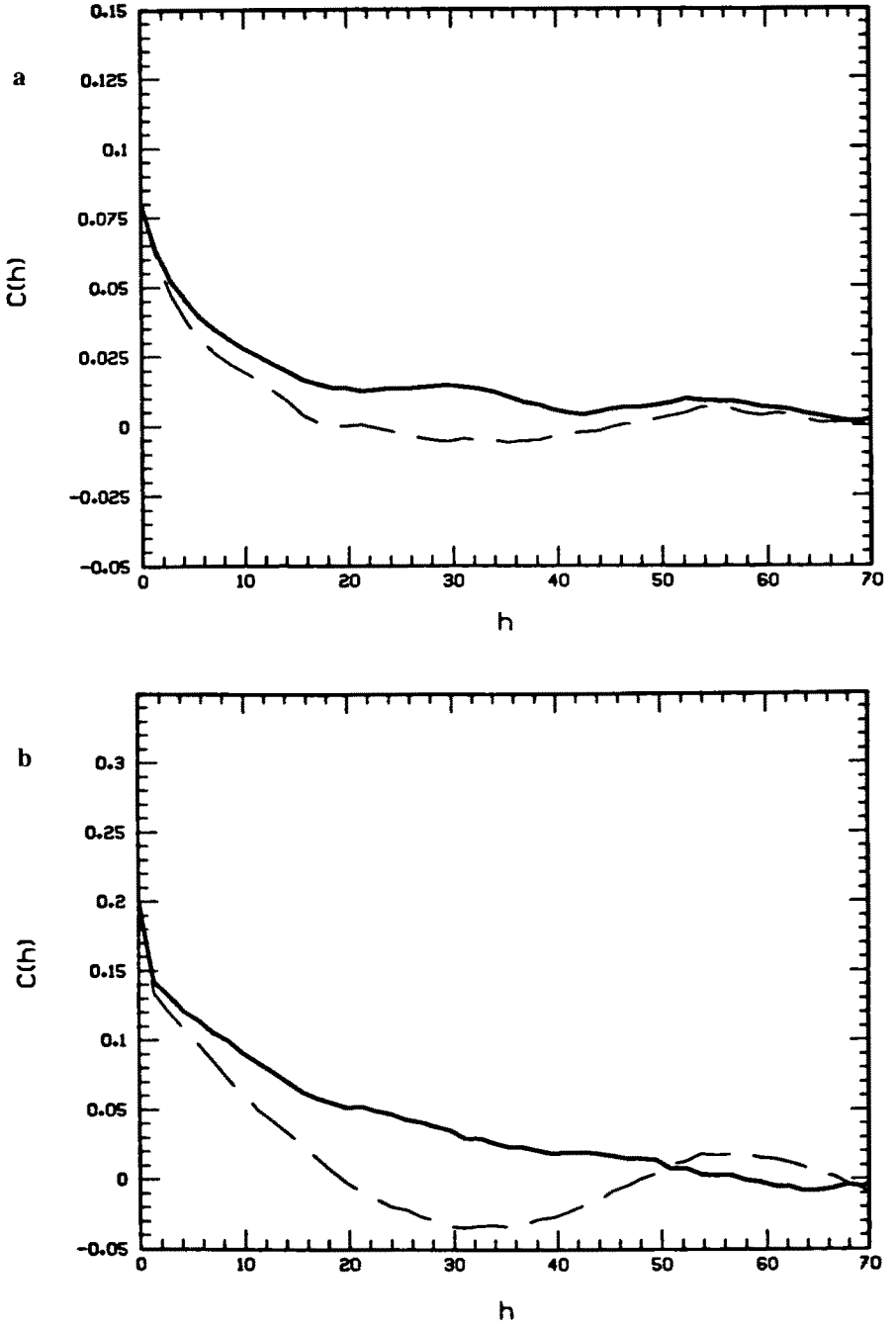


Fig. 5. Indicator crosscovariances. (a) First and second decile; (b) median and sixth decile.

account for such indicator cross-covariances in modeling the conditional distributions. The CoIK model, using the full discrete bivariate structural information, is now compared with the models provided by IPCK and IK.

Indicator Principal Component Kriging

The indicator principal components $Y_i(\mathbf{x})$ are obtained by expression (12). The indicator covariance matrix was orthogonalized at $\mathbf{h}_1$, with $\mathbf{h}_1$ being the unit lag in the N–S direction. In real practice, $\mathbf{h}_1$ could be chosen as zero or the smallest lag considered for the experimental covariances. The corresponding orthogonal matrix **A** is:

$$\mathbf{A} = \begin{bmatrix} -0.156 & 0.348 & -0.508 & 0.608 & -0.433 & 0.191 & -0.015 & 0.000 & 0.001 \\ -0.281 & 0.492 & -0.364 & -0.119 & 0.539 & -0.486 & 0.066 & -0.002 & -0.009 \\ -0.369 & 0.422 & 0.083 & -0.501 & -0.047 & 0.625 & -0.180 & 0.011 & 0.007 \\ -0.426 & 0.180 & 0.456 & -0.047 & -0.488 & -0.386 & 0.422 & -0.091 & 0.005 \\ -0.440 & -0.096 & 0.356 & 0.367 & 0.148 & -0.119 & -0.610 & 0.348 & 0.067 \\ -0.411 & -0.294 & 0.035 & 0.250 & 0.319 & 0.235 & 0.152 & -0.625 & -0.330 \\ -0.355 & -0.381 & -0.225 & -0.019 & 0.125 & 0.177 & 0.410 & 0.303 & 0.607 \\ -0.258 & -0.355 & -0.353 & -0.275 & -0.212 & -0.104 & -0.007 & 0.390 & -0.629 \\ -0.148 & -0.240 & -0.298 & -0.295 & -0.308 & -0.273 & -0.469 & -0.484 & 0.346 \end{bmatrix}$$

The columns of matrix **A** correspond to the weights $a_{k',k}$ in expression (13). The first indicator principal component explains about 74% of the indicators variability, while the last component has almost zero variance.

Figure 6a–c shows the two first and the last indicator principal component autocovariances. The spatial correlation range decreases with the higher order of the indicator principal components, and essentially vanishes for the 9th component. Transformation (12) appears to order the indicator principal components not only according to their variances, but also according to their spatial correlation.

Figure 7a and b shows the two most significant cross-covariances (in the sense that they are different from zero). In both cases, the magnitude of the cross-covariances appears to be negligible when compared to the corresponding autocovariances. Therefore, the constitutive approximation that the Y-covariance matrix is diagonal for all $\mathbf{h}$ is reasonably well-supported by the Walker Lake data. In other cases where the indicator covariance matrix was factorized for $\mathbf{h}_1$ close or equal to zero, similar results were obtained. For example, the cases $\mathbf{h}_1 = [0\ 0]^T$, $[1\ 0]^T$, $[1\ 1]^T$, $[0\ -1]^T$, yield an almost diagonal Y-covariance

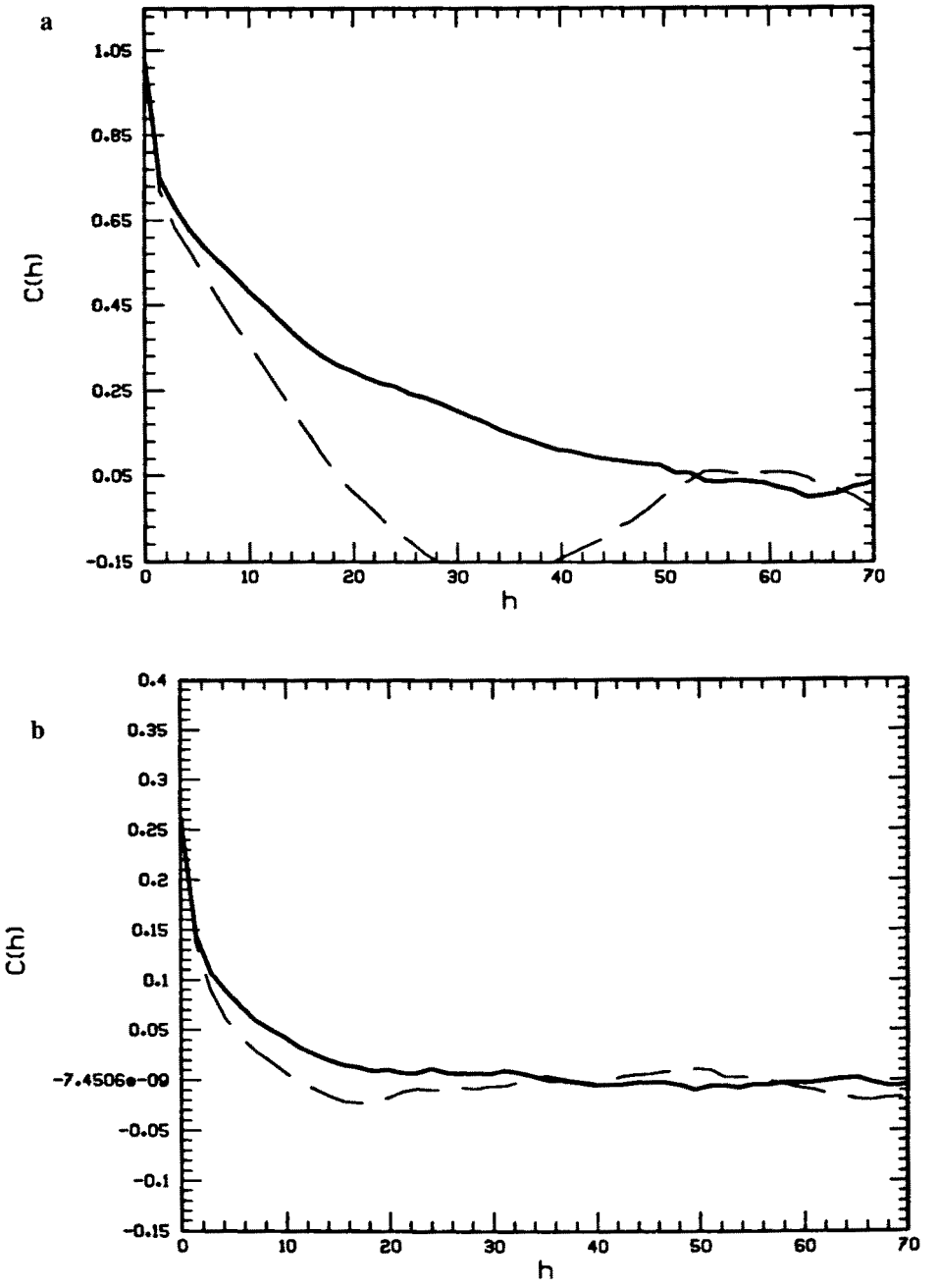


Fig. 6. Indicator principal component covariances. (a) First component; (b) second component; (c) ninth component.

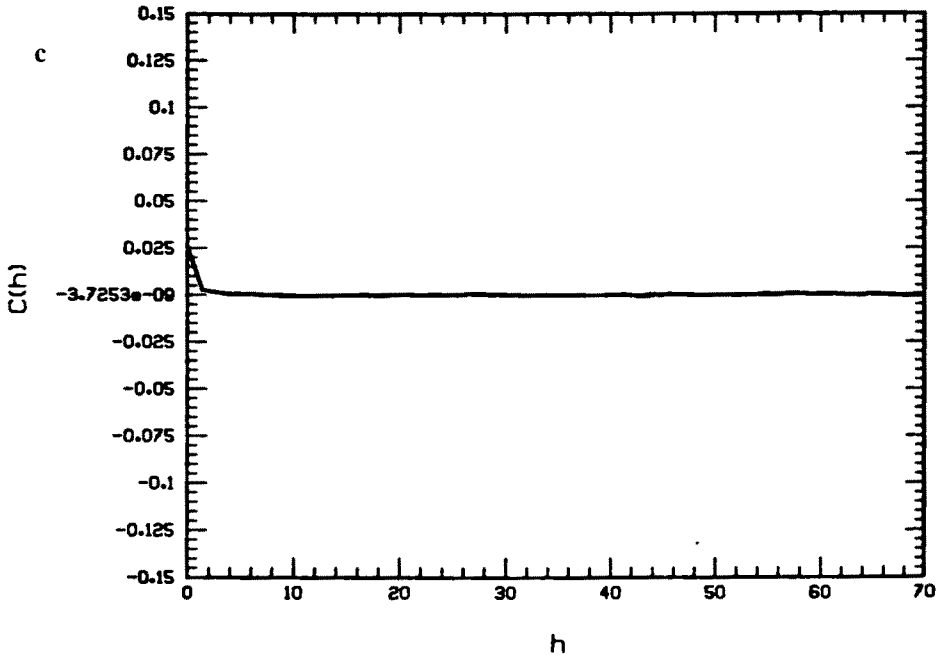


Fig. 6. Continued.

matrix for any $\mathbf{h}$ with negligible indicator principal component cross-covariances. This confirms the conjecture that orthogonalizing the indicator covariance matrix for small $\mathbf{h}_1$ yields approximately diagonal Y-covariance matrices.

For this exercise only, the three first indicator principal components were considered as significantly spatially correlated; the other six were modeled as pure nugget effect. Consequently, kriging is needed only for the first three components; thus, three equation systems are required. For the other six, a moving average estimate suffices without any need for solving further systems of equations.

A comparison of the conditional cdf's modeled from CoIK and IPCK is shown on the scattergrams of Figure 8a–c. Figure 9a–c shows the conditional cdf models obtained from CoIK and IK. The three models appear to be very similar. Apparently, introduction of the indicator cross-covariances in CoIK does not result in models of conditional cdf's much different from the models obtained using IPCK or IK. Table IV shows the average of the absolute differences between the models provided by IPCK-CoIK and IK-CoIK. Notice that, in general, IPCK approximates better CoIK; however, for the first decile IK does a better job than IPCK.

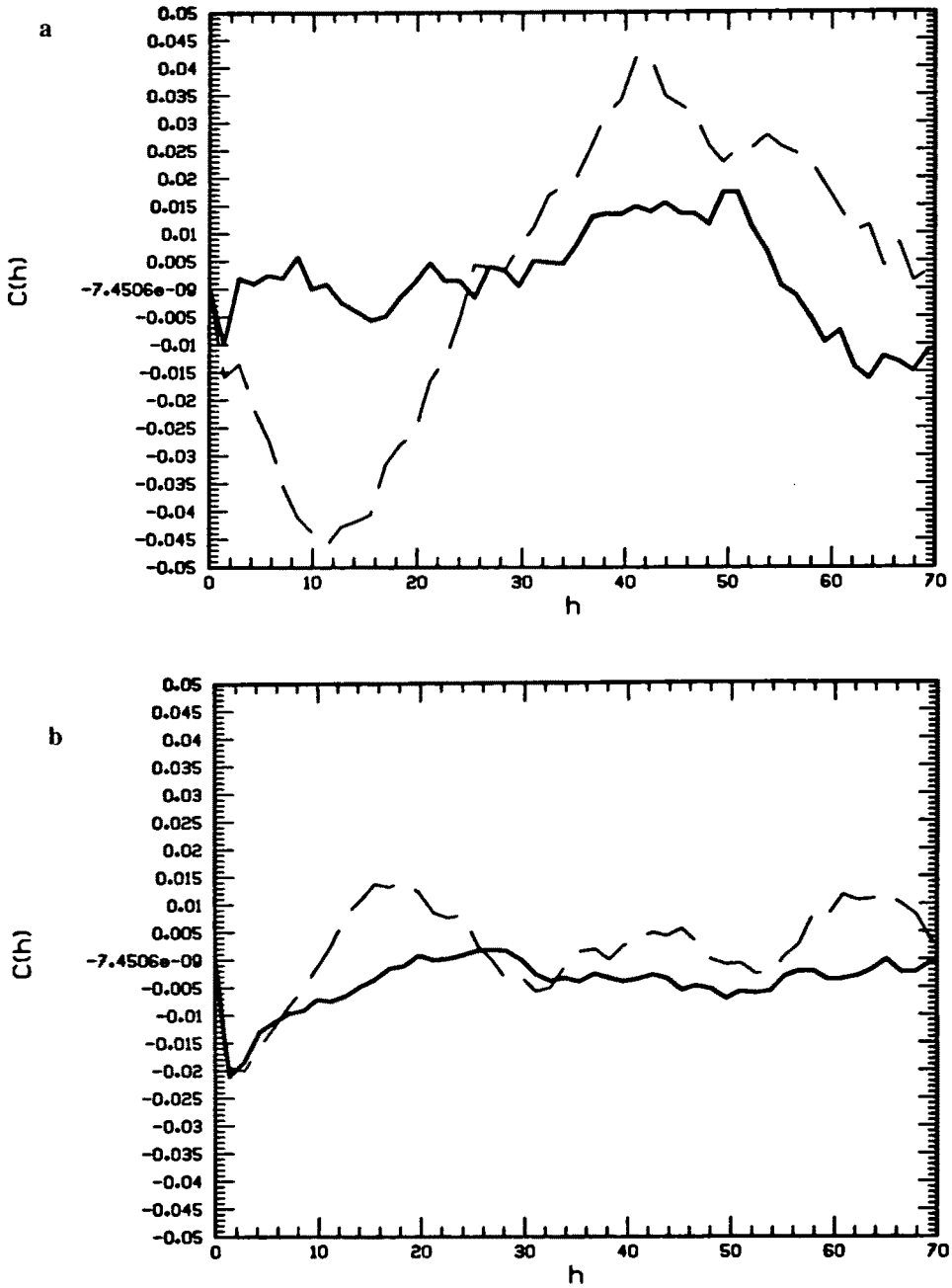


Fig. 7. Indicator principal component cross-covariances. (a) First and second component; (b) second and third components.

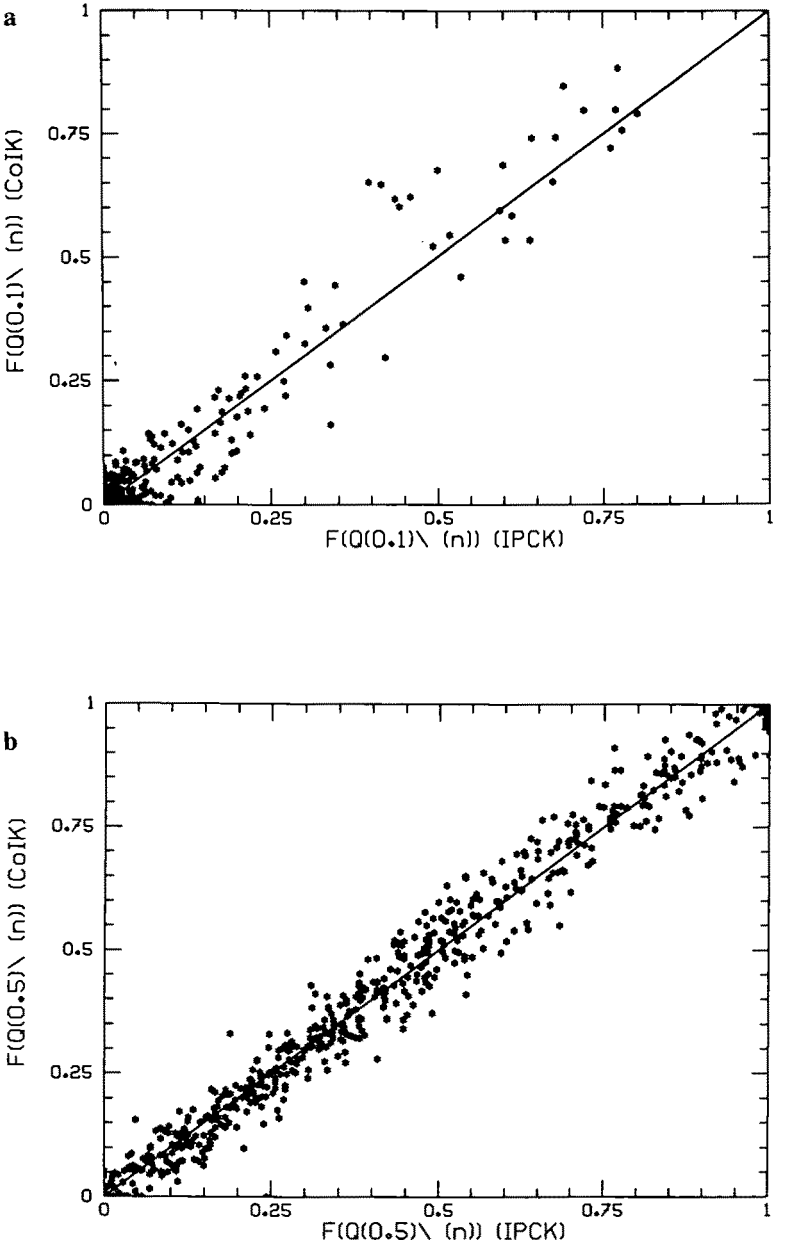


Fig. 8. Models of conditional distributions derived from CoIK and IPCK. The solid line refers to the 45° line. (a) First decile. The correlation coefficient is 0.97. (b) Median. The correlation coefficient is 0.99. (c) Ninth decile. The correlation coefficient is 0.99.

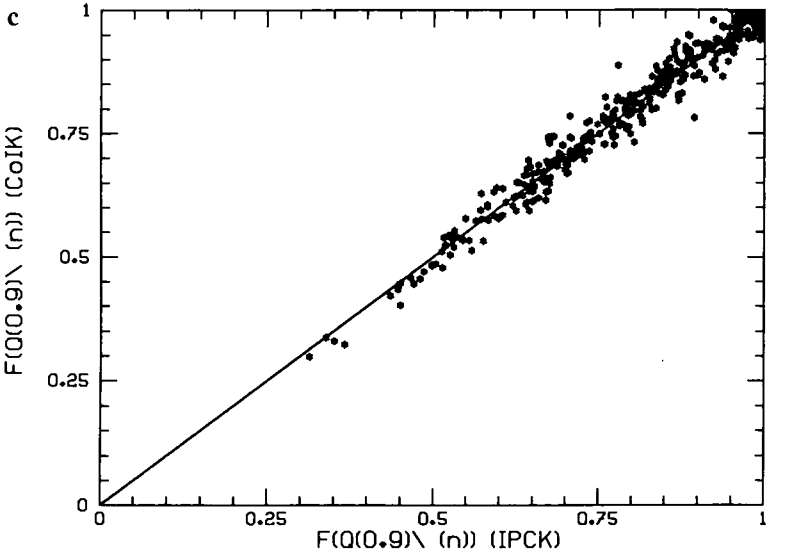


Fig. 8. Continued.

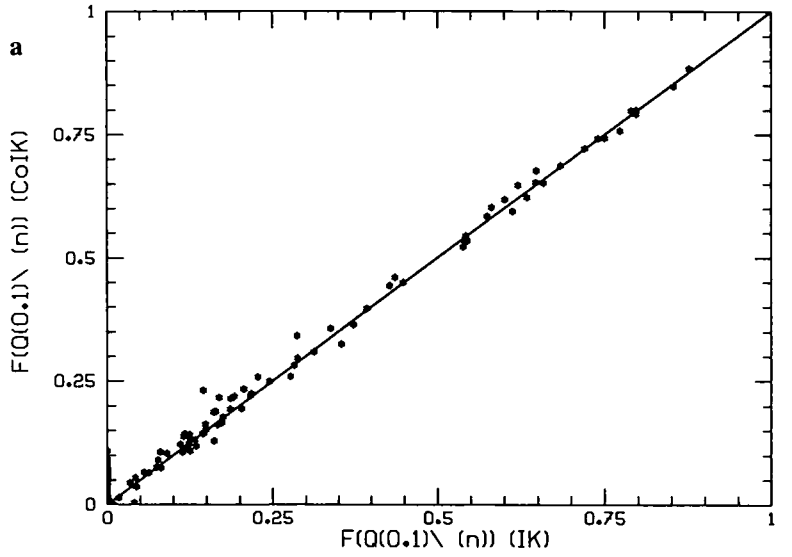
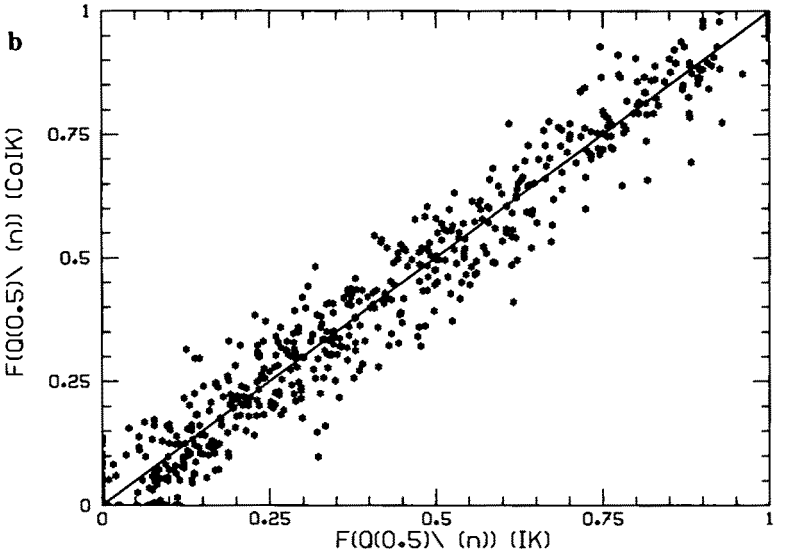


Fig. 9. Models of conditional distributions derived from CoIK and IK. The solid line is the 45° line. (a) First decile. The correlation coefficient is 0.99. (b) Median. The correlation coefficient is 0.98. (c) Ninth decile. The correlation coefficient is 0.97.



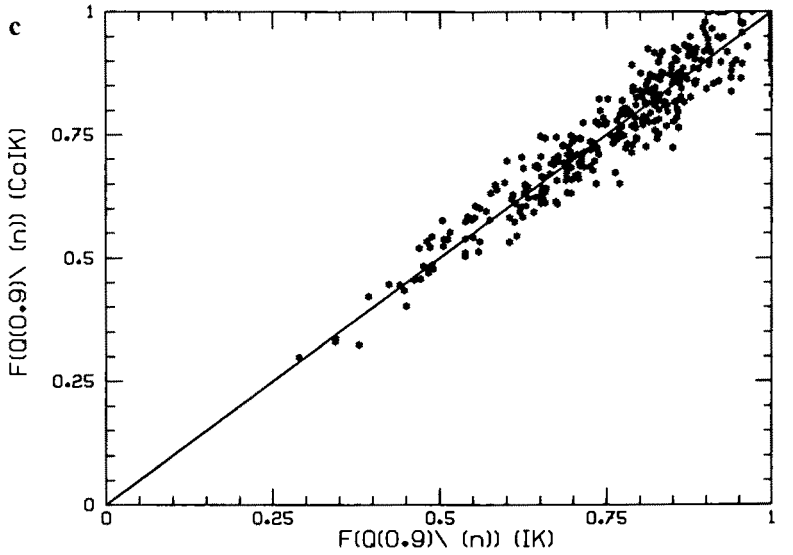


Fig. 9. Continued.

Table IV. Average Absolute Differences[a]

z_k	Average$\lvert CoIK - IPCK\rvert$	Average$\lvert CoIK - IK\rvert$
0.057	0.019	0.012
1.636	0.022	0.020
7.957	0.023	0.031
17.898	0.028	0.038
35.225	0.031	0.044
74.396	0.027	0.052
156.072	0.024	0.048
323.342	0.021	0.042
593.257	0.014	0.023

[a]The column Average$\lvert CoIK - I\rvert$ is the average of the absolute difference at 660 locations (Fig. 3) between the conditional cdf provided by CoIK and the model I, for the cutoff z_k. I corresponds either to the IPCK or the IK model.

Predicting Proportions and Optimal Estimates

From each conditional cdf, two symmetric quantile values, $q^*_{p_1}(\mathbf{x})$ and $q^*_{p_2}(\mathbf{x})$, are retrieved. The predicted probability for $Z(\mathbf{x})$ to be in the interval $[q^*_{p_1}(\mathbf{x}), q^*_{p_2}(\mathbf{x})]$ is then:

$$t = p_2 - p_1 \qquad \text{with } p_2 = 1 - p_1$$

Quantile values were interpolated linearly (i.e., within-class conditional distributions are considered as uniform). In a practical situation, better models for within-class distributions could stem from the corresponding data.

Table V shows the performance in terms of predicted proportion for the three techniques. In this table, t_I corresponds to the actual proportion of values $Z(\mathbf{x})$ falling in the interval $[q^*_{p_1}(\mathbf{x}), q^*_{p_2}(\mathbf{x})]$ obtained from the model I (CoIK, IPCK, and IK). If these models are reliable, then on average over all 660 estimated locations, the actual proportion t_I should be close to the predicted value t.

The scores are practically the same for proportions less than 0.5. For the 0.5 predicted proportion, the actual proportion derived from IK is closer to the predicted one than either CoIK or IPCK. However, for the predicted proportions 0.7 and 0.8, the actual proportions obtained from CoIK and IPCK are closer to the actual ones than IK. For these proportions, the IK model of conditional distribution overestimates the proportion of values $Z(\mathbf{x})$ by predicting a 70% and 80%, when actually 66% and 74% of the data are contained in the corresponding probability intervals.

For the 90% proportion, the three methods overestimate the proportion of $Z(\mathbf{x})$ values inside of the corresponding probability interval. However, IK overestimate the most.

Table V. Proportions Prediction (Walker Lake)[a]

t	t_{CoIK}	t_{IPCK}	t_{IK}
0.1	0.086	0.098	0.083
0.2	0.210	0.195	0.177
0.3	0.303	0.309	0.290
0.4	0.418	0.418	0.390
0.5	0.536	0.542	0.495
0.6	0.639	0.628	0.580
0.7	0.730	0.727	0.663
0.8	0.819	0.810	0.748
0.9	0.874	0.862	0.801

[a] t and t_l are the predicted and actual proportion, respectively. l corresponds to the model *CoIK*, *IPCK*, or *IK*.

Next, consider for loss function the mean absolute deviation:

$$L(Z(\mathbf{x}) - z^*(\mathbf{x})) = |Z(\mathbf{x}) - z^*(\mathbf{x})|$$

The value that minimizes the corresponding expected loss (32) is the median or $q^*_{0.5}(\cdot)$. That median value is taken from each of the three conditional distribution models. In both cases, a within-class linear interpolation was used to obtain the median.

Two different criteria (Zhu and Journel, 1990) are considered to compare the three different optimal estimates. The first one concentrates on conditional bias. Selection is usually done by applying some cutoff on the optimal estimate (here the conditional median) and the selection loss of accuracy can be measured as:

$$E[Z^*(\mathbf{x}) - Z(\mathbf{x})|Z^*(\mathbf{x}) < z_{k'}] \tag{35}$$

with $z_{k'}$ being a selected cutoff. This expression (35) measures the average error given that the estimate is less than a particular cutoff. If selection is accurate, this measure is close to zero. Table VI shows the (loss of) accuracy scores considering nine cutoffs. The scores are similar and there appears to be no clear advantage of one method over the others.

The second criterion corresponds to the efficiency of a selection performed on the estimated value:

$$E\{Z(\mathbf{x})|Z(\mathbf{x}) > z_{k'}\} - E\{Z(\mathbf{x})|Z^*(\mathbf{x}) > z_{k'}\} \tag{36}$$

This index compares the average of the $Z(\mathbf{x})$ under perfect selection, $E\{Z(\mathbf{x})|Z(\mathbf{x}) > z_{k'}\}$, with the one obtained by doing the selection using the estimate $E\{Z(\mathbf{x})|Z^*(\mathbf{x}) > z_{k'}\}$. Table VII shows the resulting (loss of) efficiency scores. Again, the scores are similar, indicating that estimates derived from CoIK, IPCK, and IK provide about the same efficiency in selection.

Table VI. Accuracy Performance (Walker Lake Data)[a]

Cutoff	*CoIK*	*IPCK*	*IK*
$q_{0.2}$	−2.30	−2.13	−2.31
$q_{0.3}$	−2.05	−1.89	−2.31
$q_{0.4}$	−29.66	−34.63	−34.58
$q_{0.5}$	−56.07	−58.70	−57.44
$q_{0.6}$	−68.48	−68.10	−77.26
$q_{0.7}$	−73.62	−75.90	−77.09
$q_{0.8}$	−95.25	−93.71	−101.41

[a]The scores have been obtained by using expression (35). Best is zero.

Table VII. Efficiency Performance (Walker Lake Data)[a]

Cutoff	*CoIK*	*IPCK*	*IK*
$q_{0.2}$	19.32	15.85	18.57
$q_{0.3}$	36.52	37.27	38.69
$q_{0.4}$	45.01	49.19	47.12
$q_{0.5}$	71.24	78.12	70.84
$q_{0.6}$	100.76	101.85	106.20
$q_{0.7}$	76.03	66.77	64.68
$q_{0.8}$	228.41	215.08	238.02

[a]Expression (36) has been used to obtain these scores. Best is zero.

The last two criteria indicate no advantage of CoIK over IPCK or IK, or vice versa. For predicting proportions, CoIK and IPCK perform better than IK.

CONCLUSION

IPCK is a nonparametric technique for modeling the conditional distribution $F(\mathbf{x}; \mathbf{z}|\{n\})$. This model is obtained by kriging a limited number of principal components obtained from a linear transformation of the indicators $\mathbf{I}(\mathbf{x}; \mathbf{z})$. A comparison of the models of conditional distribution derived from CoIK, IPCK, and IK has shown only small differences between the three models. If CoIK is considered the reference model, then, IPCK approximates it better than IK. Furthermore, because indicator principal components are continuous variables as opposed to binary indicators, inference of their covariances or variograms should be easier.

The Walker Lake data set has shown that under a practical bivariate distribution, not biGaussian, IPCK performs as well as CoIK or IK. In addition,

because the higher indicator principal components are uncorrelated, there is no need to krige them. This property makes IPCK faster than IK and much faster than CoIK, since a smaller number of systems needs to be solved and a smaller number of indicator principal component variograms needs to be modeled.

IPCK is an approximation to CoIK since, in general, it considers only part of the full discrete bivariate distribution. However, in the biGaussian case, and also for the real data set analyzed in this paper, there is no significant loss of information by using IPCK.

There appears to be no major practical implementation problems to obtain the model $F_{\text{IPCK}}(\mathbf{x}; \mathbf{z}|\{n\})$. Any ordinary kriging program suffices to develop it, with indicator principal component data replacing indicator data. The transformation matrix $\mathbf{A}^T$ is obtained from any public-domain singular value decomposition or spectral decomposition subroutine.

ACKNOWLEDGMENTS

The first author is grateful to Consejo Nacional de Ciencia y Tecnología (México) for financial support through a scholarship. Although additional support was provided by the U.S. Environmental Protection Agency through assistance agreement CR-814899-01-0 to Leland Stanford Jr. University, this research has not been subjected to Agency review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

REFERENCES

Anderson, T. W., 1984, Multivariate Statistical Analysis: Wiley, New York, 675 p.

Borgman, L., and Frahme, R. B., 1976, Multivariate Properties of Bentonite in Northeastern Wyoming, Guarascio, M. et al. (Eds.), Advanced Geostatistics in the Mining Industry: D. Reidel, Dordrecht, p. 381–390.

Davis, B. M., and Greenes, K., 1983, Estimation Using Spatially Distributed Multivariate Data: An Example with Coal Quality Data: Math. Geol., v. 15, n. 3, p. 287–300.

Englund, E. J., 1990, A Variance of Geostatisticians: Math. Geol, v. 22, n. 4, p. 417–456.

Golub, G., and Van Loan Ch., 1983, Matrix Computations: The John Hopkins University Press, 476 p.

Guibal, D., and Remacre, A., 1984, Local Estimation of Recoverable Reserves: Comparing Various Methods with the Reality on a Porphiry Copper Deposit, in Verly, G., et al. (Eds.), Geostatistics for Natural Resources Characterization: D. Reidel, Dordrecht, p. 435–448.

Howarth, R. J. (Ed.), 1983, Statistics and Data Analysis in Geochemical Prospecting: Elsevier, New York, 437 p.

Isaaks, E., and Srivastava, M. R., 1988, Spatial Continuity Measures for Probabilistic and Deterministic Geostatistics: Math. Geol., v. 20, n. 4, p. 313–341.

Isaaks, E., and Srivastava, M. R., 1989, An Introduction to Applied Geostatistics: Oxford University Press, New York, 561 p.

Journel, A. G., 1977, Kriging in Terms of Projections: Math. Geol., v. 9, n. 6, p. 563–586.

Journel, A. G., 1983, Nonparametric Estimation of Spatial Distributions: Math. Geol., v. 15, n. 3, p. 445-468.

Journel, A. G., 1984, mAD and Conditional Quantile Estimators, in Verly, G., et al. (Eds), Geostatistics for Natural Resources Characterization: D. Reidel, Dordrecht, p. 261-270.

Journel, A. G., 1989, Fundamentals of Geostatistics in Five Lessons: Short Course in Geology, V. 8: American Geophysical Union, Washington, D. C., 40 p.

Marcotte, D., and David, M., 1985, The BiGaussian Approach: A Simple Method for Recovery Estimation: Math. Geol., v. 17, n. 6, p. 625-644.

Matheron, G., 1976, A Simple Substitute to Conditional Expectation: The Disjunctive Kriging, in Guarascio, M., et al. (Eds.), Advanced Geostatistics in the Mining Industry: NATO ASI, D. Reidel, Dordrecht, p. 221-236.

Myers, D. E., 1982, Matrix Formulation of Co-Kriging: Math. Geol. v. 17, n. 6, p. 625-644.

Rivoirard, J., 1989, Models with Orthogonal Indicator Residuals, in Armstrong, M. (Ed.), Geostatistics: Kluwer Academic, Dordrecht, p. 91-108.

Sullivan, J., 1984, Conditional Recovery Estimation Through Probability Kriging, in Verly, G., et al. (Eds.), Geostatistics for Natural Resources Characterization: D. Reidel, Dordrecht, p. 365-384.

Suro-Pérez, V., 1988, Indicator Kriging Based on Principal Component Analysis: M. Sc. thesis, Branner Earth Sciences Library, Stanford University, 168 p.

Suro-Pérez, V. and Journel, A. G., 1990, Indicator Principal Component Kriging: Two Case Studies: Stanford Center for Reservoir Forecasting, Stanford University, unpublished Annual Report.

Verly, G., 1983, The Multigaussian Approach and Its Applications to the Estimation of Local Reserves: Math. Geol., v. 15, n. 2, p. 259-286.

Wilkinson, J. H., 1965, The Algebraic Eigenvalue Problem, Oxford University Press, 662 p.

Zhu, H. and Journel, A. G., 1990, Indicator-conditioned Estimator (ICE): SME paper #89-151, in Mining Engineering Transactions, v. 286.