WATER RESOURCES RESEARCH, VOL. 22, NO. 5, PAGES 615–621, MAY 1986

Disjunctive Kriging
1. Overview of Estimation and Conditional Probability

S. R. YATES AND A. W. WARRICK

Department of Soil and Water Science, The University of Arizona, Tucson

D. E. MYERS

Department of Mathematics, The University of Arizona, Tucson

The disjunctive kriging method described in this paper produces a nonlinear unbiased estimator with the characteristic minimum variance of errors. Disjunctive kriging is as good, or otherwise better than linear estimators in the sense of reduced kriging variance and exactness of estimation. It does not suffer from the difficulties associated with computing the conditional expectation and can be thought of as its estimator. Disjunctive kriging also provides an estimate of the conditional probability that a random variable located at a point or averaged over a block in two-dimensional space is above some specified cutoff or tolerance level and this can be written in terms of the probability distribution or the density function. The method has important implications in aiding management decisions by providing a quantitative input (which is not readily obtained from the linear kriging estimators), based on the available data, which is the best nonlinear unbiased estimator short of the conditional expectation. A major disadvantage in using disjunctive kriging is the increased computational time. This, however, is mitigated by increased information about the estimate.

INTRODUCTION

Linear kriging methods such as simple, ordinary and universal kriging have been used in soil and water science to estimate a spatially distributed random variable at an unsampled location. The final product of such analyses is often a contour map showing the spatial distribution of the property of interest. Many examples occur in the literature. For example, *Burgess and Webster* [1980*a*] applied punctual kriging to three data sets. The first two data sets, which were found to be isotropic, were the sodium content and the thickness of cover loam. The third data set, which was shown to be anisotropic, was the stone content. In two following papers [*Burgess and Webster*, 1980*b*; *Webster and Burgess*, 1980] they demonstrated the use of block and universal kriging, respectively. *Warrick et al.* [1986] give an example of kriging for the natural logarithm of the electrical conductivity for a Typic Haplargid [see *Al-Sanabani*, 1982] which was randomly sampled in a 10 ha field in southwest Arizona. *Vieira et al.* [1981] obtained 1280 measurements of the limiting infiltration rate over a 0.9 ha field and showed the correlation between the kriged and actual values for various number of sample values used in the estimation process. They found that approximately 128 samples was sufficient to provide about the same amount of information as the original 1280. *Vauclin et al.* [1983] used ordinary kriging to estimate the spatial distribution of the available water content and sand content of a Tunisian field. They also used cokriging to improve the estimation of the available water content by including the sand content in the estimation. The improvement results from the added information about the spatial distribution of sand and enters the problem via the cross-correlation function. These studies represent but a small sample of the works which have used linear kriging estimators in the analysis of the spatial variability of a physical property.

Alternatively, geostatistics can be used to help in planning and/or decision making. A few examples include attempts to answer questions about the minimum number of samples necessary to produce a certain level of confidence or accuracy in the estimate [*Burgess et al.*, 1981; *Vieira et al.*, 1981]; methods directed toward answering questions about the number of samples and the placement of the samples in space such that the estimation of the variogram is optimized [*Russo*, 1984] and how best to deal with the effects of scale [*Burrough*, 1983*a*, *b*].

To our knowledge, no one in the soil and water sciences has attempted to use disjunctive kriging as a management or decision-making tool. Through disjunctive kriging one can obtain an estimate of the conditional probability (CP) that a measured indicator variable goes above some prescribed tolerance level (see Figure 1). With respect to hazardous waste, the tolerance level could be a threshold level where leakage of a contained material near an impoundment becomes toxic to humans. For agriculture, it could be a situation where the level of salts in field increases beyond the tolerance level of the plants present. Combining a known tolerance level with disjunctive kriging's (DK's) ability to estimate the conditional probability (that the indicator variable is greater than the tolerance level) offers a method where management decisions (or action) may be based on quantitative information. The two examples given above are but a few of numerous possibilities. Therefore we define the unit on which management decisions are based as the "management unit," which is the smallest unit area that can be managed separately from the surrounding units.

In terms of the management unit, the problem becomes a matter of determining when the indicator variable has become larger than the allowed tolerance level. To complicate matters, sampling may be on one support (i.e., punctual) whereas management decisions may be based on another support such as a block. An alternative solution would be to extensively sample

Paper number 5W4133.
0043-1397/86/005W-4133$05.00

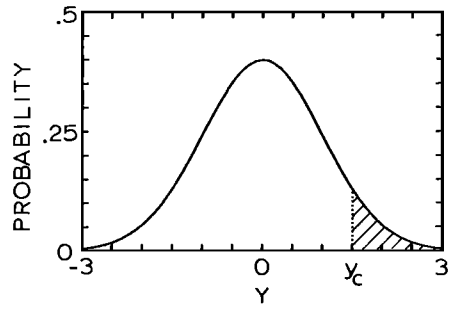

Fig. 1. Idealized conditional probability density function in terms of the transformed variable. The shaded area is the conditional probability that the indicator variable is above y_c.

the indicator variable over the entire region and attempt to construct the probability density functions (PDF) for each management unit from the samples that fall into each of the units. Then one could determine the probability that the block is above the cutoff level from the PDF. This is not a practical method for routine analysis. A simpler and less costly approach would be to estimate the PDF using traditional sampling and the DK method as an estimator. In either case the management decision is based on three inputs, the tolerance level and the probability level that will cause an action (which are assumed to be prescribed at the onset) and the CP calculated for a given data set (i.e., at a given time). With these three pieces of information one can make a management decision which is quantitatively based. Also, since the tolerance level and the probability level that will initiate an action are generally preset and constant (or known for all stages of interest such as growth of a plant) only one input, the CP, needs to be calculated.

Compared to a nonlinear estimator the linear kriging estimator is a compromise. The minimum variance unbiased estimator of a random variable Y in terms of random variable $X_1, X_2, \ldots, X_n$ is the conditional expectation of Y given $X_1, X_2, \ldots, X_n$. To compute this conditional expectation knowledge of the joint density of $Y, X_1, X_2, \ldots, X_n$ is required. In general, this density is not estimable from data except possibly in the multivariate normal case, where the conditional expectation is a linear function. This is part of the motivation that led to the usual (linear) kriging estimator. The conditional expectation is a function of n variables, that is,

$$Y^* = g(X_1, X_2, \ldots, X_n) \tag{1}$$

The linear kriging estimator has a special form

$$Y_K^* = \sum_{i=1}^{n} \lambda_i X_i \tag{2}$$

which could be viewed as a form of the more general estimator

$$Y_{DK}^* = \sum_{i=1}^{n} f_i(X_i) \tag{3}$$

where each f_i is a function of one X variable only. In using linear kriging, each f_i is a linear function and only the coefficients need to be determined. For each function, f_i, in the more general case, as opposed to finding coefficients, the appropriate functions must be found. This requires a slightly stronger hypothesis but has the advantage that the nonlinear estimator in (3) will have (in general) a smaller estimation variance than (2) and will be easier to determine than the most general form given by (1).

The purpose of the first paper of this series is to review the theoretical basis of the (DK) method as it relates to estimating the value of a random function at an unsampled location and obtaining the conditional probability that the value of a random function at a point will be above a given cutoff level. Although the information is available in scattered materials, to our knowledge a complete derivation using consistent terminology is not readily available in one location for immediate application to soil-water problems.

The second section of this paper discusses the conditional probability which makes use of the DK estimator directly but is important in its own right since, for many applications, information about the conditional probability may be a more important consideration than only obtaining an estimate.

The second paper in this series [*Yates et al.*, 1986] illustrates the DK method with examples and includes a comparison between DK and ordinary kriging (OK) with respect to the accuracy of the estimators and the computational requirements.

THEORY

Hermite Polynomials

Before describing the DK method, it is helpful to describe some of the properties of Hermite polynomials. Hermite polynomials are functions orthogonal with respect to the weighting function $\exp[-y^2/2]$ on the interval $[-\infty, \infty]$. Many functions can be represented by an infinite series of Hermite polynomials of the form

$$\phi(y) = \sum_{k=0}^{\infty} C_k H_k(y) \tag{4}$$

where the C_k's can be found using the orthogonality properties.

A sufficient condition for a function, $\phi(y)$, to have a Hermite representation (which converges mean-square) is that

$$\int_{-\infty}^{\infty} [\phi(y)]^2 \exp(-y^2/2)\, dy \tag{5}$$

be finite. In fact it is this condition which allows us to compute the coefficients, C_k.

A Hermite polynomial of order k is defined by Rodriques' formula [*Abramowitz and Stegun*, 1965] as

$$H_k(y) = (-1)^k \exp[y^2/2] d^k(\exp[-y^2/2])/dy^k \tag{6}$$

and may be evaluated by the recursive relationship

$$H_{k+1}(y) = yH_k(y) - kH_{k-1}(y) \tag{7}$$

where $H_0 = 1$ and $H_1 = y$.

The orthogonality relationship for nonnormalized Hermite polynomials is

$$\begin{aligned} &\int_{-\infty}^{\infty} H_k(y)H_{k'}(y) \exp[-y^2/2]\, dy = 0 \qquad k \neq k' \\ &\int_{-\infty}^{\infty} H_k(y)H_{k'}(y) \exp[-y^2/2]\, dy = (k!)(2\pi)^{1/2} \qquad k = k' \end{aligned} \tag{8}$$

Using the orthogonality relationship, the coefficients C_k can be determined by

$$C_k = \left[\int_{-\infty}^{\infty} \phi(y) H_k(y) \exp[-y^2/2]\, dy\right] \Big/ (k!(2\pi)^{1/2}) \quad (9)$$

where $\phi(y)$ is the function represented by the infinite series of Hermite polynomials. Since for DK, $\phi(y)$ is not generally known, we can not evaluate this integral analytically. Standard numerical integration techniques would require knowledge of $\phi(y_i)$ for a large number of points $y_1, y_2, \ldots, y_n$. However, Hermite integration [*Abramowitz and Stegun*, 1965] provides an alternative method of integration which uses only a few special points

$$C_k = 1/(k!(2\pi)^{1/2}) \sum_{i=1}^{J} w_i \phi(v_i) H_k(v_i) \exp[-v_i^2/2] \quad (10)$$

The abscissa's, v_i, and weight factors, w_i, for Hermite integration are given by *Abramowitz and Stegun* [1965].

The Mean and Variance

If $\phi(X)$ is a function of a standard normal random variable X, then the mean and variance of $Z = \phi(X)$ given by (4) can be obtained from the coefficients C_k as

$$\mu_z = E[\phi(X)] = C_0 \quad (11)$$

and

$$\sigma_z^2 = \text{Var}[\phi(X)] = \sum_{k=1}^{\infty} k! C_k^2 \quad (12)$$

where μ_z is the mean and σ_z^2 is the variance of Z. Note that the series in (12) begins at $k = 1$.

TABLE 1. C_k's for Exactly Normal and Lognormal Distributions

k	Normal C_k	Lognormal C_k
0	μ	$e^{\mu}\left[1 + \frac{\sigma^2}{2} + \frac{\sigma^4}{8} + \frac{\sigma^6}{48} + \frac{\sigma^8}{384} + \frac{\sigma^{10}}{3840} + \cdots\right]$
1	σ	$e^{\mu}\left[\sigma + \frac{\sigma^3}{2} + \frac{\sigma^5}{8} + \frac{\sigma^7}{48} + \frac{\sigma^9}{384} + \cdots\right]$
2	0	$e^{\mu}\left[\frac{\sigma^2}{2} + \frac{\sigma^4}{4} + \frac{\sigma^6}{16} + \frac{\sigma^8}{96} + \frac{\sigma^{10}}{768} + \cdots\right]$
3	0	$e^{\mu}\left[\frac{\sigma^3}{6} + \frac{\sigma^5}{12} + \frac{\sigma^7}{48} + \frac{\sigma^9}{288} + \cdots\right]$
4	0	$e^{\mu}\left[\frac{\sigma^4}{24} + \frac{\sigma^6}{48} + \frac{\sigma^8}{192} + \frac{\sigma^{10}}{1152} + \cdots\right]$
5	0	$e^{\mu}\left[\frac{\sigma^5}{120} + \frac{\sigma^7}{240} + \frac{\sigma^9}{960} + \cdots\right]$
6	0	$e^{\mu}\left[\frac{\sigma^6}{720} + \frac{\sigma^8}{1440} + \frac{\sigma^{10}}{5760} + \cdots\right]$
7	0	$e^{\mu}\left[\frac{\sigma^7}{7!} + \frac{\sigma^9}{10080} + \cdots\right]$
8	0	$e^{\mu}\left[\frac{\sigma^8}{8!} + \frac{\sigma^{10}}{80640} + \cdots\right]$
9	0	$e^{\mu}\left[\frac{\sigma^9}{9!} + \cdots\right]$

Bivariate Densities

Bivariate densities can be expressed in the form

$$f(x, y) = g(x, y) f_1(x) f_2(y) \quad (13)$$

where f_1 and f_2 are the marginal densities. In particular, if X and Y are bivariate normal with zero mean and unit variance then the joint density is given by

$$f(x, y) = \sum_{k=0}^{\infty} (\rho_{XY})^k H_k(x) H_k(y) g(x) g(y)/k! \quad (14)$$

where ρ_{XY} is the correlation coefficient and g is the standard normal density function. Given in this form it is easy to compute the conditional expectation of a function, $\phi(X)$, given by (4)

$$E[\phi(X)|Y] = \sum_{k=0}^{\infty} (\rho_{XY})^k C_k H_k(Y) \quad (15)$$

Recall, however, that bivariate normality for each pair of $n + 1$ variables does not imply $n + 1$ multivariate normality. In the particular case where $Y(x)$ is a stationary random function (i.e., each pair $Y(x_i)$, $Y(x_j)$ has the same bivariate density as $Y(x_i + h)$, $Y(x_j + h)$), and $Y(x)$ is bivariate normal then the correlation coefficient, ρ_{ij}, depends only on the particular pair. Generally, the correlation coefficient is given via the covariance function.

Normal or Lognormal Random Variables

As special cases we could consider $\phi(Y)$ to be normally or lognormally distributed. For such cases the C_k's for the Hermite expansion can be calculated from the mean and variance of $\phi(Y)$. The C_k's for these cases are given in Table 1. As an alternative to DK for the lognormal case we could use lognormal kriging [*Journel*, 1980; *Journel and Huijbregts*, 1978].

DISJUNCTIVE KRIGING

Disjunctive kriging represents a form of nonlinear kriging (i.e., results in a nonlinear estimator) which in general offers an improvement over linear kriging methods, yet doesn't require knowledge of the $n + 1$ joint probability distributions necessary for the conditional expectation. *Matheron* [1976] proposed the disjunctive kriging method as a simplified alternative which only requires that the bivariate distributions for the $n + 1$ variables be known. When a random variable is uni- and bivariate normally or lognormally distributed, the linear or lognormal kriging estimator for a known mean is identical to the DK estimator [*Rendu*, 1980; *Journel and Huijbregts*, 1978]. By adding an additional assumption that the distribution is multivariate normal, then the linear kriging estimator and the DK estimator are the same as the conditional expectation.

One assumption implicit in the DK method described here is that a gaussian transform, $\phi(Y)$, exists, is unique and invertible. Furthermore, this gaussian transform produces a random function which is univariate normal with mean zero and unit variance from a random variable of arbitrary distribution. Another assumption is that the random function produced by the transform also has a bivariate normal distribution for each pair of locations. These assumptions are necessary only in that they allow the conditional expectation, $E[\phi[Y(x_0)]|Y(x_i)]$, to be written in terms of the correlation function.

The transformation is applied to $Z(x)$ for three reasons: (1) the normal distribution is convenient to work with in defining

relationships such as the conditional expectation, (2) a transformation always exists between the arbitrary $Z(x)$ and the normally distributed $Y(x)$ [*Kim et al.*, 1977]; therefore the disjunctive kriging method is generalized in the sense that only one formulation is necessary for all possible distributions of the random variable $Z(x)$, and (3) it eliminates the need for a limiting assumption (i.e., that the distribution must be specified such as is the case with lognormal kriging).

Disjunctive Kriging Estimator

Consider a second-order stationary random function, $Z(x)$, which is sampled on a point support at n locations: $x_1, x_2, \ldots, x_n$ where x represents a vector in two-dimensional space. Applying the gaussian transform to each $Z(x_i)$ produces a standard normal random variable $Y(x_i)$. The disjunctive kriging estimator is made up of a sum of nonlinear functions where each function depends on only one normalized sample value, $Y(x_i)$,

$$Z_{DK}^*(x_0) = \sum_{i=1}^{n} f_i[Y(x_i)] = \sum_{i=1}^{n} \sum_{k=0}^{\infty} f_{ik} H_k[Y(x_i)] \qquad (16)$$

where n is the number of samples, $f_i[Y(x_i)]$ is a function to be determined and expressed on the right-hand side of (16) as a series of Hermite polynomials where f_{ik} is a constant which depends on i and k. (Note this is only one place where Hermite polynomials are used in DK).

The two conditions imposed to produce the "best" estimator are unbiasedness and minimum variance of errors:

$$E[Z(x_0) - Z_{DK}^*(x_0)] = 0 \qquad (17)$$

and

$$\text{Var}\,[Z(x_0) - Z_{DK}^*(x_0)] = \min \qquad (18)$$

or

$$E[Z(x_0) - Z_{DK}^*(x_0)]^2 = \min \qquad (19)$$

The minimum for (18) and (19) occurs when $Z(x_0) - Z_{DK}^*(x_0)$ is perpendicular to any function, $f[Z(x_i)]$, in the hyper-plane defined by the measurable functions $f_i[Z(x_i)]$ [see *Journel and Huijbregts*, 1978, p. 568] given in (16). Therefore, using the perpendicular projection, namely that the vectors $Z(x_0) - Z_{DK}^*(x_0)$ and $f[Z(x_i)]$ are orthogonal, we can then write

$$E[\{Z(x_0) - Z_{DK}^*(x_0)\} f[Z(x_i)]] = 0 \qquad (20)$$

or

$$E[Z(x_0) f[Z(x_i)]] = E[Z_{DK}^*(x_0) f[Z(x_i)]] \qquad (21)$$

In terms of the conditional expectation (21) can be written as

$$E[Z(x_0)|Z(x_j)] = E[Z_{DK}^*(x_0)|Z(x_j)] \qquad j = 1, 2, \ldots, n \qquad (22)$$

The next step in the DK process is to determine the unknowns in (22) (i.e., $Z(x_0)$ and f_i's which make up Z_{DK}^*). Since no assumptions have been made about the distribution of $Z(x)$ (and in fact if $Z(x)$ is uni- and bivariate normal or lognormal there is no advantage in using DK for estimation) a transform function, $\phi[Y(x)]$, is necessary to transform $Z(x)$ to a random function with a standard normal distribution. Therefore we write this function in terms of Hermite polynomials

$$\phi[Y(x)] = Z(x) = \sum_{k=0}^{\infty} C_k H_k[Y(x)] \qquad (23)$$

where $Y(x)$ is assumed to be bivariate normal and the C_k's are the coefficients which are determined by using (9) or (10).

Incorporating (16) and the left most part of (23) into (22) gives

$$E[\phi[Y(x_0)]|Y(x_j)] = \sum_{i=1}^{n} E[f_i[Y(x_i)]|Y(x_j)] \qquad (24)$$

$$j = 1, 2, \ldots, n$$

The unknowns are found by applying (15) to (24) and rearranging

$$\sum_{k=0}^{K} C_k H_k[Y(x_j)] \left\{ (\rho_{oj})^k - \sum_{i=1}^{n} f_{ik} (\rho_{ij})^k / C_k \right\} = 0 \qquad (25)$$

where the infinite series have been truncated to K terms and $j = 1, 2, \ldots, n$.

Defining $b_{ik} = f_{ik}/C_k$ and noting that (25) must be satisfied for all k gives the DK system (equations (26), (27), and (28))

$$Z_{DK}^*(x_0) = \sum_{k=0}^{K} C_k H_k^*[Y(x_0)] \qquad (26)$$

where $H_k^*[Y(x_0)]$ is the estimated value of $H_k[Y(x_0)]$ and is written as a weighted sum of Hermite polynomials of the sample values

$$H_k^*[Y(x_0)] = \sum_{i=1}^{n} b_{ik} H_k[Y(x_i)] \qquad (27)$$

and the weights, b_{ik}, are determined by solving the linear kriging system

$$(\rho_{oj})^k = \sum_{i=1}^{n} b_{ik} (\rho_{ij})^k \qquad j = 1, 2, \ldots, n \qquad (28)$$

For $k = 0$, (28) reduces to the unbiasedness condition $\sum_{i=1}^{n} b_{io} = 1$ (or $Z_{DK}^* = C_0 = \mu_z$ from (11)).

Disjunctive Kriging Variance

The disjunctive kriging variance is found by combining (18) and (20) which gives

$$\text{Var}\,[Z(x_0) - Z_{DK}^*(x_0)] = E[Z(x_0)^2] - E[Z(x_0) Z_{DK}^*(x_0)] \qquad (29)$$

Substituting (11) and (12) into (29), using (13) to determine the covariance and rearranging gives

$$\sigma_{DK}^2 = \sum_{k=1}^{K} k! C_k^2 \left[1 - \sum_{i=1}^{n} b_{ik} (\rho_{oi})^k \right] \qquad (30)$$

Block Disjunctive Kriging

If a block estimate is required, the only necessary modification is to use $\bar{\rho}_{oj}$ in place of ρ_{oj} on the left-hand side of (28) where the bar indicates that the correlation coefficient is averaged over the block, V,

$$\bar{\rho}_{oj} = (1/V) \int_V \rho(x - x_j)\, dx \qquad (31)$$

Several observations about the DK system can be made: (1) (28) is in the form of a simple kriging equation (i.e., no Lagrange multiplier); (2) $k = 0$ results in the unbiasedness condition; and (3) as k increases, $\rho \to 0$ as well as the weights b_{ik}, therefore K need not be very large.

THE CONDITIONAL PROBABILITY

Estimating a conditional probability (termed the transfer function by *Matheron* [1976]; *Marechal* [1976]; *Journel and Huijbregts* [1978]; and *Kim et al.* [1977]) at a given location is possible since the disjunctive kriging estimator is nonlinear. The method used here consists of two steps. The first is to find the conditional probability that the point value of $Z(x)$ at a randomly located point, x_0, inside the block V is above a prescribed cutoff value z_c. This is called the point transfer function. Next, to obtain the conditional probability that the block value is above the cutoff value, the probability function is then integrated over the entire block.

Point Conditional Probability

In terms of a point estimate, the conditional probability that $Z(x_0)$ is above the cutoff value z_c (where y_c is the associated transformed cutoff value) is

$$P[Z(x_0) \geq z_c|Z(x_i)] = P[Y(x_0) \geq y_c|Y(x_i)] \tag{32}$$

and for the probability averaged over a block, the desired probability is

$$P[\bar{Z}(x_0) \geq z_c|Z(x_i)] = 1/V \int_V P[Z(x) \geq z_c|Z(x_i)]\, dx \tag{33}$$

where $\bar{Z}(x_o)$ indicates the block averaged value, x_0 is the randomly selected point in the block V, z_c is the cutoff value, and the $Z(x_i)$ for $i = 1, 2, 3, \ldots, n$ are the data used in the estimation process (i.e., the conditioning).

In order to determine the conditional probability using DK, the problem must be recast so that the DK estimator, which estimates the conditional expectation, will also estimate the conditional probability. Therefore, we define an indicator function $\Theta_{y_c}[Y(x_i)]$ based on the transformed cutoff value, y_c, as

$$\begin{aligned} \Theta_{y_c}(Y) &= 1 \qquad Y \geq y_c \\ \Theta_{y_c}(Y) &= 0 \qquad Y < y_c \end{aligned} \tag{34}$$

This allows (34) to be written as

$$P[Y(x_0) \geq y_c|Y(x_i)] = P[\Theta_{y_c}[Y(x_0)] = 1|Y(x_i)] \tag{35}$$

From (35), the conditional expectation can be written

$$E[\Theta_{y_c}[Y(x_0)]|Y(x_i)] = P[\Theta_{y_c}[Y(x_0)] = 1|Y(x_i)] \tag{36}$$

since $\Theta_{y_c}(Y) = 0$ for $-\infty < Y < y_c$ and 1 for $y_c < Y < \infty$. Therefore, the conditional probability that $Y(x)$ (or $Z(x)$) is greater than y_c (or z_c) is given as a conditional expectation of the indicator function $\Theta_{y_c}(Y)$. Expanding $\Theta_{y_c}(Y)$ in terms of an infinite series of Hermite polynomials with respect to the sample locations gives

$$\Theta_{y_c}[Y(x_0)] = \sum_{k=0}^{K} \theta_k H_k[Y(x_0)] \tag{37}$$

which is the conditional probability estimator, $P^*(x_0)$, and the θ_k's are the coefficients in the expansion. Applying orthogonality gives a method for evaluating the θ_k's

$$\theta_k = \int_{-\infty}^{\infty} \Theta_{y_c}(u) H_k(u) g(u)\, du \tag{38}$$

or

$$\theta_k = \int_{y_c}^{\infty} H_k(u) g(u)\, du \tag{39}$$

where $g(u)$ is the normal distribution density function and u is a dummy variable of integration. Using the definitions for $H_k(u)$ and $g(u)$ in (38) or (39) allows us to calculate the coefficients. For $k = 0$, the coefficient is

$$\theta_0 = 1 - G(y_c) \tag{40}$$

where $G(u)$ is the cumulative gaussian distribution function of $g(u)$. For $k > 0$ (note H_k is defined by Rodrigues' formula, [*Abramowitz and Stegun*, 1965]), the other coefficients are given by

$$\theta_k = g(y_c) H_{k-1}(y_c)/k! \tag{41}$$

Substituting (40) and (41) into (37) gives the conditional probability estimator, $P^*(x_0)$, in terms of the indicator function, $\Theta_{y_c}[Y(x_0)]$

$$\Theta_{y_c}[Y(x_0)] = 1 - G(y_c) + g(y_c) \sum_{k=1}^{K} H_{k-1}(y_c) H_k[Y(x_0)]/k! \tag{42}$$

The only unknown in (42) is $H_k[Y(x_0)]$. Using the estimator for $H_k[Y(x_0)]$ in (27) and substituting into (42) gives the point conditional probability estimator

$$P^*(x_0) = 1 - G(y_c) + g(y_c) \sum_{k=1}^{K} H_{k-1}(y_c) H_k^*[Y(x_0)]/k! \tag{43}$$

The density function for $\Theta_{y_c}(y)$ is

$$\Theta_{y_c}'[u] = g(u)\left\{1 + \sum_{k=1}^{K} H_k(u) H_k^*[Y(x_0)]/k!\right\} \tag{44}$$

and gives the probability density as a function of u. Using (44) we can write the estimator for the conditional probability estimator in an alternate form as

$$P^*[Y(x_0)] = \int_{y_c}^{\infty} \Theta_{y_c}'(u)\, du = \Theta_{y_c}[Y(x_0)] \tag{45}$$

Block Conditional Probability

In order to obtain estimates of the conditional probability for the block, $\bar{P}$, we integrate over V

$$\bar{P}[Y(x_0) \geq y_c|Y(x_i)] = 1/V \int_V \Theta_{y_c}[Y(x)]\, dx \tag{46}$$

Combining (43) and (46) gives the block estimator for the conditional probability, $\bar{P}^*(x_0)$

$$\bar{P}^*(x_0) = 1/V \int_V \left\{1 - G(y_c) + g(y_c) \cdot \sum_{k=1}^{K} H_{k-1}(y_c) H_k^*[Y(x_0)]/k!\right\} dx \tag{47}$$

Since only the b_{ik}'s depend on the location x, (47) can be rewritten

$$\bar{P}^*(x_0) = 1 - G(y_c) + g(y_c) \sum_{k=1}^{K} H_{k-1}(y_c) \sum_{i=1}^{n} \bar{b}_{ik} H_k[Y(x_i)]/k! \tag{48}$$

where the $\bar{b}_{ik}$ denotes that the block average value, $\bar{\rho}_{oj}$ (equation (31)), is used on the left-hand side of (28).

Factors Affecting the Conditional Probability

Some of the factors that affect the resulting P^* include the cutoff value y_c, the estimated value of $H_k[Y(x_0)]$, and the cor-

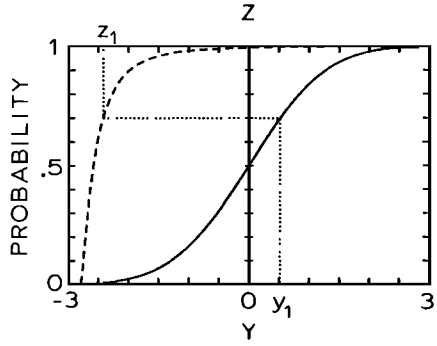

Fig. 2. Mapping of arbitrarily distributed $Z(x)$ (dashed curve) into normally distributed $Y(x)$ (solid curve).

relation function. If y_c is large and negative then $G(y_c)$ and $g(y_c)$ will approach zero. This will result in a P^* which approaches unity. When y_c is a large positive number, $G(y_c)$ approaches unity and $g(y_c)$ zero; this will result in a P^* which will approach zero.

The estimated value of $H_k[Y(x_0)]$ and the correlation function also affect the result but not in a manner easily treated analytically. Generally, as the distance between the samples and the estimate becomes large (i.e., larger than the range of the correlation function) two possibilities arise. The first occurs if the distance between all the samples is greater than the range. For this case all the b_{ik}'s for $k > 0$ are zero and $P^* = 1 - G(y_c)$. The alternative is when some (or all) of the samples are closer than the range of the correlation function; for this case the third term of (48) must be included.

The effect due to the correlation function is not amenable to generalization since the third term in (48) may be positive or negative depending on the correlation function and the sample values. However, if the correlation function has a long range (everything else constant) the importance of the third term will tend to increase.

In summary, (27), (28), and (43) are used for an estimation of the conditional probability, P^*, that $Z(x)$ at a randomly located point x_0 is greater than the cutoff value z_c, whereas (27), (28), (31), and (48) are used if the conditional probability, $\bar{P}^*$, over the block V is required.

STEPS FOR A SOLUTION

The following steps are required to estimate the value of a random variable at an unsampled location using the disjunctive kriging method. Steps 1 through 5 are preliminary to the actual estimation process.

1. Transform the original data, $Z(x)$, into a new variable, $Y(x)$, which is assumed uni- and bivariate normal. This mapping is illustrated in Figure 2. One method to approximate this function is to place the original data (i.e., $Z(x)$) in order from smallest to largest in magnitude. The cumulative frequency distribution (while keeping track of the corresponding x's) and an approximation for the probability of $Z(x_i)$ is used to obtain a transformed value of $Y(x_i)$. One possibility for calculating the probability of $Z(x_i)$ is to use $P[Z \le z_0] \approx (i - 0.5)/n$ where i is the total number of $Z(x_i)$ less than or equal to z_0 and n is the total number of samples. The associated $Y(x_i)$ is then calculated by inverting the gaussian probability function

$$Y(x) = P^{-1}[Z \le z_0] \tag{49}$$

2. Once the $Y(x)$'s are known, the coefficients C_k are calculated from (9) or (10) where the intermediate values (i.e., the v_i's in equation (10)) for the transform relationship $\phi[Y(x)] = Z(x)$ are calculated by linear interpolation, fitting an nth order polynomial to the $[Z(x), Y(x)]$ data pairs or some other method. The interpolation relationship for the nth order polynomial is then $\phi(x) = a_0 + a_1x + \cdots + a_nx^n$, where the a's are determined from the data pairs.

3. Calculate values of $Y(x_i)$ where $k = 1, 2, \ldots, K$ for each sample location (i.e., $i = 1, 2, \ldots, n$) by inverting (23). This is necessary because in general a truncated series will be used for the transform and this step will assure that the DK system will interpolate exactly at the sample locations.

4. Compute the sample mean and variance using (11) and (12) to verify that appropriate C_k's in step 2 were calculated and to aid in determining the number of k's necessary. The $\phi[Y(x)]$ function should also be plotted with the original data to help to determine the number of C_k's necessary.

5. Calculate the sample correlation function using the transformed data [i.e., $Y(x_i)$]. If the variogram for the $Z(x_i)$ is available then one can estimate the correlation function using

$$\rho(h) = 1 - \gamma(h)/C(O) \tag{50}$$

since to use DK, second-order stationarity must be satisfied [*Kim et al.*, 1977].

The remaining steps are required to calculate an estimate at each sample location and the entire sequence (steps 6 through 10) is repeated for each estimate desired.

6. Set k index to zero.

7. Calculate $H_k^*[Y(x_0)]$ in (27) by solving the simple kriging system given in (28), using the values of $H_k[Y(x_i)]$ that correspond to the appropriate samples.

8. Use (27) and (28) to calculate the kth term of (26).

9. Increment k by 1.

10. Repeat steps 7 through 10 for all K. Note that as k increases, $(\rho_{oi})^k$ goes to 0 and likewise b_{ik} goes to zero. Therefore, K need not be very large [see *Rendu*, 1980, Table 3].

Once the DK estimator has been obtained it is a simple matter to calculate the conditional probability, since $G(y_c)$, $g(y_c)$, and $H_{k-1}(y_c)$ in (43) are known. Using the result for $H_k^*[Y(x_0)]$ from step 7 in (43) completes the calculation.

Disjunctive kriging uses considerably more computer time than ordinary kriging since in general the equations must be solved K times for each estimate as well as requiring more complex calculations (i.e., calculation of Hermite polynomials and taking powers).

CONCLUSIONS

The disjunctive kriging (DK) method produces a nonlinear estimator which is suitable for estimating a spatially variable property. Generally, the DK estimator is better in the sense of reduced kriging variance compared to linear kriging estimators but is not as good as the conditional expectation (unless the random variable is multivariate normal). This property will be demonstrated in the second paper of this series [*Yates et al.*, 1986]. Disjunctive kriging has the advantage over the conditional expectation in that only the bivariate probability distributions need be known.

The calculations necessary to make an estimate are basically the same as the simple kriging method only performed K times per estimate. The other differences are (1) a transform is defined by solving for the coefficients C_k which are appropriate for the data set and (2) Hermite polynomials of order k must be calculated for each sample used in the estimation.

These latter two steps need only be done once. A third difference is that the interpolation is with respect to $H_k[Y(x)]$ for DK whereas it is based on $Z(x)$ for ordinary (linear) kriging methods. This property is also used in calculation of the conditional probability.

Since disjunctive kriging uses considerably more computer time than linear kriging, in terms of estimation alone, DK may not be justified. However, when information is required about the probability distribution of the estimation, DK offers an attractive alternative.

The disjunctive kriging method offers a simple method for calculating the conditional probability that the point or block value at a randomly located point in V is greater than an arbitrarily assigned cutoff value. This is an important result since it offers one geostatistical method whereby quantitative information is available to aid in management decisions.

Three requirements are necessary in order to use the method described in this paper. The first is that $Z(x)$ must be second order stationary. Next, a gaussian transform must exist which will transform an arbitrarily distributed $Z(x)$ into a normally distributed $Y(x)$. This requirement is easily satisfied since a gaussian transform has been shown to always exists [*Kim et al.*, 1977]. The final requirement is that the transformation produce a bivariate distribution that is jointly normal in $Y(x)$. This assumption is not required by the method in general but is adopted in order to write the conditional expectation in terms of the correlation function. If $Z(x)$ is such that the two marginal densities are normal but not jointly normal it is still possible to write the DK equations for this situation [see *Kim et al.*, 1977]. The second paper of this series will illustrate the DK method with examples.

Acknowledgment. Technical paper 4048, Arizona Agricultural Experimental Station, Tucson, Arizona.

REFERENCES

Abramowitz, M., and A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1965.

Al-Sanabani, M., Spatial variability of salinity and sodium adsorption ratio in a typic haplargid soil, M.S. thesis, Univ. of Ariz., Tucson, 1982.

Burgess, T. M., and R. Webster, Optimal interpolation and isarithmic mapping of soil properties, I, The semivariogram and punctual kriging, *J. Soil Sci.*, *31*, 315–331, 1980*a*.

Burgess, T. M., and R. Webster, Optimal interpolation and isarithmic mapping of soil properties, II, Block kriging, *J. Soil Sci.*, *31*, 333–341, 1980*b*.

Burgess, T. M., R. Webster, and A. B. McBratney, Optimal interpolation and isarithmic mapping of soil properties, IV, sampling strategy, *J. Soil Sci.*, *32*, 643–659, 1981.

Burrough, P. A., Multiscale sources of spatial variation in soil, I, The application of fractal concepts to nested levels of soil variation, *J. Soil Sci.*, *34*, 577–597, 1983*a*.

Burrough, P. A., Multiscale sources of spatial variation in soil, II, A Non-Brownian fractal model and its application in soil survey, *J. Soil Sci.*, *34*, 599–620, 1983*b*.

Journel, A. G., The lognormal approach to predicting local distributions of selective mining unit grades, *Math Geol.*, *12*, 285–303, 1980.

Journel, A. G., and C. J. Huijbregts, *Mining Geostatistics*, Academic, Orlando, Fla., 1978.

Kim, Y. C., D. E. Myers, and H. P. Knudsen, Advanced Geostatistics in Ore Reserve Estimation and Mine Planning Practitioner's Guide), Report to the U.S. Energy Research and Development Administration, Subcontract No. 76-003-E, Phase II, 1977.

Marechal, A., The practice of transfer functions: Numerical methods and their application, in *Proceedings, NATO A.S.I. "Geostat 75,"* D. Reidel, Hingham, Mass., 1976.

Matheron, G., A simple substitute for conditional expectation: The disjunctive kriging, *Proceedings, NATO A.S.I. "Geostat 75," D. Reidel*, Hingham, Mass., 1976.

Rendu, J. M., Disjunctive kriging: A simplified theory, *Math. Geol.*, *12*, 306–321, 1980.

Russo, D., Design of an optimal sampling network for estimating the variogram, *Soil Sci. Soc. Am. J.*, *48*, 708–716, 1984.

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

Vieira, S. R., D. R. Nielsen, and J. W. Biggar, Spatial variability of field-measured infiltration rate, *Soil Sci. Soc. Am. J.*, *45*, 1040–1048, 1981.

Warrick, A. W., D. E. Myers, and D. R. Nielsen, Geostatistical methods applied to soil science, in *Methods of Soil Analysis*, Part 1, 2nd ed., American Society of Agronomy, Madison, Wis., in press, 1986.

Webster, R., and T. M. Burgess, Optimal interpolation and isarithmic mapping of soil properties, III, Changing drift and universal kriging, *J. Soil Sci.*, *31*, 505–524, 1980.

Yates, S. R., A. W. Warrick, and D. E. Myers, Disjunctive kriging, 2, Examples, *Water Resour. Res.*, *22*, this issue, 1986.

D. E. Myers, Department of Mathematics, The University of Arizona, Tucson, AZ 85721.

A. W. Warrick, Department of Soil and Water Science, The University of Arizona, Tucson, AZ 85721.

S. R. Yates, U.S. Environmental Protection Agency, R. S. Kerr Environmental Research Laboratory, P.O. Box 1198, Ada, OK 74820.

(Received May 3, 1985;
revised January 29, 1985;
accepted February 7, 1986.)