Mathematical Geology, Vol. 21, No. 2, 1989

Some Asymptotic Properties of Kriging When the Covariance Function Is Misspecified[1]

Michael L. Stein[2] and Mark S. Handcock[2]

The impact of using an incorrect covariance function on kriging predictors is investigated. Results of Stein (1988) show that the impact on the kriging predictor from not using the correct covariance function is asymptotically negligible as the number of observations increases if the covariance function used is "compatible" with the actual covariance function on the region of interest R. *The definition and some properties of compatibility of covariance functions are given. The compatibility of generalized covariances also is defined. Compatibility supports the intuitively sensible concept that usually only the behavior near the origin of the covariance function is critical for purposes of kriging. However, the commonly used spherical covariance function is an exception: observations at a distance near the range of a spherical covariance function can have a nonnegligible effect on kriging predictors for three-dimensional processes. Finally, a comparison is made with the perturbation approach of Diamond and Armstrong (1984) and some observations of Warnes (1986) are clarified.*

KEY WORDS: prediction, intrinsic random fields, compatibility, robustness.

INTRODUCTION

Kriging is a method of prediction for random fields popular in mining and hydrology (Journel and Huijbregts, 1978; Kitanidis, 1983). The basic model is that $Z(x)$ is the value of a quantity of interest at location x, and $Z(\cdot)$ is a random field on $\mathbf{R}^d$ with mean function

$$EZ(x) = \beta' f(x) \tag{1}$$

where $f(\cdot)$ is a known vector-valued function and β is a vector of unknown coefficients, and covariance function

$$\operatorname{cov}\left[Z(x), Z(y)\right] = K(x, y)$$

[1]Manuscript received 19 March 1987, accepted 10 December 1987.
[2]Department of Statistics, University of Chicago, Chicago, Illinois 60637.

We observe $Z(x_1), \cdots, Z(x_N)$ and wish to predict $Z(x_0)$. A commonly used class of predictors are linear combinations of the form

$$\sum_{i=1}^{N} \lambda_i Z(x_i)$$

The kriging predictor, $\hat{Z}(x_0)$, is the unbiased linear predictor that minimizes the variance of the prediction error. The weights $\lambda = (\lambda_1, \cdots, \lambda_N)'$ defining $\hat{Z}(x_0)$ are given by

$$\lambda = C^{-1}c + C^{-1}F'(FC^{-1}F')^{-1}[f(x_0) - FC^{-1}c] \tag{2}$$

where

$$F = [f(x_1), \cdots, f(x_N)]$$

$$c = [K(x_0, x_1), \cdots, K(x_0, x_N)]'$$

C is an $N \times N$ matrix with ijth element $K(x_i, x_j)$, and F and C have full rank (Goldberger, 1962). The prediction error is denoted by $e(x_0) = \hat{Z}(x_0) - Z(x_0)$; its variance is

$$K(x_0, x_0) - c'C^{-1}c + [f(x_0) - FC^{-1}c]'(FC^{-1}F')^{-1}[f(x_0) - FC^{-1}c] \tag{3}$$

A concise introduction to kriging methods is given by Rendu (1978).

In practice, the covariance function $K(\cdot, \cdot)$ is not specified and must be estimated from data. The effect of not having the correct covariance function in Eqs. (2) and (3) has been considered by Diamond and Armstrong (1984), Sukhatme (1985), Yakowitz and Szidarovszky (1985), Warnes (1986), Armstrong and Myers (1984), and Stein (1988). Diamond and Armstrong (1984), Sukhatme (1985), Warnes (1986), and Armstrong and Myers (1984) investigate the effect of small perturbations of the covariance function on the kriging predictor given a fixed set of observations. Yakowitz and Szidarovszky (1985), Stein (1988), and the present work study the behavior of kriging predictors based on an incorrect covariance function as the number of observations in some fixed region R increases.

In many situations, $Z(x)$ is an intrinsic random function of nonnegative order so that $K(x, y)$ can not be written as $K(x - y)$, but certain linear combinations of $Z(x)$ (i.e., increments) have an explicitly defined (generalized) covariance function that is homogeneous (Delfiner, 1976). This paper will focus on fields for which the covariance function is homogeneous [i.e., $K(x, y) = K(x - y)$].

In the next section, results of Stein (1988) are summarized, showing that the impact on the kriging predictor from not using the correct covariance func-

tion is asymptotically negligible as the number of observations increases if the covariance function used is "compatible" with the actual covariance function on the region of interest R. The definition and some conditions for the compatibility of covariance functions on a region R are given. For homogeneous covariance functions, two covariance functions must behave similarly at the origin in order to be compatible. Covariance functions that behave similarly at the origin are, in many practical cases, compatible. Thus, the intuitively sensible concept that only the behavior of the covariance function at the origin needs to be estimated well is supported. However, similar behavior at the origin does not guarantee compatibility. In particular, the commonly used spherical covariance function (Journel and Huijbregts, 1978, p. 164), for large enough regions in three dimensions, is shown not to be compatible with an exponential covariance function that behaves similarly at the origin. An example is given in which the kriging predictor based on a spherical covariance function exhibits what would usually be considered physically unrealistic behavior. This suggests that the spherical covariance function is an inappropriate model for most three-dimensional fields. Finally, the compatibility perspective on the misspecification problem is compared to the perturbation approach of Diamond and Armstrong (1984). When compatibility can be brought to bear, it provides sharp practically useful solutions. This is demonstrated in clarifying some examples given by Warnes (1986).

ASYMPTOTICALLY EFFICIENT PREDICTION

Consider a random field $Z(\cdot)$ with a continuous covariance function on a bounded region R in $\mathbf{R}^d$ with mean function as in Eq. (1). Let x_0 be a point in R and $\{x_i\}_{i=1}^{\infty}$ a sequence of points in R that has x_0 as a limit point but does not include x_0. The asymptotic behavior of the kriging predictor of $Z(x_0)$ based on $Z(x_1), \cdots, Z(x_N)$ as $N \to \infty$ is considered when the covariance function used in Eqs. (2) and (3) is incorrect. The asymptotic approach of an increasing number of observations in a fixed region also is used by Yakowitz and Szidarovszky (1985) and Stein (1988) and is a natural one for spatial phenomena. Often the behavior of some field $Z(\cdot)$ in a fixed region of interest is important, and, at least in principle, taking an increasing number of observations in this fixed region can be considered. Of course, in practice, such additional sampling is rare. As with all asymptotic statistical results, the real utility of this formulation should be judged in terms of how appropriate it is for samples of the sizes seen in practice.

Now, suppose $K_0(x, y)$ and $K_1(x, y)$ are compatible covariance functions on R. An exact mathematical definition and some interpretations of compatibility are given in the next section. Let $e_i(N)$ measure the error of the kriging predictor of $Z(x_0)$ based on $Z(x_1), \cdots, Z(x_N)$ and the covariance function

$K_i(\cdot, \cdot)$, $i = 0, 1$. Let $V_i(\cdot)$ be the variance of a linear combination of values of $Z(x)$ under $K_i(\cdot, \cdot)$. If

$$\lim_{N \to \infty} V_0[e_0(N)] = 0 \tag{4}$$

then Stein (1988) proves

$$\lim_{N \to \infty} \frac{V_0[e_0(N)]}{V_0[e_1(N)]} = 1 \quad \text{and} \tag{5}$$

$$\lim_{N \to \infty} \frac{V_1[e_1(N)]}{V_0[e_1(N)]} = 1 \tag{6}$$

Equation (5) says that, if our predictor of $Z(x_0)$ is based on K_1 when in fact K_0 is the actual covariance function, then as long as K_0 and K_1 are compatible on R, an asymptotically efficient predictor of $Z(x_0)$ will be obtained. That is, the ratio of the actual variance of the prediction error using K_1 (i.e., $V_0[e_1(N)]$) to the variance of the prediction error using the correct covariance function K_0 (i.e., $V_0[e_0(N)]$) tends to one as $N \to \infty$. Equation (6) says that the ratio of the value of the variance of the prediction error obtained by using K_1 in Eq. (3) (i.e., $V_1[e_1(N)]$) to the actual variance of the error of this predictor when K_0 is the correct covariance function (i.e., $V_0[e_1(N)]$) tends to one as $N \to \infty$. In other words, the value of the variance of the prediction error based on the incorrect covariance function K_1 will be close to the actual variance for large N as long as K_1 is compatible with the true covariance function K_0 on R. The assumption in Eq. (4) just says that the kriging predictor using the correct covariance function is consistent as $N \to \infty$. Because x_0 is a limit point of the observation sites, this assumption is weak. For example, this condition is satisfied when $f(x) \equiv 1$ and $K_0(\cdot, \cdot)$ is bounded and continuous on $R \times R$ (Stein, 1988). If K_0 and cK_1 are compatible for some $c > 0$, Eq. (5) still will hold and the limit in (6) will be c^{-1}. That is, an asymptotically efficient predictor will result and the value for variance of the prediction error is, asymptotically, off by a constant multiple. In summary, these results say that, for purposes of kriging, the difference between using the correct covariance function K_0 and a compatible covariance function K_1 is asymptotically negligible.

These results apply to both Gaussian and non-Gaussian fields. However, caution against using linear kriging for highly non-Gaussian fields is recommended for two reasons. First, nonlinear predictors of $Z(x_0)$ may perform much better than the best linear predictor. More importantly, the formula for variance of the prediction error in Eq. (3) gives the unconditional variance, which may be very different from the conditional variance for a non-Gaussian field. Because predicting $Z(x_0)$ *given* the observed values $Z(x_1), \cdots, Z(x_N)$ is of interest, the conditional distribution of the prediction error really is wanted.

Hence, the unconditional variance of the prediction error can be misleading for non-Gaussian fields.

DEFINITION AND DISCUSSION OF COMPATIBILITY

Introduction

In order to define compatibility of covariance functions, some properties of probability measures need to be introduced. Let R be a bounded region in $\mathbf{R}^d$ and $[Z(x) : x \in R]$ be a random field. For the moment, consider continuous random fields. Suppose $[Z(x) : x \in R]$ has covariance function $K(x, y)$ and mean function $m(x)$. For any such mean function and covariance function, a unique Gaussian random field with those characteristics (Doob, 1953, p. 72) can be produced. To simplify the notation, denote the unique Gaussian probability measure corresponding to $m(x)$ and $K(x, y)$ by $[m, K]$.

Mutual Singularity and Mutual Absolute Continuity of Probability Measures

Let P_0 and P_1 be arbitrary probability measures with respect to a given sample space Ω (and an appropriate σ-field). A *support* of P_0 is defined to be any event A with $P_0(A) = 1$. P_0 and P_1 are called mutually *singular* if they have disjoint supports. That is, disjoint events A and B exist that satisfy

$$P_0(A) = 1 \qquad P_1(B) = 1, \qquad \text{and} \qquad A \cap B = \varnothing$$

where $\varnothing$ is the null event. This is denoted by $P_0 \perp P_1$. P_0 and P_1 are called mutually *absolutely continuous* if they have the same set of supports. That is, an event A satisfies $P_0(A) = 1$ if and only if $P_1(A) = 1$. This is denoted by $P_0 \sim P_1$.

Now suppose that a realization of the field under either P_0 or P_1 is given, but which one is the correct measure is not known. If $P_0 \perp P_1$ then, with probability one, which law produced this single realization can be determined. If $P_0 \sim P_1$, which law produced this single realization cannot be determined with certainty. These definitions denote different ends of the spectrum: $P_0 \sim P_1$ when supports are the same and $P_0 \perp P_1$ when they have a disjoint support. Although two general probability measures may be neither mutually singular nor mutually absolutely continuous, any two Gaussian measures are necessarily either mutually singular or mutually absolutely continuous (Hajek, 1958, p. 615). This property can be used to show that

$$[m_0, K_0] \sim [m_1, K_1] \Leftrightarrow \begin{cases} [m_0 - m_1, K_1] \sim [0, K_1] & \text{and} \\ [0, K_0] \sim [0, K_1] \end{cases}$$

Hence, to establish mutual absolute continuity in the Gaussian case, two simpler cases may be considered:

(1) K_0 is identical to K_1, but m_0 differs from m_1, and

(2) K_0 differs from K_1, but m_0 is identical to m_1.

For homogeneous covariance functions, both of these cases are considered at length by Rozanov (1968), Skorokhod and Yadrenko (1973), Ibragimov and Rozanov (1978), and Krasnitskii (1973). Here, only the second case, where the covariance function is misspecified, is considered.

Definition of Compatibility

Definition. Let K_0 and K_1 be covariance functions defined on R, a bounded region in $\mathbf{R}^d$. Then K_0 is *compatible* with K_1 on R if $[0, K_0] \sim [0, K_1]$.

Compatibility is well-defined as each covariance and mean function pair characterizes a Gaussian probability measure and vice versa. Because Gaussian measures are either mutually absolutely continuous or mutually singular, K_0 is not compatible with K_1 on R if and only if $[0, K_0] \perp [0, K_1]$. The presence of R in the definition is essential. If K_0 is compatible with K_1 on R, K_0 is also compatible with K_1 on any subset of R. On the other hand, if K_0 is not compatible with K_1 on R, K_0 still may be compatible with K_1 on some subset of R.

Compatibility is a property of the covariance functions on a specified domain and applies to non-Gaussian as well as Gaussian fields. The definition does make use of properties of the Gaussian random fields that are characterized by the covariance functions. However, the idea of compatibility of covariance functions is appropriate whatever the random field $Z(\cdot)$ that the covariance functions describe. In particular, validity of Eqs. (5) and (6) does not depend in any way on $Z(\cdot)$ being Gaussian.

Now consider a random field on R with unknown covariance function K_0. All covariance functions on R can be partitioned into a set that is compatible with K_0 and a set that is not compatible with K_0. Covariance functions in different sets are not compatible on R whereas those in the first set are compatible with each other. Results given by Eqs. (5) and (6) indicate that, when estimating K_0, it is asymptotically not important to distinguish between covariance functions in this first set.

Behavior of the Covariance Function at the Origin and Local Behavior of a Random Field

Local variations of a random field about its mean function are governed by the behavior of its covariance function at the origin (Ripley, 1981, p. 44–75). Let $Z(x)$ be a random field with constant mean and homogeneous covariance function $K(x)$. $K(x)$ is continuous at the origin if and only if $Z(x)$ is mean-square continuous (Yaglom, 1962, p. 22). If $K(x)$ has continuous second de-

rivative, $Z(x)$ is mean-square differentiable. More generally, if $K(x)$ has $2m$ derivatives, $Z(x)$ is m times mean-square differentiable. Hence, the smoothness of $K(x)$ at the origin translates directly to the local smoothness of $Z(x)$.

Let $K_0(x)$ and $K_1(x)$ be homogeneous covariance functions on a region R. If K_0 is compatible with K_1, they have similar behavior at the origin. For example, suppose $R = [0, T] \subset \mathbf{R}^1$ and K_0 has $2m$ continuous derivatives on R with $K_0^{(2m+1)}(0^+) \neq 0$. If K_1 is compatible with K_0, it must also have $2m$ continuous derivatives on R and $K_0^{(2m+1)}(0^+) = K_1^{(2m+1)}(0^+)$. As will be seen in the next section, this condition is not generally sufficient for compatibility, although it is in many practical cases. For instance, $K_0(x) = \lambda_0 e^{-|x|/\lambda_0}$ and $K_1(x) = \lambda_1 e^{-|x|/\lambda_1}$ are compatible on $[0, T]$ for all T and positive λ_0, λ_1. Note that $K_0(x)$ and $K_1(x)$ are not differentiable at zero and $K_0'(0^+) = K_1'(0^+) = -1$. An example of a pair of covariance functions that are compatible and do not have linear behavior at the origin is $K_0(x) = 5e^{-|x|}[\cos(|x|) + \sin(|x|)]$ and $K_1(x) = e^{-2|x|}[\cos(|x|) + 2\sin(|x|)]$. They satisfy the above condition with $K_0^{(3)}(0^+) = K_1^{(3)}(0^+) = 20$. As a last example, $K_0(x) = e^{-|x|^2/\lambda}$ is not compatible with $K_1(x) = e^{-|x|}$ for any λ because the first is parabolic at the origin and the second is linear at the origin.

In higher dimensions, similar results hold. Precise necessary and sufficient conditions for arbitrary covariance functions are known (Skorokhod and Yadrenko, 1973). Many results are stated better in terms of the corresponding spectral densities. For example, suppose K_0 and K_1 are isotropic [$K_0(x) = K_0(|x|)$] and have spectral densities $f_0(\lambda)$ and $f_1(\lambda)$, respectively, then necessarily $f_0(\lambda) = f_0(|\lambda|)$ and $f_1(\lambda) = f_1(|\lambda|)$. Using Theorem 7 of Skorokhod and Yadrenko (1973), K_0 is compatible with K_1 on any bounded region in $\mathbf{R}^d$ if

$$\int_0^\infty \left[1 - \frac{f_1(\lambda)}{f_0(\lambda)}\right]^2 \lambda^{d-1}\, d\lambda < \infty \tag{7}$$

and $f_0(\lambda)$ satisfies a technical condition. Compatibility of covariance functions with geometric anisotropies (Journel and Huijbregts, 1978, p. 177) can be described in terms of compatibility of the corresponding isotropic covariance functions. For example, for a given matrix V, $K_0(|Vx|)$ and $K_1(|Vx|)$ are compatible on R if $K_0(|x|)$ and $K_1(|x|)$ are compatible on the set $Q = \{Vx : x \in R\}$.

Finally, up to now, the covariance functions have been assumed to be continuous. Often the covariance function has a discontinuity at the origin [i.e., "nugget effect" (Ripley, 1981, p. 50)] caused by measurement error or microstructures. If K_0 is compatible with K_1 on R, $K_0(x) + \theta_0 I\{x = 0\}$ can be shown to be compatible with $K_1(x) + \theta_1 I\{x = 0\}$ on R if and only if $\theta_0 = \theta_1$.

Compatibility of Generalized Covariances

Extending the notion of compatibility to intrinsic random functions (IRFs) and generalized covariances (Matheron, 1973) is considered briefly. Every IRF

has an infinite number of representations as a (possibly nonstationary) random field (Matheron, 1973, p. 445). Let $Z_i(x)$ be a continuous IRF on R, a bounded region in $\mathbf{R}^d$, with generalized covariance $G_i(r)$, $i = 0, 1$. G_0 and G_1 are defined to be compatible on R if compatible covariance functions K_0 and K_1 exist (possibly inhomogeneous) such that K_i is the covariance function of a representation of Z_i. The compatibility of K_0 and K_1 implies that Eqs. (5) and (6) will hold for that pair of representations. The variance of any contrast depends only on the generalized covariance and not on the particular representation. Therefore, because all prediction errors are contrasts, Eqs. (5) and (6) apply for compatible generalized covariances G_0 and G_1.

Determining the compatibility of generalized covariances is straightforward when they are bounded. If G_0 is bounded on $\mathbf{R}^d$, Matheron (1973, Theorem 2.4) proved that representations of Z_0 have homogeneous covariance functions that differ by constants. If $f(\lambda)$ is the spectral density of a homogeneous covariance function $K_0(x)$ and $\exists \alpha$, $\lambda_0 > 0$, and n a positive integer such that

$$f(\lambda) \geq \alpha\left(1 + |\lambda|^2\right)^{-n} \qquad \text{for} \quad |\lambda| > \lambda_0$$

K_0 can be shown, by applying Theorem 3 of Skorokhod and Yadrenko (1973), to be compatible (on R) with every covariance function that differs from it by a constant. Subject to this mild condition, the compatibility of bounded generalized covariances can be determined directly from the compatibility of any pair of covariance functions corresponding to respective stationary representations. This result will allow us to determine the compatibility of variograms (IRF-0s) used in the upcoming examples. The condition is not satisfied by the covariance function $K_G(x; a, \sigma^2) = \sigma^2 \exp(-|x|^2/a^2)$ that is used in Warnes (1986). $K_G(x; a, \sigma^2)$ is not compatible with any other covariance function (Ibragimov and Rozanov, 1978, p. 95), so it is incompatible with $\theta^2 + K_G(x; a, \sigma^2)$ for $\theta^2 > 0$. This covariance function corresponds to a Gaussian field with "analytic" realizations, so that, for example, the entire realization along a line can be reconstructed from knowing the realization on any segment of that line. Thus, this covariance function normally would be considered a physically unrealistic model.

The focus on bounded generalized covariances is not overly limiting because, in most practical situations, attention is restricted to some natural bounded region in $\mathbf{R}^d$. If a bounded generalized covariance on $\mathbf{R}^d$ coinciding with our unbounded model on this bounded region can be found, this bounded model may be used and the above result can be applied. As an example, suppose that kriging is contemplated on $[0, 1]$ and the unbounded generalized covariance model $G_L(x) = -x$ is being considered. The alternative

$$G_T(x) = \begin{cases} -x & \text{if } |x| < 1 \\ -1 & \text{if } |x| \geq 1 \end{cases}$$

coincides with G_L on $[0, 1]$ and is bounded on $\mathbf{R}^1$. Just as with covariance functions, the behavior at the origin of generalized covariance is critical in determining compatibility. For example, the order 1 generalized covariances $-x + \alpha x^3$ and βx^3 can be shown to be incompatible on any bounded region for any $\alpha > 0$ and $\beta > 0$, and $-x + \alpha x^3$ and $-x$ are conjectured to be compatible on any bounded region for any $\alpha > 0$.

INCOMPATIBLE COVARIANCE FUNCTIONS WITH SIMILAR BEHAVIOR AT THE ORIGIN

A One-Dimensional Example

The "triangular" covariance function with range b for a one-dimensional field is given by

$$K(x) = (b - |x|)^+$$

where the plus indicates the positive part. Let $K_0(x)$ be the triangular covariance function with range one. This covariance function has the same behavior at the origin as

$$K_1(x) = e^{-|x|}$$

However, if $R = [0, T]$, $K_0(\cdot)$ and $K_1(\cdot)$ are compatible on R if and only if $T \leq 1$ (Ibragimov and Rozanov, 1978, p. 100). The incompatibility for $T > 1$ is due to $K_0(x)$ not being differentiable at $x = 1$, in contrast to the behavior of $K_1(x)$.

Kriging predictors based on $K_0(\cdot)$ can be shown to behave in a manner that usually would be considered physically unrealistic. Throughout this section, the mean of $Z(\cdot)$ is taken to be an unknown constant. Suppose that $Z(\cdot)$ is observed at the N locations,

$$\frac{1}{N-1}, \frac{2}{N-1}, \cdots, 1, \frac{N}{N-1}$$

and $Z(0)$ is to be predicted. Assuming $K_0(\cdot)$ is the covariance function, the kriging predictor of $Z(0)$ is

$$Z[(N-1)^{-1}] - \frac{1}{2}Z(1) + \frac{1}{2}Z[N/(N-1)] \tag{8}$$

Two observations, $Z(1)$ and $Z[N/(N-1)]$, that are far from $Z(0)$ get substantial weights in the kriging predictor for all N. These two observations are both about a distance of one from $Z(0)$; $K_0(x)$ not being differentiable at $x = 1$ leads to this phenomenon. The value of the variance of the prediction error under $K_0(\cdot)$ is $1.5/(N-1)$. The variance of the kriging predictor using $Z[(N$

$-1)^{-1}]$ alone is $2/(N-1)$, so $Z(1)$ and $Z[N/(N-1)]$ have a nonnegligible impact on the kriging variance. Suppose, in fact, that $Z(x)$ has a triangular covariance function with a range of two, so $K_0(\cdot)$ has the same slope at the origin. The variance of the error of the predictor in Eq. (8) is then $2.5/(N-1)$, which is $5/3$ times the variance of the prediction error using $K_0(\cdot)$. Also note that, if $Z(\cdot)$ has a triangular covariance function with a range of two, the kriging predictor of $Z(0)$ is $Z[1/(N-1)]$, which does not put weight on distant observations. The predictor in Eq. (8) is peculiar, and thus the triangular covariance function would usually be an inappropriate model when the region of interest is an interval whose length is greater than the range of the covariance function. This effect is not due to the fact that the mean is unknown because the simple kriging predictor of $Z(0)$ (i.e., assuming the mean is zero) is given by

$$\frac{2N-3}{2(N-1)} Z[(N-1)^{-1}] - \frac{1}{2} Z(1) + \frac{N-2}{2(N-1)} Z[N/(N-1)]$$

The Spherical Covariance Function

The triangular covariance function has a physical interpretation as a moving average of white noise: if $W(x)$ is Brownian motion so that $dW(x)$ is white noise, and

$$Z(x) = \int_{x-1/2}^{x+1/2} dW(t) = W(x+\tfrac{1}{2}) - W(x-\tfrac{1}{2})$$

$Z(x)$ has the triangular covariance function with range of one. The commonly used spherical covariance function has a similar interpretation in three dimensions. If $Y(x)$ is three-dimensional white noise and $Z(x)$ is obtained by integrating $Y(\cdot)$ over a ball of radius 0.5 centered at x and normalizing by the appropriate constant, $Z(\cdot)$ has covariance function

$$K(x) = \begin{cases} \frac{2}{3} - |x| + \frac{1}{3}|x|^3 & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases} \tag{9}$$

the spherical covariance function with range one. If R contains a ball of diameter greater than one, the spherical covariance function in Eq. (9) can be shown, using Theorem 2 in Skorokhod and Yadrenko (1973) to be incompatible with $K(x) = \exp(-|x|)$ on R, even though these covariance functions behave similarly at the origin. The incompatibility is caused by the lack of a third derivative at $|x| = 1$ for the spherical covariance function. This theorem also shows that, for any bounded region in one or two dimensions, spherical and exponential covariance functions that behave similarly at the origin are compatible. Thus, one might expect to find unusual behavior in a kriging predictor based on a spherical covariance function with three-dimensional fields, but not with one- or two-dimensional fields.

Consider the situation, in three dimensions, where $Z(0)$ is to be predicted based on observing the field on four spheres of radii ϵ, 1, $1 + \epsilon/2$, and $1 + \epsilon$ units and centered at the origin (Fig. 1). As in the case of the triangular covariance function in one dimension, behavior of the field at distances near the range of the spherical covariance function from $Z(0)$ can have a nonnegligible impact on the kriging predictor even when the field near $Z(0)$ is observed. The following example is a three-dimensional analog of the triangular example. Define $\bar{Z}(r)$ to be the average of $Z(x)$ over the surface of a sphere of radius r centered at the origin. By symmetry, the predictor based on the four spheres is of the form

$$\lambda_\epsilon \bar{Z}(\epsilon) + \lambda_1 \bar{Z}(1) + \lambda_{1+\epsilon/2} \bar{Z}(1 + \epsilon/2) + \lambda_{1+\epsilon} \bar{Z}(1 + \epsilon) \tag{10}$$

$\bar{Z}(1)$, $\bar{Z}(1 + \epsilon/2)$, and $\bar{Z}(1 + \epsilon)$ will be demonstrated to have a substantial impact on the kriging predictor of $Z(0)$ for small ϵ even though $\bar{Z}(\epsilon)$ is observed. Now, if $Z(\cdot)$ is isotropic with covariance function $K(\cdot)$,

$$\operatorname{cov}\left[\bar{Z}(r_1), \bar{Z}(r_2)\right] = \frac{1}{2r_1r_2} \int_{|r_1 - r_2|}^{r_1 + r_2} yK(y)\, dy \tag{11}$$

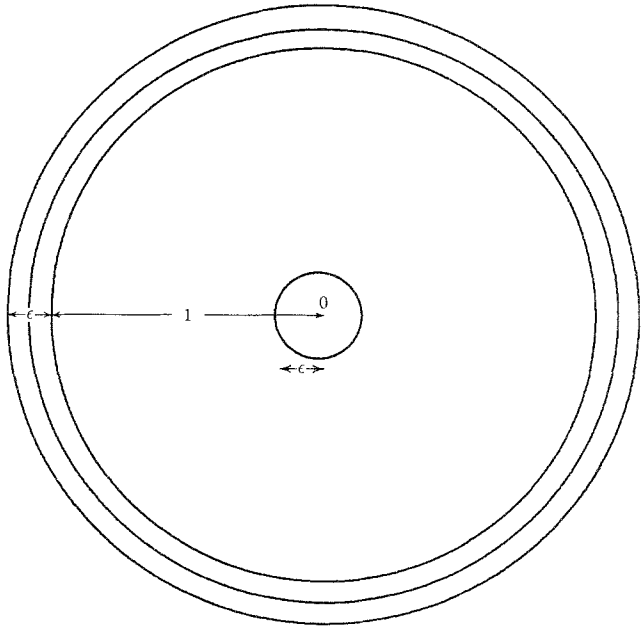

Fig. 1. Cross section of the three-dimensional data configuration in which the random field $Z(\cdot)$ is observed everywhere on spheres of radii ϵ, 1, $1 + \epsilon/2$, and $1 + \epsilon$. The object is to predict $Z(0)$.

If $K(\cdot)$ is the spherical covariance function in Eq. (9), the kriging weights in Eq. (10) can be shown using Eq. (11) to be

$$\begin{aligned} \lambda_\epsilon &= 1 + O(\epsilon) \\ \lambda_1 &= -\frac{3}{4\epsilon} - \frac{2389}{2944} + O(\epsilon) \\ \lambda_{1+\epsilon/2} &= \frac{3}{2\epsilon} + \frac{1833}{1472} + O(\epsilon) \\ \lambda_{1+\epsilon} &= -\frac{3}{2\epsilon} - \frac{1277}{2944} + O(\epsilon) \end{aligned} \tag{12}$$

The variance of the prediction error when the spherical covariance function in Eq. (9) is correct is $(55/96)\epsilon + O(\epsilon^2) \approx 0.573\epsilon$. The kriging predictor of $Z(0)$ based just on $\overline{Z}(\epsilon)$ is, of course, $\overline{Z}(\epsilon)$, and the variance of its prediction error is $(2/3)\epsilon + O(\epsilon^2) \approx 0.667\epsilon$. As $\epsilon \downarrow 0$, $\overline{Z}(1)$, $\overline{Z}(1 + \epsilon/2)$, and $\overline{Z}(1 + \epsilon)$ have a nonnegligible impact on the kriging predictor of $Z(0)$. Note that $\overline{Z}(1)$, $\overline{Z}(1 + \epsilon/2)$, and $\overline{Z}(1 + \epsilon)$ do not get zero weight even though they are uncorrelated with $Z(0)$.The point is that the partial correlation between $Z(0)$ and $\overline{Z}(1)$, $\overline{Z}(1 + \epsilon/2)$, or $\overline{Z}(1 + \epsilon)$ induced by $\overline{Z}(\epsilon)$ is not zero, and this correlation determines the importance of the observation in the predictor.

Now consider this example using the covariance function $\exp(-|x|)$, which behaves similarly at the origin as the covariance function in Eq. (9), but is incompatible with it on large enough regions in three dimensions. As $\epsilon \downarrow 0$, the kriging weights can be shown to be

$$\begin{aligned} \lambda_\epsilon &= 1 + \frac{2e^2}{3(3e^2 - 4e - 3)}\epsilon + O(\epsilon^2) \\ \lambda_1 &= \frac{-2e^2}{3(3e^2 - 4e - 3)}\epsilon + O(\epsilon^2) \end{aligned}$$

whereas $\lambda_{1+\epsilon/2}$ and $\lambda_{1+\epsilon}$ are $O(\epsilon^2)$. The variance of the prediction error when the exponential covariance function is correct is $(2/3)\epsilon + O(\epsilon^2) \approx 0.667\epsilon$. The variance of the prediction error using only $\overline{Z}(\epsilon)$ to predict $Z(0)$ is also $(2/3)\epsilon + O(\epsilon^2)$. So, using an exponential covariance function, the distant observations have an asymptotically negligible impact on the kriging predictor of $Z(0)$. The variance of the error for the predictor using weights in Eq. (12) when $\exp(-|x|)$ is the correct covariance function is $(73/96)\epsilon + O(\epsilon^2) \approx 0.760\epsilon$. Thus, the actual variance is 32% greater than the predicted variance for small ϵ. In contrast, if the covariance function is presumed to be exponential, the actual variance of the prediction error is approximately $(2/3)\epsilon$ whether the true covariance function is exponential or the spherical covariance function in Eq. (9). The numerical results presented here are summarized in Table 1.

Table 1. An Example of Unusual Behavior of a Kriging Predictor Based on a Spherical Covariance Function

	Actual covariance	
Presumed covariance	Spherical	Exponential
Spherical	0.5729ϵ	0.7604ϵ
Exponential	0.6667ϵ	0.6667ϵ

Note: Each entry gives the variance of the prediction error when predicting $Z(0)$ for the situation described in the text. The presumed covariance function is used to compute the kriging predictor [see Eq. (2)], and the actual covariance function is used to compute the variance of the error of that predictor. Errors in the table entries are all $O(\epsilon^2)$ as $\epsilon \downarrow 0$.

This example suggests that the spherical covariance function is an inappropriate model for most fields in three dimensions unless some reason exists for thinking that the observed field is a spherical moving average of white noise. In most practical situations, using spherical covariance functions may not lead to serious problems. However, to use a model that is known to lead to such physically unrealistic results is dubious, especially when alternative models (e.g., the exponential covariance function) exist that do not, as far as known, have this problem.

COMPARISON WITH PERTURBATION METHODS

The three factors that affect the kriging predictor are

(1) the covariance model,

(2) the number and spatial arrangement of observations, and

(3) the relationship between the support for the observations and support for the quantity being kriged.

Small perturbations in any of these will lead to changes in the kriging predictor. Diamond and Armstrong (1984) (from now on, DA) attempt to define a neighborhood of variograms "close" to a given variogram based on relative deviation from that variogram. The effect of perturbing the variogram with the observations and geometric relationships fixed is investigated. This perturbation is passed through the matrix equations defining the kriging system to give bounds for the effect on kriging weights and estimation variance. Armstrong and Meyers (1984) extend the work, showing that the effect of the above factors on kriging weights can be determined by solving the appropriate perturbed kriging equations. In this context, Sukhatme (1985) derives approximations to kriging

weights explicitly in terms of known covariance quantities. Finally, Warnes (1986) approximates the effects of covariance perturbations on the prediction surfaces of kriging.

The perturbation approach is contrasted with compatibility in several examples. Although results given in Eqs. (5) and (6) do not provide explicit error bounds, they provide useful insights in these examples. In the notation of DA, let $N_\delta(\gamma) = \{ g \in \mathfrak{A} : |(g/\gamma) - 1| < \delta \}$ where $\mathfrak{A}$ is the class of valid variograms in $\mathbf{R}^d$ and $|\cdot|$ is the sup norm on continuous functions from $\mathbf{R}^+$ to $\mathbf{R}^+$. In situations where the support of the variograms is a subset of $\mathbf{R}^+$, the variogram is extended to $\mathbf{R}^+$ in a manner consistent with the restriction to conditional positive definiteness.

Example 1

Consider the class of variograms $\gamma_\lambda(x) = \lambda(1 - e^{-x/\lambda})$, $\lambda > 0$, $x \geq 0$. Using Eq. (7), these can be shown to be compatible on any bounded region in $\mathbf{R}^d$. Also note that $\gamma_\lambda(x) \in N_\delta(\gamma_1)$ if and only if $|\lambda - 1| < \delta$. Hence, λ may be chosen to make γ_λ as "close" or as "far" from γ_1 as desired in the sense of DA, whereas any γ_λ is "close" to γ_1 in the compatibility sense.

Example 2

Consider the class of variograms

$$\gamma_\alpha(x) = 1 - \frac{e^{-3\alpha x/4}}{\beta}\left[3\alpha \sin(\beta x/4) + \beta \cos(\beta x/4)\right]$$

$$= x^2 - \frac{\alpha}{2}x^3 + \frac{(9\alpha^2 - 8)}{48}x^4 + O(x^5)$$

for small x, where $\beta = \sqrt{32 - 9\alpha^2}$ and $0 \leq \alpha \leq 1$. Here, γ_α can be shown to be the variogram of the AR(2) process $U(x)$ defined by

$$d\dot{U}(x) + \tfrac{3}{2}\alpha\dot{U}(x)dx + 2U(x)dx = dW(x)$$

where $W(x)$ is a Wiener process. Although $U(x)$ is differentiable, $\dot{U}(x)$ is nowhere differentiable and hence $\ddot{U}(x)$ does not exist. Thus $\gamma_\alpha \sim c\gamma_{\alpha_1}$ if and only if $c = \alpha/\alpha_1$, where $0 \leq \alpha, \alpha_1 \leq 1$ so that the predictors based on γ_α and γ_{α_1} will be asymptotically the same. Furthermore, values for the kriging variance will, asymptotically, differ only by the constant multiple α/α_1.

In the simple situation (Fig. 2), a constant mean random field is to be predicted at 3ϵ based on observations at 0 and ϵ. Suppose that the predictions under the true variogram γ_1 are compared to those under the compatible variogram model $2\gamma_{1/2}$ (Table 2). Each ϵ represents a different kriging situation, with the region of interest shrinking as $\epsilon \downarrow 0$. The exact numerical solutions for

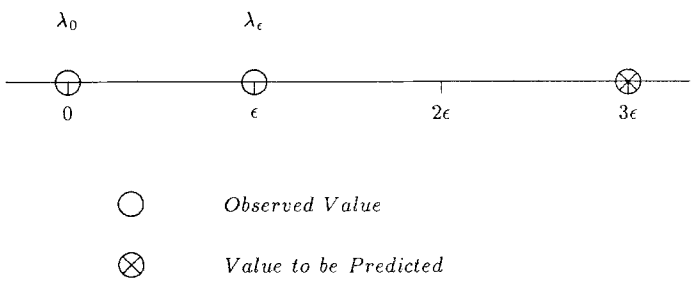

Fig. 2. The kriging situation for Example 2. $Z(3\epsilon)$ is to be predicted on the basis of observations at 0 and ϵ. Weights given $Z(0)$ and $Z(\epsilon)$ are λ_0 and λ_ϵ, respectively.

the particular cases $\epsilon = 0.1, 0.01, 0.001$ (Table 3) show that both kriging weights and prediction variances are similar and converge rapidly. Note that, for small, x,

$$\frac{\gamma_{1/2}(x)}{\gamma_1(x)} = 1 + \frac{x}{4} - \frac{x^2}{64} + O(x^3)$$

so that $2\gamma_{1/2}(x) \notin N_1(\gamma_1)$. Although $2\gamma_{1/2}$ is not "close" to γ_1 in the DA sense, they provide similar predictions. On the other hand, when the region of interest is the interval $[0, 3\epsilon]$, $\gamma_{1/2}(x) \in N_{3\epsilon/4}(\gamma_1)$, so that the DA approach suggests $\gamma_{1/2}$ and γ_1 are "closer" than $2\gamma_{1/2}$ and γ_1. The kriging weights using $\gamma_{1/2}$ will be the same as for $2\gamma_{1/2}$ whereas the presumed prediction variance will be about half its actual value under γ_1.

Interestingly, the condition number $\kappa(\Gamma) = 2^{1/2}\epsilon^{-2} + 2^{-1/2}\epsilon^{-1} + O(1)$ as $\epsilon \downarrow 0$. Hence, the necessary condition [DA equation (12)],

$$\delta\kappa(\Gamma) < 1 \tag{13}$$

Table 2. A Comparison Between the True Variogram and a Misspecified Variogram

Quantity	Presumed variogram	
	Correct, $[\gamma_1(x)]$	Incorrect, $[2\gamma_{1/2}(x)]$
Variogram	$x^2 - x^3/2 + x^4/48 + O(x^5)$	$2x^2 - x^3/2 - 23x^4/96 + O(x^5)$
λ_0	$-2 + 7\epsilon/2 + 9\epsilon^2/8 + O(\epsilon^3)$	$-2 + 7\epsilon/4 + 49\epsilon^2/16 + O(\epsilon^3)$
λ_ϵ	$3 - 7\epsilon/2 - 9\epsilon^2/8 + O(\epsilon^3)$	$3 - 7\epsilon/4 - 49\epsilon^2/16 + O(\epsilon^3)$
Presumed prediction variance	$24\epsilon^3 - 29\epsilon^4 + O(\epsilon^5)$	$24\epsilon^3 + 79\epsilon^4/4 + O(\epsilon^5)$
Actual prediction variance	$24\epsilon^3 - 29\epsilon^4 + O(\epsilon^5)$	$24\epsilon^3 - 183\epsilon^4/8 + O(\epsilon^5)$

Note: Kriging weights $(\lambda_0, \lambda_\epsilon)$, and presumed prediction variance are determined relative to the presumed variogram for the kriging situation given in Fig. 2. The actual prediction variance is relative to the correct variogram, γ_1. All entries are in terms of the distance ϵ, as $\epsilon \downarrow 0$.

Table 3. Numerical Values of Prediction Variances

	Presumed based on the model		Actual based on the model
ϵ	Correct, (γ_1)	Incorrect, $(2\gamma_{1/2})$	Incorrect, $(2\gamma_{1/2})$
0.1	2.0821×10^{-2}	2.6739×10^{-2}	2.1107×10^{-2}
0.01	2.3706×10^{-5}	2.4383×10^{-5}	2.3765×10^{-5}
0.001	2.3971×10^{-8}	2.4039×10^{-8}	2.3977×10^{-8}

Note: Each entry gives the variance of the prediction error when predicting $Z(3\epsilon)$ for the situation described in Fig. 2. Refer to Table 2 and Example 2 for definitions of $2\gamma_{1/2}$ and γ_1. Actual prediction variances under $2\gamma_{1/2}$ are relative to the correct variogram, γ_1. Actual prediction variances under γ_1 are, of course, the same as the presumed values given in the second column.

is not satisfied for ϵ sufficiently small. This could be interpreted as indicating that $\gamma_{1/2}$ is not "close" to γ_1 for sufficiently small ϵ. For this specific prediction problem (Tables 2 and 3), this is not the case. This emphasizes the essential role that locations of observations play in interpretation of these neighborhoods. Moving the observations closer together increases the condition number, so that Eq. (13) will not be satisfied for ϵ sufficiently small. As this example indicates, predictions under two variograms can be similar even when Eq. (13) is not satisfied.

Example 3

Consider a stochastic process $Z(\cdot)$ defined on $[0, 1]$ with constant unknown mean and variogram $\gamma_0(x) = x$. Suppose observations are taken at 0, $1/N$, $2/N$, $\cdots$, $(N-1)/N$ and the value of $Z(1)$ is to be predicted. The kriging predictor is

$$\hat{Z}_1 = Z\left(\frac{N-1}{N}\right)$$

Suppose that γ_0 is misspecified by

$$\gamma_\alpha(x) = x - \alpha x^2/2 \qquad \text{for} \quad 0 \le x \le 1$$

where $0 \le \alpha \le 1$ is given. Under γ_α, the kriging predictor is

$$\hat{Z}_2 = (1-\beta)Z\left(\frac{N-1}{N}\right) + \beta Z(0)$$

where $\beta = \alpha/[\alpha + N(2-\alpha)]$. By Theorem 13 (Ibragimov and Rozanov 1978, p. 99), γ_0 is compatible with γ_α on $[0, 1]$. Hence, asymptotically similar predictions can be expected. As $N \to \infty$, $\beta \downarrow 0$ so that $\hat{Z}_2$ becomes decreasingly dependent on the relatively distant $Z(0)$. Predictions under γ_0 and γ_α (Table 4) clearly converge quickly for any $0 \le \alpha \le 1$. All of these results are consistent

Table 4. A Comparison Between the True Variogram and a Misspecified Variogram

	Actual covariance	
Presumed covariance	Linear (γ_0)	Quadratic (γ_α)
Linear	$2/N$	$2/N - \alpha/N^2$
Quadratic	$\dfrac{2(2-\alpha)^2N - 2\alpha^2 + 8\alpha}{(2-\alpha)^2N^2 + 2\alpha(2-\alpha)N + \alpha^2}$ $= \dfrac{2}{N} - \dfrac{2\alpha}{(2-\alpha)} \cdot \dfrac{1}{N^2} + O\left(\dfrac{1}{N^3}\right)$	$\dfrac{2(2-\alpha)}{(2-\alpha)N + \alpha}$ $= \dfrac{2}{N} + \dfrac{2\alpha^2}{(2-\alpha)^2} \cdot \dfrac{1}{N^2} + O\left(\dfrac{1}{N^3}\right)$

Note: Each entry gives the variance of prediction error when predicting $Z(1)$ for the situation described in Example 3. The presumed covariance function is used to compute the kriging predictor [see Eq. (2)], and the actual covariance function is used to compute the variance of the error of that predictor. Values in the last row are asymptotic expansions, to allow easy comparisons.

with the compatibility paradigm. Using the DA neighborhood paradigm, $\gamma_\alpha \in N_{\alpha/2}(\gamma_0)$ and vice versa. The DA bounds for differences in kriging weights and prediction variances are compared to the actual differences for $\alpha = 0.0002$ and $N = 2, 5, 10, 25, 45$ (Table 5). This small value of α is chosen so that Eq. (13) is satisfied and the DA bounds are defined. In this situation, the asymptotics give a good approximation for as few as two points (columns 3 and 5). The actual differences are based on the formulas

$$\|\Delta X\|^2 = 2\beta^2 + \frac{\alpha^2}{4N^2} \approx \frac{\alpha^2(\alpha^2 - 4\alpha + 12)}{4(2 - \alpha)^2 N^2}$$

$$\sigma^2_{\gamma_\alpha} - \sigma^2_{\gamma_o} = \frac{2\alpha^2(N - 1)}{(2 - \alpha)^2 N^3 + 2\alpha(2 - \alpha)N^2 + \alpha^2 N} \approx \frac{2\alpha^2}{(2 - \alpha)^2 N^2}$$

using the notation of DA. Note that $\|\Delta X\|$ goes to zero like $1/N$ and $\sigma^2_{\gamma_\alpha} - \sigma^2_{\gamma_o}$ goes to zero like $1/N^2$ as $N \to \infty$. The comparison between bounds and actual differences should be made for fixed locations (i.e., fixed N). This table indicates that, for any particular value of N, the bounds are so conservative as to be meaningless and, furthermore, as N grows, the bounds tend to increase whereas the actual differences decrease rapidly.

The bounds, of course, are correct and could possibly be attained by some member of $N_{\alpha/2}(\gamma_0)$ for some set of observations and quantity to be predicted; the difficulty lies with the choice of δ-neighborhood definition. As is noted by DA, these bounds can be extremely conservative. What is clear from this example, however, is that the bounds indicate the situation gets worse as N increases, whereas, in fact, the situation rapidly improves. Hence, the DA neighborhood paradigm, by attempting to include all possible variograms in the δ-neighborhood, does not describe the situation correctly. It does not represent a meaningful partition of the variogram models; members of different δ-neighborhoods easily can produce similar predictions (e.g., Examples 2 and 3) and the bounds obtained for members of the same δ-neighborhood easily can be

Table 5. Numerical Comparison of Diamond and Armstrong Bounds to Actual Differences

	$\|\Delta X\|$		$\lvert\sigma^2_{\gamma_{0.0002}} - \sigma^2_{\gamma_0}\rvert$	
N	Bound	Actual value	Bound	Actual value
2	6.747×10^{-4}	7.501×10^{-9}	1.633×10^{-3}	2.500×10^{-9}
5	1.604×10^{-2}	1.200×10^{-9}	4.086×10^{-2}	6.401×10^{-10}
10	4.089×10^{-2}	3.000×10^{-10}	1.351×10^{-1}	1.800×10^{-10}
25	4.562×10^{-1}	4.801×10^{-11}	2.304	3.073×10^{-11}
45	5.871	1.482×10^{-11}	3.954×10^{1}	9.659×10^{-12}

Note: Refer to Table 4 and Example 3 for definitions of bounds and method of calculating exact actual differences.

misleading. A more meaningful neighborhood definition is the neighborhood of all variograms compatible with the given variogram. Hence, when estimating variograms, a procedure should be used that is able to distinguish strongly between variograms that are not compatible and choose a variogram that is "nearly" compatible with the true variogram.

Explanation of Examples from Warnes (1986)

Warnes (1986) has investigated how prediction surfaces change when the variogram model is perturbed. His results (p. 665–669) can be explained easily in terms of compatibility. Let $\gamma_e(x;\ \alpha,\ \sigma^2) = \sigma^2(1 - e^{-x/a})$ denote the exponential variogram class. Then, $\gamma_e(x;\ 2,\ \sigma_1^2)$ is compatible with $c\gamma_e(x;\ a,\ \sigma_2^2)$ if and only if $c = (a\sigma_1^2)/(2\sigma_2^2)$. Warnes finds that the predicted surface for an exponential variogram with $a = 2$ is close to the predicted surface for an exponential variogram with other values of a. Indeed, the compatibility results in the second section show that predictions are asymptotically the same and prediction error variances will be approximate multiples of each other.

Let $\gamma_G(x;\ a,\ \sigma^2) = \sigma^2(1 - e^{-x^2/a^2})$ denote the so-called "Gaussian" variogram class. Then, $\gamma_G(x;\ a_1,\ \sigma_1^2)$ is compatible with $c\gamma_G(x;\ a_2,\ \sigma_2^2)$ if and only if $a_1 = a_2$ and $c = \sigma_1^2/\sigma_2^2$ (Ibragimov and Rozanov, 1978, p. 95). Gaussian models with different range parameters cannot be made compatible by multiplication by a constant, so Eq. (5) does not apply for Gaussian models with different range parameters. Warnes finds the predicted surface for the Gaussian with $a = 2$ and compares it to models with a varied. He finds that even small changes in a produce substantially different prediction surfaces, which is not surprising in light of the incompatibility of these models. Finally, Warnes compares the exponential model with $a = 2$ to the Gaussian with $a = 2$. As these models are always incompatible and because one corresponds to a nondifferentiable field and the other to an infinitely differentiable field, the predicted surfaces are vastly different.

CONCLUSIONS

The effect of using an incorrect covariance function on kriging has been investigated by an asymptotic approach in which the number of observations in some fixed bounded region tends to infinity. Stein (1988) showed that the impact of using an incorrect covariance function is asymptotically negligible if the covariance function used in the kriging procedure is compatible with the true covariance function. These general results show that the commonly used spherical covariance function is an inappropriate model for most three-dimensional fields. The perturbation paradigm of Diamond and Armstrong (1984) is shown to result in substantially different conclusions than the paradigm based on compatibility considerations. Finally, the compatibility approach is used to clarify some examples of Warnes (1986).

ACKNOWLEDGMENTS

This manuscript was prepared using computer facilities supported in part by National Science Foundation (NSF) Grant DMS-8404941 to the Department of Statistics at the University of Chicago. M. L. S. was supported in part by a NSF Mathematical Sciences Postdoctoral Research Fellowship. M. S. H. acknowledges gratefully the support of IBM through the 1986–88 IBM Predoctoral Fellowship.

REFERENCES

Armstrong, M. and Myers, D., 1984, Robustness of Kriging; Variogram Modelling and its Applications: Technical Report, Centre de Geostatistique, Fontainebleau, France.

Delfiner, P., 1976, Linear Estimation of Non-Stationary Spatial Phenomena: p. 49–68 *in* M. Guarascio, M. David, and C. Huijbregts (Eds.), Advanced Geostatistics in the Mining Industry: Reidel, Dordrecht.

Diamond, P. and Armstrong, M., 1984, Robustness of Variograms and Conditioning of Kriging Matrices: Math. Geol., v. 16, p. 809–822.

Doob, J. L., 1953, Stochastic Processes: John Wiley, New York, 654 p.

Goldberger, A., 1962, Best Linear Unbiased Prediction in the Generalized Linear Regression Model: J. Amer. Stat. Assoc., v. 57, p. 369–375.

Hajek, J., 1958, On a Property of Normal Distributions of an Arbitrary Stochastic Process: Czech. Math. J., v. 8, p. 610–618.

Ibragimov, I. A. and Rozanov, Y. A., 1978, Gaussian Random Processes: (trans. Aries, A. B.), Springer-Verlag, New York, 275 p.

Journel, A. G. and Huijbregts, C. J., 1978, Mining Geostatistics: Academic Press, London, 600 p.

Kitanidis, P. K., 1983, Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Applications: Water Resour. Res., v. 19, p. 909–921.

Krasnitskii, S. M., 1973, On Conditions of Equivalence and Perpendicularity of Measures Corresponding to Homogeneous Gaussian Fields: Theory Probab. Its Appl., v. 18, p. 588–592.

Matheron, G., 1973, The Intrinsic Random Functions and their Applications: Adv. Appl. Probab., v. 5, p. 437–468.

Rendu, J-M., 1978, An Introduction to Geostatistical Methods of Mineral Evaluation: Monogr. S. Afr. Inst. Mining Metall., 100 p.

Ripley, B. D., 1981, Spatial Statistics: John Wiley, New York, 252 p.

Rozanov, Y. A., 1968, Infinite Dimensional Gaussian Distributions: Proceedings of the Steklov Institute of Mathematics, Amer. Math. Soc. v. 108, p. 1–136.

Skorokhod, A. V. and Yadrenko, M. I., 1973, On Absolute Continuity of Measures Corresponding to Homogeneous Gaussian Fields: Theory Probab. Its Appl., v. 8, p. 27–40.

Stein, M., 1988, Asymptotically Efficient Spatial Interpolation with a Misspecified Covariance Function: Ann. Stat., v. 16, p. 55–63.

Sukhatme, S., 1985, Kriging with Perturbed Variogram: Proceedings of the Social Statistics Section of the American Statistical Association Annual Meetings, p. 296–299.

Warnes, J. J., 1986, A Sensitivity Analysis for Universal Kriging: Math. Geol., v. 18, p. 653–676.

Yaglom, A. M., 1962, An Introduction to the Theory of Stationary Random Functions: Prentice-Hall, New Jersey, 235 p.

Yakowitz, S. J. and Szidarovszky, F., 1985, A Comparison of Kriging with Nonparametric Regression Methods: J. Multivariate Anal. v. 16, p. 21–53.