The Annals of Statistics
1988, Vol. 16, No. 1, 55–63

ASYMPTOTICALLY EFFICIENT PREDICTION OF A RANDOM FIELD WITH A MISSPECIFIED COVARIANCE FUNCTION[1]

BY MICHAEL L. STEIN

University of Chicago

Best linear unbiased predictors of a random field can be obtained if the covariance function of the random field is specified correctly. Consider a random field defined on a bounded region R. We wish to predict the random field $z(\cdot)$ at a point x in R based on observations $z(x_1), z(x_2), \ldots, z(x_N)$ in R, where $\{x_i\}_{i=1}^{\infty}$ has x as a limit point but does not contain x. Suppose the covariance function is misspecified, but has an equivalent (mutually absolutely continuous) corresponding Gaussian measure to the true covariance function. Then the predictor of $z(x)$ based on $z(x_1), \ldots, z(x_N)$ will be asymptotically efficient as N tends to infinity.

1. Introduction. Kriging is a method of prediction for spatial processes that is popular in mining, hydrology and petroleum engineering [Journel and Huijbregts (1978)]. Basically, the kriging predictor is just the best linear unbiased predictor of a random field and, thus, kriging is the same as optimal linear prediction [Wiener (1949) and Kolmogorov (1941)]. Specifically, consider a random field $z(\cdot)$ satisfying

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

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

$$\operatorname{Cov}(z(x), z(y)) = K(x, y).$$

If we observe $z(x_1), \ldots, z(x_N)$ and wish to predict $z(x)$, an unbiased (for all possible β) linear predictor will be of the form

$$\hat{z}(x) = \sum_{i=1}^{N} \lambda_i z(x_i),$$

where

$$\sum_{i=1}^{N} \lambda_i f(x_i) = f(x).$$

If $K(\cdot,\cdot)$ is specified, among all such predictors (assuming one exists), we can obtain the linear unbiased predictor that minimizes $\operatorname{Var}(z(x) - \hat{z}(x))$; this is the kriging predictor. It is given by [Goldberger (1962)]

$$\left\{c'C^{-1} + \left(f(x) - FC^{-1}c\right)'\left(FC^{-1}F'\right)^{-1}FC^{-1}\right\}\left(z(x_1), \ldots, z(x_N)\right)', \tag{2}$$

Received January 1986; revised October 1986.

[1]Manuscript prepared using computer facilities supported in part by National Science Foundation Grant MCS-84-04941 to the Department of Statistics at the University of Chicago.

AMS 1980 *subject classifications*. Primary 62M20; secondary 60G30, 60G60.

Key words and phrases. Kriging, mutual absolute continuity of Gaussian measures.

assuming all inverses exist, where

$$F = (f(x_1), \ldots, f(x_N)),\ c = (K(x, x_1), \ldots, K(x, x_N))'$$

and C is an $N \times N$ matrix with ijth element $K(x_i, x_j)$. In practice, the covariance function is not specified and must be estimated from the observations.

The effect on kriging of using a misspecified covariance function has been investigated by Diamond and Armstrong (1984) and Yakowitz and Szidarovszky (1985). In particular, this second paper considers the following problem: Let R be a bounded subset of $\mathbb{R}^n$ and $\{x_i\}_{i=1}^{\infty}$ a sequence of points in R that have $x \in R$ as a limit point. If $z(x)$ is predicted on $z(x_1), \ldots, z(x_N)$ assuming $\mathrm{Cov}(z(x), z(y)) = K(x - y)$ is the covariance function and $f(x) \equiv 1$, then (2) yields a consistent predictor of $z(x)$ even if the covariance function is incorrect, as long as the observed process is continuous almost surely, and there exist positive numbers q and D such that

$$\liminf |\omega|^q S(\omega) \geq D \quad \text{as } |\omega| \to \infty,$$

where $S(\cdot)$ is the spectral density for $K(\cdot)$ [Yakowitz and Szidarovszky (1985)].

In this paper, we will show that one obtains not only a consistent predictor of $z(x)$, but an asymptotically efficient predictor, as long as the specified and actual covariance functions are *compatible* on R. We will say that two covariance functions K_0 and K_1 are *compatible* on R if the probability measures P_0 and P_1 of two Gaussian processes on R with equal mean functions and covariance functions K_0 and K_1, respectively, are mutually absolutely continuous. This definition of compatibility will be used whether or not the observed process is actually Gaussian. We will also show that one obtains a value for the variance of the prediction error that has an asymptotically negligible error if K_1 is assumed to be the covariance function when K_0 is the actual covariance function, as long as K_0 and K_1 are compatible. While these results make use of the notion of mutual absolute continuity of Gaussian measures in order to define "compatible" covariance functions, the results in no way depend on the random field being Gaussian.

For stationary covariance functions ($K(x, x + h) = K(h)$), if $K_0(\cdot)$ and $K_1(\cdot)$ behave "similarly" at the origin, they will commonly be compatible, although exceptions exist. Thus, the intuitively sensible concept that it is sufficient to specify the covariance function well only over relatively short distances to obtain a nearly optimal predictor now has mathematical support in the context of prediction of spatial processes. For example, if $K_1(\cdot)$ has $2m$ derivatives on $(-T, T)$ and $K_1^{(2m+1)}(0^+) \neq 0$ [so that the Gaussian process $z(\cdot)$ with covariance function $K_1(\cdot)$ has m, but not $m + 1$ derivatives almost surely], then for $K_0(\cdot)$ and $K_1(\cdot)$ to be compatible on $[0, T]$, $K_0(\cdot)$ must also have $2m$ derivatives on $(-T, T)$ and $K_0^{(2m+1)}(0^+) = K_1^{(2m+1)}(0^+)$. This result follows by noting that under P_1, $z^{(m)}(\cdot)$ exists almost surely and has covariance function $(-1)^m K_1^{(2m)}(\cdot)$ and then applying (1.11) of Ibragimov and Rozanov [(1978), page 67] to $z^{(m)}(\cdot)$. In many common cases, two covariance functions satisfying this property will be compatible; for example, two covariance functions for autoregressive processes of

the same order are mutually absolutely continuous on $[0, T]$ for all finite T if they are properly normalized, which is a special case of Theorem 13 of Ibragimov and Rozanov [(1978), Chapter 3]. These are exceptions to this rule: If

$$K_0(x) = e^{-|x|} \quad \text{and} \quad K_1(x) = \max(0, 1 - |x|),$$

then $K_0(\cdot)$ and $K_1(\cdot)$ are compatible on $[0, T]$ if and only if $T \leq 1$ [Ibragimov and Rozanov (1978), page 100], despite the fact that $K_0'(0^+) = K_1'(0^+)$.

General necessary and sufficient conditions for mutual absolute continuity of Gaussian measures are given for stationary processes in one dimension in Ibragimov and Rozanov [(1978), Chapter 3] and in higher dimensions in Skorokhod and Yadrenko (1973). The examples in Krasnitskii (1979) show that these conditions can be difficult to apply in practice for stationary processes in more than one dimension. Fortunately, we will not need to use any specific properties of mutually absolutely continuous Gaussian measures, because our results will follow essentially from definitions.

In Section 2, the main results on asymptotically efficient prediction with a misspecified covariance function are stated and proved. In Section 3, the connection between these results and kriging with an estimated covariance function is briefly discussed.

2. Main results. Consider a continuous random field $z(\cdot)$ defined on a bounded region R in $\mathbb{R}^n$. Suppose, on R, that $z(\cdot)$ has a mean function as given in (1) and bounded second moments. Let x be a point in R and $\{x_i\}_{i=1}^{\infty}$ a sequence of points in R that has x as a limit point but does not contain x. An alternative asymptotic approach would be to "fix" the distance between neighboring observations and allow the size of the observation region to grow with the number of observations, as is done in Mardia and Marshall (1984). There are both practical and theoretical reasons for considering the approach taken here. For example, we may be interested in predicting a mineral concentration in a fixed region, and we wish to investigate the effect of increasing the number of measurements taken in this region on our ability to predict. Also, this approach allows us to handle the common situation in which the covariance function is not near zero at distances a substantial fraction of the dimensions of the observation region. Finally, the model of an increasing number of observations in a fixed region is used by Stone (1977) in the context of nonparametric regression and, thus, taking the same approach here allows us to compare kriging and nonparametric regression, as is done in Yakowitz and Szidarovszky (1985).

Let us use $\hat{z}_i(N)$ to denote the best (minimum variance) linear unbiased predictor of $z(x)$ based on $z(x_1), \ldots, z(x_N)$, assuming $K_i(\cdot, \cdot)$ is the actual covariance function of the process. Let us also define $e_i(N) = \hat{z}_i(N) - z(x)$, the error of this predictor. Denote $E_i(\cdot)$, $V_i(\cdot)$ and $\mathrm{Cov}_i(\cdot, \cdot)$ to be the expectation, variance and covariance under P_i, the measure of a Gaussian process with mean $\beta' f(x)$ and covariance function $K_i(\cdot, \cdot)$. We have the result:

THEOREM 1. *Suppose $K_0(\cdot, \cdot)$ and $K_1(\cdot, \cdot)$ are two compatible covariance functions on a bounded set R. Let x be a point in R and $\{x_i\}_{i=1}^{\infty}$ a sequence of*

points in R not containing x but having x as a limit point. Assume

$$\lim_{N\to\infty} V_0(e_0(N)) = 0. \tag{3}$$

Then

$$\lim_{N\to\infty} \frac{V_0(e_0(N))}{V_0(e_1(N))} = 1. \tag{4}$$

We will say that a predictor $\hat{z}_1(N)$ is asymptotically efficient if (4) is satisfied. Then Theorem 1 says that if we base our predictor of $z(x)$ on $K_1(\cdot,\cdot)$ when, in fact, $K_0(\cdot,\cdot)$ is the actual covariance function, then as long as $K_0(\cdot,\cdot)$ and $K_1(\cdot,\cdot)$ are compatible on R, we will obtain an asymptotically efficient predictor of $z(x)$, assuming (3) holds. Equation (3) just says that the sequence of best linear unbiased predictors under $K_0(\cdot,\cdot)$ is consistent. Implicit in this assumption is that a linear unbiased predictor of $z(x)$ exists for some value of N. Since x is a limit point of $\{x_i\}_{i=1}^{\infty}$, (3) is met under quite weak conditions on $K_0(\cdot,\cdot)$ and the mean function. For example, if $f(x) \equiv 1$ and $K_0(\cdot,\cdot)$ is bounded and continuous on $R \times R$, we can easily show that (3) is satisfied [follows by the argument for Theorem 2.1 of Yakowitz and Szidarovszky (1985)]. Under weak conditions, (3) also holds with more general mean functions when the covariance function is stationary [Yakowitz and Szidarovszky (1985)].

PROOF OF THEOREM 1. Let us assume, for now, that $z(\cdot)$ is a Gaussian process and let P_i be the measure of a Gaussian process on R with mean function as given in (1) and covariance function $K_i(\cdot,\cdot)$. Clearly,

$$V_0(e_0(N))/V_0(e_1(N)) \le 1$$

for all N, since $\hat{z}_0(N)$ is optimal under P_0. Also,

$$\frac{V_0(e_0(N))}{V_0(e_1(N))} = \frac{V_0(e_0(N))}{V_1(e_0(N))} \frac{V_1(e_0(N))}{V_1(e_1(N))} \frac{V_1(e_1(N))}{V_0(e_1(N))}. \tag{5}$$

Now, $V_1(e_0(N))/V_1(e_1(N)) \ge 1$ for all N, so it suffices to show

$$\liminf_{N\to\infty} \frac{V_0(e_0(N))}{V_1(e_0(N))} \ge 1, \tag{6}$$

since this implies

$$\liminf_{N\to\infty} \frac{V_1(e_1(N))}{V_0(e_1(N))} \ge 1.$$

Define

$$y(N) = e_0(N)/(V_0(e_0(N)))^{1/2},$$

so that $E_0 y(N)^2 = 1$ for all N. The crux of the proof is contained in

LEMMA. *Any subsequence* $N_1, N_2, \ldots$ *contains a further subsequence* $N_{k(1)}, N_{k(2)}, \ldots$ *satisfying*

$$M^{-1} \sum_{l=1}^{M} y\left(N_{k(l)}\right)^2 \to 1 \quad \text{as } M \to \infty$$

in probability under P_0.

PROOF. Trivially, for any $k(\cdot)$,

$$E_0 M^{-1} \sum_{l=1}^{M} y\left(N_{k(l)}\right)^2 = 1 \quad \text{for all } M.$$

Hence, it suffices to find $k(\cdot)$ such that

$$V_0\left(M^{-1} \sum_{l=1}^{M} y\left(N_{k(l)}\right)^2 \right) \to 0 \quad \text{as } M \to \infty. \tag{7}$$

Now, using the assumption that $z(\cdot)$ is a Gaussian process and $e_0(N)$ has mean zero, we obtain

$$\begin{aligned} V_0\left(M^{-1} \sum_{l=1}^{M} y\left(N_{k(l)}\right)^2 \right) &= M^{-2} \sum_{l=1}^{M} V_0\left(y\left(N_{k(l)}\right)^2 \right) \\ &\quad + 2M^{-2} \sum_{m=2}^{M} \sum_{l=1}^{m-1} \operatorname{Cov}_0\left(y\left(N_{k(m)}\right)^2, y\left(N_{k(l)}\right)^2 \right) \\ &= 2M^{-1} + 4M^{-2} \sum_{m=2}^{M} \sum_{l=1}^{m-1} \left\{ E_0\left(y\left(N_{k(m)}\right) y\left(N_{k(l)}\right) \right) \right\}^2. \end{aligned} \tag{8}$$

By definition,

$$\begin{aligned} &E_0\left\{ y\left(N_{k(m)}\right) y\left(N_{k(l)}\right) \right\} \\ &\quad = E_0\left\{ e_0\left(N_{k(m)}\right) e_0\left(N_{k(l)}\right) \right\} / \left\{ V_0\left(e_0\left(N_{k(m)}\right) \right) V_0\left(e_0\left(N_{k(l)}\right) \right) \right\}^{1/2}. \end{aligned}$$

For $l < m$,

$$\begin{aligned} &E_0\left\{ e_0\left(N_{k(m)}\right) e_0\left(N_{k(l)}\right) \right\} \\ &\quad = E_0\left[e_0\left(N_{k(m)}\right) \left(e_0\left(N_{k(l)}\right) - e_0\left(N_{k(m)}\right) \right) + e_0\left(N_{k(m)}\right)^2 \right] \\ &\quad = E_0\left[e_0\left(N_{k(m)}\right) \left(e_0\left(N_{k(l)}\right) - e_0\left(N_{k(m)}\right) \right) \right] + V_0\left(e_0\left(N_{k(m)}\right) \right) \\ &\quad = V_0\left(e_0\left(N_{k(m)}\right) \right), \end{aligned}$$

since, under P_0, $e_0(N_{k(m)})$ is uncorrelated with $e_0(N_{k(l)}) - e_0(N_{k(m)})$. This follows from the projection property of best linear unbiased predictors, which says that $e_0(N)$ is uncorrelated with

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

under P_0 if

$$\sum_{i=1}^{N} \lambda_i f(x_i) = 0;$$

see Journel (1977), for example. Thus, for $l < m$,

$$E_0\{y(N_{k(l)})y(N_{k(m)})\} = \left[V_0(e_0(N_{k(m)}))/V_0(e_0(N_{k(l)}))\right]^{1/2}.$$

By (3), we can choose $k(\cdot)$ so that

$$V_0(e_0(N_{k(l+1)}))/V_0(e_0(N_{k(l)})) \leq 1/2, \quad \text{for all } l,$$

for example. Substituting this inequality into (8) establishes (7) and, hence, the lemma. □

Now, let us suppose (6) is untrue. Then there exists a subsequence $N_1, N_2, \ldots$ satisfying

$$\lim_{l\to\infty} \frac{V_0(e_0(N_l))}{V_1(e_0(N_l))} = c < 1. \tag{9}$$

Equivalently,

$$V_1(y(N_l)) \to c^{-1}.$$

We note that we must have $c > 0$, since $c = 0$ would contradict the mutual absolute continuity of P_0 and P_1 [see Ibragimov and Rozanov (1978), page 70, formula (1.27)]. Now pick a further subsequence such that the lemma holds. We have

$$\lim_{M\to\infty} E_1\left\{M^{-1}\sum_{l=1}^{M} y(N_{k(l)})^2\right\} = c^{-1}. \tag{10}$$

Since P_0 and P_1 are mutually absolutely continuous, by the lemma we must have

$$M^{-1}\sum_{l=1}^{M} y(N_{k(l)})^2 \to 1 \quad \text{in probability under } P_1. \tag{11}$$

As $M \to \infty$, we also have, using the fact that $z(\cdot)$ is Gaussian,

$$\begin{aligned} V_1\left(M^{-1}\sum_{l=1}^{M} y(N_{k(l)})^2\right) &= 2M^{-2}\sum_{m=1}^{M}\sum_{l=1}^{M}\left\{\mathrm{Cov}_1(y(N_{k(l)}), y(N_{k(m)}))\right\}^2 \\ &\leq 2M^{-2}\sum_{m=1}^{M}\sum_{l=1}^{M} V_1(y(N_{k(l)}))V_1(y(N_{k(m)})) \to 2c^{-2}, \end{aligned}$$

where the inequality is by Cauchy–Schwarz. Applying Theorem 4.5.2 in Chung (1974) to (11), we have

$$\lim_{M\to\infty} E_1\left\{M^{-1}\sum_{l=1}^{M} y(N_{k(l)})^2\right\} = 1.$$

Thus, we have a contradiction to (10) and the theorem follows when $z(\cdot)$ is Gaussian. Now suppose $z(\cdot)$ is a non-Gaussian process on R with true covariance function K_0, but we use the incorrect but compatible covariance function K_1 in (2). Consider a Gaussian process $z^*(\cdot)$ on R with the same mean function and true covariance function K_0 as $z(\cdot)$. Again we use the incorrect but compatible covariance function K_1 in (2). We observe both $z(\cdot)$ and $z^*(\cdot)$ at the same sequence of points in R. If (3) holds for $z^*(\cdot)$, then we have just shown that (4) follows for $z^*(\cdot)$. Since (3) and (4) are only statements about the first two moments of a process, (3) and (4) must also hold for the non-Gaussian process $z(\cdot)$. We note that this proof has used properties of Gaussian processes. The use of such properties is relevant to non-Gaussian processes because we have defined compatibility of covariance functions in terms of mutual absolute continuity of corresponding Gaussian measures whether or not the process is Gaussian. □

From (4) and (5), we have

$$1 = \lim\left\{\frac{V_0(e_0(N))}{V_0(e_1(N))}\,\frac{V_1(e_1(N))}{V_1(e_0(N))}\right\}$$

$$= \lim\left\{\frac{V_0(e_0(N))}{V_1(e_0(N))}\,\frac{V_1(e_1(N))}{V_0(e_1(N))}\right\}.$$

Applying (6), we see that we must have

THEOREM 2. *Under the same conditions as in Theorem* 1,

$$\lim_{N\to\infty}\frac{V_0(e_0(N))}{V_1(e_0(N))} = 1.$$

That is, if we assume $K_0(\cdot,\cdot)$ is the covariance function when, in fact, $K_1(\cdot,\cdot)$ is the covariance function, we not only get an asymptotically efficient predictor of $z(x)$, we also get a value for the variance of the prediction error whose relative error is asymptotically negligible. As with Theorem 1, Theorem 2 is valid whether or not $z(\cdot)$ is Gaussian.

In many applications we are interested in predicting some linear functional of the $z(\cdot)$ field, for example,

$$\int_R z(x)\,dx. \tag{12}$$

We can use kriging to predict such a linear functional, and Theorems 1 and 2 will apply as long as the variance of the prediction error under K_0 converges to zero (3). For predicting the quantity in (12), (3) will be satisfied if the best linear unbiased predictor of $z(x)$ is consistent for all x in R, since

$$\operatorname{Var}\left(\int_R(\hat{z}_N(x) - z(x))\,dx\right) \le \left(\int_R[\operatorname{Var}(\hat{z}_N(x) - z(x))]^{1/2}\,dx\right)^2$$

by the Cauchy–Schwarz inequality. As noted earlier, if $\{x_i\}_{i=1}^{\infty}$ has x as a limit

point, $f(x) \equiv 1$ and $K_0(\cdot, \cdot)$ is bounded on $R \times R$, $\hat{z}_N(x)$ will be consistent. Thus, in this case, if $\{x_i\}_{i=1}^{\infty}$ is dense in R, (3) will be satisfied for the expression in (12).

3. Discussion. We have considered the effect of misspecifying the covariance function on the performance of kriging. In practice, the covariance function is estimated from the data. Thus, if we have a large number of observations in a bounded region R, we see that it is important to obtain an estimated covariance function that is "nearly" compatible to the actual covariance function, and it is unimportant to distinguish between covariance functions that are compatible.

As an example, suppose

$$K(x; \theta) = \theta_1 e^{-\theta_2 |x|} \quad .$$

is our model for a covariance function on $[0, 1]$, where θ_1 and θ_2 are positive. In this case, if $\theta = (\theta_1, \theta_2)$ and $\theta^* = (\theta_1^*, \theta_2^*)$ are two possible values of the parameters, $K(x; \theta)$ and $K(x; \theta^*)$ are compatible if and only if $\theta_1\theta_2 = \theta_1^*\theta_2^*$ [Ibragimov and Rozanov (1978), Chapter 3, Theorem 13]. Theorems 1 and 2 say that, at least asymptotically, for purposes of prediction, the value of $\theta_1\theta_2$ is critical, but the values of θ_1 and θ_2 are not. Kitanidis (1983, 1985), Mardia and Marshall (1984) and Marshall and Mardia (1985) investigate the properties of estimators of θ, although they do not specifically address the problem of an increasing number of observations in a fixed bounded region. The results in Stein (1987) suggest that, for Gaussian processes, there are available techniques that consistently estimate those functions of the parameters of a model for a covariance function that determine compatibility, at least in the case where $K(x; \theta)$ is linear in θ. However, these results, unlike the ones presented here, do depend on the process being Gaussian.

4. Conclusions. We have derived two quite general theorems on the asymptotic efficiency of kriging predictors based on an increasing number of observations in a bounded region when the covariance function is misspecified. The first says that as long as the covariance function used to obtain the kriging predictor is compatible with the actual covariance function, the obtained predictor will have the same asymptotic efficiency as the optimal predictor based on the actual covariance function. The second says that we also obtain a value for the variance of the prediction error that has asymptotically negligible relative error using the incorrect (but compatible with the correct) covariance function.

REFERENCES

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic, New York.

DIAMOND, P. and ARMSTRONG, M. (1984). Robustness of variograms and conditioning of kriging matrices. *J. Internat. Assoc. Math. Geol.* **16** 809–822.

GOLDBERGER, A. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.* **57** 369–375.

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes*. Springer, New York.

JOURNEL, A. (1977). Kriging in terms of projections. *J. Internat. Assoc. Math. Geol.* **9** 563–586.

JOURNEL, A. and HUIJBREGTS, C. (1978). *Mining Geostatistics*. Academic, New York.

KITANIDIS, P. K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Res.* **19** 909–921.

KITANIDIS, P. K. (1985). Minimum-variance unbiased estimation of covariances of regionalized variables. *J. Internat. Assoc. Math. Geol.* **17** 195–208.

KOLMOGOROV, A. N. (1941). Interpolation und Extrapolation von stationären zuffäligen Folgen. *Izv. Akad. Nauk SSSR* **5** 3–14.

KRASNITSKII, S. M. (1979). Some examples of equivalent and orthogonal Gaussian distributions. *Theory Probab. Appl.* **24** 161–165.

MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.

MARSHALL, R. J. and MARDIA, K. V. (1985). Minimum norm quadratic estimation of the components of spatial covariance. *J. Internat. Assoc. Math. Geol.* **17** 517–525.

SKOROKHOD, A. V. and YADRENKO, M. I. (1973). On absolute continuity of measures corresponding to homogeneous Gaussian fields. *Theory Probab. Appl.* **18** 27–40.

STEIN, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *J. Amer. Statist. Assoc.* **82** 765–772.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.

WIENER, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Wiley, New York.

YAKOWITZ, S. and SZIDAROVSZKY, F. (1985). A comparison of kriging with nonparametric regression methods. *J. Multivar. Anal.* **16** 21–53.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637