The Annals of Statistics
1990, Vol. 18, No. 2, 850–872

UNIFORM ASYMPTOTIC OPTIMALITY OF LINEAR PREDICTIONS OF A RANDOM FIELD USING AN INCORRECT SECOND-ORDER STRUCTURE[1]

BY MICHAEL STEIN

The University of Chicago

For a random field $z(t)$ defined for $t \in R \subseteq \mathbb{R}^d$ with specified second-order structure (mean function m and covariance function K), optimal linear prediction based on a finite number of observations is a straightforward procedure. Suppose (m_0, K_0) is the second-order structure used to produce the predictions when in fact (m_1, K_1) is the correct second-order structure and (m_0, K_0) and (m_1, K_1) are "compatible" on R. For bounded R, as the points of observation become increasingly dense in R, predictions based on (m_0, K_0) are shown to be uniformly asymptotically optimal relative to the predictions based on the correct (m_1, K_1). Explicit bounds on this rate of convergence are obtained in some special cases in which $K_0 = K_1$. A necessary and sufficient condition for the consistency of best linear unbiased predictors is obtained, and the asymptotic optimality of these predictors is demonstrated under a compatibility condition on the mean structure.

1. Introduction. This paper investigates the effect of misspecifying the mean and covariance function of a random field on optimal linear predictions of the random field. Optimal linear prediction is commonly used in geology and hydrology, where it is known as kriging [Journel and Huijbregts (1978)]. Consider a random field $z(t)$, $t \in R \subseteq \mathbb{R}^d$, with $Ez(t) = m(t)$ and $\operatorname{cov}(z(t), z(t')) = K(t, t')$. The pair of functions (m, K) defines the second-order structure of the field. If (m, K) is known, then for any finite set of observations, the optimal linear predictor of any unobserved value of $z(\cdot)$ can be readily calculated. For example, if $Z_N = (z(t_1), \ldots, z(t_N))'$, then the optimal (minimum mean-squared error) linear predictor of $z(t_0)$ is

$$Ez(t_0) + \operatorname{cov}\bigl(z(t_0), Z_N'\bigr)\bigl[\operatorname{cov}\bigl(Z_N, Z_N'\bigr)\bigr]^{-1}(Z_N - EZ_N).$$

In practice, (m, K) is at least partially unknown and must be estimated from the data. We will not directly address the effect of using estimated second-order structures on linear prediction. Instead, we will investigate the effect of using a fixed but incorrect second-order structure. This work focuses on the situation where R is a bounded region and $t_1, t_2, \ldots$ is a dense sequence of points in R; the behavior of predictions of $z(\cdot)$ based on $z(t_1), \ldots, z(t_N)$ as N increases is

Received June 1987; revised March 1989.

[1]The author was supported by an NSF Mathematical Sciences Postdoctoral Fellowship. This manuscript was prepared using computer facilities supported in part by NSF Grants DMS-8601732 and DMS-8404941 to the Department of Statistics at The University of Chicago.

AMS 1980 *subject classifications*. Primary 62M20; secondary 41A25, 60G60.

Key words and phrases. Kriging, spatial statistics, approximation in Hilbert spaces.

the object of study. This asymptotic approach of an increasing number of observations over a fixed region is also used by Yakowitz and Szidarovszky (1985) and Stein (1988). Yakowitz and Szidarovszky (1985) show that the kriging predictors of $z(t_0)$, $t_0 \in R$, based on an incorrect covariance function K is mean-square consistent when $z(\cdot)$ is presumed to be second-order stationary with the spectral density $S(\omega)$ of the presumed covariance function K satisfying

$$\liminf_{|\omega| \to \infty} S(\omega)|\omega|^q > 0$$

for some q. In this paper, we show that an asymptotically optimal linear predictor is obtained if the second-order structure used to produce the prediction is incorrect but "compatible" (defined below) with the actual second-order structure. This result can also be obtained using a very general theorem by Blackwell and Dubins (1962) which compares conditional distributions under equivalent measures. We will give a different derivation for the particular case of linear prediction which will enable us to obtain bounds on the rate of convergence to optimality. A rather different approach to the problem of using incorrect second-order structures is taken by Diamond and Armstrong (1984), Sukhatme (1985) and Warnes (1986), who consider the effect on kriging predictions of perturbing the covariance function a small amount given a fixed set of observations.

Corresponding to every second-order structure (m_i, K_i) on R there is a unique Gaussian measure P_i with mean function m_i and covariance function K_i. We will say that (m_0, K_0) and (m_1, K_1) are compatible on R if P_0 and P_1 are mutually absolutely continuous. This definition is a natural extension of the one given by Stein (1988) in which it was assumed $m_0 = m_1$. Compatibility is a property of the second-order structures (m_0, K_0) and (m_1, K_1), and the region R; it is not necessary for the random field $z(\cdot)$ to be Gaussian for the definition to apply. Conditions for the mutual absolute continuity of Gaussian measures in terms of the mean and covariance functions are given for homogeneous covariance functions in one dimension by Ibragimov and Rozanov (1978), Chapter 3, and in higher dimensions by Yadrenko (1983), Chapter 3. By definition, these conditions also allow us to determine the compatibility of second-order structures.

In Section 3, we derive some general asymptotic optimality properties of predictors based on an incorrect second-order structure that is compatible with the actual second-order structure. Suppose (m_0, K_0) and (m_1, K_1) are compatible on R, where (m_1, K_1) is the correct second-order structure, and $t_1, t_2, \ldots$ is dense in R. We consider predictions based on $z(t_1), \ldots, z(t_N)$ as N increases. Then the mean-squared difference between predictions based on (m_0, K_0) and (m_1, K_1) relative to the mean-squared prediction error based on (m_1, K_1) goes to 0 uniformly over all linear functionals of $z(\cdot)$ on R with finite second moments (Corollary 3.1). Moreover, the difference between the mean-squared prediction error obtained by using (m_0, K_0) both to predict and to

evaluate the mean-squared prediction error and the actual mean-squared error of this prediction tends to 0 uniformly relative to the presumed mean-squared prediction error under (m_0, K_0) (Theorem 3.1).

In Section 4, upper bounds on the rates of this uniform convergence are computed under various conditions when $K_0 = K_1$; that is, only the mean function is misspecified. When $z(\cdot)$ is in fact Gaussian, these bounds are related to the variation distance between the conditional distributions of $z(\cdot)$ under the two mean functions. The bounds can be obtained by approximating an element of a Hilbert space by an element of a finite-dimensional subspace. This problem has been considered as it applies to optimal design for estimating the regression coefficients of a stochastic process in one dimension by Sacks and Ylvisaker (1966, 1968, 1970, 1971), Wahba (1971, 1974), Eubank, Smith and Smith (1981), and in two dimensions by Ylvisaker (1975). In this work, we obtain results for less smooth mean functions than are considered in previous works. An order-of-magnitude sharper bound is obtained for differentiable processes when only the value of the process at each observation point is available, and not the derivatives of the process at these points. These results are extended to two dimensions, and apply to a much broader class of processes than considered by Ylvisaker (1975).

In Section 5, two general theorems on asymptotic properties of best linear unbiased predictors are obtained. We suppose $Ez(t) = \beta' g(t)$, where β is an unknown vector of coefficients and $g(t) = (g^1(t) \cdots g^p(t))'$ is a known function. The best linear unbiased predictor has minimum error variance among all linear predictors with expected error of 0 for all values of β; it is given explicitly in (5.2). A necessary and sufficient condition for the mean-squared consistency of a best linear unbiased predictor is given. Under the stronger condition that $(0, K)$ and (g^i, K) are compatible for $i = 1, \ldots, p$, the best linear unbiased predictor is shown to be uniformly asymptotically optimal relative to the best linear predictor (i.e., assuming β is known).

The basic message here is that as the observations in some bounded region get increasingly dense, it becomes inconsequential to distinguish between compatible second-order structures for purposes of linear prediction in that region. While it is not necessarily true that incompatible second-order structures give predictions that are asymptotically different, the examples in Stein and Handcock (1989) show that they can give predictions with very different asymptotic behavior in many cases. Thus, in order to guarantee good linear predictors based on an estimated second-order structure, it would be of value to be able to distinguish between incompatible second-order structures with high probability based on observations within the region of interest. When the random field $z(\cdot)$ is Gaussian, there is some evidence that this can be done. Gaussian measures are either mutually absolutely continuous or orthogonal [Ibragimov and Rozanov (1978), page 77]. Thus, if $z(\cdot)$ is observed everywhere on R, (m_1, K_1) is the correct second-order structure and (m_0, K_0) is an incompatible second-order structure on R, this fact can be determined with probability 1. If $z(\cdot)$ is observed at a large number of locations throughout R, we would thus expect to be able to make this distinction with high probability.

Results of the author (1987, 1989) on the behavior of minimum norm quadratic estimators of the parameters of a model for a covariance function give some support for this position. Thus, a plausible conjecture would be that for Gaussian random fields, those distinctions between second-order structures that are possibly asymptotically important to make for purposes of prediction (i.e., between incompatible second-order structures) are distinctions that can be made well.

2. Compatibility and Hilbert spaces. Suppose $z(t)$, $t \in R$, is a random field with finite second moments and second-order structure (m, K), with m and K continuous. Let $H_R(m, K)$ be the closed linear hull of $z(t)$, $t \in R$, with respect to the inner product

$$(z(t), z(t'))_{m,K} = K(t, t') + m(t)m(t').$$

Then $H_R(m, K)$ is a separable Hilbert space of real random variables. Moreover, there exist mean and covariance operators, which we will also call m and K, such that for $h_1, h_2 \in H_R(m, K)$, $Eh_1 = m(h_1)$ and $\operatorname{cov}(h_1, h_2) = K(h_1, h_2)$. We will freely switch between the functions m and K and the operators m and K, the meaning being apparent from context.

If $(0, K)$, (m, K) and (m_1, K_1) are compatible on R, then $H_R(m, K) = H_R(m_1, K_1) = H_R(K)$ [where we define $H_R(K) = H_R(0, K)$], which follows from elementary considerations on the equivalence of Gaussian measures [Ibragimov and Rozanov (1978), page 71]. Thus, in this case, it makes sense to talk about the second moments of elements of $H_R(K)$ under (m, K) or (m_1, K_1). The subscript i will be used to denote expectations under (m_i, K_i), so that, for example, $E_i h_1 = m_i(h_1)$ and $\operatorname{cov}_i(h_1, h_2) = K_i(h_1, h_2)$ for $h_1, h_2 \in H_R(K)$.

3. General results. Suppose $(0, K_0)$, (m_0, K_0) and (m_1, K_1) are all compatible on R. Consider a linearly independent sequence $\eta_1, \eta_2, \ldots \in H_R(K_0)$ and let $\psi_1, \psi_2, \ldots$ be the Gram–Schmidt orthogonalization of the sequence under $(0, K_0)$. Let $H_\Psi(K_0)$ be the closed linear hull of $\psi_1, \psi_2, \ldots$ using the inner product defined by $(0, K_0)$. If $\eta_1, \eta_2, \ldots$ form a basis for $H_R(K_0)$, which will be the case if $\eta_i = z(t_i)$, $t_1, t_2, \ldots$ are dense in R, and K_0 is continuous on $R \times R$, then of course $H_\Psi(K_0) = H_R(K_0)$. For $\psi \in H_R(K_0)$, the best linear predictor of ψ based on $\Psi_N = (\psi_1, \ldots, \psi_N)'$ under (m_0, K_0) is

$$\hat{\psi}_N = E_0(\psi) + k_N'(\Psi_N - E_0(\Psi_N)),$$

where $k_N = (K_0(\psi, \psi_1), \ldots, K_0(\psi, \psi_N))'$. Let $e_0(\psi, N) = \hat{\psi}_N - \psi$ be the error of this prediction. Similarly define $e_1(\psi, N)$ to be the prediction error of the best linear predictor under (m_1, K_1). Suppose we assume (m_0, K_0) gives the second-order structure when in fact (m_1, K_1) is the correct second-order structure. Then

$$\frac{E_1 e_0(\psi, N)^2 - E_0 e_0(\psi, N)^2}{E_0 e_0(\psi, N)^2} \tag{3.1}$$

is a measure of the discrepancy between the presumed mean-squared prediction error $(E_0 e_0(\psi, N)^2)$ and the actual mean-squared prediction error $(E_1 e_0(\psi, N)^2)$. If both the numerator and the denominator of (3.1) are 0 (so that the prediction is both presumed to be perfect and is in fact perfect), define (3.1) to equal 0. We will use the convention that $0/0 = 0$ throughout this paper.

We now consider the behavior of (3.1) for $\psi \in H_\Psi(K_0)$ as $N \to \infty$. First, $\psi \in H_\Psi(K_0)$ implies

$$\psi = \sum_{i=1}^{\infty} c_i \psi_i,$$

the limit existing in L^2 under either (m_0, K_0) or (m_1, K_1), where $\Sigma c_i^2 < \infty$. We can then write

$$e_0(\psi, N) = \sum_{i=N+1}^{\infty} c_i(\psi_i - E_0\psi_i).$$

Define

$$b_{jk} = K_1(\psi_j, \psi_k) - K_0(\psi_j, \psi_k)$$

and

$$\mu_j = E_1\psi_j - E_0\psi_j.$$

Now, (m_0, K_0) and (m_1, K_1) compatible on $H_R(K_0)$ imply their compatibility on $H_\Psi(K_0)$; hence [Ibragimov and Rozanov (1978), page 81]

$$\sum_{jk=1}^{\infty} b_{jk}^2 < \infty \tag{3.2}$$

and [Ibragimov and Rozanov (1978), page 78]

$$\sum_{j=1}^{\infty} \mu_j^2 < \infty. \tag{3.3}$$

Thus, for all $\psi \in H_\Psi(K_0)$,

$$\left| \frac{E_1 e_0(\psi, N)^2 - E_0 e_0(\psi, N)^2}{E_0 e_0(\psi, N)^2} \right| = \frac{\left|\Sigma_{jk=N+1}^{\infty} c_j c_k (b_{jk} + \mu_j \mu_k)\right|}{\Sigma_{j=N+1}^{\infty} c_j^2}$$

$$\leq \left[\sum_{jk=N+1}^{\infty} (b_{jk} + \mu_j \mu_k)^2 \right]^{1/2}$$

by the Cauchy–Schwarz inequality. From (3.2) and (3.3), it follows that

$$\lim_{N \to \infty} \sup_{\psi \in H_\Psi(K_0)} \left| \frac{E_1 e_0(\psi, N)^2 - E_0 e_0(\psi, N)^2}{E_0 e_0(\psi, N)^2} \right| = 0.$$

This result can also be derived from a very general theorem due to Blackwell and Dubins (1962) on the variation distance between conditional distributions

on equivalent measures. The present derivation will allow us to make statements about the rate of convergence. The operator on l^2 defined by the infinite matrix $(b_{jk} + \mu_j\mu_k)_{jk=1}^{\infty}$ satisfies

$$\sum_{jk=1}^{\infty} (b_{jk} + \mu_j\mu_k)^2 < \infty,$$

so it is completely continuous [Akhiezer and Glazman (1961), page 58]. Let D_N be the operator on l^2 defined by the infinite matrix $(b_{jk} + \mu_j\mu_k)_{jk=N+1}^{\infty}$. Then D_N is also completely continuous, so its spectrum is discrete away from 0 [Akhiezer and Glazman (1961), page 118]; hence, for $c = (c_1, c_2, \dots)$, we have [Friedrichs (1973), Section 27]

$$\sup_{c \in l^2} \frac{\sum_{jk=N+1}^{\infty} c_j c_k (b_{jk} + \mu_j\mu_k)}{\sum_{j=N+1}^{\infty} c_j^2} = \Lambda_N \downarrow 0$$

and

$$\inf_{c \in l^2} \frac{\sum_{jk=N+1}^{\infty} c_j c_k (b_{jk} + \mu_j\mu_k)}{\sum_{j=N+1}^{\infty} c_j^2} = \lambda_N \uparrow 0$$

as $N \to \infty$, where Λ_N and λ_N are, respectively, the largest and smallest eigenvalues of D_N. We obtain the following theorem.

THEOREM 3.1. *As* $N \to \infty$,

$$\sup_{\psi \in H_\Psi(K_0)} \frac{E_1 e_0(\psi, N)^2 - E_0 e_0(\psi, N)^2}{E_0 e_0(\psi, N)^2} = \Lambda_N \downarrow 0$$

and

$$\inf_{\psi \in H_\Psi(K_0)} \frac{E_1 e_0(\psi, N)^2 - E_0 e_0(\psi, N)^2}{E_0 e_0(\psi, N)^2} = \lambda_N \uparrow 0.$$

Switching the roles of (m_0, K_0) and (m_1, K_1), we can define the operator $\tilde{D}_N$ with largest and smallest eigenvalues $\tilde{\Lambda}_N$ and $\tilde{\lambda}_N$, respectively, such that

$$\begin{aligned} &\sup_{\psi \in H_\Psi(K_0)} \frac{E_0 e_1(\psi, N)^2 - E_1 e_1(\psi, N)^2}{E_1 e_1(\psi, N)^2} = \tilde{\Lambda}_N \downarrow 0, \\ &\inf_{\psi \in H_\Psi(K_0)} \frac{E_0 e_1(\psi, N)^2 - E_1 e_1(\psi, N)^2}{E_1 e_1(\psi, N)^2} = \tilde{\lambda}_N \uparrow 0. \end{aligned} \tag{3.4}$$

These results can be used to obtain a bound for the difference in predictions under (m_0, K_0) and (m_1, K_1). For $E_0e_0(\psi, N)^2 > 0$,

$$\begin{aligned}\frac{E_0e_1(\psi, N)^2}{E_0e_0(\psi, N)^2} &= \frac{E_0e_1(\psi, N)^2}{E_1e_1(\psi, N)^2}\frac{E_1e_1(\psi, N)^2}{E_1e_0(\psi, N)^2}\frac{E_1e_0(\psi, N)^2}{E_0e_0(\psi, N)^2} \\ &\le \frac{E_0e_1(\psi, N)^2}{E_1e_1(\psi, N)^2}\frac{E_1e_0(\psi, N)^2}{E_0e_0(\psi, N)^2},\end{aligned} \tag{3.5}$$

where the inequality uses the fact that $e_i(\psi, N)$ minimizes the expected squared prediction error under (m_i, K_i). Since $e_0(\psi, N)$ is uncorrelated with Ψ_N under (m_0, K_0),

$$\begin{aligned}E_0e_1(\psi, N)^2 &= E_0\big[(e_1(\psi, N) - e_0(\psi, N)) + e_0(\psi, N)\big]^2 \\ &= E_0(e_1(\psi, N) - e_0(\psi, N))^2 + E_0e_0(\psi, N)^2,\end{aligned}$$

and it follows from (3.5) that

$$\begin{aligned}\frac{E_0(e_1(\psi, N) - e_0(\psi, N))^2}{E_0e_0(\psi, N)^2} &\le \left(1 + \frac{E_0e_1(\psi, N)^2 - E_1e_1(\psi, N)^2}{E_1e_1(\psi, N)^2}\right) \\ &\quad \times \left(1 + \frac{E_1e_0(\psi, N)^2 - E_0e_0(\psi, N)^2}{E_0e_0(\psi, N)^2}\right) - 1,\end{aligned}$$

and by applying Theorem 3.1 and (3.4), we obtain the following corollary.

COROLLARY 3.1. *As* $N \to \infty$,

$$\sup_{\psi \in H_\Psi(K_0)} \frac{E_0(e_1(\psi, N) - e_0(\psi, N))^2}{E_0e_0(\psi, N)^2} \le \Lambda_N + \tilde{\Lambda}_N + \Lambda_N\tilde{\Lambda}_N \to 0$$

and

$$\sup_{\psi \in H_\Psi(K_0)} \frac{E_1(e_0(\psi, N) - e_1(\psi, N))^2}{E_1e_1(\psi, N)^2} \le \Lambda_N + \tilde{\Lambda}_N + \Lambda_N\tilde{\Lambda}_N \to 0.$$

That is, for $\psi \in H_\Psi(K_0)$, the predictions obtained under (m_0, K_0) are uniformly close to the predictions obtained under (m_1, K_1).

4. Explicit bounds with equal covariance functions

4.1. *General discussion.* Obtaining bounds on Λ_N (or λ_N) appears to be a rather difficult problem in general. The situation simplifies considerably when $K_0 = K_1 = K$, so that the two possible models only differ in their mean

functions. In this case, it is easy to show that

$$\Lambda_N = \sum_{j=N+1}^{\infty} \mu_j^2 \quad \text{and} \quad \lambda_N = 0.$$

Without loss of generality, take $m_0 = 0$ and $m_1 = m$.

The fact that $(0, K)$ and (m, K) are compatible is equivalent to the existence of a $\nu \in H_R(K)$ such that for all $\psi \in H_R(K)$ [Parzen (1963)]

$$E_1\psi = E_0\nu\psi, \tag{4.1}$$

so that $E_0\nu z(t) = m(t)$ for $t \in R$. In fact, ν is just the Radon–Nikodym derivative of P_1 with respect to P_0. For given $\psi_1, \ldots, \psi_N \in H_R(K)$, let H_N be the subspace of $H_R(K)$ generated by $\psi_1, \ldots, \psi_N$ and H_{-N} all elements of $H_R(K)$ orthogonal to H_N. Then

$$\Lambda_N = \sup_{\psi \in H_{-N}} \frac{(E_0(\nu\psi))^2}{E_0\psi^2}.$$

If $\nu_N \in H_N$, then

$$\begin{aligned} \Lambda_N &= \sup_{\psi \in H_{-N}} \frac{(E_0(\nu - \nu_N)\psi)^2}{E_0\psi^2} \le \sup_{\psi \neq 0} \frac{(E_0(\nu - \nu_N)\psi)^2}{E_0\psi^2} \\ &= E_0(\nu - \nu_N)^2, \end{aligned} \tag{4.2}$$

where the inequality is an equality if ν_N is the element of H_N satisfying $E_0\nu\psi = E_0\nu_N\psi$ for all $\psi \in H_N$. That is, Λ_N can be bounded by approximating ν with elements of H_N.

If in fact $z(\cdot)$ is Gaussian and $K_0 = K_1 = K$, it is not difficult to show that the variation distance between the measures P_0 and P_1 given $\psi_1, \ldots, \psi_N$ is independent of $\psi_1, \ldots, \psi_N$ and equals $2\Phi(0.5\Lambda_N^{1/2}) - 1$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable. If Λ_N is small, then the variation distance is approximately $(2\pi)^{-1/2}\Lambda_N^{1/2}$. Blackwell and Dubins (1962) showed that the variation distance between conditional distributions based on equivalent measures converges to 0 almost surely. In the case of equivalent Gaussian measures with equal covariance functions, the results in the next sections allow us to give explicit bounds on this variation distance.

4.2. *One-dimensional process.* Sacks and Ylvisaker (1966, 1968, 1970, 1971), Wahba (1971, 1974) and Eubank, Smith and Smith (1981) have considered the problem of approximating an element of a Hilbert space by an element of a finite-dimensional subspace as it relates to choosing design points for estimating the regression coefficients of a continuous-time stochastic process. As noted in the review paper by Cambanis (1985), the same problem occurs in designs for the prediction of integrals of random quantities or for

signal detection. These papers all assume that $m(t)$ can be represented as

$$m(t) = \int_0^1 K(t, s)\rho(s)\, ds, \tag{4.3}$$

where $\rho(\cdot)$ satisfies some smoothness condition. The fact that $m(t)$ can be expressed as in (4.3) implies that (m, K) and $(0, K)$ are compatible on $[0, 1]$ [follows from comments by Wahba (1971), page 1036]. Suppose that $z(t)$ has exactly $n - 1$ mean-squared derivatives and $H_{N,p}$ is the subspace generated by $z^{(j)}(t_i)$ for $j = 0, \ldots, p$; $0 = t_0 < t_1 < \cdots < t_N = 1$, where $p \leq n - 1$. Then under certain conditions on K, an asymptotically valid expression for Λ_N as $N \to \infty$ is obtained for $p = n - 1$ which is of order N^{-2n} [see Sacks and Ylvisaker (1968) and Wahba (1971, 1974)].

In this section we will derive upper bounds on Λ_N under conditions on m and K not covered in previous works as well as sharper bounds when $p = 1$. While upper bounds on Λ_N are not sufficient for obtaining optimal or asymptotically optimal designs for estimating regression coefficients, they do allow one to obtain bounds on the efficiencies of certain designs [Sacks and Ylvisaker (1971)] besides being of interest in the problem at hand. We will assume that K is homogeneous so that

$$K(t) = \int e^{i\omega t} F(d\omega),$$

where $F(d\omega)$ is the spectral distribution. Let $L_R(F)$ be the real linear hull of the functions $e^{i\omega t}$ of ω, $t \in R$, closed with respect to the scalar product

$$\langle \varphi_1, \varphi_2 \rangle_F = \int \varphi_1(\omega)\overline{\varphi_2(\omega)} F(d\omega).$$

If $\Phi(d\omega)$ is the stochastic spectral measure corresponding to F then $\psi \in H_R(K)$ implies there exists $\varphi(\omega) \in L_R(F)$ satisfying

$$\psi = \int \varphi(\omega)\Phi(d\omega).$$

The correspondence between ψ and $\varphi(\omega)$ is a unitary isomorphism of the Hilbert spaces $H_R(K)$ and $L_R(F)$ [Ibragimov and Rozanov (1978), page 17]:

$$E_0\psi_1\psi_2 = \langle \varphi_1, \varphi_2 \rangle_F.$$

We can make use of the correspondence to obtain bounds on Λ_N under certain conditions on F.

Suppose $R = [0, 1]$ and $F(d\omega) = f(\omega)\, d\omega$, where for a positive integer n,

$$f(\omega) \leq \beta(1 + \omega^2)^{-n} \tag{4.4}$$

and

$$\liminf_{\omega \to \infty} f(\omega)(1 + \omega^2)^n > 0. \tag{4.5}$$

Under these conditions, all elements of $L_R(F)$ can be written in the form

$$\sum_{j=0}^{n-1} c_j(i\omega)^j + (1+i\omega)^n \int_0^1 c(t)e^{i\omega t}\,dt, \tag{4.6}$$

where the c_j's are real and $c(t)$ is square-integrable on $[0,1]$ [Ibragimov and Rozanov (1978), page 30]. Let $\varphi(\omega)$ be the element of $L_R(F)$ corresponding to $\nu \in H_R(K)$ as defined in (4.1). Then $\varphi(\omega)$ can be represented as in (4.6). Thus it is natural to investigate how well $\varphi(\omega)$ can be approximated by an element of some finite-dimensional subspace of $L_R(F)$ under additional restrictions on the c_j's and $c(t)$.

For $f(\cdot)$ satisfying (4.4) and (4.5), the process $z(\cdot)$ will have exactly $n-1$ mean-squared derivatives. The element in $L_R(F)$ corresponding to $z^{(j)}(t)$ for $j \le n-1$ is $(i\omega)^j e^{i\omega t}$. Let $L_{N,p}$ be the subspace of $L_R(F)$ isomorphic to $H_{N,p}$ and define $L_N = L_{N,0}$. Moreover, let $P_{N,p}$ be the operator that projects elements of $L_R(F)$ onto $L_{N,p}$ and P_N the operator that projects onto L_N, so that

$$\inf_{\varphi_N \in L_{N,p}} \|\varphi - \varphi_N\|_F^2 = \|\varphi - P_{N,p}\varphi\|_F^2.$$

From Theorem 3.1 and (4.2), for any $\varphi_N \in L_{N,p}$, $\|\varphi - \varphi_N\|_F^2$ is a uniform bound on the ratio of the squared bias of a prediction error to its variance when the mean function is taken to be 0 when in fact the mean function is [see Ibragimov and Rozanov (1978), page 91, or Yadrenko (1983), page 137]

$$m(t) = \int e^{-i\omega t}\varphi(\omega)\,f(\omega)\,d\omega. \tag{4.7}$$

Define $\Delta_k = t_k - t_{k-1}$. Assuming $f(\cdot)$ satisfies (4.4), we have the following bounds.

THEOREM 4.1. *Suppose $m(\cdot)$ is as in (4.7), $c(\cdot)$ as given in (4.6) has an absolutely continuous $m-1$th derivative and $c^{(m)}(t)$ is square-integrable on $[0,1]$, where m is a positive integer and $m \le n$. Then for $g(t) = c(t)e^{-t}$,*

$$\|\varphi - P_{N,n-1}\varphi\|_F^2 \le \frac{2\pi\beta e^2}{[(m-1)!]^2} \sum_{k=1}^{N} \left(\frac{\Delta_k}{2}\right)^{2m} \int_{t_{k-1}}^{t_k} g^{(m)}(t)^2\,dt,$$

which if $\Delta_k = 1/N$ for all k,

$$= \frac{\pi\beta e^2}{[(m-1)!]^2 2^{2m-1} N^{2m}} \int_0^1 g^{(m)}(t)^2\,dt.$$

THEOREM 4.2. *Assume $\varphi(\omega)$ is of the form*

$$\varphi(\omega) = \int_0^1 c(t)e^{i\omega t}\,dt,$$

which is equivalent to assuming $c(t) = \rho(t)$ in (4.3), and suppose $|c(t)|$ is

uniformly bounded by C on $[0, 1]$. *Let* $\gamma = N \max \Delta_k$. *Then for* $n > 1$,

$$\|\varphi - P_N\varphi\|_F^2 \leq \frac{16\beta n C^2}{2n-1}[2\gamma(n-1)]^{2n} \max\bigl(1, (4\gamma(n-1)/n!)^2\bigr) N^{-2n+1}.$$

THEOREM 4.3. *Suppose* $\Delta_k = N^{-1}$ *and*

$$\varphi(\omega) = \sum_{j=0}^{q} c_j (i\omega)^j + (1 + i\omega)^r \int_0^1 c(t) e^{i\omega t}\, dt,$$

where the c_j*'s are real,* $c(t)$ *is bounded and* $q, r \leq n - 1$. *Then*

$$\|\varphi - P_N\varphi\|_F^2 = O(N^{-2n+2\max(q,r)+1}).$$

Before proving these results, let us compare them to previous work. If $m(\cdot)$ is as in (4.3) with homogeneous K then from (4.3) and the definition of $f(\cdot)$,

$$m(t) = \int_0^1 \left[\int_{-\infty}^{\infty} e^{i\omega(s-t)} f(\omega)\, d\omega\right] \rho(s)\, ds,$$

and after changing the order of integration we see that

$$\varphi(\omega) = \int_0^1 \rho(s) e^{i\omega s}\, ds. \tag{4.8}$$

Letting $r(s) = \rho(s)e^{-s}$ and defining $R(s)$ by $R'(s) = r(s)$ and $R(1) = 0$,

$$\begin{aligned}\varphi(\omega) &= \int_0^1 r(s) e^{(1+i\omega)s}\, ds \\ &= -R(0) - (1 + i\omega)\int_0^1 R(s) e^{(1+i\omega)s}\, ds.\end{aligned}$$

Repeating the integration by parts, we conclude that $\varphi(\omega)$ can be written in the form (4.6), where $\rho(\cdot)$ square-integrable implies $c^{(n)}(\cdot)$ is square-integrable. Thus, in the case where $\varphi(\omega)$ can be expressed as in (4.8), the N^{-2n} convergence rate in Theorem 4.1 easily follows from Wahba (1974). The main advance here, then, is the bound on the approximation of φ by φ_N when $c(t)$ is not so smooth, which corresponds to $m(t)$ not being as smooth as is implied by (4.3). Typically, for $f(\omega)$ satisfying (4.4) and (4.5), $m(t)$ satisfying (4.3) translates into conditions on the $2n$th derivative of $m(t)$ [Wahba (1976), page 179]. In contrast, using Theorem 10 by Ibragimov and Rozanov (1978), page 92, we see that conditions on $c^{(m)}(t)$ translate into conditions on the $n + m$th derivative of $m(t)$.

Theorems 4.2 and 4.3 give results when only $z(\cdot)$ and not its mean-squared derivatives are observed at $t_0, \ldots, t_N$. Under weaker conditions on the covariance function, Cambanis and Masry (1983), Equation (2.7), obtained

$$\|\varphi - P_N\varphi\|_F^2 \leq A(\gamma/N)^{2n-2},$$

where A is a constant independent of N. Thus, Theorem 4.2 lowers the bound by order of N^{-1}, although it is possible that the correct order of magnitude of the bound is N^{-2n} [Cambanis and Masry (1983), page 86, and Eubank, Smith and Smith (1981), Theorem 5.1]. Theorem 4.3 gives a bound for mean functions that cannot be expressed as in (4.3).

PROOF OF THEOREM 4.1. For $l = 0, 1, \ldots, n-1$,

$$\begin{aligned}(1+i\omega)^n &\int_0^{t_k} t^l e^{(1+i\omega)t}\,dt \\ &= e^{(1+i\omega)t_k}\sum_{j=0}^{l}(1+i\omega)^{n-j-1}t_k^{l-j}(-1)^j l!/(l-j)! \\ &\quad +(-1)^{l+1}(1+i\omega)^{n-l-1}l!,\end{aligned}$$

which is clearly in $L_{N,n-1}$, so that for b_{jk}'s constants,

$$(1+i\omega)^n\sum_{k=1}^{N}\sum_{j=0}^{n-1}b_{jk}\int_0^{t_k}t^j e^{(1+i\omega)t}\,dt \in L_{N,n-1}.$$

Then

$$\begin{aligned}\|\varphi - P_{N,n-1}\varphi\|_F^2 &\le \int_{-\infty}^{\infty} f(\omega)\left|(1+i\omega)^n\int_0^1 e^{(1+i\omega)t}g(t)\,dt\right. \\ &\qquad \left.-(1+i\omega)^n\sum_{k=1}^{N}\sum_{j=0}^{n-1}b_{jk}\int_0^{t_k}t^j e^{(1+i\omega)t}\,dt\right|^2 d\omega \\ &\le \beta\int_{-\infty}^{\infty}\left|\int_0^1 e^{i\omega t}\left\{e^t g(t) - \sum_{k=1}^{N}\sum_{j=0}^{n-1}b_{jk}t^j e^t I_{\{t\le t_k\}}\right\}dt\right|^2 d\omega \\ &= 2\pi\beta\int_0^1\left\{e^t g(t) - \sum_{k=1}^{N}\sum_{j=0}^{n-1}b_{jk}t^j e^t I_{\{t\le t_k\}}\right\}^2 dt \\ &= 2\pi\beta\sum_{k=1}^{N}\int_{t_{k-1}}^{t_k} e^{2t}\left\{g(t) - \sum_{l=k}^{N}\sum_{j=0}^{n-1}b_{jl}t^j\right\}^2 dt,\end{aligned}$$

where the first equality is by Parseval's relation. Defining $\tau_k = (t_k + t_{k-1})/2$, it is clear that we can choose the b_{jk}'s such that

$$\sum_{l=k}^{N}\sum_{j=0}^{n-1}b_{jl}t^j = \sum_{j=0}^{m-1}g^{(j)}(\tau_k)(t-\tau_k)^j/j!,$$

so

$$\|\varphi - P_{N,n-1}\varphi\|_F^2 \le 2\pi\beta \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} e^{2t}\left[g(t) - \sum_{j=0}^{m-1} g^{(j)}(\tau_k)(t-\tau_k)^j/j!\right]^2 dt$$

$$= 2\pi\beta \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} e^{2t}\left[\int_{\tau_k}^{t} \frac{(t-s)^{m-1}}{(m-1)!} g^{(m)}(s)\, ds\right]^2 dt$$

$$\le \frac{2\pi\beta}{[(m-1)!]^2} \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} e^{2t}(t-\tau_k) \times \left[\int_{\tau_k}^{t} (t-s)^{2(m-1)} g^{(m)}(s)^2\, ds\right] dt,$$

by Jensen's inequality. Letting $\Delta_k = t_k - t_{k-1}$,

$$\int_{t_{k-1}}^{t_k} e^{2t}(t-\tau_k)\left[\int_{\tau_k}^{t} (t-s)^{2(m-1)} g^{(m)}(s)^2\, ds\right] dt$$

$$\le \left(\frac{\Delta_k}{2}\right)^{2m-1} \int_{t_{k-1}}^{t_k} e^{2t}\left|\int_{\tau_k}^{t} g^{(m)}(s)^2\, ds\right| dt$$

$$\le \left(\frac{\Delta_k}{2}\right)^{2m-1}\left[\int_{t_{k-1}}^{\tau_k} e^{2t}\, dt \int_{t_{k-1}}^{\tau_k} g^{(m)}(s)^2\, ds + \int_{\tau_k}^{t_k} e^{2t}\, dt \int_{\tau_k}^{t_k} g^{(m)}(s)^2\, ds\right]$$

$$\le \left(\frac{\Delta_k}{2}\right)^{2m} e^2 \int_{t_{k-1}}^{t_k} g^{(m)}(t)^2\, dt,$$

and Theorem 4.1 follows. □

PROOF OF THEOREM 4.2. For $n > 1$, we have

$$\|\varphi - P_N\varphi\|_F^2 \le \beta \int_{-\infty}^{\infty} (1+\omega^2)^{-n} \left|\int_0^1 c(t) e^{i\omega t}\, dt - \sum_{k=0}^{N} a_k e^{i\omega t_k}\right|^2 d\omega, \tag{4.9}$$

for any set of real a_k's. We will bound this integral using polynomial interpolating quadratures [Krylov (1962), Chapter 6]. For $k = 0, \ldots, n-1$, let

$$b_k = \int_{t_0}^{t_{n-1}} c(t) \frac{u(t)\, dt}{(t-t_k)u'(t_k)}, \tag{4.10}$$

where $u(t) = (t-t_0)\cdots(t-t_{n-1})$. For a real function $h(t)$ whose nth derivative is bounded by M [Krylov (1962), page 81, Equation (6.1.9)]

$$\left|\int_{t_0}^{t_{n-1}} c(t)h(t)\, dt - \sum_{k=0}^{n-1} b_k h(t_k)\right| \le \frac{M}{n!} \int_{t_0}^{t_{n-1}} |c(t)u(t)|\, dt$$

$$\le \frac{MC}{n!}(t_{n-1} - t_0)^{n+1}.$$

Applying this bound to the real and imaginary parts of $e^{i\omega t}$ separately, we have

$$\left|\int_{t_0}^{t_{n-1}} c(t)e^{i\omega t}\,dt - \sum_{k=0}^{n-1} b_k e^{i\omega t_k}\right| \le \frac{2C}{n!}|\omega|^n (t_{n-1} - t_0)^{n+1}. \tag{4.11}$$

We can similarly approximate the integrals from t_{n-1} to $t_{2(n-1)}$ up through $t_{(j-1)(n-1)}$ to $t_{j(n-1)}$, where $j(n-1) \le N < (j+1)(n-1)$. To approximate the integral from $t_{j(n-1)}$ to $t_N = 1$, let

$$b_k = \int_{t_{j(n-1)}}^{1} c(t)\frac{u(t)\,dt}{(t - t_{N-k})u'(t_{N-k})}$$

for $k = 0, \ldots, n-1$, where now $u(t) = (t - t_N)(t - t_{N-1}) \cdots (t - t_{N-n+1})$. Then

$$\left|\int_{t_{j(n-1)}}^{1} c(t)e^{i\omega t}\,dt - \sum_{k=0}^{n-1} b_k e^{i\omega t_{N-k}}\right| \le \frac{2C}{n!}|\omega|^n (t_N - t_{N-n+1})^{n+1}. \tag{4.12}$$

Thus, using (4.11) and (4.12), it is possible to choose the a_k's in (4.9) such that

$$\begin{aligned}&\left|\int_0^1 c(t)e^{i\omega t}\,dt - \sum_{k=0}^{N} a_k e^{i\omega t_k}\right| \\ &\quad\le \frac{2C}{n!}|\omega|^n\left[(n-1)\max_{1\le k\le N}\Delta_k\right]^{n+1}\left(\frac{N}{n-1}+1\right).\end{aligned} \tag{4.13}$$

We need a better bound for large ω. Considering b_k as defined in (4.10),

$$\begin{aligned}|b_k| &\le C\int_{t_0}^{t_{n-1}}\left|\frac{u(t)}{(t-t_k)u'(t_k)}\right|dt \\ &\le C(t_{n-1} - t_0)^n \Big/ \left(\min_{1\le k\le n-1}\Delta_k\right)^{n-1}.\end{aligned}$$

Since each a_k is made up of at most two "b_k's,"

$$|a_k| \le 2C\left[(n-1)\max_{1\le k\le N}\Delta_k\right]^n \Big/ \left(\min_{1\le k\le N}\Delta_k\right)^{n-1}. \tag{4.14}$$

It is easy to see that it is possible to eliminate some of $t_1, \ldots, t_{N-1}$ such that the remaining points $0 = t'_0 < t'_1 < \cdots < t'_{N'} = 1$ satisfy

$$\max_{1\le k\le N'}(t'_k - t'_{k-1}) \le \frac{2\gamma}{N},$$

$$\min_{1\le k\le N'}(t'_k - t'_{k-1}) \ge \frac{1}{2N},$$

and

$$N' \ge \frac{N}{2\gamma}.$$

Thus, using (4.14), it is possible to choose a'_k's such that

$$|a'_k| \le \frac{C[4\gamma(n-1)]^n}{N},$$

so that

$$\left|\int_0^1 c(t)e^{i\omega t}\,dt - \sum_{k=0}^{N'} a'_k e^{i\omega t_k}\right| \le C + \sum_{k=0}^{N'} |a'_k|$$
$$\le 2C[4\gamma(n-1)]^n = C_1,$$

independent of ω and N. Moreover, from (4.13), the a'_k's can also be chosen so that

$$\left|\int_0^1 c(t)e^{i\omega t}\,dt - \sum_{k=0}^{N'} a'_k e^{i\omega t_k}\right| \le \frac{4C}{n!}[2\gamma(n-1)]^{n+1}\left(\frac{|\omega|}{N}\right)^n$$
$$= C_2\left(\frac{|\omega|}{N}\right)^n.$$

Using (4.9), Theorem 4.2 then follows from

$$\|\varphi - P_N\varphi\|_F^2 \le \beta\int_{-\infty}^{\infty}(1+\omega^2)^{-n}\left[\min\left(C_1, C_2\left(\frac{|\omega|}{N}\right)^n\right)\right]^2 d\omega$$
$$\le \frac{4\beta n}{2n-1}\max(C_1^2, C_2^2)N^{-2n+1}. \qquad \square$$

PROOF OF THEOREM 4.3. We only outline the derivation, as many of the details are similar to the previous case. The assumption that $\Delta_k = N^{-1}$ simplifies the proof, but this result can be generalized to allow for unequally spaced observations. First we consider approximating $(i\omega)^m$ by an element of L_N. Let

$$(\lambda_0 \cdots \lambda_{n-1})' = m!\,N^m A_n^{-1} e_{m+1},$$

where A_n is the $n \times n$ matrix with the abth element $(b-1)^{a-1}$ (define $0^0 = 1$), and $e_p = (0 \cdots 0\,1\,0 \cdots 0)'$, where the 1 is in the pth place. Define

$$h_{m,N}(\omega) = \sum_{j=0}^{n-1} \lambda_j e^{i\omega j/N}.$$

Then the λ_j's have been defined so that

$$N^{-m}h_{m,N}(N\nu) = \sum_{j=0}^{n-1} \lambda_j e^{i\nu j}$$

has the Taylor series expansion $(i\nu)^m + O(|\nu|^n)$. Since $\Sigma\,\lambda_j e^{i\nu j}$ is bounded, it follows that there exists a constant C_1 such that

$$\left|(i\nu)^m - N^{-m}h_{m,N}(N\nu)\right| \le C_1 \min(|\nu|^n, |\nu|^m)$$

for all ν. Letting $\omega = N\nu$, we obtain

$$\left|(i\omega)^m - h_{m,N}(\omega)\right| \le C_1 \min(|\omega|^n N^{-n+m}, |\omega|^m)$$

for all N and ω. It follows that for $m \le n-1$ and $f(\omega)$ satisfying (4.4),

$$\begin{aligned}&\int_{-\infty}^{\infty} f(\omega)\left|(i\omega)^m - h_{m,N}(\omega)\right|^2 d\omega \\ &\quad = O\left(\int_0^{\infty} (1+\omega^2)^{-n} \min(\omega^{2n}N^{-2n+2m}, \omega^{2m})\, d_\omega\right) \\ &\quad = O(N^{-2n+2m+1}).\end{aligned} \tag{4.15}$$

Define

$$b_j = \int_0^{(n-m-1)/N} c(t)\frac{v(t)\,dt}{(t-jN^{-1})v'(j/N)},$$

where

$$v(t) = \prod_{j=0}^{n-m-1}(t - jN^{-1}).$$

Then there exists a constant C_2 such that

$$\begin{aligned}&\left|(i\omega)^m \int_0^{(n-m-1)/N} c(t)e^{i\omega t}\,dt - h_{m,N}(\omega)\sum_{j=0}^{n-m-1} b_j e^{i\omega j/N}\right| \\ &\quad \le \left|(i\omega)^m \int_0^{(n-m-1)/N} c(t)e^{i\omega t}\,dt - (i\omega)^m \sum_{j=0}^{n-m-1} b_j e^{i\omega j/N}\right| \\ &\qquad + \left|\left\{(i\omega)^m - h_{m,N}(\omega)\right\}\sum_{j=0}^{n-m-1} b_j e^{i\omega j/N}\right| \\ &\quad \le \frac{C_2}{N}\min(|\omega|^n N^{-n+m}, |\omega|^m).\end{aligned}$$

It follows that there exist $d_0, \ldots, d_N$ and a constant C_3 such that

$$\left|(i\omega)^m \int_0^1 c(t)e^{i\omega t}\,dt - \sum_{j=0}^{N} d_j e^{i\omega j/N}\right| \le C_3 \min(|\omega|^n N^{-n+m}, |\omega|^m).$$

Hence, for $f(\omega)$ satisfying (4.4),

$$\begin{aligned}&\int_{-\infty}^{\infty} f(\omega)\left|(i\omega)^m \int_0^1 c(t)e^{i\omega t}\,dt - \sum_{j=0}^{N} d_j e^{i\omega j/N}\right|^2 d\omega \\ &\quad = O(N^{-2(n-m)+1}).\end{aligned} \tag{4.16}$$

Using (4.15) and (4.16), Theorem 4.3 follows. □

4.3. *Two-dimensional process.* We now give an analog to Theorem 4.2 for two-dimensional processes. Assume $z(t)$, $t \in R \subseteq \mathbb{R}^2$, is a random field with homogeneous covariance function whose spectral density $f(\omega)$, $\omega \in \mathbb{R}^2$, satisfies

$$f(\omega) \leq \beta(1 + |\omega|^2)^{-n}, \tag{4.17}$$

where $n > 1$. Moreover, suppose the mean of $z(\cdot)$ corresponds to $\varphi(\omega) \in L_R(F)$, where $\varphi(\omega)$ can be written as

$$\varphi(\omega) = \int_R c(t) e^{i\omega' t}\, dt, \tag{4.18}$$

where $c(\cdot)$ is bounded on R. We only consider R rectangular, and without loss of generality, assume $R = [0, 1] \times [0, 1]$. We further suppose that the N observations are on a rectangular grid T_N. Specifically, $z(\cdot)$ is observed at (x_i, y_j), where $i = 0, \ldots, N_1$ and $j = 0, \ldots, N_2$ where $0 = x_0 < x_1 < \cdots < x_{N_1} = 1$, $0 = y_0 < y_1 < \cdots < y_{N_2} = 1$ and $N = (N_1 + 1)(N_2 + 1)$. Define L_N as the span of $e^{i\omega' t}$, $t \in T_N$,

$$\Delta_1 = \max_{1 \leq k \leq N_1} (x_k - x_{k-1}),$$

$$\Delta_2 = \max_{1 \leq k \leq N_2} (y_k - y_{k-1}),$$

and $\gamma = N^{1/2} \max(\Delta_1, \Delta_2)$.

THEOREM 4.4. *Suppose $f(\cdot)$ satisfies* (4.17) *and $\varphi(\cdot)$ is given by* (4.18), *where $|c(\cdot)|$ is bounded by C on R. Then there exists a constant C' depending only on n and C such that*

$$\|\varphi - P_N \varphi\|_F^2 \leq \beta C' \gamma^{2n+2} N^{-n+1}.$$

PROOF. A direct generalization of (4.9) is given by

$$\|\varphi - P_N \varphi\|_F^2 \leq \beta \int_{\mathbb{R}^2} (1 + |\omega|^2)^{-n} \left| \int_R c(t) e^{i\omega' t}\, dt - \sum_{j=0}^{N_1} \sum_{k=0}^{N_2} a_{jk} e^{i(\omega_1 x_j + \omega_2 y_k)} \right|^2 d\omega,$$

where $\omega = (\omega_1, \omega_2)'$. For $j, k = 0, \ldots, n-1$, let

$$b_{jk} = \int_{y_0}^{y_{n-1}} \int_{x_0}^{x_{n-1}} c(x, y) \prod_{\substack{l=1 \\ l \neq j}}^{n-1} \prod_{\substack{m=1 \\ m \neq k}}^{n-1} \left[\frac{(x - x_l)(y - y_m)}{(x_j - x_l)(y_k - y_m)} \right] dx\, dy.$$

Defining b_{jk} in this way gives exact quadrature for any real function of the form

$$h(x, y) = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} h_{jk} x^j y^k; \tag{4.19}$$

that is,

$$\int_{y_0}^{y_{n-1}}\int_{x_0}^{x_{n-1}} c(x,y)h(x,y)\,dx\,dy = \sum_{j,k=0}^{n-1} b_{jk}h(x_j,y_k).$$

Thus, for a given ω,

$$\left|\int_{y_0}^{y_{n-1}}\int_{x_0}^{x_{n-1}} c(x,y)\cos(\omega_1 x+\omega_2 y)\,dx\,dy - \sum_{j,k=0}^{n-1} b_{jk}\cos(\omega_1 x_j+\omega_2 y_k)\right|$$

$$\le C\int_{y_0}^{y_{n-1}}\int_{x_0}^{x_{n-1}} |R_\omega(x,y)|\,dx\,dy,$$

where $R_\omega(x,y)$ is the difference between $\cos(\omega_1 x+\omega_2 y)$ and the polynomial of the form (4.19) that interpolates $\cos(\omega_1 x+\omega_2 y)$ at (x_j,x_k) for $j,k=0,\ldots,n-1$. Using the expression for the error of this interpolatory polynomial [Steffensen (1927), page 206], it follows that

$$|R_\omega(x,y)| \le \frac{(x_{n-1}-x_0)^n}{n!}|\omega_1|^n + \frac{(y_{n-1}-y_0)^n}{n!}|\omega_2|^n + \frac{(x_{n-1}-x_0)^n(y_{n-1}-y_n)^n}{(n!)^2}|\omega_1\omega_2|^n.$$

Thus,

$$\int_{y_0}^{y_{n-1}}\int_{x_0}^{x_{n-1}} |R_\omega(x,y)|\,dx\,dy \le \frac{1}{n!}(x_{n-1}-x_0)^{n+1}(y_{n-1}-y_0)|\omega_1|^n + \frac{1}{n!}(y_{n-1}-y_0)^{n+1}(x_{n-1}-x_0)|\omega_2|^n + \frac{1}{(n!)^2}(x_{n-1}-x_0)^{n+1}(y_{n-1}-y_0)^{n+1}|\omega_1\omega_2|^n.$$

Arguing as in the one-dimensional case, there exist a_{jk}'s and a constant C_1 independent of ω and N such that

$$\left|\int_R c(t)e^{i\omega' t}\,dt - \sum_{j=0}^{N_1}\sum_{k=0}^{N_2} a_{jk}' e^{i(\omega_1 x_j+\omega_2 y_k)}\right| \le C_1 N\Delta_1\Delta_2\left(\Delta_1^n|\omega_1|^n + \Delta_2^n|\omega_2|^n + \Delta_1^n\Delta_2^n|\omega_1\omega_2|^n\right).$$

We can also show there exists a constant C_2 independent of N and ω such that

$$|a_{jk}| \le C_2\frac{(\Delta_1\Delta_2)^n}{(\delta_1\delta_2)^{n-1}},$$

where

$$\delta_1 = \min_{1 \le k \le N_1} (x_k - x_{k-1})$$

and

$$\delta_2 = \min_{1 \le k \le N_2} (y_k - y_{k-1}).$$

Then, as in the previous section, it is possible to find a_{jk}'s and a constant C_3 independent of N, ω and γ such that

$$\left| \int_R c(t) e^{i\omega' t}\, dt - \sum_{j=0}^{N_1} \sum_{k=0}^{N_2} a_{jk} e^{i(\omega_1 x_j + \omega_2 y_k)} \right| \le C_3 \gamma^{n+1} \min\left(1, \left(\frac{|\omega_1|}{N^{1/2}}\right)^n + \left(\frac{|\omega_2|}{N^{1/2}}\right)^n + \left(\frac{|\omega_1 \omega_2|}{N}\right)^n\right).$$

Theorem 4.4 then follows from

$$\|\varphi - P_N \varphi\|_F^2 \le \beta C_3 \gamma^{2n+2} \left[\int_{|\omega| \le N^{1/2}} \left(1 + |\omega|^2\right)^{-n} \left\{ \left(\frac{|\omega_1|}{N^{1/2}}\right)^n + \left(\frac{|\omega_2|}{N^{1/2}}\right)^n + \left(\frac{|\omega_1 \omega_2|}{N}\right)^n \right\}^2 d\omega + \int_{|\omega| > N^{1/2}} \left(1 + |\omega|^2\right)^{-n} d\omega \right]. \quad \square$$

Generalizing Theorems 4.2 and 4.4 to a d-dimensional hypercube, the appropriate bound is of the order $N^{-(2n/d)+1}$, where N is the total number of observations. Ylvisaker (1975) has obtained an asymptotic expression for the best approximation of an element of $H_R(K)$ that can be expressed as in (4.3) (where s and t are now vectors of length 2) based on observations on a rectangular grid. He also assumes that the covariance function is of a rather restricted form which does not include differentiable processes.

5. Consistency and asymptotic optimality of best linear unbiased predictors. Consider a random field $z(t)$ defined for $t \in R$,

$$z(t) = \beta' g(t) + e(t),$$

where $g(t) = (g^1(t), \ldots, g^p(t))'$ is a known vector-valued function, β a vector of unknown coefficients and $e(t)$ a zero mean random field with known covariance function $K(t, t')$. Let $H_R(g, K)$ be the closed linear hull $z(t)$, $t \in R$, with respect to the inner product

$$(z(t), z(t'))_{g,K} = K(t, t') + g(t)' g(t'). \tag{5.1}$$

Suppose $\psi_0, \psi_1, \ldots \in H_R(g, K)$, and let $g_i = g(\psi_i)$ and $K_{ij} = K(\psi_i, \psi_j)$. Define $\Psi_N = (\psi_1, \ldots, \psi_N)'$, $G_N = (g_1, \ldots, g_N)$, K_N the $N \times N$ matrix with ijth element K_{ij} and $k_N = (K_{01}, \ldots, K_{0N})'$. Assume K_N is positive definite for all N. Then the best linear unbiased predictor (BLUP) of ψ_0 based on Ψ_N is defined to be the predictor of the form $v'\Psi_N$ that minimizes $\operatorname{var}(\psi_0 - v'\Psi_N)$ subject to the unbiasedness constraint $g_0 = G_N v$. The BLUP exists if and only if $g_0 \in C(G_N)$, where $C(A)$ means the column space of a matrix A. In this case the

BLUP is given by

$$\hat{\psi}_n = \left\{k_N' K_N^{-1} + \left(g_0 - G_N K_N^{-1} k_N\right)' \left(G_N K_N^{-1} G_N'\right)^{-} G_N K_N^{-1}\right\} \Psi_N, \tag{5.2}$$

where $(G_N K_N^{-1} G_N')^{-}$ is any generalized version of $G_N K_N^{-1} G_N'$. Let $H_\Psi(g, K)$ be the closed linear hull of $\psi_1, \psi_2, \ldots$ with respect to the inner product defined by (5.1).

THEOREM 5.1. *The BLUP of ψ_0 is mean-square consistent if and only if $\psi_0 \in H_\Psi(g, K)$.*

PROOF. Suppose the BLUP does not exist for any N so that $g_0 \notin C(G_N)$ for all N. Let $P_N g_0$ be the projection of g_0 onto $C(G_N)$ with respect to Euclidean distance. Then it is clear that $\|g_0 - P_N g_0\| \nrightarrow 0$ as $N \to \infty$ so that $\psi_0 \notin H_\Psi(g, K)$, completing this case.

Now suppose $g_0 \in C(G_N)$ for some N, in which case, there exists N_0 such that $g_0 \in C(G_{N_0})$ and $\operatorname{rank}(G_N) = \operatorname{rank}(G_{N_0}) = q$ for all $N \geq N_0$. For all $N \geq N_0$, we have the spectral decomposition

$$G_N K_N^{-1} G_N' = \sum_{j=1}^{q} \varphi_{jN} \beta_{jN} \beta_{jN}',$$

where $\varphi_{1N} \geq \varphi_{2N} \geq \cdots \geq \varphi_{qN} > 0$ and $\beta_{1N}, \ldots, \beta_{qN}$ are orthonormal vectors of length p. Consider the particular generalized inverse of $G_N K_N^{-1} G_N'$ given by

$$J_N = \sum_{j=1}^{q} \varphi_{jN}^{-1} \beta_{jN} \beta_{jN}'.$$

For $N \geq N_0$, from (5.2),

$$\operatorname{var}\left(\psi_0 - \hat{\psi}_N\right) = K_{00} - k_N' K_N^{-1} k_N + \sum_{j=1}^{q} \varphi_{jN}^{-1} \left[\left(g_0 - G_N K_N^{-1} k_N\right)' \beta_{jN}\right]^2. \tag{5.3}$$

Now, $\psi_0 \in H_\Psi(g, K)$ means

$$\lim_{N \to \infty} \min_{v \in \mathbb{R}^N} \|\psi_0 - v' \Psi_N\|_{g,K}^2 = 0,$$

and by straightforward calculation,

$$\begin{aligned}
&\min_{v \in \mathbb{R}^N} \|\psi_0 - v' \Psi_N\|_{g,K}^2 \\
&\quad = K_{00} + g_0' g_0 - \left(G_N' g_0 + k_N\right)' \left(G_N' G_N + K_N\right)^{-1} \left(G_N' g_0 + k_N\right) \\
&\quad = K_{00} + g_0' g_0 - \left(G_N' g_0 + k_N\right)' \\
&\qquad \times \left\{K_N^{-1} - K_N^{-1} G_N' \left(I + G_N K_N^{-1} G_N'\right)^{-1} G_N K_N^{-1}\right\} \left(G_N' g_0 + k_N\right) \qquad (5.4) \\
&\quad = K_{00} - k_N' K_N^{-1} k_N + \left(g_0 - G_N K_N^{-1} k_N\right)' \left(I + G_N K_N^{-1} G_N'\right)^{-1} \\
&\qquad \times \left(g_0 - G_N K_N^{-1} k_N\right) \\
&\quad = K_{00} - k_N' K_N^{-1} k_N + \sum_{j=1}^{q} \left(1 + \varphi_{jN}\right)^{-1} \left[\left(g_0 - G_N K_N^{-1} k_N\right)' \beta_{jN}\right]^2.
\end{aligned}$$

For $N \geq N_0$, φ_{jN} is bounded from below by $\varphi_{qN_0} > 0$. It follows that (5.4) tends to 0 if and only if (5.3) tends to 0, so Theorem 5.1 is proved. □

As an application of this result, suppose $\psi_i = z(t_i)$, $K(\cdot, \cdot)$ is continuous at (t_0, t_0), $g(\cdot)$ is continuous at t_0 and t_0 is a limit point of $\{t_i\}_{i=1}^{\infty}$. Define $t_{(N)}$ to be the nearest neighbor to t_0 among $t_1, \ldots, t_N$. Obviously, $\|z(t_0) - z(t_{(N)})\|_{g,K}^2 \to 0$ as $N \to \infty$, so that $z(t_0) \in H_\Psi(g, K)$. By Theorem 5.1, the BLUP of $z(t_0)$ based on $z(t_1), \ldots, z(t_N)$ exists for N sufficiently large and it is mean-square consistent.

If (g^i, K) and $(0, K)$ are compatible on R for $i = 1, \ldots, p$, then for $\psi_0 \in H_\Psi(K)$, the BLUP of ψ_0 based on Ψ_N is asymptotically optimal relative to the best linear predictor (the minimum variance linear predictor when β is known). More specifically, assume without loss of generality that $\mathrm{Cov}(\Psi_N) = I_N$, in which case the error of the best linear predictor is

$$e(\psi, N) = \psi_0 - \left[\beta' g_0 + k_N'(\Psi_N - G_N\beta)\right] \tag{5.5}$$

with

$$\mathrm{var}(e(\psi, N)) = K_{00} - k_N' k_N,$$

and the error of the BLUP is

$$\tilde{e}(\psi, N) = \psi_0 - \left[k_N' + (g_0 - G_N k_N)' J_N G_N\right]\Psi_N \tag{5.6}$$

with

$$\mathrm{var}(\tilde{e}(\psi, N)) = K_{00} - k_N' k_N + (g_0 - G_N' k_N)' J_N (g_0 - G_N' k_N),$$

where

$$J_N \downarrow J = \sum_{j=1}^{q} \varphi_j \beta_j \beta_j'.$$

Define $\Lambda_N(\beta)$ to be the largest eigenvalue of the operator D_N as defined in Section 3 where $(m_0, K_0) = (0, K)$ and $(m_1, K_1) = (\beta' g, K)$.

THEOREM 5.2. *There exists $\varepsilon_N \to 0$ such that*

$$\sup_{\psi \in H_\Psi(K)} \frac{E\tilde{e}(\psi, N)^2 - Ee(\psi, N)^2}{Ee(\psi, N)^2} \leq \sum_{j=1}^{q} (\varphi_j + \varepsilon_N)\Lambda_N(\beta_j).$$

Since $\Lambda_N(\beta_j) \downarrow 0$ as $N \to \infty$ for $j = 1, \ldots, q$ (see Theorem 3.1), we have that BLUP's are uniformly asymptotically efficient relative to best linear predictors over all elements of $H_\Psi(K)$. The error of the BLUP given in (5.6) can be obtained by replacing β in (5.5) by $\hat{\beta}$, where $\hat{\beta}$ is any least squares estimator of β based on Ψ_N. Thus, Theorem 5.2 is not surprising since Theorem 3.1 essentially says that using a fixed but incorrect value of β yields asymptotically optimal predictors.

PROOF OF THEOREM 5.2. Since $J_N \downarrow J$ and $\operatorname{rank}(J_N) = q$ for N sufficiently large, there exists $\varepsilon_N \to 0$ such that

$$J_N \leq \sum_{j=1}^{q} (\varphi_j + \varepsilon_N)\beta_j\beta_j'.$$

Then

$$\begin{aligned}(g_0 - G_N'k_N)J_N(g_0 - G_N'k_N) &\leq \sum_{j=1}^{q} (\varphi_j + \varepsilon_N)\left[(g_0 - G_N'k_N)'\beta_j\right]^2 \\ &= \sum_{j=1}^{q} (\varphi_j + \varepsilon_N)\left[E_j e(\psi, N)\right]^2,\end{aligned}$$

where $E_j(\cdot)$ indicates expectation under $((\beta_j + \beta)'g, K)$. Since

$$\sup_{\psi \in H_\Psi(K)} \frac{E_j e(\psi, N)^2 - E_\beta e(\psi, N)^2}{E_\beta e(\psi, N)^2} = \Lambda_N(\beta_j),$$

Theorem 5.2 follows from Theorem 3.1. □

Stein (1988) showed that if in fact K_1 is the correct covariance function and $(0, K)$ and $(0, K_1)$ are compatible on R, then the BLUP of ψ_0 based on K is asymptotically efficient relative to the BLUP based on the correct covariance function K_1 if the BLUP based on K_1 is consistent. Theorem 5.1 resolves the issue of the consistency of the BLUP. If $(0, K)$ and (g^i, K) are compatible on R for $i = 1, \ldots, p$, then the results in Stein (1988) can be extended to show that the BLUP based on K is uniformly optimal over all $\psi \in H_\Psi(K)$ relative to the BLUP based on K_1. This result can be combined with Theorem 5.2 yielding that the BLUP based on K is uniformly optimal relative to the best linear predictor based on K_1 over all $\psi \in H_\Psi(K)$.

Acknowledgments. I would like to thank the referee and associate editor for their correction of errors and other helpful criticisms of earlier versions of this paper.

REFERENCES

AKHIEZER, N. I. and GLAZMAN, I. M. (1961). *Theory of Linear Operators in Hilbert Space* **1**. Ungar, New York.

BLACKWELL, D. and DUBINS, L. (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33** 882–886.

CAMBANIS, S. (1985). Sampling designs for time series. In *Time Series in the Time Domain. Handbook of Statistics* (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) **5** 337–362. Elsevier, Amsterdam.

CAMBANIS, S. and MASRY, E. (1983). Sampling designs for the detection of signals in noise. *IEEE Trans. Inform. Theory* **29** 83–104.

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

EUBANK, R. L., SMITH, P. L. and SMITH, P. W. (1981). Uniqueness and eventual uniqueness of optimal designs in some time series models. *Ann. Statist.* **9** 486–493.

FRIEDRICHS, K. O. (1973). *Spectral Theory of Operators in Hilbert Space*. Springer, New York.

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

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

KRYLOV, V. I. (1962). *Approximate Calculation of Integrals*. Macmillan, New York.

PARZEN, E. (1963). Probability density functionals and reproducing kernel Hilbert spaces. In *Proc. Symp. Time Series Anal.* (M. Rosenblatt, ed.) 155–169. Wiley, New York.

SACKS, J. and YLVISAKER, D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* **37** 66–89.

SACKS, J. and YLVISAKER, D. (1968). Designs for regression problems with correlated errors: Many parameters. *Ann. Math. Statist.* **39** 49–69.

SACKS, J. and YLVISAKER, D. (1970). Designs for regression problems with correlated errors. III. *Ann. Math. Statist.* **41** 2057–2074.

SACKS, J. and YLVISAKER, D. (1971). Statistical designs and integral approximation. In *Proc. Twelfth Biennial Canad. Math. Soc. Sem.* (R. Pyke, ed.) 115–136.

STEFFENSEN, J. F. (1927). *Interpolation*. Williams and Wilkins, Baltimore.

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

STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63.

STEIN, M. L. (1989). Asymptotic distributions of minimum norm quadratic estimators of the covariance function of a Gaussian random field. *Ann. Statist.* **17** 980–1000.

STEIN, M. L. and HANDCOCK M. S. (1989). Some asymptotic properties of kriging when the covariance function is misspecified. *Math. Geol.* **21** 171–190.

SUKHATME, S. (1985). Kriging with perturbed variogram. In *Proc. Soc. Statist. Sec. Amer. Statist. Assoc. Annual Meetings* 296–299.

WAHBA, G. (1971). On the regression problem of Sacks and Ylvisaker. *Ann. Math. Statist.* **42** 1035–1053.

WAHBA, G. (1974). Regression design for some equivalence classes of kernels. *Ann. Statist.* **2** 925–934.

WAHBA, G. (1976). On the optimal choice of nodes in the collocation–projection method for solving linear operator equations. *J. Approx. Theory* **16** 175–186.

WARNES, J. J. (1986). A sensitivity analysis for universal kriging. *Math. Geol.* **18** 653–676.

YADRENKO, M. I. (1983). *Spectral Theory of Random Fields*. Optimization Software, New York.

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

YLVISAKER, D. (1975). Designs on random fields. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) 593–608. North-Holland, Amsterdam.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637