The Annals of Statistics
2002, Vol. 30, No. 1, 298–323

THE SCREENING EFFECT IN KRIGING[1]

BY MICHAEL L. STEIN

When predicting the value of a stationary random field at a location **x** in some region in which one has a large number of observations, it may be difficult to compute the optimal predictor. One simple way to reduce the computational burden is to base the predictor only on those observations nearest to **x**. As long as the number of observations used in the predictor is sufficiently large, one might generally expect the best predictor based on these observations to be nearly optimal relative to the best predictor using all observations. Indeed, this phenomenon has been empirically observed in numerous circumstances and is known as the screening effect in the geostatistical literature. For linear predictors, when observations are on a regular grid, this work proves that there generally is a screening effect as the grid becomes increasingly dense. This result requires that, at high frequencies, the spectral density of the random field not decay faster than algebraically and not vary too quickly. Examples demonstrate that there may be no screening effect if these conditions on the spectral density are violated.

1. Introduction. Kriging, which is effectively optimal linear prediction, is a popular method for predicting random fields based on observations of the random field at some set of locations. Geostatisticians have long noted that, when predicting at a particular location **x**, it is often the case that those observations nearest to **x** have the largest impact on the kriging predictor and that the kriging predictor based on only these nearest observations is nearly optimal relative to the kriging predictor based on all of the observations. This phenomenon is known as the screen or screening effect in the geostatistical literature [Armstrong (1998), Chilès and Delfiner (1999), Cressie (1993), Journel and Huijbregts (1978), Wackernagel (1995)]. In addition to its theoretical interest, the screening effect is of practical importance because it provides a justification of the common practice of using only those observations nearest to **x** when predicting at **x** as a way of reducing computations or because of concerns about the lack of validity of a model over larger spatial scales. Except for some highly specific special cases in which certain observations have no impact on the kriging predictor, the evidence to date for the screening effect is empirical in the spatial setting [Chilès and Delfiner (1999), Section 3.6] although results such as Theorem 10 in Chapter 3 and Theorem 12 in Chapter 4 of Stein (1999), which show that models with similar local behavior commonly have similar optimal linear predictors, provide indirect

Received October 2000; revised August 2001.
[1]Supported in part by NSF Grant DMS-99-71127.
AMS 2000 subject classifications. Primary 62M20; secondary 62M40, 60G25.
Key words and phrases. Random field, best linear prediction, asymptotics, regular variation, self-similar, self-affine.

evidence that only local observations should generally have a nonnegligible impact on kriging predictions.

It is not clear how one could obtain useful and general results on the screening effect for any fixed set of observations. As in many statistical problems, it will prove helpful to consider an asymptotic approach to the problem. It will also prove helpful to choose a scenario in which it is possible to do some exact analysis on the behavior of kriging predictors. The following setup meets these criteria. For $\delta > 0$, suppose we observe a mean 0, weakly stationary and mean square continuous random field Z at $\delta\mathbf{j}$ for all $\mathbf{j} \in \mathbb{Z}^d$, the d-dimensional integer lattice. For fixed $\mathbf{x}$ in the unit cube $[0, 1]^d$ but not a vertex of this cube, consider predicting $Z(\delta\mathbf{x})$. It turns out (see Section 2) that we can explicitly characterize the best linear predictor of $Z(\delta\mathbf{x})$ based on this infinite lattice of observations.

One possible mathematical embodiment of a screening effect would be to prove that, if one uses only those locations $\delta\mathbf{j}$ in some fixed neighborhood of the origin to predict at $\delta\mathbf{x}$, as $\delta \downarrow 0$, the best linear predictor based on these observations is asymptotically optimal relative to the best linear predictor based on the infinite lattice. More specifically, for $A \subseteq \mathbb{R}^d$, define $e(\mathbf{x}, A)$ to be the error of the optimal linear predictor of $Z(\mathbf{x})$ when Z is observed at all $\mathbf{y} \in A$. If $B \subseteq \mathbb{R}^d$ contains some neighborhood of the origin, then

$$\lim_{\delta\downarrow 0} \frac{Ee(\delta\mathbf{x}, B \cap \delta\mathbb{Z}^d)^2}{Ee(\delta\mathbf{x}, \delta\mathbb{Z}^d)^2} = 1 \tag{1}$$

would be the asymptotic result we seek. In fact, since, for B bounded, the number of observations in $B \cap \delta\mathbb{Z}^d$ grows like δ^{-d} as $\delta \downarrow 0$, we might hope for a stronger result that restricts consideration to some large but fixed number of observations near $\delta\mathbf{x}$. Again consider $B \subseteq \mathbb{R}^d$ containing some neighborhood of the origin. The idea now is that, if r is large, then $\lim_{\delta\downarrow 0}\{Ee(\delta\mathbf{x}, \delta(rB \cap \mathbb{Z}^d))^2/Ee(\delta\mathbf{x}, \delta\mathbb{Z}^d)^2\}$ should be near 1, or,

$$\lim_{r\to\infty} \lim_{\delta\downarrow 0} \frac{Ee(\delta\mathbf{x}, \delta(rB \cap \mathbb{Z}^d))^2}{Ee(\delta\mathbf{x}, \delta\mathbb{Z}^d)^2} = 1. \tag{2}$$

Of course, (2) implies (1). The main result of this paper, Theorem 1 in Section 2, proves (2) under fairly general conditions on the spectral density for Z.

One might imagine that (2), or at least the weaker (1), is essentially always true for stationary random fields. To see that this is not the case and to motivate the assumptions on the spectral density in Theorem 1, let us consider several examples. We first give some definitions. For two positive functions f and g on a domain D, we will say $f \ll g$ if f/g is bounded on D. Furthermore, we will say $f(t) \ll g(t)$ as $t \to t_0$ if, for some neighborhood D' of t_0, $f \ll g$ on D'. We will say that $f \asymp g$ on D if $f \ll g$ and $g \ll f$ on D and that $f(t) \asymp g(t)$ as $t \to t_0$ if $f(t) \ll g(t)$ as $t \to t_0$ and $g(t) \ll f(t)$ as $t \to t_0$.

In the first example, suppose that Z is a weakly stationary process on $\mathbb{R}$ and, for some $\beta > 0$, has spectral density f satisfying $f(v) \asymp e^{-\beta|v|}$ on $\mathbb{R}$. This condition implies that Z has mean square derivatives of all orders. Consider predicting $Z(\frac{1}{2}\delta)$ based on observing $Z(\delta j)$ for all $j \in \mathbb{Z}$. From results in 3.8 of Stein (1999) (see also Section 2),

$$\begin{aligned} Ee\left(\tfrac{1}{2}\delta, \delta\mathbb{Z}^d\right)^2 &= \int_{-\pi/\delta}^{\pi/\delta} \left| 1 - \frac{\sum_{j=-\infty}^{\infty}(-1)^j f(v + 2\pi\delta^{-1}j)}{\sum_{j=-\infty}^{\infty} f(v + 2\pi\delta^{-1}j)} \right|^2 \\ &\qquad \times \sum_{j=-\infty}^{\infty} f(v + 2\pi\delta^{-1}j)\, dv \\ &= 4\int_{-\pi/\delta}^{\pi/\delta} \frac{\left\{\sum_{j \text{ odd}} f(v + 2\pi\delta^{-1}j)\right\}^2}{\sum_{j=-\infty}^{\infty} f(v + 2\pi\delta^{-1}j)}\, dv \\ &\asymp \int_0^{\pi/\delta} e^{\beta v}\left\{e^{-\beta(2\pi\delta^{-1}+v)} + e^{-\beta(2\pi\delta^{-1}-v)}\right\}^2 dv \asymp e^{-\pi\beta/\delta} \end{aligned} \tag{3}$$

as $\delta \downarrow 0$. If, for some fixed $m > 0$, one considers using just those observations at δj for $j = -m+1, -m+2, \dots, m$, then the mean squared prediction error of $\sum_{j=-m+1}^{m} c_j(\delta) Z(\delta j)$ is

$$\int_{-\infty}^{\infty} \left| e^{i\delta v/2} - \sum_{j=-m+1}^{m} c_j(\delta) e^{i\delta j v} \right|^2 f(v)\, dv. \tag{4}$$

For any f bounded away from 0 in a neighborhood of the origin, it is possible to show that (4) cannot tend to 0 faster than δ to a finite power (details available from author). Comparing this result to (3), we see that, for any bounded set B, $\lim_{r\to\infty} \lim_{\delta\downarrow 0}\{Ee(\frac{1}{2}\delta, \delta(rB \cap \mathbb{Z}))^2 / Ee(\frac{1}{2}\delta, \delta\mathbb{Z})^2\} = \infty$, so that (2) is false for $f(v) \asymp e^{-\beta|v|}$. I do not know whether (1) is true in this case.

The "Cauchy" model for covariance functions [Chilès and Delfiner (1999), page 86] has autocovariance function $K(x) = \operatorname{cov}\{Z(y), Z(y+x)\} \propto (1 + x^2/\beta^2)^{-1}$ and spectral density $f(v) \propto e^{-\beta|v|}$ for a process on $\mathbb{R}$ and hence a process with this autocovariance function will not satisfy (2). By an argument very similar to the one leading to (3), it is possible to show that a process in one dimension with "Gaussian" autocovariance function will not satisfy (2). A Gaussian autocovariance function is proportional to $e^{-\beta x^2}$ for some $\beta > 0$ and has corresponding spectral density proportional to $e^{-v^2/(4\beta)}$. Gaussian autocovariance functions have been used in applications in various disciplines [Chilès and Delfiner (1999), page 85], but Stein [(1999), pages 29 and 69] has argued against their practical use, in part because of the pronounced lack of a screening effect found in a numerical study.

One can get an even more extreme result by considering f with bounded support. Specifically, according to the sampling theorem for random fields [Jerri (1977)], if the support of f is contained in $[-\pi T, \pi T]^d$ and $\delta \le T^{-1}$, then the random field can be perfectly predicted throughout $\mathbb{R}^d$ based on observations on $\delta\mathbb{Z}^d$. Furthermore, the fact that the process has a spectral density implies that no finite set of distinct observations has a singular covariance matrix, so that $Ee(\delta\mathbf{x}, B \cap \delta\mathbb{Z}^d)^2 > 0$ for all bounded B and $\delta > 0$. Thus, $Ee(\delta\mathbf{x}, B \cap \delta\mathbb{Z}^d)^2/Ee(\delta\mathbf{x}, \delta\mathbb{Z}^d)^2$ is a positive number over 0 for any bounded B and any $\delta \in (0, T^{-1})$, so (1) is badly false.

For a different kind of example of when (1) is false, suppose Z on $\mathbb{R}$ has autocovariance function $K(x) = \operatorname{cov}\{Z(y), Z(y+x)\} = 1 - |x|$ for $|x| < 1$ and 0 otherwise, which is known as the triangle [Chilès and Delfiner (1999), page 61] or triangular [Stein (1999), page 30] autocovariance function. From the example on page 67 of Stein (1999), one should expect a problem with this autocovariance function. When predicting $Z(\frac{1}{2}\delta)$ based on all $Z(\delta j)$ in $B = [-b, b]$ for $b < 1$, for all δ sufficiently small, the best linear predictor of $Z(\frac{1}{2}\delta)$ is just $\frac{1}{2}\{Z(0) + Z(\delta)\}$ and the mean squared error of this prediction is $\frac{1}{2}\delta$. Now suppose $\delta = \frac{2}{2n+1}$ for some positive integer n. Then

$$\begin{aligned} &Ee\left(\frac{1}{2n+1}, \frac{2}{2n+1}\mathbb{Z}\right)^2 \\ &\quad \le E\Bigg[Z\left(\frac{1}{2n+1}\right) - \frac{1}{2}\left\{Z\left(\frac{2}{2n+1}\right) + Z(0)\right\} \\ &\qquad\qquad - \frac{1}{8}\left\{Z\left(\frac{2n}{2n+1}\right) + Z\left(\frac{-2(n-1)}{2n+1}\right) - Z\left(\frac{2(n+1)}{2n+1}\right) - Z\left(\frac{-2n}{2n+1}\right)\right\}\Bigg]^2 \\ &\quad = \frac{7}{8(2n+1)}, \end{aligned}$$

so that $\overline{\lim}_{\delta\downarrow 0}\{Ee(\frac{1}{2}\delta, B \cap \delta\mathbb{Z})^2/Ee(\frac{1}{2}\delta, \delta\mathbb{Z})^2\} \ge \frac{8}{7}$ and (1) is false. The problem is caused by the lack of smoothness of the autocovariance function at ± 1. The corresponding spectral density is $f(v) = (1 - \cos v)/(\pi v^2)$ [Stein (1999), page 68], which has substantial oscillations at high frequencies.

The spherical autocovariance function, which Chilès and Delfiner [(1999), page 225] call "the geostatistician's best friend," also has a spectral density with substantial oscillations at high frequencies when used as a model for a random field in three dimensions. For this model, covariances only depend on the distance r between observations and equal 0 beyond a range parameter a and are proportional to $1 - \frac{3}{2}\frac{r}{a} + \frac{1}{2}\frac{r^3}{a^3}$ for $r < a$. For $d = 3$, the corresponding spectral density depends only on $v = |\mathbf{v}|$ and is proportional to $\cos^2(\frac{1}{2}av)/v^4$ plus a term of order v^{-5} as $v \to \infty$. I would conjecture that (1) is false in this setting for any bounded set B, although I cannot prove it. However, Stein and Handcock (1989) demonstrate that the spherical model lacks a screening effect when one observes averages of the random field over certain concentric spheres and wishes to predict the random field at their common center.

These examples suggest that we will need to exclude at least some spectral densities that either decay too fast at high frequencies or are too wiggly at high frequencies. We can capture both of these characteristics by requiring that for some constant $\alpha > d$, the spectral density be what is known as regularly varying at infinity with exponent $-\alpha$ as the frequency increases along any ray from the origin. As an example, if $f(\mathbf{v}) \sim |\mathbf{A}\mathbf{v}|^{-\alpha}$ as $|\mathbf{v}| \to \infty$ for some nonsingular matrix $\mathbf{A}$, then f is regularly varying at infinity with exponent $-\alpha$ along all rays from the origin. If, in addition, $f \asymp 1$ on bounded sets, then it satisfies the conditions of Theorem 1 in the next section and (2) holds. One way to think about the conditions on f in Theorem 1 is that they require that the corresponding random field not be too different from a self-affine random field. We will return to this issue in Section 5.

Another asymptotic formulation we could use to study the screening effect would be to fix the observation grid by, for example, setting $\delta = 1$, selecting $\mathbf{x}$ and some bounded set B containing a neighborhood of the origin and looking at how $Ee(\mathbf{x}, (rB) \cap \mathbb{Z}^d)^2$ behaves as $r \to \infty$. We then trivially have a screening effect of sorts for any mean 0 random field Z with finite second moments: $Ee(\mathbf{x}, (rB) \cap \mathbb{Z}^d)^2 \to Ee(\mathbf{x}, \mathbb{Z}^d)^2$ as $r \to \infty$. It is exactly because such a result holds so generally that it is not informative and that Theorem 1, which excludes the counterexamples considered in this section, is.

Section 2 states and proves the main result of this paper on the screening effect. A critical step in this proof is Theorem 2, which shows that predictions under a presumed but incorrect spectral density can be asymptotically optimal even if the presumed spectral density is not asymptotically proportional to the correct spectral density as the norm of the frequency increases. Theorem 2 thus goes beyond Theorem 10 in Stein [(1999), Chapter 3] which proves asymptotically optimal prediction occurs when the ratio of the presumed to the true spectral densities tends to a positive constant as the norm of the frequency increases. The results in Section 2 all assume that the mean of the random field is known to be 0, which corresponds to what is called simple kriging in the geostatistical literature [Chilès and Delfiner (1999)]. It is common in practice to assume the mean is an unknown constant and then predict the random field at unobserved locations using what is known as ordinary kriging [Chilès and Delfiner (1999)], which is just a special case of best linear unbiased prediction. Section 3 shows that Theorem 2 also applies to ordinary kriging. Section 4 provides numerical results quantifying the screening effect in some limited circumstances, including a case in which the process is observed with measurement error, which is not treated by the theoretical results herein. Section 5 provides some discussion on the assumptions about the spectral densities, the nature of the asymptotic regime in Theorems 1 and 2 and some possible extensions of these results.

2. Main result. Suppose $\angle\mathbf{v} = \mathbf{v}/|\mathbf{v}|$, $b_d(r)$ is the d-dimensional ball of radius r centered at the origin and $\partial b_d(r)$ is the ball's surface. Throughout this work, we will assume that the spectral density f possesses the following properties:

(A1) $f \asymp 1$ on bounded subsets of $\mathbb{R}^d$.

(A2) There exists a positive function g on $\mathbb{R}^d$ such that $f \asymp g$ on $\mathbb{R}^d$, $f(\mathbf{v}) \sim g(\mathbf{v})$ as $|\mathbf{v}| \to \infty$ and $g(\mathbf{v}) = \tilde{g}(|\mathbf{v}|)\theta(\angle\mathbf{v})$ for some functions $\tilde{g}$ on $[0,\infty)$ and θ on $\partial b_d(1)$, where $\theta \asymp 1$ on $\partial b_d(1)$ and, for some $\alpha > d$,

$$\tilde{g}(r) = \begin{cases} r^{-\alpha}L(r), & \text{for } r \geq 1, \\ L(1), & \text{for } 0 \leq r < 1, \end{cases}$$

where L is slowly varying at ∞.

A positive function L on $[0,\infty)$ is said to be slowly varying at ∞ if, for every $r > 0$, $L(tr)/L(t) \to 1$ as $t \to \infty$. For example, $\log^b t$ is slowly varying at infinity for any value of b. The function $\tilde{g}$ given in (A2) is said to be regularly varying at ∞ with exponent $-\alpha$. Everything we will need about slowly varying functions is contained in Chapter 1 of Bingham, Goldie and Teugels (1987), to which we will refer henceforth as BGT.

From the Representation Theorem for slowly varying functions (BGT, page 12), the discussion on page 14 of BGT and the fact that we only require $f(\mathbf{v}) \sim g(\mathbf{v})$ as $|\mathbf{v}| \to \infty$ in (A2), we can assume without loss of generality that

$$L(r) = L_1 \exp\left\{\int_1^r \frac{\xi(u)}{u}\,du\right\} \tag{5}$$

for all $r \geq 1$, where $L_1 > 0$ and ξ is a bounded, measurable function on $(0,\infty)$ satisfying $\xi(u) \to 0$ as $u \to \infty$. For definiteness, set $L(r) = L_1$ for $0 \leq r < 1$, so that $L(1) = L_1$.

Suppose Z is a mean 0 weakly stationary random field on $\mathbb{R}^d$ with spectral density f. For $\mathbf{x} \in \mathbb{R}^d$ and $A \subseteq \mathbb{R}^d$, let $\hat{Z}_f(\mathbf{x}, A)$ be the best linear predictor of $Z(\mathbf{x})$ based on observing $Z(\mathbf{y})$ for all $\mathbf{y} \in A$ under the spectral density f. The error of this predictor is denoted by $e_f(\mathbf{x}, A) = Z(\mathbf{x}) - \hat{Z}_f(\mathbf{x}, A)$ and write $e_f(\mathbf{x}, \delta)$ as shorthand for $e_f(\mathbf{x}, \delta\mathbb{Z}^d)$. Denote expectations under the spectral density f by E_f.

THEOREM 1. *Suppose f satisfies* (A1) *and* (A2), $\mathbf{x}$ *is a nonvertex of* $[0,1]^d$ *and $B \subseteq \mathbb{R}^d$ contains some neighborhood of the origin. Then*

$$\lim_{r\to\infty}\lim_{\delta\downarrow 0} \frac{E_f e_f(\delta\mathbf{x}, \delta(rB \cap \mathbb{Z}^d))^2}{E_f e_f(\delta\mathbf{x}, \delta)^2} = 1.$$

Allowing the grid of observations to extend infinitely may appear unnatural for practical applications. The following corollary, in which observations are restricted to some bounded set, follows immediately from Theorem 1.

COROLLARY 1. *Suppose f satisfies* (A1) *and* (A2), $\mathbf{x}$ *is a nonvertex of* $[0,1]^d$ *and* $A, B \subseteq \mathbb{R}^d$ *each contain some neighborhood of the origin. Then*

$$\lim_{r\to\infty}\lim_{\delta\downarrow 0}\frac{E_f e_f\big(\delta\mathbf{x}, \delta(rB\cap\mathbb{Z}^d)\big)^2}{E_f e_f(\delta\mathbf{x}, A\cap\delta\mathbb{Z}^d)^2}=1.$$

The proof of Theorem 1 is rather roundabout and some of the intermediate steps are of independent interest, so it is worthwhile to outline the argument before giving details. The first step is to show that, asymptotically, there is no difference between using f and g to do the prediction and to evaluate their mean squared errors. This step is easy, since f and g satisfy all the conditions of Theorem 10 in Stein [(1999), page 102], so that, as $\delta \downarrow 0$,

$$E_f e_f(\delta\mathbf{x}, \delta)^2 \sim E_f e_g(\delta\mathbf{x}, \delta)^2 \sim E_g e_g(\delta\mathbf{x}, \delta)^2. \tag{6}$$

The next step, proven later in this section, is to evaluate the order of magnitude of $E_g e_g(\delta\mathbf{x}, \delta)^2$ as $\delta \downarrow 0$.

LEMMA 1. *Under the conditions of Theorem* 1, *as* $\delta \downarrow 0$, $E_g e_g(\delta\mathbf{x}, \delta)^2 \asymp \delta^{\alpha-d}L(\delta^{-1})$.

Now consider the function $\gamma(\mathbf{v}) = |\mathbf{v}|^{-\alpha}\theta(\angle\mathbf{v})$. Although γ is not integrable in a neighborhood of the origin, as we will describe later in this section, it can be thought of as the spectral "density" for a nonstationary random field. Furthermore, as we will show in Section 5, this random field is self-affine in a well-defined sense, so that (A2) can be thought of as saying that Z is approximately self-affine. Despite the fact that γ is not integrable, it is possible to give a sensible definition to $\hat{Z}_\gamma(\delta\mathbf{x}, \delta)$, the "best" linear predictor of $Z(\delta\mathbf{x})$ under γ, and we will do so later in this section. Under this definition, we will prove the following result.

THEOREM 2. *Under the conditions of Theorem* 1, *as* $\delta \downarrow 0$, $E_g\{\hat{Z}_g(\delta\mathbf{x}, \delta) - \hat{Z}_\gamma(\delta\mathbf{x}, \delta)\}^2 = o(E_g e_g(\delta\mathbf{x}, \delta)^2)$.

Theorem 2 is equivalent to $E_g e_\gamma(\delta\mathbf{x}, \delta)^2 - E_g e_g(\delta\mathbf{x}, \delta)^2 = o(E_g e_g(\delta\mathbf{x}, \delta)^2)$ as $\delta \downarrow 0$, so it implies that $\hat{Z}_\gamma(\delta\mathbf{x}, \delta)$ is an asymptotically optimal linear predictor under g. We also have that predictions under γ are asymptotically optimal when f is true:

COROLLARY 2. *Under the conditions of Theorem* 1, *as* $\delta \downarrow 0$, $E_f\{\hat{Z}_f(\delta\mathbf{x}, \delta) - \hat{Z}_\gamma(\delta\mathbf{x}, \delta)\}^2 = o(E_f e_f(\delta\mathbf{x}, \delta)^2)$.

PROOF. This result follows readily from Theorem 2, (6), $f \asymp g$ and

$$\begin{aligned} E_f\{\hat{Z}_f(\delta\mathbf{x}, \delta) - \hat{Z}_\gamma(\delta\mathbf{x}, \delta)\}^2 &\le 2E_f\{\hat{Z}_f(\delta\mathbf{x}, \delta) - \hat{Z}_g(\delta\mathbf{x}, \delta)\}^2 \\ &\quad + 2E_f\{\hat{Z}_g(\delta\mathbf{x}, \delta) - \hat{Z}_\gamma(\delta\mathbf{x}, \delta)\}^2. \qquad \square \end{aligned}$$

The advantage of considering γ is that the form of the optimal predictor of $Z(\delta\mathbf{x})$ based on observing $Z(\delta\mathbf{j})$ for $\mathbf{j} \in \mathbb{Z}^d$ under γ is independent of δ, which essentially follows from the fact that $\gamma(\delta\mathbf{v}) = \delta^{-\alpha}\gamma(\mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^d$ and all $\delta > 0$, although some care is needed since γ is not integrable. When $L \asymp 1$, this lack of dependence on δ, combined with Theorem 2, yields Theorem 1 rather directly. A somewhat more involved argument is needed to handle more general slowly varying L.

Corollary 2 provides a nontrivial advance over results in Stein [(1999), Chapter 3] on the asymptotic optimality of predictions based on a misspecified spectral density. In that work, it is always assumed that the ratio of the misspecified spectral density to the correct spectral density tends to a positive constant as the norm of the frequency tends to ∞, but here we allow this ratio to be a slowly varying function, which is not necessarily bounded away from either 0 or ∞.

Before we can proceed with the proofs, we need some definitions. Let $\mathcal{H}^0$ be the real linear hull of the random variables $Z(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^d$. The spectral density f defines an inner product on $\mathcal{H}^0$: for $h_1, h_2 \in \mathcal{H}^0$, $\langle h_1, h_2\rangle_f = E_f(h_1h_2)$. Let $\mathcal{H}(f)$ be the closure of $\mathcal{H}^0$ with respect to this inner product, so that $\mathcal{H}(f)$ is a Hilbert space. Now let $\mathcal{L}^0$ be the real linear hull of functions on $\mathbb{R}^d$ of the form $e^{i\mathbf{v}^T\mathbf{x}}$ for $\mathbf{x} \in \mathbb{R}^d$. For $\ell_1, \ell_2 \in \mathcal{L}^0$, define the inner product $\langle \ell_1, \ell_2\rangle_f = \int_{\mathbb{R}^d} \ell_1(\mathbf{v})\overline{\ell_2(\mathbf{v})} f(\mathbf{v})\,d\mathbf{v}$ and let $\mathcal{L}(f)$ be the closure of $\mathcal{L}^0$ with respect to this inner product. Identifying $Z(\mathbf{x})$ with $e^{i\mathbf{v}^T\mathbf{x}}$ and extending this identification to all elements in $\mathcal{H}(f)$ and $\mathcal{L}(f)$, we see that any statement we wish to make about the covariances of random variables in $\mathcal{H}(f)$ can be restated in terms of the inner products of functions in $\mathcal{L}(f)$.

Write $\mathcal{H}_\delta(f)$ for the subspace of $\mathcal{H}(f)$ generated by $Z(\delta\mathbf{j})$ for $\mathbf{j} \in \mathbb{Z}^d$, and let $\mathcal{L}_\delta(f)$ be the corresponding subspace of $\mathcal{L}(f)$. The best linear predictor $\hat{Z}_f(\delta\mathbf{x}, \delta)$ is in $\mathcal{H}_\delta(f)$ and the corresponding element in $\mathcal{L}_\delta(f)$ is [Stein (1999), page 99]

$$\hat{H}_f(\mathbf{v}; \delta\mathbf{x}, \delta) = \sum_{\mathbf{j}\in\mathbb{Z}^d} e^{i(\mathbf{v}+2\pi\delta^{-1}\mathbf{j})^T\mathbf{x}} \frac{f(\mathbf{v}+2\pi\delta^{-1}\mathbf{j})}{f_\delta(\mathbf{v})}, \tag{7}$$

where $f_\delta(\mathbf{v}) = \sum_{\mathbf{j}\in\mathbb{Z}^d} f(\mathbf{v}+2\pi\delta^{-1}\mathbf{j})$. As a function of $\mathbf{v}$, $\hat{H}_f(\mathbf{v}; \delta\mathbf{x}, \delta)$ has period $2\pi\delta^{-1}$ in each coordinate. Thus, defining $A_d(t) = (-\pi t, \pi t]^d$ and writing $\sum'_{\mathbf{j}}$ to indicate summation over all $\mathbf{j} \in \mathbb{Z}^d$ except the origin, we have, for example,

$$E_f e_g(\delta\mathbf{x}, \delta)^2 = \int_{A_d(\delta^{-1})} \left| \sum_{\mathbf{j}}{}' (1 - e^{2\pi i\mathbf{j}^T\mathbf{x}}) g(\mathbf{v}+2\pi\delta^{-1}\mathbf{j}) \right|^2 \frac{f_\delta(\mathbf{v})}{g_\delta(\mathbf{v})^2}\,d\mathbf{v}. \tag{8}$$

PROOF OF LEMMA 1. Let $\mathbf{e}_\ell \in \mathbb{R}^d$ be the unit vector along the ℓth coordinate axis. Now, as $\delta \downarrow 0$, $\sum'_{\mathbf{j}} g(\mathbf{v}+2\pi\delta^{-1}\mathbf{j}) \ge g(\mathbf{v}+2\pi\delta^{-1}\mathbf{e}_1) \asymp \delta^\alpha L(\delta^{-1})$. Define Q_d to be the subset of $\mathbb{R} \times \mathbb{R}^d$ given by $\{(\delta, \mathbf{v}) : 0 < \delta < 1, \mathbf{v} \in A_d(\delta^{-1})\}$. Since L is slowly varying, $g(\mathbf{v}+2\pi\delta^{-1}\mathbf{j}) \asymp \delta^\alpha \mathbf{j}^{-\alpha} L(\delta^{-1}|\mathbf{j}|)$ for $(\delta,\mathbf{v},\mathbf{j}) \in Q_d \times (\mathbb{Z}^d \setminus \{\mathbf{0}\})$, so that

$$\begin{aligned} \sum_{\mathbf{j}}{}' g(\mathbf{v}+2\pi\delta^{-1}\mathbf{j}) &\ll \delta^\alpha \sum_{\mathbf{j}}{}' |\mathbf{j}|^{-\alpha} L(\delta^{-1}|\mathbf{j}|) \ll \delta^\alpha \int_1^\infty r^{-\alpha+d-1} L(\delta^{-1}r)\,dr \\ &= \delta^d \int_{\delta^{-1}}^\infty r^{-\alpha+d-1} L(r)\,dr \ll \delta^\alpha L(\delta^{-1}) \end{aligned} \tag{9}$$

for $(\delta,\mathbf{v}) \in Q_d$ by 1.5.10 in BGT. Thus, for $(\delta,\mathbf{v}) \in Q_d$,

$$\sum_{\mathbf{j}}{}' g(\mathbf{v}+2\pi\delta^{-1}\mathbf{j}) \asymp \delta^\alpha L(\delta^{-1}) \tag{10}$$

and $g(\mathbf{v}) \asymp g_\delta(\mathbf{v})$. Since $f_\delta(\mathbf{v}) \asymp g_\delta(\mathbf{v})$ for $(\delta,\mathbf{v}) \in Q_d$ and $|1-e^{2\pi i\mathbf{j}^T\mathbf{x}}| \le 2$, from (8) (with $f=g$) and (10), $E_g e_g(\delta\mathbf{x},\delta)^2 \ll \delta^{2\alpha} L(\delta^{-1})^2 \int_{A_d(\delta^{-1})} g(\mathbf{v})^{-1}\,d\mathbf{v}$ as $\delta \downarrow 0$. By Proposition 1.5.8 in BGT, as $\delta \downarrow 0$,

$$\int_{A_d(\delta^{-1})} \frac{1}{g(\mathbf{v})}\,d\mathbf{v} \asymp 1 + \int_1^{\delta^{-1}} \frac{r^{d-1}}{r^{-\alpha}L(r)}\,dr \asymp \frac{\delta^{-\alpha-d}}{L(\delta^{-1})}, \tag{11}$$

so $E_g e_g(\delta\mathbf{x},\delta)^2 \ll \delta^{\alpha-d} L(\delta^{-1})$ as $\delta \downarrow 0$.

To complete the proof, we need to show $\delta^{\alpha-d}L(\delta^{-1}) \ll E_g e_g(\delta\mathbf{x},\delta)^2$ as $\delta \downarrow 0$. Since $\mathbf{x} = (x_1,\dots,x_d)^T$ is a nonvertex of $[0,1]^d$, it has a component, say the ℓth, with $x_\ell \in (0,1)$. As $\delta \downarrow 0$, just taking the term $\mathbf{j} = \mathbf{e}_\ell$ in the sum over $\mathbf{j}$ on the right side of (8),

$$\begin{aligned} E_g e_g(\delta\mathbf{x},\delta)^2 &\ge \{1-\cos(2\pi x_\ell)\}^2 \int_{A_d(\delta^{-1})} \frac{g(\mathbf{v}+2\pi\delta^{-1}\mathbf{e}_\ell)^2}{g_\delta(\mathbf{v})}\,d\mathbf{v} \\ &\gg \delta^{2\alpha} L(\delta^{-1})^2 \int_{A_d(\delta^{-1})} \frac{1}{g(\mathbf{v})}\,d\mathbf{v} \gg \delta^{\alpha-d} L(\delta^{-1}) \end{aligned}$$

and Lemma 1 follows. □

The following lemma is helpful in proving Theorems 1 and 2:

LEMMA 2. *If $\xi : (1,\infty) \to \mathbb{R}$ is bounded and measurable, $\xi(u) \to 0$ as $u \to \infty$ and, for some $C > 0$, $L(r) = C\exp\{\int_1^r \frac{\xi(u)}{u}\,du\}$ for all $r \ge 1$, then there exists a positive increasing function σ on $(0,\infty)$ such that $\sigma(r) \to \infty$, $r/\sigma(r) \to \infty$ as $r \to \infty$ and*

$$\lim_{r\to\infty} \sup_{x,y\in[r/\sigma(r),r\sigma(r)]} \left|\log\frac{L(x)}{L(y)}\right| = 0.$$

PROOF. For $r \geq 1$, define $\eta(r) = \sup_{s>r} |\xi(s)|$, so that η is decreasing on $[1, \infty)$ and $\eta(r) \downarrow 0$ as $r \to \infty$. Suppose, without loss of generality, that $x > y > 1$. Then

$$\left|\log \frac{L(x)}{L(y)}\right| \leq \int_y^x \frac{|\xi(u)|}{u}\, du \leq \eta(y) \log \frac{x}{y}. \tag{12}$$

It is possible to find a continuous strictly increasing function τ on $(0, \infty)$ that increases sufficiently rapidly to make $\tau(x)/x \to \infty$ as $x \to \infty$ and

$$\lim_{x\to\infty} \eta\bigl(\tau(x)/x\bigr) \log x = 0.$$

Let σ be the inverse of τ, so that $\sigma(r) \to \infty$ as $r \to \infty$ as required.

Setting $s = \sigma(r)$, we see that $r/\sigma(r) = \tau(s)/s$, so $\lim_{r\to\infty} r/\sigma(r) = \lim_{s\to\infty} \tau(s)/s = \infty$ as required. In addition, by (12),

$$\begin{aligned}\lim_{r\to\infty} \sup_{x,y\in[r/\sigma(r), r\sigma(r)]} \left|\log \frac{L(x)}{L(y)}\right| &\leq \lim_{r\to\infty} 2\eta\bigl(r/\sigma(r)\bigr) \log\{\sigma(r)\} \\ &= \lim_{s\to\infty} 2\eta\bigl(\tau(s)/s\bigr) \log s = 0,\end{aligned}$$

proving the lemma. $\square$

PROOF OF THEOREM 2. We have

$$\begin{aligned}&\bigl|\hat{H}_g(\mathbf{v}; \delta\mathbf{x}, \delta) - \hat{H}_\gamma(\mathbf{v}; \delta\mathbf{x}, \delta)\bigr| \\ &\quad= \left|\sum_{\mathbf{j}} e^{2\pi i\delta^{-1}\mathbf{j}^T\mathbf{x}} \left\{\frac{g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j})}{g_\delta(\mathbf{v})} - \frac{\gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j})}{\gamma_\delta(\mathbf{v})}\right\}\right| \\ &\quad\leq \sum_{\mathbf{j}} \left|\frac{g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j})\gamma_\delta(\mathbf{v}) - \gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}) g_\delta(\mathbf{v})}{g_\delta(\mathbf{v})\gamma_\delta(\mathbf{v})}\right| \\ &\quad\leq \frac{1}{g_\delta(\mathbf{v})\gamma_\delta(\mathbf{v})} \sum_{\mathbf{j}\neq\mathbf{k}} \bigl|\gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{k}) g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}) \\ &\qquad\qquad - \gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}) g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{k})\bigr|.\end{aligned} \tag{13}$$

Now, for $(\delta, \mathbf{v}) \in Q_d$,

$$\begin{aligned}&\sum_{\mathbf{j}\neq\mathbf{k}} \bigl|\gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{k}) g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}) - \gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}) g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{k})\bigr| \\ &\quad\leq 2 \sum_{\mathbf{j}\neq\mathbf{k}} \gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{k}) g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}) \\ &\quad\leq 2\gamma(\mathbf{v}) {\sum_{\mathbf{j}}}' g(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}) + 2g(\mathbf{v}) {\sum_{\mathbf{j}}}' \gamma(\mathbf{v} + 2\pi\delta^{-1}\mathbf{j})\end{aligned}$$

$$
\begin{aligned}
&+2{\sum_{\mathbf{j}}}' \gamma(\mathbf{v}+2\pi\delta^{-1}\mathbf{j}){\sum_{\mathbf{j}}}' g(\mathbf{v}+2\pi\delta^{-1}\mathbf{j}) \\
&\ll |\mathbf{v}|^{-\alpha}\delta^{\alpha}L(\delta^{-1})+|\mathbf{v}|^{-\alpha}L(|\mathbf{v}|)\delta^{\alpha}+\delta^{2\alpha}L(\delta^{-1}) \\
&\ll |\mathbf{v}|^{-\alpha}\delta^{\alpha}\{L(\delta^{-1})+L(|\mathbf{v}|)\}.
\end{aligned}
$$

It follows that, for $(\delta, \mathbf{v}) \in Q_d$,

$$
\bigl|\hat{H}_g(\mathbf{v};\delta\mathbf{x},\delta)-\hat{H}_\gamma(\mathbf{v};\delta\mathbf{x},\delta)\bigr|^2 g_\delta(\mathbf{v}) \ll \delta^{2\alpha}\left\{\frac{L(\delta^{-1})^2}{L(|\mathbf{v}|)}+L(|\mathbf{v}|)\right\}\max(1,|\mathbf{v}|^\alpha).
$$

Define $\mu(\delta)=\delta^{-1}/\sigma(\delta^{-1})$. Then as $\delta \downarrow 0$,

$$
\begin{aligned}
&\int_{b_d(\mu(\delta))\setminus b_d(1)} \bigl|\hat{H}_g(\mathbf{v};\delta\mathbf{x},\delta)-\hat{H}_\gamma(\mathbf{v};\delta\mathbf{x},\delta)\bigr|^2 g_\delta(\mathbf{v})\,d\mathbf{v} \\
&\qquad \ll \delta^{2\alpha}\int_1^{\mu(\delta)}\left\{\frac{L(\delta^{-1})^2}{L(r)}+L(r)\right\} r^{\alpha+d-1}\,dr \\
&\qquad \ll \frac{\delta^{\alpha-d}}{\sigma(\delta^{-1})^{\alpha+d}}\left\{\frac{L(\delta^{-1})^2}{L(\mu(\delta))}+L(\mu(\delta))\right\} \\
&\qquad \ll \frac{\delta^{\alpha-d}L(\delta^{-1})}{\sigma(\delta^{-1})^{\alpha+d}}
\end{aligned}
\tag{14}
$$

using 1.5.10 in BGT and Lemma 2. Furthermore,

$$
\begin{aligned}
\int_{b_d(1)} \bigl|\hat{H}_g(\mathbf{v};\delta\mathbf{x},\delta)-\hat{H}_\gamma(\mathbf{v};\delta\mathbf{x},\delta)\bigr|^2 g_\delta(\mathbf{v})\,d\mathbf{v} &\ll \delta^{2\alpha}\{1+L(\delta^{-1})^2\} \\
&= o\bigl(\delta^{\alpha-d}L(\delta^{-1})\bigr)
\end{aligned}
\tag{15}
$$

as $\delta \downarrow 0$.

Next, define

$$
\begin{aligned}
\phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) = |\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|^{-\alpha}|\mathbf{v}+2\pi\delta^{-1}\mathbf{k}|^{-\alpha}\bigl|&L(|\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|) \\
&-L(|\mathbf{v}+2\pi\delta^{-1}\mathbf{k}|)\bigr|.
\end{aligned}
$$

From (13), for $\mathbf{v} \in A_d(\delta^{-1}) \setminus b_d(1)$,

$$
\bigl|\hat{H}_g(\mathbf{v};\delta\mathbf{x},\delta)-\hat{H}_\gamma(\mathbf{v};\delta\mathbf{x},\delta)\bigr| \ll \frac{1}{g_\delta(\mathbf{v})\gamma_\delta(\mathbf{v})}\sum_{\mathbf{j}\neq\mathbf{k}}\phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k})
\tag{16}
$$

as $\delta \downarrow 0$, and, for any $j_0 > 0$, using $\phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) = \phi_\delta(\mathbf{v};\mathbf{k},\mathbf{j})$,

$$
\sum_{\mathbf{j}\neq\mathbf{k}}\phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) \le 2\sum_{|\mathbf{j}|>j_0}\sum_{\mathbf{k}\in\mathbb{Z}^d}\phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) + 2\sum_{0<|\mathbf{j}|\le j_0}\sum_{|\mathbf{k}|\le j_0}\phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}).
\tag{17}
$$

Now, for any $\mathbf{v} \in A_d(\delta^{-1}) \setminus b_d(1)$,

$$
\begin{aligned}
&\sum_{|\mathbf{j}|>\sigma(\delta^{-1})} \sum_{\mathbf{k}\in\mathbb{Z}^d} \phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) \\
&\quad \le \sum_{|\mathbf{j}|>\sigma(\delta^{-1})} |\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|^{-\alpha} L(|\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|) \sum_{\mathbf{k}\in\mathbb{Z}^d} |\mathbf{v}+2\pi\delta^{-1}\mathbf{k}|^{-\alpha} \\
&\qquad + \sum_{|\mathbf{j}|>\sigma(\delta^{-1})} |\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|^{-\alpha} \sum_{\mathbf{k}\in\mathbb{Z}^d} |\mathbf{v}+2\pi\delta^{-1}\mathbf{k}|^{-\alpha} L(|\mathbf{v}+2\pi\delta^{-1}\mathbf{k}|).
\end{aligned}
\tag{18}
$$

Define $Q'_d = \{(\delta,\mathbf{v}) : 0<\delta<1, \mathbf{v} \in A_d(\delta^{-1}) \setminus b_d(1)\}$. Refining the analysis in (9), for $(\delta,\mathbf{v}) \in Q'_d$,

$$
\begin{aligned}
&\sum_{|\mathbf{j}|>\sigma(\delta^{-1})} |\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|^{-\alpha} L(|\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|) \\
&\quad \ll \delta^d \int_{\sigma(\delta^{-1})\delta^{-1}}^{\infty} r^{-\alpha+d-1} L(r)\,dr \ll \frac{\delta^\alpha L(\delta^{-1})}{\sigma(\delta^{-1})^{\alpha-d}}
\end{aligned}
\tag{19}
$$

and

$$
\sum_{|\mathbf{j}|>\sigma(\delta^{-1})} |\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|^{-\alpha} \ll \frac{\delta^\alpha}{\sigma(\delta^{-1})^{\alpha-d}}. \tag{20}
$$

From (18)–(20), we see that, for $(\delta,\mathbf{v}) \in Q'_d$,

$$
\sum_{|\mathbf{j}|>\sigma(\delta^{-1})} \sum_{\mathbf{k}\in\mathbb{Z}^d} \phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) \ll \frac{\delta^\alpha |\mathbf{v}|^{-\alpha}}{\sigma(\delta^{-1})^{\alpha-d}} \{L(|\mathbf{v}|) + L(\delta^{-1})\}.
$$

It follows that

$$
\begin{aligned}
&\int_{A_d(\delta^{-1})\setminus b_d(\mu(\delta))} \left| \sum_{|\mathbf{j}|>\sigma(\delta^{-1})} \sum_{\mathbf{k}\in\mathbb{Z}^d} \phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) \right|^2 \frac{1}{g_\delta(\mathbf{v})\gamma_\delta(\mathbf{v})^2}\,d\mathbf{v} \\
&\quad \ll \frac{\delta^{2\alpha}}{\sigma(\delta^{-1})^{2(\alpha-d)}} \int_1^{\delta^{-1}} \{L(r)^2 + L(\delta^{-1})^2\} \frac{r^{\alpha+d-1}}{L(r)}\,dr \\
&\quad \ll \frac{\delta^{\alpha-d} L(\delta^{-1})}{\sigma(\delta^{-1})^{2(\alpha-d)}}
\end{aligned}
\tag{21}
$$

as $\delta \downarrow 0$. Next, define

$$
C(\delta) = \sup_{\mathbf{v}\in A_d(\delta^{-1})\setminus b_d(\mu(\delta))} \max_{|\mathbf{j}|,|\mathbf{k}|\le\sigma(\delta^{-1})} \left| L(|\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|) - L(|\mathbf{v}+2\pi\delta^{-1}\mathbf{k}|) \right|.
$$

Lemma 2 implies that $\lim_{\delta\downarrow 0} C(\delta)/L(\delta^{-1}) = 0$. Thus,

$$\begin{aligned}
&\int_{A_d(\delta^{-1})\backslash b_d(\mu(\delta))} \left\{ \sum_{0<|\mathbf{j}|\le\sigma(\delta^{-1})} \sum_{|\mathbf{k}|\le\sigma(\delta^{-1})} \phi_\delta(\mathbf{v};\mathbf{j},\mathbf{k}) \right\}^2 \frac{1}{g_\delta(\mathbf{v})\gamma_\delta(\mathbf{v})^2}\, d\mathbf{v} \\
&\qquad \ll C(\delta)^2 \int_{A_d(\delta^{-1})\backslash b_d(1)} \left\{ {\sum_{\mathbf{j}}}' |\mathbf{v}+2\pi\delta^{-1}\mathbf{j}|^{-\alpha} \sum_{\mathbf{k}\in\mathbb{Z}^d} |\mathbf{v}+2\pi\delta^{-1}\mathbf{k}|^{-\alpha} \right\}^2 \\
&\qquad\qquad\qquad \times \frac{1}{g_\delta(\mathbf{v})\gamma_\delta(\mathbf{v})^2}\, d\mathbf{v} \qquad (22)\\
&\qquad \ll \frac{C(\delta)^2}{L(\delta^{-1})}\delta^{\alpha-d} = o\big(L(\delta^{-1})\delta^{\alpha-d}\big)
\end{aligned}$$

as $\delta \downarrow 0$. From (16), (17), (21) and (22), we get

$$\int_{A_d(\delta^{-1})\backslash b_d(\mu(\delta))} \big|\hat{H}_g(\mathbf{v};\delta\mathbf{x},\delta) - \hat{H}_\gamma(\mathbf{v};\delta\mathbf{x},\delta)\big|^2 g_\delta(\mathbf{v})\, d\mathbf{v} = o\big(\delta^{\alpha-d}L(\delta^{-1})\big)$$

as $\delta \downarrow 0$, which, together with (14), (15) and Lemma 1, implies Theorem 2. □

PROOF OF THEOREM 1. To prove that Theorem 1 follows from Theorem 2, we need to consider in what sense $\hat{H}_\gamma(\cdot;\delta\mathbf{x},\delta)$ corresponds to an optimal linear predictor of $Z(\delta\mathbf{x})$ based on observing $Z(\delta\mathbf{j})$ for $\mathbf{j}\in\mathbb{Z}^d$ under the model γ. Set $p = \lfloor \frac{1}{2}(\alpha-d)\rfloor$, where $\lfloor\cdot\rfloor$ is the greatest integer function. Although γ is not integrable, we see that $\int_{\mathbb{R}^d} |\mathbf{v}|^{2p+2}(1+|\mathbf{v}|)^{-(2p+2)}\gamma(\mathbf{v})\, d\mathbf{v} < \infty$, which implies that γ is the spectral "density" of what is known as an intrinsic random function of order p, or IRF-p [Chilès and Delfiner (1999), Stein (1999)]. Define an authorized linear combination of order p, or ALC-p, to be a random variable of the form $\sum_{j=1}^n \lambda_j Z(\mathbf{x}_j)$ for which $\sum_{j=1}^n \lambda_j P(\mathbf{x}_j) = 0$ for all polynomials $P(\mathbf{x})$ of order at most p. If $\sum_{j=1}^n \lambda_j Z(\mathbf{x}_j)$ is an ALC-p, then γ defines its second moment through

$$E_\gamma\left\{\sum_{j=1}^n \lambda_j Z(\mathbf{x}_j)\right\}^2 = \int_{\mathbb{R}^d} \left|\sum_{j=1}^n \lambda_j e^{i\mathbf{v}^T\mathbf{x}_j}\right|^2 \gamma(\mathbf{v})\, d\mathbf{v},$$

which is finite. If $\sum_{j=1}^n \lambda_j Z(\mathbf{x}_j)$ is not an ALC-p, then

$$\int_{\mathbb{R}^d} \left|\sum_{j=1}^n \lambda_j e^{i\mathbf{v}^T\mathbf{x}_j}\right|^2 \gamma(\mathbf{v})\, d\mathbf{v} = \infty$$

and γ does not define the second moment of $\sum_{j=1}^n \lambda_j Z(\mathbf{x}_j)$.

Fixing $\mathbf{x} \in [0,1]^d$, define the linear manifold $\mathcal{L}_\delta^0$ to be the class of functions of the form $\sum_{\mathbf{j}\in B} \beta_j e^{i\delta\mathbf{v}^T\mathbf{j}}$ for which the β_js are real, $B \subseteq \mathbb{Z}^d$ is finite and

$$\int_{\mathbb{R}^d} \left| e^{i\delta\mathbf{v}^T\mathbf{x}} - \sum_{\mathbf{j}\in B} e^{i\delta\mathbf{v}^T\mathbf{j}} \right|^2 \gamma(\mathbf{v})\,d\mathbf{v} < \infty.$$

For every $\mathbf{x} \in [0,1]^d$ and γ satisfying $\int_{\mathbb{R}^d} |\mathbf{v}|^{2p+2}(1+|\mathbf{v}|)^{-(2p+2)}\gamma(\mathbf{v})\,d\mathbf{v} < \infty$, it is possible to show that $\mathcal{L}_\delta^0$ is not empty. For any two functions $\hat{H}_1, \hat{H}_2$ in $\mathcal{L}_\delta^0$, define $\|\hat{H}_1 - \hat{H}_2\|_\gamma^2 = \int_{\mathbb{R}^d} \left|\hat{H}_1(\mathbf{v}) - \hat{H}_2(\mathbf{v})\right|^2 \gamma(\mathbf{v})\,d\mathbf{v}$ and let $\mathcal{L}_\delta(\gamma)$ be the completion of $\mathcal{L}_\delta^0$ with respect to the metric $\|\cdot - \cdot\|_\gamma$. By an easy generalization of the argument leading to (7) [Stein (1999), page 99], we can show that $\hat{H}_\gamma(\cdot;\delta\mathbf{x},\delta)$ as defined in (7) minimizes $\|e^{i\delta\cdot^T\mathbf{x}} - \hat{H}\|_\gamma$ among all $\hat{H} \in \mathcal{L}_\delta(\gamma)$. Defining $\hat{Z}_\gamma(\delta\mathbf{x},\delta)$ to be the random variable corresponding to $\hat{H}_\gamma(\cdot;\delta\mathbf{x},\delta)$, we then have that $\hat{Z}_\gamma(\delta\mathbf{x},\delta)$ is an optimal linear predictor of $Z(\mathbf{x})$ in the sense that it minimizes the mean squared prediction error among all linear predictors of $Z(\delta\mathbf{x})$ whose prediction errors are ALC-ps.

Let us now consider approximating $\hat{H}_\gamma = \hat{H}_\gamma(\cdot;\mathbf{x},1)$ by an element of $\mathcal{L}_1^0$. Specifically, given $\varepsilon > 0$, there is a function $\hat{H}_\gamma^\varepsilon$ of the form

$$\hat{H}_\gamma^\varepsilon(\mathbf{v}) = \sum_{\mathbf{j}\in B_\varepsilon} c_{\mathbf{j}}(\varepsilon) e^{i\mathbf{v}^T\mathbf{j}},$$

where $B_\varepsilon \subseteq \mathbb{Z}^d$ is finite and

$$\|\hat{H}_\gamma^\varepsilon - \hat{H}_\gamma\|_\gamma^2 < \varepsilon \|\hat{H}_\gamma - e^{i\cdot^T\mathbf{x}}\|_\gamma^2. \tag{23}$$

Define

$$\hat{H}_\gamma^\varepsilon(\mathbf{v};\delta\mathbf{x},\delta) = \sum_{\mathbf{j}\in B_\varepsilon} c_{\mathbf{j}}(\varepsilon) e^{i\delta\mathbf{v}^T\mathbf{j}},$$

so that $\hat{H}_\gamma^\varepsilon(\cdot;\mathbf{x},1) = \hat{H}_\gamma^\varepsilon$. If $\sum_{j=1}^n \lambda_j Z(\mathbf{x}_j)$ is an ALC-p, then $\sum_{j=1}^n \lambda_j Z(\delta\mathbf{x}_j)$ is also an ALC-p for any $\delta > 0$ and, furthermore,

$$E_\gamma \left\{ \sum_{j=1}^n \lambda_j Z(\delta\mathbf{x}_j) \right\}^2 = \delta^{\alpha-d} E_\gamma \left\{ \sum_{j=1}^n \lambda_j Z(\mathbf{x}_j) \right\}^2.$$

Thus, letting $\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta)$ be the random variable corresponding to $\hat{H}_\gamma^\varepsilon(\cdot;\delta\mathbf{x},\delta)$, (23) implies

$$E_\gamma\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - \hat{Z}_\gamma(\delta\mathbf{x},\delta)\}^2 < \varepsilon E_\gamma\{\hat{Z}_\gamma(\delta\mathbf{x},\delta) - Z(\delta\mathbf{x})\}^2, \tag{24}$$

so that $\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta)$ is a nearly optimal predictor of $Z(\delta\mathbf{x})$ under γ.

The last major hurdle in proving Theorem 1 is to show that something like (24) holds for δ sufficiently small when the E_γs are replaced by E_gs. Specifically, we seek to prove that there exists a $C < \infty$ independent of ε and a $\delta(\varepsilon) > 0$ such that

$$E_g\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x}, \delta) - \hat{Z}_\gamma(\delta\mathbf{x}, \delta)\}^2 < C\varepsilon E_g\{\hat{Z}_\gamma(\delta\mathbf{x}, \delta) - Z(\delta\mathbf{x})\}^2 \tag{25}$$

for all $\delta < \delta(\varepsilon)$. Now

$$\hat{H}_\gamma(\mathbf{v}; \delta\mathbf{x}, \delta) = \hat{H}_\gamma(\delta\mathbf{v}; \mathbf{x}, 1)$$

and

$$\hat{H}_\gamma^\varepsilon(\mathbf{v}; \delta\mathbf{x}, \delta) = \hat{H}_\gamma^\varepsilon(\delta\mathbf{v}; \mathbf{x}, 1),$$

so, defining $\theta_0 = \sup_{\mathbf{v}} \theta(\angle\mathbf{v})$ and recalling that $L(r) = L(1)$ for $0 < r < 1$,

$$\begin{aligned} E_g\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x}, \delta) - \hat{Z}_\gamma(\delta\mathbf{x}, \delta)\}^2 &= \int_{\mathbb{R}^d} \big|\hat{H}_\gamma(\mathbf{v}; \delta\mathbf{x}, \delta) - \hat{H}_\gamma^\varepsilon(\mathbf{v}; \delta\mathbf{x}, \delta)\big|^2 g(\mathbf{v})\, d\mathbf{v} \\ &\le \theta_0 \delta^{\alpha-d} \int_{\mathbb{R}^d} \big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2 |\mathbf{v}|^{-\alpha} L(\delta^{-1}|\mathbf{v}|)\, d\mathbf{v}. \end{aligned}$$

From Theorem 2, we see that (25) follows if there exists a $C < \infty$ independent of ε and a $\delta(\varepsilon) > 0$ [not necessarily the same C and $\delta(\varepsilon)$ as in (25)] such that

$$\begin{aligned} &\int_{\mathbb{R}^d} \big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2 |\mathbf{v}|^{-\alpha} L(\delta^{-1}|\mathbf{v}|)\, d\mathbf{v} \\ &\quad = \int_{A_d(1)} |\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})|^2 \sum_{\mathbf{j}\in\mathbb{Z}^d} |\mathbf{v} + 2\pi\delta^{-1}\mathbf{j}|^{-\alpha} L(\delta^{-1}|\mathbf{v} + 2\pi\mathbf{j}|)\, d\mathbf{v} \\ &\quad < C\varepsilon L(\delta^{-1}) \end{aligned} \tag{26}$$

for all $\delta < \delta(\varepsilon)$. Along the lines of (9), it is possible to show that, for $(\delta, \mathbf{v}) \in Q_d$,

$$\sum_{\mathbf{j}} |\mathbf{v} + 2\pi\mathbf{j}|^{-\alpha} L(\delta^{-1}|\mathbf{v} + 2\pi\mathbf{j}|) \ll |\mathbf{v}|^{-\alpha} L(\delta^{-1}|\mathbf{v}|),$$

so, to prove (26), and hence (25), it suffices to show there exists a $C < \infty$ independent of ε and a $\delta(\varepsilon) > 0$ such that

$$\int_{A_d(1)} \big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2 |\mathbf{v}|^{-\alpha} L(\delta^{-1}|\mathbf{v}|)\, d\mathbf{v} < C\varepsilon L(\delta^{-1}) \tag{27}$$

for all $\delta < \delta(\varepsilon)$. If $L \asymp 1$, then (27) is a trivial consequence of (23).

The general case needs greater care. For any $\varepsilon > 0$, by Lemma 2,

$$\begin{aligned} &\int_{A_d(1)\backslash b_d(1/\sigma(\delta^{-1}))} \big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2 |\mathbf{v}|^{-\alpha} L(\delta^{-1}|\mathbf{v}|)\, d\mathbf{v} \\ &\quad \le 2L(\delta^{-1}) \int_{A_d(1)\backslash b_d(1/\sigma(\delta^{-1}))} \big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2 |\mathbf{v}|^{-\alpha}\, d\mathbf{v} \\ &\quad < 2\theta_0 L(\delta^{-1})\varepsilon \|\hat{H}_\gamma - \hat{H}_\gamma^\varepsilon\|_\gamma^2 \end{aligned} \tag{28}$$

for all δ sufficiently small. Now

$$(29)\qquad \big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2 \le 2\big|\hat{H}_\gamma(\mathbf{v}) - e^{i\mathbf{v}^T\mathbf{x}}\big|^2 + 2\big|\hat{H}_\gamma^\varepsilon(\mathbf{v}) - e^{i\mathbf{v}^T\mathbf{x}}\big|^2$$

and there exists $C_1 < \infty$ such that

$$(30)\qquad \big|\hat{H}_\gamma(\mathbf{v}) - e^{i\mathbf{v}^T\mathbf{x}}\big|^2 \le C_1|\mathbf{v}|^{2\alpha}$$

for all $\mathbf{v}$. In addition, I claim that, given $\varepsilon > 0$, there exists $K_\varepsilon < \infty$ such that

$$(31)\qquad \big|\hat{H}_\gamma^\varepsilon(\mathbf{v}) - e^{i\mathbf{v}^T\mathbf{x}}\big|^2 \le K_\varepsilon|\mathbf{v}|^{2p+2}.$$

To prove (31), notice that, if it is false, then it is false in a neighborhood of the origin and, by considering a Taylor series in $\mathbf{v}$ about the origin for

$$\hat{H}_\gamma^\varepsilon(\mathbf{v}) = \sum_{\mathbf{j}\in B_\varepsilon} c_{\mathbf{j}}(\varepsilon)e^{i\mathbf{v}^T\mathbf{j}},$$

we see that, for some $\ell \in \{1,\dots,d\}$ and some $C_2 > 0$, $|\hat{H}_\gamma^\varepsilon(\mathbf{v}) - e^{i\mathbf{v}^T\mathbf{x}}|^2 \ge C_2 v_\ell^{2p}$ for all $\mathbf{v}$ in some neighborhood of the origin. However, this lower bound, together with the equality $p = \lfloor\frac{1}{2}(\alpha - d)\rfloor$, contradicts the finiteness of $\int_{\mathbb{R}^d}|\hat{H}_\gamma^\varepsilon(\mathbf{v}) - e^{i\mathbf{v}^T\mathbf{x}}|^2|\mathbf{v}|^{-\alpha}d\mathbf{v}$, so (31) must be true. Define S_d to be the surface area of $b_d(1)$ and apply (29)–(31) and Lemma 2 to obtain, for all δ sufficiently small,

$$\begin{aligned}
&\int_{b_d(1/\sigma(\delta^{-1}))}\big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2|\mathbf{v}|^{-\alpha}L(\delta^{-1}|\mathbf{v}|)\,d\mathbf{v}\\
&\quad\le 2\int_{b_d(1/\sigma(\delta^{-1}))}\big\{C_1|\mathbf{v}|^{2\alpha} + K_\varepsilon|\mathbf{v}|^{2p+2}\big\}|\mathbf{v}|^{-\alpha}L(\delta^{-1}|\mathbf{v}|)\,d\mathbf{v}\\
&\quad= 2S_d\delta^d\int_0^{\mu(\delta)}\big\{C_1\delta^\alpha r^{\alpha+d-1} + K_\varepsilon\delta^{2p-\alpha+2}r^{2p-\alpha+d+1}\big\}L(r)\,dr\\
&\quad\le 4S_d\delta^d\left\{C_1\delta^\alpha\frac{\mu(\delta)^{\alpha+d}}{\alpha+d} + K_\varepsilon\delta^{2p-\alpha+2}\frac{\mu(\delta)^{2p-\alpha+d+2}}{2p-\alpha+d+2}\right\}L\left(\frac{1}{\delta\sigma(\delta^{-1})}\right)\\
&\quad\le 8S_dL(\delta^{-1})\left\{\frac{C_1}{(\alpha+d)\sigma(\delta^{-1})^{\alpha+d}} + \frac{K_\varepsilon}{(2p-\alpha+d+2)\sigma(\delta^{-1})^{2p-\alpha+d+2}}\right\}.
\end{aligned}$$

Since $\sigma(\delta^{-1}) \to 0$ as $\delta \downarrow 0$ and $2p+2 > \alpha - d$,

$$\lim_{\delta\downarrow 0}\frac{1}{L(\delta^{-1})}\int_{b_d(1/\sigma(\delta^{-1}))}\big|\hat{H}_\gamma(\mathbf{v}) - \hat{H}_\gamma^\varepsilon(\mathbf{v})\big|^2|\mathbf{v}|^{-\alpha}L(\delta^{-1}|\mathbf{v}|)\,d\mathbf{v} = 0,$$

which, together with (28), implies (27) and hence (25).

From Lemma 1 and (25) we can choose a $C < \infty$ such that, given $\varepsilon > 0$,

$$\varlimsup_{\delta\downarrow 0}\frac{E_g\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - Z(\delta\mathbf{x})\}^2}{E_g e_g(\delta\mathbf{x},\delta)^2} \le 1 + C\varepsilon,$$

or equivalently,

$$\overline{\lim_{\delta\downarrow 0}} \frac{E_g\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - \hat{Z}_g(\delta\mathbf{x},\delta)\}^2}{E_g e_g(\delta\mathbf{x},\delta)^2} \le C\varepsilon. \tag{32}$$

Now

$$\frac{E_g\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - \hat{Z}_f(\delta\mathbf{x},\delta)\}^2}{E_g e_g(\delta\mathbf{x},\delta)^2} \le 2\frac{E_g\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - \hat{Z}_g(\delta\mathbf{x},\delta)\}^2}{E_g e_g(\delta\mathbf{x},\delta)^2} + 2\frac{E_g\{\hat{Z}_g(\delta\mathbf{x},\delta) - \hat{Z}_f(\delta\mathbf{x},\delta)\}^2}{E_g e_g(\delta\mathbf{x},\delta)^2}$$

and the second term on the right side tends to 0 as $\delta \downarrow 0$ by Theorem 10 in [Stein (1999), Chapter 3], so by (32), there exists $C < \infty$ independent of ε such that

$$\overline{\lim_{\delta\downarrow 0}} \frac{E_g\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - \hat{Z}_f(\delta\mathbf{x},\delta)\}^2}{E_g e_g(\delta\mathbf{x},\delta)^2} \le C\varepsilon.$$

Using (6) and the fact that $f \asymp g$ on $\mathbb{R}^d$, we can thus choose a $C < \infty$ independent of ε such that

$$\begin{aligned}\overline{\lim_{\delta\downarrow 0}} &\frac{E_f\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - Z(\delta\mathbf{x})\}^2}{E_f e_f(\delta\mathbf{x},\delta)^2} \\ &= 1 + \overline{\lim_{\delta\downarrow 0}} \frac{E_f\{\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta) - \hat{Z}_f(\delta\mathbf{x},\delta)\}^2}{E_f e_f(\delta\mathbf{x},\delta)^2} \le 1 + C\varepsilon.\end{aligned} \tag{33}$$

The set B in the statement of Theorem 1 contains a neighborhood of the origin, so one can choose $r_\varepsilon < \infty$ so that $B_\varepsilon \subseteq rB$ for all $r \ge r_\varepsilon$. We then have $E_f e_f(\delta\mathbf{x}, \delta(rB \cap \mathbb{Z}^d))^2 \le E_f\{Z(\delta\mathbf{x}) - \hat{Z}_\gamma^\varepsilon(\delta\mathbf{x},\delta)\}^2$ for all $r \ge r_\varepsilon$, so there exists a $C < \infty$ independent of ε such that

$$\lim_{\delta\downarrow 0} \frac{E_f e_f(\delta\mathbf{x}, \delta(rB \cap \mathbb{Z}^d))^2}{E_f e_f(\delta\mathbf{x},\delta)^2} \le 1 + C\varepsilon$$

for all $r \ge r_\varepsilon$. Since ε is arbitrary, Theorem 1 follows. □

3. Ordinary kriging. Until now, we have assumed the mean of Z is known to be 0. It is common in the geostatistical literature to assume the mean of a random field is of the form $EZ(\mathbf{x}) = \sum_{j=1}^q \beta_j m_j(\mathbf{x})$ for known functions $m_1, \dots, m_q$ and unknown coefficients $\beta_1, \dots, \beta_q$. Prediction is then done using using what is called universal kriging [Chilès and Delfiner (1999), Cressie (1993)] which reduces to ordinary kriging when the mean is an unknown constant. The universal kriging predictor is just the best linear unbiased predictor; that is, it is the linear predictor that minimizes the prediction variance among all linear predictors whose error has mean 0 regardless of the values of the β_js. Let $\tilde{Z}_f(\mathbf{x}, A)$ be the best linear unbiased predictor of $Z(\mathbf{x})$ based on observing Z on A under the spectral density f and with mean function of the form $\sum_{j=1}^{q} \beta_j m_j(\mathbf{x})$; $\tilde{Z}_f(\mathbf{x}, A)$ exists whenever any linear unbiased predictor exists. If f satisfies the conditions of Theorem 1, $p = \lfloor \frac{1}{2}(\alpha - d) \rfloor$ and $m_1, \dots, m_q$ are all monomials of degree at most p, then $\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x}, \delta)$ is a linear unbiased predictor. Hence, $\tilde{Z}_f(\mathbf{x}, \delta B_\varepsilon)$ exists and it is a better predictor than $\hat{Z}_\gamma^\varepsilon(\delta\mathbf{x}, \delta)$ under f, so under the conditions of Theorem 1, following the argument from (33) to the end of the proof of Theorem 1,

$$\lim_{r\to\infty} \lim_{\delta \downarrow 0} \frac{E_f\{\tilde{Z}_f(\delta\mathbf{x}, \delta(rB \cap \mathbb{Z}^d)) - Z(\delta\mathbf{x})\}^2}{E_f e_f(\delta\mathbf{x}, \delta)^2} = 1.$$

Since $p \geq 0$ whenever $\alpha > d$, we get that there is an asymptotic screening effect for ordinary kriging predictors under the conditions of Theorem 1.

I believe a screening effect holds for best linear unbiased predictors whenever the m_js are sufficiently smooth. Such a result should follow from the asymptotic optimality of best linear unbiased predictors relative to best linear predictors [see Theorem 5.2 in Stein (1999)].

4. Numerical results. Theorem 1 provides no indication as to how fast the limit in (2) is approached as $r \to \infty$. Furthermore, the method of proof used here does not appear to be amenable to obtaining such results. There is a limitless array of possibilities one could examine to see how strong the screening effect is in particular situations and here we choose to examine only a small number of one-dimensional settings. Specifically, we consider stationary processes Z on $\mathbb{R}$ with covariance function in the Matérn class [Stein (1999), page 31]: $\operatorname{cov}(Z(x), Z(y + x)) \propto (\theta|y|)^\nu \mathcal{K}_\nu(\theta|y|)$ for positive constants θ and ν, where $\mathcal{K}_\nu$ is a modified Bessel function of order ν [Abramowitz and Stegun (1992)]. For a process on $\mathbb{R}$, the corresponding spectral density is proportional to $(\theta^2 + v^2)^{-\nu-1/2}$, which is a regularly varying function with exponent $-2\nu - 1$. In the present setting, $\alpha = 2\nu + 1$ is more pertinent than ν, so we will report all results in terms of α. Furthermore, we will set $\theta = 4\nu$, which has the effect of approximately fixing the rate of decay of the covariance function at larger distances as ν varies [Stein (1999), page 49]. Defining $A_n = \{-n + 1, -n + 2, \dots, n\}$, we will compare predictions of $Z(0.5\delta)$ based on observing Z at δA_n to predictions based on observing Z at $\delta\mathbb{Z}$ for various n and δ.

We first consider how the mean squared error of the simple kriging predictor of $Z(0.5\delta)$ changes with n and δ for various values of α. Define $R(n, \delta, \alpha) = Ee(0.5\delta, \delta A_n)^2 / Ee(0.5\delta, \delta\mathbb{Z})^2 - 1$. Numerical results suggest that replacing $Ee(0.5\delta, \delta\mathbb{Z})^2$ by $Ee(0.5\delta, \delta A_{200})^2$ in $R(n, \delta, \alpha)$ provides an excellent approximation to $R(n, \delta, \alpha)$ for the values of n, δ and α considered here. This approximation is used in Figure 1, which plots $R(n, \delta, \alpha)$ for $n = 1, \dots, 12$, $\delta = 0.1$, 0.05,

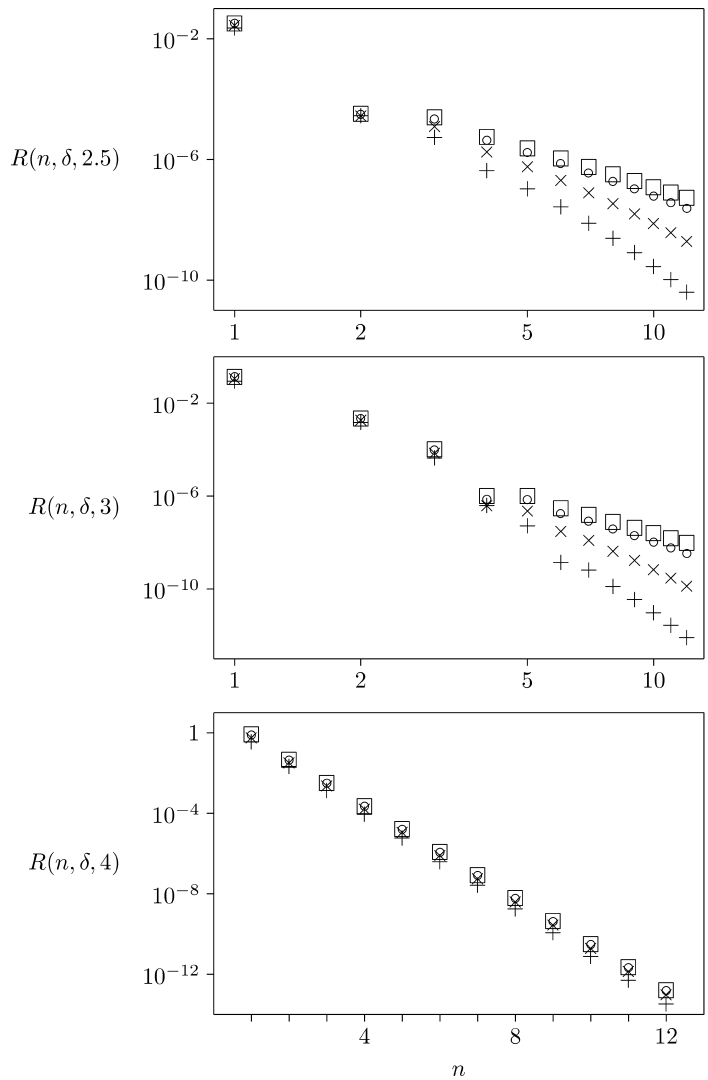

FIG. 1. *Plot of $R(n, \delta, \alpha)$ versus n. For $\alpha = 2.5$ and 3, both axes are on the log scale. For $\alpha = 4$, only the vertical axis is on the log scale. $R(n, \delta, \alpha)$ is the relative increase in mean squared error when using simple kriging to predict at 0.5δ due to using observations at $\delta\{-n+1, -n+2, \dots, n\}$ versus at $\delta\mathbb{Z}$. The covariance functions for the process are all from the Matérn class; the corresponding spectral densities decay like $|v|^{-\alpha}$ at high frequencies v (see text for details). The symbols $+$, $\times$, $\circ$ and $\square$ correspond to $\delta = 0.1, 0.05, 0.02$ and 0.01, respectively.*

0.02 and 0.01 and $\alpha = 2.5$, 3 and 4. We see that even for n quite small, $R(n, \delta, \alpha)$ is very close to 0. Indeed, $R(3, \delta, \alpha) < 0.005$ for all α and δ values considered here.

For most practical purposes, being within 0.5% of optimal would be sufficient. Nevertheless, there are at least theoretically interesting differences between how

$R(n, \delta, \alpha)$ decreases as n increases for the various α values. For $\alpha = 2.5$ or 3, the screening effect substantially weakens as δ decreases. Note that the theoretical results in Section 2 do not say anything about how $R(n, \delta, \alpha)$ should behave for fixed n as δ decreases. For $\alpha = 2.5$ or 3, the plot of $R(n, \delta, \alpha)$ versus n with both axes on the log scale is fairly linear for n sufficiently large, particularly for smaller δ. Based on a purely empirical analysis of the cases considered in Figure 1 and other values for α and δ, I would make the following conjecture: $R(n, \alpha) = \lim_{\delta \downarrow 0} R(n, \delta, \alpha)$ exists for all $\alpha > 1$ and $n > \frac{1}{2}(\alpha - 1)$ and $R(n, \alpha) \asymp n^{-\alpha-1}$ as $n \to \infty$ for any fixed α not an even integer.

For $\alpha = 4$, $R(n, \delta, \alpha)$ behaves qualitatively differently than for $\alpha = 2.5$ or 3. First, $R(n, \delta, 4)$ depends only weakly on δ. Furthermore, the plot of $\log R(n, \delta, 4)$ versus n is very nearly linear for all δ, suggesting that $R(n, \delta, 4)$ decays exponentially as n increases. When $\alpha = 2$, Z is a continuous time AR(1) process and $R(n, \delta, 2) = 0$ for all positive n and δ and there is "perfect" screening. These examples and others not reported on here suggest that there is something dramatically different about the screening effect when α is or is not an even integer, at least for Matérn models. Under the Matérn model, when $\alpha = 2m$ for a positive integer m, Z is a continuous time AR(m) process and hence has a second-order Markov property: conditional on knowing $Z, Z', \ldots Z^{(m-1)}$ at the present time, the past and the future of Z are uncorrelated. Using the close connection between kriging and splines [Wahba (1990)] and results from spline theory [see, e.g., Schoenberg (1969)], I believe it should be possible to prove that, when α is an even integer, for any given δ, $R(n, \delta, \alpha)$ decays to 0 exponentially fast in n and, furthermore, $R(n, \alpha)$ exists for $n > \frac{1}{2}(\alpha - 1)$ and decays exponentially in n.

Let us consider further why $R(n, \alpha)$ might decay only algebraically when α is not an even integer. The idea behind Theorems 1 and 2 is that as $\delta \downarrow 0$, the prediction problem under a spectral density f satisfying the conditions of Theorem 1 becomes asymptotically indistinguishable from predicting under the IRF with spectral density proportional to $|v|^{-\alpha}$. The singularity at the origin in $|v|^{-\alpha}$ for all $\alpha > 1$ implies that an IRF with such a spectral density has dependence over large scales. However, when α is an even integer, this large-scale dependence can be removed by taking a simple linear transformation of the process. Specifically, defining Δ to be the forward difference operator, so that $\Delta Z(x) = Z(x + 1) - Z(x)$, then if $\alpha = 2m$ for a positive integer m, the process $\Delta^m Z$ is stationary with a spectral density that is bounded away from 0 and ∞ in a neighborhood of the origin. When α is not an even integer, there is no linear transformation of Z of the form $\sum_{j=1}^{N} \lambda_j Z(\cdot - x_j)$ with N finite that has spectral density bounded away from 0 and ∞ in a neighborhood of the origin. Thus, one can argue that the large-scale dependence of Z is of a different and more intrinsic nature when α is not an even integer, so that a greater dependence of the optimal linear predictor on distant observations is natural.

Let us next consider ordinary kriging. Using $\bar{e}$ to indicate the error of an ordinary kriging predictor, define $\bar{R}(n, \delta, \alpha) = E\bar{e}(0.5\delta, \delta A_n)^2 / E\bar{e}(0.5\delta, \delta \mathbb{Z})^2 - 1$.

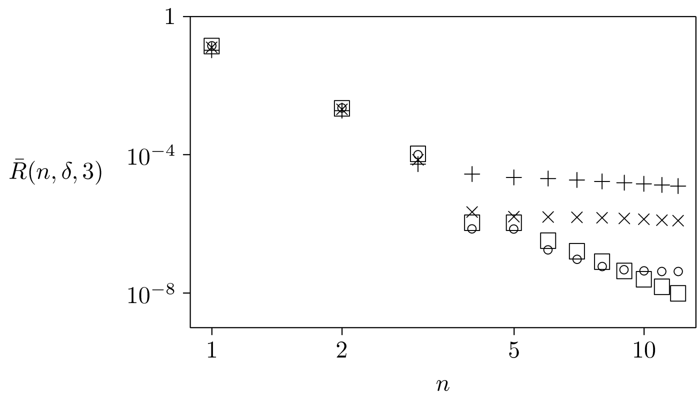

FIG. 2. *Plot of* $\bar{R}(n, \delta, 3)$ *versus* n; *both axes are on the log scale.* $\bar{R}(n, \delta, \alpha)$ *is the relative increase in mean squared error when using ordinary kriging to predict at* 0.5δ *due to using observations at* $\delta\{-n+1, -n+2, \ldots, n\}$ *versus at* $\delta\mathbb{Z}$. *The covariance function for the process is the same as in the middle panel of Figure* 1. *The symbols* $+$, $\times$, $\circ$ *and* $\square$ *correspond to* $\delta = 0.1, 0.05, 0.02$ *and* 0.01, *respectively.*

For larger δ, $n = 200$ may not be sufficiently large for $E\bar{e}(0.5\delta, \delta A_n)^2$ to provide an adequate approximation to $E\bar{e}(0.5\delta, \delta\mathbb{Z})^2$ for our purposes. Since the mean of Z can be estimated exactly with probability 1 from observations on $\delta\mathbb{Z}$ whenever Z has a spectral density, we have $E\bar{e}(0.5\delta, \delta\mathbb{Z})^2 = Ee(0.5\delta, \delta\mathbb{Z})^2$. Furthermore, $Ee(0.5\delta, \delta A_{n'})^2$ monotonically decreases to $E\bar{e}(0.5\delta, \delta\mathbb{Z})^2$ as $n' \to \infty$ and numerical evidence indicates that $Ee(0.5\delta, \delta A_{200})^2$ provides an excellent approximation to $E\bar{e}(0.5\delta, \delta\mathbb{Z})^2$ for all situations considered here. This approximation is used in Figure 2, which plots $\bar{R}(n, \delta, 3)$ for the same δ values as in Figure 1. These results should be compared to the middle panel in Figure 1, which plots $R(n, \delta, 3)$. For larger n, the screening effect now gets stronger as δ increases, which is the opposite of what occurred for simple kriging. For larger δ and n, $\bar{R}(n, \delta, 3)/R(n, \delta, 3)$ can be very large; for example, $\bar{R}(12, 0.1, 3)/R(12, 0.1, 3) = 1.55 \times 10^7$. For smaller δ, $\bar{R}(n, \delta, 3)$ is much closer to $R(n, \delta, 3)$. Indeed, $\bar{R}(n, 0.01, 3)/R(n, 0.01, 3)$ is between 1 and 1.1 for $n = 1, 2, \ldots, 12$. Apparently, the effect of having to estimate the mean on prediction is greater for larger δ, which is consistent with theoretical results in Stein (1999), page 107.

Finally, we briefly examine the impact of measurement errors on the screening effect. Let us suppose that the measurement errors are independent with mean 0 and common variance σ^2. One might expect that the presence of measurement errors weakens the screening effect, since the observations nearest to the prediction location are no longer as strongly correlated with the predictand. This belief is part of the conventional geostatistical wisdom: an often quoted aphorism in the geostatistical literature due to Matheron (1968) is “the nugget effect lifts the screening effect” [Chilès and Delfiner (1999), page 204] (the nugget effect is a geostatistical term for process variation on scales much smaller than the distances

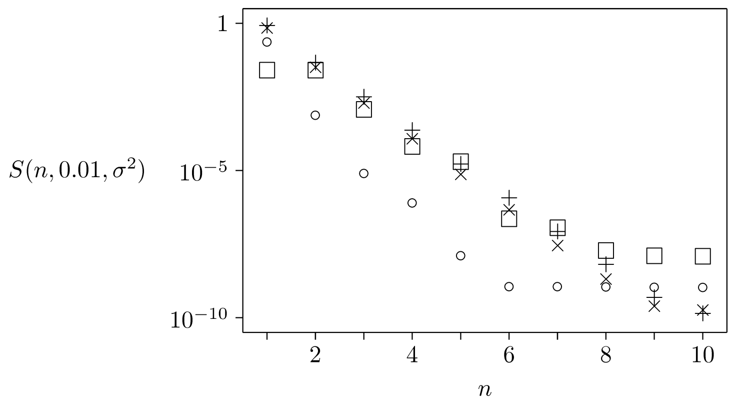

FIG. 3. *Plot of* $S(n, 0.01, \sigma^2)$ *versus* n; *vertical axis on the log scale.* $S(n, 0.01, \sigma^2)$ *is the relative increase in mean squared error when using ordinary kriging to predict at* 0.005 *due to using observations at* $0.01\{-n+1, -n+2, \ldots, n\}$ *versus at* $0.01\mathbb{Z}$ *for a stationary process on* $\mathbb{R}$ *with covariance function* $e^{-6|y|}(1+6|y|)$ *observed with measurement errors of variance* σ^2. *The symbols* $+$, $\times$, $\circ$ *and* $\square$ *correspond to* $\sigma^2 = 0, 10^{-6}, 10^{-5}$ *and* 10^{-4}, *respectively.*

between neighboring observations and, for purposes of the discussion here, is indistinguishable from measurement error). When using ordinary kriging, one should expect the screening effect to be reduced for sufficiently large σ^2, since the ordinary kriging predictor can be easily shown to converge to the sample mean of the observations as $\sigma^2 \to \infty$. However, I know of no argument that shows it should, by some measure, monotonically decrease as σ^2 increases. Indeed, the following example shows this is not always the case, at least if one measures the strength of the screening effect by the relative difference between the mean squared error of the ordinary kriging predictor based on some set of observations near to the predictand and the mean squared error of the ordinary kriging predictor based on all of the available observations. Suppose Z has covariance function $e^{-6|y|}(1 + 6|y|)$, which is just the case $\alpha = 4$ used in Figure 1. Observe Z with independent measurement errors of variance σ^2 and define $S(n, \delta, \sigma^2) = E\bar{e}(0.5\delta, \delta A_n)^2 / E\bar{e}(0.5\delta, \delta\mathbb{Z})^2 - 1$. Figure 3 plots $S(n, 0.01, \sigma^2)$ for $n = 1, \ldots, 10$ and $\sigma^2 = 0, 10^{-6}, 10^{-5}$ and 10^{-4} and shows that for $n \le 8$, the screening effect is actually stronger ($S(n, 0.01, \sigma^2)$ is smaller) for $\sigma^2 = 10^{-6}$ and 10^{-5} than when it is 0, substantially so when $\sigma^2 = 10^{-5}$. Although these measurement error variances may seem small, they are large enough to have a substantial impact on $E\bar{e}(0.005, 0.01\mathbb{Z})^2$, which has values 9.44×10^{-6}, 1.12×10^{-5}, 2.65×10^{-5} and 1.63×10^{-4} for $\sigma^2 = 0, 10^{-6}, 10^{-5}$ and 10^{-4}, respectively.

5. Discussion. This section briefly discusses the conditions on the spectral density and the nature of the limiting operation in Theorems 1 and 2. In particular, the sense in which these theorems require that the random field not be too different from a self-affine random field is addressed. We also discuss two possible extensions of the results in this work: uniformity in the predictand location $\mathbf{x}$ in Theorem 1 and a screening effect when the observations are not restricted to a cubic lattice.

Let us first examine the slowly varying aspect of assumption (A2). If we assume that L in (A2) is a positive constant, then it is possible to give a much shorter proof of Theorem 1. Furthermore, among models for spectral densities that are used in practice, any spectral density f satisfying (A1) and (A2) would also satisfy (A1) and (A2) with L constant. For example, all spectral densities in the Matérn class satisfy (A1) and (A2) with L constant. Thus, it is worthwhile to consider what is gained by allowing L in (A2) to be slowly varying.

First, the fact that spectral densities with nontrivial slowly varying components are not used in practice is not, by itself, a reason for dismissing such models. Indeed, I am unaware of any previous theoretical or empirical basis for excluding, for example, a spectral density of the form $f(\mathbf{v}) \sim |\mathbf{v}|^{-\alpha} \log |\mathbf{v}|$ as $|\mathbf{v}| \to \infty$. Theorem 2 and Corollary 2 prove that predictions under the simpler model γ are asymptotically optimal when the more complex models g or f are correct, and thus provide a theoretical argument against including nontrivial slowly varying components in models for spectral densities when one is only interested in prediction.

A second advantage of including the slowly varying component in (A2) is the insight it provides about the prediction process. The proofs of Theorems 1 and 2 effectively work by showing that for a stationary random field on a grid with spacing δ, the frequencies of the spectral density f that matter for prediction are those that are of the order of magnitude δ^{-1}. On this range of frequencies, if f satisfies (A2), then, roughly speaking, $f(\mathbf{v}) \approx L(|\mathbf{v}|)\gamma(\mathbf{v}) \approx L(\delta^{-1})\gamma(\mathbf{v})$ for δ small, since $L(|\mathbf{v}|)$ hardly varies for $|\mathbf{v}| \asymp \delta^{-1}$. This is the essential reason optimal predictions under γ are nearly the same as optimal predictions under f. Furthermore, since the screening effect holds trivially for γ, it is also the essential reason the screening effect holds under f.

Next we consider the sense in which (A2) implies that Z is nearly self-affine. Assumption (A2) says that Z has spectral density whose high frequency behavior is not too different from a function of the form $\gamma(\mathbf{v}) = \theta(\mathbf{v})|\mathbf{v}|^{-\alpha}$. As already noted, γ can be thought of as the spectral density of an IRF-p with $p = \lfloor \frac{1}{2}(\alpha - d) \rfloor$. An IRF-$p$ Y with γ as its spectral density has the following self-affinity property: for $c > 0$, define the random field Y_c by $Y_c(\mathbf{x}) = c^{-(\alpha-d)/2} Y(c\mathbf{x})$; then for any ALC-$p$ $\sum_{j=1}^{n} \lambda_j Y(\mathbf{x}_j)$,

$$E\left\{\sum_{j=1}^{n} \lambda_j Y_c(\mathbf{x}_j)\right\}^2 = E\left\{\sum_{j=1}^{n} \lambda_j Y(\mathbf{x}_j)\right\}^2.$$

A random field whose properties are invariant after rescaling in both the the $\mathbf{x}$ and Y directions is often called self-similar [Mandelbrot and Van Ness (1968), Kent and Wood (1997), Chan and Wood (2000)]. However, since at least the early

1980's, Mandelbrot has called such a random field self-affine and reserved the term self-similar for processes whose law is invariant after rescaling just $\mathbf{x}$ [Mandelbrot (1982)] and we follow this usage of terminology here.

A Gaussian random field on $\mathbb{R}^d$ with spectral density proportional to $|\mathbf{v}|^{-\alpha}$ and $\alpha \in (d, d+2)$ is often called a fractional Brownian motion [Molz, Liu and Szulga (1997)], although perhaps fractional Brownian field would be a better name when $d > 1$. In this case, the corresponding random field Z satisfies $E\{Z(\mathbf{x}) - Z(\mathbf{y})\}^2 \propto |\mathbf{x} - \mathbf{y}|^{\alpha - d}$. Fractional Brownian motions and fields have found wide application in image modeling [Saupe (1988)], hydrology [Molz, Liu and Szulga (1997)] and throughout geophysics [Malamud and Turcotte (1999)]. Theorem 2 provides some theoretical backing for using this model for prediction purposes whenever one is willing to assume that Z is an isotropic random field whose actual spectral density f satisfies $f(|\mathbf{v}|)|\mathbf{v}|^\alpha$ is slowly varying at infinity.

Let us now consider two possible extensions of the results in this work. The first would be to obtain some sort of uniformity in $\mathbf{x}$ in Theorems 1 and 2. The results in Section 2 all consider $\mathbf{x}$ to be a fixed nonvertex of $[0, 1]^d$. In contrast, results in Stein (1999) on asymptotically optimal predictions (e.g., Theorem 10 in Chapter 3 and Theorem 12 in Chapter 4) show that predictions based on a spectral density that is asymptotically correct at high frequencies are uniformly asymptotically optimal over the region of observation. Thus, one might expect some sort of uniform asymptotic optimality holds in Theorems 1 and 2 here and their corollaries. For example, I conjecture that under the conditions in Theorem 1 on f, if for some $\varepsilon > 0$ and all $\mathbf{x}$ in a set A, the ball of radius ε centered at $\mathbf{x}$ is contained in B, then

$$\lim_{r\to\infty} \lim_{\delta\downarrow 0} \sup_{\mathbf{x}\in\delta A} \frac{E_f e_f\big(\mathbf{x}, \delta(rB \cap \mathbb{Z}^d)\big)^2}{E_f e_f(\mathbf{x}, \delta)^2} = 1, \tag{34}$$

as long as we define $0/0 = 1$. The obstacle in proving this result is that Lemma 1, which gives the order of magnitude of the mean square prediction error, does not necessarily hold uniformly on δA.

If (34) were true, then under the conditions of Theorem 1 on f, we could immediately obtain a limit theorem analogous to (1) when predicting at a fixed $\mathbf{x}$, rather than letting the place at which we predict change with δ. Specifically, as long as we define $0/0 = 1$, (34) implies for any fixed $\mathbf{x}$ in the interior of B,

$$\lim_{\delta\downarrow 0} \frac{E_f e_f\big(\mathbf{x}, B \cap \delta\mathbb{Z}^d\big)^2}{E_f e_f(\mathbf{x}, \delta)^2} = 1.$$

A more ambitious extension of Theorem 1 would be to observations other than on a cubic lattice. First, if (A1) and (A2) hold for some spectral density f, they still hold after a nonsingular linear transformation of the coordinates, so all of the results in Section 2 hold for observations on any lattice. Extensions to observations not on a lattice are more difficult but perhaps not out of reach.

In particular, if X is a measurable subset of $\mathbb{R}^d$ and B is some set containing a neighborhood of the origin, then $\gamma(c|\mathbf{v}|) \propto \gamma(|\mathbf{v}|)$ as a function of $\mathbf{v}$ implies $E_\gamma e_\gamma\big(\delta\mathbf{x}, \delta(rB \cap X)\big)^2 / E_\gamma e_\gamma(\delta\mathbf{x}, \delta X)^2$ is independent of δ, so that as long as we define $0/0 = 1$, we trivially have

$$\lim_{r\to\infty} \lim_{\delta\downarrow 0} \frac{E_\gamma e_\gamma\big(\delta\mathbf{x}, \delta(rB \cap X)\big)^2}{E_\gamma e_\gamma(\delta\mathbf{x}, \delta X)^2} = 1.$$

That is, the screening effect holds for γ with essentially any arrangement of observations. To prove that a screening effect holds for other spectral densities, one would need some analog to Theorem 2 or Corollary 2. If $f(\mathbf{v}) \sim \gamma(\mathbf{v})$ as $|\mathbf{v}| \to \infty$, such a result might be obtainable by an extension of Theorem 12 in Chapter 4 of Stein (1999).

Acknowledgments. I would like to thank the former Editor, Hans Künsch, and the referees for their careful reading of this work and their many helpful suggestions.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. (1992). *Handbook of Mathematical Functions*. Dover, New York.

ARMSTRONG, M. (1998). *Basic Linear Geostatistics.* Springer, New York.

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation.* Cambridge Univ. Press.

CHAN, G. and WOOD, A. T. A. (2000). Increment-based estimators of fractal dimension for two-dimensional surface data. *Statist. Sinica* **10** 343–376.

CHILÈS, J. and DELFINER, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* Wiley, New York.

CRESSIE, N. (1993). *Statistics for Spatial Data*, rev. ed. Wiley, New York.

JERRI, A. J. (1977). The Shannon sampling theorem—its various extensions and applications: A tutorial review. *Proc. IEEE* **65** 1565–1596.

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

KENT, J. T. and WOOD, A. T. A. (1997). Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. Roy. Statist. Soc. Ser. B* **59** 679–699.

MALAMUD, B. D. and TURCOTTE, D. L. (1999). Self-affine time series: I. Generation and analysis. In *Advances in Geophysics* (R. Dmowska and B. Saltzman, eds.) **40** 1–90. Academic Press, New York.

MANDELBROT, B. B. (1982). *The Fractal Geometry of Nature*. Freeman, San Francisco.

MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

MATHERON, G. (1968). *Osnovy Prikladnoï Geostatistiki (Treatise of Geostatistics)*. Mir, Moscow.

MOLZ, F. J., LIU, H. H. and SZULGA, J. (1997). Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions. *Water Resources Research* **33** 2273–2286.

SAUPE, D. (1988). Algorithms for random fractals. In *The Science of Fractal Images* (H.-O. Peitgen and D. Saupe, eds.) 71–136. Springer, New York.

SCHOENBERG, I. J. (1969). Cardinal interpolation and spline functions. *J. Approx. Theory* **2** 167–206.

STEIN, M. L. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Statist.* **18** 850–872.

STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York.

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

WACKERNAGEL, H. (1995). *Multivariate Geostatistics.* Springer, Berlin.

WAHBA, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637