Biometrika (2002), **89,** 1, *pp.* 197–210

Printed in Great Britain

Spectral methods for nonstationary spatial processes

BY MONTSERRAT FUENTES

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.

fuentes@stat.ncsu.edu

SUMMARY

We propose a nonstationary periodogram and various parametric approaches for estimating the spectral density of a nonstationary spatial process. We also study the asymptotic properties of the proposed estimators via shrinking asymptotics, assuming the distance between neighbouring observations tends to zero as the size of the observation region grows without bound. With this type of asymptotic model we can uniquely determine the spectral density, avoiding the aliasing problem. We also present a new class of nonstationary processes, based on a convolution of local stationary processes. This model has the advantage that the model is simultaneously defined everywhere, unlike 'moving window' approaches, but it retains the attractive property that, locally in small regions, it behaves like a stationary spatial process. Applications include the spatial analysis and modelling of air pollution data provided by the US Environmental Protection Agency.

Some key words: Anisotropy; Covariance; Edge effect; Fourier transform; Geostatistics; Kernel smoothing; Periodogram; Spatial statistics; Variogram.

1. INTRODUCTION

The periodogram, a nonparametric estimate of the spectral density, is a powerful tool for studying the properties of stationary processes observed on a d-dimensional lattice. Use and properties of spatial periodograms for stationary processes have been investigated by Stein (1995), Guyon (1982, 1992), Ripley (1981), Rosenblatt (1985) and Whittle (1954) among others. Pawitan & O'Sullivan (1994) proposed a nonparametric spectral density estimator using a penalised Whittle likelihood for a stationary time series. Guyon (1982) studied the asymptotic properties of various parameter estimation procedures for a general stationary process on a d-dimensional lattice, using spectral methods. The spectral representation of a stationary process Z is interpreted as its representation in the form of superposition of sine and cosine waves of different frequencies.

However, spatial processes in environmental sciences, oceanography, soil science and many other disciplines are generally nonstationary, in the sense that the spatial structure depends on location. In this paper we introduce spectral methods for studying the spatial structure of a nonstationary process, using a Fourier–Stieltjes representation of the process. New fitting algorithms are developed. In cases where the data are observed on a d-dimensional lattice, it is argued that these spectral approaches have computational benefits compared with maximum likelihood on the space domain. We also present a new class of nonstationary processes.

In § 2, we introduce a spectral representation for a nonstationary spatial process, and

we define an asymptotic model that we call shrinking asymptotics, which assumes that the distance between neighbouring observations tends to zero and the size of the observation region grows without bound. In § 3, we propose a nonstationary periodogram for estimating the spectral density of a nonstationary spatial process, and we study the properties of the proposed estimator. In § 4, we present a new class of nonstationary processes, based on a convolution of local stationary processes, and we propose nonparametric and parametric estimators of the spectral density using this new nonstationary model. Section 5 contains an application of the methodology presented in this paper to atmospheric air pollution data. The data come from the regional scale air quality models, known as Models-3, that are run by the US Environmental Protection Agency. Models-3 provide pollutants' concentrations and fluxes in regular grids in parts of the United States; see Fig. 1 below. The lattice structure of Models-3 is perfectly suited to our spectral methods.

2. REPRESENTATION OF A NONSTATIONARY PROCESS

2·1. *Spectral function of a nonstationary process*

A random field Z in $\mathbb{R}^2$ is called weakly stationary if it has finite second moments, its mean function is constant and it possesses an autocovariance function C such that $C(x-y) = \mathrm{cov}\{Z(x), Z(y)\}$. If Z is a weakly stationary random field with autocovariance function C, then we can represent the process in the form of the following Fourier–Stieltjes integral:

$$Z(x) = \int_{\mathbb{R}^2} \exp(ix^{\mathrm{T}}\omega)\, dY(\omega), \tag{1}$$

where Y are random functions with uncorrelated increments; see Jaglom (1987, p. 100), for example. Now we abandon the requirement that the function $Y(\omega)$ in the integrand of (1) necessarily have uncorrelated increments, thereby obtaining the Fourier–Stieltjes representation of a nonstationary random process $Z(x)$ (Jaglom, 1987, p. 460). It is easy to see that then the covariance function C of the nonstationary process $Z(x)$ is given by

$$\mathrm{cov}\{Z(x_1), Z(x_2)\} = C(x_1, x_2) = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp\{i(x_1^{\mathrm{T}}\omega_1 - x_2^{\mathrm{T}}\omega_2)\}\, d^2F(\omega_1, \omega_2), \tag{2}$$

where $F(\omega_1, \omega_2)$ is the spectral function

$$E\{Y(\omega_1)Y^c(\omega_2)\} = F(\omega_1, \omega_2), \tag{3}$$

and Y^c denotes the conjugate of Y. It is clear that, if $F(\omega_1, \omega_2)$ is a function of bounded variation in the plane, i.e.

$$\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} |d^2F(\omega_1, \omega_2)| = V_F < \infty,$$

then the integral (2) is necessarily convergent, and Z is called a harmonisable random process.

If F has a density with respect to Lebesgue measure, this density is the spectral density f which is again the Fourier transform of the autocovariance function:

$$f(\omega_1, \omega_2) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \exp\{-i(x_1^{\mathrm{T}}\omega_1 - x_2^{\mathrm{T}}\omega_2)\}C(x_1, x_2)\, dx_1\, dx_2. \tag{4}$$

2·2. *Aliasing*

The spectral functions f and F are closely related to the spectral decomposition of the process Z into a superposition of harmonic oscillations. However, it is easy to see that such a decomposition cannot be uniquely restored from observations of a continuous process Z on a lattice. If Z is observed only at uniformly spaced spatial locations Δ units apart, the spectrum of observations of the sample sequence $Z(\Delta x)$, for $x \in \mathbb{Z}^2$, is concentrated within the finite frequency band $-\pi/\Delta \leqslant \omega < \pi/\Delta$. The whole frequency spectrum is partitioned into bands of length $2\pi/\Delta$ by so-called fold points $(2y+1)\pi/\Delta$, with $y \in \mathbb{Z}^2$, and the power distribution within each of the bands distinct from the principal band, $-\pi/\Delta \leqslant \omega < \pi/\Delta$, is superimposed on the power distribution within the principal band. Thus, if we wish the spectral characteristics of the process Z to be determined accurately enough from the observed sample, then the Nyquist frequency π/Δ must necessarily be so high that still higher frequencies ω make only a negligible contribution to the total power of the process.

The spectral density f_Δ of the process on the lattice can be written in terms of the spectral density f of the continuous process Z as

$$f_\Delta(\omega_1, \omega_2) = \sum_{Q_1 \in \mathbb{Z}^2} \sum_{Q_2 \in \mathbb{Z}^2} f\left(\omega_1 + \frac{2\pi Q_1}{\Delta}, \omega_2 + \frac{2\pi Q_2}{\Delta}\right),$$

for $\omega_1, \omega_2 \in \prod_\Delta^2 = [-\pi/\Delta, \pi/\Delta]^2$. If Z is stationary with spectral density f, then

$$f_\Delta(\omega) = \sum_{Q \in \mathbb{Z}^2} f\left(\omega + \frac{2\pi Q}{\Delta}\right). \tag{5}$$

3. Nonparametric spectral density estimation

3·1. *Stationary periodograms*

Consider a stationary spatial process Z. We observe the process at N equally spaced locations in a regular grid D $(n_1 \times n_2)$, where $N = n_1 n_2$ and the spacing between observations is Δ. The periodogram is a nonparametric estimator of the spectral density. We define $I_N(\omega_0)$ to be the periodogram at a frequency ω_0:

$$I_N(\omega_0) = \Delta^2 (2\pi)^{-2} (n_1 n_2)^{-1} \left| \sum_{s_1=1}^{n_1} \sum_{s_2=1}^{n_2} Z(\Delta s) \exp(-i \Delta s^{\mathrm{T}} \omega_0) \right|^2, \tag{6}$$

where $s^{\mathrm{T}} = (s_1, s_2)$. In practice, the periodogram estimator for ω is computed at the set of Fourier frequencies $2\pi f/(\Delta n)$ where $f/(\Delta n) = (f_1/\Delta n_1\ , f_2/\Delta n_2)$, and $f \in J_N$, for

$$J_N = \{\lfloor -(n_1 - 1)/2 \rfloor, \ldots, n_1 - \lfloor n_1/2 \rfloor\} \times \{\lfloor -(n_2 - 1)/2 \rfloor, \ldots, n_2 - \lfloor n_2/2 \rfloor\}, \tag{7}$$

in which $\lfloor u \rfloor$ denotes the largest integer less than or equal to u.

In some instances, consistency demands that the length of the grid $(n_1 \times n_2)$ over which the process is observed increase as the number of observations increases. This ensures that the amount of information in the data increases. We should also ask that the spacing Δ between neighbouring observations goes to 0 as the number of observations increases. This guarantees that an accurate picture of the covariance function and spectral density can be developed nonparametrically, assuming only smoothness conditions. Therefore, our asymptotic model supposes that the observed data represent a realisation of $Z(x\Delta)$

for $x^{\mathrm{T}} = (x_1, x_2)$, where $0 \leqslant x_1 \leqslant n_1$ and $0 \leqslant x_2 \leqslant n_2$. We assume that Δ goes to zero as $n_1 \to \infty$, $n_2 \to \infty$ and $n_1/n_2 \to \lambda$, for a constant $\lambda > 0$, and that $\Delta n_1 \to \infty$, $\Delta n_2 \to \infty$. We call this type of asymptotics shrinking asymptotics. This is a mixture of increasing-domain (Cressie, 1993, p. 480) and fixed-domain asymptotics (Stein, 1995) in that the distance between neighbouring observations tends to zero and the size of the observation region grows without bound. Constantine & Hall (1994) used this asymptotic model for one-dimensional processes, and Lahiri et al. (1999) used it in the context of subsampling.

In Theorem 1 we study the asymptotic behaviour of the periodogram in the case where Z is Gaussian.

For each $x^{\mathrm{T}} = (x_1, x_2) \in \mathbb{R}^2$, let $\| x \|$ denote the Euclidean norm $(x_1^2 + x_2^2)^{\frac{1}{2}}$.

THEOREM 1. *We assume the following conditions.*

(1a) *The rate of decay of the spectral density $f(\omega)$ at high frequencies is proportional to $\| \omega \|^{-\tau}$, for $\tau > 2$.*

(1b) *The covariance function satisfies the inequality $\int \| u \| \, |C(u)| \, du < \infty$.*

(1c) *We require that $\Delta \to 0$, $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \lambda$ for a constant $\lambda > 0$, $\Delta n_1 \to \infty$ and $\Delta n_2 \to \infty$.*

Then

(i) *the expected value of the periodogram, $I_N(\omega)$, for $\omega \in [-\pi/\Delta, \pi/\Delta]^2$, is asymptotically $f(\omega)$;*

(ii) *the asymptotic variance of $I_N(\omega)$ is $f^2(\omega)$; and*

(iii) *the periodogram values $I_N(\omega)$ and $I_N(\omega')$, for $\omega \neq \omega'$, are asymptotically uncorrelated.*

Condition (1a) is always satisfied by the spectral density (22) corresponding to a Matérn covariance function (12). Condition (1b) implies that the spectral density f has a bounded derivative, and Condition (1c) corresponds to the shrinking asymptotic model.

By part (i) the periodogram I_N is asymptotically an unbiased estimator of f, the spectral density of the continuous process Z. If we use a different asymptotic model, increasing-domain asymptotic for instance, then I_N is not asymptotically an unbiased estimator for f but for f_Δ, the spectral density of the sampled sequence $Z(\Delta x)$ (Brillinger, 1975, pp. 122–3). By part (ii) the variance of the periodogram at ω is asymptotically $f^2(\omega)$. Thus, since its variance is independent of N it will not be a very useful estimator in practice. The traditional way of dealing with this lack of consistency of the periodogram is to smooth the periodogram across frequencies. In an unpublished North Carolina State University technical report, H. J. Kim and M. Fuentes present a new method for choosing an appropriate smoothing parameter for the data taper by taking into account the trade-off between bias and variance of the tapered periodogram. By part (iii) the periodogram values are approximately uncorrelated. This property allows us to fit easily a parametric model to the periodogram values by using least squares in the spectral domain. However, in the space domain, the correlation among the empirical covariance or variogram values thwarts the use of ordinary least squares.

3·2. *Nonstationary periodograms*

We consider a spatial nonstationary process Z with covariance function $C(x, y)$. As in § 2·1, we observe the process at $N = n_1 n_2$ equally spaced locations in a regular $n_1 \times n_2$ grid D; the spacing between observations is Δ. We define a nonstationary periodogram, $I_N(\omega_1, \omega_2)$, that is a nonparametric estimator of the spectral density, $f(\omega_1, \omega_2)$, which is

the Fourier transform of the nonstationary covariance function:

$$I_N(\omega_1, \omega_2) = \frac{\Delta^2}{(2\pi)^2(n_1 n_2)} \sum_{x_1=1}^{n_1} \sum_{x_2=1}^{n_2} Z(\Delta x) \exp(-i\Delta x^{\mathrm{T}} \omega_1) \sum_{y_1=1}^{n_1} \sum_{y_2=1}^{n_2} Z(\Delta y) \exp(i\Delta y^{\mathrm{T}} \omega_2), \tag{8}$$

with $x^{\mathrm{T}} = (x_1, x_2)$ and $y^{\mathrm{T}} = (y_1, y_2)$.

We define $J(\omega)$, an approximation to $Y(\omega)$, the Fourier transform of Z, by

$$J(\omega) = \Delta(2\pi)^{-1}(n_1 n_2)^{-\frac{1}{2}} \sum_{x_1=1}^{n_1} \sum_{x_2=1}^{n_2} Z(\Delta x) \exp(-i\Delta x^{\mathrm{T}} \omega).$$

Thus, we obtain

$$I_N(\omega_1, \omega_2) = J(\omega_1) J^c(\omega_2), \tag{9}$$

and this expression for I_N is consistent with the definition of the spectral distribution F in (3), as a function of the spectral processes Y and Y^c.

We now study the asymptotic behaviour of the nonstationary periodogram $I_N(\omega_1, \omega_2)$ using shrinking asymptotics, assuming Z to be Gaussian. In Theorem 2, for each $x^{\mathrm{T}} = (x_1, x_2) \in \mathbb{R}^2$ and $y^{\mathrm{T}} = (y_1, y_2) \in \mathbb{R}^2$ we write $\|(x, y)\|$ to denote the Euclidean norm $(x_1^2 + x_2^2 + y_1^2 + y_2^2)^{\frac{1}{2}}$.

THEOREM 2. *We assume the following conditions.*

(2a) *The rate of decay of the spectral density $f(\omega_1, \omega_2)$ at high frequencies is proportional to $\|(\omega_1, \omega_2)\|^{-\tau}$, for $\tau > 4$.*

(2b) *The covariance function satisfies the inequality*

$$\int \|(u_1, u_2)\| \, |C(u_1, u_2)| \, du_1 \, du_2 < \infty.$$

(2c) *We require that $\Delta \to 0$, $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \lambda$ for a constant $\lambda > 0$, $\Delta^2 n_1 \to \infty$ and $\Delta^2 n_2 \to \infty$.*

Then

(i) *the expected value of the periodogram, $I_N(\omega_1, \omega_2)$, with $\omega_i \in [-\pi/\Delta, \pi/\Delta]^2$ for $i = 1, 2$, is asymptotically $f(\omega_1, \omega_2)$;*

(ii) *the asymptotic variance of $I_N(\omega_1, \omega_2)$ is $f^2(\omega_1, \omega_2) + f(\omega_1, -\omega_2) f(-\omega_1, \omega_2)$; and*

(iii) $\operatorname{cov}\{I_N(\omega_1, \omega_2), I_N(\omega_1', \omega_2')\}$, *for $\omega_1 \neq \omega_1'$ and $\omega_2 \neq \omega_2'$ is asymptotically equal to*

$$f(\omega_1, -\omega_1') f(-\omega_2, \omega_2') + f(\omega_1, \omega_2') f(\omega_1', \omega_2).$$

By part (i) the nonparametric periodogram is asymptotically an unbiased estimator of the spectral density f.

In the stationary case when $\omega' \neq \omega$ we have that $I_N(\omega')$ and $I_N(\omega)$ are asymptotically uncorrelated; see Theorem 1(iii). However, in the nonstationary case, Theorem 2(iii) shows that $I_N(\omega_1, \omega_2)$ and $I_N(\omega_1', \omega_2')$ are not asymptotically uncorrelated. We propose in § 4·2 another nonstationary periodogram, $\tilde{I}_N$, with better asymptotic properties. In particular, the values of $\tilde{I}_N$ at different frequencies are asymptotically uncorrelated.

4. A NEW MODEL FOR NONSTATIONARY PROCESSES

4·1. *Nonstationary spatial model*

In this section, we propose a new class of nonstationary processes, based on a convolution of local stationary processes. This new model has the advantage that the model is

simultaneously defined everywhere, unlike moving window approaches (Haas, 1995), but it retains the attractive property that, locally in small regions, it behaves like a stationary spatial process.

We model a nonstationary process Z observed on a region D as a convolution of weakly stationary processes:

$$Z(x) = \int_D K(x-s) Z_{\theta(s)}(x)\, ds, \tag{10}$$

where K is a kernel function and $Z_{\theta(s)}(x)$, for $x \in D$, is a family of independent stationary Gaussian processes indexed by $\theta(s)$.

The covariance function $C(x_1, x_2; \theta)$ of Z is a convolution of the covariance functions $C_{\theta(s)}(x_1 - x_2)$ of the stationary processes $Z_{\theta(s)}$:

$$C(x_1, x_2; \theta) = \int_D K(x_1 - s) K(x_2 - s) C_{\theta(s)}(x_1 - x_2)\, ds. \tag{11}$$

The covariance function of $Z_{\theta(s)}$ is stationary with parameter $\theta(s)$, and we assume that $\theta(s)$ is a continuous function on s. The process $Z_{\theta(s)}$ could have a Matérn stationary covariance function of the form

$$C_{\theta(s)}(x) = \frac{\pi^{d/2} \phi_s}{2^{\nu_s - 1} \Gamma(\nu_s + d/2) \alpha_s^{2\nu_s}} (\alpha_s |x|)^{\nu_s} \mathcal{K}_{\nu_s}(\alpha_s |x|), \tag{12}$$

where $\mathcal{K}_{\nu_s}$ is a modified Bessel function, d is the dimension of s, and $\theta(s) = (\nu_s, \alpha_s, \phi_s)$. The parameter α_s^{-1} can be interpreted as the autocorrelation range, ϕ_s is a scale parameter, and the parameter ν_s measures the degree of smoothness of the process $Z_{\theta(s)}$: the higher the value of ν_s the smoother $Z_{\theta(s)}$ would be. For example, when $\nu_s = \frac{1}{2}$, we obtain the exponential covariance function, $C_{\theta(s)}(x) = \pi \phi_s \alpha_s^{-1} \exp(-\alpha_s |x|)$.

In (11) every entry requires an integration. Since each such integration is actually an expectation with respect to a uniform distribution, we propose Monte Carlo integration to approximate the integral (11). We propose to draw an independent set of locations s_i $(i = 1, 2, \ldots, k)$ on D. Hence, we replace $C(x_1, x_2; \theta)$ with

$$\hat{C}(x_1, x_2; \theta) = k^{-1} \sum_{i=1}^{k} K(x_1 - s_i) K(x_2 - s_i) C_{\theta(s_i)}(x_1 - x_2). \tag{13}$$

In this notation, the 'hat' denotes a Monte Carlo integration that can be made arbitrarily accurate and has nothing to do with the data Z. The kernel function $K(x - s_i)$ centred at s_i could be positive for all $x \in D$, or could have compact support. In the latter case, $K(x - s_i)$ would be positive only when x is in a subregion S_i centred at s_i, and this would simplify the calculations.

It is useful to note that, if we define a process $\hat{Z}$ for any $s \in D$ by

$$\hat{Z}(s) = k^{-1} \sum_{i=1}^{k} K(s - s_i) Z_{\theta(s_i)}(s), \tag{14}$$

then $\hat{Z}(s)$ is a Monte Carlo integration for $Z(s)$ as given in (10). Assume that $C(x_1, x_2, \theta)$ is a continuous function of x_1 and x_2. As we increase k and the number of observations $s_1, \ldots, s_k$ becomes more dense on the domain D, the covariance function (13) of the process $\hat{Z}$ converges to the covariance function (11) of the process Z. Thus, $\hat{Z}(s)$ as defined in (14) converges in distribution to $Z(s)$. Therefore, we still obtain asymptotically optimal

predictors for $Z(x_0)$ by using $\hat{C}$ as in (13) and the model (14) to represent the process, instead of the integral representation in (10).

The size of the sample, k, is selected using the following iterative algorithm. We first start with a systematic sample of size k, where k is small, and we increase k by adding a new sample point at a time. At each step of the iterative approach we draw a new sample point in between two neighbouring points in the current sample sequence. Thus, in each iteration we decrease by half the distance between two neighbouring draws. We iterate this process until an Akaike information criterion (Akaike, 1974) suggests no significant improvement in the estimation of the nonstationary covariance of Z by increasing k, equivalent to decreasing the distance between draws in the sample sequence.

Throughout the rest of this paper we simplify the notation by writing Z_i to denote $Z_{\theta(s_i)}$ and $w_i(x)$ to represent $K(x - s_i)$, the kernel or weight function centred at s_i.

4·2. *Spectrum for a new class of nonstationary processes*

The nonstationary process Z is modelled here as in (14), as a mixture of weakly stationary processes Z_i $(i = 1, \ldots, k)$, with $\operatorname{cov}\{Z_i(x), Z_j(y)\} = 0$ for $i \neq j$:

$$Z(x) = \sum_{i=1}^{k} Z_i(x) w_i(x), \tag{15}$$

and we choose k using the AIC approach discussed in § 4·1; see Fuentes (2001) for further discussion about how to model a nonstationary process as a mixture of stationary random fields.

Each stationary process Z_i has the representation

$$Z_i(x) = \int_{\mathbb{R}^2} \exp(ix^{\mathrm{T}}\omega)\, dY_i(\omega), \tag{16}$$

where the Y_i are random functions with uncorrelated increments.

Thus, the spectral representation of Z is $Z(x) = \int_{\mathbb{R}^2} \exp(ix^{\mathrm{T}}\omega)\, dY(\omega)$, where

$$Y(\omega) = \sum_{i=1}^{k} \bar{w}_i * Y_i(\omega), \tag{17}$$

$\bar{w}_i$ is the Fourier transform of w_i and $*$ denotes the following convolution function:

$$\bar{w}_i * Y_i(\omega) = \int_{\mathbb{R}^2} \bar{w}_i(h) Y_i(\omega - h)\, dh.$$

The covariance function of Z can be defined in terms of the covariance function of the orthogonal stationary processes Z_i:

$$\operatorname{cov}\{Z(x_1), Z(x_2)\} = \sum_{i=1}^{k} w_i(x_1) w_i(x_2) \operatorname{cov}\{Z_i(x_1), Z_i(x_2)\}. \tag{18}$$

This is a valid nonstationary covariance function. Then the corresponding spectral density is

$$f(\omega_1, \omega_2) = \sum_{i=1}^{k} f_i * \{\bar{w}_i(\omega_1)\bar{w}_i(\omega_2)\}, \tag{19}$$

where $\bar{w}$ is the fast Fourier transform of w, and $*$ denotes again a convolution function:

$$f_i * \{\bar{w}_i(\omega_1)\bar{w}_i(\omega_2)\} = \int_{\mathbb{R}^2} f_i(\omega)\bar{w}_i(\omega_1 - \omega)\bar{w}_i(\omega_2 - \omega)\, d\omega.$$

4·3. *Nonparametric spectral estimation*

We present here an asymptotically unbiased nonparametric estimator, $\tilde{I}_N$, of the spectral density f of a nonstationary process Z. We model Z as in (15). Thus, a natural way of defining $\tilde{I}_N$ is as a convolution of the periodograms $I_{i,N}$ of the stationary processes Z_i with domain D:

$$\tilde{I}_N(\omega_1, \omega_2) = \sum_{i=1}^{k} I_{i,N} * \{\bar{w}_i(\omega_1)\bar{w}_i(\omega_2)\}, \tag{20}$$

where $*$ denotes the convolution

$$I_{i,N} * \{\bar{w}_i(\omega_1)\bar{w}_i(\omega_2)\} = \sum_{\omega \in J_N} I_{i,N}(\omega)\bar{w}_i(\omega_1 - \omega)\bar{w}_i(\omega_2 - \omega),$$

with J_N the set of the Fourier frequencies (7). The weights w_i have compact support and they help to identify the processes Z_i that are being used. By the definition of f in (19) as a function of the spectral densities f_i $(i = 1, \ldots, k)$ and the fact that the periodograms $I_{i,N}$ are asymptotically unbiased estimators of f_i, we obtain that $\tilde{I}$ is asymptotically unbiased. The asymptotic variance of $\tilde{I}_N(\omega_1, \omega_2)$ can be easily obtained because the processes Z_i are orthogonal. Thus, when $n_i \to \infty$, for $i = 1, 2$, $\Delta \to 0$ and $\Delta n_1 \to \infty$, $\Delta n_2 \to \infty$, the variance of $\tilde{I}_N(\omega_1, \omega_2)$ becomes

$$\sum_{i=1}^{k} f_i^2 * \{\bar{w}_i^2(\omega_1)\bar{w}_i^2(\omega_2)\}.$$

Furthermore, since $\tilde{I}_N$ is a convolution of independent stationary periodograms, we obtain $\operatorname{cov}\{\tilde{I}_N(\omega_1, \omega_2), \tilde{I}_N(\omega_1', \omega_2')\} = 0$ asymptotically.

4·4. *Parametric spectral estimation*

Suppose again that Z takes the form (15), so that we use the expression in (19) for f. The spectral density f is modelled then as a function of the spectral densities f_i $(i = 1, \ldots, k)$.

A parametric estimator $\hat{f}$ of the spectral density is easily obtained from parametric estimators of the spectral densities f_i $(i = 1, \ldots, k)$:

$$\hat{f}(\omega_1, \omega_2) = \sum_{i=1}^{k} \hat{f}_i * \{\bar{w}_i(\omega_1)\bar{w}_i(\omega_2)\}. \tag{21}$$

We now study parametric models for the f_i. A class of practical variograms and autocovariance functions for the stationary processes Z_i can be obtained from the Matérn class of spectral densities

$$f_i(\omega) = \phi_i(\alpha_i^2 + \|\omega\|^2)^{(-\nu_i - \frac{1}{2}d)}, \tag{22}$$

with parameters $\nu_i > 0$, $\alpha_i > 0$ and $\phi_i > 0$, where d is the dimensionality of Z_i. Here, the vector of covariance parameters is $\theta_i = (\phi_i, \nu_i, \alpha_i)$ The parameter α_i^{-1} can be interpreted as the autocorrelation range. The parameter ν_i measures the degree of smoothness of the

process Z_i, in that the higher the value of v_i the smoother Z_i would be, and ϕ_i is the variance σ_i^2 times α_i^{2v}. The corresponding covariance function for the Matérn class is given in (12). For further discussion about the Matérn class see Stein (1999, pp. 48–51).

If we assume that we are sampling from a Gaussian process, then it is straightforward in principle to write down the exact likelihood function and hence to maximise it numerically with respect to the unknown parameters (Kitanidis, 1983; Mardia & Marshall, 1984). The evaluation of the likelihood function requires the computation of the inverse and determinant of the model covariance matrix. In general, environmental datasets are very large and calculation of such determinants is often infeasible, though, if we use the nonstationary model (15), the covariance matrix is a block matrix and calculations are simpler; see for instance, Mardia et al. (1979, formula A.2.3j, p. 457).

5. Application: air pollution data

The goal is to understand and quantify the weekly spatial structure of air pollutants using the output of the regional scale air quality models, Models-3. Models-3 estimate hourly concentrations and fluxes of different pollutants. The primary objective of Models-3 is to improve the environmental management community's ability to evaluate the impact of air quality management practices for multiple pollutants at multiple scales, as part of the regulation process of the air pollutants standards. As an example we examine nitric acid. The spatial domain D is a regular 81×87 grid, where the dimensions of each cell on the grid are 36 km $\times$ 36 km. The 81×87 lattice for Models-3 is a two-dimensional grid that takes account of the earth's curvature. Models-3 provide the estimated concentration for the middle point of each cell. In this example we analyse the spatial structure of the hourly averaged nitric acid concentrations for the week starting 11 July 1995. We fit model (10), taking K to be the Epanechnikov kernel $K(u) = 2\pi^{-1}(1 - \|u\|^2/h^2)$, for $0 < \|u\| < h$, with h an arbitrary bandwidth, and replacing the integral over D by a sum over a grid of cells covering the observation region.

In practice, the choice of h is crucial. The bandwidth should be small to preserve the general 'shape' of the data (Clark, 1977). In a regression setting, reducing the size of the bandwidth reduces the bias but increases the variance. In our spatial setting, since the variance might change with location we do not gain much by increasing h. The shape of the process is represented by the parameter θ, which accounts for the lack of stationarity of Z. Thus, we need to choose h as small as possible to preserve this general shape. However, we also need to ensure that for all $x \in D$ there is at least one s_i, with $K(x - s_i) > 0$, where the $\{s_i\}$ are the k draws on D to calculate the covariance (13). In this application we choose the smallest value of h that satisfies this condition. When the distance between neighbouring points of the sample sequence $s_1, \ldots, s_k$ varies, we could also allow the bandwidth to change with location. If we have k draws from a systematic sample with a distance l between sampling points, then the recommended value for h is $l/\sqrt{2}$. Note that the value of h depends on k.

In this example $k = 9$, which is the optimal value for k based on the AIC criterion. The sample points $s_1, \ldots, s_k$ are a systematic sample and they are plotted in Fig. 1. The distance l between the sampling points is 972 km. The value of h in this application is $h = l/\sqrt{2} =$ 687 km. We used a likelihood approach to estimate the parameters of the nonstationary covariance matrix, which is a mixture of 9 stationary Matérn models of the form (22). Since the kernel function K has compact support, the covariance matrix of Z is approximately a block matrix, which simplifies the calculations; otherwise the evaluation of the

likelihood function requires us to compute the inverse and determinant of a 7209×7209 matrix.

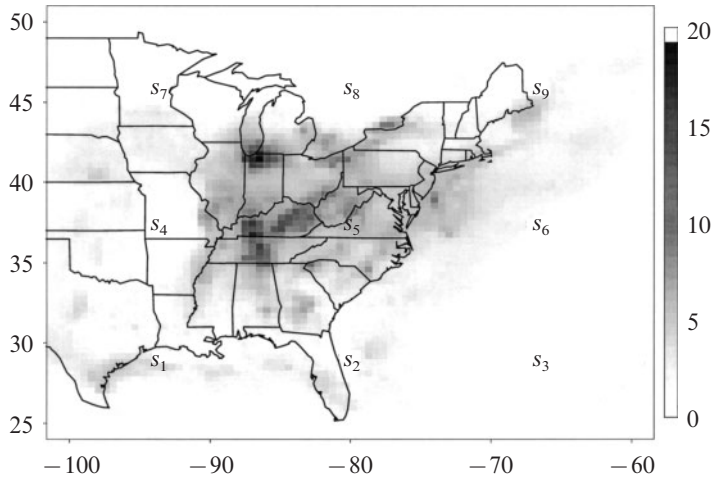

Fig. 1. Air pollution data. Output of Models-3, showing the hourly average concentration of nitric acid in parts per billion for the week starting 11 July 1995. The resolution is 36 km × 36 km. This graph also shows the locations of the sampling points $s_1, \ldots, s_9$.

Figure 2 shows the fitted models for the spectral densities, f_i, of the stationary processes Z_i, for $i = 1, \ldots, 9$. The nonstationary spectral density $f(\omega_1, \omega_2)$, defined in (19), is obtained as a convolution of the densities f_i $(i = 1, \ldots, 9)$. Table 1 shows the estimated parameters for the spectral densities f_i and the corresponding standard errors.

The smoothing parameter represents the rate of decay of the spectral density at high frequencies; this is an indication of how smooth the corresponding process is. The smoothing parameter is approximately 0·5, corresponding to the exponential model for the processes Z_1, Z_4, Z_7 and Z_8. These processes explain the spatial structure of the nitric acid concentrations on the eastern part of our domain; see the location of the sampling points in Fig. 1. We observe a relatively faster rate of decay at high frequencies for the processes Z_5, Z_6, Z_9, Z_2 and Z_3, with a smoothing parameter of approximately 1, corresponding to the Whittle model. These processes explain the spatial structure of the nitric acid concentrations on the western part of our domain mainly over water; the nitric acid seems to be a smoother process over water than over the land surface. The nitric acid is a secondary pollutant, in being the result of photochemical reactions in the atmosphere rather than being emitted directly from sources on the surface. It therefore usually remains in the atmosphere for long periods of time and travels long distances across water.

When the range parameter is large, for example for process Z_8, there is a faster decay of the spectral density at short frequencies. We can appreciate this phenomenon by comparing the spectral density of Z_6, large range, to the spectral density of Z_5, small range. In general we observe larger ranges of autocorrelation on the western part of the grid. Furthermore, on the eastern part we should not expect large ranges because of the discontinuity of the nitric acid concentration that results from transition from land to ocean.

The variance of the process, also called the sill parameter, is the integral of the spectral density function, $\int_{\mathbb{R}^2} f(\omega)\, d\omega$. In this example, the sill is relatively large for Z_8. There is higher spatial variability, large sill, mainly on the Great Lakes area, process Z_8, since the area is downwind from sources of pollution, primarily Chicago.

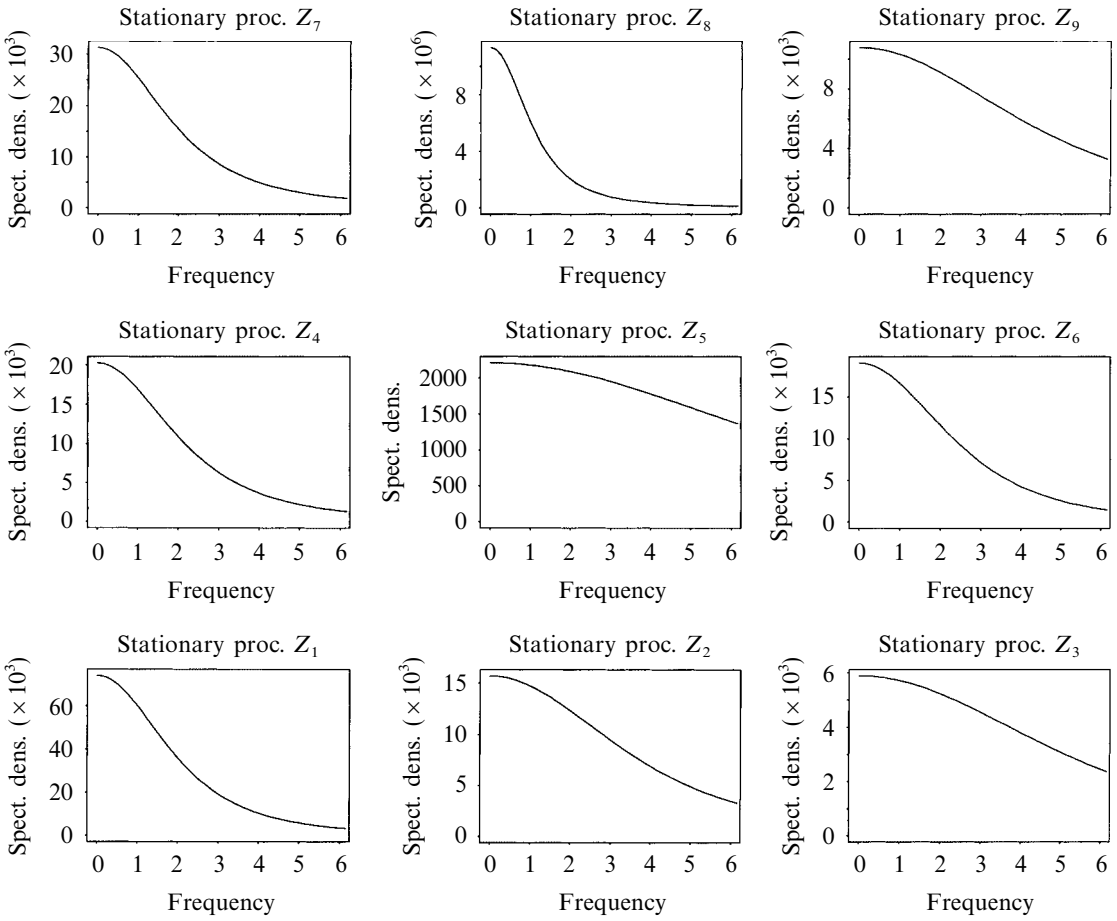

Fig. 2. Air pollution data. Fitted models for the spectral densities, f_i, of the stationary processes Z_i, for $i = 1, \ldots, 9$. Table 1 shows the estimated parameters with the corresponding standard errors for the Matérn spectral densities of Z_i for $i = 1, \ldots, 9$. The estimates and standard errors are obtained using a likelihood approach.

Table 1. *Estimated parameters, and standard errors in parentheses, for the spectral densities of the processes Z_i. The parameters have been estimated using a likelihood approach*

Process	Sill	Range	Smoothness
Z_1	2·9 (1·1)	566 (282)	0·74 (0·03)
Z_2	2·1 (0·5)	315 (131)	0·98 (0·10)
Z_3	1·8 (0·1)	231 (18)	1·01 (0·05)
Z_4	1·10 (0·3)	480 (193)	0·63 (0·03)
Z_5	1·22 (0·07)	150 (20)	1·19 (0·22)
Z_6	1·17 (0·01)	476 (17)	0·91 (0·01)
Z_7	1·4 (0·3)	500 (139)	0·46 (0·03)
Z_8	15 (3)	975 (292)	0·67 (0·01)
Z_9	3 (0·2)	252 (45)	0·84 (0·08)

ACKNOWLEDGEMENT

This research was sponsored by grants from the National Science Foundation and the United States Environmental Protection Agency. The author would like to acknowledge the helpful insight and assistance provided by the US Environmental Protection Agency, Office of Research Development at Research Triangle Park, North Carolina, in particular by Dr Peter Finkelstein and Dr John Pleim. The author also thanks Dr Richard Smith, Dr Peter Bloomfield and Dr Dennis Boos for their valuable comments.

APPENDIX

Proofs of Theorems

Proof of Theorem 1. (i) We define

$$J(\omega) = \Delta(2\pi)^{-1}(n_1 n_2)^{-1/2} \sum_{x_1=1}^{n_1} \sum_{x_2=1}^{n_2} Z(\Delta x) \exp(-i\Delta x^{\mathrm{T}}\omega), \tag{A1}$$

for $\omega \in [-\pi/\Delta, \pi/\Delta]^2$. Then $I_N(\omega) = J(\omega)J^c(\omega)$. We obtain the following expression for the expected value of I_N in terms of C_Δ (Brillinger, 1970):

$$\begin{aligned} E\{J(\omega)J^c(\omega)\} &= \left(\frac{\Delta}{2\pi}\right)^2 \sum_{\tau_1=-(n_1-1)}^{n_1-1} \sum_{\tau_2=-(n_2-1)}^{n_2-1} (1-\tau_1/n_1)(1-\tau_2/n_2) C_\Delta(\tau) \exp(-i\Delta\tau^{\mathrm{T}}\omega) \\ &= f_\Delta(\omega) + \varepsilon_{N,\Delta}, \end{aligned} \tag{A2}$$

where

$$C_\Delta(\tau) = \mathrm{cov}[Z\{\Delta(s+\tau)\}, Z(\Delta s)].$$

We prove that, as $n_i \to \infty$ and $\Delta n_i \to \infty$ for $i = 1, 2$, the term $\varepsilon_{N,\Delta}$ goes uniformly to zero. First, by Lemma P4.3 in Brillinger (1975, p. 403) and under condition (1b), the residual term in (A2) can be bounded as follows:

$$|\varepsilon_{N,\Delta}| \leqslant \frac{1}{n_1 n_2} \sum_{\tau_1=-2(n_1-1)}^{2(n_1-1)} \sum_{\tau_2=2(n_2-1)}^{2(n_2-1)} \|\Delta\tau\| \, |C_\Delta(\tau)| \leqslant \frac{1}{N\Delta^2} \int_{\mathbb{R}^2} \|u\| \, |C(u)| \, du = 0(\Delta^{-2}N^{-1}). \tag{A3}$$

Thus, by combining results (A2) and (A3), we obtain that, for $\omega \in [-\pi/\Delta, \pi/\Delta]^2$, $E\{J(\omega)J^c(\omega)\}$ converges uniformly to $f_\Delta(\omega)$, so that

$$E\{I_N(\omega)\} = f_\Delta(\omega) + 0(\Delta^{-2}N^{-1}).$$

We now obtain that, as $\Delta \to 0$ and $n_i \to \infty$ for $i = 1, 2$, the spectral density of the sampled data $Z(\Delta x)$, for $x \in \mathbb{Z}^2$, becomes the spectral density of $Z(s)$ for $s \in \mathbb{R}^2$; this means that $f_\Delta(\omega)$ converges to $f(\omega)$. We also address the question of how fast $f_\Delta(\omega)$ converges to $f(\omega)$. Define $\mathbb{Z}_{-0} = \mathbb{Z} - \{0\}$ and $\mathbb{Z}^2_{-0} = \mathbb{Z}^2 - \{(0,0)\}$. Then

$$f_\Delta(\omega) - f(\omega) = \sum_{Q \in \mathbb{Z}^2} f\left(\omega + \frac{2\pi Q}{\Delta}\right) - f(\omega) = \sum_{Q \in \mathbb{Z}^2_{-0}} f\left(\omega + \frac{2\pi Q}{\Delta}\right).$$

Under condition (1a), for small enough Δ, the previous expression is proportional to

$$\sum_{Q \in \mathbb{Z}^2_{-0}} \left\| \omega + \frac{2\pi Q}{\Delta} \right\|^{-\tau}. \tag{A4}$$

Under condition (1a) for $\tau > 2$, by the integral test (Stein, 1967, p. 319) the series (A4) converges to zero uniformly in ω at a rate of $O(\Delta^\tau)$. Therefore,

$$E\{J(\omega)J^c(\omega)\} = f(\omega) + 0(\Delta^{-2}N^{-1}) + O(\Delta^\tau). \tag{A5}$$

(ii) By combining (A2) and (A3), as a generalisation of Theorem 5.2.4 in Brillinger (1975, p. 125), we obtain

$$\text{var}\{I_N(\omega)\} = E\{J(\omega)J(\omega)\}E\{J^c(\omega)J^c(\omega)\} + |E\{J(\omega)J^c(\omega)\}|^2 = f_\Delta^2(\omega) + 0(\Delta^{-2}N^{-1}); \tag{A6}$$

the residual term is uniform in ω. Using the same argument as in part (i), we obtain that, as $\Delta \to 0$, $\Delta n_i \to \infty$ and $n_i \to \infty$ for $i = 1, 2$, the rate of convergence of the function $f_\Delta^2(\omega)$ to $f^2(\omega)$ is of order $O(\Delta^{2\tau})$. Thus,

$$\text{var}\{I_N(\omega)\} = f^2(\omega) + 0(\Delta^{-2}N^{-1}) + 0(\Delta^{2\tau}). \tag{A7}$$

(iii) By (A3) and Theorem 5.2.4 in Brillinger (1975, p. 125), we obtain

$$\text{cov}\{I_N(\omega), I_N(\omega')\} = E\{J(\omega)J(\omega')\}E\{J^c(\omega)J^c(\omega')\} + |E\{J(\omega)J^c(\omega')\}|^2 = 0(\Delta^{-2}N^{-1}),$$

the residual term being uniform in ω. □

Proof of Theorem 2. (i) By a similar argument to that in the proof of Theorem 1(i), we obtain

$$E\{I_N(\omega_1, \omega_2)\} = E\{J(\omega_1)J^c(\omega_2)\} = f_\Delta(\omega_1, \omega_2) + \varepsilon_{N,\Delta}.$$

Under condition (2b), the residual term $\varepsilon_{N,\Delta}$ can be uniformly bounded:

$$|\varepsilon_{N,\Delta}| \leqslant \frac{1}{N\Delta^4(2\pi)^4} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \|(u_1, u_2)\| \, |C(u_1, u_2)| \, du_1 \, du_2 = 0(\Delta^{-4}N^{-1}).$$

Thus,

$$E\{I_N(\omega_1, \omega_2)\} = f_\Delta(\omega_1, \omega_2) + 0(\Delta^{-4}N^{-1}). \tag{A8}$$

By using the same argument as in Theorem 1(i) the rate of convergence of $f_\Delta(\omega_1, \omega_2)$ to $f(\omega_1, \omega_2)$ is $0(\Delta^\tau)$. Therefore,

$$f_\Delta(\omega_1, \omega_2) - f(\omega_1, \omega_2) = 0(\Delta^\tau). \tag{A9}$$

Combination of (A8) and (A9) establishes Theorem 2(i).

(ii) We have that

$$\text{var}\{I_N(\omega_1, \omega_2)\} = E\{J(\omega_1)J(\omega_1)\}E\{J^c(\omega_2)J^c(\omega_2)\} + |E\{J(\omega_1)J^c(\omega_2)\}|^2.$$

As in the proof of Theorem 2(i), we obtain that $E\{J(\omega_1)J(\omega_1)\}$ is asymptotically $f(\omega_1, -\omega_1)$, $E\{J^c(\omega_2)J^c(\omega_2)\}$ is $f(-\omega_1, \omega_2)$, and

$$E\{J(\omega_1)J^c(\omega_2)\} = f_\Delta^2(\omega_1, \omega_2) + \varepsilon_{N,\Delta}, \tag{A10}$$

which by (A9) is asymptotically $f^2(\omega_1, \omega_2)$.

(iii) That the covariance $\text{cov}\{I_N(\omega_1, \omega_2), I_N(\omega_1', \omega_2')\}$ is asymptotically

$$f(\omega_1, -\omega_1')f(-\omega_2, \omega_2') + f(\omega_1, \omega_2')f(\omega_1', \omega_2)$$

can be easily proven using the same argument as in (A6). □

References

Akaike, H. (1974). A new look at statistical model identification. *IEEE Trans. Auto. Contr.* **AC-19**, 716–22.

Bloomfield, P. (2000). *Fourier Analysis of Time Series*, 2nd ed. New York: Wiley.

Brillinger, D. R. (1970). The frequency analysis of relations between stationary spatial series. In *Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress*, Ed. R. Pyke , pp. 39–81. Montreal: Canadian Math. Congress.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.

Clark, R. M. (1977). Non-parametric estimation of a smooth regression function. *J. R. Statist. Soc.* B **39**, 107–13.

CONSTANTINE, A. G. & HALL, P. (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *J. R. Statist. Soc.* B **56**, 96–113.
CRESSIE, N. (1993). *Statistics for Spatial Data*, rev. ed. New York: Wiley.
FUENTES, M. (2001). A high frequency kriging approach for nonstationary environmental processes. *Envirometrics* **12**, 469–83.
GUYON, X. (1982). Parameter estimation for a stationary process on a *d*-dimensional lattice. *Biometrika* **69**, 95–105.
GUYON, X. (1992). *Champs Aléatoires sur un Réseau.* Paris: Masson.
HAAS, T. C. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J. Am. Statist. Assoc.* **90**, 1189–99.
JAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions.* New York: Springer-Verlag.
KITANIDIS, P. K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resour. Res.* **19**, 909–21.
LAHIRI, S. M., KAISER, M. K., CRESSIE, N. & HSU, N. J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *J. Am. Statist. Assoc.* **94**, 86–97.
MARDIA, K. V. & MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–46.
MARDIA, K. V., KENT, J. T. & BIBBY, J. M. (1979). *Multivariate Analysis.* New York: Academic Press.
PAWITAN, Y. & O'SULLIVAN F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *J. Am. Statist. Assoc.* **89**, 600–10.
RIPLEY, B. D. (1981). *Spatial Statistics.* New York: John Wiley.
ROSENBLATT, M. R. (1985). *Stationary Sequences and Random Fields.* Boston: Birkhäuser.
STEIN, M. L. (1995). Fixed-domain asymptotics for spatial periodograms. *J. Am. Statist. Assoc.* **90**, 1277–88.
STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* New York: Springer-Verlag.
STEIN, S. K. (1967). *Calculus in the First Three Dimensions.* New York: McGraw Hill, Inc.
WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434–49.

[*Received January* 2001. *Revised May* 2001]