Biometrika (2011), **98**, 4, *pp.* 775–789
doi: 10.1093/biomet/asr056
Advance Access publication 3 November 2011

Nonparametric estimation of the variogram and its spectrum

BY CHUNFENG HUANG

Department of Statistics, Indiana University, Bloomington, Indiana 47408, U.S.A.
huang48@indiana.edu

TAILEN HSING

Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
thsing@umich.edu

AND NOEL CRESSIE

Department of Statistics, The Ohio State University, Columbus, Ohio 43210, U.S.A.
ncressie@stat.osu.edu

SUMMARY

In the study of intrinsically stationary spatial processes, a new nonparametric variogram estimator is proposed through its spectral representation. The methodology is based on estimation of the variogram's spectrum by solving a regularized inverse problem through quadratic programming. The estimated variogram is guaranteed to be conditionally negative-definite. Simulation shows that our estimator is flexible and generally has smaller mean integrated squared error than the parametric estimator under model misspecification. Our methodology is applied to a spatial dataset of decadal temperature changes.

Some key words: Bochner's Theorem; Decadal temperature change; Generalized crossvalidation; Mean integrated squared error; Smoothing spline.

1. INTRODUCTION

Modelling of statistical dependence in spatial data is often achieved through the variogram. Let $\{Y(s): s \in D \subset \mathbb{R}^d\}$ be a real-valued stochastic process defined on a subset D of Euclidean space for which $\mathrm{var}\{Y(s)\} < \infty$, for all $s \in D$. The variogram function is defined as $2G(s,t) = \mathrm{var}\{Y(s) - Y(t)\}$, for $s, t \in D$. Estimation of G from a single, incomplete sample from Y is not practical without further assumptions. Intrinsic stationarity is commonly assumed, i.e., $2G(s,t) = 2\gamma(s-t)$, where the function $2\gamma(\cdot)$ is known as the variogram. The variogram has achieved prominence because of its role in kriging; see Matheron (1963), Cressie (1993, Ch. 3), Chilés J. & Delfiner (1999, Ch. 2) and Stein (1999, Ch. 1).

One crucial property of the variogram is conditional negative-definiteness, here defined as

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j 2\gamma(s_i - s_j) \leqslant 0\,, \tag{1}$$

for any finite number of spatial locations $s_1, \ldots, s_m$ and real numbers $a_1, \ldots, a_m$ satisfying $\sum_{i=1}^{m} a_i = 0$. When the variogram $2\gamma(\cdot)$ is considered as a parameter to be estimated, the usual

method-of-moments estimate based on a sample $\{Y(s_i): i = 1, \ldots, n\}$ (Cressie, 1993, p. 69), is not guaranteed to be conditionally negative-definite. In general, neither are kernel-based nonparametric estimators, such as those found in Yu et al. (2007).

To ensure (1), one strategy has been to develop parametric variogram models that are fitted to various nonparametric estimators at a finite number of spatial lags using weighted least squares (Cressie, 1985). When a time-indexed sequence of spatial processes can be treated as replicates, Sampson & Guttorp (1992) develop a spatially nonstationary model and estimator through multidimensional scaling, although replications of spatial data are not typically encountered in spatial statistics. On the basis of our literature review, there is a clear need for a smooth, nonparametric estimator of the variogram, in the purely spatial setting, that is guaranteed to be conditionally negative definite.

In this article, we develop a new, nonparametric method to estimate a valid variogram based on its spectral representation. This is in contrast to a parametric method to estimate a valid covariance function, developed by Shapiro & Botha (1991). As well as assuming intrinsic stationarity, in this article we assume isotropy, namely $2\gamma(s) = 2\gamma^0(\|s\|)$, $s \in \mathbb{R}^d$. While this appears to be a rather specialized property, anisotropic variograms can be built from it. For any spatial lag s, a geometrically anisotropic variogram satisfies $2\gamma(s) = 2\gamma^0(\|As\|)$, where A is a $d \times d$ nonsingular matrix. For convenience, $2\gamma^0$ will also be referred to as the variogram and henceforth denoted simply as 2γ or $2\gamma(\cdot)$.

A spectral representation of an isotropic variogram in $\mathbb{R}^d$ is Yaglom (1987, § 25.3) $2\gamma(h) = \int_0^\infty \{1 - Y_d(\omega h)\} dF(\omega)$ $(h \geqslant 0)$, where $Y_d(t) = (2/t)^{(d-2)/2}\Gamma(d/2)J_{(d-2)/2}(t)$, $J_\nu(\cdot)$ is the Bessel function of the first kind of order ν, and F is a nondecreasing function on $(0, \infty)$ such that $\int_0^\infty \{1/r(\omega)\} dF(\omega) < \infty$, where $r(\omega) = (1 + \omega^2)/\omega^2$, for $\omega \in (0, \infty)$. The regularization function $r(\omega)$ in the last integral is not unique; any continuous function with the same asymptotic behaviour at 0 and ∞ will work. However, for our purpose this particular choice is a convenient one and we adhere to it throughout. In this article, we consider mostly the planar case, whence $d = 2$, $Y_2(t) = J_0(t)$, and

$$2\gamma(h) = \int_0^\infty \{1 - J_0(\omega h)\} dF(\omega) \quad (h \geqslant 0). \tag{2}$$

We shall further assume that F has a derivative, which we write as $r(\omega) f(\omega)$, and hence (2) becomes

$$2\gamma(h) = \int_0^\infty \{1 - J_0(\omega h)\} r(\omega) f(\omega) d\omega, \tag{3}$$

where $f(\omega) \geqslant 0$ and $\int_0^\infty f(\omega) d\omega < \infty$, and is termed the variogram spectrum. The spectral representation (3) of the variogram 2γ is the analogue of Bochner's theorem for a spatial covariance function (Cressie, 1993, p. 84).

The equivalence between the covariance function and its spectrum has not received as much exposure in spatial statistics as it has in time series analysis. However, there have been articles that have recognized its utility for estimation of the covariance function through Bochner's theorem and modelling the covariance spectrum (Shapiro & Botha, 1991; Hall et al., 1994; Fuentes, 2002, 2007; Pyrcz & Deutsch, 2006; Im et al., 2007; Powojowski, 2008; Matsuda & Yajima, 2009 and Huang et al., 2011). For example, assuming an isotropic covariance function, the covariance spectrum is modelled by a linear combination of B-splines plus a power function in Im et al. (2007), while it is modelled by a piecewise constant function in Powojowski (2008); the Hankel transform is used in both cases to compute the covariance function. In contrast, our approach

is nonparametric in the variogram spectral domain; since generalized covariance functions (Matheron, 1973) also have spectral representations, our research also offers a general approach to their nonparametric estimation.

Our approach is to pose the problem of estimation of 2γ through (3) as a form of an ill-posed integral equation, and to employ general spline methodology to estimate the spectrum. Provided the estimated variogram spectrum is nonnegative, substituting it into (3) results in a variogram estimate that is conditionally negative-definite.

2. Estimation methodology

2·1. *General spline estimation*

Let Y be an intrinsically stationary process with isotropic variogram 2γ and variogram spectrum $f(\omega)$, for $\omega \in (0, \infty)$, defined in (3), which means that the increments have first moment zero and second moment given by the variogram. In this article, we consider the spatial-data process with measurement error,

$$X(t) = Y(t) + \epsilon(t) \quad (t \in \mathbb{R}^2),$$

where $\epsilon(\cdot)$ is a mean zero, white noise, measurement error process that is independent of the process $Y(\cdot)$, and $\text{var}\{\epsilon(t)\} = \sigma_\epsilon^2$. Assume that the data are observed at spatial locations $t_1, \ldots, t_N$. Therefore, we have $E\{X(t_i) - X(t_j)\}^2 = 2\gamma(\|t_i - t_j\|) + 2\sigma_\epsilon^2$. That is, $x_{i,j} \equiv \{X(t_i) - X(t_j)\}^2$ is an unbiased estimator of

$$2\gamma(\|t_i - t_j\|) + 2\sigma_\epsilon^2 = \int_0^\infty \{1 - J_0(\omega\|t_i - t_j\|)\} r(\omega) f(\omega) d\omega + 2\sigma_\epsilon^2.$$

Intuitively, the following sum of squares will be small for a function g close to f and a real number c close to $2\sigma_\epsilon^2$:

$$\sum_{i \neq j}^{N} \left[x_{i,j} - \int_0^\infty \{1 - J_0(\omega\|t_i - t_j\|)\} r(\omega) g(\omega) d\omega - c \right]^2.$$

Conversely, any nonnegative function g and nonnegative number c that make this sum small can be considered as candidate estimators of f and $2\sigma_\epsilon^2$, respectively. However, searching for estimators in this manner constitutes an ill-posed inverse problem (O'Sullivan, 1986), for which regularization is essential.

Before we proceed, note that while our aim is to estimate the spectrum on the whole real line, for a variety of considerations, it makes sense to conduct the estimation on a bounded interval $[0, \nu]$. The first consideration is that there is a very well established approximation theory for functions on a finite interval that can be readily adapted for the problem at hand. The second consideration is more fundamental and can be most easily understood within the context of stationary processes. If the observations of a stationary process are made on a grid, then the spectrum is not identifiable in that there is not enough information in the data to distinguish the true spectrum on an unbounded interval from a folded spectrum on a bounded interval. This is known as the aliasing effect; see, for example, Yaglom (1987, p. 187). A discussion of the choice of ν in that setting can be found in Huang et al. (2011), where one can choose ν to be π divided by the grid size when data are equally spaced and ν to be π divided by the averaged grid size when data are irregularly spaced in $\mathbb{R}^1$. The aliasing effect is less studied in intrinsically stationary spatial processes. As indicated in § 3, our nonparametric estimates of the variogram spectrum and the

estimated variogram are quite robust to the choice of ν. As a working rule, we have had success with $\nu = \pi/\tau$, where τ is the length of the diagonal of a grid cell.

An effective way to address the penalized-least-squares problem described above is through the general smoothing-spline approach of Wahba (1990, Ch. 1). Then, accounting for the constraints on g and c, the problem is to estimate f and $2\sigma_\epsilon^2$ through finding the minimizer associated with

$$\min_{c\geqslant 0, g\in W_m[0,\nu], g\geqslant 0} \left\{ \sum_{i\neq j}^{N} \left(x_{i,j} - L_{i,j}g - c\right)^2 + \lambda J(g) \right\}, \tag{4}$$

for some smoothing or regularization parameter $\lambda > 0$. In (4), the functional $L_{i,j}$ is defined as

$$L_{i,j}g = \int_0^\nu \{1 - J_0(\omega h_{i,j})\} r(\omega) g(\omega) d\omega,$$

where $h_{i,j} = \|t_i - t_j\|$, $W_m[0, \nu]$ is the Sobolev space of order m, of functions on $[0, \nu]$ that are m-times differentiable with square integrable mth derivative; and

$$J(g) = \int_0^\nu \left\{ g^{(m)}(\omega) \right\}^2 d\omega. \tag{5}$$

The Sobolev space $W_m[0, \nu]$ has the structure $\mathscr{H}_0 \oplus \mathscr{H}_1$, where $\mathscr{H}_0$ is an m-dimensional space of polynomials of degree $m - 1$ spanned by $\phi_j(x) = x^{j-1}$ $(j = 1, \ldots, m)$, $\mathscr{H}_1$ is the collection of functions on $[0, \nu]$ such that $g^{(\tau)}(0) = 0$ $(\tau = 0, \ldots, m - 1)$, and $\int_0^\nu \left(g^{(m)}\right)^2 < \infty$. See Wahba (1990, Ch. 1) for more details, where it is shown that $W_m[0, \nu]$ is a reproducing kernel Hilbert space. Let $R(\cdot, \cdot)$ be the reproducing kernel of $W_m[0, \nu]$, and define $R_t(\cdot) \equiv R(t, \cdot)$. The constraint $g \geqslant 0$ can be approximated by $g(\omega_1) \geqslant 0, \ldots, g(\omega_L) \geqslant 0$ for a set of densely distributed $\{\omega_1, \ldots, \omega_L\}$ in $[0, \nu]$; without loss of generality, we take them to be equally spaced. By the reproducing property, these constraints can be expressed as

$$\langle g, R_{\omega_l} \rangle \geqslant 0 \quad (l = 1, \ldots, L), \tag{6}$$

where $\langle \cdot, \cdot \rangle$ is the inner product of $W_m[0, \nu]$. Since $1 - J_0(\omega) \sim \omega^2/4$, it is easy to see that $L_{i,j}$ is a bounded linear functional on $W_m[0, 1]$. Let $\eta_{i,j}$ be the representer of $L_{i,j}$, namely, $\langle \eta_{i,j}, g \rangle = L_{i,j}g$, for $g \in W_m[0, 1]$. Also, let $\xi_{i,j} = \mathscr{P}\eta_{i,j}$, and $\rho_j = \mathscr{P}R_{\omega_j}$, where $\mathscr{P}$ is the projection operator onto $\mathscr{H}_1$.

A referee has suggested that the positivity of g could be handled by writing it as $g(\omega) = \exp\{h(\omega)\}$ for some function h, which would lead to a different approach that may not involve reproducing kernel Hilbert spaces. This offers an interesting possibility that is best pursued elsewhere. In spatial statistics, computational issues are generally quite challenging, and a key advantage of our approach is the ease of computation. Another advantage is that there is already a sizable literature studying the various issues in our optimization algorithm.

From Wahba (1990, § 9.4), replacing the constraint $g \geqslant 0$ in (4) by (6) leads to the unique minimizer

$$\hat{g} = \sum_{i\neq j}^{N} c_{i,j}\xi_{i,j} + \sum_{\tau=1}^{m} d_\tau \phi_\tau + \sum_{j=1}^{L} b_j \rho_j;$$

the coefficients can be found through quadratic programming if $\xi_{i,j}$, $L_{i,j}\phi_\tau$, and $\langle \xi_{i,j}, \xi_{i',j'} \rangle$ are known. However, we are unable to find closed-form expressions for $\xi_{i,j}$, $L_{i,j}\phi_\tau$ and $\langle \xi_{i,j}, \xi_{i',j'} \rangle$, and hence some form of numerical approximation becomes necessary. A method of matched quadrature in Nychka et al. (1984) is a possibility but, unfortunately, the complexity of the expressions for $\xi_{i,j}$ and ϕ_τ complicates the implementation of this method. In the next subsection, we present a simple, intuitively motivated numerical solution.

2·2. *Numerical solution*

Let

$$\ell_{i,j} = (v/L)\left[\{1 - J_0(\omega_1 h_{i,j})\}r(\omega_1), \ldots, \{1 - J_0(\omega_L h_{i,j})\}r(\omega_L)\right]^{\mathrm{T}},$$

and $g_\omega = \left\{g(\omega_1), \ldots, g(\omega_L)\right\}^{\mathrm{T}}$, where $\{\omega_1, \ldots, \omega_L\}$ are as in (6). For a sufficiently dense set of $\{\omega_l\}$, $L_{i,j}g$ is well approximated by the Riemann sum, $\ell_{i,j}^{\mathrm{T}} g_\omega$. Hence, if $L_{i,j}g$ is replaced by $\ell_{i,j}^{\mathrm{T}} g_\omega$, and $g \geqslant 0$ is replaced by $g_\omega \geqslant 0$, elementwise, in (4), then the solution $\hat{g}$ must be a natural spline with knots $\omega_1, \ldots, \omega_L$, and the roughness penalty $J(g)$ can be expressed as $g_\omega^{\mathrm{T}} K g_\omega$, for some $L \times L$ matrix K (Green & Silverman, 1994, Ch. 2). Specifically, $\hat{g}$ is the unique natural spline that interpolates $(\omega_l, \hat{g}_l)$ $(\ell = 1, \ldots, L)$, where $\hat{g}_\omega = (\hat{g}_1, \ldots, \hat{g}_L)^{\mathrm{T}}$ is the minimizer associated with

$$\min_{c \geqslant 0, g_\omega \in \mathbb{R}^L, g_\omega \geqslant 0} \left\{ \sum_{i \neq j}^{N} \left(x_{i,j} - \ell_{i,j}^{\mathrm{T}} g_\omega - c \right)^2 + \lambda g_\omega^{\mathrm{T}} K g_\omega \right\}. \tag{7}$$

For convenience of notation, let $n = N(N-1)/2$, and write $\{x_{i,j}\}$ and $\{\ell_{i,j}\}$ as $\{x_1, \ldots, x_n\}$ and $\{\ell_1, \ldots, \ell_n\}$, respectively; also, define $x = (x_1, \ldots, x_n)^{\mathrm{T}}$, and

$$B = \begin{pmatrix} \ell_1^{\mathrm{T}} & 1 \\ \vdots & \vdots \\ \ell_n^{\mathrm{T}} & 1 \end{pmatrix}, \quad \Psi = \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix}.$$

Then (7) can be expressed as

$$\min_{v \in \mathbb{R}^{L+1}} \left\{ v^{\mathrm{T}}(B^{\mathrm{T}}B + \lambda\Psi)v - 2v^{\mathrm{T}}B^{\mathrm{T}}x \right\}, \text{ subject to } v \equiv (v_1, \ldots, v_L, v_{L+1})^{\mathrm{T}} \geqslant 0, \tag{8}$$

which can be solved numerically by quadratic programming in a standard way (Gill et al., 1981, Ch. 13). Let the minimizing vector of (8) be $\hat{v} = (\hat{v}_1, \ldots, \hat{v}_L, \hat{v}_{L+1})^{\mathrm{T}}$. Then, the measurement error variance estimate is $\hat{\sigma}_\epsilon^2 = \hat{v}_{L+1}/2$. The values of the spectrum at the knots can be estimated as $\hat{f}(\omega_1) = \hat{v}_1, \ldots, \hat{f}(\omega_L) = \hat{v}_L$, and the entire function of f is then estimated by the unique natural spline that interpolates $\{\omega_l, \hat{f}(\omega_l)\}$ $(\ell = 1, \ldots, L)$. The question arises whether that entire function is nonnegative. For L large enough and $\{\omega_l\}$ equally spaced on $[0, \nu]$, we have not encountered a single example for which $\hat{f}$ is negative.

To estimate the variogram, we substitute $\hat{f}(\cdot)$ into (3) and define

$$2\hat{\gamma}(h) = \int_0^\nu \{1 - J_0(\omega h)\}r(\omega)\hat{f}(\omega)d\omega \quad (h > 0).$$

Now $\hat{f}(\omega)$ is a degree $2m-1$ polynomial on each interval $[\omega_i, \omega_{i+1})$, for $i = 1, \ldots, L-1$, and it is a degree $m-1$ polynomial outside $[\omega_1, \omega_L]$, where the coefficients of each polynomial

are determined by $\hat{f}(\omega_1), \ldots, \hat{f}(\omega_L)$ (de Boor, 1978, Ch. 4; Green & Silverman, 1994, Ch. 2). Therefore, within each interval, the computation of $2\hat{\gamma}(h)$ can be carried out through the integration of a product of the Bessel function and the spline polynomials. In practice, we use as our variogram estimator the Riemann sum

$$2\hat{\gamma}(h) = \frac{\nu}{L}\sum_{l=1}^{L}\{1 - J_0(\omega_l h)\}r(\omega_l)\hat{f}(\omega_l) \quad (h > 0). \tag{9}$$

Since the $\{\hat{f}(\omega_\ell)\}$ are all nonnegative, (9) is always conditionally negative-definite. An alternative approach that can be used to perform numerical integration here, as in (7), is the trapezoidal rule; see Press et al. (2007, Ch. 4). If the integrand has bounded second derivative, the rates of approximation for both of these methods are $O(L^{-2})$ using equally spaced $\{\omega_l\}$. More refined numerical integration recipes, such as Simpson's rule, may also be considered. A related question is the choice of L. Theoretically, a large L leads to a better approximation in a numerical integration. However, that has to be balanced by the numerical error that is generated in solving (8). In our simulation study, presented in § 3, we have experimented using both the Riemann sum and trapezoidal rule and choices of L between 200 and 500. No clear differences have resulted from the different combinations of these methods and choices.

When the data are located on a regular grid, each distance $h_{i,j} = \|t_i - t_j\|$ will be duplicated a large number of times. For example, suppose data are observed at every node of an $N_0 \times N_0$ grid with grid spacing h_1. Then $N = N_0^2$, there are $2N_0(N_0 - 1)$ pairs $\{X(t_i), X(t_j)\}$ having the same distance h_1, and we can define

$$w_1 = 2N_0(N_0 - 1), \quad a_1 = \frac{1}{w_1}\sum_{h_{i,j}=h_1} x_{i,j}\,. \tag{10}$$

Going through all possible distances in this manner, we obtain a sequence $\{(a_i, h_i, w_i) : i = 1, \ldots, N_0(N_0 - 1)\}$, and the double summation in (4) reduces to a single weighted summation,

$$\min_{c\geqslant 0, g_\omega \in \mathbb{R}^L, g_\omega \geqslant 0}\left\{\sum_{i=1}^{n_0} w_i(a_i - \ell_i^{\mathrm{T}} g_\omega - c)^2 + \lambda g_\omega^{\mathrm{T}} K g_\omega\right\}, \tag{11}$$

where $n_0 = N_0(N_0 - 1)$, and

$$\ell_i = (\nu/L)\left[\{1 - J_0(\omega_1 h_i)\}r(\omega_1), \ldots, \{1 - J_0(\omega_L h_i)\}r(\omega_L)\right]^{\mathrm{T}}.$$

Let $a = (a_1, \ldots, a_{n_0})^{\mathrm{T}}$ and $W = \mathrm{diag}(w_1, \ldots, w_{n_0})$. Then (11) can be expressed as

$$\min_{v\in\mathbb{R}^{L+1}}\{v^{\mathrm{T}}(B^{\mathrm{T}}WB + \lambda\Psi)v - 2a^{\mathrm{T}}WBv\}, \text{ subject to } v \geqslant 0. \tag{12}$$

Thus, the computational burden can be significantly reduced when the observations are on a regular grid.

If there are missing data on the regular grid, there may be some h_i for which no squared differences are computed. Then $w_i = 0$ and the term does not contribute to the summation in (11), in which case the dimensions of the matrices in (12) will decrease.

At most, the dimension of B is $\{N(N - 1)/2\} \times (L + 1)$. This optimization in (7) can be carried out efficiently when the sample size N is small, but there could be cases where the size of the matrix B overwhelms the computations. It is worth noting that $\{a_1, \ldots, a_{n_0}\}$ defined by (10)

are the method-of-moments variogram estimates, sometimes called the experimental variogram (Matheron, 1963) referred to in § 1. As the experimental variogram extends to irregularly located data using tolerance regions (Journel & Huijbregts, 1978, p. 194; Omre, 1984), similar ideas can be employed to extend (12) to the case where observations are not on a grid, for example, rounding the distance $h_{i,j}$ up to a given number of significant digits. Hence, our approach is numerically viable for data that are both regularly located and irregularly located.

2·3. *Theoretical perspective*

The theoretical issue of rate of convergence, or even consistency, of our procedure is difficult and will not be addressed directly in this paper. There have been a number of papers that discuss the rate of convergence of regularized solutions of Fredholm integral equations of the first kind based on noisy data, including Wahba (1973), Nychka et al. (1984), Cox (1988), Lukas (1988) and Nychka & Cox (1989). They all highlight the difficulty of a solution to this problem that has checkable conditions. Generically, we wish to estimate f on $[0, 1]$ based on data $\{(z_i, s_i)\}$, where

$$z_i = \int_0^1 H(s_i, s) f(s) ds + \varepsilon_i,$$

with some known function H and mean zero error ε_i. In our case, let

$$H(u, s) = \{1 - J_0(us)\} r(u), \tag{13}$$

where the sequences $\{z_i\}$ and $\{s_i\}$ are defined in § 2·1 and denoted there by $\{x_{i,j}\}$ and $\{h_{i,j}\}$, respectively. Assume that $f \in W_m[0, 1]$ and let R_1 be the reproducing kernel of $\mathscr{H}_1$. Then the papers cited above show that one of the keys to computing the rate of convergence of the regularized estimator is obtaining the rate of decay of the eigenvalues of the operator,

$$\mathscr{Q} : f \mapsto \int_0^1 H_1(s, t) f(t) dt,$$

where

$$H_1(s, t) \equiv \int_0^1 \int_0^1 H(s, u) R_1(u, v) H(t, v) du dv.$$

We have not found a general theoretical way to study the rate of decay of the eigenvalues, and hence we compute them numerically, as suggested by Nychka et al. (1984, pp. 843–4). For example, when $m = 2$ and hence we consider the Sobolev space $W_2[0, \nu]$, then

$$R_1(s, t) = st \min(s, t) - \frac{1}{2}(s + t) \min(s, t)^2 + \frac{1}{3} \min(s, t)^3.$$

Let H be given by (13). A discrete version of H_1 is the matrix $Q = \{H_1(u_i, u_j)\}_{i,j=1}^n$, where $\{u_i\}$ are equally spaced. The eigenvalues of Q approximate those of $\mathscr{Q}$. A log-log plot of the ordered eigenvalues of Q versus index $k = 1, \ldots, n = 1000$, is presented in Fig. 1. The main part of the plot can be approximated by a line with slope $-5{\cdot}74$. This suggests that $\lambda_k = O(k^{-\beta})$ where $\beta \simeq 6 > 1$, which would result in a regularized solution consistent in some sense given a proper choice of the smoothing parameter (Lukas, 1988; Nychka & Cox, 1989). We emphasize again that this is strong numerical evidence that $\hat{f}$ is consistent, but we have no theoretical result to that effect.

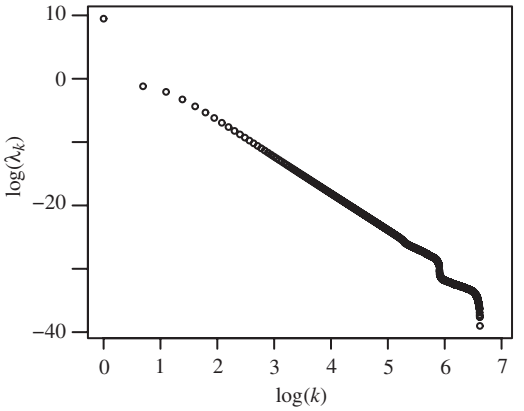

Fig. 1. Log-log plot of eigenvalues of the matrix Q for $k = 1, \ldots, n = 1000$; the slope of the linear part of the plot is estimated to be $-5{\cdot}74$.

2·4. *Smoothing-parameter selection*

To implement our procedure in practice, a data-driven choice of smoothing parameter λ will be needed. A generalized crossvalidation approach for smoothing splines has been proposed by Villalobos & Wahba (1987), which involves appropriate minimization with linear constraints and assumes independent observations.

In our case, for each candidate smoothing parameter λ, we can solve the quadratic programming problem in (12). Let the minimizing vector be $\hat{v}(\lambda)$; then, $\hat{y}(\lambda) = B\hat{v}(\lambda)$. Let $\tilde{B}$ and $\tilde{\Psi}$ be the respective matrices B and Ψ with rows and columns that correspond to the active constraints removed, where an active constraint is one that causes the solution of the quadratic programming problem to equal 0. This can be obtained by examining the solution (Villalobos & Wahba, 1987). Consider $\mathrm{RSS}(\lambda) = \sum_{i=1}^{n_0} w_i(y_i - \hat{y}_i)^2$, and $A(\lambda) = \tilde{B}(\tilde{B}^{\mathrm{T}} W \tilde{B} + \lambda\tilde{\Psi})^{-1}\tilde{B}^{\mathrm{T}} W$. Further, in our case, the spatial data are correlated, for which several modified generalized crossvalidation criteria can be found in Wang (1998). One of the modified generalized crossvalidation functions given there (Wang, 1998, equation (15), $k = 0$) is

$$\frac{\mathrm{RSS}(\lambda)}{[1 - (1/p_0)\mathrm{tr}\{\Xi^{-1}A(\lambda)\}]^2}, \tag{14}$$

where Ξ is the covariance matrix of y and $p_0 = \mathrm{tr}(\Xi^{-1})$.

In Wang (1998), the covariance Ξ in (14) is assumed to have a parametric form, and the parameters are estimated through maximum likelihood. Since we wish to estimate the variogram nonparametrically, we replace Ξ with W^{-1} and p_0 with p, see below, and we propose to minimize the following criterion with respect to λ:

$$V(\lambda) \equiv \frac{\mathrm{RSS}(\lambda)}{[1 - (1/p)\mathrm{tr}\{WA(\lambda)\}]^2}, \tag{15}$$

where $p = \mathrm{tr}\{WA(0)\} = \mathrm{tr}\{W\tilde{B}(\tilde{B}^{\mathrm{T}}W\tilde{B})^{+}\tilde{B}^{\mathrm{T}}W\}$, and M^{+} denotes the Moore–Penrose generalized inverse of M. Finally, we choose the smoothing parameter λ that minimizes $V(\lambda)$. The procedure is implemented in our simulation study and produces good results; see § 3.

Table 1. *Formulae of spectra and variograms for four parametric models*

	Spectrum $f(\omega)$	Variogram $2\gamma(h)$
Model 1	$\dfrac{c_0\omega^2 e^{-a\omega}}{1+\omega^2}$	$c_0\left\{\dfrac{1}{a}-\dfrac{1}{(a^2+h^2)^{1/2}}\right\}+2\sigma_\epsilon^2$
	$0<\omega<\infty, \quad a>0, \quad c_0>0$	
Model 2	$c_0\dfrac{2\kappa\omega^3}{a^{2\kappa}(\omega^2+1/a^2)^{\kappa+1}(1+\omega^2)}$	$c_0\left\{1-\dfrac{1}{2^{\kappa-1}\Gamma(\kappa)}(h/a)^\kappa K_\kappa(h/a)\right\}+2\sigma_\epsilon^2$
	$0<\omega<\infty, \quad a>0, \quad c_0>0, \quad \kappa>0$	
Model 3	$\dfrac{c_0a^2\omega^3}{(1-a^2\omega^2)^{1/2}(1+\omega^2)}$	$c_0\left\{1-\dfrac{a}{h}\sin\left(\dfrac{h}{a}\right)\right\}+2\sigma_\epsilon^2$
	$0<\omega<1/a, \quad a>0, \quad c_0>0$	
Model 4	$\dfrac{a\omega^{1/2}}{1+\omega^2}$	$(1\cdot 911955)ah^{1/2}+2\sigma_\epsilon^2$
	$0<\omega<\infty, \quad c_0>0$	

The spectra are given for the measurement error free process Y.

3. Simulation study

To demonstrate the efficacy of our methodology, a simulation study is conducted. We consider four isotropic intrinsically stationary processes on $\mathbb{R}^2$ with measurement error having variance σ_ϵ^2. The formulae for the spectra and variograms for these four models are presented in Table 1. The variogram spectrum, $f(\omega)$, given in Table 1 is the measurement error free version.

The variogram of Model 2 is in the Matérn family, where $K_\kappa(\cdot)$ denotes the modified Bessel function of the second kind of order κ, and the variograms of Models 3 and 4 are sometimes referred to as the wave or hole-effect model and the power model, respectively (Cressie, 1993, p. 63). The variogram and variogram spectrum for Model 2 can be found in Stein (1999, § 2.10) and Yaglom (1987, § 22.2), the correspondences between variograms and variogram spectra for Models 1, 3 and 4 can be obtained through Gradshteyn & Ryzhik (2007, equations (6.611.1), (6.554.2)), and Abramowitz & Stegun (1972, equation (11.4.18)), respectively. The variograms in Models 1–3 are bounded, while the variogram in Model 4 increases with h according to $O(h^{1/2})$, and hence it is unbounded. Plots of the variogram spectra and variograms are shown in Fig. 2.

The integrated squared error (Yu et al., 2007) is used to assess the spectrum and variogram estimators:

$$\text{ISE}(f)=\int_0^\nu\{\hat{f}(\omega)-f(\omega)\}^2d\omega, \quad \text{ISE}(2\gamma)=\int_{h_l}^{h_u}\{2\hat{\gamma}(h)-2\gamma(h)\}^2dh,$$

where h_l and h_u are the smallest and largest lags, respectively, for which variogram estimates are available. Means and standard errors of the integrated squared error are obtained by simulation. In most applications in spatial statistics, variogram estimation is carried out parametrically. It is of interest to compare the parametric and nonparametric approaches, where parameter estimates are obtained here through the weighted-least-squares approach due to Cressie (1985). In our comparisons, we use variograms of the four models as parametric models and fit them to the simulated data. Comparisons are made through means and standard deviations of the integrated squared error based on 100 simulation runs.

For each individual simulation, the numerical method outlined in § 2 is used to estimate the spectrum, the variogram, and σ_ϵ^2. We choose $m=2$ in (5), set $L=200$ and estimate $f(\cdot)$ on

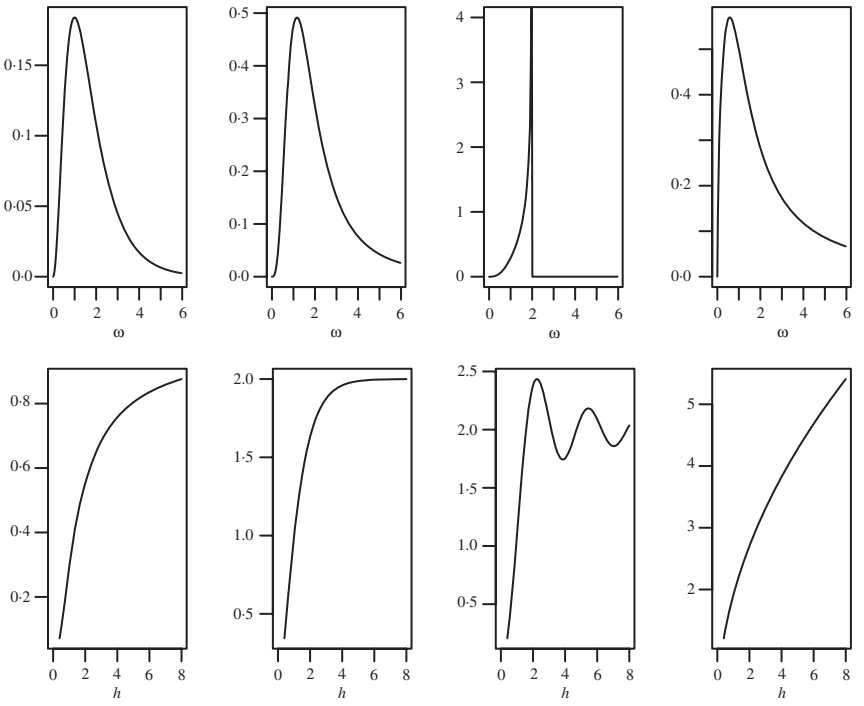

Fig. 2. The spectra and variograms of Models 1–4. Upper panels are the spectra and lower panels are the corresponding variograms. The parameters are $a = 1$, $c_0 = 1$ for Model 1, $\kappa = 1$, $c_0 = 2$, $a = 0{\cdot}8$ for Model 2, $a = 0{\cdot}5$, $c_0 = 2$ for Model 3 and $a = 1$ for Model 4.

the bounded interval $[0, \nu]$, where ν is either 6 or 10. The smoothing parameter is selected by minimizing $V(\lambda)$ in (15). The processes are observed on a 60×60 regular grid with grid spacing equal to 0·4, and hence $N = 3600$ and the domain is $[0, 24] \times [0, 24]$.

We first simulate the process from the parametric model, Model 1. Different settings of σ_ϵ^2 and ν are used, and the results are shown in Table 2. For variogram estimation, the mean $\text{ISE}(2\gamma)$ values for Model 1 are the smallest; these represent the benchmark for this part of the study, since the data are simulated from Model 1. If one chooses an incorrect parametric model, the estimate may be quite different from the true variogram. For example, we know that Models 3 and 4 are the wrong models, and notice that their mean $\text{ISE}(2\gamma)$ values are substantially larger than those of Model 1. The mean $\text{ISE}(2\gamma)$ values of our nonparametric estimate consistently outperform those for Models 3 and 4, are the same as those for Model 2, and are relatively close to the benchmark given by fitting Model 1. The measurement error variance estimates also show a similar pattern. It is interesting to note that the mean $\text{ISE}(2\gamma)$ values for Model 2 compare well with the benchmark, although the measurement error variance is better estimated through our nonparametric procedure. The mean $\text{ISE}(f)$ values appear to be relatively small; given that $\int_0^\infty f(\omega)^2 d\omega = 0{\cdot}046$, this shows that our nonparametric estimates of the variogram spectrum are close to the true variogram spectrum. In addition, Table 2 shows that the mean integrated squared error values of both the estimated variogram spectrum and the estimated variogram are quite robust to the choice of ν. In this study, we also considered the irregular-grid case, where $N = 3600$ locations in $\mathbb{R}^2$ are generated uniformly on $[0, 24] \times [0, 24]$. The simulation results mimic the regular-grid case, and hence are not included in Table 2.

To illustrate the new methodology that results in the nonparametric estimates in Table 2, one realization from a simulation of Model 1 is analysed and presented in Fig. 3, where the simulation

Table 2. *Simulation results based on true model, Model 1*

		$\sigma_\epsilon^2 = 0{\cdot}16$		$\sigma_\epsilon^2 = 0{\cdot}36$	
		$\nu = 10$	$\nu = 6$	$\nu = 10$	$\nu = 6$
		Regular grid			
Model 1	ISE(2γ)	5·34 (0·80)		6·12 (0·90)	
	$\hat{\sigma}_\epsilon^2$	15·70 (0·29)		35·56 (0·39)	
Model 2	ISE(2γ)	5·47 (0·83)		6·22 (0·90)	
	$\hat{\sigma}_\epsilon^2$	7·00 (0·65)		19·83 (1·48)	
Model 3	ISE(2γ)	12·15 (0·83)		14·63 (0·92)	
	$\hat{\sigma}_\epsilon^2$	29·61 (0·54)		48·65 (0·49)	
Model 4	ISE(2γ)	9·88 (0·86)		10·74 (0·95)	
	$\hat{\sigma}_\epsilon^2$	22·59 (0·56)		43·30 (0·55)	
NP	ISE(2γ)	5·50 (0·80)	5·52 (0·80)	6·24 (0·89)	6·26 (0·89)
	$\hat{\sigma}_\epsilon^2$	12·07 (0·44)	15·44 (0·10)	31·05 (0·52)	35·29 (0·13)
	ISE(f)	0·32 (0·03)	0·20 (0·02)	0·42 (0·04)	0·22 (0·02)

Entries in the table show the means and standard deviations, in parentheses, of the respective integrated squared errors and $\hat{\sigma}_\epsilon^2$ values. The entries in rows labelled Models 1–4 show the results from parametric fitting when those models are selected as the parametric model, and the entries in rows labelled NP show the results from our proposed nonparametric procedure. The entries are in units of 10^{-2}.

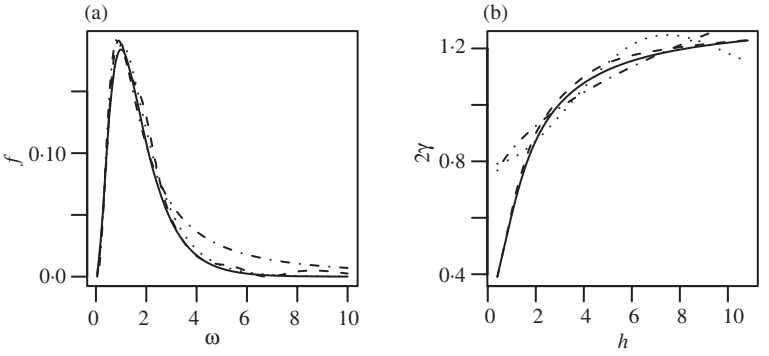

Fig. 3. Panels (a) and (b) show, respectively, the estimates of the variogram spectrum and the variogram. The true model is Model 1. In (a), the solid line is the true spectrum; the dashed line is our nonparametric estimate; the dotted line is the estimate assuming Model 1; the dot-dashed line is the estimate assuming Model 2. Panel (a) does not contain the fits assuming Models 3 and 4, since if they were included it would distort the scale. In panel (b), the solid line is the true variogram; the dashed line is our nonparametric estimate; the dotted line is the estimate assuming Model 3; the dot-dashed line is the estimate assuming Model 4. The fits assuming Models 1 and 2 are not included, since they are almost identical to our nonparametric estimate.

is based on $\sigma_\epsilon^2 = 0{\cdot}16$ in the regular-grid case, and $\nu = 10$ is used in the nonparametric estimation. The nonparametric estimate of the variogram spectrum and its associated variogram are shown to be quite close to the true functions, and $\hat{\sigma}_\epsilon^2 = 0{\cdot}149$, which is close to the true value of 0·16. The fitted variograms assuming Models 1 and 2, along with our nonparametric variogram estimate, are all close to the true variogram, and they are hard to tell apart. The parametric fitting for Model 1 gives parameter estimates, $\hat{a} = 0{\cdot}926$, $\hat{c}_0 = 0{\cdot}948$ and $\hat{\sigma}_\epsilon^2 = 0{\cdot}151$, which are close to their true values, $a = 1$, $c_0 = 1$, $\sigma_\varepsilon^2 = 0{\cdot}16$. However, the two fitted variograms assuming Models 3 and 4

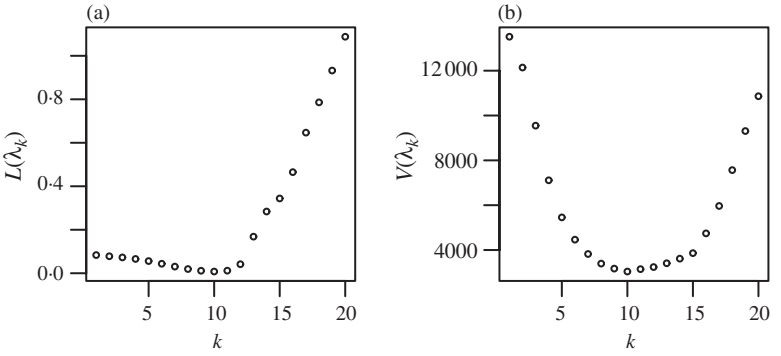

Fig. 4. Smoothing parameter selection plots. Panel (a) is $L(\lambda_k)$ versus k; panel (b) is $V(\lambda_k)$ versus k.

deviate substantially from the true variogram. These results are in line with the results in Table 2 from our complete simulation study.

We also compare the estimated variogram spectra in Fig. 3(a). The estimated spectra from the parametric approach are obtained by substituting the estimated variogram parameters into the variogram-spectrum formulae given in Table 1. The true spectrum is estimated very well by both the parametric approach assuming Model 1 and by the nonparametric procedure, but not so well by the parametric approach assuming Model 2. The results are much worse for the parametric approach assuming Models 3 and 4 and are not included in this figure since they distort the scale.

For smoothing-parameter selection, we chose $\lambda_k = 10^{u_k}$, where $u_k = 6(k-1)/19$ $(k = 1, \ldots, 20)$, and we computed both $V(\lambda_k)$ given in (15) and the true loss $L(\lambda_k)$, namely,

$$L(\lambda_k) = \frac{\nu}{L} \sum_{l=1}^{L} \{\hat{f}(\omega_l, \lambda_k) - f(\omega_l)\}^2,$$

where $\hat{f}(\cdot, \lambda_k)$ is our nonparametric estimate of the variogram spectrum when the smoothing parameter is λ_k. It is not possible to compute $L(\lambda_k)$ in practice, since the true f is not available. These two criteria plotted against k are shown in Fig. 4 for one typical simulation run. Both $V(\cdot)$ and $L(\cdot)$ reach similar minimum points, and $V(\cdot)$ provides a satisfactory choice of smoothing parameter. When 100 replications in this setting are drawn, the mean and standard error of the optimal $\log(\lambda)$ chosen by $V(\lambda)$ are 3·21 and 0·06, compared with 2·84 and 0·04 when $L(\lambda)$ is utilized. This indicates that minimizing λ according to the criterion $V(\lambda)$ approximates what one would obtain by minimizing the true loss.

Expanding the comparisons made under Model 1, we also examined the scenarios where the true models are Models 2–4. The results are shown in Table 3. When the process is simulated from Model 2, the mean of the ISE(2γ) values is the smallest for the parametric estimate based on assuming Model 2 is the true model; however, our nonparametric estimate outperforms all parametric estimates based on incorrect models. When the process is simulated from Model 3, our nonparametric estimate outperforms all parametric estimates, including that from the correct model. When the process is simulated from Model 4, the mean of the ISE(2γ) values is the smallest for the parametric estimate assuming Model 4 is the true model; however, our nonparametric estimate outperforms all parametric estimates assuming incorrect models. Since

Table 3. *Simulation results based on true models, Models 2–4*

		Model 2	Model 3	Model 4
Model 1	ISE(2γ)	19·86 (2·89)	96·73 (4·94)	230·6 (36·85)
	$\hat{\sigma}_\epsilon^2$	6·54 (0·70)	0·38 (0·38)	94·10 (1·05)
Model 2	ISE(2γ)	18·05 (2·88)	55·90 (4·44)	221·7 (36·85)
	$\hat{\sigma}_\epsilon^2$	16·01 (0·93)	3·14 (0·16)	43·06 (2·31)
Model 3	ISE(2γ)	43·56 (3·24)	38·61 (8·40)	261·4 (37·38)
	$\hat{\sigma}_\epsilon^2$	58·44 (1·10)	19·33 (1·76)	118·6 (0·47)
Model 4	ISE(2γ)	73·13 (3·63)	214·8 (7·26)	199·8 (32·34)
	$\hat{\sigma}_\epsilon^2$	70·93 (1·09)	114·4 (1·00)	23·44 (2·14)
NP	ISE(2γ)	18·79 (2·89)	29·95 (4·56)	221·5 (36·63)
	$\hat{\sigma}_\epsilon^2$	22·23 (0·14)	7·84 (0·26)	63·35 (0·21)
	ISE(f)	1·03 (0·08)	6·18 (0·56)	3·68 (0·44)

The rows of Models 1–4 contain the results of parametric fitting based on those models, and the row NP shows the results from our proposed nonparametric procedure. All simulations were conducted in the regular-grid case with $\sigma_\epsilon^2 = 0{\cdot}16$, and $\nu = 6$ in our nonparametric procedure. The entries are in units of 10^{-2}.

the variogram spectrum of Model 3 has a pole at 2, the values of ISE(f) are computed on the interval $[0, 1{\cdot}8]$.

4. Decadal temperature change over the Americas

To illustrate our methodology, we consider a temperature dataset obtained from the Climate System Model at the National Center for Atmospheric Research. The complete dataset consists of yearly averages of 2-metre air temperature for the period of 1980–1999, over the whole globe on 128×64 equi-angular longitude-latitude grid cells. Each cell is roughly $2{\cdot}8^o$ in longitude by $2{\cdot}8^o$ in latitude. Parts of this dataset, have been analysed by Shen et al. (2002), Zhang et al. (2008) and Cressie & Wikle (2011, Ch. 4), who used false discovery rates, spatial exceedances and kriging, respectively, to look for evidence of temperature change. We also focus on a reduced spatial domain containing most of North and South America, where the domain is from 40°N to 40°S and 125°W to 45°W. We are interested in the variogram spectrum of temperature change between the 1980s and 1990s in this region. Consequently, for each grid cell, we calculate the pixelwise average monthly temperatures in the 1980s and subtract them from their respective pixelwise average monthly temperatures in the 1990s. A map of these differences is given in Cressie & Wikle (2011, Figure 4.5a).

All observations are on a regular grid. We apply the methodology given in § 2 and show the results in Fig. 5. Figure 5(a) is the nonparametric estimate of the variogram spectrum, and Fig. 5(b) shows, amongst other things, the nonparametric estimate of the variogram. Since the temperature field is actually computer output from the Climate System Model, it is smoother than one might encounter from measured data, and hence we expect σ_ε^2 to be small or zero. We see this in Fig. 5, and it is consistent with the findings of Zhang et al. (2008).

It can be seen from Fig. 5(b) that our nonparametric estimate of the variogram fits the method-of-moments variogram estimate very well. Because the Matérn model, namely Model 2, is very close to Model 1, it is not shown in the figure. Our nonparametric estimate fits the method-of-moments variogram estimate better than either Model 1 or Model 4, especially for small lags. The

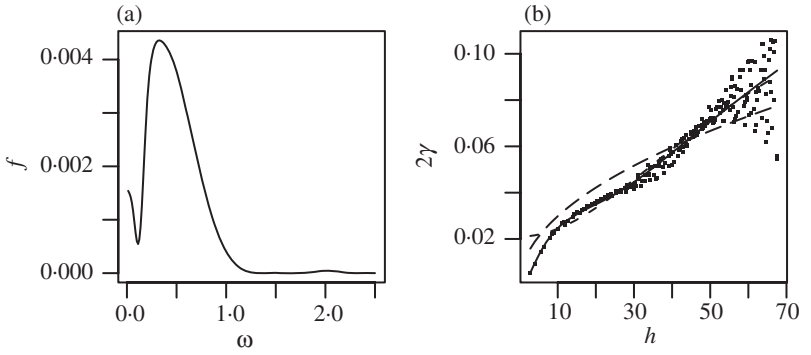

Fig. 5. Decadal temperature change data over the Americas. Panel (a) shows the nonparametric estimate of the variogram spectrum. Panel (b) shows three variogram estimates. The solid line is our nonparametric variogram estimate; the dashed line is the fitted variogram estimate of Model 1 with $\hat{a} = 49{\cdot}53$, $\hat{c}_0 = 8{\cdot}42$ and $\hat{\sigma}^2_\epsilon = 0{\cdot}021$; and the long-dashed line is the fitted variogram estimate of Model 4 with $\hat{a} = 0{\cdot}0049$ and $\hat{\sigma}^2_\epsilon = 0{\cdot}000$. Superimposed on (b) is the individual method-of-moments variogram estimates $\{x_{i,j}\}$ in solid dots, plotted against their spatial lag distances $\{h_{i,j}\}$.

variogram model Model 3 is not considered here, since the data do not support the wave shape evident in its variogram. Our nonparametric variogram estimate could be used to characterize the spatial pattern in the decadal differences and compared with estimates from other decadal differences to quantify how the temperature field is changing over time.

ACKNOWLEDGEMENT

C. H. and T. H.'s research was supported by the National Science Foundation. N. C.'s research was partially supported by the Office of Naval Research. We would like to express our sincere appreciation to a referee whose comments helped us to improve the paper substantially. The kind assistance of the editor is also gratefully acknowledged.

REFERENCES

ABRAMOWITZ, M. & STEGUN, I. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications.
CHILÉS, J. & DELFINER, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley.
COX, D. D. (1988). Approximation of method of regularization estimators. *Ann. Statist.* **16**, 694–712.
CRESSIE, N. (1985). Fitting variogram models by weighted least squares. *J. Int. Assoc. Math. Geol.* **17**, 563–86.
CRESSIE, N. (1993). *Statistics for Spatial Data*, rev. ed. New York: Wiley.
CRESSIE, N. & WIKLE, C. (2011). *Statistics for Spatio-Temporal Data*. Hoboken: Wiley.
DE BOOR, C. (1978). *A Practical Guide to Splines*. New York: Springer.
FUENTES, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89**, 197–210.
FUENTES, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Am. Statist. Assoc.* **102**, 321–31.
GILL, P. E., MURRAY, W. & WRIGHT, M. H. (1981). *Practical Optimization*. London: Academic Press.
GRADSHTEYN, I. M. & RYZHIK, I. S. (2007). *Table of Integrals, Series, and Products*, 7th ed. London: Academic Press.
GREEN, P. J. & SILVERMAN, S. W. (1994). *Nonparametric Regression and Generalized Linear Models*. New York: Chapman & Hall.
HALL, P., FISHER, N. I. & HOFFMANN, B. (1994). On the nonparametric estimation of covariance functions. *Ann. Statist.* **22**, 2115–34.
HUANG, C., HSING, T. & CRESSIE, N. (2011). Spectral density estimation through a regularized-inverse problem. *Statist. Sinica* **21**, 1115–44.

Im, H. K., Stein, M. L. & Zhu, Z. (2007). Semiparametric estimation of spectral density with irregular observations. *J. Am. Statist. Assoc.* **102**, 726–35.
Journel, A. G. & Huijbregts, C. J. (1978). *Mining Geostatistics*. London: Academic Press.
Lukas, M. A. (1988). Convergence rates for regularized solutions. *Math. Comp.* **51**, 101–37.
Matheron, G. (1963). Principles of geostatistics. *Econ. Geol.* **58**, 1246–63.
Matheron, G. (1973). The intrinsic random functions and their applications. *Adv. Appl. Prob.* **5**, 439–68.
Matsuda, Y. & Yajima, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. R. Statist. Soc.* B **71**, 191–217.
Nychka, D. & Cox, D. D. (1989). Convergence rates for regularized solutions of integral equalities from discrete noisy data. *Ann. Statist.* **17**, 556–72.
Nychka, D., Wahba, G., Goldfarb, S. & Pugh, T. (1984). Cross-validated spline methods for the estimation of three-dimensional tumor size distribution from observations on two-dimensional cross sections. *J. Am. Statist. Assoc.* **79**, 832–46.
Omre, H. (1984). The variogram and its estimation. In *Geostatistics for Natural Resources Characterization, Part I*, Ed. G. Verly, M. David, A. Journel and A. Marechal, pp. 107–125. Dordrecht: Reidel.
O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statist. Sci.* **1**, 502–27.
Powojowski, M. R. (2008). Isotropic spectral additive models of the covariogram. *J. R. Statist. Soc.* B **70**, 739–53.
Press, W. H., Teukolsky, S., Vetterling, W. T. & Flannery, B. (2007). *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge: Cambridge University Press.
Pyrcz, M. J. & Deutsch, C. V. (2006). Spectral corrected semivariogram models. *Math. Geol.* **38**, 891–9.
Sampson, P. & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Assoc.* **87**, 108–19.
Shapiro, A. & Botha, J. D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Comp. Statist. Data Anal.* **11**, 87–96.
Shen, X., Huang, H. C. & Cressie, N. (2002). Nonparametric hypothesis testing for a spatial signal. *J. Am. Statist. Assoc.* **97**, 1122–40.
Stein, M. (1999). *Interpolation of Spatial Data—Some Theory for Kriging*. New York: Springer.
Villalobos, M. & Wahba, G. (1987). Inequality constrained multivariate smoothing splines with application to the estimation of posterior probabilities. *J. Am. Statist. Assoc.* **82**, 239–48.
Wahba, G. (1973). Convergence rates of certain approximation solution to Fredholm integral equations of the first kind. *SIAM J. Contr.* **11**, 64–79.
Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia, PA: SIAM.
Wang, Y. (1998). Smoothing spline models with correlated random errors. *J. Am. Statist. Assoc.* **93**, 341–8.
Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions*. New York: Springer.
Yu, K., Mateu, J. & Porcu, E. (2007). A kernel-based method for nonparametric estimation of variogram. *Statist. Neer.* **61**, 173–97.
Zhang, J., Craigmile, P. & Cressie, N. (2008). Loss function approaches to predict a spatial quantile and its exceedance region. *Technometrics* **50**, 216–27.

[*Received July* 2009. *Revised July* 2011]