

Nonparametric Estimation of Variogram and its Spectrum

Chunfeng Huang^a, Tailen Hsing^b, and Noel Cressie^c

October 30, 2008

Technical Report 08-05

Department of Statistics

Indiana University

Bloomington, IN 47405

^aDepartments of Statistics, Indiana University. huang48@indiana.edu

^bDepartment of Statistics, University of Michigan. thsing@umich.edu

^cDepartments of Statistics, The Ohio State University. ncressie@stat.osu.edu

Nonparametric Estimation of Variogram and its Spectrum

by

Chunfeng Huang, Tailen Hsing, and Noel Cressie

October 30, 2008

Abstract. In the study of isotropic intrinsically stationary spatial processes, a new nonparametric variogram estimator is proposed through its spectral representation. The spectrum estimation is formulated in terms of solving a regularized inverse problem. A numerical implementation is presented through quadratic programming. We demonstrate our method in a simulation study and a dataset of temperature changes over America.

1. Introduction.

Modeling of statistical dependence is often achieved through the covariance. In the spatial setting, where $\{Y(\mathbf{s}): \mathbf{s} \in D \subset \mathbb{R}^d\}$ is a real-valued stochastic process defined on a subset D of Euclidean space where $\text{var}(Y(\mathbf{s})) < \infty$ for all $\mathbf{s} \in D$, the *covariance function* is defined as,

$$K(\mathbf{s}, \mathbf{t}) \equiv \text{cov}(Y(\mathbf{s}), Y(\mathbf{t})); \quad \mathbf{s}, \mathbf{t} \in D.$$

Estimation of K from a single, incomplete sample from Y can be difficult without further assumptions. The *second-order stationarity assumption* is commonly made, namely,

$$\begin{aligned} E(Y(\mathbf{s})) &\equiv \mu; & \mathbf{s} \in D \\ \text{cov}(Y(\mathbf{s}), Y(\mathbf{t})) &\equiv C(\mathbf{s} - \mathbf{t}); & \mathbf{s}, \mathbf{t} \in D, \end{aligned}$$

where μ and C are unknown.

Let $\{Y(\mathbf{s}_i): i = 1, \dots, n\}$ denote an incomplete sample from Y . An often-used method-of-moments estimator of C is:

$$\widehat{C}(\mathbf{h}) \equiv \text{ave}\{(Y(\mathbf{s}_i) - \bar{Y})(Y(\mathbf{s}_j) - \bar{Y}): \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\},$$

where $\text{ave}\{\cdot\}$ denotes the average over all terms in the set $\{\cdot\}$ and $\bar{Y} \equiv \text{ave}\{Y(\mathbf{s}_i)\}$. While \widehat{C} is a reasonable estimator, it does have some unsatisfactory properties: Think of $C(\cdot)$ as a parameter to be estimated; then the parameter space is the class of all positive-definite functions (e.g., Yaglom, 1987, page 58.). Unfortunately, $\widehat{C}(\cdot)$ is not guaranteed to belong to the parameter space. Moreover, the residuals $\{Y(\mathbf{s}_i) - \bar{Y}: i = 1, \dots, n\}$ sum to 0, implying a negative covariance between them even when the data are independent; hence, $\widehat{C}(\mathbf{h})$ is generally negatively biased.

Another measure of spatial statistical dependence is the *variogram*,

$$2G(\mathbf{s}, \mathbf{t}) \equiv \text{var}(Y(\mathbf{s}) - Y(\mathbf{t})); \quad \mathbf{s}, \mathbf{t} \in D.$$

The variogram is easily related to the covariance function as follows:

$$2G(\mathbf{s}, \mathbf{t}) = K(\mathbf{s}, \mathbf{s}) + K(\mathbf{t}, \mathbf{t}) - 2K(\mathbf{s}, \mathbf{t}).$$

Similar to estimation of K , estimation of G from a single, incomplete sample from Y can be difficult without further assumptions. The *intrinsic-stationarity* assumption is commonly made, namely,

$$\begin{aligned} E(Y(\mathbf{s})) &\equiv \mu; & \mathbf{s} \in D, \\ \text{var}(Y(\mathbf{s}) - Y(\mathbf{t})) &\equiv 2\gamma(\mathbf{s} - \mathbf{t}); & \mathbf{s}, \mathbf{t} \in D, \end{aligned}$$

where μ and 2γ are unknown. The variogram $2\gamma(\cdot)$ has achieved prominence because of its role in kriging (i.e., optimal spatial prediction); see Matheron (1963). Note that the covariance function $C(\cdot)$ has an equivalent role in kriging (Cressie, 1993, p. 123), although $C(\cdot)$ may not exist when $2\gamma(\cdot)$ does.

A method-of-moments estimator of $2\gamma(\cdot)$ from the sample $\{Y(\mathbf{s}_i): i = 1, \dots, n\}$, is

$$2\hat{\gamma}(\mathbf{h}) \equiv \text{ave}\{(Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2: \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}.$$

Notice that the estimator does not involve \bar{Y} nor the residuals, and hence is unbiased. Cressie (1993), Sect. 2.4.1 compares estimation of these two measures of spatial dependence, C and 2γ , and demonstrates that 2γ is a desirable parameter to estimate. As mentioned above, it exists for processes Y for which C does not exist, and it is central to optimal spatial prediction (e.g., Cressie, 1993; Chilés and Delfiner, 1999; and Stein, 1999).

One crucial property of the variogram is conditional negative-definiteness, namely,

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0, \tag{1}$$

for any finite number of spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_m$ and real numbers a_1, \dots, a_m satisfying $\sum_{i=1}^m a_i = 0$. Armstrong and Jabin (1981) present examples to show that a negative variance can happen when the variogram is not valid. We write the method-of-moments estimator $2\hat{\gamma}(\mathbf{h})$ explicitly as:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2,$$

where $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j): \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$ and $|N(\mathbf{h})|$ is the number of pairs in $N(\mathbf{h})$. This method can be extended by tolerance region method (Omre, 1984, Journel and Huijbregts, 1978). A kernel-based nonparametric estimation of variogram can be found in Yu, Mateu and Porcu (2007). However, these methods do not employ the consideration of (1) and may produce the estimates that are not conditional negative definite. To ensure this property, many parametric models are

developed and have been widely utilized in data analysis (Cressie, 1993). The choices of such models are limited, and in practice it is not always clear which model to use for a given situation. When there are temporal information available, Sampson and Guttorp (1992) developed an estimator through multidimensional scaling.

In this article, we develop a new method to estimate a valid nonparametric variogram when it is isotropic. A variogram is isotropic when $2\gamma(\mathbf{h}) = 2\gamma^0(\|\mathbf{h}\|)$. While isotropy is a rather specialized property, its simplicity makes it a popular model in practice. Often, in anisotropic situations, a simple transformation of coordinates, $2\gamma(\mathbf{h}) = 2\gamma^0(A\mathbf{h})$, where A is a $d \times d$ non-singular matrix, can convert the model to isotropy. For notation convenience, $2\gamma^0(\cdot)$ will also be referred to the variogram and will be denoted simply as 2γ . A spectral representation of a variogram for $d = 2$ is (cf. Yaglom, 19867 section 25.3)

$$2\gamma(h) = \int_0^\infty (1 - J_0(\omega h)) dF(\omega), \quad h \geq 0, \quad (2)$$

where $J_0(\cdot)$ is a Bessel function of the first kind of order 0, and F is a non-decreasing function on $(0, \infty)$ such that $\int_0^\infty \omega^2/(1 + \omega^2) dF(\omega) < \infty$. We will further assume that F has a derivative f , which is called spectrum. The spectral representation (2) is to be compared to Bochner's Theorem for the covariance function $C(\cdot)$.

Spectrum is less explored in spatial statistics as its counterpart in time series, although there have been articles that have recognized its utility for estimation of the covariance function C through Bochner's Theorem (Hall, Fisher and Hoffmann, 1994; Shapiro and Botha, 1991; Fuentes, 2002, 2003; Im, et. al., 2007). Such method depends on the inverse of Fourier transformation and appears not applicable for (2). In this article, we convert the problem to a form of ill-posed integral equation, and employ general spline approach to estimate the spectrum. With the constraints of nonnegativeness, the resulting variogram estimate is shown to be conditional negative definite. This is presented in Section 2. A simulation study to demonstrate our method is given in Section 3 with a data analysis for temperature data in Section 4. Some discussions are given in Section 5.

2. Estimation Methodology. Let Y be an intrinsically stationary process with a constant mean μ , an isotropic variogram $2\gamma(\cdot)$, and a spectrum $f(\omega) > 0, \omega \in (0, \infty)$. Let

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

where Z is a white noise measurement error process that is independent of process Y with $\text{Var}(Z(\mathbf{t})) = \sigma^2$. Assume that the data are observed at locations $\mathbf{t}_1, \dots, \mathbf{t}_N$. In Section 2.1, we will first discuss a general approach, while a numerical approach will be given in Section 2.2, with a theoretical justification of the approach given in Section 2.3. Smoothing parameter selection is discussed in Section 2.4.

2.1. General Spline Estimation. Since X has a constant mean, we have

$$E[X(\mathbf{t}_i) - X(\mathbf{t}_j)]^2 = 2\gamma(\|\mathbf{t}_i - \mathbf{t}_j\|) + 2\sigma^2.$$

That is, $x_{i,j} := [X(\mathbf{t}_i) - X(\mathbf{t}_j)]^2$ is an unbiased estimator of

$$2\gamma(\|\mathbf{t}_i - \mathbf{t}_j\|) + 2\sigma^2 = 2 \int_0^\infty (1 - J_0(\omega\|\mathbf{t}_i - \mathbf{t}_j\|))f(\omega)d\omega + 2\sigma^2.$$

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

$$\sum_{i \neq j}^N \left[x_{i,j} - c - \int_0^\infty (1 - J_0(\omega\|\mathbf{t}_i - \mathbf{t}_j\|))g(\omega)d\omega \right]^2. \quad (3)$$

Conversely, any nonnegative function g and real number c that make this sum small can be considered as candidate estimators of f and $2\sigma^2$, respectively. However, searching for estimators in this manner constitutes an ill-posed inverse problem (cf. O’Sullivan, 1986) for which regularization is essential. To obtain a regularized solution, one possible approach is to add a roughness penalty term of g to the sum of squares in (3).

Before we proceed, it is worth noting 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. 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 on hand. The second consideration is more fundamental and can be most easily understood from the context of stationary process. If the observations of a stationary process are made on a grid, then the spectrum is not identifiable in that there isn’t 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 Yaglom (1987), page 187. The aliasing effect is more difficult to pin down theoretically in the case of intrinsically stationary process, but we have observed it in our computational work. As a compromise, we will choose a large enough ν in the estimation algorithm so that little generality will be lost.

In this setting, an effective way to address the penalized least squares problem is through the general spline smoothing approach of Wahba (cf. Wahba, 1990). Consider the problem with constraints to estimate f and $2\sigma^2$ through the solution of

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

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

$$L_{h_{i,j}}g = \int_0^\nu (1 - J_0(\omega h_{i,j}))g(\omega)d\omega, \quad h_{i,j} = \|\mathbf{t}_i - \mathbf{t}_j\|,$$

$W_m[0, \nu]$ is the Sobolev space of order m of functions on $[0, \nu]$, i.e., the space of functions on $[0, \nu]$ that are m -times differentiable with square integrable m -th derivatives, and

$$J(g) = \int_0^\nu [g^{(m)}(\omega)]^2 d\omega.$$

The Sobolev space $W_m[0, \nu]$ has the structure $\mathcal{H}_0 \oplus \mathcal{H}_1$, where \mathcal{H}_0 is a m -dimensional space of polynomials of degree $m - 1$ spanned by $\phi_j = x^{j-1}, j = 1, \dots, m$, and \mathcal{H}_1 is the collection of functions on $[0, \nu]$ such that $g^{(\tau)}(0) = 0, \tau = 0, 1, \dots, m - 1$ and $\int_0^\nu [g^{(m)}]^2 < \infty$. See Chapter 1 of Wahba (1990). It can be 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 $R_t(\cdot) = R(t, \cdot)$. The constraint $g \geq 0$ can be approximated by $g(\omega_1) \geq 0, \dots, g(\omega_L) \geq 0$ for a set of densely distributed ω_l 's. Without loss of generality we take the ω_l 's to be equally spaced in $[0, \nu]$. By the reproducing property, these constraints can be expressed as

$$\langle g, R_{\omega_l} \rangle \geq 0, 1 \leq l \leq L, \quad (5)$$

where $\langle \cdot, \cdot \rangle$ is the inner product of $W_m[0, \nu]$. Let $\eta_{i,j}$ be the representer of $L_{h_{i,j}}$, namely, $\langle \eta_{i,j}, f \rangle = L_{h_{i,j}}f, f \in \mathcal{H}$; also let $\xi_{i,j} = \mathcal{P}\eta_{i,j}$, and $\rho_j = \mathcal{P}R_{\omega_j}$, where \mathcal{P} is the projection operator onto \mathcal{H}_1 . By Section 9.4 of Wahba (1990), replacing the constraint $g \geq 0$ in (4) by (5) leads to the unique minimizer

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

where 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. The closed-form expressions of $\xi_{i,j}, L_{i,j}\phi_\tau$ can be obtained manually or by Maple and are shown in Appendix A. However, we are unable to find a closed-form expression for $\langle \xi_{i,j}, \xi_{i',j'} \rangle$, and some form of approximation will be necessary. A method of matched quadrature in Nychka, et. al. (1984) is a possibility, but, unfortunately, the complexity of the expressions of $\xi_{i,j}, \phi_\tau$ complicates the implementation of that method. In the next subsection we present a simple and intuitive numerical solution.

2.2. The Numerical Method. Let

$$\boldsymbol{\ell}_{i,j} = \frac{\nu}{L} \left(1 - J_0(\omega_1 h_{i,j}), \dots, 1 - J_0(\omega_L h_{i,j}) \right) \quad \text{and} \quad \mathbf{g} = (g(\omega_1), \dots, g(\omega_L))^T,$$

where the ω_l are as in (5). For a sufficiently dense set of ω_l , $L_{i,j}g$ is well approximated by the Riemann sum $\ell_{i,j}\mathbf{g}$. If $L_{i,j}g$ is replaced by $\ell_{i,j}\mathbf{g}$ and $g \geq 0$ by $\mathbf{g} \geq \mathbf{0}$ in (4), then the solution \hat{g} must be a natural spline with knots $\omega_1, \dots, \omega_L$, and the roughness penalty $J(g)$ can be expressed as $\mathbf{g}^T K \mathbf{g}$ for some $L \times L$ matrix K (cf. Green and Silverman, 1994). Specifically, \hat{g} is the unique natural spline that interpolates $(\omega_j, \hat{g}_l), 1 \leq l \leq L$, where $\hat{\mathbf{g}} = (\hat{g}_1, \dots, \hat{g}_L)$ is the minimizer of

$$\min_{c \geq 0, \mathbf{g} \in \mathbb{R}^L, \mathbf{g} \geq \mathbf{0}} \left\{ \sum_{i \neq j}^N (x_{i,j} - c - \ell_{i,j}\mathbf{g})^2 + \lambda \mathbf{g}^T K \mathbf{g} \right\}. \quad (6)$$

For convenience of notation, let $n = N(N-1)/2$ and vectorize $x_{i,j}, \ell_{i,j}$ as x_k, ℓ_k ; also, let $\mathbf{x} = (x_1, \dots, x_n)^T$,

$$\mathbf{B} = \begin{pmatrix} 1 & \ell_1 \\ \vdots & \vdots \\ 1 & \ell_n \end{pmatrix} \quad \text{and} \quad \mathbf{\Psi} = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{K} \end{pmatrix}.$$

Then, (6) can be expressed as

$$\min_{\mathbf{v} \in \mathbb{R}^{L+1}} \{ \mathbf{v}^T (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{\Psi}) \mathbf{v} - 2 \mathbf{v}^T \mathbf{B}^T \mathbf{x} \} \quad \text{subject to} \quad \mathbf{v} \geq \mathbf{0}, \quad (7)$$

which can be solved numerically by quadratic programming (QP) in a standard way. Let the minimizing vector of (7) be $\hat{\mathbf{v}} = (\hat{v}_1, \dots, \hat{v}_L, \hat{v}_{L+1})^T$. Then, the measurement error variance estimate is $2\hat{\sigma}^2 = \hat{v}_{L+1}$. The values of the spectrum at the knots can be estimated as $\hat{f}(\omega_1) = \hat{v}_1, \dots, \hat{f}(\omega_L) = \hat{v}_L$, and the entire function of f is then estimated by the unique natural spline that interpolates $(\omega_l, \hat{f}(\omega_l)), 1 \leq l \leq L$. Assuming that L is large enough and the ω_l are evenly distributed in $[0, \nu]$, it follows intuitively that $\hat{f}(\omega) \geq 0$ for all ω . In our numerical work, we have not encountered a single example for which this is not the case.

To estimate the variogram, we may apply (2) and let

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

Note that $\hat{f}(\omega)$ is a degree $2m-1$ polynomial on each interval $[\omega_i, \omega_{i+1}), i = 1, \dots, L-1$ and is linear outside $[\omega_1, \omega_L]$. The coefficients of each polynomial are determined by $\hat{f}(\omega_1), \dots, \hat{f}(\omega_L)$ (cf. de boor, 1978, Green and Silverman, 1994). Therefore, within each interval, the computation of $2\hat{\gamma}(h)$ can be carried out through the integration of the combinations of Bessel function and polynomials (cf. Appendix).

Alternatively, the variogram can be estimated by the Riemann sum:

$$2\hat{\gamma}(h) = \frac{\nu}{L} \sum_{l=1}^L (1 - J_0(\omega_l h)) \hat{f}(\omega_l). \quad (8)$$

This approach has the clearly advantage of computing ease, and, more importantly, produces a legitimate, i.e., conditionally negative definite, variogram.

Note that the dimension of \mathbf{B} is $N(N-1)/2 \times (L+1)$. This optimization in (6) can be carried out efficiently when the sample size N is small. When N is large, the size of the matrix \mathbf{B} could overwhelm the computations. When the observations are on a grid, each distance $h_{i,j} = \|\mathbf{t}_i - \mathbf{t}_j\|$ will be duplicated a large number of times. For example, suppose data are observed at every point of an $N_0 \times N_0$ grid with grid spacing h . Then there are $2N_0(N_0 - 1)$ pairs of $X(\mathbf{t}_i), X(\mathbf{t}_j)$ having the same distance h , and we can define

$$w_1 = 2N_0(N_0 - 1), \quad y_1 = \frac{1}{w_1} \sum_{h_{i,j}=h} x_{i,j}. \quad (9)$$

Going through all possible distances between grid points in this manner, we obtain a sequence of (y_i, h_i, w_i) , and the double summation in (4) reduces to a weighted single summation

$$\min_{c \geq 0, \mathbf{g} \in \mathbb{R}^L, \mathbf{g} \geq 0} \left\{ \sum_{i=1}^{n_0} w_i (y_i - c - \boldsymbol{\ell}_i \mathbf{g})^2 + \lambda \mathbf{g}^T \mathbf{K} \mathbf{g} \right\}, \quad (10)$$

where $n_0 = N_0(N_0 + 1)/2$, and

$$\boldsymbol{\ell}_i = \frac{\nu}{L} (1 - J_0(\omega_1 h_i), \dots, 1 - J_0(\omega_L h_i)).$$

Let $\mathbf{y} = (y_1, \dots, y_{n_0})^T$ and $\mathbf{W} = \text{diag}(w_1, \dots, w_{n_0})$. Then (10) can be expressed as

$$\min_{\mathbf{v} \in \mathbb{R}^{L+1}} \{ \mathbf{v}^T (\mathbf{B}^T \mathbf{W} \mathbf{B} + \lambda \boldsymbol{\Psi}) \mathbf{v} - 2 \mathbf{y}^T \mathbf{W} \mathbf{B} \mathbf{v} \} \quad \text{subject to} \quad \mathbf{v} \geq 0. \quad (11)$$

Thus, the computation task is significantly reduced.

It is worth noting that the y_i 's defined by (9) are method-of-moments variogram estimates (Matheron, 1963). As this method extends to irregularly spaced data by the tolerance region method (Omre, 1984, Journel and Huijbregts, 1978), similar ideas can be employed to extend (11) to the case where observations are not on a grid. For example, one can round the distance $h_{i,j}$ up to a given number of significant digits, and (4) can be approximated by a weighted version similar to (11).

As for the choice of L, ν and λ , it is clear that one should choose the largest possible L that can be handled by the computing device that one is using. It can be seen from our simulation studies in Section 4 that the choice of ν does not have much impact of the estimation of the spectrum and the variogram. However, the smoothing parameter λ is important, and will be discussed in detail in Section 2.4.

2.3. A Theoretical Considerations. The theoretical issue of rate of convergence, or even consistency, of our procedure is a difficult one, which we are not ready to address 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 (1989) and Nychka and Cox (1989), to name a few. They all highlight the difficulty of a practical solution of this problem. The generic problem is estimating f on $[0, 1]$ based on the data (z_i, t_i) where

$$z_i = \int_0^1 K(t_i, t)f(t)dt + \varepsilon_i,$$

and some known function K and “error” ε_i with mean zero. In the context of our problem,

$$K(u, t) = 1 - J_0(ut) \tag{12}$$

and the z_i and t_i are the $x_{i,j}$ and $h_{i,j}$, respectively. Assume that $f \in W_m[0, 1]$ and let R_1 be the reproducing kernel of \mathcal{H}_1 . Then 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

$$\mathcal{Q} : f \mapsto \int_0^1 KRK^*(s, t)f(t)dt,$$

where

$$KRK^*(s, t) := \int_0^1 \int_0^1 K(s, u)R_1(u, v)K(t, v)dudv.$$

Unfortunately, there is no systematic way to study the rate of decay of the eigenvalues. This is in particular true for our problem. As such, the next best thing is to compute the eigenvalues numerically as suggested by Nychka et al. (1984), pp. 843-844. Let us try this in our problem. Assume now that $m = 2$ so that

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

and we take K as in (12). A discrete version of KRK^* is the matrix $\mathbf{Q} = \{KRK^*(t_i, t_j)\}_{i,j=1}^n$ where the t_i 's are equally spaced. The eigenvalues of \mathbf{Q} approximate those of \mathcal{Q} . A log-log plot of eigenvalue of \mathbf{Q} versus its own index for $n = 1000$ is presented in Figure 1. This figure suggests that $\lambda_\nu = O(\nu^{-\beta})$ where β is in a region, namely, $(1/2, \infty)$, such that the regularized solution will be consistent given a proper choice of the smoothing parameter (cf. Lukas, 1986, Nychka and Cox, 1989, the slope is -5.14 in Figure 1).

2.4. Smoothing Parameter Selection. To implement our procedure in practice, a data-driven choice of smoothing parameter λ will be needed. A generalized cross validation (GCV) is proposed by Villalobos and Wahba (1987) for the minimization with linear constraints.

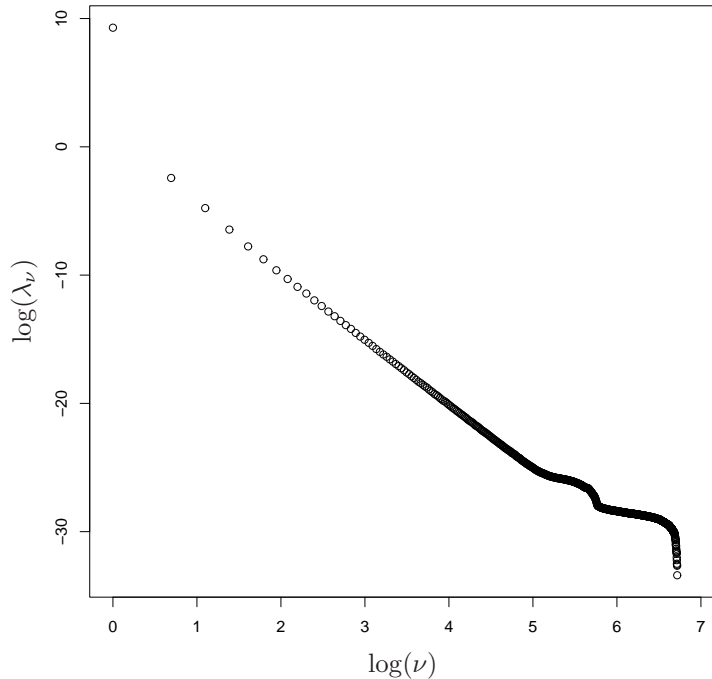


Figure 1: log-log plot of eigenvalue for $n = 1000$, the slope is -5.14 .

For each candidate smoothing parameter λ , solve the QP problem in (11), let the minimizing vector be $\hat{\mathbf{v}}$ and $\hat{\mathbf{y}}(\lambda) = \mathbf{B}\hat{\mathbf{v}}$. Let $\tilde{\mathbf{B}}, \tilde{\Psi}$ be the matrix \mathbf{B} and Ψ with rows and columns that correspond to the active constraints removed. An active constraint is one that causes the solution of the QP problem to equal 0. This can be done by examining the solution. Consider

$$\text{GCV}(\lambda) := \frac{\text{RSS}(\lambda)}{(1 - (1/n_0)\text{tr}A(\lambda))^2},$$

where $\text{RSS}(\lambda) = \sum_{i=1}^{n_0} w_i (y_i - \hat{y}_i)^2$, and

$$A(\lambda) = \tilde{\mathbf{B}}(\tilde{\mathbf{B}}^T \mathbf{W} \tilde{\mathbf{B}} + \lambda \tilde{\Psi})^{-1} \tilde{\mathbf{B}}^T \mathbf{W}.$$

Observations are assumed to be independent in this approach. Our data are highly correlated, and it is known that GCV is likely to fail in this situation (Wahba, 1990). For dependent data, several modified GCVs can be found in Wang (1998). One form of GCV (Wang, 1998, equation (15), $k = 0$) adopts:

$$\frac{\text{RSS}(\lambda)}{\{1 - (1/p_0)\text{tr}(\Xi^{-1}A(\lambda))\}^2}, \quad (13)$$

where Ξ is the covariance matrix of \mathbf{y} and $p_0 = \text{tr}\Xi^{-1}$. In this paper, the covariance matrix Ξ is unknown. We propose the following criterion:

$$V(\lambda) := \frac{\text{RSS}(\lambda)}{\{1 - (1/p)\text{tr}(\mathbf{W}A(\lambda))\}^2} \quad (14)$$

where $p = \lim_{\lambda \rightarrow 0} \text{tr}\{\mathbf{W}A(\lambda)\}$. Then, pick λ that minimizes $V(\lambda)$.

Remark : In Wang (1998), the covariance Ξ in (13) is assumed to have a parametric form, and the parameters are estimated through maximum likelihood. As the goal of our method is to estimate the variogram nonparametrically, this approach does not seem to be directly applicable in our situation. As a practical solution, we replace Ξ and p_0 with W^{-1} and p and propose $V(\lambda)$. This procedure is implemented in our simulation study and produces satisfactory results (see Figure 3 and discussion in Section 3).

3. Simulations. To demonstrate how our method works numerically, a simulation study is conducted. Two isotropic intrinsically stationary process on \mathbb{R}^2 with measurement error (variance is σ^2) are considered. They are

$$\begin{aligned} \text{Model I (MI):} \quad & \text{Spectrum} \quad f(\omega) = c_0 e^{-a\omega}, \quad 0 < \omega < \infty, \quad a > 0, \quad c_0 > 0 \\ & \text{Variogram} \quad 2\gamma(h) = c_0 \left(\frac{1}{a} - \frac{1}{\sqrt{a^2 + h^2}} \right) + 2\sigma^2, \end{aligned}$$

and

$$\begin{aligned} \text{Model II (MII):} \quad & \text{Spectrum} \quad f(\omega) = a\omega^{-3/2}, \quad 0 < \omega < \infty, \quad c_0 > 0 \\ & \text{Variogram} \quad 2\gamma(h) \doteq 1.911955a\sqrt{h} + 2\sigma^2. \end{aligned}$$

These two theoretical spectrums are chosen for stationary and intrinsically stationary processes respectively. The variogram in MI is bounded, therefore the process is actually stationary. The variogram in MII is unbounded from above, that is, the process is intrinsically stationary. The plots of these two models can be seen in Figure 2 and Figure 4. The processes are observed at a 60×60 grid with spacing to be 0.4.

When the process is from MI with parameters $a = 1, c_0 = 1$ and $\sigma^2 = .16$, a procedure according to last section is carried out. We estimate spectrum on $[0, 10]$, i.e., $\nu = 10$, and $L = 200$. Figure 2 shows both spectrum and variogram estimates along with the true functions. It is clear that our spectrum is close to the truth, so is the variogram estimate. The measurement error variance estimate performs also quite well ($\hat{\sigma}^2 = 0.157, \sigma^2 = 0.16$). Note that our procedure is through quadratic programming with $\hat{f}(\omega_l) \geq 0, l = 1, \dots, L$, therefore, the variogram estimate is always conditional negative definite.

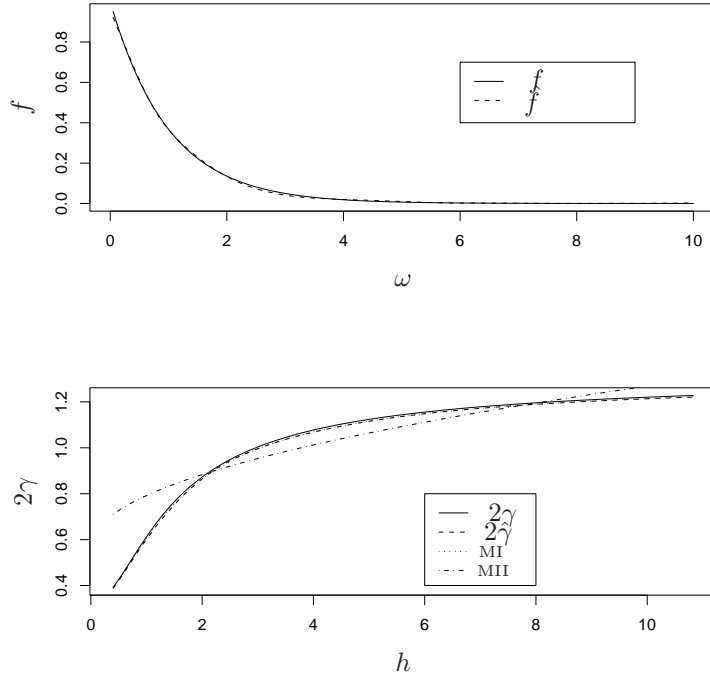


Figure 2: (Grid case) The upper plot spectrum estimation; the lower plot is variogram estimation. The true model is MI.

For the smoothing parameter selection, we let $\lambda_i = 10^{u_i}$, where $u_i = 6(i - 1)/19, i = 1, \dots, 20$, and compute both the true loss $L(\lambda_i)$ and $V(\lambda_i)$. These two criteria versus i are shown in Figure 3. It can be seen that $V(\cdot)$ mimics the shape of $L(\cdot)$ and provides a satisfactory choice of smoothing parameter. When 50 replications in this setting are conducted, $\log(\lambda)$ chosen by $V(\lambda)$ yields the mean (standard error) to be 3.07(0.10), compared to 2.79(0.11) when $L(\lambda)$ is utilized. This confirms that the criterion $V(\lambda)$ is an acceptable approach in this article.

In most applications, variogram estimations are computed parametrically. It is of interest to compare the parametric and nonparametric approaches. The parameters estimation is obtained through the weighted least square method (Cressie, 1993). We use both MI and MII as parametric model fitting. For MI, $\hat{a} = 1.033, \hat{c}_0 = 1.021, \hat{\sigma}^2 = 0.164$. For MII, $\hat{a} = 0.115, \hat{\sigma}^2 = 0.286$. Both fits are also shown in Figure 2. As MI is the true parametric model, the parameters estimates are close to their true values, and it outperforms MII. If one chooses a wrong parametric model, the estimate can be quite away from the true variogram, for example, in this case, MII is a wrong model.

In addition, 50 replications each for different settings are carried out in our study. The results

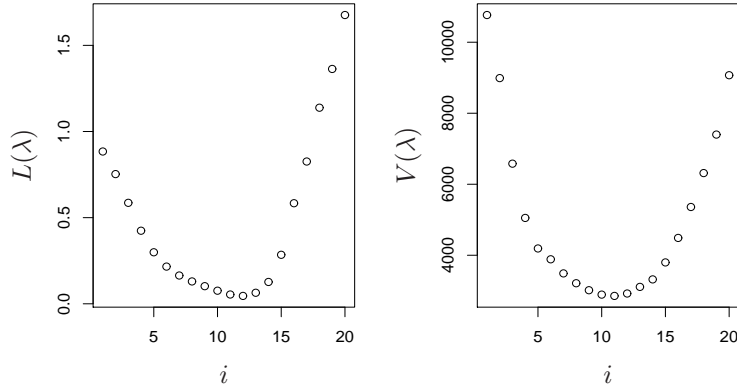


Figure 3: The left plot is $L(\lambda_i)$ versus i ; the right plot is $V(\lambda_i)$ versus i

are shown in Table 1. The mean integrated squared error (MISE, Yu, et. al. 2007) is used to assess spectrum and variogram:

$$\text{MISE}(f) = \int_0^\nu (\hat{f}(\omega) - f(\omega))^2 d\omega,$$

and

$$\text{MISE}(2\gamma) = \int_{h_l}^{h_u} (2\hat{\gamma}(h) - 2\gamma(h))^2 dh,$$

where h_l, h_u are smallest and largest lags for variogram estimates. Mean and standard error values are reported for each set of simulations. Parametric modeling results are also shown in the table, where $\text{MISE}(2\gamma, \text{MI})$ and $\text{MISE}(2\gamma, \text{MII})$ are for Model I and II, respectively. As the data are simulated from MI, the $\text{MISE}(2\gamma, \text{MI})$ always has the smallest value and should be the benchmark. Our estimate $\text{MISE}(2\gamma)$ consistently outperforms the $\text{MISE}(2\gamma, \text{MII})$, and is relatively close to the benchmark. The $\text{MISE}(f)$ appears to be relatively small. The table also shows the effect of choosing different ν . The $\text{MISE}(2\gamma)$ is quite robust against the choice of ν , while $\text{MISE}(f)$ changes a bit. Our measurement error variance estimates is much better than MII.

Model II is called power variogram, as the variogram does not level off as lag increases. Such process is intrinsic stationary, but not stationary. Similar simulation study is performed. One simulation result is shown in Figure 4. Our estimate $2\hat{\gamma}$ and MII (true model) are quite close, the estimate MI is also close, off a bit in the smaller lags. For all 50 simulations, we show the result in Table 1. Again, it appears that the choice of ν does not effect $\text{MISE}(2\gamma)$ much, the $\text{MISE}(f)$ changes a bit. In addition, as $\text{MISE}(2\gamma, \text{MII})$ becomes the benchmark here, our estimate $\text{MISE}(2\gamma)$ beats the wrongly chosen parametric model $\text{MISE}(2\gamma, \text{MI})$.

setting	MISE(f)	MISE(2γ)	MISE(2γ ,MI)	MISE(2γ ,MII)
$\sigma^2 = 0.16, \nu = 6$	0.367(0.059)	0.173(0.024)	0.144(0.024)	0.944(0.029)
$\hat{\sigma}^2$		0.158(0.001)	0.160(0.002)	0.298(0.004)
$\sigma^2 = 0.16, \nu = 10$	0.208(0.029)	0.171(0.024)	0.144(0.024)	0.944(0.029)
$\hat{\sigma}^2$		0.148(0.002)	0.160(0.002)	0.298(0.004)
$\sigma^2 = 0.36, \nu = 6$	0.380(0.056)	0.329(0.047)	0.302(0.046)	1.087(0.047)
$\hat{\sigma}^2$		0.356(0.002)	0.360(0.004)	0.502(0.004)
$\sigma^2 = 0.36, \nu = 10$	0.232(0.038)	0.329(0.046)	0.302(0.046)	1.087(0.047)
$\hat{\sigma}^2$		0.337(0.002)	0.360(0.004)	0.502(0.004)

Table 1: MI: Mean and standard error values of MISE values and $\hat{\sigma}^2$ values.

setting	MISE(f)	MISE(2γ ,M)	MISE(2γ ,MI)	MISE(2γ ,MII)
$\sigma^2 = 3, \nu = 6$	2.285(0.150)	10.314(1.468)	10.612(1.451)	9.264(1.454)
		3.446(0.013)	3.836(0.015)	3.007(0.019)
$\sigma^2 = 3, \nu = 10$	2.146(0.164)	10.319(1.466)	10.612(1.451)	9.264(1.451)
		3.151(0.025)	3.836(0.015)	3.007(0.019)
$\sigma^2 = 5, \nu = 6$	2.936(0.220)	19.026(2.674)	28.399(3.516)	17.445(2.643)
		5.419(0.020)	4.822(0.296)	5.009(0.016)
$\sigma^2 = 5, \nu = 10$	3.870(0.404)	19.114(2.678)	28.399(3.516)	17.445(2.643)
		5.004(0.039)	4.822(0.296)	5.009(0.016)

Table 2: MII: Mean and standard error values of MISE values and $\hat{\sigma}^2$ values. MISE(f) integrate on $[1, \nu]$.

To demonstrate our procedure in nongrid situation. In the simulation study, 3,600 random locations in \mathbb{R}^2 are generated uniformly in $[0, 24] \times [0, 24]$ space. As discussed in Section 2.2, some sort of alignment is necessary. Instead of lining each location to its nearest grid, we first compute the distance of two locations, then round it to the nearest lag as in the previous grid case. In this study, we use grid= 0.4. Again, both MI and MII simulations are conducted and shown in Table 3 and Table 4. The tables clearly show that the procedure performs well in irregular space situations. While MISE(f)s are slightly higher than those in grid situation, MISE(2γ)s are comparable, so are the measurement error variance estimates.

4. Decadal Temperature Change Over the Americas. To illustrate our methodology,

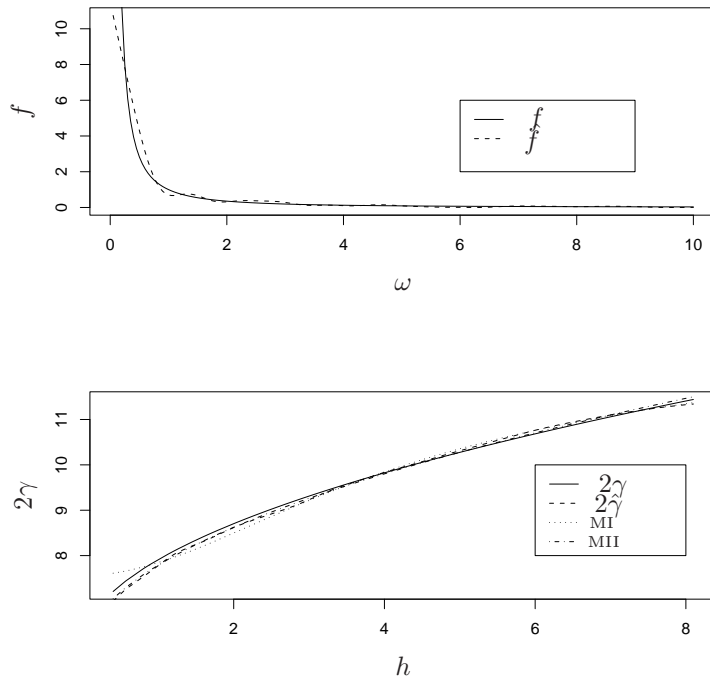


Figure 4: (Grid case) The upper plot spectrum estimation; the lower plot is variogram estimation. The true model is MII.

we consider a global temperature dataset, obtained from the Climate System Model (CSM) at the National Center for Atmospheric Research (NCAR). The complete dataset consists of yearly average of 2-meter air temperature for the period of 1980 through 1990, over the whole globe on 128×64 equi-angular longitude-latitude grid cells. Each cell is roughly 2.8° in longitude by 2.8° in latitude. This dataset, or parts of it, has been analyzed by Shen et al. (2002) and Zhang (2007), who used false discovery rates and spatial exceedances, respectively, to look for evidence of temperature change. We are interested in the spatial spectrum of temperature changes between the 1980s and 1990s. Consequently, for each grid cell, we calculate the average yearly temperature in the 1980s and subtract it from the average yearly temperature ($^\circ C$) between the two decades.

All observations are on the grid, we can directly apply the procedure in Section 2. Figure 5 show the results. The upper plot is the spectrum estimation. The lower plot is the variogram estimation. The measurement error variance turns out to be almost zero. This is due to that the temperature field is actually computer output from the CSM, it is smoother than one might encounter in practice. The solid dot in the figure is method of moment variogram estimation plot.

setting	MISE(f)	MISE(2γ)	MISE($2\gamma, \text{MI}$)	MISE($2\gamma, \text{MII}$)
$\sigma^2 = 0.16, \nu = 6$	0.537(0.045)	0.175(0.022)	0.137(0.023)	0.649(0.035)
		0.151(0.001)	0.155(0.002)	0.161(0.004)
$\sigma^2 = .16, \nu = 10$	0.274(0.027)	0.164(0.022)	0.137(0.023)	0.649(0.035)
		0.127(0.004)	0.155(0.002)	0.161(0.004)
$\sigma = .36, \nu = 6$	0.666(0.046)	0.312(0.050)	0.268(0.052)	0.738(0.062)
		0.340(0.002)	0.353(0.003)	0.371(0.004)
$\sigma = .36, \nu = 10$	0.353(0.037)	0.295(0.050)	0.268(0.052)	0.738(0.062)
		0.305(0.007)	0.353(0.003)	0.371(0.004)

Table 3: MI: (nongrid) Mean and standard error values of MISE values and $\hat{\sigma}^2$ values.

setting	MISE(f)	MISE(2γ)	MISE($2\gamma, \text{MI}$)	MISE($2\gamma, \text{MII}$)
$\sigma^2 = 3, \nu = 6$	3.038(0.147)	8.107(1.278)	8.271(1.421)	6.975(1.300)
		3.435(0.017)	3.567(0.083)	2.992(0.017)
$\sigma^2 = 3, \nu = 10$	3.344(0.416)	8.155(1.275)	8.271(1.421)	6.975(1.300)
		3.088(0.040)	3.567(0.083)	2.992(0.017)
$\sigma^2 = 5, \nu = 6$	3.820(0.202)	16.420(2.724)	24.752(2.786)	15.140(2.758)
		5.396(0.027)	3.222(0.351)	4.983(0.025)
$\sigma^2 = 5, \nu = 10$	6.212(0.943)	16.497(2.724)	24.752(2.786)	15.140(2.758)
		4.905(0.059)	3.222(0.351)	4.983(0.025)

Table 4: MII: (nongrid) Mean and standard error values of MISE values and $\hat{\sigma}^2$ values.

Our nonparametric estimate fits those points very well. If one assumes a parametric model I, which shows in the long dotted line with ($\hat{a} = 7.793, \hat{c}_0 = 3.33, \hat{\sigma}^2 = 0$). These two estimates are quite close to each other. This might indicate that the parametric model I might be a good choice for model fitting. However, parametric model II, which shows in the short dotted line with ($\hat{a} = 0.013, \hat{\sigma}^2 = 0.088$) is clearly off the chart. This again confirms that if one chooses a wrong parametric model to begin with, the estimation may be quite off.

5. Discussion and Conclusion. In this article, we present a nonparametric estimation for variogram and its spectrum. As the simulation and data analysis demonstrate, this method can provide a conditionally negative definite variogram estimate. It will add more tools when we explore the spatial dependencies.

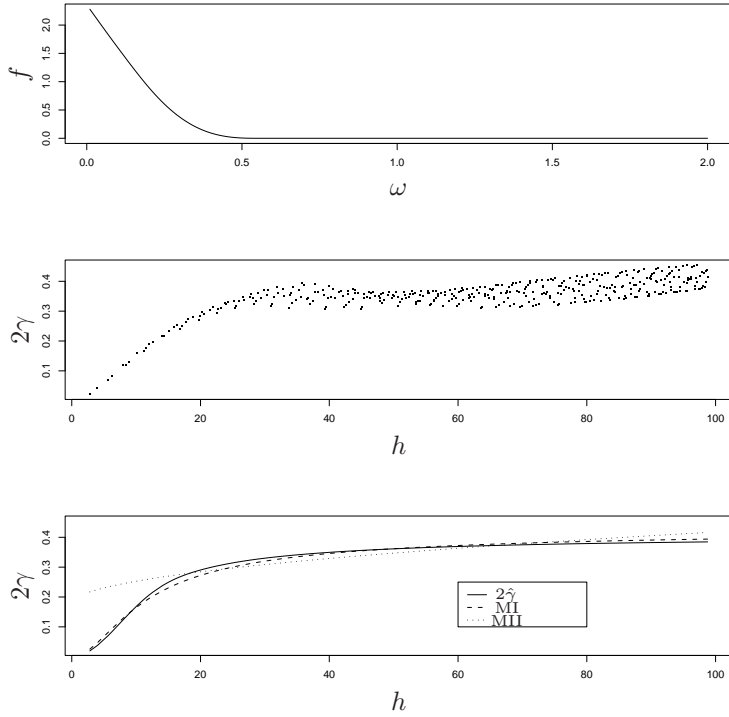


Figure 5: Data Analysis: Upper plot is spectrum estimate; middle plot is variogram cloud; lower plot is variogram estimates. $2\hat{\gamma}$ is the nonparametric fitting ($\lambda = 2.64$, $\hat{\sigma}^2 = 0$); parametric Model I gives $\hat{a} = 7.793$, $\hat{c}_0 = 3.33$, $\hat{\sigma}^2 = 0$; parametric Model II gives $\hat{a} = 0.013$, $\hat{\sigma}^2 = 0.088$.

Spectrum is less explored in spatial statistics as its counterpart in time series. As in time series, many situations show a strong periodic trend. While in spatial statistics, such physical meaning is not so clear. More studies might be needed to understand more about spectrum, its physical interpretation in spatial statistics.

The similar regularized inverse procedure in stationary process on \mathbb{R}^1 is discussed in Huang, Hsing and Cressie. There, a closed form solution is obtained, which paves a way for an efficient algorithm and the study of consistency. A short discussion on inference is provided in Section 2.3. The theoretical issue of consistency of general spline problem is difficult, and we are not ready to address it in this article.

There have not been much work on the smoothing parameter selection for dependent data; see Wang (1998) and references therein. In this article, a modified GCV procedure is proposed, which shows a satisfactory result in our simulation. Finding a general way to address such issue will need further investigation.

Acknowledgements

Chunfeng Huang and Tailen Hsing's research was supported by the National Science Foundation Grant DMS 0808993.

References.

- [1] Amstrong, M. and Jabin, R. Variogram models must be positive-definite. *Mathematical Geology* 13, 455-459, 1981
- [2] de Boor, C. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.
- [3] Chilès, J. and Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 1999.
- [4] Cox, D. D. (1988). Approximation of method of regularization estimators. *The Annals of Statistics* 16, 694-712.
- [5] Cressie, N. Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology* 17, 563-586, 1985.
- [6] Cressie, N. *Statistics for spatial data*. Wiley, New York, 1993.
- [7] Fuentes, M. Spectral methods for nonstationary spatial processes. *Biometrika* 89, 197-210, 2002.
- [8] Fuentes, M. Approximation likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association* 102, 321-331, 2007.
- [9] Green, P. J. and Silverman, S. W. *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall, New York, 1994.
- [10] Hall, P., Fisher, N. I. and Hoffmann, B. On the nonparametric estimation of covariance functions. *The Annals of Statistics*, 22, 2115-2134, 1994.
- [11] Im, H. K., Stein, M. L. and Zhu, Z. Semiparametric estimation of spectral density with irregular observations. *Journal of the American Statistical Association* 102, 726-735, 2007.
- [12] Journel, A. G. and Huijbregts, C. J. Robust kriging-a proposal, *Journal of the International Association for Mathematical Geology* 16, 61-75, 1978.
- [13] Kimeldorf, G. and Wahba, G. Some results of Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications* 33, 82-95, 1971.

- [14] Lukas, M. A. Convergence rates for regularized solutions. *Mathematics of Computation* 51, 107-131, 1988.
- [15] Matheron, G. Principles of geostatistics, *Economic Geology* 58, 1246-1266, 1963.
- [16] Nychka, D. and Cox, D. D. (1989). Convergence rates for regularized solutions of integral equalities from discrete noisy data. *Annals of Statistics* 17, 556-572.
- [17] Nychka, D., Wahba, G., Goldfarb, S. and Pugh, T. Cross-validated spline methods for the estimation of three-dimensional tumor size distribution from observations on two-dimensional cross sections. *Journal of the American Statistical Association* 79, 832-846, 1984.
- [18] Lukas, M. A. (1988). Convergence rates for regularized solutions. *Mathematics of Computation* 51, 101-137.
- [19] Omre, H. The variogram and its estimation, in: G. Verly, M. David, A. Journel and A. Marechal (eds), *Geostatistics for Natural Resources Characterization, Part I*, Reidel, Dordrecht, 107-125, 1984.
- [20] F. O'Sullivan, A statistical perspective on ill-posed inverse problems. *Stat. Sci.* 1, 502-527, 1986.
- [21] Sampson, P. and Guttorp, P. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* 87, 108-119, 1992.
- [22] Shaprio, A. and Botha, J. D. Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics & Data Analysis* 11, 87-96, 1991.
- [23] Shen, X., Huang, H. C. and Cressie, N. Nonparametric hypothesis testing for a spatial signal. *Journal of the American Statistical Association* 97, 1122-1140, 2002.
- [24] Stein, M. *Interpolation of spatial data - some theory for kriging*. Springer Verlag, New York, 1999.
- [25] Villalobos, M. and Wahba, G. Inequality constrained multivariate smoothing splines with application to the estimation of posterior probabilities. *Journal of the American Statistical Association* 82, 239-248, 1987.
- [26] Wahba, G. Convergence rates of certain approximation solution to Fredholm integral equations of the first kind. *SIAM Journal of Control* 11, 64-79, 1973.
- [27] Wahba, G. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.

- [28] Wang, Y. Smoothing spline models with correlated random errors. *J. Amer. Statist. Assoc.* **93**, 341-348, 1998.
- [29] Yaglom, A. M. *Correlation Theory of Stationary and Related Random Functions*. Springer, New York, 1987.
- [30] Yu, K., Mateu, J. and Porcu, E. A kernel-based method for nonparametric estimation of variogram. *Statistica Neerlandica* 61, 173-197, 2007.
- [31] Zhang, J. *Loss Function Approaches to Predict a Spatial Quantile and its Exceedance Region*. Ph. D. thesis, Department of Statistics, The Ohio State University, Columbus, OH, 2007.

Appendix. We first develop computational formulas for $\xi_{i,j}$ and $L_{i,j}\phi_k$. For convenience let $\nu = 1$. By integral formulas of Bessel function (Gradshteyn and Ryzhik, 1980), it follows that

$$\begin{aligned}
\xi_{i,j}(s) &= L_{i,j}R_1(s, \cdot) \\
&= \int_0^1 (1 - J_0(h_{i,j}t))R_1(s, t)dt \\
&= -\frac{4}{2h_{i,j}^4} + \frac{2\sqrt{\pi}}{h_{i,j}^3}s + \frac{1}{4}s^2 - \frac{1}{6}s^3 + \frac{1}{24}s^4 \\
&\quad - \frac{1}{2h_{i,j}}s^2J_1(h_{i,j}) + \frac{1}{6}s^3J_0(h_{i,j}) + \frac{1}{2h_{i,j}}s^3J_1(h_{i,j}s) - \frac{1}{6}s^4J_0(h_{i,j}s) \\
&\quad + \frac{1}{6h_{i,j}^3}s\{2J_0(h_{i,j}s)S_{2,-1}(h_{i,j}s) - J_{-1}(h_{i,j}s)S_{3,0}(h_{i,j}s)\} \\
&\quad - \frac{1}{2h_{i,j}^2}\{J_0(h_{i,j}s)S_{1,-1}(h_{i,j}s) - J_{-1}(h_{i,j}s)S_{2,0}(h_{i,j}s)\} \\
&\quad + \frac{\pi}{12}s^3\{J_1(h_{i,j})H_0(h_{i,j}) - J_0(h_{i,j})H_1(h_{i,j})\} \\
&\quad - \frac{\pi}{12}s^4\{J_1(h_{i,j}s)H_0(h_{i,j}s) - J_0(h_{i,j}s)H_1(h_{i,j}s)\},
\end{aligned}$$

where $J_\nu(x)$ is the Bessel function of the first kind, $H_\nu(x)$ is the Struve function and $S_{\mu,\nu}(x)$ is the Lommel function.

With $m = 2$, $\phi_1 = 1$, $\phi_2 = t$, we have

$$\begin{aligned}
L_{i,j}\phi_1 &= 1 - J_0(h_{i,j}) - \frac{\pi}{2}\{J_1(h_{i,j})H_0(h_{i,j}) - J_0(h_{i,j})H_1(h_{i,j})\} \\
L_{i,j}\phi_2 &= \frac{1}{2} - \frac{1}{h_{i,j}}J_1(h_{i,j}).
\end{aligned}$$