Semiparametric Estimation of Spectral Density With Irregular Observations

Hae Kyung IM, Michael L. STEIN, and Zhengyuan ZHU

We propose a semiparametric method for estimating spectral densities of isotropic Gaussian processes with scattered data. The spectral density function (Fourier transform of the covariance function) is modeled as a linear combination of *B*-splines up to a cutoff frequency and, from this point, a truncated algebraic tail. We calculate an analytic expression for the covariance function and tackle several numerical issues that arise when calculating the likelihood. The parameters are estimated by maximizing the likelihood using the simulated annealing method. Our method directly estimates the tail behavior of the spectral density, which has the greatest impact on interpolation properties. The use of the likelihood in parameter estimation takes the correlations between observations fully into account. We compare our method with a kernel method proposed by Hall et al. and a parametric method using the Matérn model. Simulation results show that our method outperforms the other two by several criteria. Application to rainfall data shows that our method outperforms the kernel method.

KEY WORDS: *B*-spline; Covariance function; Kriging; Maximum likelihood.

1. INTRODUCTION

Estimating the covariance structure of physical processes observed at a finite set of locations is fundamental to understanding the behavior of such processes and interpolating to locations where measurements are not available. Kriging, an interpolation method widely used by the geophysical community, is based on knowledge of the covariances between observed and interpolated locations.

Here we concentrate on Gaussian isotropic processes, which are invariant under all rigid motions. Under this assumption, the covariance of the process at two locations depends only on the distance between them, so a covariance function on $\mathbb{R}^+$ fully describes the second-order properties of the process. This function must be positive definite to ensure that the variance of any linear combinations of values of the process at various locations is positive. The most common solution to this problem is to restrict the estimation to parametric forms known to be positive definite.

Some work has been done on using nonparametric methods or a broad class of positive definite functions based on the spectral representation of covariance functions. Before describing these methods, we provide a brief review of positive definite functions. Bochner's theorem (Yaglom 1987) states that a function is continuous and positive definite if and only if it is the Fourier transform of a positive bounded measure on $\mathbb{R}^d$, that is,

$$C(\mathbf{x}) = \int_{\mathbb{R}^d} \exp(i\mathbf{w}\mathbf{x}) F(\mathbf{dw}). \tag{1}$$

For isotropic processes, (1) can be reduced to a one-dimensional integral,

$$C(r) = 2^{(d-2)/2}\Gamma(d/2) \int_0^\infty (ru)^{-(d-2)/2} J_{(d-2)/2}(ru)\, dG(u), \tag{2}$$

where $G(u) = \int_{|\mathbf{w}|<u} dF(\mathbf{dw})$ is a bounded positive measure on $\mathbb{R}$, $\Gamma(\cdot)$ is the gamma function, and $J_\nu(\cdot)$ is the Bessel function of the first kind of order ν (Abramowitz and Stegun 1965, p. 355).

Shapiro and Botha (1991) proposed using a finite discrete measure with nodes placed at $t_1, \ldots, t_n$ so that the integral in (2) is reduced to a finite sum,

$$\tilde{C}(r) = \sum_{j=1}^{m} p_j \Omega_d(t_j h), \tag{3}$$

where the p_j's are positive and $\Omega_d(x) = (\frac{2}{x})^{(d-2)/2}\Gamma(\frac{d}{2}) \times J_{(d-2)/2}(x)$. For a random field Z on $\mathbb{R}^d$, they used a raw covariogram estimate given by

$$\hat{C}(h) = \frac{1}{N(h)} \sum_{N(h)} (Z(\mathbf{x}_i) - \bar{Z})(Z(\mathbf{x}_j) - \bar{Z}), \tag{4}$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are observation sites, $\bar{Z}$ is the average of the observations, the sum in (4) runs over all pairs of observations that are approximately distance h apart, and $N(h)$ is the total number of such pairs. Shapiro and Botha estimated the values of p_j by minimizing the mean squared difference between the raw covariogram estimate at different lags and their estimator, with positivity constraint.

Genton and Gorsich (2002) followed this idea but proposed using the 0's of the Bessel functions as the nodes of the discrete measure and showed that their method is computationally simpler, needs fewer nodes, and does not show spurious oscillations. A problem with this choice of nodes is that these numbers are nondimensional. It is not clear what scale should be used to translate these nodes into nodes in the frequency domain (which have dimension 1/unit of distance). Although not totally explicit, Genton and Gorsich seemed to propose using $1/r_{\max}$ as their scale. This would mean that if we added an additional observation at distance $1.5r_{\max}$ from the most distant existing observation, then the nodes would be shifted by a factor of 1.5 in the frequency domain. This behavior seems problematic.

Hall, Fisher, and Hoffmann (1994) proposed using a kernel estimator for a preliminary covariogram estimate and, to ensure positive definiteness, proposed Fourier-transforming the

Hae Kyung Im is Research Associate, CISES, University of Chicago, Chicago, IL 60637 (E-mail: *haky@uchicago.edu*). Michael L. Stein is Professor, Department of Statistics, University of Chicago, Chicago, IL 60637. Zhengyuan Zhu is Assistant Professor, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599 (E-mail: *zhuz@email.unc.edu*). The research described herein was funded wholly or in part by the United States Environmental Protection Agency through STAR Cooperative Agreement #R-82940201-0 to the University of Chicago and National Science Foundation grant DMS-06-05434. It has not been subjected to the EPA's required peer and policy review and thus does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. The authors thank Ken Wilder for providing the C code needed to accurately calculate the hypergeometric functions, and Dr. Pavel Groisman from the National Climatic Data Center for providing the rainfall dataset.

Journal of the American Statistical Association
June 2007, Vol. 102, No. 478, Theory and Methods
DOI 10.1198/016214507000000220

kernel estimator, setting the negative values to 0, and Fourier-transforming back to the spatial domain. For $d = 2$, with $\hat{Z}_{ij} = (Z(\mathbf{x}_i) - \bar{Z})(Z(\mathbf{x}_j) - \bar{Z})$, h_{ij} the distance between the observations Z_i and Z_j, K a kernel function (a positive symmetric probability density), and δ the bandwidth, the first-step estimator of the covariogram is

$$\tilde{C}(h) = \frac{\sum_{i,j} \hat{Z}_{ij} K(\frac{h-h_{ij}}{\delta})}{\sum_{i,j} K(\frac{h-h_{ij}}{\delta})}. \tag{5}$$

The final estimate of the covariogram is

$$\bar{C}(h) = \int_0^\infty \left(\int_0^\infty \tilde{C}(x) x J_o(wx)\, dx \right)_+ w J_o(wh)\, dw, \tag{6}$$

where the subscript "+" means to take the positive part of the expression. We call this function the kernel estimator of the covariance function.

All three methods use the raw covariogram as the basis for estimation, which does not take into account the correlation between different $\hat{Z}_{ij}$'s. Furthermore, it is well known that the high-frequency properties of the spectral density determine the performance of interpolation procedures (Stein 1999, p. 65). None of the aforementioned methods gives proper consideration to the tail properties.

We propose a flexible family of models for the spectral density that is a linear combination of B-splines of order 4 (cubic splines) up to a cutoff frequency w_t and an algebraically decaying tail from w_t to infinity. We use positive coefficients for the B-splines, which ensures positiveness of the spectral density and, as a consequence, positive definiteness of the covariance function. Assuming that the process is well described by a Gaussian random field, we find the parameters that maximize the likelihood. This method estimates the tail property of the spectral density in an explicit way. Although this approach excludes exponential decay of the tail, we consider this restriction beneficial, because such a fast decay would imply an unrealistically smooth process (Stein 1999, p. 29; 2002). It also excludes oscillatory tails such as $w^{-\gamma} \cos^2 w$, which generally are undesirable (Stein 1999, pp. 67–68; 2002). In addition, through the use of likelihood, our method takes all of the correlations between observations into account. To the best of our knowledge, this is the first work that uses a likelihood approach for scattered spatial data without a parametric model.

Section 2 presents our model and the methodology for estimating the covariance function. Section 3 describes how to deal with the numerical challenges that arise when calculating the likelihood. Section 4 describes several performance measures and, through a simulation study, compares our method with a parametric method using the Matérn model and the nonparametric kernel method of Hall et al. (1994). Section 5 compares the three methods using a rainfall dataset. The simulation study shows that our method is substantially better than the kernel method and often better than the Matérn model. A real data example shows that, in terms of predictive performance, our method is superior to the kernel method and essentially identical to using the Matérn model. Section 6 summarizes and discusses possible further work.

2. METHODOLOGY

We assume that the observations come from realizations of a Gaussian random field whose value at location x is of the form

$$Z(\mathbf{x}) = \mathbf{m}(\mathbf{x})^T \boldsymbol{\beta} + \epsilon(\mathbf{x}), \tag{7}$$

where $\mathbf{m}(\mathbf{x})$ is a known vector valued function, $\boldsymbol{\beta}$ is a vector of unknown coefficients, ϵ has mean 0 with covariance function $C(\epsilon(\mathbf{x}), \epsilon(\mathbf{y})) = C_{\boldsymbol{\theta}}(|\mathbf{x} - \mathbf{y}|)$, and $\boldsymbol{\theta}$ is the vector of unknown parameters of the covariance function.

2.1 The Splines + Tail (S + T) Model and the Matérn Model for the Spectral Density

Letting $f(w)$ be the spectral density of $\epsilon(\mathbf{x})$ in (7), our S + T model can be written as

$$f_{\boldsymbol{\theta}}(w) = \sigma^2 \sum_{i=-1}^{l+1} b_i B_i(w) \mathbf{1}_{[0,w_t]}(w) + C_f \left(\frac{w_t}{w} \right)^\gamma \mathbf{1}_{(w_t,\infty)}(w), \tag{8}$$

where $\mathbf{1}_A(w)$ is an indicator function of value 1 if $w \in A$ and 0 otherwise. The B_is are B-splines of order 4 (de Boor 2001, chap. IX) with node sequence $(w_0, \ldots, w_l)$ on the interval $[0, w_t]$, where w_t is the threshold frequency. The sum goes from -1 to $l + 1$ to include all B-splines that have support on the interval $[0, w_t]$. We chose order 4 because of the flexibility that cubic splines give to represent a wide range of smooth functions. B-splines of other orders could be used with minor adjustments. We require the spectral density to be continuous and have continuous derivative at w_t. The constant C_f is chosen to achieve continuity at w_t; more explicitly, $C_f = \sigma^2 \sum_{i=-1}^{l+1} b_i B_i(w_t)$. The coefficients of the B-splines are constrained to be positive except for b_{l+1}, which is chosen so that the derivative of $f_{\boldsymbol{\theta}}(w)$ is continuous at w_t. The function is still positive, as was shown by Im, Stein, and Zhu (2006, app. D). Restricting the other coefficients to be positive is a simple way of ensuring positivity of the function.

To specify a S + T model, we first need to determine the number and location of the nodes for the B-splines. In this article we restrict the B-splines to have uniformly distributed nodes between 0 and w_t; that is, given the number of nodes $(l+1)$ and the cutoff frequency (w_t), we place the nodes at locations iw_t/l for $i = -1, 0, \ldots, l, l + 1$.

Conditional on knowing the number of nodes, the parameter $\boldsymbol{\theta}$ of our S + T model includes smoothness parameter γ, sill σ^2, cutoff frequency w_t, and the coefficients of B-splines (b_i for $i = 0, \ldots, l$). The coefficients b_{-1} and b_{l+1} determine the derivatives of the function at the end points. We chose b_{-1} to equal b_1 to make $f'(0) = 0$ and b_{l+1} to be such that the derivative of the function is continuous at w_t (see Im et al. 2006, app. C).

We briefly describe the Matérn model to compare it with our model. This class is considered a sensible model for a wide range of processes arising in environmental problems (Stein 1999; Handcock and Wallis 1994). With only three easily interpretable parameters (σ^2, ϕ, and ν), the Matérn class allows considerable flexibility in the type of processes that it can represent. The parameter σ^2 is simply the variance of the process at a given location, ϕ is the inverse range parameter, and ν is a measure of the smoothness of the process. The spectral density

of the Matérn class has the form

$$f(w) = \frac{\sigma^2 \lambda(\phi, \nu)}{(\phi^2 + w^2)^{\nu + d/2}}, \tag{9}$$

with $\lambda(\phi, \nu) = \Gamma(\nu + d/2)/(\pi^{d/2}\Gamma(\nu))\phi^{2\nu}$ such that the variance, $C(0)$, is σ^2. At high frequencies, both the Matérn and our S + T model approach 0 at the rate $1/w^\gamma$, with $\gamma = 2\nu + d$. In the simulation studies, we use ν as the parameter for the S + T model.

2.2 Computation of the Covariance Function Using the Hankel Transform

The covariance function can be calculated from the spectral density $f_{\boldsymbol{\theta}}(w)$ by applying (2) for $d = 2$,

$$C_{\boldsymbol{\theta}}(r) = 2\pi \int_0^\infty w J_o(rw) f_{\boldsymbol{\theta}}(w)\, dw. \tag{10}$$

The transform in (10) is called the Hankel transform of order 0. A function $f_{\boldsymbol{\theta}}(w)$ in the S + T family is a linear combination of B-splines and an algebraic tail; the Hankel transform of both components can be calculated analytically. The Hankel transform of B-splines requires calculating two Bessel functions of the first kind of orders 1 and 2 and two Struve functions of orders 1 and 2 (Abramowitz and Stegun 1965, p. 495) for each node. The computation of this part is straightforward, albeit moderately time-consuming (see Im et al. 2006, app. B, for details). The Hankel transform of the truncated algebraic tail is

$$\begin{aligned}\int_{w_t}^\infty & w^{1-\gamma} J_o(wr)\, dw \\ &= r^{\gamma-2}\left(\frac{\gamma\Gamma(-\gamma/2)}{2^\gamma\Gamma(\gamma/2)}\right. \\ &\quad \left. + \frac{(w_t r)^{2-\gamma}\, {}_1F_2(1-\gamma/2; 1, 2-\gamma/2; -(rw_t)^2/4)}{\gamma - 2}\right),\end{aligned} \tag{11}$$

where ${}_1F_2(a; b, c; z)$ is a generalized hypergeometric function with series representation $\sum_{k=0}^\infty \frac{(a)_k}{(b)_k (c)_k}\frac{z^k}{k!}$. Here $(\cdot)_k$ represents Pochhammer's symbol (Abramowitz and Stegun 1965, p. 256), which is defined by $(z)_0 = 1$ and $(z)_k = z(z+1)(z+2)\cdots(z+n-1) = \frac{\Gamma(z+n)}{\Gamma(z)}$.

Several features make the Hankel transform of the tail (11) numerically hard to compute. First, there is no easy way of evaluating this hypergeometric function accurately without resorting to summing a large number of terms of its series expansion, which can lead to severe numerical errors. Second, the Γ function is infinite when the argument is a negative integer, and we have no reason to exclude negative integer values for $-\gamma/2$. Third, the first term, $\gamma\Gamma(-\gamma/2)/2^\gamma\Gamma(\gamma/2)$, is the limit of the second term, $(w_t r)^{2-\gamma}\, {}_1F_2(1-\gamma/2; 1, 2-\gamma/2; -(rw_t)^2/4)/\gamma - 2$, as $rw_t \to \infty$, so when rw_t is large, we need to take the difference of two very similar numbers.

The first problem is addressed by using arbitrary precision arithmetic libraries (code downloaded from *http://www.mpfr.org/*). For the second problem, we note that the divergence of the Γ function is compensated for by the divergence of one of the terms of the series expansion of the hypergeometric function. When $\gamma/2$ is an integer, we use an asymptotic expansion of the Γ function when the argument is close to $-\gamma/2$ and subtract it from the series expansion of the hypergeometric function. Only one term in each series expansion diverges as the arguments approaches $-\gamma/2$, and we get a modified series expansion for the difference between the two terms, which can be computed in the same fashion as the hypergeometric function, that is, by adding the series until convergence is achieved and using arbitrary precision libraries to avoid numerical errors. The final expression for the tail integral when $\gamma/2 = n + 1$ is

$$\begin{aligned}\int_{w_t}^\infty & w^{1-\gamma} J_o(wr)\, dw \\ &= r^{2n}\frac{\log(2) - \log(rw_t) + \psi(n+1)}{(-4)^n n!^2} \\ &\quad + \frac{w_t^{2n}}{2n}\sum_{k=0, k\neq n}^\infty \frac{-n}{-n+k}\frac{(-(rw_t)^2/4)^k}{k!^2},\end{aligned} \tag{12}$$

where $\psi(n)$ is the digamma function (Abramowitz and Stegun 1965, p. 258), which for positive integer arguments can be evaluated as $\sum_{k=1}^{n-1}\frac{1}{k} - \gamma_{eg}$, where $\gamma_{eq} = .577216\ldots$ is the Euler's constant. The details have been given by Im et al. (2006, app. E).

To solve the third problem of subtracting two very similar numbers, we again use an asymptotic expansion of the hypergeometric function whose leading term is $\gamma\Gamma(-\gamma/2)(\gamma-2)/(2^\gamma\Gamma(\gamma/2)(rw_t)^{2-\gamma})$. As a result, we are left with an expression that directly calculates the difference. For large values of rw_t, the Hankel transform of the truncated tail is approximated by

$$\begin{aligned}\int_{w_t}^\infty & w^{1-\gamma} J_o(wr)\, dw \\ &\approx \left(\frac{\cos(rw_t) - \sin(rw_t)}{\sqrt{\pi}(rw_t)^{\gamma-1/2}}\right. \\ &\quad - \frac{(-15 + 16\gamma + 128\gamma^2)(\cos(rw_t) - \sin(rw_t))}{128\sqrt{\pi}(rw_t)^{\gamma+3/2}} \\ &\quad \left. + \frac{(-3 + 8\gamma)(\cos(rw_t) + \sin(rw_t))}{8\sqrt{\pi}(rw_t)^{\gamma+1/2}} + \cdots\right) r^{\gamma-2}\end{aligned} \tag{13}$$

(see the derivation in Im et al. 2006, app. F). The approximation (13) also applies when $\gamma/2$ is an integer.

2.3 Likelihood-Based Parameter Estimation

Let $\mathbf{Z} = (Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n))^T$ be our observation and let $\mathbf{M} = (\mathbf{m}(\mathbf{x}_1), \ldots, \mathbf{m}(\mathbf{x}_n))^T$. Under model (7), the log-likelihood has the form

$$\begin{aligned}l(\boldsymbol{\theta}, \boldsymbol{\beta}; \mathbf{Z}) = -\frac{1}{2}\log|\det \boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2}(\mathbf{Z} - \mathbf{M}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{Z} - \mathbf{M}\boldsymbol{\beta}) \\ - \frac{n}{2}\log(2\pi),\end{aligned} \tag{14}$$

where $\boldsymbol{\Sigma}_{ij} = C_{\boldsymbol{\theta}}(|\mathbf{x}_i - \mathbf{x}_j|)$ can be computed using (10) and (8). If the mean parameter $\boldsymbol{\beta}$ can be assumed to be known, then the parameter $\boldsymbol{\theta}$ of the S + T model can be estimated by maximizing (14). When $\boldsymbol{\beta}$ is unknown, $\boldsymbol{\theta}$ can be estimated by maximizing the following restricted maximum likelihood (REML) (Stein 1999, p. 170; McCullagh and Nelder 1989, sec. 7.2),

which can be calculated by (ignoring an additive constant)

$$rl(\boldsymbol{\theta}; \mathbf{Z}) = -\frac{1}{2}\log|\det \boldsymbol{\Sigma}_{\boldsymbol{\theta}}| - \frac{1}{2}\log|\det \mathbf{W}| - \frac{1}{2}\mathbf{Z}^T(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{M}\mathbf{W}^{-1}\mathbf{M}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})\mathbf{Z}, \quad (15)$$

where $\mathbf{W} = \mathbf{M}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{M}$.

3. NUMERICAL IMPLEMENTATION

Because the need for high computational speed forced us to discretize some of the parameters (smoothness and cutoff frequencies), the usual continuous optimization routines are not applicable. We use the simulated annealing method to maximize the likelihood (Givens and Hoeting 2005, p. 67). The details of the implementations have been given by Im et al. (2006).

The number of nodes can be determined using some model selection criteria. We use the Akaike information criterion (AIC) (Akaike 1974), which penalizes large number of parameters. We first fix the number of nodes l and optimize using simulated annealing as described by Im et al. (2006). We then repeat this optimization for a few different values for l and choose the one with the smallest AIC. We compare the performance of the AIC and the Bayes information criterion (BIC) (Schwarz 1978). Our simulation results indicate that AIC is more likely to select the correct number of nodes when the data are simulated using the S + T model, and the prediction performance using AIC is also slightly better than BIC for all of the models that we used for simulation. We use AIC to select nodes in all of the simulation studies presented in Section 4 and in the application to rainfall data in Section 5.

3.1 Tabulation of Hankel Transforms

To speed up the computation of the Hankel transforms of the truncated tail, we resort to some further approximations and shortcuts. We calculated the covariance function for $n_r = 100$ equispaced values between $r_{\min}$ and $r_{\max}$ and interpolated using cubic spline interpolation for distances between these points. Also, we restricted the values of the threshold frequency w_t and the power of the tail γ to $n_w = 100$ and $n_\gamma = 100$ discrete values, namely $w_t(j) = \frac{1}{r_{\max}} + \frac{j}{n_w}(\frac{1}{r_{\min}} - \frac{1}{r_{\max}})$ and $\gamma(j) = 2 + 2(.05 + \frac{j}{n_\gamma}(5 - .05))$.

The Hankel transform of the truncated tail, $t(i, j, k) = \int_{w_t(k)}^{\infty} w^{1-\gamma(j)} J_o(wr(i))\, dw$, was tabulated into an array of dimensions $n_r \times n_\gamma \times n_w = 100 \times 100 \times 100$. The Hankel transform of piecewise polynomials of the form $\mathbf{1}_{[w_i, w_{i+1})}(w - w_i)^m$ for $m = 0, 1, 2$, and 3 were tabulated into an array of dimensions $l \times 4 \times n_r \times n_w = l \times 4 \times 100 \times 100$, where l is the number of polynomial pieces used in the representation of the spectral density. To take advantage of this tabulation when we calculated the transform, we converted the splines $S(w)$ (linear combination of B-splines) into a piecewise polynomial form,

$$S(w) = \sum_{i=0}^{l}\sum_{j=0}^{3} a_{ij}\mathbf{1}_{[w_i, w_{i+1})}(w - w_i)^j. \quad (16)$$

Hence the Hankel transform was reduced to multiplying the tabulated values by the corresponding coefficients.

The frequency thresholding in the tabulation is unlikely to have an effect in the results because in all of the simulations, the estimated cutoff frequency was about one order of magnitude smaller than the maximum tabulated value.

4. SIMULATIONS

We first simulated Gaussian random fields with mean 0 and various covariance functions and estimated the spectral density using the S + T family of functions. For comparison purposes, we also estimated the covariance function using the Matérn model, as well as the kernel method proposed by Hall et al. (1994). The locations were chosen to be where the National Acid Deposition Program sites are situated. We used a total of 63 sites, as are shown in Figure 1. The smallest distance between sites is 14 km, the largest distance is 2,000 km, and the median distance is 802 km. For simplicity, we use chordal distance and ignore the fact that the surface is spherical.

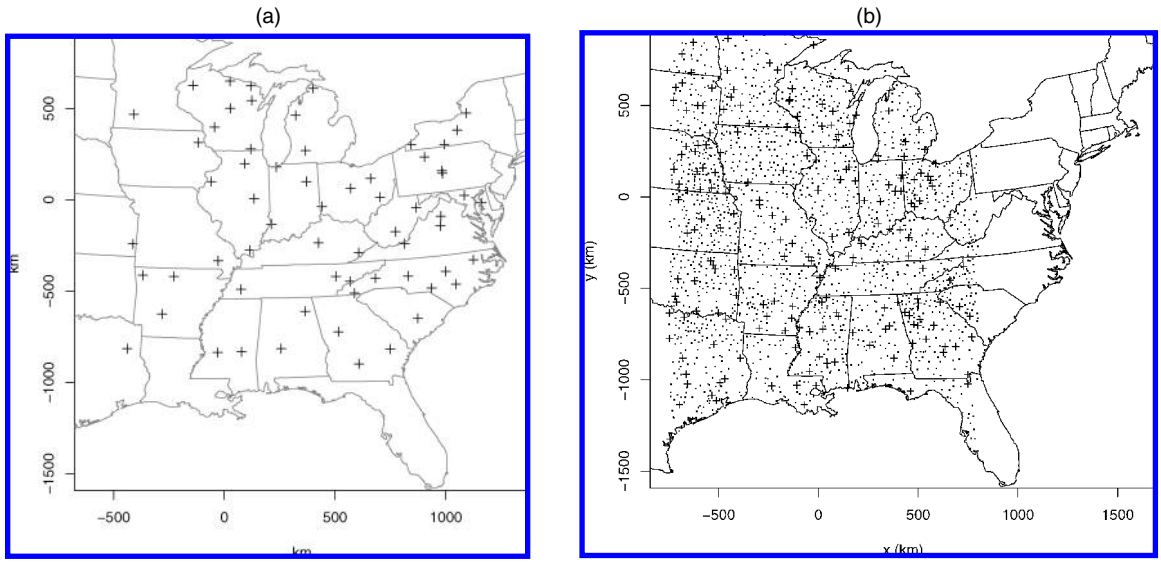

Figure 1. Plot of NADP Monitoring Sites Used for Simulations (a) and the Monitoring Sites for the Rainfall Dataset (b).

The covariance models that we used to simulate the data are Matérn, polynomial Matérn, S + T, and spectral exponential. Polynomial Matérn is a family of spectral densities given by the product of Matérn spectral density and a positive polynomial, $f(w) = ((w-u)^2+v^2)((w+u)^2+v^2)/(a^2+w^2)^{\nu+d/2}$. This function is positive on $\mathbb{R}^+$ and thus is a valid spectral density. The spectral density of the spectral exponential model has the form $f(w) = \exp(-w/\phi)$, which should not be confused with the exponential correlation function that belongs to the Matérn family and is exponential in the spatial domain. In the simulation study, we set $\phi = w_t$. Matérn, S + T, and polynomial Matérn share the same high-frequency behavior, namely $1/\omega^\gamma$. The spectral exponential model has a much faster decay and has analytic realizations of the process. Although we do not consider this type of behavior to be reasonable for modeling natural physical processes, we include it here to test the method.

Each simulation includes 200 independent realizations of a Gaussian random field at 63 locations, totaling 12,600 observations with the given covariance functions. The covariance matrix corresponding to this type of dataset is block diagonal, which allows us to have a large number of observations (so that the parameters can be well estimated) while keeping the computational load at a manageable level. For each simulation, we estimated the maximum likelihood (ML) parameters for S + T and Matérn models, and calculated the kernel estimate of the covariance function.

To assess the performance of each method, we evaluated the following quantities:

- Parameter values. When the true model and the model used to estimate the covariance functions have common parameters (e.g., σ^2 is common to all models), the difference between the true and the estimated parameters is an obvious measure of performance.
- Likelihood values. The value of the likelihood also gives us an indication of how good the models are fitting the data. Although it is a bit unfair to compare methods that seek maximizing the likelihood with methods that seek to optimize other criteria, large deviations from the true likelihood should give us an idea of how good the estimated function is.
- Covariance function. Comparing the distance between the true and estimated covariance function or spectral densities seems to be an obvious and esthetically pleasing way of assessing the performance of various methods. We use the $\mathcal{L}^2$ norm between the true and estimated covariance functions as a measure of performance.
- Prediction performance measures. In most applications, the ultimate goal of estimating the covariance functions is the prediction of the random field at unobserved locations. In this context the $\mathcal{L}^2$ norm can be a misleading benchmark, as illustrated by Stein (1999, p. 66). Following Stein (1999, p. 58), we define two quantities that are more useful for interpolation purposes. Let $\hat{Z}_i(\mathbf{x})$ be the predicted value at location $\mathbf{x}$ using covariance function C_i, and let $e_i(x) = Z(x) - \hat{Z}_i(x)$ be the prediction error; E_i and E_0 are the expectations under the estimated covariance function C_i and under the true covariance function C_0. With this notation, $E_0 e_0^2$ is the mean squared prediction error (MSPE) of the best linear unbiased predictor or the kriging variance [as in Stein 1999, p. 8, eq. (11)], which is computed using the true covariance function; $E_i e_i^2$ is the estimated prediction variance (EPV), computed by plugging in the estimated covariance function C_i in the usual kriging variance; and $E_0 e_i^2$ is the actual variance of the prediction error. It is easy to show that

$$\frac{E_0 e_i^2(x)}{E_0 e_0^2(x)} = 1 + \frac{E_0(\hat{Z}_i(x) - \hat{Z}_0(x))^2}{E_0 e_0^2(x)}. \tag{17}$$

We estimate the numerator on the right side of (17) by computing the sample mean (over 100 simulations and 200 replications in time) of the squared difference between the interpolated values with the misspecified covariance function and the interpolated values with the true covariance function,

$$\frac{1}{200}\frac{1}{100}\sum_{\text{simul.}}^{100}\sum_{\text{repl.}}^{200}(\hat{Z}_i(x) - \hat{Z}_0(x))^2.$$

Thus, we use the following prediction performance measures:

- The increase in prediction error, $IPE(\mathbf{x})$, is the second term on the right side of (17), which represents the extra MSPE incurred by using an estimated (misspecified) covariance function instead of the true covariance function,
- The log variance ratio, $LVR(x) = |\log(E_i e_i^2(\mathbf{x})/E_0 e_i^2(\mathbf{x}))|$, the absolute log ratio between the estimated and actual variance of the prediction error at location $\mathbf{x}$.

For testing, we use 100 prediction locations on a square grid inside the observation region. The medians of IPE and LVR over the prediction locations are used as performance measures.

Table 1 gives the average results of running 100 simulations for each of the following models: Matérn ($\nu = 3$, $\phi = 9.4$), polynomial Matérn ($\nu = 3$, $\phi = 9.4$, $u = 4.7$, $v = .94$), S + T [$\nu = 3$, $w_t = 9.4$, $l = 4$, $\mathbf{b} = (1, .2, 2, .6, .4)$], and spectral exponential ($\phi = 9.4$). For all of the models, $\sigma^2 = 1$, and the unit for ϕ and w_t is 1/1,000 km.

The first three rows show the smoothness parameters, true and estimated, using the S + T and Matérn models. When the true model is Matérn or polynomial Matérn, the estimated smoothness for the S + T model is around 2.2, a bit smaller than the truth, which is 3. This is expected for the Matérn model, because the rate of decay of the estimating tail function $1/w^{2\nu+2}$ is faster than the rate of decay of the true tail function $1/(a^2+w^2)^{\nu+1}$ (see Im et al. 2006, p. 53, for more details).

When simulating under the S + T model, there is no approximation in the tail, and the S + T model gives an estimated value that is very close to the truth. In the case of the spectral exponential model, we do not have a true parameter to compare it with. Note that the S + T and Matérn methods give similar estimates of the smoothness when the true model is spectral exponential.

When the true models are polynomial Matérn and S + T, the Matérn method yields very large estimates for smoothness: 10 and 7.47. The comparison with the S + T method may seem a bit unfair, because for the latter we have restricted the smoothness parameter to be less than 5. But when we restricted the

Table 1. Summary of Simulation Results for $\nu = 3$, $\sigma^2 = 1$, and Inverse Range (or w_t) = 9.4 (1/1,000 km)

		Truth			
		Matérn	*polMatérn*	*S + T*	*Spectral exponential*
ν	True	3	3	3	
	S + T	$2.15_{(.10)}$	$2.21_{(.26)}$	$3.00_{(.06)}$	$1.82_{(.70)}$
	Matérn	$2.98_{(.09)}$	$10.00_{(2.39)}$	$7.47_{(.31)}$	$1.63_{(.18)}$
σ^2	True	1	1	1	1
	S + T	$1.00_{(.02)}$	$1.00_{(.02)}$	$1.00_{(.02)}$	$1.00_{(.01)}$
	Matérn	$1.00_{(.02)}$	$1.00_{(.02)}$	$1.15_{(.03)}$	$1.00_{(.01)}$
	Kernel	$.99_{(.03)}$	$.98_{(.02)}$	$.99_{(.02)}$	$.97_{(.01)}$
$w_t \times 1{,}000$	True			9.4	
	S + T	$12.3_{(1.2)}$	$18.7_{(2.1)}$	$10.6_{(1.4)}$	$32.7_{(5.1)}$
Inverse range	True	9.4			
$\times$ 1,000	Matérn	$9.4_{(.3)}$	$35.4_{(4.6)}$	$15.3_{(.5)}$	$41.0_{(195.7)}$
$\mathcal{L}^2$-norm	S + T	$8.7_{(4.7)}$	$7.9_{(3.9)}$	$7.4_{(2.7)}$	$5.2_{(1.7)}$
covariance	Matérn	$7.4_{(5.0)}$	$53.6_{(.3)}$	$123.9_{(7.8)}$	$10.4_{(16.0)}$
$\times$1,000	Kernel	$20.6_{(6.7)}$	$21.9_{(4.8)}$	$20.9_{(6.2)}$	$9.2_{(1.6)}$
Log-likelihood	Matérn	$4_{(4)}$	$-181_{(22)}$	$-296_{(22)}$	$-20_{(96)}$
ratio	Kernel	$-1{,}093_{(167)}$	$-254_{(34)}$	$-3{,}173_{(246)}$	$-28_{(9)}$

NOTE: Average estimates from 100 simulations are shown. Each simulation consists of 12,600 observations (200 replications of 63 spatially correlated observations). Standard deviations are shown in parentheses. Columns correspond to true models, and rows correspond to estimating methods: ML using S + T, ML using Matérn, and kernel. Log-likelihoods are differences from S + T.

Matérn's smoothness to be less than 5, the estimated smoothness was exactly 5 for all 100 simulations, and we found no substantial change in the performance of the method. Thus the thresholding of the smoothness did not contribute to the advantage of our method over the Matérn.

The estimated values of the sill (σ^2) are very close to the true value of 1 for all three methods except for the Matérn method when the true model is S + T, which yields a mean of 1.15 with standard deviation .03, and the kernel method when the truth is spectral exponential, which yields a mean of .97 with standard deviation .01. The S + T method gives 1.00 with error .02 when the true model is Matérn, polynomial Matérn, and S + T. When the true model is spectral exponential, we get 1.00 with standard deviation .01.

The next two blocks of rows show the cutoff frequency w_t and the inverse range. For the S + T model, the average estimated value of w_t is 10.6 (in 1/1,000 km units) with standard deviation of 1.4, compared with the true value of 9.4. For the Matérn model, the average estimated inverse range parameter is 9.4, with standard deviation of .3.

The $\mathcal{L}^2$ norms between the true and estimated covariance functions are shown next. The S + T method is better by factors of 2.3, 2.8, 2.8, and 1.8 compared with the kernel method when the true models are Matérn, polynomial Matérn, S + T, and spectral exponential. The S + T method also gives smaller $\mathcal{L}^2$ norms compared to Matérn method except when the Matérn model is the truth. The $\mathcal{L}^2$ norms were calculated by averaging the squared differences between the true and estimated covariance function values at 100 equispaced points in the range 0 to $r_{\max}$.

The final block shows the log-likelihood ratio between alternative models and the S + T model. The S + T model gives a larger likelihood in all cases except when the Matérn model is used to estimate and the data were simulated from the Matérn model. Even then, the advantage of using the correct parametric model is modest.

Table 2 gives the prediction performance measures of each method. The first three rows compare the median IPE, the increase in MSPE achieved by using estimated covariance functions instead of the truth. The IPEs of the predictors using the S + T method are less than .2% for all four simulated models: .16%, .12%, .05%, and .12%. The kernel method gives corresponding prediction errors that are 17.5%, 3.8%, 68.5%, and .3% higher than the best linear predictor, whereas the Matérn method gives errors that are .01%, 1.8%, 4.2%, and .4% higher.

The next three rows of Table 2 compare the LVR, log ratios of the estimated (through plug-in), and actual variances of the prediction error. Our method gives EPVs that differ by about .5% from the actual prediction variance. For comparison, the kernel method gives EPVs that differ from the actual prediction variance by 59.1%, 27.0%, 100.8%, and 5.8% for the four models, and the Matérn method's differences were .3%, .9%, 8.3%, and

Table 2. Simulations as in Table 1

		Matérn	*polMatérn*	*S + T*	*Spectral exponential*
$E_0 e_1^2 / E_0 e_0^2 - 1$	S + T	$.16_{(.10)}$	$.12_{(.05)}$	$.05_{(.07)}$	$.12_{(.08)}$
$E_0 e_2^2 / E_0 e_0^2 - 1$	Matérn	$.01_{(0)}$	$1.77_{(3.17)}$	$4.24_{(3.17)}$	$.40_{(.63)}$
$E_0 e_3^2 / E_0 e_0^2 - 1$	Kernel	$17.51_{(32.30)}$	$4.10_{(9.10)}$	$68.54_{(118.61)}$	$.26_{(.45)}$
$\lvert\log(E_1 e_1^2 / E_0 e_1^2)\rvert$	S + T	$.93_{(2.83)}$	$.39_{(.68)}$	$.33_{(.52)}$	$.31_{(1.29)}$
$\lvert\log(E_2 e_2^2 / E_0 e_2^2)\rvert$	Matérn	$.30_{(.76)}$	$.86_{(2.28)}$	$8.57_{(17.24)}$	$1.21_{(2.16)}$
$\lvert\log(E_3 e_3^2 / E_0 e_3^2)\rvert$	Kernel	$59.08_{(94.75)}$	$27.06_{(43.31)}$	$100.78_{(157.13)}$	$5.79_{(16.41)}$

NOTE: The first block of three rows show the median over 100 prediction locations of IPE, the increase in MSPE by using estimated covariance function instead of the truth. The second block of three rows show the median of the log ratio of the EPV and the actual prediction variance. In all rows the interquartile range is shown in parentheses. The values are in percentages.

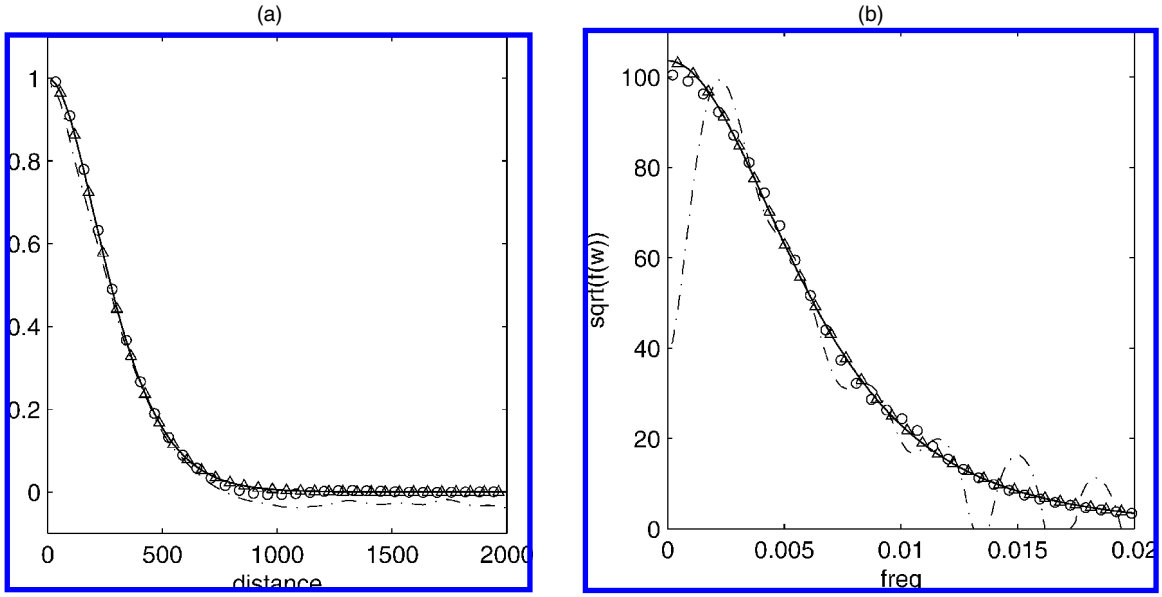

Figure 2. True and Estimated Covariance Function (a) and Spectral Density (b) (——— true; ○ ML S + T; △ ML Matérn; · — · — kernel). The true model is Matérn with $\nu = 3$, $\sigma^2 = 1.00$, and inverse range = 9.4 (1/1,000 km).

1.2%, all substantially worse than the S + T method, except for the Matérn method when the truth is Matérn.

We plotted (not shown in the article) the smoothed difference between the MSPE of the S + T and the other two methods versus the minimum distance to observations. In all cases except when the true model was Matérn, our method yielded smaller MSPEs than the other two for all distances. Most of the time, the kernel method performed substantially worse than the other two methods. For the polynomial Matérn and S + T models, the difference between our method and Matérn increased with distance, which has a natural connection with the fact that our method gives a better fit at small frequencies for all simulated models. In contrast, the difference with the kernel method decreased with distance. We attribute this finding to the kernel method's inability to capture the high-frequency properties of the spectral density. When the true model was spectral exponential, the performance of all three methods was very good, with IPE less than 1%, and we found no obvious pattern in the differences as functions of distance.

We have conducted simulation studies for other parameter values and achieved qualitatively similar results. In practice, we often need to estimate the mean. When the mean is unknown, we can use restricted likelihood to estimate the parameters, and the simulation results are very similar to the case when the mean is assumed known. These numerical results have been given by Im (2005) and are not presented here.

Figures 2–5 show the covariance function (left) and the spectral density (right) for the Matérn, polynomial Matérn, S + T, and spectral exponential models, with one typical simulation from each model. As suggested by the $\mathcal{L}^2$ norm values in Table 1, the covariance functions estimated using the S + T method are closer to the true covariance functions than the kernel estimators. The kernel estimator of the covariance function becomes wiggly for large distances, mainly because there are fewer pairs of observations that contribute to this region. Plot (b) of Figures 2–5 show that our method yields estimates of the spectral densities that are quite close to the truth except near the origin. The Matérn method is not able to reproduce the structure of the spectral density at low and middle frequencies, but it captures the tail behavior reasonably well. The kernel estimates of the spectral density follow the overall shape of the function but are very wiggly. We also see several intervals of frequencies in which the kernel estimator takes value 0, due to the truncation of the function necessary to ensure positive definiteness of the corresponding covariance function. When we calculated the spectral density of the kernel estimator we followed the ad hoc solution proposed by Hall et al. (1994): From some point T_1, use a straight line that goes from the value of the estimator at T_1 to 0 at some other distance T_2. We chose $T_1 = 1{,}500$ km and $T_2 = 2{,}000$ km. The actual transformation of the kernel estimator was done using the fast Hankel transform method proposed by Siegman (1977).

5. APPLICATION TO RAINFALL DATA

In this section we applied the three aforementioned methods to an annual rainfall dataset in the eastern United States (Groisman 2000) to compare their performance in prediction. We chose the study region to be between latitudes 27.1 and 49.0 and longitudes −100.5 and −80.2, which includes 2,742 stations (represented by dots on the right of Fig. 1). The period of data collection is 1960–1999. Detailed documentation of this dataset can be found at *http://www1.ncdc.noaa.gov/pub/data/documentlibrary/tddoc/td9651.pdf*. We used 2,082 stations that had no missing data. For each station's data, we subtracted the mean over the 40 years of data and modeled the difference as a Gaussian random process. We looked at the normal quantile plots of the difference for all of the stations at each year and found that in most years, the data are consistent with the Gaussian assumption. Some of those normal quantile plots appear to have a few large outliers. The number of possible outliers is never larger than .5% of the stations. We made no attempt to identify those outliers; instead, we used robust mea-

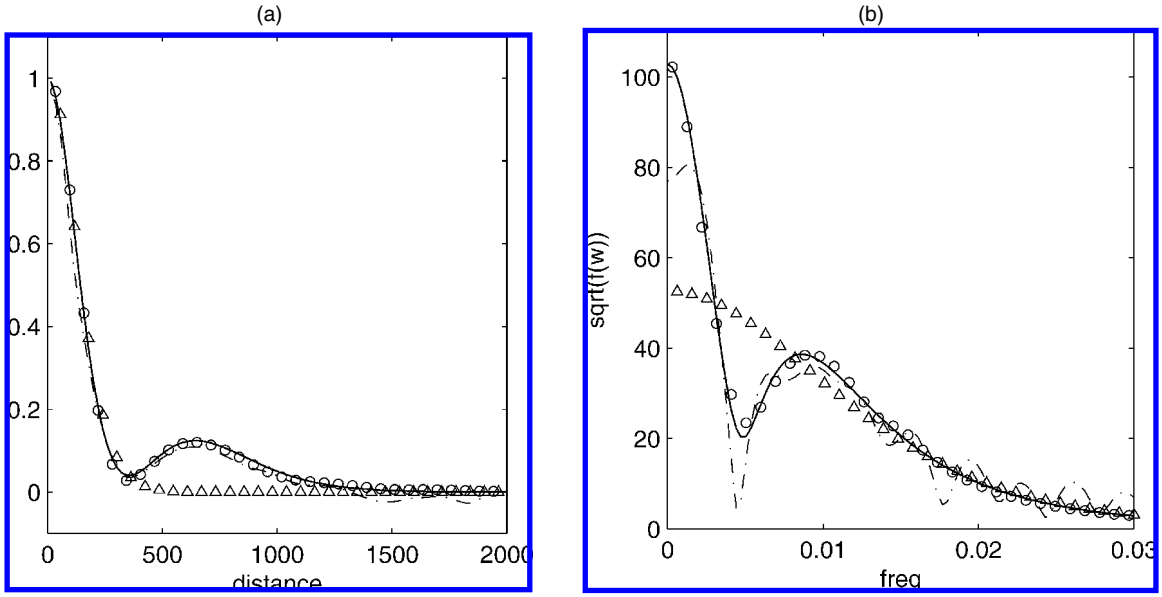

Figure 3. True and Estimated Covariance Function (a) and Spectral Density (b) (——— true; ○ ML S + T; △ ML Matérn; · — · — kernel). The true model is polynomial Matérn with $\nu = 3$, $\sigma^2 = 1.00$, inverse range (or w_t) = 9.4 (1/1,000 km), $u = .5w_t$, and $v = .1w_t$.

sures to compare the three methods. An examination of the autocorrelation and cross-correlation of the observations at different stations reveals no significant time dependence; thus we model the observations from each year as independent realizations of the same process with a different mean. The empirical variogram suggested the need to include a nugget term in the covariance model. This is straightforward for the two likelihood methods (S + T and Matérn). The kernel method needs some modification to handle the nugget effect. One possible way to do this is to estimate the covariance function using only distinct pairs, that is, $(Z_i - \bar{Z})(Z_j - \bar{Z})$, with $i \neq j$. This leaves out the nugget term. We estimate the nugget by subtracting the estimated covariance function at 0 from the sample variance.

We randomly chose 200 training stations among 2,082 stations to estimate the covariance function using all three methods (S + T, Matérn, and kernel). With the estimated covariance functions, we kriged the data at the remaining 1,882 test stations. We repeated this process for 10 different samples of 200 stations. The locations of the first sample of 200 stations are shown as "+"'s in Figure 1(b).

We calculated the mean over 40 years of the prediction error, $MPE(x) = \sqrt{\sum_{t=1}^{40}(Z(\mathbf{x}) - \hat{Z}_i(\mathbf{x}))^2/40}$, and an empirical estimate of the LVR as $|\log(MPE^2(x)/EPV(x))|$. Table 3 shows the median over locations and samples of the MPEs and LVRs. On average, our method yields 1.5% better prediction errors than

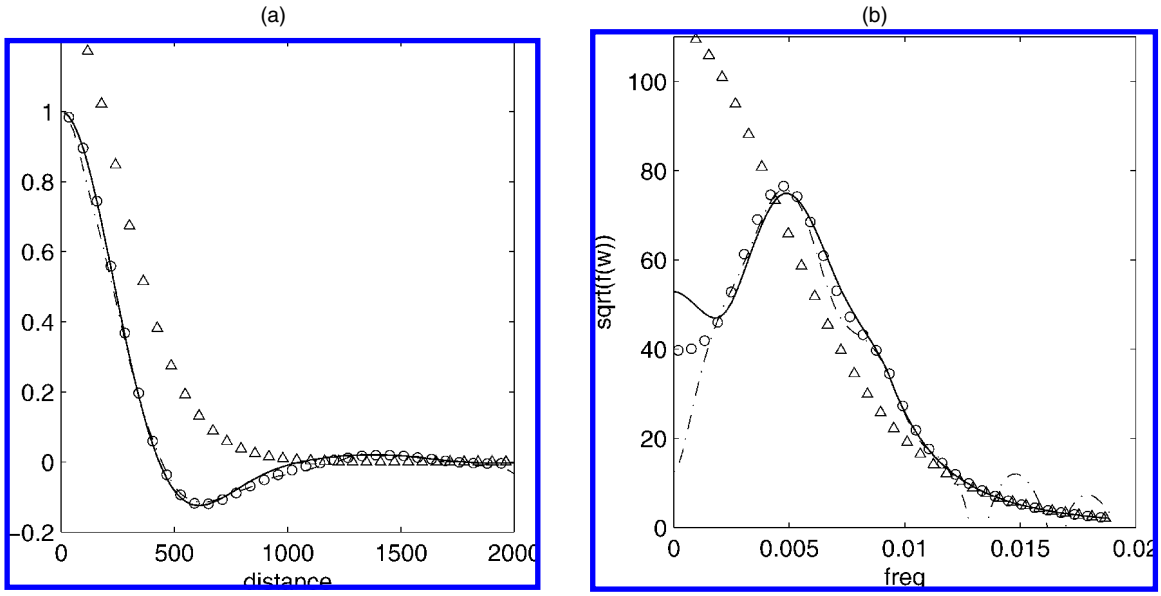

Figure 4. True and Estimated Covariance Function (a) and Spectral Density (b) (——— true; ○ ML S + T; △ ML Matérn; · — · — kernel). The true model is S + T with $\nu = 3$, $\sigma^2 = 1.00$, $w_t = 9.4$ (1/1,000 km), and coefficients $\mathbf{b} = (1, .2, 2, .6, .4)$.

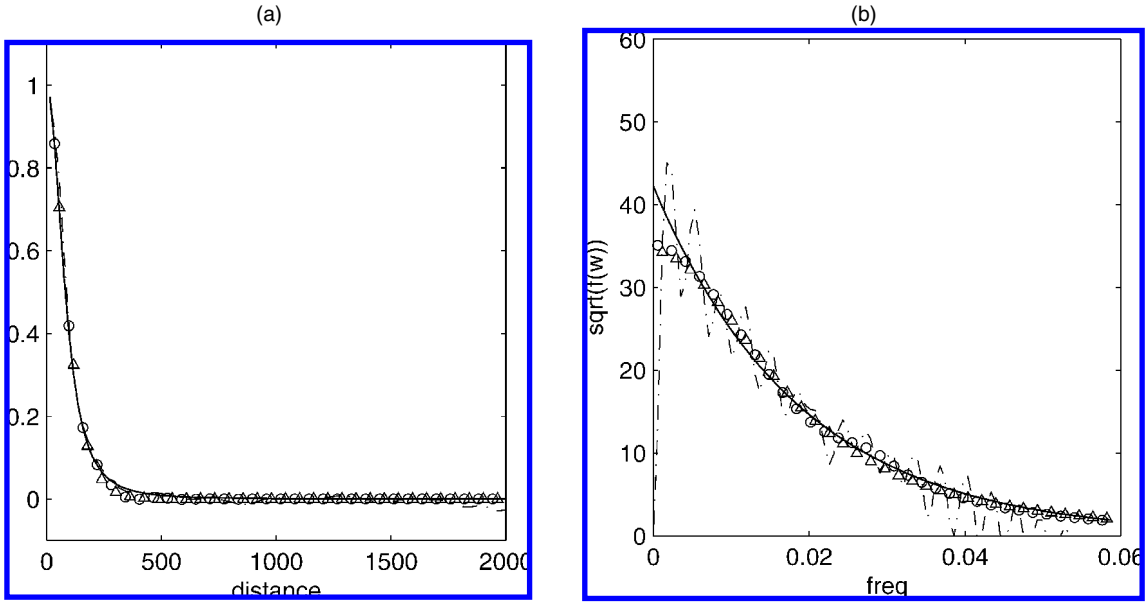

Figure 5. True and Estimated Covariance Function (a) and Spectral Density (b) (——— true; ○ ML S + T; △ ML Matérn; · — · — kernel). The true model is spectral exponential with $\sigma^2 = 1.00$ and inverse range = 9.40 (1/1,000 km).

the kernel method. The LVR of our method is about 4% better than that of the kernel method. There is no noticeable difference in performance between our method and the Matérn method. The estimated spectral density using our method did not show any interesting structure in the middle frequencies that could be missed by the Matérn spectral function. This could explain the lack of advantage in using our method.

We tried using 100 training stations and found similar results. We plotted (not shown) smoothed MPEs and LVRs versus the closest distance to observations and found no obvious pattern.

6. SUMMARY AND DISCUSSION

In this article we propose a new method to estimate spectral densities of isotropic Gaussian processes with scattered data using a flexible semiparametric S + T model, whose parameters can be estimated using ML or REML methods. We have calculated explicit expressions of the Hankel transform of the spectral density and tackled several numerical issues arising during the computation of the covariance function. Simulated annealing is used to maximize the likelihood.

To compare our method with other existing methods for estimating spectral density, we simulated observations with Matérn, polynomial Matérn, S + T, and spectral exponential spectral densities. Our simulation results showed that our method (S + T) outperforms the nonparametric kernel method in terms of estimated sill, $\mathcal{L}^2$ norms of the covariance functions, likelihood values, MSPE, and errors in the EPVs. Our method also outperforms the parametric method using the Matérn covariance model when the true model is not Matérn by all of these performance criteria. We applied all three methods to a rainfall dataset to compare the prediction performance and found that our method performed better than the kernel method and as well as the Matérn method.

Table 3. Prediction Performance Comparison for Annual Rainfall Data in the Eastern United States

	S + T	*Matérn*	*Kernel*
Median MPE	3.12	3.12	3.16
Median |log(MPE^2/EPV)|	.50	.50	.54

NOTE: We used 10 samples of 200 training stations (chosen randomly) to estimate the covariance function and predicted the values at the remaining 1,882 stations. The median over samples and prediction locations of the MPE and the empirical LVR are shown for each method.

The IPEs and LVRs are the most relevant measures of performance when our ultimate goal is interpolation to locations where there are no observations. With these criteria, the Matérn method outperforms the kernel method, although it generally has larger $\mathcal{L}^2$ norm values of the covariance function than the kernel method when the truth is not Matérn. The reason for the better prediction properties of the Matérn method is that the tail properties of the spectral function play a fundamental role in the prediction. This result points out the inappropriateness of methods that rely on minimizing the distance to empirical variograms, such as the kernel method, when the goal is interpolation.

Our method does well because it directly estimates the tail property just like the Matérn method does, while also offering more flexibility for modeling the lower frequencies. A point to note is that our method outperforms the kernel method even when the true model is spectral exponential, which has an exponential tail, whereas our method assumes an algebraic tail.

We have performed simulations with smaller number of replicates (20 and 1 instead of 200 of the 63 sites) per simulation. Using the prediction performance criteria, we found that our method outperforms the Matérn method when the number of replicates is 20, but not so when we only have one replicate of the spatial process. Thus our method should be applied with caution when a large amount of data are not available.

We applied all three methods to 40 years of annual rainfall data in the eastern United States and found that our method outperforms the kernel method in terms of prediction error and

estimated uncertainty. However, we did not find any noticeable difference in performance between our method and the Matérn method when applied to this dataset. We may need to use a dataset with a more complex middle frequency structure to demonstrate an advantage of using our method.

In this work we used equally spaced nodes for the *B*-splines and did not attempt to estimate the optimal spacing of the nodes. Our method can be modified to allow the spacing of the nodes to be estimated from the data. One possible way of doing this efficiently is to use the information contained in the kernel estimator. Even though the kernel method fails to capture the tail behavior of the spectral density, it does seem to have useful information about the mid-frequency shape of the function. We plan to address this possibility in a separate article.

One interesting extension of our work would be to use transdimensional Markov chains, such as reversible-jump Markov chains, to optimize over both parameters and number and location of nodes. We would need a prior distribution that discourages large numbers of knots when using the Markov chain approach.

[Received May 2006. Revised January 2007.]

REFERENCES

Abramowitz, M., and Stegun, I. (1965), *Handbook of Mathematical Functions* (9th ed.), New York: Dover.

Akaike, H. (1974), "A New Look at the Statistical Identification Model," *IEEE Transactions on Automatic Control*, 19, 716–723.

de Boor, C. (2001), *A Practical Guide to Splines*, New York: Springer.

Genton, M. G., and Gorsich, D. J. (2002), "Nonparametric Variogram and Covariogram Estimation With Fourier–Bessel Matrices," *Computational Statistics & Data Analysis*, 41, 47–57.

Givens, G. H., and Hoeting, J. A. (2005), *Computational Statistics*, New York: Wiley.

Groisman, P. Y. (2000), "Data Documentation for TD-3721: Gridded US Daily Precipitation and Snowfall Time Series," National Climatic Data Center, Asheville, NC.

Hall, P., Fisher, N., and Hoffmann, B. (1994), "On the Nonparametric Estimation of Covariance Functions," *The Annals of Statistics*, 22, 2115–2134.

Handcock, M. S., and Wallis, J. R. (1994), "An Approach to Statistical Spatial Temporal Modeling of Meteorological Fields," *Journal of the American Statistical Association*, 89, 368–378.

Im, H. K. (2005), "Two Problems in Environmetrics," unpublished doctoral thesis, University of Chicago, available at *http://galton.uchicago.edu/˜stein/supmaterial.html*.

Im, H. K., Stein, M. L., and Zhu, Z. (2006), "Semiparametric Estimation of Spectral Densities With Scattered Data," technical report, University of Chicago, Center for Integrating Statistical and Environmental Sciences, available at *http://galton.uchicago.edu/˜stein/supmaterial.html*.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman & Hall.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Shapiro, A., and Botha, J. D. (1991), "Variogram Fitting With a General Class of Conditionally Nonnegative Definite Functions," *Computational Statistics & Data Analysis*, 11, 87–96.

Siegman, A. E. (1977), "Quasi-Fast Hankel Transform," *Optics Letters*, 1, 13–15.

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

——— (2002), "The Screening Effect in Kriging," *The Annals of Statistics*, 30, 298–323.

Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions*, Vol. 1, New York: Springer-Verlag.

This article has been cited by:

1. Zhengyuan Zhu, Yichao Wu. 2010. Estimation and Prediction of a Class of Convolution-Based Spatial Nonstationary Models for Large Spatial DataEstimation and Prediction of a Class of Convolution-Based Spatial Nonstationary Models for Large Spatial Data. *Journal of Computational and Graphical Statistics* **19**:1, 74-95. [Abstract] [PDF] [PDF Plus] [Supplementary material]