Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Approximate Likelihood for Large Irregularly Spaced Spatial Data

Montserrat Fuentes

To cite this article: Montserrat Fuentes (2007) Approximate Likelihood for Large Irregularly Spaced Spatial Data, Journal of the American Statistical Association, 102:477, 321-331, DOI: 10.1198/016214506000000852

To link to this article: http://dx.doi.org/10.1198/016214506000000852

Published online: 01 Jan 2012.

Approximate Likelihood for Large Irregularly Spaced Spatial Data

Montserrat FUENTES

Likelihood approaches for large, irregularly spaced spatial datasets are often very difficult, if not infeasible, to implement due to computational limitations. Even when we can assume normality, exact calculations of the likelihood for a Gaussian spatial process observed at n locations requires $O(n^3)$ operations. We present a version of Whittle's approximation to the Gaussian log-likelihood for spatial regular lattices with missing values and for irregularly spaced datasets. This method requires $O(n \log_2 n)$ operations and does not involve calculating determinants. We present simulations and theoretical results to show the benefits and the performance of the spatial likelihood approximation method presented here for spatial irregularly spaced datasets and lattices with missing values. We apply these methods to estimate the spatial structure of sea surface temperatures using satellite data with missing values.

KEY WORDS: Anisotropy; Covariance; Fourier transform; Periodogram; Satellite data; Sea surface temperatures; Spatial likelihood; Spatial statistics.

1. INTRODUCTION

Statisticians are frequently involved in the spatial analysis of huge datasets. In this situation, calculating the likelihood function to estimate the covariance parameters or to obtain the predictive posterior density is very difficult due to computational limitations. Even if normality can be assumed, calculating the likelihood function involves $O(N^3)$ operations, where N is the number of observations. However, if the observations are on a regular complete lattice, then it is possible to compute the likelihood function with fewer calculations using spectral methods (Whittle 1954; Guyon 1982; Dahlhaus and Küsch 1987; Stein 1995, 1999). These spectral methods are based on the likelihood approximation proposed by Whittle (1954). Whittle's approximation is for Gaussian random fields observed on a regular lattice without missing observations. In practice, very often the data are irregularly spaced or are not on a complete regular lattice, so we could not use Whittle's approximation to the likelihood; thus there is need for new methodology to overcome this problem.

Spectral methods for irregular time series have been studied by, for example, Parzen (1963), Bloomfield (2000), Neave (1970), Clinger and Van Ness (1976), and Priestley (1981, p. 585), in the context of estimating the periodogram of a time series. However, little research has been done on spatial likelihood approximation methods using spectral tools for irregularly spaced spatial datasets. In a spatial setting, it is worth mentioning the simple likelihood approximation introduced by Vecchia (1988), which is based on partitioning the data into clusters and assuming that the clusters are conditionally independent. Pardo-Igúzquiza and Dowd (1997) wrote a computer program for Vecchia's approximation method. Stein, Chi, and Welty (2004) adapted Vecchia's approach to approximate the restricted likelihood of a Gaussian process and showed that Vecchia's approximation gives unbiased estimating equations. Caragea (2003) presented a similar clustering framework for likelihood approximation in which the clusters were assumed to be conditionally independent after conditioning on the cluster mean. For Markov fields, Besag (1974) proposed a coding method for parameter estimation. Besag and Moran (1975) introduced an exact likelihood method for rectangular autonormal processes.

The new methods presented in this article assume that the locations at which we have missing data are not random but instead are due to the spatial design of the data or other nonrandom phenomena, which seems to be the case in most environmental problems involving large spatial datasets with missing data. For instance, when working with monitoring air pollution data or meteorologic data from weather stations, the location of the sites is not random; thus, the location of the sites without data is generally not random either. Another example is satellite data. For instance, satellite microwave radiometers that measure sea surface temperature (SST) cannot provide information over land (only over water), and clearly the locations with land surface are not random. The important feature of microwave retrievals is that SST can be measured through clouds. In this work we apply our methods to estimate the spatial structure of SST fields using the tropical rainfall measuring mission (TRMM) microwave imager (TMI) satellite data for the Pacific Ocean. Likelihood methods are often not feasible, because satellite data are very large datasets. Standard spectral methods also cannot be applied, because of the missing data over land surfaces.

In Section 3 we present an approach to approximate the likelihood for Gaussian lattice processes with missing values. In Section 4 we propose a method to approximate the Gaussian likelihood for irregularly spaced datasets, and in Section 5 apply our approximated likelihood method to estimate the spatial structure of SST fields using TMI satellite data. We finish with a discussion in Section 6.

2. SPECTRAL DOMAIN

A random field Z in $\mathbb{R}^2$ is called weakly stationary if it has finite second moments, its mean function is constant, and it has an autocovariance function C, such that $C(\mathbf{x} - \mathbf{y}) = \text{cov}\{Z(\mathbf{x}), Z(\mathbf{y})\}$. If the autocovariance function satisfies $\int_{\mathbb{R}^2} |C(\mathbf{x})|\, d\mathbf{x} < \infty$, then we can define the spectral density function, f, which is the Fourier transform of the autocovariance function,

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \exp(-i\mathbf{x}^T\boldsymbol{\omega}) C(\mathbf{x})\, d\mathbf{x}.$$

M. Fuentes is Associate Professor, Statistics Department, North Carolina State University, Raleigh, NC 27695 (E-mail: *fuentes@stat.ncsu.edu*). This research was sponsored in part by National Science Foundation grant DMS-03-53029 and by the U.S. National Oceanic and Atmospheric Administration grant NAO3NES4400015 through a Cooperative Agreement (Climate & Weather Impacts on Society and the Environment) with the Charleston Coastal Services Center and the National Climatic Data Center.

Journal of the American Statistical Association
March 2007, Vol. 102, No. 477, Theory and Methods
DOI 10.1198/016214506000000852

If Z is observed only at N uniformly spaced spatial locations Δ units apart, then the spectrum of observations of the sample sequence $Z(\Delta\mathbf{x})$, for $\mathbf{x} \in \mathbb{Z}^2$, is concentrated within the finite-frequency band $-\pi/\Delta \le \boldsymbol{\omega} < \pi/\Delta$ (aliasing phenomenon). The spectral density f_Δ of the process on the lattice can be written in terms of the spectral density f of the continuous process Z as

$$f_\Delta(\boldsymbol{\omega}) = \sum_{Q\in\mathbb{Z}^2} f\left(\boldsymbol{\omega} + \frac{2\pi Q}{\Delta}\right) \tag{1}$$

for $\boldsymbol{\omega} \in \Pi_\Delta^2 = [-\pi/\Delta, \pi/\Delta]^2$.

We estimate the spectral density of a lattice process, observed on a grid $(n_1 \times n_2)$, where $N = n_1 n_2$, with the periodogram,

$$I_N(\boldsymbol{\omega}) = (2\pi)^{-2}(n_1 n_2)^{-1}\left|\sum_{s_1=1}^{n_1}\sum_{s_2=1}^{n_2} Z(\mathbf{s})\exp\{-i\mathbf{s}^T\boldsymbol{\omega}\}\right|^2. \tag{2}$$

We compute (2) for $\boldsymbol{\omega}$ in the set of Fourier frequencies $2\pi\mathbf{f}/\mathbf{n}$, where $\mathbf{f}/\mathbf{n} = (f_1/n_1, f_2/n_2)$, and $\mathbf{f} \in J_N$, for

$$J_N = \left\{\left\lfloor -\frac{n_1-1}{2}\right\rfloor, \ldots, n_1 - \left\lfloor\frac{n_1}{2}\right\rfloor\right\} \times \left\{\left\lfloor -\frac{n_2-1}{2}\right\rfloor, \ldots, n_2 - \left\lfloor\frac{n_2}{2}\right\rfloor\right\}. \tag{3}$$

For large datasets, calculating the determinants in the likelihood function can be often infeasible. Spectral methods could be used to approximate the likelihood and obtain the maximum likelihood estimates (MLEs) of the covariance parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$. These spectral methods are based on Whittle's (1954) approximation to the Gaussian negative log-likelihood,

$$\frac{N}{(2\pi)^2}\sum \log f(\boldsymbol{\omega}) + I_N(\boldsymbol{\omega}) f(\boldsymbol{\omega})^{-1}, \tag{4}$$

where the sum is evaluated at the Fourier frequencies, I_N is the periodogram, and f is the spectral density of the lattice process. The approximated likelihood can be calculated very efficiently using the fast Fourier transform; this approximation requires only $O(N\log_2 N)$ operations. Simulation studies that we conducted seem to indicate that N must be at least 100 to get good estimated MLE parameters using Whittle's approximation.

The asymptotic covariance matrix of the MLE estimates of $\theta_1, \ldots, \theta_r$ is

$$\left\{\frac{2}{N}\left[\frac{1}{4\pi^2}\int_{[-\pi,\pi]}\int_{[-\pi,\pi]} \frac{\delta \log f(\boldsymbol{\omega})}{\delta\theta_j}\frac{\delta \log f(\boldsymbol{\omega})}{\delta\theta_k}\, d\boldsymbol{\omega}\right]^{-1}\right\}_{jk}. \tag{5}$$

This is much easier to compute than the inverse of the Fisher information matrix.

3. INCOMPLETE LATTICES

In this section we introduce spectral methods to approximate the likelihood for spatial processes observed on incomplete lattices. We propose an estimated spectral density for the process of interest on the incomplete lattice. We study the asymptotic properties of the estimated spectrum and the potential impact of this approximation on the likelihood approximation.

Consider Z to be a lattice process with spectral density f_Z. We assume that Z is a weakly stationary real-valued Gaussian process with mean 0 and finite moments. The process Z is defined on a rectangle $P_N = \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$ of sample size $N = n_1 n_2$. The covariance c of the process Z satisfies the following condition:

[a.1] $\Sigma_{\mathbf{x}}[1 + \|\mathbf{x}\|]|c(\mathbf{x})| < \infty$,

where $c(\mathbf{x}) = \operatorname{cov}\{Z(\mathbf{x}+\mathbf{y}), Z(\mathbf{y})\}$ and $\|\mathbf{x}\|$ denotes the l^2-norm of a vector $\mathbf{x} = (x_1, x_2)$ on the two-dimensional integer lattice. Thus the spectral density of Z exits and has uniformly bounded first derivatives.

The process Z is not observed directly; rather, we observe Y, an amplitude-modulated version of Z for the observations on the grid. We write

$$Y(\mathbf{x}) = g(\mathbf{x}/\mathbf{n})Z(\mathbf{x}), \tag{6}$$

where $\mathbf{x}/\mathbf{n} = (x_1/n_1, x_2/n_2)$.

The function g satisfies the following condition:

[a.2] $g(\mathbf{u})$ is bounded in $\mathbf{u}$; for any $\mathbf{u}$ in the two-dimensional integer lattice, it is of bounded variation and vanishes for $\mathbf{u}$ outside a bounded domain A.

The results in this article hold for general g functions satisfying condition [a.2], but without lost of generality, in the remainder of this article we consider g defined as

$$g(\mathbf{x}_j/\mathbf{n}) = \begin{cases} 0 & \text{if } Z(\mathbf{x}_j) \text{ is missing at location } \mathbf{x}_j \\ 1 & \text{if } Z(\mathbf{x}_j) \text{ is observed at location } \mathbf{x}_j. \end{cases} \tag{7}$$

The function g is assumed to be deterministic, which implies that the locations with missing values are not random.

We propose the following estimate of the spectral density of Z:

$$\tilde{I}_Z(\boldsymbol{\omega}) = \frac{1}{H_2(\mathbf{0})}\left|\sum_{i=1}^{N}\left(Y(\mathbf{x}_i) - g(\mathbf{x}_i/\mathbf{n})\tilde{Z}\right)\exp\{-i\boldsymbol{\omega}\mathbf{x}_i\}\right|^2,$$

where $H_j(\boldsymbol{\lambda}) = (2\pi)^2\sum_{i=1}^N g^j(\mathbf{x}_i/\mathbf{n})e^{i\boldsymbol{\lambda}^T\mathbf{x}_i}$; then $H_2(\mathbf{0}) = (2\pi)^2\sum_{i=1}^N g(\mathbf{x}_i/\mathbf{n})^2$ and

$$\tilde{Z} = \left(\sum_{i=1}^{N} Y(\mathbf{x}_i)\right)\Big/\left(\sum_{i=1}^{N} g(\mathbf{x}_i/\mathbf{n})\right).$$

If $g(\mathbf{x}_i/\mathbf{n}) = 1$ for all $\mathbf{x}_i$ in P_N, then $Y \equiv Z$ on the lattice, and $\tilde{I}_N$ reduces to the standard definition of the periodogram. When g takes same zero values (due to missing data), the difference between the traditional periodogram for Z, I_N as in (2), and the new definition given here, $\tilde{I}_N$, is a multiplicative factor, $(2\pi)^2 n_1 n_2/H_2(\mathbf{0})$. This is the bias adjustment that must be done to the periodogram function due to the missing values.

We now study the asymptotic properties of this estimate of f_Z, as $N \to \infty$ (increasing domain asymptotics), under conditions [a.1] and [a.2]. The expected value of $\tilde{I}_Z$ is

$$E[\tilde{I}_Z(\boldsymbol{\omega})] = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_Z(\boldsymbol{\omega}-\boldsymbol{\phi})|H_1(\boldsymbol{\phi})|^2\, d\boldsymbol{\phi}. \tag{8}$$

Thus $E[\tilde{I}_Z(\boldsymbol{\omega})]$ is a weighted integral of $f_Z(\boldsymbol{\omega})$. Asymptotically,

$$E[\tilde{I}_Z(\boldsymbol{\omega})] = f_Z(\boldsymbol{\omega}) + O(N^{-1}). \tag{9}$$

This result is obtained by using (3.12) of Brillinger (1970), because $\tilde{I}_Z(\boldsymbol{\omega})$ is the periodogram of a tapered version of Z.

Sharp changes in g make its Fourier transform and the squared modulus of its Fourier transform exhibit side lobes. The scatter associated with a large number of missing values creates very large side lobes in (8). Even if asymptotically the bias is negligible by (9), it could have some impact for small samples.

We now obtain the asymptotic variance for $\tilde{I}_Z$,

$$\text{var}\{\tilde{I}_Z(\boldsymbol{\omega})\} = |H_2(\mathbf{0})|^{-2}\{H_2(\mathbf{0})^2 + H_2(2\boldsymbol{\omega})^2\}f_Z(\boldsymbol{\omega})^2 + O(N^{-1}). \quad (10)$$

This can be proven by applying theorem 5.2.8 of Brillinger (1981) to $\tilde{I}_Z(\boldsymbol{\omega})$.

The quantity multiplying f_Z in (10) for the asymptotic variance is >1 when we have missing values and is 1 when there are no missing values. Thus a large number of missing values would increase the variance of the estimated spectrum.

We now use the estimated spectrum $\tilde{I}_Z(\boldsymbol{\omega})$ to approximate the likelihood of the spatial process Z (mean 0). As we have discussed in this section, the impact of the missing values in the periodogram of Z is simply a multiplicative factor, $(2\pi)^2 n_1 n_2 / H_2(\mathbf{0})$; clearly, this factor does not depend on $\boldsymbol{\omega}$. This new periodogram $\tilde{I}_Z(\boldsymbol{\omega})$ is asymptotically uniformly unbiased (as a function of $\boldsymbol{\omega}$), assuming that g satisfies [a.2]. The simplicity of this relationship between the periodogram with and without missing data allows us to preserve the asymptotic properties of the Whittle approximated likelihood function (Guyon 1982) when it is applied to processes with missing data. By proposition 1 of Guyon (1982), the estimated negative log-likelihood turns out to be

$$L_Z = \frac{N}{(2\pi)^2}\sum_{\mathbf{j}\in J_N}\left\{\log f_Z(2\pi\mathbf{j}/\mathbf{n}) + \tilde{I}_Z(2\pi\mathbf{j}/\mathbf{n})(f_Z(2\pi\mathbf{j}/\mathbf{n}))^{-1}\right\},$$

which (as $N \to \infty$) converges to the exact negative log-likelihood of Z (up to a constant),

$$\mathcal{L}_Z = \frac{1}{2}\log|\Sigma_N| + Z^T\Sigma_N^{-1}Z.$$

We have

$$E_{P_{\boldsymbol{\theta}}} \sup_{k=0,1,2} \sup_{\boldsymbol{\theta}} \left|L_Z^{(k)} - \mathcal{L}_Z^{(k)}\right| = O(\max(n_1, n_2)), \quad (11)$$

where $P_{\boldsymbol{\theta}}$ is a prior distribution for $\boldsymbol{\theta}$. L_Z requires only $O(N \times \log_2 N)$ operations. In addition, from Guyon (1982, props. 3 and 4), under assumptions (b.1) and (b.2) for f_Z (see Sec. A.3), our approximate MLE is consistent and asymptotically normal.

3.1 Simulation Study for Incomplete Lattices

To understand the performance with finite samples of our estimated likelihood function on an incomplete lattice, we conducted several simulation studies. We simulated a spatial lattice (15×15) with 15% missing values. The simulated process of interest is a stationary Gaussian spatial process with an exponential covariance function C, $C(\mathbf{h}) = \sigma e^{-|\mathbf{h}|/\rho}$, the sill parameter ($\sigma$) 2, and the range ($\rho$) 3. We calculated $\tilde{I}_Z$ (with g being an indicator function, i.e., using 0's to fill in for the missing values) and obtained the approximated likelihood function of the range and sill parameters using the spectral approach introduced here. We repeated the experiment 50 times by simulating 50 realizations of the spatial process. Table 1 presents results to study the performance of our likelihood estimates. The table gives the estimated parameters (mean from 50 simulations), Monte Carlo (MC) standard deviations (SDs) from the 50 simulations, and mean of the 50 estimated SDs using (5).

Table 1. Estimated Covariance Parameters Using 50 Simulations

	Exact likelihood			Approximated likelihood		
Parameter	*MC mean*	*MC SD*	$\hat{SD}$	*MC mean*	*MC SD*	$\hat{SD}$
Sill	2.1	.09	.09	1.8	.12	.18
Range	3	.12	.11	3.5	.16	.20

NOTE: The table gives the mean of the estimated parameters from the 50 simulations, the Monte Carlo SDs obtained from the 50 simulations, and the mean of the estimated SDs using asymptotic results.

With the filling-in approach, we tend to be a bit overoptimistic for the sill parameter (Table 1). But overall, the estimated pseudo-MLEs are practically the same for the sill and the range. The fact that the MC SD and the estimated SD are simular suggests that our estimates of the uncertainty associated with the parameters are also reliable. The standard errors for the sill and the range are around 0.02, which seems to indicate that in our setting 50 simulations might be sufficient.

We repeated this experiment with different percentages of missing data and found that these results no longer held when datasets had $>20\%$ missing values.

4. LIKELIHOOD FOR IRREGULARLY SPACED DATA

Assume that Z is a continuous Gaussian spatial process of interest, observed at M irregularly spaced locations, and f_Z is the stationary spectral density of Z. We define a process Y at location $\mathbf{x}$ as the integral of Z in a block of area Δ^2 centered at $\mathbf{x}$,

$$Y(\mathbf{x}) = \Delta^{-2}\int h(\mathbf{x}-\mathbf{s})Z(\mathbf{s})\,d\mathbf{s}, \quad (12)$$

where for $\mathbf{u} = (u_1, u_2)$ we have

$$h(\mathbf{u}) = \begin{cases} 1 & \text{if } |u_1| < \Delta/2,\ |u_2| < \Delta/2 \\ 0 & \text{otherwise.} \end{cases}$$

Then Y is also a stationary process with spectral density f_Y given by

$$f_Y(\boldsymbol{\omega}) = \Delta^{-2}|\Gamma(\boldsymbol{\omega})|^2 f_Z(\boldsymbol{\omega}),$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2)$ and $\Gamma(\boldsymbol{\omega}) = \int h(\mathbf{u})e^{-i\boldsymbol{\omega}\mathbf{u}} = [2\sin(\Delta\omega_1/2)/\omega_1][2\sin(\Delta\omega_2/2)/\omega_2]$.

For small values of Δ, $f_Y(\boldsymbol{\omega})$ is approximately $f_Z(\boldsymbol{\omega})$, because we have

$$\lim_{\Delta\to 0}\Delta^{-2}|\Gamma(\boldsymbol{\omega})|^2 = 1.$$

By (12), $Y(\mathbf{x})$ can be treated as a continuous spatial process defined for all $\mathbf{x} \in D$. But here we consider the process Y only on a lattice ($n_1 \times n_2$) of sample size $N = n_1 n_2$; that is, the values of $\mathbf{x}$ in (12) are the centroids of the N grid cells in the lattice, where the spacing is Δ between neighboring sites (Fig. 1). Then we have that the spectral density of the lattice process Y is

$$f_{\Delta,Y}(\omega) = \sum_{Q\in\mathbb{Z}^2}|\Gamma(\omega + 2\pi Q/\Delta)|^2 f_Z(\omega + 2\pi Q/\Delta). \quad (13)$$

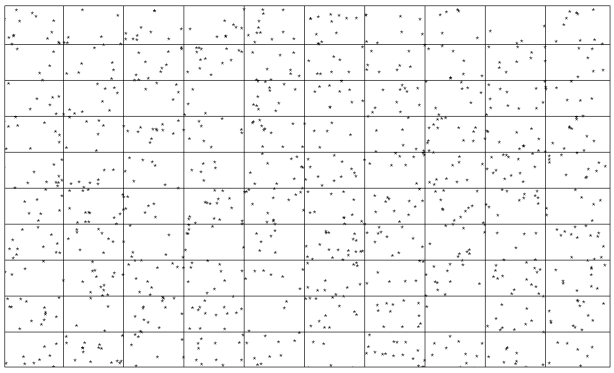

Figure 1. Simulation: Irregularly Spaced Observations. The observations are in a 10 × 10 grid, with an average of 10 observations per grid.

In practice, we truncate the sum in (13) after $2N$ terms. The justification for this is given in Section A.1.

The idea is to apply Whittle likelihood to $f_{\Delta,Y}$, written in terms of f_Z. Therefore, we can obtain the MLE for the covariance/spectral density parameters of Z by writing the likelihood of the process Y. It might help the reader to interpret this key idea in the spatial domain rather than the spectral domain.

Basic idea (Interpretation in the spatial domain). The covariance for the block averages (the lattice process Y) is defined as

$$\begin{aligned}\operatorname{cov}\big(Y(\mathbf{x}_{j_1}), Y(\mathbf{x}_{j_2})\big) &= \Delta^{-4}\int_{B_{j_1}}\int_{B_{j_2}} \operatorname{cov}(Z(\mathbf{u}), Z(\mathbf{v}))\,d\mathbf{u}\,d\mathbf{v} \\ &= \Delta^{-4}\int_{B_{j_1}}\int_{B_{j_2}} C_{\boldsymbol{\theta}}(\mathbf{u}-\mathbf{v})\,d\mathbf{u}\,d\mathbf{v},\end{aligned}$$

where $C_{\boldsymbol{\theta}}(\mathbf{u}-\mathbf{v})$ is the covariance for the continuous underlying process Z and $\boldsymbol{\theta}$ are the covariance parameters. The continuous process Z is defined in terms of a pointwise covariance $C_{\boldsymbol{\theta}}(\mathbf{h})$, but we then use the previous expression to derive the covariances of the block averages $Y(\mathbf{x}_i)$, $i = 1,\ldots,N$, in terms of the pointwise covariance $C_{\boldsymbol{\theta}}$. We then use the pointwise covariance to define a likelihood function for the parameters of the covariance function for the process Z in terms of the likelihood function of $Y(\mathbf{x}_1), Y(\mathbf{x}_2),\ldots,Y(\mathbf{x}_N)$.

To calculate the likelihood of Y, we first need to estimate $f_{\Delta,Y}$. With that purpose in mind, we define Y_N as

$$Y_N(\mathbf{x}) = \frac{1}{n_{\mathbf{x}}}\sum_{\mathbf{s}_i\in J_{\mathbf{x}}} h(\mathbf{s}_i-\mathbf{x})Z(\mathbf{s}_i), \tag{14}$$

where, for $\mathbf{x} = (x_1, x_2)$,

$$J_{\mathbf{x}} = \big\{\mathbf{s} = (s_1,s_2), |x_1-s_1| < \Delta/2, |x_2-s_2| < \Delta/2\big\}, \tag{15}$$

and the cardinal of this set is $|J_{\mathbf{x}}| = n_{\mathbf{x}}$. For locations $\mathbf{x}$, such that $n_{\mathbf{x}} = 0$, the value of $Y_N(\mathbf{x})$ is not known.

As the observations become more dense, the covariance of Y_N converges to the covariance of Y; see Section A.1. But the approximation of Y_N to Y is less reliable in grid cells with very few observations. Thus we apply a data taper to Y_N that gives less weight to grid cells with less observations. We define $g_1(\mathbf{x}) = n_{\mathbf{x}}/n$, with n the mean of the $n_{\mathbf{x}}$ values. This g_1 function plays a role similar to that of the g weight function (7) in the incomplete grid scenario.

We define $I_{g_1Y_N}(\boldsymbol{\omega})$ as the periodogram for the tapered process $g_1(\mathbf{x})Y_N(\mathbf{x})$,

$$I_{g_1Y_N}(\boldsymbol{\omega}) = |H_2^*(\mathbf{0})|^{-1}\left|\sum_{s_1=1}^{n_1}\sum_{s_2=1}^{n_2} g_1(\mathbf{s})Y_N(\mathbf{s})\exp\{-i\mathbf{s}^T\boldsymbol{\omega}\}\right|^2, \tag{16}$$

where $H_k^*(\boldsymbol{\lambda}) = 2\pi\sum_{j=1}^N g_1^k(\mathbf{x}_j)e^{i\boldsymbol{\lambda}^T\mathbf{x}_j}$.

The periodogram $I_{g_1Y_N}(\boldsymbol{\omega})$ is an asymptotically unbiased estimate of $f_{\Delta,Y}$. The bias is of order $O(N^{-1}) + O(\bar{n}^{-1})$, where $\bar{n}$ is the average of the n_x^2 values ($\bar{n} = \sum_{j=1}^N n_{\mathbf{x}_j}^2/N$),

$$E\big[I_{g_1Y_N}(\boldsymbol{\omega})\big] = f_{\Delta,Y}(\boldsymbol{\omega}) + O(N^{-1}) + O(\bar{n}^{-1}).$$

The proof of this result is included in the Appendix (Thm. A.1).

Thus, as long as $O(N/\bar{n})$ is a faster or equal rate of convergence than $O(1)$, we have that (see the App., Thm. A.2)

$$L_Y = \frac{N}{(2\pi)^2}\sum_{\mathbf{j}\in J_N}\big\{\log f_{\Delta,Y}(2\pi\mathbf{j}/\mathbf{n}) + I_{g_1Y_N}(2\pi\mathbf{j}/\mathbf{n})(f_{\Delta,Y}(2\pi\mathbf{j}/\mathbf{n}))^{-1}\big\} \tag{17}$$

converges to $\mathcal{L}_Y = \frac{1}{2}\log|\Sigma_N| + Y^T\Sigma_N^{-1}Y$ (the exact negative log-likelihood for Y). The order of convergence [in the sense of (11)] is $N^{1/2}$.

If M is the total number of observations of the process Z, then the calculation of L_Y requires $O(N\log_2 N + M)$ operations rather than $O(M^3)$ for the exact likelihood of Z. We choose $N \le M^{2/3}$ (with the equality only when there are not many empty cells) to satisfy that $O(N/\bar{n})$ is a faster or equal rate of convergence than $O(1)$. If we have $N = M^{2/3}$, then the number of operations for obtaining the likelihood function is $O(M^{2/3}\log_2 M)$.

4.1 Simulation Study for Irregularly Spaced Datasets

To understand the performance of the proposed likelihood approximation for an irregularly spaced dataset, we simulated 1,000 observations of a Gaussian spatial process with a stationary exponential covariance (range, .25; sill, 1). We gridded the observations in a 10×10 lattice and obtained an average of 10 observations per grid, as shown in Figure 1. We want to emphasize that in our approach we do not estimate the parameters of the block covariance for the gridded process, but instead write this block covariance (or the corresponding spectrum) in terms of the covariance parameters of the continuous underlying process Z, then estimate the parameters of the point covariance of Z. To study the performance of our approach, Table 2 compares the exact and approximate estimates of the sill and range parameters for 50 realizations of the process. There is no significant difference between the MLE and pseudo-MLE estimated parameters.

Table 2. Estimated Covariance Parameters Using 50 Simulations

Parameters	*Range*	*Sill*
TRUTH	.25	1
MLE (exact likelihood)	$.24_{(.05)}$	$.91_{(0.21)}$
MLE (spectral gridded)	$.30_{(.15)}$	$.86_{(0.30)}$

NOTE: This table gives the mean of the estimated parameters in the 50 simulations. The values in parentheses are Monte Carlo standard deviations obtained from the 50 simulations.

4.2 Modified Version of the Approximated Likelihood Function

The approach presented in this section to approximate the spatial likelihood for irregularly spaced datasets [see (17)] performs well when there is no nugget effect (measurement error). But we could improve the estimation of the nugget and also the smoothness parameter (which explains the degree of differentiability of Z; see Stein 1999) by adding to the likelihood function of Y information about the behavior of the process $Z(\mathbf{s}_i)$ within grid cells.

Thus we randomly choose m blocks (no more than 10–15% of the blocks) and treat them as if $n_{\mathbf{x}_i} = 0$ (i.e., we give them weight 0). We do not use the information from these m blocks in L_Y, the negative log-likelihood for Y.

Then we add to the negative log-likelihood, L_Y, the negative log-likelihood of each one of the m blocks (treating the blocks as independent),

$$L_Y + \sum_{j=1}^{m} \frac{1}{2} \log|\Sigma_j| + \mathbf{Z}_j^T \Sigma_j^{-1} \mathbf{Z}_j, \tag{18}$$

where $\mathbf{Z}_j$ is a vector with the $n_{\mathbf{x}_j}$ observations within block j and Σ_j is the covariance within the block written in terms of $C_{\boldsymbol{\theta}}(\mathbf{h})$ (covariance of Z). The calculation of (18) is very rapid because the blocks are small, approximately of order $M^{1/3}$.

4.3 Simulation Study for Irregular Datasets

To gain insight into the performance of the proposed likelihood approximation method with finite samples, we present several simulation studies. In our spatial approximated negative log-likelihood (18), we add a correction term with information within blocks. It is important to understand and study the impact on the pseudo-MLE parameters by choosing different numbers of blocks for the correction terms.

We start by simulating 1,000 observations from a Gaussian spatial process with a stationary Matérn covariance (nugget, .25; range, .25; smoothness parameter, 3; partial sill, 1),

$$C(\mathbf{h}) = \sigma_0 I(\mathbf{h}) + \frac{\sigma_1}{2^{\nu-1}\Gamma(\nu)} \left(2\nu^{1/2}|\mathbf{h}|/\rho\right)^{\nu} \mathcal{K}_{\nu}\left(2\nu^{1/2}|\mathbf{h}|/\rho\right), \tag{19}$$

where $\mathcal{K}_{\nu_s}$ is a modified Bessel function and $\boldsymbol{\theta} = (\sigma_0, \nu, \sigma_1, \rho)$. $I(\mathbf{h})$ is an indicator function that takes the value 1 when $\mathbf{h} = (0, 0)$ and 0 otherwise. The nugget parameter is σ_0 (microscale variation). The parameter ρ measures how the correlation decays with distance; generally this parameter is called the *range*. The partial sill, σ_1, is the total variance of the process minus the nugget. The parameter ν measures the degree of smoothness of the process Z. The higher the value of ν, the smoother the Z; for example, when $\nu = \frac{1}{2}$, we get the exponential covariance function. In the limit as $\nu \to \infty$, we get the Gaussian covariance. We gridded the observations in a 10×10 lattice (as in Fig. 1) and simulated 50 realizations of the spatial process. Each table in this section presents results based on 50 simulations; the parameters and standard deviations shown in Tables 4–8 and Table 10 are the mean and Monte Carlo standard deviations from 50 realizations.

Table 3. Number of Blocks Used for the Correction Term in the Spectral Likelihood, Along With an Average of the Number of Points in Each Block (in parentheses)

No. of blocks	$N = 25$	$N = 100$	$N = 225$	$N = 400$
5%	$1_{(40)}$	$5_{(10)}$	$11_{(4)}$	$20_{(2)}$
10%	$2_{(40)}$	$10_{(10)}$	$22_{(4)}$	$40_{(2)}$
15%	$3_{(40)}$	$20_{(10)}$	$44_{(4)}$	$80_{(2)}$

In the next simulations we study the impact of the number of blocks used in the correction term in (18) on the approximated MLE parameters. We calculate the likelihood for the 50 realizations of the process, varying the two parameters of interest: the number of total blocks, N, and m, the number of blocks used for the correction term. We considered N (number of blocks) to be 25 (corresponding to a 5×5 grid), 100 (10×10), 225 (15×15), and 400 (20×20). In each case m represented 5% of the blocks, 10% of the blocks, and 20% of the blocks.

Table 3 gives the different designs used in our simulation study, varying N and m. Each block had approximately 40 observations, when N was 25 versus only 2 when N was 400.

Table 4 presents the approximated MLE for the nugget parameter using the 12 different designs introduced in Table 3. The Monte Carlo standard deviations from 50 simulations are given in parentheses. The parameter was obtained using the likelihood approximation proposed in Section 4.2. The designs with $N < 100$ offer larger variability. On the other hand, designs with N very large might not very helpful for estimating the nugget, because the number of points within each block is too small. It seems that when the number of blocks for the correction term is 10%, we obtain better results for any value of N.

Table 5 studies the impact of N and m on the approximated MLE for the smoothness parameter. Estimating this parameter is always very difficult. It seems that is important to have a sufficiently large number of observations within blocks to estimate this parameter; the designs with 10 and 40 observations within block seem to perform better. The results do not appear to be very sensitive to the value of m. N seems to play a more relevant role than m.

Table 6 gives the results for the partial sill parameter. The results for the sill parameter are similar to those for the nugget. N needs to be at least 100, and designs with 10% of blocks always perform better except when N is too large, in which case the blocks only have two observations, and this design does not perform very well either.

Table 7 presents the results for the range parameter. This parameter is not very well estimated when N is <100, also the variability associated with the nugget is larger for designs with $N < 100$. The designs with 10 or 4 observations per block seem

Table 4. Approximated MLE for the Nugget Parameter and the Monte Carlo Standard Deviations (in parentheses) Using the Proposed Spectral Likelihood Method for Irregular Data

Nugget	$N = 25$	$N = 100$	$N = 225$	$N = 400$
$m = 5\%$	$.12_{(.31)}$	$.14_{(.21)}$	$.11_{(.12)}$	$0_{(.01)}$
$m = 10\%$	$.27_{(.34)}$	$.21_{(.24)}$	$.14_{(.11)}$	$.10_{(.01)}$
$m = 15\%$	$.34_{(.40)}$	$.30_{(.27)}$	$.13_{(.12)}$	$0_{(.17)}$

NOTE: The true nugget parameter is .25.

Table 5. Approximated MLE for the Smoothness Parameter, Along With the Monte Carlo Standard Deviations (in parentheses) Using the Proposed Spectral Likelihood Method for Irregular Data

Smoothness	$N = 25$	$N = 100$	$N = 225$	$N = 400$
$m = 5\%$	$4.7_{(5)}$	$5_{(4.2)}$	$6_{(4)}$	$.1_{(9)}$
$m = 10\%$	$3_{(4.1)}$	$4.7_{(3.2)}$	$1_{(2)}$	$.8_{(4)}$
$m = 15\%$	$3.3_{(3.8)}$	$3.9_{(3.4)}$	$2_{(3)}$	$5_{(2)}$

NOTE: The true smoothness parameter is 3.

Table 7. Approximated MLE for the Range Parameter, Along With the Monte Carlo Standard Deviations (in parentheses) Using the Proposed Spectral Likelihood Method for Irregular Data

Range	$N = 25$	$N = 100$	$N = 225$	$N = 400$
$m = 5\%$	$.77_{(.62)}$	$.10_{(.11)}$	$.23_{(.12)}$	$.21_{(.19)}$
$m = 10\%$	$.59_{(.31)}$	$.32_{(.20)}$	$.12_{(.15)}$	$.51_{(.27)}$
$m = 15\%$	$.15_{(.56)}$	$.21_{(.39)}$	$.19_{(.29)}$	$.13_{(.80)}$

NOTE: The true value of the range parameter is .25.

to perform better, and the results in that setting do not seem to be so sensitive to the value of m.

As another way to compare the different designs, we calculate the mean relative absolute error (MRAE), defined as

$$MRAE = Mean_i\left\{\frac{|\text{true parameter}_i - \text{Estimated parameter}_i|}{\text{true parameter}_i}\right\},$$

where i is the index for the different covariance parameters (a total of 4). Table 8 presents the results. It seems that when N is 100 (10 observations per block) we obtain better results, and designs with $m = 15\%$ and $m = 10\%$ perform better.

Overall, it seems that we need to have at least 100 blocks, and the number of observations within each block should be around 10 (not <4). Using 10% or 15% of blocks for the correction term would be our recommendation; $<10\%$ does not seem sufficient.

4.4 Comparisons With Other Likelihood Approaches for Continuous Spatial Processes

In this section we compare our spectral method for approximating the likelihood of a continuous Gaussian spatial processes with other known approaches proposed by Vecchia (1988) and Stein et al. (2004). Vecchia (1988) proposed a simple approximation to the likelihood function for spatial processes by writing the probability density function as a product of conditional densities. If p denotes the density function, then assume that $\mathbf{Z} = (Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n))$ are the observed values of the process, then we denote the joint density $p(\mathbf{z}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the unknown covariance parameters. We partition $\mathbf{Z}$ into k subvectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_k$ and define $\mathbf{Z}_{(i)} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_i)$. Thus we have

$$p(\mathbf{z}; \boldsymbol{\theta}) = p(\mathbf{z}_1; \boldsymbol{\theta}) \prod_{i=1}^{k} p\big(\mathbf{z}_{(i)} \big| \mathbf{z}_{(i-1)}; \boldsymbol{\theta}\big).$$

Vecchia proposed conditioning on not all components of $\mathbf{z}_{(i-1)}$ to calculate $p(\mathbf{z}_{(i)}|\mathbf{z}_{(i-1)}; \boldsymbol{\theta})$, but only of the nearest neighbors; thus

$$p(\mathbf{z}; \boldsymbol{\theta}) \sim p(\mathbf{z}_1; \boldsymbol{\theta}) \prod_{i=1}^{k} p\big(\mathbf{z}_{(i)} \big| \mathbf{s}_{(i-1)}; \boldsymbol{\theta}\big),$$

where $\mathbf{s}_{(i-1)}$ is a subset of $\mathbf{z}_{(i-1)}$. Vecchia considered only prediction vectors $\mathbf{z}_{(i)}$ of length 1.

Recently, Stein et al. (2004) proposed a method to reduce the computation of the approach of Vecchia (1988) and improve the efficiency of the estimated parameters, by considering prediction vectors of longer length and conditioning sets $\mathbf{s}_{(i-1)}$ that include not only the nearest neighbors, but also some rather distant observations from the prediction vectors.

One of the key contributions of Stein et al. (2004) is the significant improvement obtained by adding in the conditioning set some other observations, rather than just the nearest neighbors, and also the reduction of the computational effort by using prediction vectors of length >1. There is quite a lot of latitude about how to select the prediction and conditioning sets. Our simulation study uses a design that seems to perform very well based on the results presented by Stein et al. (2004).

We simulated 1,000 irregularly spaced observations of a spatial Gaussian process. We gridded the data in 100 blocks, each (prediction set) with approximately 10 observations. All of the observations were separated by at least a distance of .5 to avoid numerical difficulties. This setting is similar to that presented by Stein et al. (2004). The conditioning sets were of size 20, and, following the recommendation given by Stein et al., we selected 50% of those as nearest neighbors. This design would minimize the computationally effort using the method of Stein et al. (2004), and also, based on the simulations presented by Stein et al. (2004), seems to be a good choice for obtaining more reliable estimated covariance parameters. We simulated the same setting 50 times. The simulations presented by Stein et al. (2004) are based on an exponential model without nugget, and a power model with nugget. Here we study the exponential model but also examine a more general model, the Matérn model with a smoothness parameter and a nugget effect. To calculate our approximated spectral likelihood, we use the same 1,000 observations gridded in 100 blocks, having 10% of the blocks for the correction term.

We use the computer code provided by Pardo-Igúzquiza and Dowd (1997) to calculate the approximated likelihood of Vecchia (1988), with a conditioning set of size 20. Table 9 presents the results. The parameters in Table 9 are the mean from 50 realizations and the Monte Carlo standard deviations (in parentheses) from the 50 realizations. Our results are

Table 6. Approximated MLE for the Partial Sill Parameter, Along With the Estimated Standard Deviation (in parentheses) Using the Proposed Spectral Likelihood Method for Irregular Data

Partial sill	$N = 25$	$N = 100$	$N = 225$	$N = 400$
$m = 5\%$	$.47_{(.9)}$	$.81_{(.14)}$	$.89_{(.39)}$	$.70_{(.20)}$
$m = 10\%$	$.70_{(.4)}$	$.90_{(.20)}$	$.87_{(.30)}$	$.63_{(.13)}$
$m = 15\%$	$.72_{(.43)}$	$1.10_{(.11)}$	$.79_{(.32)}$	$.84_{(.21)}$

NOTE: The true partial sill parameter is 1.

Table 8. MRAE for Each Design

MRAE	$N = 25$	$N = 100$	$N = 225$	$N = 400$
$m = 5\%$	.85	.47	.43	.60
$m = 10\%$	.44	.27	.43	.68
$m = 15\%$	.28	.19	.31	.57
Average MRAE	.51	.31	.39	.61

Table 9. Comparisons of the Approach of Vecchia (1988), the Method of Stein et al. (2004), and the Spectral Likelihood Method Presented Herein for Estimating the Sill and Nugget Parameters of an Exponential Covariance

	Sill	*Range*	*Sill/Range*	*MRAE*
Truth	1	.25	4	0
Exact likelihood	$.97_{(.07)}$	$.24_{(.1)}$	4	.07
Spectral method	$.90_{(.1)}$	$.31_{(.14)}$	2.9	.34
Stein's method	$.80_{(.32)}$	$.20_{(.19)}$	4	.40
Vecchia's method	$.59_{(.3)}$	$.25_{(.1)}$	2.3	.45

NOTE: The parameters are the means from 50 realizations and the Monte Carlo standard deviations (in parentheses) from the 50 realizations.

consistent with to those of Stein et al. (2004). Having some observations in the conditioning set that are not the nearest neighbors improves the performance of the approximated likelihood method. If in the conditioning sets of size 20 we select 10 observations that are not nearest neighbors but instead are among the "past" observations (and we use prediction sets of size 10), then we can reduce the MRAE from .45 to .34. The estimated range parameter (mean from 50 realizations) of an exponential covariance (Table 9) using the approach of Vecchia (1988) with 20 neighbors is .25 (the truth is .25); when we reduce the number of neighbors to 10, we obtain a range of .44. Increasing the number of neighbors beyond 20 does not improve the results (the estimated range is .20 with 30 neighbors). The main problem is estimating the sill parameter, for which we obtain a value of .59 (truth, 1). When the number of neighbors increases to 30, the estimated sill parameter does not change, remaining .59.

The results obtained using the approach of Stein et al. (2004) are very similar to those obtained using the spectral method, at least in terms of estimating the range and sill parameters. The MRAE is .34 for the spectral method and slightly larger for the method of Stein et al. (2004) (.40). However, if we consider the ratio of the sill to range, which is the important parameter for interpolation (Stein 1999), then the approach of Stein et al. has much better performance. The truth is 4, the exact likelihood gives a value of 4, and the method of Stein et al. (2004) also gives a value of 4 (Table 9).

In terms of capturing the behavior at very short distances, the method of Stein et al. (2004) performs much better. The performance obtained with the approach of Vecchia (1988) at short distances is very poor.

We implemented the likelihood approximation approach proposed by Stein et al. (2004) using the previous design (conditioning sets of size 20 and prediction sets of size 10), but with a Matérn covariance (as in Tables 4–8). Here we also present results using the exact likelihood. Table 10 presents the results for the method of Stein et al. (2004); Tables 4–8 give the results for the spectral method. The estimated parameters are the mean of the results obtained using 50 simulations of an underlying Gaussian process with a Matérn covariance (1,000 observations, prediction sets of size 10, conditioning sets of size 20 with 10 nearest neighbors). We present sample standard deviation of the parameters from the 50 simulations in parentheses. The MRAE is .36 for the exact likelihood and .57 for the method of Stein et al. (2004). The results obtained using the spectral method proposed here are also given in Tables 4–8 (for different designs). The design with 100 blocks and 10% of blocks for the correction term has an MRAE of .27.

Table 10. Estimated Covariance Parameters, Along With Monte Carlo Standard Deviations (in parentheses)

Parameter	*Nugget*	*Partial sill*	*Range*	*Smoothness*	*MRAE*
TRUTH	.25	1	.25	3	0
MLE (exact likelihood)	$.24_{(.2)}$	$.9_{(.20)}$	$.50_{(.25)}$	$2.0_{(3.5)}$	.36
Stein et al. (2004)	$.23_{(.1)}$	$.79_{(.43)}$	$.15_{(.4)}$	$7.8_{(7.9)}$	.57

In terms of computational considerations, Stein et al. (2004) indicated that to minimize the computational effort of the approach that they proposed, the size of the conditioning sets (c) should be twice the size of the prediction sets (d). Thus, in our setting with 100 blocks and prediction sets of 10 observations, if the conditioning sets are chosen to be of size 20, then there should be 10 nearest neighbors among these 20; then the number of floating point operations (flops) needed to calculate the likelihood approximation proposed by Stein et al. (2004) is 9×10^5. To compare the efficiency of choosing the prediction sets with more than one observation to prediction sets of size 1 as proposed by Vecchia (1988), consider now $d = 1$ and $c = 20$ (vs. $d = 10$ and $c = 20$). A total of 3×10^6 flops are needed. Thus the design suggested by Vecchia (1988) requires about three times the amount of computation as involved in the approach suggested by Stein et al. (2004). Our approach has a clear advantage in terms of computational effort; with our approach, we need only approximately 10^4 flops (see Sec. 4.2).

5. APPLICATION

Global SST fields are very useful in monitoring climate change, as oceanic boundary conditions for numerical atmospheric models, and for comparison with the SSTs produced by ocean numerical models. SSTs can be estimated from satellites using, for example, the TMI. The spatial scales and structure of SST fields are the main factor in identifying phenomena such as El Nino and La Nina that occur in the equatorial Pacific and influence weather in the Western Hemisphere. Spatial patterns of SSTs in the Pacific Ocean are also one of the main climate factors in identifying tropical cyclones (hurricanes) that form in south Mexico and strike Central America and Mexico from June to October. Studying the spatial structure of SSTs in the Pacific Ocean is also important to gain insight into the exchange of water between the north and south equatorial currents. A good understanding of the SST's spatial variability is crucial for guiding future research on the variability and predictability of the world ocean SST and the climate that it influences.

In this work we apply our methods to estimate the spatial structure of TMI SST data over the Pacific Ocean. Currently, most of the operational approaches to estimate the covariance parameters of the TMI SST fields, particularly the mesoscale and zone scale parameters (ranges of correlation; Reynolds and Smith 1994) are empirical methods, and there is no reliable measure of the uncertainty associated with the estimated parameters. Implementing likelihood methods is difficult because the satellite datasets are very large. Traditional spectral methods can not be applied directly because the data are not on a complete regular grid; the TMI sensor only retrieves SST data

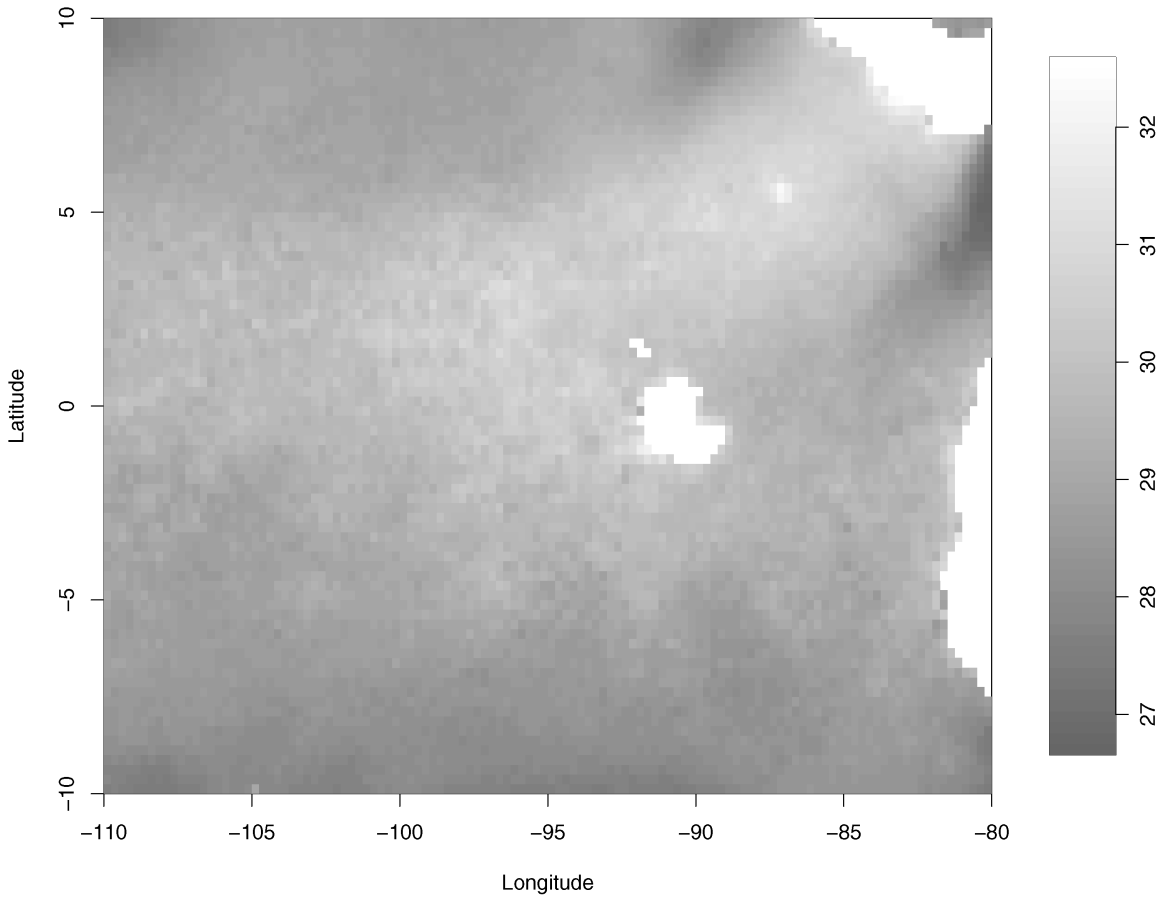

Figure 2. Satellite SSTs (in degrees Celsius) Over the Western Pacific Ocean. Missing values (in white) correspond to land surfaces (Central and South America and the Galapagos Islands).

from the surface through areas of open water (ocean or lakes) (Fig. 2). However, our methodology can be easily applied to these large satellite data.

The satellite observations in this application are obtained from a radiometer onboard the TRMM satellite. This radiometer, the TMI is well calibrated and contains lower-frequency channels required for SST retrievals. The measurement of SST through clouds by satellite microwave radiometers has been an elusive goal for many years. The TMI data have a spatial resolution of 25 km $\times$ 25 km and are available at *http://www.remss.com/tmi/tmi_browse.html*. Here we present results for March 1998. The study of the mesoscale spatial dependency for SST fields is generally done using SST anomalies (SSTA), that is, the SST field after removal of the spatial trend. The SSTA fields here are obtained from the monthly SST values after subtracting the yearly SST climatology (temporal mean for each year) at each location. In our analysis we have 4% of missing values that correspond to land surface (Central and South America and the Galapagos Islands). The domain of interest is presented in Figure 2. In this image we have 9,600 pixels (120×80).

In our analysis, we first project the spatial coordinates using a Lambert projection to take into account the curvature of the Earth when calculating spatial dependencies. Some parameters of interest when studying the SSTA spatial structure are the meridional scale (i.e., the range of correlation along the north–south direction) and the zonal scale (i.e., the range of correlation in the east–west direction). These parameters explain the potential effect of directionality in the spatial dependency (anisotropy effect). We estimate this effect by allowing for geometric anisotropy in our covariance model and estimating the rotation and stretching effects. First, we convert from Cartesian coordinates $\mathbf{x} = (x_1, x_2)$ (spatial domain) and $\boldsymbol{\omega} = (\omega_1, \omega_2)$ (spectral domain) to polar coordinates (r, θ) and (u, ϕ), where $x_1 = r\cos\theta$, $x_2 = r\sin\theta$, $\omega_1 = u\cos\phi$, and $\omega_2 = u\sin\phi$. Then the covariance function of the SSTA process, $C(x_1, x_2)$, and the corresponding spectral densities, $f(\omega_1, \omega_2)$, become $C(r, \theta)$ and $f(r, \theta)$. We incorporate a parameter α that explains the stretching effect and a parameter θ_0 that explains the rotation. We estimate these two parameters and the rest of the covariance parameters (smoothness, nugget, range, and partial sill) using the spectral likelihood approximation method presented here for incomplete regular lattices, which can be calculated very rapidly and efficiently for these large satellite datasets. The covariance function with the stretching and rotation parameters becomes $C(\alpha r, \theta + \theta_0)$, and the corresponding spectral density is $\frac{1}{|\alpha|} f(u/\alpha, \phi + \theta_0)$, where f is a Matérn spectral density. The pseudo-MLE estimated parameters and their standard deviations [eq. (5)] have been obtained using the likelihood method proposed in this article.

The anisotropy angle parameter is defined here as the azimuth angle (in radians) of the direction with greater spatial continuity, that is, the angle between the y-axis and the direction with the maximum range. It is estimated as 0 radians (0 de-

Table 11. Estimated MLE Parameters and Standard Deviations Using the Proposed Likelihood Approximation Method for Incomplete Lattices

Parameters	*Spectral MLE*	*SD*
Nugget	.001	.002
Range	312 km	70
Partial sill	.57	.02
Smoothness	1	1.2
Anisotropy rotation angle	0	.02
Anisotropy stretching	1.05	.2

grees), with a SD of .02. The anisotropy stretching parameter is defined here as the ratio between the ranges of the directions with greater and smaller continuity, i.e., the ratio between maximum and minimum ranges. Therefore, its value is always ≥ 1. In this case it is 1.05 (.2). Therefore, the fact that there are no significant rotation and/or stretching effects indicate that there is not a meridional versus zonal effect in the spatial scales. This also suggests that the water exchange between the north equatorial current and the south equatorial current is not affecting the mesoscale spatial dependency structure of the SSTA field. The nugget is 0, which is expected because the data are not real measurements. The range of correlation is 312 km, indicating that the SSTA field has significant spatial continuity. The variability of the process is relatively small, .57°C^2. The smoothness parameter is 1 (1.2). Estimating the smoothness parameter tends to be difficult in practice; this is also the case here, because the uncertainty associated with this parameter is quite large. The methods presented here are being applied to other time windows and oceans.

6. DISCUSSION

In this article we have introduced a new spectral method for estimating the covariance parameters of a spatial process. Our approach is likelihood-based and offers enormous computational benefits. This is the first time that an extension of Whittle approximated likelihood has been applied to lattice data with missing values and irregularly spaced datasets. Some other alternative spectral approaches for irregularly sampled processes are more computationally expensive:

- A spectral likelihood based on a periodogram for irregularly sampled processes obtained using generalized prolate spheroidal sequences (Bronez 1988)
- the EM algorithm (Dempster, Laird, and Rubin 1977), which is a very well-known technique to find maximum likelihood estimates in parametric models with incomplete data. In the EM algorithm, we could first impute the values of the process at the locations in the grid where we have no data and then calculate the complete-data likelihood using spectral methods. We would need to iterate through these two steps.

The spectral likelihood approach presented here is attractive because of its simplicity and because it is very computationally efficient and fast compared with any other known likelihood approximation method for spatial data that gives consistent estimates.

The weight function g introduced here to handle incomplete lattices could have a more sophisticated structure than that used in this article. For instance, instead of just taking values 1 and 0, g could go smoothly from 1 to 0 using a cosine function or a spline function to better capture the transition zones in the areas with missing values. This can be particularly helpful when missing values are clustered together (e.g., clouds in AVHRR satellite data). Alternatively, we could use a moving average of the observations to fill in for the missing values of Z. Using then the approach introduced in Section 4 to obtain an approximation to the likelihood function of Z, by defining the process Y in (6) as a moving average of the observed values of Z [as in (14)] rather than an amplitude-modulated process.

In terms of calculating the likelihood for lattice processes, note that if we use a correlation function on the lattice rather than a continuous one, such as the Matérn and others suggested in this article then there are other approaches to approximating the likelihood (e.g., Besag and Moran 1975; Rue 2001; Rue and Tjerlmeland 2002) that can be evaluated using only $O(n^{3/2})$ flops or $O(n^{3/2}\log n)$ depending on the setup, rather than $O(n^3)$ using a continuous covariance. For a review of all these approximation techniques on the lattice, we recommend the new book by Rue and Held (2005, chap. 5).

In this article we have assumed that the spatial process has a stationary covariance function. There are many models for nonstationary covariances (e.g., Sampson and Guttorp 1992; Fuentes 2001, 2002). In particular, by modeling the nonstationary process as a mixture of independent stationary processes (Fuentes 2001, 2005), the likelihood framework presented herein can be easily adopted, because then the likelihood function can be written in terms of stationary likelihood functions.

APPENDIX: PROOFS OF THEOREMS

A.1 Truncation of $f_{\Delta,Y}$

Assume that for large frequencies (as $|\boldsymbol{\omega}| \to \infty$), the spectral density of a continuous spatial process Z satisfies

$$f_Z(\boldsymbol{\omega}) \propto |\boldsymbol{\omega}|^{-\alpha}, \quad \text{with } \alpha > 2. \tag{A.1}$$

The spectral density models generally used for continuous spatial processes (i.e., Matérn) satisfy condition (A.1). Under this condition, we need to prove that the residual term in the expression for $f_{\Delta,Y}$ given in (13), when we truncate the sum the sum after $2N$ terms, is negligible compared with $O(N^{-1})$, which is the bias of our estimated function of $f_{\Delta,Y}$ (Sec. 4).

The spectral density of the lattice process Y, $f_{\Delta,Y}(\boldsymbol{\omega})$ for $\boldsymbol{\omega} \in [-\pi/\Delta, \pi/\Delta]^2$, is defined in (13) in terms of f_Z. Here we study the order of the residual term R,

$$f_{\Delta,Y}(\boldsymbol{\omega}) = \sum_{Q\in N_Y} |\Gamma(\boldsymbol{\omega} + 2\pi Q/\Delta)|^2 f_Z(\boldsymbol{\omega} + 2\pi Q/\Delta) + R(\boldsymbol{\omega}, N_Y), \tag{A.2}$$

where $N_Y = \{(q_1, q_2) \in \mathbb{Z}^2; -n_1 < q_1 < n_1, -n_2 < q_2 < n_2\}$. We have that

$$\begin{aligned}
&\sum_{q_1=n_1}^{+\infty}\sum_{q_2=n_2}^{+\infty} \big|\Gamma\big(\boldsymbol{\omega} + 2\pi(q_1,q_2)/\Delta\big)\big|^2 f_Z\big(\boldsymbol{\omega} + 2\pi(q_1,q_2)/\Delta\big) \\
&\quad \leq \int_{-\pi/\Delta+2\pi n_1/\Delta}^{+\infty}\int_{-\pi/\Delta+2\pi n_2/\Delta}^{+\infty} |\omega_1|^{-1}|\omega_2|^{-1} \\
&\qquad\qquad \times |\omega_1^2+\omega_2^2|^{-\alpha/2}\, d\omega_1\, d\omega_2 \\
&\quad = O\big(N^{-\alpha/2}\big).
\end{aligned} \tag{A.3}$$

We also have

$$\sum_{q_1=-n_1}^{-\infty}\sum_{q_2=-n_2}^{-\infty}\left|\Gamma\left(\boldsymbol{\omega}+2\pi(q_1,q_2)/\Delta\right)\right|^2 f_Z\left(\boldsymbol{\omega}+2\pi(q_1,q_2)/\Delta\right)$$
$$\le\int_{-\infty}^{\pi/\Delta-2\pi n_1/\Delta}\int_{-\infty}^{\pi/\Delta-2\pi n_2/\Delta}|\omega_1|^{-1}|\omega_2|^{-1}\times|\omega_1^2+\omega_2^2|^{-\alpha/2}\,d\omega_1\,d\omega_2$$
$$=O\left(N^{-\alpha/2}\right). \tag{A.4}$$

Similarly, the same bound is obtained for the other six subregions, with the pair (q_1,q_2) in $N_1=\{(q_1,q_2)\in\mathbb{Z}^2; q_1>n_1 \text{ and } q_2<-n_2\}$, $N_2=\{(q_1,q_2)\in\mathbb{Z}^2; q_1<-n_1 \text{ and } q_2>n_2\}$, $N_3=\{(q_1,q_2)\in\mathbb{Z}^2; |q_1|<n_1 \text{ and } q_2<-n_2\}$, $N_4=\{(q_1,q_2)\in\mathbb{Z}^2; |q_1|<n_1 \text{ and } q_2<-n_2\}$, $N_5=\{(q_1,q_2)\in\mathbb{Z}^2; q_1<-n_1 \text{ and } |q_2|<n_2\}$, or $N_6=\{(q_1,q_2)\in\mathbb{Z}^2; q_1>n_1 \text{ and } |q_2|<n_2\}$.

Therefore, the order of convergence to 0 of the residual term in (A.2) is faster than $O(N^{-1})$, which is the bias of $I_{g_1Y_N}$ [defined in (16)]; it is our estimate of $f_{\Delta,Y}$.

A.2 Asymptotic Properties of Estimated Spectrum

Definitions. A *spatially random pattern* on a domain of interest D is synonymous with a homogeneous Poisson process in D, this means that any location on D has equal probability of receiving a point (see, e.g., Cressie 1993, p. 586).

A *spatially regular pattern* on D indicates regular spacings between points (see, e.g., Cressie 1993, p. 590).

Theorem A.1. Consider a continuous weakly stationary spatial process Z observed at M locations in a domain D of interest. We define two lattice processes, Y [as in (12)] and Y_N [as in (14)], both written in terms of the process Z, and defined on a lattice $n_1\times n_2$ covering D, with spacing Δ between neighboring observations. We define $f_{\Delta,Y}$, the spectral density of the process Y. We propose $I_{g_1Y_N}$, defined in (16), as an estimate of $f_{\Delta,Y}$. As $N\to\infty$ and $\bar{n}\to\infty$, where $\bar{n}=\frac{1}{N}\sum_i n_{\mathbf{x}_i}^2$ and $n_{\mathbf{x}_i}$ is the number of observations of the process Z in the grid cell i. We assume that the locations of the $n_{\mathbf{x}_i}$ observations within each grid cell i form a spatially random or regular pattern.

We then have

$$E\left[I_{g_1Y_N}(\boldsymbol{\omega})\right]=f_{\Delta,Y}(\boldsymbol{\omega})+O(N^{-1})+O(\bar{n}^{-1}).$$

Proof.

$$I_{g_1Y_N}(\boldsymbol{\omega})=|H_2^*(\mathbf{0})|^{-1}\left|\sum_{s_1=1}^{n_1}\sum_{s_2=1}^{n_2}g_1(\mathbf{s})Y_N(\mathbf{s})\exp\{-i\mathbf{s}^T\boldsymbol{\omega}\}\right|^2.$$

First, we need to study the convergence of the second-order moments of Y_N to those of the process Y as the observations become more dense, that is, as each $n_{\mathbf{x}_i}\to\infty$. Assume that $C_{\boldsymbol{\theta}}(\mathbf{h})$ is the stationary covariance of the process Z at a distance $\mathbf{h}$, with parameters $\boldsymbol{\theta}$. We have

$$\operatorname{cov}\left(Y_N(\mathbf{x}_1),Y_N(\mathbf{x}_2)\right)=\frac{1}{n_{\mathbf{x}_1}n_{\mathbf{x}_2}}\sum_{\mathbf{s}_i\in J_{\mathbf{x}_1}}\sum_{\mathbf{s}_j\in J_{\mathbf{x}_2}}C_{\boldsymbol{\theta}}(\mathbf{s}_i-\mathbf{s}_j),$$

where $J_{\mathbf{x}}$ is defined in (15) and

$$\operatorname{cov}(Y(\mathbf{x}_1),Y(\mathbf{x}_2))=\Delta^{-2}\int\int h(\mathbf{u}-\mathbf{x}_1)h(\mathbf{v}-\mathbf{x}_2)C_{\boldsymbol{\theta}}(\mathbf{u}-\mathbf{v})\,d\mathbf{u}\,d\mathbf{v}$$
$$=\Delta^{-2}\int_{B_\Delta}\int_{B_\Delta}C_{\boldsymbol{\theta}}\left((\mathbf{x}_1-\mathbf{x}_2)-(\mathbf{u}-\mathbf{v})\right)d\mathbf{u}\,d\mathbf{v},$$

where $B_\Delta=[-\frac{1}{2}\Delta,\frac{1}{2}\Delta]^2$. Clearly, the covariance of Y is stationary. By the foregoing expressions for the covariance functions of Y_N and Y, and the fact that the locations of the $n_{\mathbf{x}_i}$ observations within each grid cell i is spatially random or regular, we have that the covariance of Y_N between points $\mathbf{x}_1$ and $\mathbf{x}_2$ converges to $\operatorname{cov}(Y(\mathbf{x}_1),Y(\mathbf{x}_2))$ as $n_{\mathbf{x}_1}\to\infty$ and $n_{\mathbf{x}_2}\to\infty$. The order of convergence is $O(n_{\mathbf{x}_1}^{-1}n_{\mathbf{x}_2}^{-1})$.

Thus it is straightforward to see that the order of convergence of the covariance of the tapered process g_1Y_N to the covariance of g_1Y is $O(n^{-1}n^{-1})$ (n average of the $n_{\mathbf{x}_i}$ values). Then

$$E[g_1(\mathbf{x}_1)Y_N(\mathbf{x}_1)g_1(\mathbf{x}_2)Y_N(\mathbf{x}_2)]=E[g_1(\mathbf{x}_1)Y(\mathbf{x}_1)g_1(\mathbf{x}_2)Y(\mathbf{x}_2)]+O(n^{-1}n^{-1}),$$

uniformly in $\mathbf{x}_1$ and $\mathbf{x}_2$. This is a uniform convergence, because C is a uniformly bounded function and we assume that the variance of Z is finite.

Thus, because

$$E\left[I_{g_1Y_N}(\boldsymbol{\omega})\right]=|H_2^*(\mathbf{0})|^{-1}\left|\sum_{s_1=1}^{n_1}\sum_{s_2=1}^{n_2}g_1(\mathbf{s})Y_N(\mathbf{s})\exp\{-i\mathbf{s}^T\boldsymbol{\omega}\}\right|^2,$$

we have

$$E\left[I_{g_1Y_N}(\boldsymbol{\omega})\right]=E\left[I_{g_1Y}(\boldsymbol{\omega})\right]+\epsilon_N(\boldsymbol{\omega}), \tag{A.5}$$

where I_{g_1Y} is the periodogram of the tapered process g_1Y. As $N\to\infty$, $E[I_{g_1Y}(\boldsymbol{\omega})]$ in the foregoing expression converges uniformly to $f_{\Delta,Y}(\boldsymbol{\omega})$,

$$E\left[I_{g_1Y}(\boldsymbol{\omega})\right]=f_{\Delta,Y}(\boldsymbol{\omega})+O(N^{-1}),$$

and the residual term $\epsilon_N(\boldsymbol{\omega})$ in expression (A.5) is of the following order:

$$\epsilon_N(\boldsymbol{\omega})\le|H_2^*(\mathbf{0})|^{-1}\left|\sum_{s_1=1}^{n_1}\sum_{s_2=1}^{n_2}\frac{1}{n^2}\exp\{-i\mathbf{s}^T\boldsymbol{\omega}\}\right|^2=O(\bar{n}^{-1});$$

therefore,

$$E\left[I_{g_1Y_N}(\boldsymbol{\omega})\right]=f_{\Delta,Y}(\boldsymbol{\omega})+O(N^{-1})+O(\bar{n}^{-1}).$$

A.3 Convergence to the Exact Likelihood Function

Assume that the spectral density of the lattice process Y [as in (12)], $f_{\Delta,Y}$ with parameters $\boldsymbol{\theta}$, satisfies the following conditions:

(b.1) $f_{\Delta,Y}(\boldsymbol{\omega})$ is rational with respect to $e^{i\boldsymbol{\omega}}$, without 0's or poles.
(b.2) The second derivative of $f_{\Delta,Y}$ in $\boldsymbol{\theta}$ is continuous in $\boldsymbol{\theta}$.

All classical spectral density models satisfy these two conditions.

Theorem A.2. Assume that the order of convergence to 0 of $\bar{n}$ as $N\to\infty$ is at least $O(N^{-1})$; this means that $O(N/\bar{n})$ is a faster or equal rate of convergence than $O(1)$. Then, under conditions (b.1) and (b.2), we have that

$$L_Y=\frac{N}{(2\pi)^2}\sum_{\mathbf{j}\in J_N}\left\{\log f_{\Delta,Y}(2\pi\mathbf{j}/\mathbf{n})+I_{g_1Y_N}(2\pi\mathbf{j}/\mathbf{n})(f_{\Delta,Y}(2\pi\mathbf{j}/\mathbf{n}))^{-1}\right\} \tag{A.6}$$

converges to $\mathcal{L}_Y=\frac{1}{2}\log|\Sigma_N|+Y^T\Sigma_N^{-1}Y$ (exact likelihood function for the lattice process Y), and if n_1 and n_2 are of the same order, then the rate of approximation [in the sense of (11)] is $N^{1/2}$.

Proof. By the condition that $O(N/\bar{n})$ is a faster or equal rate of convergence than $O(1)$, the proposed periodogram function, $I_{g_1Y_N}$, approximates the spectral density $f_{\Delta,Y}$ with a bias of the same order (Thm. A.1) than if we used I_{g_1Y}, the periodogram of a tapered version of Y. Then, by proposition 1 of Guyon (1982), in which the convergence of the spectral Whittle likelihood function for a tapered process, g_1Y to the exact likelihood of Y is proven, we obtain that (A.6) holds and that the order is $N^{1/2}$.

[Received August 2005. Revised May 2006.]

REFERENCES

Besag, J. E. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 36, 76–86.

Besag, J. E., and Moran, P. A. P. (1975), "On the Estimation and Testing of Spatial Interaction in Gaussian Lattice Processes," *Biometrika*, 62, 555–562.

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

Brillinger, D. R. (1970), "The Frequency Analysis of Relations Between Stationary Spatial Series," in *Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress*, ed. R. Pyke, Montreal: Canadian Mathematical Congress, pp. 39–81.

——— (1981), *Time Series: Data Analysis and Theory* (expanded ed.), San Francisco: Holden-Day.

Bronez, T. P. (1988), "Spectral Estimation of Irregularly Sampled Multidimensional Processes by Generalized Prolate Spheroidal Sequences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36, 1862–1873.

Caragea, P. (2003), "Approximate Likelihoods for Spatial Processes," doctoral dissertation, University of North Carolina, Dept. of Statistics, available at *http://www.stat.unc.edu/postscript/rs/caragea.pdf*.

Clinger, W., and Van Ness, J. W. (1976), "On Unequally Spaced Time Points in Time Series," *The Annals of Statistics*, 4, 736–745.

Cramér, H., and Leadbetter, M. R. (1967), *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*, New York: Wiley.

Cressie, N. (1993), *Satistics for Spatial Data* (rev. ed.), New York: Wiley.

Dahlhaus, R., and Küsch, H. (1987), "Edge Effects and Efficient Parameter Estimation for Stationary Random Fields," *Biometrika*, 74, 877–882.

Dempster, A. D., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, 39, 1.

Fuentes, M. (2001), "A New High-Frequency Kriging Approach for Nonstationary Environmental Processes," *Environmetrics*, 12, 469–483.

——— (2002), "Spectral Methods for Nonstationary Spatial Processes," *Biometrika*, 89, 197–210.

——— (2005), "A Formal Test for Nonstationarity of Spatial Stochastic Processes," *Journal of Multivariate Analysis*, 96, 30–55.

Fuentes, M., Chen, L., and Davis, J. M. (2005), "Modeling and Predicting Complex Space-Time Structures and Patterns of Coastal Wind Fields," *Environmetrics*, 16, 449–464.

Guyon, X. (1982), "Parameter Estimation for a Stationary Process on a d-Dimensional Lattice," *Biometrika*, 69, 95–105.

Letelier, J., Pizarro, O., Núñez, S., and Arcos, D. (2004), "Spatial and Temporal Variability of Thermal Fronts Off Central Chile (33-40S)," *Gayana*, 68, 358–362.

Marcotte, D. (1996), "Fast Variogram Computation With FFT," *Computers and Geosciences*, 22, 1175–1186.

Matérn, B. (1986), *Spatial Variation* (2nd ed.), New York: Springer-Verlag.

Neave, H. R. (1970), "Spectral Analysis of a Stationary Time Series Using Initially Scarce Data," *Biometrika*, 57, 111–122.

Pardo-Igúzquiza, E., and Dowd, P. A. (1997), "AMLE3D: A Computer Program for the Inference of Spatial Covariance Parameters by Approximate Maximum Likelihood Estimation," *Computational Geosciences*, 23, 793–805.

Park, K. A., and Chung, J. Y. (1999), "Spatial and Temporal Scale Variations of Sea Surface Temperature in the East Sea Using NOAA/AVHRR Data," *Journal of Oceanography*, 55, 271–288.

Parzen, E. (1963), "On Spectral Analysis With Missing Observations and Amplitude Modulation," *Sankhyā*, Ser. A, 25, 383–392.

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, London: Academic Press.

Reynolds, R. W., and Smith, T. M. (1994), "Improved Global Sea Surface Temperature Analyses," *Journal of Climate*, 7, 929–948.

Rue, H. (2001), "Fast Sampling of Gaussian Markov Random Fields," *Journal of the Royal Statistical Society*, Ser. B, 63, 325–338.

Rue, H., and Held, H. (2005), *Gaussian Markov Random Fields: Theory and Applications*, London: Chapman & Hall.

Rue, H., and Tjelmeland, H. (2002), "Fitting Gaussian Markov Random Fields to Gaussian Fields," *Scandinavian Journal of Statistics*, 29, 31–49.

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119.

Stein, M. L. (1995), "Fixed Domain Asymptotics for Spatial Periodograms," *Journal of the American Statistical Association*, 90, 1277–1288.

——— (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag.

Stein, M. L., Chi, Z., and Welty, L. (2004), "Approximating Likelihoods for Large Spatial Data Sets," *Journal of the Royal Statistical Society*, Ser. B, 66, 275–296.

Vecchia, A. V. (1988), "Estimation and Model Identification for Continuous Spatial Processes," *Journal of the Royal Statistical Society*, Ser. B, 50, 297–312.

Whittle, P. (1954), "On Stationary Processes in the Plane," *Biometrika*, 41, 434–449.

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