Biometrika (2019), **106**, 2, *pp.* 267–286
Printed in Great Britain
doi: 10.1093/biomet/asz004
Advance Access publication 3 April 2019

Spectral density estimation for random fields via periodic embeddings

By JOSEPH GUINNESS

Department of Statistical Science, Cornell University, 1178 Comstock Hall, Ithaca, New York 14853, U.S.A.

guinness@cornell.edu

Summary

We introduce methods for estimating the spectral density of a random field on a d-dimensional lattice from incomplete gridded data. Data are iteratively imputed onto an expanded lattice according to a model with a periodic covariance function. The imputations are convenient computationally, in that circulant embedding and preconditioned conjugate gradient methods can produce imputations in $O(n \log n)$ time and $O(n)$ memory. However, these so-called periodic imputations are motivated mainly by their ability to produce accurate spectral density estimates. In addition, we introduce a parametric filtering method that is designed to reduce periodogram smoothing bias. The paper contains theoretical results on properties of the imputed-data periodogram and numerical and simulation studies comparing the performance of the proposed methods to existing approaches in a number of scenarios. We present an application to a gridded satellite surface temperature dataset with missing values.

Some key words: Circulant embedding; Conjugate gradient; Covariance function; Gaussian process; Nonparametric estimation; Semiparametric estimation; Spatial statistics.

1. Introduction

Random fields defined on the integer lattice have wide applications in modelling gridded spatial and spatial-temporal datasets. They also form the basis for some models for non-gridded data (Nychka et al., 2015). The large sizes of modern spatial and spatial-temporal datasets entail an enormous computational burden when using traditional methods for estimating random field models. Modelling data on a grid provides a potential solution to the computational issue, since there exist some methods based on the discrete Fourier transform which can be computed efficiently with fast Fourier transform algorithms. However, there are some pitfalls associated with discrete Fourier transform-based methods related to edge effects and the handling of missing data. This paper provides an accurate and computationally efficient estimation framework for addressing those issues.

Let $Y(x) \in \mathbb{R}, x \in \mathbb{Z}^d$, be a zero-mean stationary process on the d-dimensional integer lattice, that is, $E\{Y(x)\} = 0$ and $\mathrm{cov}\{Y(x), Y(x+h)\} = K(h)$ for every x and h in $\mathbb{Z}^d$. Herglotz's theorem states that the covariance function has a Fourier transform representation,

$$K(h) = \int_{[0,1]^d} \exp(2\pi \mathrm{i}\omega \cdot h)\, \mathrm{d}F(\omega), \tag{1}$$

where $\mathrm{i} = \sqrt{-1}$ and $\cdot$ is the dot product. The function F is a spectral measure, and we assume throughout that it has a continuous derivative f, called a spectral density. We focus on estimation

of f, which encodes the covariance function and thus is crucial for prediction of missing values and for regressions when Y is used as a model for residuals. We restrict our attention to stationary models and note that stationary models often form the basis for more flexible nonstationary models that are needed to accurately model many physical processes (Fuentes, 2002).

Suppose that we observe vector $U = \{Y(x_1), \ldots, Y(x_n)\}$ at a distinct set of n locations $\mathbb{S}_1 = \{x_1, \ldots, x_n\}$. If f or K has a known parametric form and we assume that $Y(x)$ is a Gaussian process, then we can use likelihood-based methods for estimating the parameters, which generally requires $O(n^2)$ memory and $O(n^3)$ floating point operations. If the locations form a complete rectangular subset of the integer lattice, we can use Whittle's likelihood approximation (Whittle, 1954), which leverages fast Fourier transform algorithms in order to approximate the likelihood in $O(n \log n)$ FLOPs and $O(n)$ memory. Guyon (1982) showed that, due to edge effects, the Whittle likelihood parameter estimates are not root-n consistent when the dimension d of the field is greater than 1. Dahlhaus & Künsch (1987) suggested the use of data tapers to reduce edge effects and proved that the tapered version of the likelihood approximation is asymptotically efficient when $d \leqslant 3$. Stroud et al. (2017) and Guinness & Fuentes (2017) suggested the use of periodic embeddings and demonstrated their accuracy in numerical studies. Sykulski et al. (2019) introduced a debiased Whittle likelihood.

If one is not willing to assume that f or K has a known parametric form, and if the data are observed on a complete rectangular grid, nonparametric methods can be used to estimate f. The standard approach uses the discrete Fourier transform,

$$J(\omega) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Y(x_j) \exp(-2\pi \mathrm{i} \omega \cdot x_j),$$

and estimates the spectrum with a smoothed version of the periodogram $|J(\omega)|^2$,

$$\hat{f}(\omega) = \sum_{\nu} |J(\nu)|^2 \alpha(\omega - \nu),$$

where α is a smoothing kernel. Selection of the kernel bandwidth has been studied by Lee (1997), Ombao et al. (2001) and Lee (2001). Alternatively, one can smooth using penalized likelihoods (Wahba, 1980; Chow & Grenander, 1985; Pawitan & O'Sullivan, 1994) or smooth priors in a Bayesian setting (Zheng et al., 2009). Politis & Romano (1995) provided a method for reducing bias in the smoothed periodogram. Heyde & Gay (1993) studied asymptotic properties of the periodogram in an increasing domain setting, while Stein (1995) studied them in an increasing resolution setting, noting the importance of data filtering. Lim & Stein (2008) considered the multivariate case.

The nonparametric methods discussed above apply when a complete dataset is available on a rectangular grid. However, even when available on a grid, spatial datasets often have many missing values; for example, it is common to encounter gridded satellite datasets with some values obscured by clouds. Missing values complicate two aspects of periodogram-based estimators. The first is that a surrogate for the missing values must be substituted. Fuentes (2007, § 3) suggested replacing missing values with zeros and scaling the periodogram by the number of observed grid cells. Also of relevance is the extensive theoretical literature on spectral domain analysis for irregularly sampled spatial data (Matsuda & Yajima, 2009; Bandyopadhyay & Lahiri, 2009; Bandyopadhyay et al., 2015; Deb et al., 2017; Subba Rao, 2018), which can be applied to incomplete gridded data as well. All of these approaches use a discrete Fourier transform of the sampled data, which for gridded datasets is equivalent to the zero-infill approach in Fuentes (2007,

§ 3). Numerical comparisons between a zero-infill approach and our new approach are given in § 4. A second problem for spatial data is that scattered missing values seriously disrupt the use of differencing filters. For example, two-dimensional differencing at an observed location (j, k) can be applied only if observations at $(j+1, k)$, $(j, k+1)$, and $(j+1, k+1)$ are observed as well.

To address these issues, this paper introduces computationally efficient methodology for estimating the spectrum based on imputing missing values with conditional simulations and iteratively updating the spectrum estimate, in a similar vein to the method proposed by Lee & Zhu (2009) for time series data. The novelty of our approach is that the missing values are imputed onto an expanded lattice under a covariance function that is periodic on the expanded lattice. These periodic imputations or periodic conditional simulations are convenient computationally, since circulant embedding and preconditioned conjugate gradient methods can be employed for efficient imputations, but their main appeal is their ability to produce accurate estimates via the amelioration of edge effects. We provide thorough numerical studies and theoretical results describing when the imputed-data spectrum is expected to give an estimate with a smaller bias than the spectrum used for imputation, which suggests that existing spectral density estimates can be improved through periodic embedding.

The theoretical results provide a sound basis for the nonparametric estimation methods and give some insight into why the parametric methods in Guinness & Fuentes (2017) perform so well in simulations. Additionally, this paper introduces a parametric filtering method based on fitting simple parametric models within the iterative method. The fitted parametric models can be used to filter the data, which is effective for reducing bias due to periodogram smoothing. Taken together, this work develops accurate and computationally efficient methods for estimating spectral densities when the gridded data have arbitrary missingness patterns. We present thorough numerical and simulation studies for the methods and demonstrate that even a small amount of lattice expansion provides substantial bias and correlation reduction. We apply the methods to a gridded but incomplete land surface temperature dataset.

2. Methodology

2.1. *Notation and background*

Let $y = (y_1, \ldots, y_d)$ with $y_i \in \mathbb{N}$, and define the hyper-rectangle $\mathbb{J}_y \subset \mathbb{Z}^d$, where

$$\mathbb{J}_y = \{(a_1, \ldots, a_d) \mid a_j \in \{1, \ldots, y_j\} \text{ for all } j\}.$$

If $d = 2$, this is simply a rectangular lattice of size (y_1, y_2). We assume that the observation locations $\mathbb{S}_1$ form a subset of $\mathbb{J}_y$, and so we call $\mathbb{J}_y$ the observation lattice. Define V to be the vector containing the process at the remaining locations $\mathbb{J}_y \setminus \mathbb{S}_1$. Throughout, we asssume that $Y(x)$ is missing at random, meaning that the missingness is potentially related to x but not related to the value of $Y(x)$ (Little & Rubin, 2014). This section describes several existing and new iterative methods for estimating a spectral density f. All of these methods proceed by updating the spectrum estimate at the kth iteration, f_k, to the next estimate, f_{k+1}. Although the specific updating formulas vary, we use the notation f_{k+1} for all of them to keep the number of symbols manageable.

For time series data, Lee & Zhu (2009) proposed an iterative method for obtaining nonparametric estimates of the spectrum. Let E_k denote expectation in the zero-mean multivariate normal distribution for (U, V) under f_k with covariance given by (1). Their method can be extended from one dimension to general dimensions with the updating formula

$$f_{k+1}(\omega) = \sum_{\nu \in \mathbb{F}_y} E_k\left\{|J(\nu)|^2 \,\middle|\, U\right\} \alpha(\omega - \nu), \tag{2}$$

where $\mathbb{F}_y$ is the set of Fourier frequencies associated with a grid of size y. The procedure is then iterated over k until convergence. Here, we use a smoothing kernel, but Lee & Zhu (2009) noted that any smoothing method can be applied. The conditional expectation of the periodogram under f_k is computationally expensive, so Lee & Zhu (2009) proposed replacing the expected value with an average over L independent realizations of V given U, as in

$$f_{k+1}(\omega) = \sum_{\nu \in \mathbb{F}_y} \frac{1}{L} \sum_{\ell=1}^{L} |J^{(\ell)}(\nu)|^2 \alpha(\omega - \nu), \tag{3}$$

where $J^{(\ell)}$ is the discrete Fourier transform derived from $(U, V^{(\ell)})$, with $V^{(1)}, \ldots, V^{(L)}$ being independent Gaussian conditional simulations of V given U under f_k. Replacing the conditional expectation with a sample average is analogous to the approach taken in the iterative method in Tanner & Wong (1987) for Bayesian estimation of parametric statistical models. In this case, using a sample average creates a convergence issue, in that the Monte Carlo error causes the spectra in (3) to fluctuate indefinitely. In § 2.2, we propose an alternative averaging scheme, as well as imputation under a periodic model.

2.2. *Periodic imputation*

When $d > 1$, edge effects become a prominent issue (Guyon, 1982); in particular, the Whittle likelihood can be interpreted as the exact likelihood for a model in which the field is periodic on the observation lattice (Guinness & Fuentes, 2017). Data tapers have been proposed to alleviate the issue, but tapering can lead to loss of information from data near the boundaries or near missing values. In this paper, we propose extending the hyper-rectangle in each dimension and performing the imputations under a periodic approximation to the covariance function. Surprisingly, using the periodic approximation to the covariance function for the imputations, rather than the true covariance function, leads to improved spectral density estimates. This is demonstrated numerically in § 4. Periodic models also facilitate straightforward implementation of circulant embedding techniques to simulate from the conditional distributions efficiently.

Let $\tau \geqslant 1$, and define $z_j = \lceil \tau y_j \rceil$ so that $z_j \geqslant y_j$ for $j = 1, \ldots, d$. Define $m = z_1 \cdots z_d$ to be the total number of locations in $\mathbb{J}_z$, which we refer to as the embedding lattice. Let W denote the vector of missing values on $\mathbb{J}_z \setminus \mathbb{S}_1$ and $\tilde{E}_k$ denote expectation in the zero-mean multivariate normal distribution for (U, W) with covariance function $R_k(\cdot)$, defined as

$$R_k(h) = \frac{1}{m} \sum_{\omega \in \mathbb{F}_z} f_k(\omega) \exp(2\pi \mathrm{i} \omega \cdot h), \quad h = (h_1, \ldots, h_d), \tag{4}$$

where $\mathbb{F}_z$ are the Fourier frequencies associated with $\mathbb{J}_z$. For every $\omega \in \mathbb{F}_z$, the function $\exp(2\pi \mathrm{i} \omega \cdot h)$ is periodic in h_j with period z_j. This ensures that $R_k(\cdot)$ is periodic on $\mathbb{J}_z$ in each dimension and is not the integral Fourier transform of f_k that appears in (1). We refer to a draw of W under $R_k(\cdot)$ as a periodic conditional simulation or a periodic imputation. Figure 1 contains an example with $\tau = 1.15$.

Using conditional expectations, the update in the periodic model is

$$f_{k+1}(\omega) = \sum_{\nu \in \mathbb{F}_z} \tilde{E}_k \left\{ |J(\nu)|^2 \mid U \right\} \alpha(\omega - \nu). \tag{5}$$

The conditional expectation in the Lee & Zhu (2009) estimator in (2) is calculated on the observation lattice and using the correct model, whereas in (5) we use the conditional expectation under

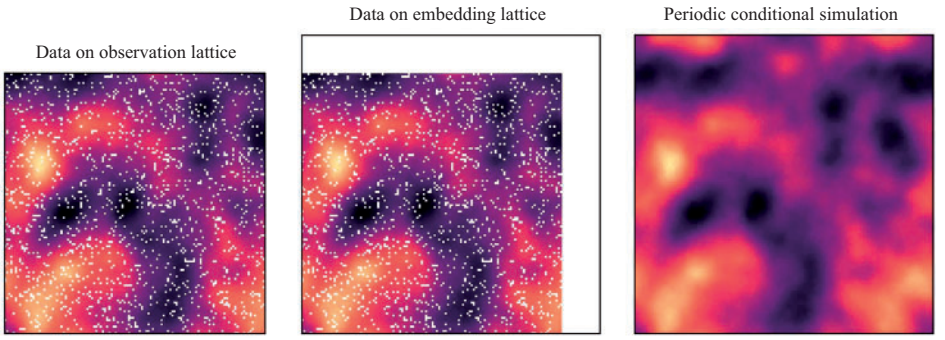

Fig. 1. Data on the observation lattice $\mathbb{J}_y$, data on the embedding lattice $\mathbb{J}_z$, and a periodic conditional simulation.

a model that is periodic on the embedding lattice. As before, the conditional expectation can be replaced by the average over one or several conditional simulations. To address the convergence issue mentioned in § 2.1, we propose an alternative updating formula consisting of a burn-in period of B iterations and convergence monitoring based on the asymptotic standard deviation of the complete-data smoothed periodogram,

$$S_k(\omega) = \Big\{ \sum_{\nu \in \mathbb{F}_z} f_k(\nu)^2 \alpha(\omega - \nu)^2 \Big\}^{1/2}.$$

Our full proposed estimation algorithm is as follows. Initialize $f_1(\omega)$ as a constant flat spectrum, and given spectrum f_k, update as follows.

Step 1. For $\ell = 1, \ldots, L$, conditionally simulate $W^{(\ell)}$ given U under f_k.
Step 2. For $\ell = 1, \ldots, L$, compute $J^{(\ell)}(\omega)$ from $(U, W^{(\ell)})$.
Step 3. Update spectrum as

$$f_{k+1}(\omega) = \begin{cases} \dfrac{1}{L} \displaystyle\sum_{\ell=1}^{L} \sum_{\nu \in \mathbb{F}_z} |J^{(\ell)}(\nu)|^2 \alpha(\omega - \nu), & k \leqslant B, \\ \dfrac{k-B}{k-B+1} f_k(\omega) + \dfrac{1}{k-B+1} \dfrac{1}{L} \displaystyle\sum_{\ell=1}^{L} \sum_{\nu \in \mathbb{F}_z} |J^{(\ell)}(\nu)|^2 \alpha(\omega - \nu), & k > B. \end{cases}$$

The algorithm is stopped when

$$\max_{\omega \in \mathbb{F}_z} \frac{|f_{k+1}(\omega) - f_k(\omega)|}{S_k(\omega)} < \varepsilon.$$

To summarize, during the B burn-in iterations, we use the sample average version of (5). After burn-in, the updating formula uses a weighted average of the previous spectrum and the current smoothed periodogram. Using a burn-in period avoids averaging over spectra from the first few iterations. Convergence is relative to the asymptotic standard deviation of the complete-data smoothed spectrum and a tolerance criterion ε, which we take to be 0.05 or 0.01 in practice. We typically take $L = 1$ in practice. The Appendix contains details on how circulant embedding and preconditioned conjugate gradient methods can be employed to efficiently compute the periodic conditional simulations.

2.3. *Variant with parametric filter*

Even if $|J(\omega)|^2$ is unbiased for $f(\omega)$, the smoothing step can introduce some bias in the spectral density estimate. For spectral densities with large dynamic range, data filters have been proposed to pre-whiten the data prior to smoothing (Stein, 1995). Missing data pose a challenge for data filters, but filters can easily be applied to the imputed data at each iteration. In this subsection, we propose a parametric filtering method that we show in simulations is successful in reducing smoothing bias.

Let f_θ be a parametric spectral density. The imputed-data Whittle likelihood approximation is

$$\ell(\theta) = -\frac{m}{2}\log 2\pi - \frac{1}{2}\sum_{\omega\in\mathbb{F}_z}\left[\log f_\theta(\omega) + \frac{\tilde{E}_k\{|J(\omega)|^2 \mid U\}}{f_\theta(\omega)}\right].$$

Let θ_k be the maximizer of $\ell(\theta)$. Then update as

$$f_{k+1}(\omega) = f_{\theta_k}(\omega)\sum_{\nu\in\mathbb{F}_z}\frac{\tilde{E}_k\{|J(\nu)|^2 \mid U\}}{f_{\theta_k}(\nu)}\alpha(\omega-\nu).$$

As before, in practice we replace $\tilde{E}_k\{|J(\nu)|^2 \mid U\}$ with a sample average that can be computed efficiently. The completely nonparametric variant is a special case with $f_\theta(\omega)$ constant. Using the parametric step in the smoothing serves to flatten the periodogram, which we show in simulation studies is helpful for reducing smoothing bias. This allows for the use of wider smoothing kernels, which reduces variance as well.

The parametric Matérn covariance is a popular choice for modelling spatial data, and so we recommend using some form of the Matérn covariance for the parametric model. Guinness & Fuentes (2017) described a quasi-Matérn covariance, whose spectral density can be evaluated quickly without aliasing calculations. Based on their results, we recommend using the quasi-Matérn covariance in practice. A special case of it is explored in § 4.

3. THEORY

This section studies bias in the imputed-data periodogram and correlation in the imputed-data discrete Fourier transform vector. We use the notation that f is the true spectrum and f_1 is a spectrum to be used for imputation. The theorem should be interpreted as a statement about how the discrete Fourier transform vector behaves given a particular imputation spectrum, not about the iterative procedure itself. Section 4 contains a numerical exploration of the iterative procedure, and § 6 discusses issues related to the theoretical study of the iterative procedure.

Let R, without parentheses, be the covariance matrix for (U, W) under the periodic covariance function $R(\cdot)$ in (4) with spectrum f. Partition R as $[A\ B; B^{\mathrm{T}}\ C]$, so that A and C are the covariance matrices for U and W, respectively. Let K denote the covariance matrix for U under the true nonperiodic covariance function $K(\cdot)$ in (1). Note that R is $m\times m$, while K is $n\times n$. Define R_1 to be the covariance matrix for (U, W) under periodic covariance function $R_1(\cdot)$ with spectrum f_1, and define A_1, B_1, and C_1 accordingly. Throughout, we assume that both f and f_1 are bounded above and below by positive constants. If W_1 is a periodic conditional simulation given observations U under f_1, then the true covariance matrix for (U, W_1) is

$$S = \mathrm{cov}\left\{\begin{pmatrix} U \\ W_1\end{pmatrix}\right\} = \begin{pmatrix} K & KA_1^{-1}B_1 \\ B_1^{\mathrm{T}}A_1^{-1}K & C_1 - B_1^{\mathrm{T}}A_1^{-1}B_1 + B_1^{\mathrm{T}}A_1^{-1}KA_1^{-1}B_1\end{pmatrix}. \tag{6}$$

The matrix S is a key object of study, and it is of interest to understand its Fourier spectrum. To this end, define the $m \times 1$ vector $g(\nu)$ to have entries $m^{-1/2}\exp(-2\pi i \nu \cdot x)$, where $\nu \in \mathbb{F}_z$ is a Fourier frequency and $x \in \mathbb{J}_z$, with the entries of $g(\nu)$ ordered as they are in (U, W). Define

$$\tilde{f}_1(\nu,\omega) = E[\tilde{E}_1\{J(\nu)J(\omega)^* \mid U\}] = g(\nu)^\dagger S g(\omega),$$

where $*$ is complex conjugate and $\dagger$ is conjugate transpose, so that $\tilde{f}_1(\omega,\omega)$ is the Fourier spectrum of S from which we construct our estimates of the spectrum. Likewise, we define $f(\nu,\omega) = f(\nu)$ if $\omega = \nu$ and 0 if $\omega \neq \nu$. This notation is useful for succinct theorem statements and reflects the fact that the true bispectrum is zero off the diagonal for stationary models. It is of interest to study $\tilde{f}_1(\nu,\omega) - f(\nu,\omega)$, which for $\omega = \nu$ corresponds to the bias of the periodogram and for $\omega \neq \nu$ measures dependence in the periodogram, both of which should ideally be near zero.

The difference $\tilde{f}_1(\nu,\omega) - f(\nu,\omega)$ will be exactly zero for every ν and ω if and only if $S = R$ due to the uniqueness of the Fourier transform. Inspection of (6) suggests that S approaches R if both K and A_1 approach A. The entries of K come from the true covariance function $K(h)$, and the entries of A come from the periodic covariance function $R(h)$. To see when K approaches A, consider the multi-dimensional Poisson summation formula,

$$R(h) = \sum_{j \in \mathbb{Z}^d} K(h + j \circ z) = K(h) + \sum_{j \neq 0} K(h + j \circ z),$$

where $j \circ z$ is the elementwise product $(j_1 z_1, \ldots, j_d z_d)$. This says that $R(h) - K(h)$, and thus $K - A$, approaches zero whenever $K(h + j \circ z)$ decays quickly enough, which can be ensured by placing smoothness conditions on the spectrum. We now state the main result.

THEOREM 1. *Let f have p continuous partial derivatives, $y_j = O(n^{1/d})$, and $z_j = O(\tau y_j)$ for $\tau > 1$. Define $\Delta = \max_{\omega \in [0,1]^d} |f(\omega) - f_1(\omega)|$. Then for every $\nu, \omega \in \mathbb{F}_z$,*

$$\tilde{f}_1(\nu,\omega) - f(\nu,\omega) = O(n^{-p/d+1}) + O\left\{\Delta\left(\frac{m-n}{m}\right)^{1/2}\right\},$$

meaning that the difference contains two terms with the respective rates.

The first term in the rate derives from the decay of the covariances $K(h)$. This term decays quickly with n when the spectrum is smooth and the dimension of the domain is small. The second term concerns the proportion of missing values relative to the number of observed values, which, when small, overwhelms the fact that $A_1 \neq A$. The assumptions about how the observation grid grows with n are standard assumptions that ensure that each dimension grows at the same rate with n. When the spectrum is smooth enough, the first decay rate is better than the usual $n^{-1/d}$ (Guyon, 1982) or even n^{-1} rate for the bias of the non-imputed periodograms. The proof is given in the Appendix along with intermediate results that assume a correct imputation spectrum.

The implication of the theorem is that when n is large enough and $(m-n)/m$ is small enough, we can initialize the iterative algorithm with any estimate of the spectrum (e.g., Fuentes, 2007; Matsuda & Yajima, 2009), and one step in the iterative algorithm will decrease the bias relative to the initialized estimate. The theorem does not make any claims about convergence of the iterative algorithm; these issues are explored numerically in § 4.

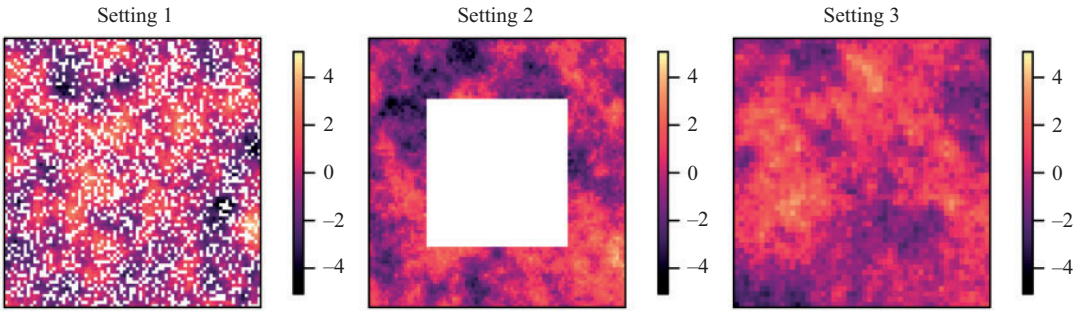

Fig. 2. Example realizations from the three missingness settings, with missing values in white.

4. NUMERICAL STUDIES AND SIMULATIONS

To provide more insight into the behaviour of the proposed estimation methods, we present a numerical study analysing the bispectrum of the imputed data and simulation results comparing the proposed estimators to other spectral density estimators. The numerical study involves calculations of the bispectrum from covariance matrices and thus involves no simulated data. In the simulation study, we estimate the spectral densities on simulated datasets, which allows us to study sampling variability and the effect of smoothing on the estimated spectral densities. In both the numerical study and the simulation study, we consider data on square grids under three missingness settings as shown in Fig. 2. The first setting has 30% scattered missing values. The second setting has a missing block in the centre of the grid, with roughly 30% of the total missing. The third setting has no missing values.

In the numerical study, we assume that the true covariance function is $K(h) = 2\exp(-\|h\|/8)$, with data on a $(32, 32)$ grid. Let f_0 be the bispectrum of K, that is,

$$f_0(\nu, \omega) = \frac{m}{n} g(\nu)^\dagger \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix} g(\omega).$$

Then for $k \geqslant 0$, let $f_{k+1}(\nu, \omega) = g(\nu)^\dagger S_k g(\omega)$, where

$$S_k = \begin{pmatrix} K & KA_k^{-1}B_k \\ B_k^{\mathrm{T}} A_k^{-1} K & C_k + B_k^{\mathrm{T}} A_k^{-1}(K - A_k)A_k^{-1}B_k \end{pmatrix}.$$

This numerical study mirrors a setting where we initialize the iterative procedure with the periodogram of the non-imputed data. This is repeated for four values of expansion factor $\tau \in \{32/32, 34/32, 36/32, 38/32\} = \{1, 1.0625, 1.125, 1.1875\}$. We quantify the error in the bispectrum with an integrated normalized squared bias

$$\frac{1}{m} \sum_{\nu \in \mathbb{F}_z} \sum_{\omega \in \mathbb{F}_z} \frac{\left\{ f_k(\nu, \omega) - f(\nu, \omega) \right\}^2}{f(\nu, \nu) f(\omega, \omega)}.$$

The results for the integrated normalized squared bias are shown in Table 1. The column for iteration 0 corresponds to bias in the non-imputed-data periodogram and has values that are quite large compared to the imputed-data periodograms, especially in Setting 1. Rows 1 and 5 correspond to imputation of missing values on the original data domain; row 9 has no missing

Table 1. *Integrated normalized squared bias under exponential covariance model, for three missingness settings, four expansion factors, including no expansion $\tau = 1$, and zero to six iterations*

		Iteration						
Setting	Expansion τ	0	1	2	3	4	5	6
1	32/32	757.6	9.584	5.946	5.457	5.457	5.516	5.562
1	34/32	866.4	5.077	1.181	0.406	0.230	0.185	0.173
1	36/32	971.6	5.663	1.466	0.432	0.152	0.069	0.043
1	38/32	1083.3	6.332	1.933	0.638	0.228	0.090	0.040
2	32/32	27.20	8.622	8.305	8.201	8.161	8.144	8.136
2	34/32	24.87	0.613	0.279	0.222	0.210	0.206	0.206
2	36/32	27.99	0.494	0.133	0.059	0.040	0.035	0.033
2	38/32	31.40	0.531	0.146	0.052	0.024	0.015	0.011
3	32/32	7.990	7.990	7.990	7.990	7.990	7.990	7.990
3	34/32	5.489	0.231	0.201	0.200	0.200	0.200	0.200
3	36/32	6.297	0.083	0.035	0.031	0.031	0.031	0.031
3	38/32	7.180	0.079	0.016	0.010	0.009	0.009	0.009

values. We see that imputing missing values on the original domain offers some improvement. However, imputing on an expanded domain gives biases that orders of magnitude smaller in many cases, and the biases decrease substantially in just a few iterations. It is also apparent that even a small amount of expansion lowers the bias; for example, expanding the domain by four pixels ($\tau = 36/32$) gives biases near zero even though the spatial range parameter is twice as large as the domain expansion.

In the simulation study, we use an $(80, 80)$ grid in Settings 1 and 2, and a $(50, 50)$ grid in Setting 3. Data are generated from a zero-mean Gaussian process model with Matérn covariance function

$$K(h) = \frac{2}{\Gamma(\nu)2^{\nu-1}}\Big(\frac{\sqrt{2\nu}\|h\|}{8}\Big)^{\nu}\mathcal{K}_{\nu}\Big(\frac{\sqrt{2\nu}\|h\|}{8}\Big),$$

with three different choices of smoothness parameter $\nu \in \{1/2, 1, 3/2\}$, range parameter 8, and variance parameter 2.

We consider several methods for estimating the spectral densities. The first method uses a smoothed periodogram computed from the discrete Fourier transform of the sampled data, scaled by the number of observations n. This method is described in § 1 and is the approach suggested by Fuentes (2007), Matsuda & Yajima (2009), Bandyopadhyay & Lahiri (2009), Bandyopadhyay et al. (2015), Deb et al. (2017), and Subba Rao (2018). The second method uses a periodogram computed from tapered data. We define one-dimensional cosine tapers T_1 and T_2 applied to 5% of the observations on each of the two edges, and the taper function is the outer product $T\{(j, k)\} = T_1(j)T_2(k)$. In Setting 1, the taper function is set to zero whenever there is a missing value. In Setting 2, which includes a square of missing values in the centre, we also taper the interior observations. The periodograms of tapered data are normalized by the sum of the squares of the taper function. Additionally, we consider the Lee & Zhu (2009) estimator described in § 2, i.e., nonperiodic imputation, and variants of their method that use lattice expansion and/or

parametric filters. Using a nonperiodic embedding method allows us to separate the effect of using a larger lattice from the effect of imputing periodically.

For the imputation-based methods proposed in this paper, we consider lattice expansion factors $\tau \in \{1.0, 1.1, 1.2\}$. We also consider two settings for the use of a parametric filter, the first being no filter, and the second with a filter of the form

$$f_\theta(\omega) = \frac{1}{1 - \frac{\theta}{2}\{\cos(2\pi\omega_1) + \cos(2\pi\omega_2)\}},$$

where $0 \leqslant \theta < 1$. This choice for the parametric model is a member of the quasi-Matérn family (Guinness & Fuentes, 2017) and is deliberately misspecified for the two cases $\nu = 1/2$ and $\nu = 3/2$. Lindgren et al. (2011) showed that this model can approximate the Matérn covariance with smoothness parameter equal to 1.

All of the imputation-based estimation methods use $L = 1$ conditional simulations, $B = 100$ burn-in iterations, and convergence criterion $\varepsilon = 0.01$. The estimate from the jth dataset is denoted by $\hat{f}^j$. All methods use a Gaussian smoothing kernel proportional to $\exp(-\|\omega - \nu\|^2/\delta^2)$, where the distance $\|\omega - \nu\|$ is defined periodically on the domain $[0, 1]^2$. We consider two metrics for evaluating the estimation methods. The first is a relative bias

$$\text{BIAS}(\omega) = \frac{1}{100}\sum_{j=1}^{100}\frac{\hat{f}^j(\omega) - f(\omega)}{f(\omega)},$$

where 100 is the total number of simulated datasets and f is the true spectrum. The second metric is a mean relative squared error

$$\text{MSE}(\omega) = \frac{1}{100}\sum_{j=1}^{100}\left\{\frac{\hat{f}^j(\omega) - f(\omega)}{f(\omega)}\right\}^2.$$

To evaluate relative bias on an equal footing, we compare all methods using a small value of $\delta = 0.02$. Figure 3 contains plots of the relative bias for the nontapered and tapered methods, and for the nonfiltered and filtered periodic and nonperiodic embedding methods with $\tau = 1.2$. Results for $\nu = 1/2$ are shown and results for larger values of ν are similar. In Setting 1, the nontapered and tapered methods have a very large relative bias at almost every frequency. They estimate far too much power at higher frequencies, due to the fact that imputing with zeros produces fields that are rougher than the underlying process. In contrast, the periodic embedding methods have small bias. In Setting 2, the nontapered and tapered biases improve, but are still larger than the periodic embedding biases, especially for low frequencies. The relative biases for nontapered and tapered methods are similar in Setting 3 and are still larger than the periodic embedding relative biases. Though not shown here, the biases for $\tau = 1.1$ and 1.3 are similar. The parametric filters serve to reduce the bias compared to not filtering. The periodic embedding methods have a small bias near $\omega = (0, 0)$; based on the accuracies shown in the numerical studies, this bias is likely due to smoothing bias because of the sharply peaked spectra near the origin. Imputing nonperiodically does not substantially improve the bias in Settings 2 and 3. It does improve bias in Setting 1, but it is not as effective as periodic embedding.

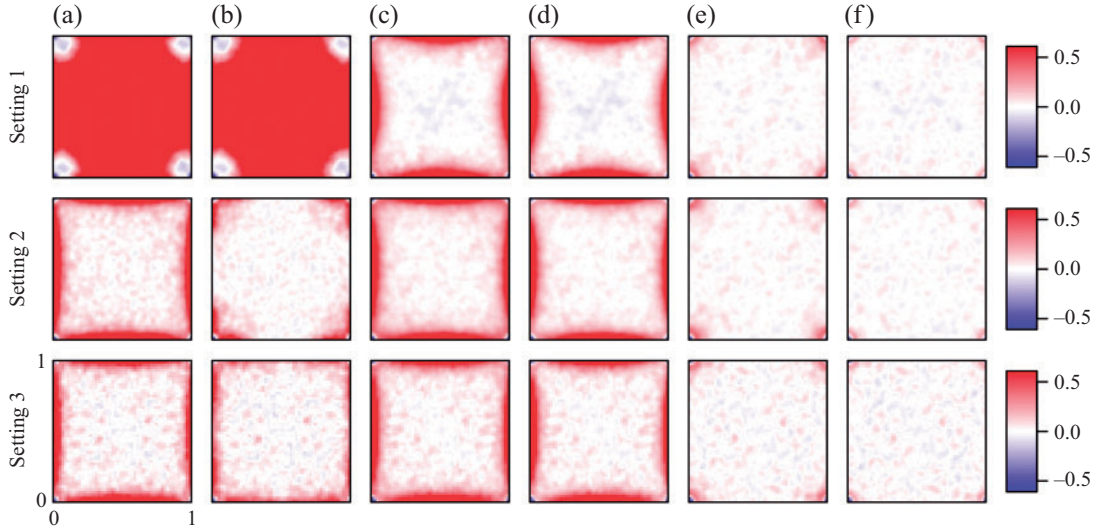

Fig. 3. Relative bias as a function of frequency for the three missingness settings under $\nu = 1/2$ and six estimation settings: (a) not tapered; (b) tapered; (c) $\tau = 1.2$, no filter, not periodic; (d) $\tau = 1.2$, parametric filter, not periodic; (e) $\tau = 1.2$, no filter, periodic; (f) $\tau = 1.2$, parametric filter, periodic.

To evaluate mean relative squared error on an equal footing, all methods were computed with a range of choices for δ; the reported results are for the value of δ that minimized

$$\left\{\frac{1}{m}\sum_{\omega\in\mathbb{F}_z}\mathrm{MSE}(\omega)\right\}^{1/2},$$

the root integrated mean relative squared error over all Fourier frequencies. Table 2 contains root integrated mean relative squared error results for the various methods. The periodic embedding methods with $\tau > 1$ are more accurate than both the nontapered and the tapered periodogram estimates in every case. In Setting 1, the nontapered and tapered estimates are quite poor, likely due to the large biases seen in Fig. 3. For periodic embedding, we see that the values improve when $\tau > 1$ but do not improve beyond $\tau = 1.1$. This is consistent with the numerical studies that showed a small amount of periodic embedding was sufficient. Filtering provides a further improvement, reducing the values by 30–40%. In Setting 2, the nontapered and tapered estimates improve substantially, and the periodic embedding methods offer further improvement. Imputing missing values is an improvement, but imputing periodically always gives better results than imputing nonperiodically. This can be seen by comparing the $\tau = 1.0$ results to the $\tau > 1$ results and by comparing the periodic to the nonperiodic imputation results. In Setting 3, the parametric filter performs similarly to tapering, but periodic embedding with parametric filtering is by far the most accurate method when $\tau > 1$.

5. Application to satellite data

To illustrate the practical usefulness of the proposed methods, we analyse a gridded land surface temperature dataset. These data were used recently in Heaton et al. (2018), a study comparing various Gaussian process approximations. The data were originally collected by the Moderate Resolution Imaging Spectrometer on board the NASA Terra Satellite. The region is a grid of 500 by 300 locations in the latitudinal range of 34.295 to 37.068 and longitudinal range of −95.912

Table 2. *Root integrated mean relative squared error results*

	Missingness setting								
	1	2	3	1	2	3	1	2	3
impute - filter - periodic		$\nu = 1/2$			$\nu = 1$			$\nu = 3/2$	
no - no - no	3.495	0.560	0.478	32.11	3.145	2.299	257.3	17.03	15.36
no - taper - no	3.498	0.291	0.342	31.83	0.472	0.900	255.8	0.920	3.735
$\tau = 1.0$ - no - no	0.389	0.423	0.462	1.913	2.093	2.294	9.107	12.49	15.36
$\tau = 1.1$ - no - no	0.362	0.379	0.412	1.734	1.930	2.097	8.691	11.05	12.28
$\tau = 1.2$ - no - no	0.397	0.402	0.439	1.858	2.098	2.313	9.669	11.88	13.35
$\tau = 1.0$ - yes - no	0.313	0.320	0.323	1.168	1.197	1.559	6.138	8.280	8.803
$\tau = 1.1$ - yes - no	0.284	0.296	0.296	1.001	1.041	1.260	5.775	6.469	7.511
$\tau = 1.2$ - yes - no	0.296	0.312	0.309	1.022	1.062	1.266	5.689	7.443	7.587
$\tau = 1.0$ - no - yes	0.367	0.423	0.462	1.684	2.092	2.294	8.418	12.48	15.36
$\tau = 1.1$ - no - yes	0.208	0.219	0.259	0.238	0.268	0.326	0.288	0.297	0.382
$\tau = 1.2$ - no - yes	0.205	0.223	0.262	0.247	0.256	0.309	0.281	0.280	0.353
$\tau = 1.0$ - yes - yes	0.253	0.319	0.323	0.908	1.195	1.559	5.348	8.266	8.803
$\tau = 1.1$ - yes - yes	0.136	0.145	0.153	0.096	0.088	0.108	0.141	0.141	0.166
$\tau = 1.2$ - yes - yes	0.133	0.143	0.153	0.097	0.091	0.109	0.142	0.137	0.156

to -91.284, roughly 450 km by 300 km with grid spacing 1100 m in the north/south direction and 900 m in the east/west direction. The values in the dataset represent land surface temperature in degrees Celsius. The dataset has 105 569 nonmissing values, which are plotted in the top left panel of Fig. 4. We can see that there is a distinct trend from the southeast to the northwest corner, so we include a linear trend in the mean function, estimated by generalized least squares.

We have found that $\varepsilon = 0.05$ is a reasonable convergence tolerance criterion, and we choose $B = 30$ burn-in iterations. We use a crossvalidation procedure to choose the smoothing parameter. A random subset of 30% of the data is held out; the iterative methods are run with a range of smoothing parameters, and the parameter that minimizes sum of squared prediction errors was chosen.

In Fig. 4, we plot the original data, the conditional expectation, an estimate of the conditional standard deviations, and three conditional simulation plots. The conditional standard deviations are estimated by computing 30 conditional simulations and finding the root mean squared difference between the conditional expectation and each of the conditional simulations at each pixel. On average, each conditional simulation took just 2.76 seconds and converged in 25 iterations with the Vecchia preconditioner, and took 15.48 seconds and converged in 159 iterations with the inverse spectrum preconditioner. The iterative spectrum estimation method took 4.86 minutes to converge. While these timings indicate that the analysis is feasible on a large dataset, a zero-infill method is much faster, taking just 0.06 seconds. All timings were carried out on a 2016 Macbook Pro with 3.3 GHz Intel Core i7 dual-core processer and 16 GB memory, running R 3.4.2 linked to Apple's Accelerate BLAS libraries.

Visually, the data appear to have a longer correlation length scale in the northeast-southwest direction than in the southeast-northwest direction. The estimate of the spectrum returned by the iterative method confirms our visual suspicions, as can be seen in Fig. 5 where the logarithm of the estimated spectrum is plotted. The estimated spectrum shows clear signs of anisotropy in that the spectrum has contours that are not circular. Maximum likelihood estimation of anisotropic models is generally difficult due to optimization over additional parameters. In contrast, the nonparametric spectral density estimation methods automatically estimated the anisotropies with no extra computational effort.

The spectral methods described in this paper were included in the Heaton et al. (2018) comparison project and compared favourably to all of the other methods on all of the prediction and timing

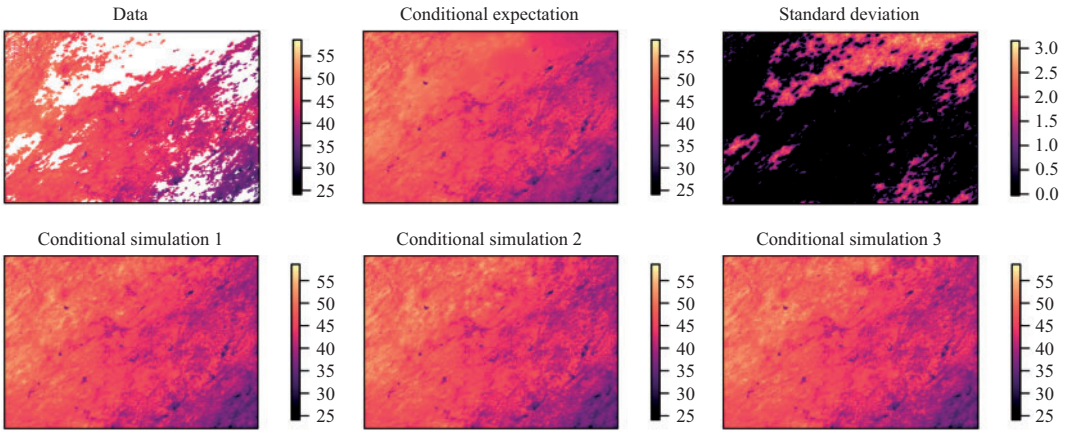

Fig. 4. Original data, predictions, standard deviations, and three conditional simulations of the missing values.

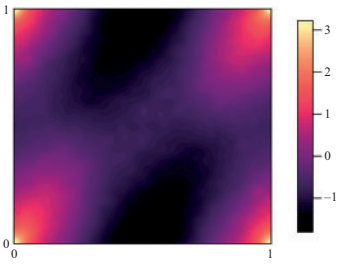

Fig. 5. Log base 10 of the spectral density estimate.

metrics, and it was the best performing method for the interval score metric (Gneiting & Raftery, 2007), which rewards forecasts that come with small prediction intervals that often contain the predictand. To gain some intuition for this result, we report some results for $(1-\alpha)\%$ prediction intervals based on a Gaussian assumption. In particular, we sort the predictions $(\hat{Y}_1, \ldots, \hat{Y}_p)$ to be increasing in the prediction standard deviation, and then report average prediction standard deviations for $(\hat{Y}_i, \ldots, \hat{Y}_j)$ for various ranges of the indices i and j. The results from the periodic spectral methods are compared in Table 3 to predictions that use an isotropic Matérn covariance model, with parameters estimated via Vecchia's approximation (Vecchia, 1988), as implemented in the `R` package `GpGp` (Guinness, 2018; Guinness & Katzfuss, 2018; R Development Core Team, 2019). Vecchia's approximation applies to parametric models and to both gridded and nongridded data. We can see that while the two methods do not differ substantially for predictions that the model expects to be uncertain, the periodic spectral methods produce smaller prediction intervals and smaller root mean squared prediction errors when the model expects small prediction errors. This is achieved with coverage rates that are larger than those produced by Vecchia's approximation with an isotropic model.

6. Discussion

The methods involve choosing the factor by which the lattice should be expanded. We have found that even very small factors that expand the lattice by a few pixels are effective at improving the spectral density estimates. We recommend expanding each dimension by an amount roughly equal to the correlation range in the data. The fact that we expand the lattice in the positive

Table 3. *Average prediction standard deviation and coverages for the specified range of predicted values, with the predicted values sorted according to the fitted models' prediction standard deviations. In other words, the first column corresponds to prediction results for the 500 predictions that the model expects to be most certain, and the last column corresponds to the predictions expected to be most uncertain*

	Index Range	1	501	1001	2001	10001	20000
		500	1000	2000	10000	20000	44431
Periodic	Avg Pred SD	0.365	0.427	0.482	0.694	1.164	1.88
Spectral	Std. Dev.	0.414	0.477	0.554	0.686	1.078	2.209
	80% Coverage	81.14	82.06	83.15	84.96	85.34	73.53
	90% Coverage	86.56	89.23	89.29	91.35	92.61	84.59
	95% Coverage	91.47	92.82	93.08	94.72	95.99	91.19
Vecchia	Avg Pred SD	0.501	0.548	0.585	0.749	1.198	1.876
	Std. Dev.	0.503	0.58	0.538	0.718	1.094	2.201
	80% Coverage	74.88	78.55	77.23	80.71	82.05	61.12
	90% Coverage	84.88	87.65	87.15	88.58	90.66	74.85
	95% Coverage	89.02	91.84	90.87	92.39	94.47	83.87

direction, rather than in the negative direction or both positive and negative directions, is not important since we assume a periodic model on the expanded lattice. As with most nonparametric spectral density estimates, the methods involve the choice of a smoothing parameter. We have not attempted to provide any new methods for selecting smoothing parameters, as this issue has been well-studied in the literature. However, the parametric filtering methods serve to flatten the periodogram, which makes the estimates less sensitive to the choice of smoothing parameter. In our application to land surface temperature data, we used a crossvalidation procedure to select the smoothing parameter. Though we have chosen $L = 1$ imputation per iteration in every example, the methods allow for $L > 1$. We suspect that choosing $L > 1$ would drive the iterative methods to converge in fewer iterations but incur a higher computational cost per iteration. Examining the details of this trade-off would be an interesting study. It may be advantageous to use $L > 1$ if the conditional simulations can be computed in parallel.

While many large datasets involve spatially gridded observations, we acknowledge that there is also a need for methods for analysing nongridded data. The nonparametric methods described in this paper may prove useful for analysing nongridded data as well; in fact Nychka et al. (2015) have a framework for analysing nongridded data that includes a lattice process as a model component. Here we have considered stationary models which can also be used as components in nonstationary models (Fuentes, 2002), and so the methods developed here could potentially be extended to be used for local nonparametric estimation of nonstationary models. The paper contains some theoretical results about the iterative procedure, but proving that the iterative algorithm converges remains elusive, partly due to pathological cases in the observed vector U, but this is an important area of future work.

ACKNOWLEDGEMENT

This research was supported by the U.S. National Science Foundation Division of Mathematical Sciences and the National Institute of Environmental Health Sciences.

SUPPLEMENTARY MATERIAL

The methods are implemented in an `R` package titled `npspec` available at `https://github.com/joeguinness/npspec`.

Appendix

Circulant embedding and inverse spectrum preconditioner

To see how the conditional simulations of W given U can be computed efficiently, define R to be the covariance matrix for (U, W) under covariance function $R(\cdot)$, and partition R as

$$R = \begin{pmatrix} A & B \\ B^{\mathrm{T}} & C \end{pmatrix},$$

where A and C are the covariance matrices for the observations U and missing values W, respectively. The conditional expectation for W given U is $\tilde{E}(W \mid U) = B^{\mathrm{T}}A^{-1}U$. The most demanding computational step for obtaining $\tilde{E}(W \mid U)$ is solving the linear system $Ax = U$. Preconditioned conjugate gradient methods for solving linear systems (Greenbaum, 1997) are efficient when the forward multiplication Ax can be computed efficiently and when we can find a matrix M, called the preconditioner, for which $MA \approx I$ and for which Mx can be computed efficiently. Below, we describe how circulant embedding can be used to compute the forward multiplication Ax efficiently. In practice, we have found that a preconditioner based on Vecchia's Gaussian process approximation (Vecchia, 1988) is effective and fast for the problems we have studied. This preconditioner was proposed in Stroud et al. (2017). At the end of this section, we give details about another preconditioner based on a submatrix of the inverse of R^{-1}.

Suppose that Q is an $m \times m$ nested block circulant matrix. Nested block circulant includes the special cases of circulant, arising from a periodic and stationary covariance in one dimension, and block circulant with circulant blocks, arising from a periodic and stationary covariance in two dimensions. The matrix Q can be written as $Q = FDF^{\dagger}$, where F is the discrete Fourier transform matrix and D is a diagonal matrix with the eigenvalues on the diagonal. Because of the discrete Fourier matrix representation, one can multiply Qv in $O(m \log m)$ time and $O(m)$ memory by taking the discrete Fourier transform of v, i.e., `fft(v)` in `R`, then multiplying the entries of the resultant vector pointwise by the eigenvalues in D, and then taking an inverse discrete Fourier transform of the result.

The multiplication Ax can be computed efficiently by embedding the multiplication inside of

$$R\begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} A & B \\ B^{\mathrm{T}} & C \end{pmatrix}\begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} Ax \\ B^{T}x \end{pmatrix}.$$

Then the appropriate entries Ax can be extracted, and the unnecessary entries $B^{T}x$ can be discarded. Note that R is not nested block circulant, but there exists a row-column permutation of R that is nested block circulant. Let P denote the permutation matrix such that $Q = PRP^{T}$ is nested block circulant. Then the multiplication can be performed as

$$R\begin{pmatrix} x \\ 0 \end{pmatrix} = P^{T}QP\begin{pmatrix} x \\ 0 \end{pmatrix}.$$

Thus, the multiplication can be carried out by an appropriate reordering of $[x \;\; 0]$ in $O(m)$ time, then an $O(m \log m)$-time multiplication by nested block circulant Q, and then an $O(m)$-time reordering of the result.

The preconditioner $M = (A - BC^{-1}B^{T})^{-1}$ is a submatrix of R^{-1}. Here, we describe how the multiplication Mx can be performed efficiently without computing the entries of M. The inverse of R is a permutation of a nested block circulant matrix and can be written as

$$R^{-1} = \begin{Bmatrix} (A - BC^{-1}B^{T})^{-1} & -(A - BC^{-1}B^{T})^{-1}BC^{-1} \\ -C^{-1}B^{T}(A - BC^{-1}B^{T})^{-1} & (C - B^{T}A^{-1}B)^{-1} \end{Bmatrix} = P^{T}FD^{-1}F^{\dagger}P.$$

This means that the multiplication Mx can be embedded in the larger multiplication

$$R^{-1}\begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{Bmatrix} (A - BC^{-1}B^{T})^{-1}x \\ -C^{-1}B^{T}(A - BC^{-1}B^{T})^{-1}x \end{Bmatrix} = P^{T}FD^{-1}F^{\dagger}P\begin{pmatrix} x \\ 0 \end{pmatrix},$$

and the multiplication can be carried out in $O(m \log m)$ time and $O(m)$ memory by a sequence of reorderings, discrete Fourier transforms, and pointwise multiplications.

Proofs of theoretical results

LEMMA A1. *If $f_1 = f$, $K(h) = 0$ for all $\|h\| > h_0$, and $|z_j - y_j| > h_0$ for all $j \in \{1, \ldots, d\}$, then for all $\nu, \omega \in \mathbb{F}_z$, $\tilde{f}_1(\nu, \omega) - f(\nu, \omega) = 0$.*

Proof. We have $f(\nu, \omega) = g(\nu)^\dagger R g(\omega)$, and so $\tilde{f}_1(\nu, \omega) - f(\nu, \omega) = g(\nu)^\dagger (S - R) g(\omega)$. The matrix $S - R$ can be written as

$$S - R = \begin{Bmatrix} K - A & (K - A)A^{-1}B \\ B^T A^{-1}(K - A) & B^T A^{-1}(K - A)A^{-1}B \end{Bmatrix}.$$

It suffices to show that $K = A$ in order to establish the result. According to the multi-dimensional Poisson summation formula, we can relate $K(\cdot)$ and $R(\cdot)$ by (Guinness & Fuentes, 2017, Lemma 1)

$$R(x_1 - x_2) = \sum_{k \in \mathbb{Z}^d} K\{x_1 - x_2 + (z \circ k)\} = K(x_1 - x_2) + \sum_{k \in \mathbb{Z}^d \setminus 0} K\{x_1 - x_2 + (z \circ k)\}, \tag{A1}$$

where $z \circ k := (z_1 k_1, \ldots, z_d k_d)$. For any $x_1, x_2 \in \mathbb{J}_y$, the observation lattice, we have $|x_{1j} - x_{2j}| < y_j$ for every $j = 1, \ldots, d$. Thus if $k_j \in \mathbb{Z} \setminus 0$, $|x_{1j} - x_{2j} + z_j k_j| > |z_j - y_j| > h_0$. Thus at least one element of $x_1 - x_2 + (z \circ k)$ has absolute value greater than h_0 when $k \in \mathbb{Z}^d \setminus 0$, and so $\|x_1 - x_2 + (z \circ k)\| > h_0$ for all $k \in \mathbb{Z}^d \setminus 0$, implying that all terms in the sum in (A1) must be zero. This gives us $R(x_1 - x_2) = K(x_1 - x_2)$ for any $x_1, x_2 \in \mathbb{J}_y$, and so $K = A$. □

LEMMA A2. *If f has p continuous partial derivatives, $f_1 = f$, $y_j = O(n^{1/d})$, and $z_j = O(\tau y_j)$ for $\tau > 1$, then for all $\nu, \omega \in \mathbb{F}_z$,*

$$\tilde{f}_1(\nu, \omega) - f(\nu, \omega) = O(n^{-p/d+1}).$$

Proof. As in the proof of Lemma A1, $\tilde{f}_1(\nu, \omega) - f(\nu, \omega) = g(\nu)^\dagger (S - R) g(\omega)$. Partitioning the vector $g(\nu)$ as $\{g_1(\nu), g_2(\nu)\}$ according to the same partition as (U, W), we have

$$\begin{aligned} \tilde{f}_1(\nu, \omega) - f(\nu, \omega) &= \begin{pmatrix} g_1(\nu)^\dagger & g_2(\nu)^\dagger \end{pmatrix} \begin{Bmatrix} K - A & (K - A)A^{-1}B \\ B^T A^{-1}(K - A) & B^T A^{-1}(K - A)A^{-1}B \end{Bmatrix} \begin{Bmatrix} g_1(\omega) \\ g_2(\omega)^\dagger \end{Bmatrix} \\ &= \{g_1(\nu)^\dagger + g_2(\nu)^\dagger B^T A^{-1}\}(K - A)\{g_1(\omega) + A^{-1}B g_2(\omega)\}, \end{aligned}$$

and so the difference can be bounded as

$$\begin{aligned} |\tilde{f}_1(\nu, \omega) - f(\nu, \omega)| &= |\{g_1(\nu)^\dagger + g_2(\nu)^\dagger B^T A^{-1}\}(K - A)\{g_1(\omega) + A^{-1}B g_2(\omega)\}| \\ &\leqslant \|g_1(\nu) + A^{-1}B g_2(\nu)\|_2 \, \|K - A\|_2 \, \|g_1(\omega) + A^{-1}B g_2(\omega)\|_2. \end{aligned}$$

We will consider each term in turn. Let $\rho(M)$ denote the spectral radius of symmetric matrix M. Then

$$\begin{aligned} \|g_1(\nu) + A^{-1}B g_2(\nu)\|_2 &\leqslant \|g_1(\nu)\|_2 + \|A^{-1}B g_2(\nu)\|_2 \\ &\leqslant (n/m)^{1/2} + \|A^{-1/2}\|_2 \, \|A^{-1/2}B\|_2 \, \|g_2(\nu)\|_2 \\ &= (n/m)^{1/2} + \rho(A^{-1})^{1/2} \rho(B^T A^{-1} B)^{1/2} \|g_2(\nu)\|_2 \\ &\leqslant (n/m)^{1/2} + f_{\min}^{-1/2} \rho(C)^{1/2} \left(\frac{m - n}{m}\right)^{1/2} \\ &\leqslant (n/m)^{1/2} + \left(\frac{f_{\max}}{f_{\min}}\right)^{1/2} \left(\frac{m - n}{m}\right)^{1/2}, \end{aligned}$$

where in the second to last inequality, we used $\rho(A^{-1}) < \rho(R^{-1}) = f_{\min}^{-1}$ and $C - B^T A^{-1} B$ is positive definite, so the largest eigenvalue of $B^T A^{-1} B$ is smaller than the largest eigenvalue of C, which is smaller than the largest eigenvalue of R, $f_{\max}$.

The previous inequality did not depend on ν, so it holds for $\|g_1(\omega) + A^{-1} B g_2(\omega)\|_2$ as well. To bound $\|K - A\|_2$, we use the fact that for symmetric matrices $\|K - A\|_2 = \rho(K - A) < \|K - A\|_1$, where

$$\begin{aligned}
\|K - A\|_1 &= \max_i \sum_{j=1}^{n} |K_{ij} - A_{ij}| \\
&= \max_{x_1 \in \mathbb{J}_y} \sum_{x_2 \in \mathbb{J}_y} \Big| K(x_1 - x_2) - R(x_1 - x_2) \Big| \\
&= \max_{x_1 \in \mathbb{J}_y} \sum_{x_2 \in \mathbb{J}_y} \Big| K(x_1 - x_2) - \sum_{k \in \mathbb{Z}^d} K\{x_1 - x_2 + (k \circ z)\} \Big| \\
&= \max_{x_1 \in \mathbb{J}_y} \sum_{x_2 \in \mathbb{J}_y} \Big| \sum_{k \neq \mathbf{0}} K\{x_1 - x_2 + (k \circ z)\} \Big| \\
&\leqslant \max_{x_1 \in \mathbb{J}_y} \sum_{x_2 \in \mathbb{J}_y} \sum_{k \neq \mathbf{0}} \Big| K\{x_1 - x_2 + (k \circ z)\} \Big|.
\end{aligned}$$

The third equality uses the multi-dimensional Poisson summation formula referenced in (A1). By assumption, for $k = (k_1, \ldots, k_d) \neq 0$ and any $x_1, x_2 \in \mathbb{J}_y$,

$$\max_{1 \leqslant j \leqslant d} |x_{1j} - x_{2j} + z_j k_j| \geqslant \min_{1 \leqslant j \leqslant d} |z_j - y_j|.$$

This is because $k_j \neq 0$ for at least one j. Define $\ell_{min} = \min_{1 \leqslant j \leqslant d} |z_j - y_j|$, which is the embedding distance in the dimension with the smallest amount of embedding. By assumption, we have

$$\ell_{\min} > \min_{1 \leqslant j \leqslant d} \lfloor (\tau - 1) y_j \rfloor > \min_{1 \leqslant j \leqslant d} \lfloor (\tau - 1) a_j n^{1/d} \rfloor.$$

This means that the sum does not contain any terms $K(h)$ for which $\max_j |h_j| < \ell_{\min}$. Define the set $G_\ell = \{h : \max_j |h_j| = \ell\}$, which is a hollowed-out cube on $\mathbb{Z}^d$ and has size $|G_\ell| = (2\ell + 1)^d - (2\ell - 1)^d = O(\ell^{d-1})$. Using this notation, the sum can be bounded as

$$\|K - A\|_1 \leqslant \sum_{\ell = \ell_{\min}}^{\infty} \sum_{h \in G_\ell} |K(h)|.$$

Lemma 9.5 in Körner (1989) states that if f has p continuous partial derivatives on T^d, with maximum pth partial derivative $Q_p(f)$, then

$$|K(h)| \leqslant Q_p(f) |h_j|^{-p}$$

for every $1 \leqslant j \leqslant d$, and so we can use the bound $|K(h)| \leqslant Q_p(f) (\max_{1 \leqslant j \leqslant d} |h_j|)^{-p}$. This gives us an explicit bound

$$\begin{aligned}
\|K - A\|_1 &\leqslant \sum_{\ell = \ell_{\min}}^{\infty} \sum_{h \in G_\ell} \frac{Q_p(f)}{\ell^p} = \sum_{\ell = \ell_{\min}}^{\infty} \Big\{ (2\ell + 1)^d - (2\ell - 1)^d \Big\} \frac{Q_p(f)}{\ell^p} \\
&= \sum_{\ell = \ell_{\min}}^{\infty} Q_p(f) \frac{a_{d-1}(\ell)}{\ell^p},
\end{aligned}$$

where $a_{d-1}(\ell)$ is a polynomial of degree $d-1$ in ℓ. Then we have

$$\|K - A\|_1 = O(\ell_{\min}^{-p+d}) = O(n^{-p/d+1})$$

since the largest exponent in $a_{d-1}(\ell)/\ell^p$ is $-p+d-1$. Combining this with $\|g_1(\nu) + A^{-1}Bg_2(\nu)\| = O(1)$ gives the desired result. □

THEOREM A1. *Let f have p continuous partial derivatives, and assume the same conditions on the observation and embedding lattice as in Lemma A2. Define $\Delta_1 = \max_{\omega\in[0,1]^d} |f(\omega) - f_1(\omega)|$. Then for all $\nu, \omega \in \mathbb{F}_z$, $\tilde{f}_1(\nu,\omega) - f(\nu,\omega) = O(n^{-p/d+1}) + O\left\{\Delta_1(m - n/m)^{1/2}\right\}$, meaning that the difference contains two terms with the respective rates.*

Proof. Define the matrix S_1 as

$$S_1 = \begin{pmatrix} A & AA_1^{-1}B_1 \\ B_1^TA_1^{-1}A & C_1 - B_1^TA_1^{-1}B_1 + B_1^TA_1^{-1}AA_1^{-1}B_1 \end{pmatrix},$$

which differs from S in that K in S is replaced by A in S_1. The difference $\tilde{f}_1(\nu,\omega) - f(\nu,\omega)$ can be written as

$$g(\nu)^\dagger(S-R)g(\omega) = g(\nu)^\dagger(S-S_1)g(\omega) + g(\nu)^\dagger(S_1 - R_1 + R_1 - R)g(\omega). \tag{A2}$$

The first term in (A2) is

$$\begin{aligned}\delta_1 &= \begin{pmatrix} g_1(\nu)^\dagger & g_2(\nu)^\dagger \end{pmatrix} \begin{Bmatrix} K-A & (K-A)A_1^{-1}B_1 \\ B_1^TA_1^{-1}(K-A) & B_1^TA_1^{-1}(K-A)A_1^{-1}B_1 \end{Bmatrix} \begin{Bmatrix} g_1(\omega) \\ g_2(\omega)^\dagger \end{Bmatrix} \\ &= \{g_1(\nu)^\dagger + g_2(\nu)^\dagger B_1^TA_1^{-1}\}(K-A)\{g_1(\omega) + A_1^{-1}B_1g_2(\omega)\}.\end{aligned}$$

This expression has a similar form to that which appears in the proof of Lemma A2. As before, we need bounds for $\|g_1(\omega)+A_1^{-1}B_1g_2(\omega)\|_2$ and $\|K-A\|_2$ in order to bound δ_1. The proof for the bound on $\|K-A\|_2$ is identical to that in Lemma A2, and the proof for the bound on $\|g_1(\omega)+A_1^{-1}B_1g_2(\omega)\|_2$ is similar, although f is replaced by f_1, which does not change the overall result that the first term in (A2) is $O(n^{-p/d+1})$.

To shorten the equations to follow, write $M_1 = A_1^{-1}B_1$. The second term in (A2) is

$$\begin{aligned}\delta_2 &= \begin{pmatrix} g_1(\nu)^\dagger & g_2(\nu)^\dagger \end{pmatrix} \begin{Bmatrix} A-A_1 & (A-A_1)M_1 \\ M_1^T(A-A_1) & M_1^T(A-A_1)M_1 \end{Bmatrix} \begin{Bmatrix} g_1(\omega) \\ g_2(\omega)^\dagger \end{Bmatrix} + f_1(\nu,\omega) - f(\nu,\omega) \\ &= \{g_1(\nu)^\dagger + g_2(\nu)^\dagger M_1^T\}(A-A_1)\{g_1(\omega) + M_1g_2(\omega)\} + f_1(\nu,\omega) - f(\nu,\omega).\end{aligned}$$

Define the discrete Fourier transform matrix F to have (j,k) entry $m^{-1/2}\exp(\mathrm{i}\omega_k \cdot x_j)$, where ω_k is a Fourier frequency in $\mathbb{F}_z$ and x_j is a location in $\mathbb{J}_z$. Partition the discrete Fourier transform matrix F into rows for the observations and missing values as $F^\dagger = [G^\dagger H^\dagger]$. We have $R = FDF^\dagger$, where D is diagonal with entries $f(\nu,\omega)$. This gives $A = GDG^\dagger$, and likewise $A_1 = GD_1G^\dagger$, where D_1 is diagonal with entries $f_1(\nu,\omega)$. Then δ_2 can be written as

$$\delta_2 = \{g_1(\nu)^\dagger G + g_2(\nu)^\dagger M_1^T G\}(D-D_1)\{G^\dagger g_1(\omega) + G^\dagger M g_2(\omega)\} + f_1(\nu,\omega) - f(\nu,\omega).$$

Note that $I = G^\dagger G + H^\dagger H$. Since $g_1(\nu)^\dagger$ is a row of $G^\dagger$ and $g_2(\nu)^\dagger$ is the same row of $H^\dagger$, we have

$$g_1(\nu)^\dagger G + g_2(\nu)^\dagger H = e(\nu),$$

where $e(\nu) = 1$ for the entry corresponding to ν and 0 otherwise. This gives

$$\delta_2 = \{e(\nu)^T + g_2(\nu)^\dagger(M_1^T G - H)\}(D - D_1)\{e(\omega) + (G^\dagger M_1 - H^\dagger)g_2(\omega)\} + f_1(\nu,\omega) - f(\nu,\omega).$$

We can see now that since $e(\nu)^T(D - D_1)e(\omega) = f(\nu,\omega) - f_1(\nu,\omega)$, there is a cancellation, giving

$$\begin{aligned}\delta_2 = {} & g_2(\nu)^\dagger(M_1^T G - H)(D - D_1)e(\omega) + e(\nu)^T(D - D_1)(G_1^\dagger M_1 - H^\dagger)g_2(\omega) \\ & + g_2(\nu)^\dagger(M_1^T G - H)(D - D_1)(G_1^\dagger M_1 - H^\dagger)g_2(\omega).\end{aligned}$$

This cancellation is the key step. Using matrix norm inequalities, we have

$$\begin{aligned}|\delta_2| \leqslant {} & \|g_2(\nu)\|_2\|M_1^T G - H\|_2\|D - D_1\|_2\|e(\omega)\|_2 \\ & + \|g_2(\omega)\|_2\|M_1^T G - H\|_2\|D - D_1\|_2\|e(\nu)\|_2 \\ & + \|g_2(\omega)\|_2\|g_2(\nu)\|\|M_1^T G - H\|_2^2\|D - D_1\|_2\|e(\nu)\|_2.\end{aligned}$$

Since $g_2(\omega)$ is of length n_2 and has entries $n^{-1/2}\exp(2\pi \mathrm{i}\omega \cdot x)$, $\|g_2(\omega)\|_2 = \sqrt{n_2/n}$. Clearly, $\|e(\omega)\| = 1$ because of its definition, and $\|D - D_1\|_2 = \Delta_1$ because both D and D_1 are diagonal with diagonal entries holding $f(\nu,\nu)$ and $f_1(\nu,\nu)$, respectively. This leaves

$$\|M_1^T G - H\|_2^2 = \rho\left\{(M_1^T G - H)(G^\dagger M_1 - H^\dagger)\right\} = \rho\Big(M_1^T M_1 + I\Big),$$

because $GG^\dagger = I$, $HH^\dagger = I$, $HG^\dagger = 0$, and $GH^\dagger = 0$. Thus, the squared 2-norm is 1 plus the largest eigenvalue of $M_1^T M$, which is

$$\|M_1^T M_1\|_2 = \|M_1\|_2^2 = \|A_1^{-1}B_1\|_2^2 \leqslant \frac{f_{1,\max}}{f_{1,\min}},$$

with the last inequality following from the proof of Lemma A2. Bringing this all together gives

$$\begin{aligned}|\delta_2| & \leqslant 2\sqrt{\frac{m-n}{m}}\sqrt{1 + \frac{f_{1,\max}}{f_{1,\min}}\Delta_1} + \frac{m-n}{m}\Big(1 + \frac{f_{1,\max}}{f_{1,\min}}\Big)\Delta_1 \\ & = O\Big(\sqrt{\frac{m-n}{m}}\Delta_1\Big),\end{aligned}$$

establishing the second term of the theorem. □

References

Bandyopadhyay, S. & Lahiri, S. (2009). Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhyā* **71**, 221–59.

Bandyopadhyay, S., Lahiri, S. N. & Nordman, D. J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *Ann. Statist.* **43**, 519–45.

Chow, Y.-S. & Grenander, U. (1985). A sieve method for the spectral density. *Ann. Statist.* **13**, 998–1010.

Dahlhaus, R. & Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74**, 877–82.

Deb, S., Pourahmadi, M. & Wu, W. B. (2017). An asymptotic theory for spectral analysis of random fields. *Electron. J. Statist.* **11**, 4297–322.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* **89**, 197–210.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *J. Am. Statist. Assoc.* **102**, 321–31.

Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Am. Statist. Assoc.* **102**, 359–78.

Greenbaum, A. (1997). *Iterative Methods for Solving Linear Systems*. Philadelphia: SIAM.

GUINNESS, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* **60**, 415–29.
GUINNESS, J. & FUENTES, M. (2017). Circulant embedding of approximate covariances for inference from Gaussian data on large lattices. *J. Comp. Graph. Statist.* **26**, 88–97.
GUINNESS, J. & KATZFUSS, M. (2018). *GpGp: Fast Gaussian Process Computation Using Vecchia's Approximation*. R package version 0.1.0, available at https://CRAN.R-project.org/package=GpGp.
GUYON, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* **69**, 95–105.
HEATON, M. J., DATTA, A., FINLEY, A., FURRER, R., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D., KATZFUSS, M., LINDGREN, F. et al. (2018). A case study competition among methods for analyzing large spatial data. *arXiv*:1710.05013v2.
HEYDE, C. & GAY, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stoch. Proces. Appl.* **45**, 169–82.
KÖRNER, T. W. (1989). *Fourier Analysis*. Cambridge: Cambridge University Press.
LEE, T. C. (1997). A simple span selector for periodogram smoothing. *Biometrika* **84**, 965–9.
LEE, T. C. (2001). A stabilized bandwidth selection method for kernel smoothing of the periodogram. *Sig. Proces.* **81**, 419–30.
LEE, T. C. & ZHU, Z. (2009). Nonparametric spectral density estimation with missing observations. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP 2009* 19–24 April, Taipei, Taiwan). Piscataway, New Jersey: IEEE, DOI: 10.1109/ICASSP.2009.4960265.
LIM, C. Y. & STEIN, M. (2008). Properties of spatial cross-periodograms using fixed-domain asymptotics. *J. Mult. Anal.* **99**, 1962–84.
LINDGREN, F., RUE, H. & LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Statist. Soc.* B **73**, 423–98.
LITTLE, R. J. & RUBIN, D. B. (2014). *Statistical Analysis with Missing Data*. Hoboken, New Jersey: John Wiley & Sons.
MATSUDA, Y. & YAJIMA, Y. (2009). Fourier analysis of irregularly spaced data on R^d. *J. R. Statist. Soc.* B **71**, 191–217.
NYCHKA, D., BANDYOPADHYAY, S., HAMMERLING, D., LINDGREN, F. & SAIN, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *J. Comp. Graph. Statist.* **24**, 579–99.
OMBAO, H. C., RAZ, J. A., STRAWDERMAN, R. L. & VON SACHS, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika* **88**, 1186–92.
PAWITAN, Y. & O'SULLIVAN, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *J. Am. Statist. Assoc.* **89**, 600–10.
POLITIS, D. N. & ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.* **16**, 67–103.
R DEVELOPMENT CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.
STEIN, M. L. (1995). Fixed-domain asymptotics for spatial periodograms. *J. Am. Statist. Assoc.* **90**, 1277–88.
STROUD, J. R., STEIN, M. L. & LYSEN, S. (2017). Bayesian and maximum likelihood estimation for Gaussian processes on an incomplete lattice. *J. Comp. Graph. Statist.* **26**, 108–20.
SUBBA RAO, S. (2018). Statistical inference for spatial statistics defined in the Fourier domain. *Ann. Statist.* **46**, 469–99.
SYKULSKI, A. M., OLHEDE, S. C., GUILLAUMIN, A. P., LILLY, J. M. & EARLY, J. J. (2019). The debiased Whittle likelihood. *Biometrika* **106**, 251–66.
TANNER, M. A. & WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc.* **82**, 528–40.
VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. R. Statist. Soc.* B **50**, 297–312.
WAHBA, G. (1980). Automatic smoothing of the log periodogram. *J. Am. Statist. Assoc.* **75**, 122–32.
WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41**, 434–49.
ZHENG, Y., ZHU, J. & ROY, A. (2009). Nonparametric Bayesian inference for the spectral density function of a random field. *Biometrika* **97**, 238–45.

[*Received on* 24 *October* 2017. *Editorial decision on* 18 *October* 2018]