Nonparametric spectral methods for multivariate spatial and spatial–temporal data

Joseph Guinness

Cornell University, Ithaca, NY, USA

ARTICLE INFO

ABSTRACT

We propose computationally efficient methods for estimating stationary multivariate spatial and spatial–temporal spectra from incomplete gridded data. The methods are iterative and rely on successive imputation of data and updating of model estimates. Imputations are done according to a periodic model on an expanded domain. The periodicity of the imputations is a key feature that reduces edge effects in the periodogram and is facilitated by efficient circulant embedding techniques. In addition, we describe efficient methods for decomposing the estimated cross spectral density function into a linear model of coregionalization plus a residual process. The methods are applied to two storm datasets, one of which is from Hurricane Florence, which struck the southeastern United States in September 2018. The application demonstrates how fitted models from different datasets can be compared, and how the methods are computationally feasible on datasets with more than 200,000 total observations.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

There are several exciting scientific campaigns to produce and observe multivariate data that vary over space and time. For example, large climate centers such as the National Center for Atmospheric Research (NCAR) produce high resolution simulations of the Earth system. These models include dozens of variables evolving in concert over space and time. The National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) have deployed numerous satellites that collect observations of the Earth surface and atmosphere. NASA and NOAA recently launched a pair of satellites in its ongoing geostationary operational environmental satellite (GOES) program, the GOES-16 and GOES-17 spacecraft, that sit in geostationary orbit, continually monitoring light reflected by the Earth surface and atmosphere in 16 separate wavelength bands. In turn, the raw data are processed to produce dozens of physically relevant variables, such as atmospheric water vapor content and land surface temperature.

A statistical framework for analyzing these data should, at a minimum, include (1) sufficiently flexible statistical models capable of capturing complex multivariate dependencies, and (2) computationally efficient tools for estimating the models from data. [10] provide a thorough review of existing modeling and estimation frameworks. [18] conducted a theoretical analysis of a number of multivariate spatial models from the literature and concluded that many of them impose oversimplistic restrictions on the coherence between pairs of variables. For example, separable and kernel convolution multivariate spatial models [22] have constant coherence. In the linear model of coregionalization (LMC) [3], when component processes have spectral densities that decay with different rates, the coherence always converges to a non-zero constant as the frequency increases.

It is desirable to have models that do not impose such restrictions, for example, to have a model class that allows the coherence to decay to zero as frequency increases. [18] showed that the multivariate Matérn models [2,12] do possess such flexibility. However, likelihood-based estimation of parameters in multivariate Matérn models is not computationally feasible for the massive datasets mentioned in our opening paragraph. Computational issues come from two sources. The first is that the data size prohibits formation and factoring of the covariance matrix, necessary operations for evaluating the likelihood function. There exist promising approximations that could in principle apply to the multivariate case, such as Vecchia's likelihood approximation [8,13,17,26] and hierarchical matrix approximations [1,21,25], but these approximations are untested for multivariate models, and it is not yet clear how to optimally implement them, and how their performance will compare to univariate cases. Such work would be a welcome development to the literature. However, a second and perhaps more serious computational problem, even for fast approximate methods, is that multivariate spatial models contain many more parameters, usually on the order of the square of the number of components. The large number of parameters poses a serious issue for optimization of the (approximate) likelihood function, which typically requires many iterations until convergence over a large parameter space. We demonstrate this problem with a simple example in Section 4.

Nonparametric spectral methods offer the potential of addressing both the modeling and computational demands for multivariate spatial and spatial–temporal data. In this framework, a discrete Fourier transform (DFT) is applied to each multivariate component, and then the periodogram vectors are smoothed over frequency to generate nonparametric estimates of the cross spectral density matrices. This approach is reasonably flexible from a modeling standpoint, since the only assumption is that the cross spectral density matrices vary smoothly with frequency. For example, the class of stationary spatial–temporal models includes non-separable models and asymmetric spatial–temporal models [11]. The approach is also computationally inexpensive, since we can evaluate the DFTs quickly with FFT algorithms. However, nonparametric spectral methods have limited applicability since they typically apply only to complete, gridded data on rectangular domains. Moreover, even if we have such data, edge effects can introduce severe biases in the spatial and especially the spatial–temporal cases [16]. Tapering [7] and differencing [20] have the potential to reduce edge effects, but differencing is not possible when there are missing values, and [14] showed that tapering can be ineffective under certain missingness scenarios and suboptimal compared to imputation-based methods.

This paper provides methods that addresses edge effects and extend the applicability of nonparametric spectral methods to incomplete gridded multivariate spatial–temporal data. This is achieved by leveraging the simple, yet powerful technique of periodic imputation introduced for the univariate case in [14]. We also introduce a method for decomposing the estimated cross spectral density function into an LMC cross spectral density function plus a residual cross spectral density function, which can be used as tool for exploring the spatial–temporal variation in the data. We apply the methods to compare the multivariate spatial–temporal covariances of data from an ordinary storm and from Hurricane Florence, demonstrating that these methods can computationally feasibly estimate multivariate spatial–temporal models with more than 200,000 observations. We also show how the estimates can be interpreted to distinguish among models.

2. Iterative spectrum estimator

2.1. Model and notation

Let $x \in \mathbb{Z}^d$ be a location on the d-dimensional integer lattice, and let $\boldsymbol{Y}(x) = (Y_1(x), \ldots, Y_p(x))^T \in \mathbb{R}^p$ be a p-variate mean-zero random process on $\mathbb{Z}^d$. The process $\boldsymbol{Y}$ is stationary if $x_1 - x_2 = x_3 - x_4$ implies that $E(\boldsymbol{Y}(x_1)\boldsymbol{Y}(x_2)^T) = E(\boldsymbol{Y}(x_3)\boldsymbol{Y}(x_4)^T)$, in which case we can define $\boldsymbol{K}(h) = E(\boldsymbol{Y}(x)\boldsymbol{Y}(x+h)^T)$ as the covariance between the process at pairs of locations separated by lag $h = (h_1, \ldots, h_d)$. Cramér's Theorem [6] states that $\boldsymbol{K}$ has the representation

$$\boldsymbol{K}(h) = \int_{[0,1]^d} \boldsymbol{f}(\omega) \exp(2\pi i \omega \cdot h) d\omega, \qquad (1)$$

where $i = \sqrt{-1}$, and $\omega \cdot h = \omega_1 h_1 + \cdots + \omega_d h_d$. The matrix-valued function $\boldsymbol{f}(\omega) \in \mathbb{C}^{p \times p}$ has (j, k) entry $f_{jk}(\omega)$, and must be a Hermitian positive definite matrix in order that $\boldsymbol{K}$ is a real and positive definite function. We call $\boldsymbol{f}$ the cross-spectral density (CSD) function.

Let $j_1, \ldots, j_n \in \{1, \ldots, p\}$ and $x_1, \ldots, x_n$ be a sequence of locations and $U = (Y_{j_1}(x_1), \ldots, Y_{j_n}(x_n))$ be the vector of observations. In this notation, n is the total number of individual observations, and we may observe between 0 and p of the components at any location. Because the process is stationary, we can assume without loss of generality that each location x_k falls "northeast" of the location $(1, \ldots, 1)$ and "southwest" of some location $\boldsymbol{a} = (a_1, \ldots, a_d)$. Formally we say that x_k falls in the observation lattice $\mathbb{X}_{\boldsymbol{a}}$, where

$$\mathbb{X}_{\boldsymbol{a}} = \{(r_1, \ldots, r_d) : r_k \in \{1, \ldots, a_k\}\}. \qquad (2)$$

The methods in this paper involve imputing the multivariate process onto a domain larger than the observation lattice $\mathbb{X}_{\boldsymbol{a}}$. To this end, define $\boldsymbol{b} = (b_1, \ldots, b_d)$ with $b_j > a_j$ for each j, and the corresponding embedding lattice $\mathbb{X}_{\boldsymbol{b}}$. Let (U, V) denote a complete vector of observations on $\mathbb{X}_{\boldsymbol{b}}$, that is (U, V) contains each of the p elements of $\boldsymbol{Y}(x)$ at each location $x \in \mathbb{X}_{\boldsymbol{b}}$, ordered with the observations U first, then the missing values V second. Fig. 1 provides a visual description of these definitions.
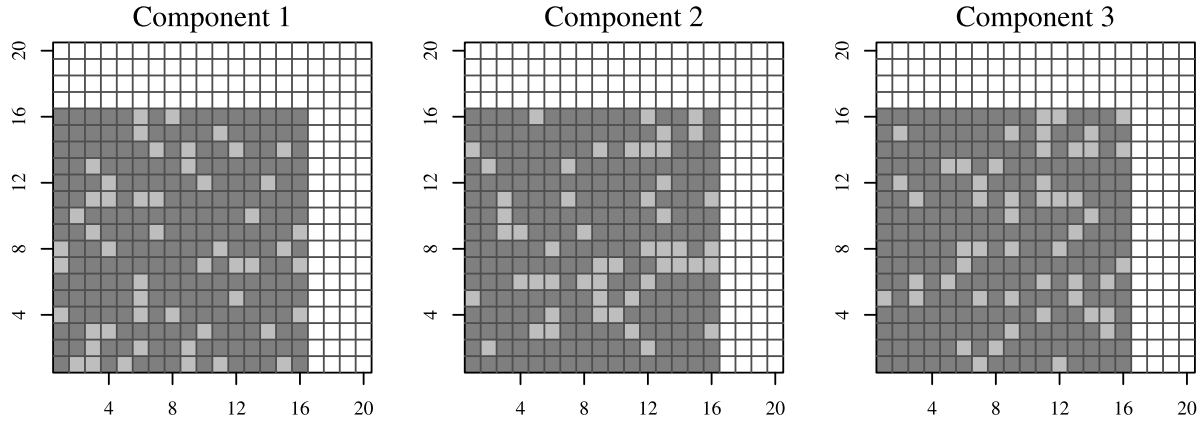
Fig. 1. For a $p = 3$ component dataset, observation lattice $\mathbb{X}_{\boldsymbol{a}}$ (light + dark gray) and embedding lattice $\mathbb{X}_{\boldsymbol{b}}$ (gray + white), for $\boldsymbol{a} = (16, 16)$ and $\boldsymbol{b} = (20, 20)$. Dark gray represents locations for observations in U, light gray and white represent locations for observations in V. Note that locations of missing data on $\mathbb{X}_{\boldsymbol{a}}$ need not be the same in each component.

In the next subsection, we describe methods for iteratively imputing the missing values V on $\mathbb{X}_{\boldsymbol{b}}$. However, we first describe how we obtain a CSD estimate given a complete set of observations (U, V) on $\mathbb{X}_{\boldsymbol{b}}$. We propose the following routine for obtaining a CSD estimate $\widehat{\boldsymbol{f}}[(U, V)]$:

$$\mathcal{Y}_j(\omega) = \frac{1}{\sqrt{m}} \sum_{x \in \mathbb{X}_{\boldsymbol{b}}} Y_j(x) \exp(-2\pi i \omega \cdot x), \tag{3}$$

$$\theta_j = \arg\max_{\theta} \sum_{\omega \in \mathbb{F}_{\boldsymbol{b}}} \left[-\log f_\theta(\omega) - \frac{|\mathcal{Y}_j(\omega)|^2}{f_\theta(\omega)} \right], \tag{4}$$

$$\widehat{f}_{jk}(\omega) = \left(f_{\theta_j}(\omega) f_{\theta_k}(\omega) \right)^{1/2} \sum_{v \in \mathbb{F}_{\boldsymbol{b}}} \frac{\mathcal{Y}_j(v) \mathcal{Y}_k(v)^*}{\left(f_{\theta_j}(v) f_{\theta_k}(v) \right)^{1/2}} \gamma(\omega - v). \tag{5}$$

The symbol $\mathbb{F}_{\boldsymbol{b}}$ refers to the set of Fourier frequencies on a grid of size b, and $*$ is complex conjugate. Each computation written in terms of ω is computed for every $\omega \in \mathbb{F}_{\boldsymbol{b}}$. If the equation is written in terms of j, it is computed for each $j \in \{1, \ldots, p\}$, and likewise, equations written in terms of j and k are computed also for each $k \in \{1, \ldots, p\}$. In (3), we compute the DFT of the complete dataset on $\mathbb{X}_{\boldsymbol{b}}$. In (4), we maximize Whittle's loglikelihood approximation for some parametric spectral density f_θ. In Sections 4 and 5 we use the quasi Matérn spectral density proposed in [15]. In (5), we smooth the sample covariances of the normalized DFT entries with kernel γ, and multiply by the square roots of the estimated parametric spectral densities.

[14] suggested the parametric normalization in (5) as a method for reducing smoothing bias, which allows for the use of wider smoothing kernels, resulting in estimates with smaller variance. If instead f_θ is assumed to be constant, the routine for obtaining a CSD estimate reduces to a standard routine described, for example, in [4] for multivariate time series. Nonparametric or semiparametric CSD estimation methods are desirable because the estimated CSD function $\widehat{\boldsymbol{f}}(\omega)$ is automatically positive definite at every frequency if the support of γ includes at least p Fourier frequencies in $\mathbb{F}_{\boldsymbol{b}}$. However, when the dimension d of the field is larger than 1, substantial bias is imparted by edge effects [16,20]. In the next subsection, we describe iterative imputation methods for reducing edge effect bias.

2.2. Iterative periodic imputation methods

[19] proposed a method for estimating a univariate time series spectrum from incomplete data by iteratively imputing the missing data on the observation domain. [14] showed that when $d > 1$, univariate spectral density estimates can be improved dramatically when data are iteratively imputed onto embedding lattice $\mathbb{X}_{\boldsymbol{b}}$ with a conditional simulation from a periodic model on $\mathbb{X}_{\boldsymbol{b}}$. Here, we propose an extension of the periodic imputation method to the estimation of multivariate cross spectral densities. At iteration ℓ, the missing values V are imputed according to a Gaussian process model with a covariance function that is periodic on $\mathbb{X}_{\boldsymbol{b}}$,

$$\boldsymbol{R}^{(\ell)}(h) = \frac{1}{m} \sum_{\omega \in \mathbb{F}_{\boldsymbol{b}}} \boldsymbol{f}^{(\ell)}(\omega) \exp(2\pi i \omega \cdot h), \tag{6}$$

where $\boldsymbol{f}^{(\ell)}$ is the CSD function at the ℓth iteration, and $m = \prod_{i=1}^{d} b_i$ is the total number of locations in $\mathbb{X}_{\boldsymbol{b}}$. With the complete dataset (U, V), we update the spectrum estimate with $\boldsymbol{f}^{(\ell+1)} = \widehat{\boldsymbol{f}}[(U, V)]$ and iterate. After a burn-in period of

B iterations, we update with a weighted average of the previous and current spectrum, and we monitor changes in $\boldsymbol{f}^{(\ell)}$ for convergence. The full estimation algorithm is as follows:

Estimation Algorithm:
Initialize with $\boldsymbol{f}^{(1)}(\omega) = \widehat{\boldsymbol{f}}[(U, 0)]$.
For $\ell \in \{1, \ldots, B\}$,

1. Simulate V given U from Gaussian process with covariance $\boldsymbol{R}^{(\ell)}$.
2. Given (U, V), set $\boldsymbol{f}^{(\ell+1)} = \widehat{\boldsymbol{f}}[(U, V)]$.

For $\ell > B$

1. Simulate V given U from Gaussian process with covariance $\boldsymbol{R}^{(\ell)}$.
2. Given (U, V), set

$$\boldsymbol{f}^{(\ell+1)} = \frac{\ell - B - 1}{\ell - B} \boldsymbol{f}^{(\ell)} + \frac{1}{\ell - B} \widehat{\boldsymbol{f}}[(U, V)].$$

3. Stop if

$$\max_{j,\omega} \frac{|f_{jj}^{(\ell+1)}(\omega) - f_{jj}^{(\ell)}(\omega)|}{f_{jj}^{(\ell)}(\omega)} < \varepsilon$$

and return CSD estimate $\widehat{\boldsymbol{f}} = \boldsymbol{f}^{(\ell+1)}$.

Drawing Gaussian V given U is the main computational challenge for the iterative methods. All of the other operations are either pointwise multiplications, divisions, DFTs, or convolutions, which can be computed with a DFT. Drawing V given U involves an unconditional simulation from $\boldsymbol{R}^{(\ell)}$ on $\mathbb{X}_{\boldsymbol{b}}$, which can be done with DFTs and pointwise multiplications. It also involves solving a linear system with the covariance matrix for U under $\boldsymbol{R}^{(\ell)}$. This is the most demanding step. We use preconditioned conjugate gradient (PCG), which is an iterative method for solving positive definite linear systems [24]. The forward multiplications in PCG can be done in $O(pm \log m + p^2 m)$ with circulant embedding techniques [27]. We use a preconditioner based on Vecchia's approximation [26] that can be computed in $O(pm)$ time.

3. Factor decompositions

The CSD $\boldsymbol{f}(\omega)$ has an interpretation as the covariance matrix for $\mathcal{Y}(\omega)$, the variation in $\boldsymbol{Y}$ at frequency ω. This is admittedly difficult to communicate to a diverse audience. One solution is to simply report the estimated cross covariance function, the inverse DFT of $\widehat{\boldsymbol{f}}$. To further facilitate interpretability of a p-dimensional multivariate model, it can be useful to consider lower-dimensional representations for the CSD functions. To this end, we propose a factor model consisting of the linear model of coregionalization (LMC) plus a residual process $\boldsymbol{Z}(x)$,

$$\boldsymbol{Y}(x) = \sum_{j=1}^{J} \boldsymbol{A}_j W_j(x) + \boldsymbol{Z}(x), \tag{7}$$

where $\boldsymbol{A}_j \in \mathbb{R}^p$, $W_1, \ldots, W_J$ are univariate process components of the LMC, independent of each other and of $\boldsymbol{Z}$, a p-variate spatial process with correlated components. The CSD function for this process is

$$\boldsymbol{f}(\omega) = \sum_{j=1}^{J} \boldsymbol{A}_j \boldsymbol{A}_j^T g_j(\omega) + \boldsymbol{h}(\omega), \tag{8}$$

where $g_j(\omega) > 0$ is the univariate spectral density for W_j, and $\boldsymbol{h}$ is the CSD function for $\boldsymbol{Z}$.

There is a clear lack of identifiability in assigning power from $\boldsymbol{f}(\omega)$ to either $\sum \boldsymbol{A}_j \boldsymbol{A}_j^T g_j(\omega)$ or $\boldsymbol{h}(\omega)$; for example, any model written as (8) can be rewritten as $\boldsymbol{f}(\omega) = \boldsymbol{h}_1(\omega)$, where $\boldsymbol{h}_1(\omega) = \sum \boldsymbol{A}_j \boldsymbol{A}_j^T g_j(\omega) + \boldsymbol{h}(\omega)$. For this reason, we argue that it is not meaningful to ask for an "optimal" or "true" value of J. We take the position here that since the purpose of the factor model is to explain variation in the response with a lower-dimensional model, we should assign as much power as possible to the factor term over a range of values of J. This is analogous to how one views the number of components in a principal components analysis. Then the amount of variation explained by the factor term can be used as one way to distinguish among models.

3.1. Profiled optimization

After the CSD function $\widehat{\boldsymbol{f}}$ has been estimated using the proposed periodic imputation methods, we aim to select $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_J$ and $g_1, \ldots, g_J$ in order to minimize the amount of power assigned to the residual process $\boldsymbol{Z}$. To quantify the

power, we consider the sum of the variances of the residual processes, which can be written in terms of the sum of traces of the residual CSD function,

$$\sum_{k=1}^{p} \text{Var}(Z_k(x)) = \frac{1}{m} \sum_{\omega \in \mathbb{F}_b} \text{Tr}(\boldsymbol{h}(\omega)) = \frac{1}{m} \sum_{\omega \in \mathbb{F}_b} \text{Tr}\left(\widehat{\boldsymbol{f}}(\omega) - \sum_{j=1}^{J} \boldsymbol{A}_j \boldsymbol{A}_j^T g_j(\omega)\right), \tag{9}$$

subject to $g_j(\omega) > 0$ and $\widehat{\boldsymbol{f}}(\omega) - \sum \boldsymbol{A}_j \boldsymbol{A}_j^T g_j(\omega)$ nonnegative definite for each ω. Even for $J = 1$, minimizing the loss function over all $g_j(\omega)$ is demanding because there are potentially thousands or millions individual frequencies $\omega \in \mathbb{F}_b$. This intractability has motivated a theoretical study to find closed-form minimizers over $g_j(\omega)$ for fixed $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_J)$. This allows us to solve the global optimization problem by profiling out $g_j(\omega)$ and numerically minimizing the sum-of-variance criterion over $\boldsymbol{A}$. Further, we can assume without loss of generality that each $\boldsymbol{A}_j$ lies on the unit sphere because the radius of $\boldsymbol{A}_j$ can be absorbed into $g_j(\omega)$, so the numerical optimization is over a $J(p-1)$-dimensional space.

The sum-of-variance criterion can be rewritten as

$$\frac{1}{m} \sum_{\omega \in \mathbb{F}_b} \sum_{k=1}^{p} \widehat{f}_{kk}(\omega) - \frac{1}{m} \sum_{\omega \in \mathbb{F}_b} \sum_{k=1}^{p} \sum_{j=1}^{J} A_{jk}^2 g_j(\omega) = \frac{1}{m} \sum_{\omega \in \mathbb{F}_b} \sum_{k=1}^{p} \widehat{f}_{kk}(\omega) - \sum_{\omega \in \mathbb{F}_b} \sum_{j=1}^{J} g_j(\omega) \sum_{k=1}^{p} A_{jk}^2. \tag{10}$$

The first term on the right does not depend on g_j or $\boldsymbol{A}_j$, and so minimizing the criterion corresponds to maximizing the second term subject to $g_j(\omega) > 0$ and $\boldsymbol{h}(\omega)$ nonnegative definite for each ω. For fixed $\boldsymbol{A}$ with $\sum_k A_{jk}^2 = \|\boldsymbol{A}_j\|_2 = 1$, this corresponds to maximizing $\sum_\omega \sum_j g_j(\omega)$. Since the nonnegative definiteness criterion applies separately to each ω, we can maximize $\sum_j g_j(\omega)$ separately for each ω.

The theorem gives a result for finding the minimizers over $g_1(\omega), \ldots, g_J(\omega)$ for fixed $\boldsymbol{A}$. Following the theorem are two corollaries that give specific results for the cases $J = 1$ and $J = 2$. To state the theorem, parameterize nonnegative $g_j(\omega)$ as $\exp(c_j)$, and define the $J \times J$ matrices

$$\boldsymbol{C} = \text{diag}(e^{c_1}, \ldots, e^{c_J}), \quad \boldsymbol{B} = \boldsymbol{A}^T \widehat{\boldsymbol{f}}(\omega)^{-1} \boldsymbol{A}. \tag{11}$$

Further, define the Lagrangian function

$$L(c_1, \ldots, c_J, \lambda) = e^{c_1} + \cdots + e^{c_J} + \lambda \det(\boldsymbol{C}^{-1} - \boldsymbol{B}) \tag{12}$$

and its gradient ∇L.

Theorem 1. *For fixed $\boldsymbol{A} = (\boldsymbol{A}_1, \ldots, \boldsymbol{A}_J)$, the minimizer of*

$$Tr(\boldsymbol{h}(\omega)) = Tr\left(\widehat{\boldsymbol{f}}(\omega) - \sum_{j=1}^{J} e^{c_j} \boldsymbol{A}_j \boldsymbol{A}_j^T\right)$$

with respect to $c_1, \ldots, c_J$, subject to $\boldsymbol{h}(\omega)$ nonnegative definite, is either a solution to $\nabla L = \boldsymbol{0}$ or for some $k \in \{1, \ldots, J\}$, the minimizer of

$$Tr\left(\widehat{\boldsymbol{f}}(\omega) - \sum_{j \neq k} e^{c_j} \boldsymbol{A}_j \boldsymbol{A}_j^T\right).$$

Proof. Assume without loss of generality that $\|\boldsymbol{A}_j\|_2 = 1$ for every j. We first show that $\boldsymbol{h}(\omega)$ is nonnegative definite but not strictly positive definite at the minimizer. To establish a contradiction, suppose that $\boldsymbol{h}(\omega)$ is strictly positive definite at the minimizer. Thus there exists $\varepsilon > 0$ such that for any $\|\boldsymbol{u}\|_2 = 1$,

$$\boldsymbol{u}^T \left(\widehat{\boldsymbol{f}}(\omega) - \sum_{j=1}^{J} \boldsymbol{A}_j \boldsymbol{A}_j^T e^{c_j}\right) \boldsymbol{u} > \varepsilon.$$

Now consider the quadratic form

$$\boldsymbol{u}^T (\widehat{\boldsymbol{f}}(\omega) - \sum_{j=1}^{J} \boldsymbol{A}_j \boldsymbol{A}_j^T (e^{c_j} + \varepsilon/J)) \boldsymbol{u} > \varepsilon - \frac{\varepsilon}{J} \sum_{j=1}^{J} (\boldsymbol{u}^T \boldsymbol{A}_j)^2 \geq \varepsilon - \varepsilon = 0.$$

This establishes that each e^{c_j} could have been increased by ε/J, and thus were not the minimizers of the trace, giving a contradiction.

This means that the determinant of $\boldsymbol{h}(\omega)$ is zero at the minimizer. Consider the matrix

$$\begin{bmatrix} \widehat{\boldsymbol{f}}(\omega) & \boldsymbol{A} \\ \boldsymbol{A}^T & \boldsymbol{C}^{-1} \end{bmatrix}.$$

Its determinant must be zero because it is equal to $\det(\boldsymbol{C}^{-1})\det(\boldsymbol{h}(\omega))$. Its determinant is also equal to

$$\det(\widehat{\boldsymbol{f}}(\omega))\det(\boldsymbol{C}^{-1} - \boldsymbol{B}).$$

We know that $\det(\widehat{\boldsymbol{f}}(\omega)) > 0$ because $\widehat{\boldsymbol{f}}(\omega)$ is strictly positive definite; therefore $\det(\boldsymbol{C}^{-1} - \boldsymbol{B}) = 0$ at the minimizer.

The theorem follows from using the method of Lagrange multipliers, that is, we seek to maximize $\sum_{j=1}^{J} e^{c_j}$ subject to $\det(\boldsymbol{C}^{-1} - \boldsymbol{B}) = 0$, and checking the endpoints $e^{c_j} = 0$. $\square$

The theorem states that the minimizers either solve the gradient of the Lagrangian or are the solution to a subproblem. This arises because we must check the boundaries of the parameter space when using the method of Lagrange multipliers. We must reparameterize $g_j(\omega) = \exp(c_j)$ because it is possible that the maximizer of $\sum_j g_j(\omega)$ subject to $\boldsymbol{h}(\omega)$ nonnegative definite includes one or more negative $g_j(\omega)$. The following two corollaries give concrete solutions for the cases $J = 1$ and $J = 2$.

Corollary 1. *For fixed $\boldsymbol{A}_1$, $Tr(\boldsymbol{f}(\omega) - \boldsymbol{A}_1\boldsymbol{A}_1^T g_1(\omega))$ is minimized by $g_1(\omega) = (\boldsymbol{A}^T\boldsymbol{f}(\omega)^{-1}\boldsymbol{A}_1^T)^{-1}$.*

Proof. When $J = 1$, $\boldsymbol{C}^{-1} = e^{-c_1}$, and $\boldsymbol{B} = B_{11} = \boldsymbol{A}_1^T\widehat{\boldsymbol{f}}(\omega)^{-1}\boldsymbol{A}_1$. Therefore, $\det(\boldsymbol{C}^{-1} - \boldsymbol{B}) = 0$ if and only if $e^{-c_1} = B_{11}$ if and only if $e^{c_1} = 1/B_{11}$. $\square$

Corollary 2. *For fixed $\boldsymbol{A}_1, \boldsymbol{A}_2$, $Tr(\boldsymbol{f}(\omega) - \boldsymbol{A}_1\boldsymbol{A}_1^T g_1(\omega) - \boldsymbol{A}_2\boldsymbol{A}_2^T g_2(\omega))$ is minimized by either*

(1) $g_1(\omega) = 0$ and $g_2(\omega) = 1/B_{22}$,
(2) $g_2(\omega) = 0$ and $g_1(\omega) = 1/B_{11}$,
(3)

$$g_1(\omega) = \frac{B_{22} - \sqrt{B_{12}B_{21}}}{\det(\boldsymbol{B})}, \quad g_2(\omega) = \frac{B_{11} - \sqrt{B_{12}B_{21}}}{\det(\boldsymbol{B})}.$$

Proof. Setting $\nabla L = \boldsymbol{0}$ yields the three equations

$$e^{c_1} + \lambda[e^{-c_1}(e^{-c_2} - B_{22})] = 0, \quad e^{c_2} + \lambda[e^{-c_2}(e^{-c_1} - B_{11})] = 0, \quad (e^{-c_1} - B_{11})(e^{-c_2} - B_{22}) - B_{12}B_{21} = 0.$$

Eliminating λ gives the following two equations

$$e^{c_1}(1 - B_{11}e^{c_1}) = e^{c_2}(1 - B_{22}e^{c_2}), \quad (e^{-c_1} - B_{11})(e^{-c_2} - B_{22}) - B_{12}B_{21} = 0.$$

One can verify that the solution is

$$e^{c_1} = \frac{B_{22} - \sqrt{B_{12}B_{21}}}{\det(\boldsymbol{B})}, \quad e^{c_2} = \frac{B_{11} - \sqrt{B_{12}B_{21}}}{\det(\boldsymbol{B})}.$$

According to Theorem 1, the solution is either the above, or the solution to one of the one-factor problems, establishing the corollary. $\square$

3.2. Conditional expectation of component processes

The conditional expectations of the jth factor component given the data are

$$E(W_j(x)|U) = E_{V|U}[E(W_j(x)|U, V)] = E_{V|U}\left[E\left(\frac{1}{\sqrt{m}} \sum_{\omega \in \mathbb{F}_{\boldsymbol{b}}} \mathcal{W}_j(\omega)e^{i\omega \cdot x} \middle| \mathcal{Y}(\mathbb{F}_{\boldsymbol{b}}) \right) \right]$$

$$= \frac{1}{\sqrt{m}} E_{V|U}\left[\sum_{\omega \in \mathbb{F}_{\boldsymbol{b}}} E\left(\mathcal{W}_j(\omega)|\mathcal{Y}(\omega) \right) \right] e^{i\omega \cdot x} = \frac{1}{\sqrt{m}} \sum_{\omega \in \mathbb{F}_{\boldsymbol{b}}} g_j(\omega)\boldsymbol{A}_j^T\boldsymbol{f}(\omega)^{-1} E[\mathcal{Y}(\omega)|U] e^{i\omega \cdot x}.$$

The conditional expectation of $\boldsymbol{Y}(\omega)$ given U is

$$E(\mathcal{Y}(\omega)|U) = E\left(\frac{1}{\sqrt{m}} \sum_{x \in \mathbb{X}_{\boldsymbol{b}}} \boldsymbol{Y}(x)e^{-i\omega \cdot x} \middle| U \right) = \frac{1}{\sqrt{m}} \sum_{x \in \mathbb{X}_{\boldsymbol{b}}} E(\boldsymbol{Y}(x)|U)e^{-i\omega \cdot x},$$

where $\mathcal{W}_j$ is the DFT of W_j. Therefore, in order to compute the conditional expectation of W_j, we simply compute $E(V|U)$, then take the DFT of $(U, E(V|U))$ to obtain $E(\mathcal{Y}(\omega)|U)$, and then take the inverse DFT of

$$g_j(\omega)\boldsymbol{A}_j^T\boldsymbol{f}(\omega)^{-1}E(\mathcal{Y}(\omega)|U).$$

We use the expected factors in Section 5 to explore the decomposition of the estimated spectrum from a thunderstorm dataset.

Table 1
True values of multivariate Matérn parameters in simulation study for three-component model..

Variance σ_{jk}			Inverse range α_{jk}			Smoothness ν_{jk}		
1.00	1.57	1.81	0.25	0.25	0.25	0.500	0.625	0.750
1.57	4.00	4.75	0.25	0.25	0.25	0.625	0.750	0.875
1.81	4.75	9.00	0.25	0.25	0.25	0.750	0.875	1.000

4. Simulations

The purpose of the simulation study is to demonstrate some of the computational issues with maximum likelihood estimation of the multivariate Matérn parameters, and show how our proposed methods are capable of producing fast and accurate estimates of the CSD function, even when we do not assume knowledge of the parametric form of the model.

The cross covariances in the multivariate Matérn model introduced by [12] can be written in terms of the Matérn function,

$$\text{Cov}(Y_j(x), Y_k(x + h)) = K_{jk}(h) = \sigma_{jk}\mathcal{M}(\|h\|\alpha_{jk} \; ; \; \nu_{jk}), \quad \mathcal{M}(r \; ; \; \nu) = \frac{r^\nu}{2^{\nu-1}\Gamma(\nu)}\mathcal{K}_\nu(r); \tag{13}$$

$\mathcal{M}(r \; ; \; \nu)$ is the Matérn function, written in terms of $\mathcal{K}_\nu$, a modified Bessel function of the second kind, with parameter ν. The parameters σ_{jj}^2, α_{jj}, and ν_{jj} represent the marginal variance, the inverse range, and the smoothness of Y_j. These marginal parameters must all be positive. To ensure positive definiteness of the multivariate Matérn, [12] give conditions on the parameters in the bivariate Matérn model, and [2] give conditions on the parameters in a more general p-variate setting.

In the simulation study, we consider $p = 2, 3$, and 4. The inverse range and smoothness parameters are $\alpha_{jk} = 0.25$ and $\nu_{jk} = 0.5 + 0.5(j + k - 2)/(2p - 2)$. To define the variance parameters, let β be a $p \times p$ matrix with $\beta_{jk} = 0.8^{|j-k|}$. The variance parameters are

$$\sigma_{jk} = jk \frac{\Gamma(\nu_{jk})}{\Gamma(\nu_{jk} + 1)} \frac{\sqrt{\Gamma(\nu_{jj} + 1)\Gamma(\nu_{kk} + 1)}}{\sqrt{\Gamma(\nu_{jj})\Gamma(\nu_{kk})}} \beta_{jk}. \tag{14}$$

This particular model parameterization is from the parsimonious multivariate Matérn family [12]. Table 1 presents the specific parameter values used when $p = 3$, and Fig. 2 shows one realization from the $p = 3$ model. Due to the symmetry constraints $\sigma_{jk} = \sigma_{kj}$, $\alpha_{jk} = \alpha_{kj}$, and $\nu_{jk} = \nu_{kj}$, the full multivariate Matérn has $3p(p + 1)/2$ unique parameters. For $p = 2$, this is 6 parameters; for $p = 3$, 18 parameters; for $p = 4$, 30 parameters. A mentioned in Section 1, the large number of free parameters is one aspect of the computational difficulty of working with the multivariate Matérn model.

We specify grid size (16, 16) and simulate 70 datasets for each value of $p \in \{2, 3, 4\}$ with no missing values on the grid. Scenarios with missing values, along with a comparison to tapering, were considered for the univariate case in [14]. To estimate multivariate Matérn parameters, we use the exact Gaussian likelihood function, maximized using the `optim` function in R with the default Nelder–Mead algorithm. The estimation procedure makes no assumptions that the true model is a member of the parsimonious multivariate Matérn family since we aim to evaluate the performance of maximum likelihood on Multivariate Matérn under weak assumptions. We run the Nelder–Mead algorithm for 6000 iterations, stopping at each 1000 iterations to report estimation progress and elapsed time.

For our proposed periodic imputation methods, we take three choices of expansion parameter $\tau \in \{1.00, 1.25, 1.50\}$, and we use the parametric variant of the iterative algorithm, run for 50 burn-in iterations, and an averaging tolerance of $\varepsilon = 0.01$. We use a Gaussian smoothing kernel, with four bandwidth choices in $\{0.15, 0.20, 0.25, 0.30\}$; a bandwidth of b corresponds to $(100b)\%$ of the frequency domain. For the parametric filter, we use the quasi-Matérn spectral density

$$f_\theta(\omega) = \sigma^2 \Big(1 + \alpha^{-2}\big(\sin^2(\omega_1/2) + \sin^2(\omega_2/2)\big)\Big)^{-\nu - d/2}. \tag{15}$$

The parameter σ^2 can be profiled out, so at each iteration, Whittle's likelihood is optimized over the two-dimensional parameter space (α, ν) for each of the p components.

To evaluate the estimators, we consider a spectral norm criterion

$$\frac{1}{m} \sum_{\omega \in \mathbb{F}_b} \lambda_{\max} \Big(\boldsymbol{f}^{-1/2}(\omega)\big(\widehat{\boldsymbol{f}}(\omega) - \boldsymbol{f}(\omega)\big)\boldsymbol{f}^{-1/2}(\omega)\Big) \tag{16}$$

where $\lambda_{\max}(\boldsymbol{M})$ is the largest absolute eigenvalue of matrix $\boldsymbol{M}$, $\boldsymbol{f}$ is the true CSD function, $\boldsymbol{f}^{-1/2}$ is the inverse symmetric square root, and $\widehat{\boldsymbol{f}}$ is the estimate.

The results of the simulation study are given in Table 2. When no periodic embedding is used ($\tau = 1.0$), the estimates are computed extremely fast but are poor relative to the other estimators. The poor performance when $\tau = 1.0$ highlights the fact that spectral estimators based on the periodogram of the observed data can suffer from severe edge effects. The
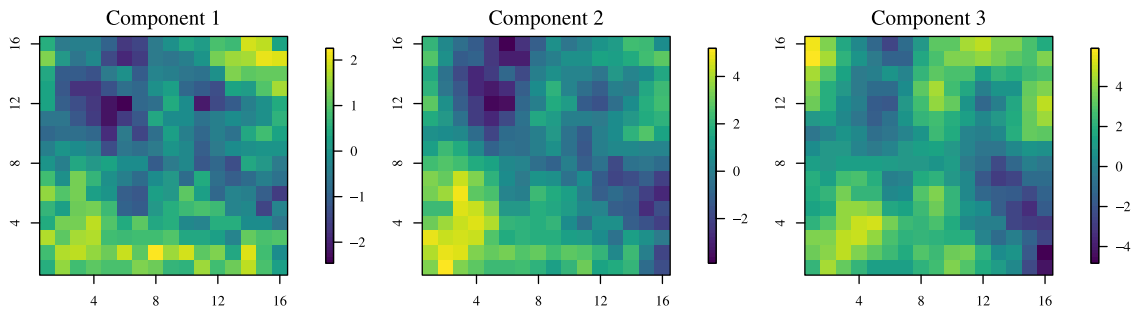
Fig. 2. One realization from the three-component multivariate Matérn model used in the simulation study.

Table 2

Simulation study results, showing 0.25, 0.50 and 0.75 quantiles of spectral norms over 70 simulation replicates for 2-, 3-, and 4-component models. Periodic imputation methods are evaluated at three expansion factors and with kernel bandwidth 0.30. Maximum likelihood is conducted in R with optim function using Nelder–Mead algorithm. Time is average time in minutes over 70 simulation replicates.

Components	Quantile	Periodic imputation expansion factor τ				Maximum likelihood multivariate Matérn Nelder–Mead iterations					
		1.00	1.25	1.50	1000	2000	3000	4000	5000	6000	
2	0.25	1.206	0.253	0.243		0.200	0.181	0.170	0.173	0.170	0.172
	0.50	1.630	0.276	0.266		0.277	0.218	0.208	0.204	0.204	0.201
	0.75	2.540	0.304	0.295		0.341	0.275	0.272	0.245	0.242	0.239
	time	0.078	0.227	0.261		16.2	33.1	50.0	66.5	83.0	98.3
3	0.25	1.300	0.306	0.302		0.579	0.415	0.392	0.386	0.361	0.360
	0.50	1.767	0.337	0.328		0.686	0.487	0.455	0.441	0.440	0.428
	0.75	2.494	0.379	0.358		0.958	0.563	0.526	0.504	0.487	0.488
	time	0.081	0.273	0.296		22.2	44.4	67.4	89.8	112.2	138.3
4	0.25	1.412	0.372	0.358		1.627	0.713	0.587	0.535	0.521	0.504
	0.50	1.888	0.398	0.383		2.096	0.927	0.720	0.667	0.605	0.594
	0.75	2.654	0.429	0.407		2.847	1.197	0.834	0.801	0.719	0.681
	time	0.090	0.363	0.416		29.8	59.8	90.2	118.1	147.4	175.9

periodically embedded estimators are nearly as accurate or more accurate than maximum likelihood, even after 6000 Nelder–Mead iterations. When $p = 2$ the periodic imputation estimators are slightly worse in terms of spectral norm, whereas the periodic imputation estimators are better when $p = 3$ and $p = 4$, even though no knowledge of the true CSD functions is assumed. Further, the periodic imputation methods are much faster. They converge on average in less than 25 s in every case, whereas the maximum likelihood estimates converge in several thousand iterations and take on the order of one to three hours.

The simulations were conducted on a single node of the Cheyenne supercomputer managed by the National Center for Atmospheric Research [5]. Each node has 36 2.3-GHz Intel Xeon E5-2697V4 processors, and each processor has 2 threads. One processor was used to manage the parallel tasks, and each of the remaining 35 processors analyzed two datasets in parallel on the two threads. The computing times in Table 2 should be understood in the context that only a single thread was used to analyze each dataset. We can expect up to, for example, a four-fold speedup for each task when running the code on a two core machine with four total threads.

5. Multivariate spatial–temporal analysis of storms

Launched in November 2016, the GOES-16 satellite now sits in geostationary orbit, allowing its Advanced Baseline Imager (ABI) to capture images of the Earth's atmosphere in 16 wavelength bands at up to 500 m resolution in space and up to 30 s in time. GOES 16 produces terabytes of data per day and provides a wealth of information about atmospheric processes. In this section, we perform a multivariate spatial–temporal analysis of images from two separate storms, one of which is an ordinary convective storm formed over Florida from July 22, 2018, while the other is from Hurricane Florence on September 14, 2018, the day it made landfall with the southeastern U.S. coastline. The analysis is demonstrative of the sort of comparisons that are possible with our proposed methods.

For each storm, we analyze 60 images from 4 wavelength bands separated by 1 min. Band 1 has a native resolution of 1 km, while the resolution of the other bands is 2 km. We locally average the Band 1 data to coincide with the resolution of the other bands. Fig. 3 contains images from Band 1 at the beginning, middle, and end of the hour. A stationary model is unrealistic for the entire scene, and so we subset the data to the black polygons in Fig. 3, where the stationary assumption is more tenable. Fig. 4 contains the subsetted images from the middle time point at the four wavelength bands. Table 3

Table 3

Information about Advanced Baseline Imager bands used in multivariate spatial–temporal analysis.

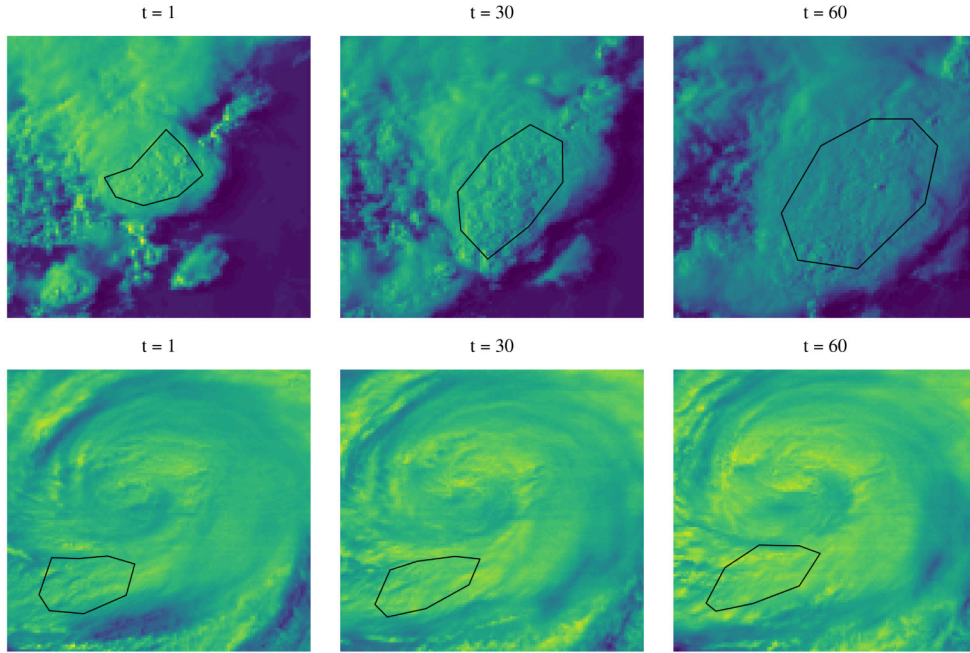| Band	Wavelength	Nickname
1	0.47 microns	Blue Band
6	2.2 microns	Cloud Particle Size Bands
7	3.9 microns	Shortwave Window Band
9	6.9 microns	Mid-Level Tropospheric Water Vapor Band

Fig. 3. Data from ordinary storm (top) and from Hurricane Florence (bottom) at three time points (of 60), with subsetted data indicated by black polygons. Band 1 ("blue" band) plotted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

contains information about the bands selected for the analyses. We choose one band in the visible spectrum since the visible spectrum bands are all highly correlated due to the grayscale color of clouds; Band 6 contains information about particle sizes; Band 7 can be used for ice detection; and Band 9 contains information about water vapor content in the mid-troposphere [23]

The ordinary storm has observation lattice size $a = (55, 57, 60)$ in longitude, latitude, time, with 253,504 total observations; Florence has $a = (47, 31, 60)$, with 167,804 total observations. For both storms, we specify expansion factor $\tau = 1.25$, extending the lattice by 25% in each of the three dimensions, and we again use the quasi Matérn parametric spectral density. Before estimation, we subtract off the sample mean $\widehat{\mu}_k$ from each component. We use $B = 20$ burn-in iterations, and we specify convergence tolerance parameter $\varepsilon = 0.005$. The ordinary storm CSD estimates converge after 6 averaging iterations, which took 3.40 h total for burn-in and convergence. Florence converged after 15 iterations and took 1.43 h. All computations are in the R programming language and run on an Intel Core i5-7200 CPU (2 cores, 4 threads at 2.50 GHz) with 8GB memory. Vecchia's preconditioner is implemented in C++ with the Rcpp package [9].

To visualize the fits, we include plots of the estimated real part of the coherences for the two storms in Fig. 5. We see that for the ordinary storm, the coherence between Bands 1, 6, and 7 persists throughout the frequency domain, while Band 9 has a slight negative coherence with the other bands throughout the domain. The situation is quite different for the Hurricane Florence data; Band 9 has a small amount of positive coherence with Bands 1 and 7 and is negatively coherent with Band 6. The coherences weaken at the highest frequencies. Bands 1, 6, and 7 are weakly coherent.

The total variation differs widely between the four bands; for example, in the ordinary storm, the estimated standard deviations are 41.15, 0.70, 0.01 and 0.20. These differences in variation are not necessarily a reflection of the relative importance of the bands, so instead of decomposing the estimated CSD functions $\widehat{\boldsymbol{f}}$ directly – which would undoubtedly return vectors $\boldsymbol{A}_j$ attempting to explain variation in Band 1 – we decompose the normalized CSD function

$$\widetilde{f}_{jk}(\omega) = \frac{\widehat{f}_{jk}(\omega)}{\sqrt{\widehat{C}_{jj}(0)\widehat{C}_{kk}(0)}}, \tag{17}$$
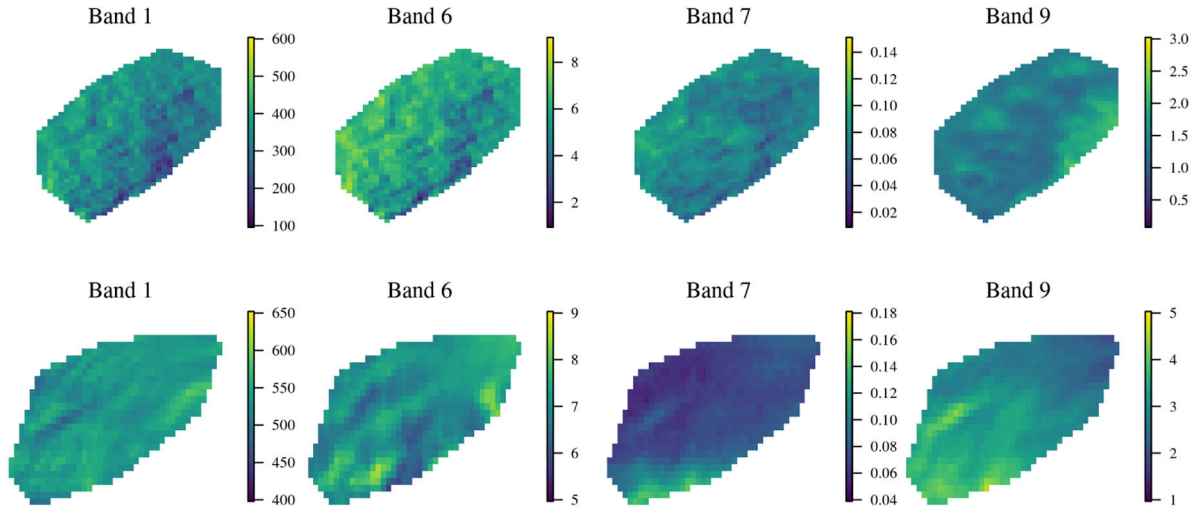
Fig. 4. Data from ordinary storm (top) and from Hurricane Florence (bottom) at $t = 30$, from Bands 1, 6, 7, and 9.
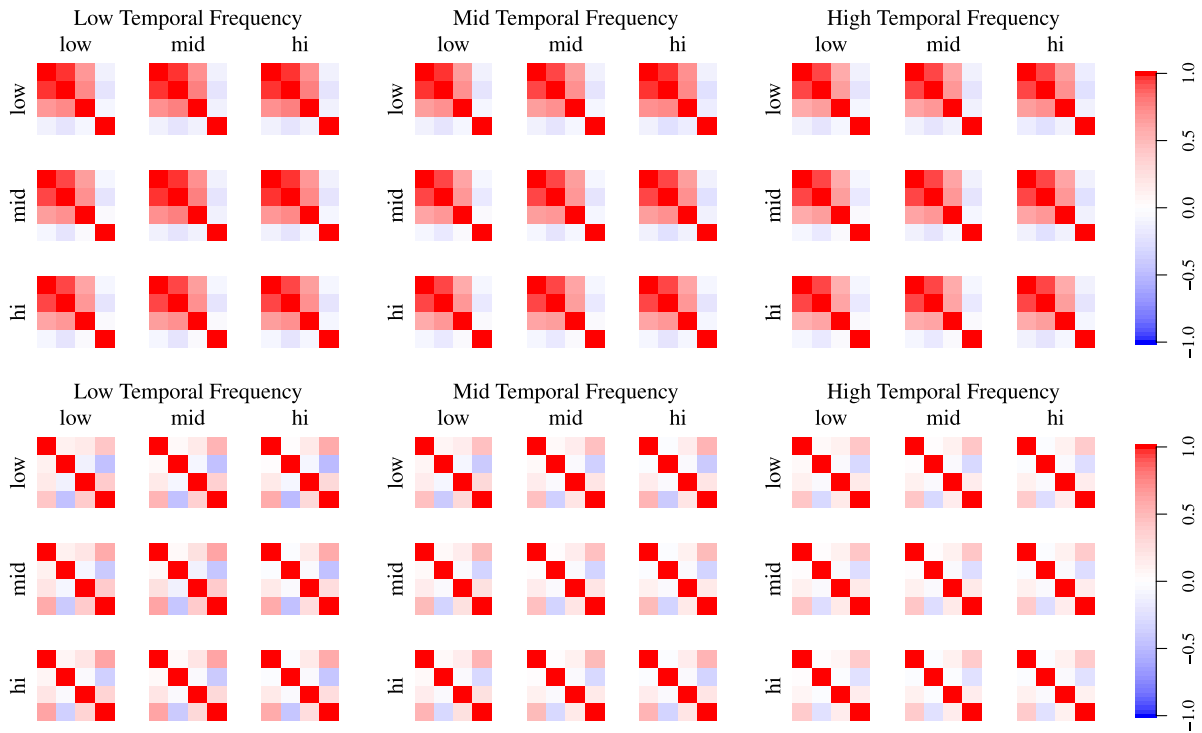

Fig. 5. Estimated real part of coherences for ordinary storm (top row) and Hurricane Florence (bottom row) at various frequencies $\omega = (\omega_1, \omega_2, \omega_3)$. "Low" refers to $\omega_j = 0$, "mid" to $\omega_j = 1/4$, and "high" to $\omega_j = 1/2$. In each 3×3 group of coherence matrices, rows correspond to latitudinal frequencies, columns to longitudinal frequencies.

where $\widehat{C}_{jj}(0)$ is estimated covariance function at lag zero, that is, the estimated variance. Table 1 provides a summary of the decompositions. For the ordinary storm, the $J = 1$ decomposition puts most of its weight on Bands 1, 6, and 7. When $J = 2$, $\boldsymbol{A}_1$ is nearly unchanged, and $\boldsymbol{A}_2$ explains variation in Band 9. For the Hurricane Florence data, when $J = 1$, $\boldsymbol{A}_1$ points mostly in the direction of Bands 1, 6, and 9. When $J = 2$, $\boldsymbol{A}_1$ changes little, and $\boldsymbol{A}_2$ explains common variation in Bands 1 and 6 (see Table 4).

Table 4

Estimated factor decompositions for the $J = 1$ factor model and the $J = 2$ factor model for both the Ordinary Storm and Hurricane Florence, namely the loadings placed on each band in the factor decompositions. The bottom row gives the percent of variance exlained by the factors.

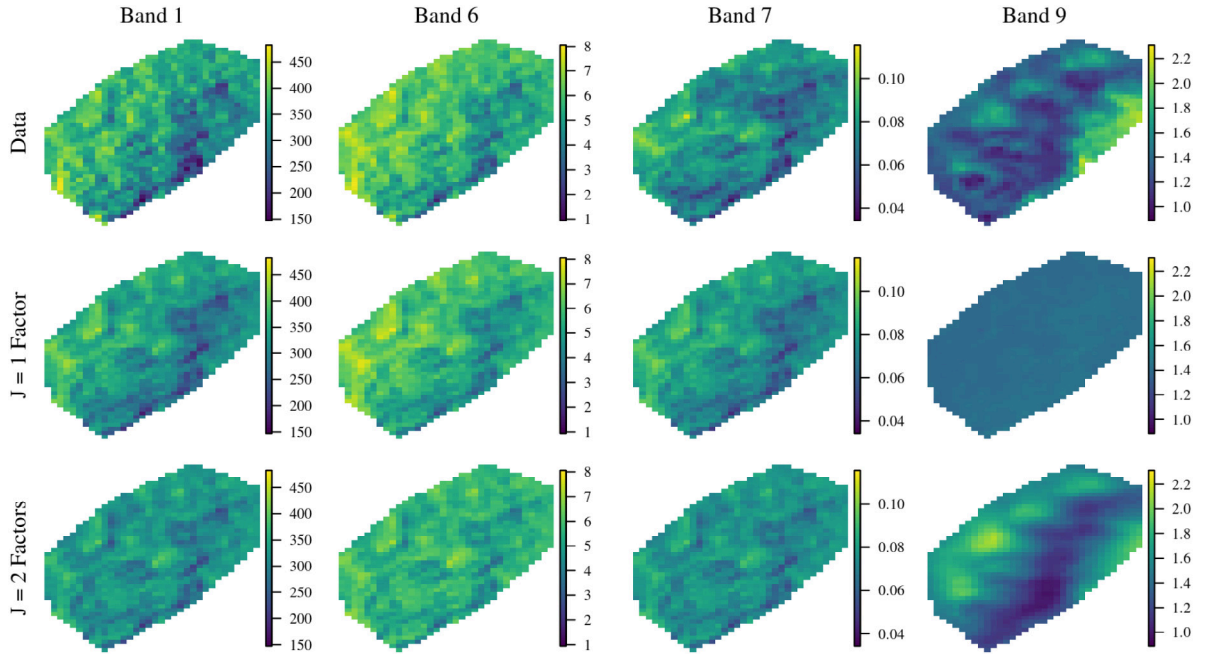| Band	Ordinary Storm			Hurricane Florence		
	$J = 1$	$J = 2$		$J = 1$	$J = 2$	
	A_1	A_1	A_2	A_1	A_1	A_2
1	−0.498	−0.492	0.000	−0.444	−0.446	−0.642
6	−0.634	−0.634	0.000	0.463	0.458	−0.763
7	−0.590	−0.594	0.000	−0.107	−0.104	−0.031
9	0.049	0.054	−1.000	−0.760	−0.762	−0.064
% explained	43.3%	58.5%		34.0%	55.1%	

Fig. 6. Data from time $t = 30$ (top row), one factor representation (middle row), and two factor representation (bottom row). Factor representations use the expected values of W_j given the data.

Lastly, in Fig. 6, we plot the original ordinary storm data from $t = 30$, and the expected $J = 1$ and $J = 2$ factor representations of each band k,

$$\widehat{\mu}_k + \widehat{C}_{kk}(0) \sum_{j=1}^{J} A_{jk} E(W_j(x)|U). \tag{18}$$

We can see that the one-factor representation ignores variation in Band 9 but captures much of the variation in Bands 1, 6, and 7. This is expected since these three bands are strongly coherent across frequencies. The two-factor captures some of the variation in Band 9 but still oversmooths.

6. Discussion

We have introduced simple, flexible, and computationally efficient methods for estimating stationary multivariate spatial–temporal models from incomplete gridded data. The methods rely on successive imputation of data onto an expanded lattice under a model that is periodic on the expanded lattice. The simulation studies demonstrate that the periodic domain expansion is crucial for addressing edge effects; when no expansion is performed, the estimates of the spectrum are poor. The new estimates are competitive with maximum likelihood – though much faster – when there are $p = 2$ multivariate components, and the estimates are faster and more accurate than maximum likelihood with $p > 2$ components. We have argued that this arises both from the computational demand of evaluating the likelihood function, and of maximizing over the large number of parameters in the multivariate Matérn model.

Table 5

Expanded simulation study results. 0.25, 0.50, and 0.75 quantiles of spectral norms for $p = 2$-, 3-, and 4-component models, for all four smoothing kernel bandwidths, 0.15, 0.20, 0.25, and 0.30, for three expansion factors τ, for the periodic imputation methods. Time is average time in minutes to convergence over 70 replicates.

| p	bandwidth	$\tau = 1.00$				$\tau = 1.25$				$\tau = 1.50$			
		0.15	0.20	0.25	0.30	0.15	0.20	0.25	0.30	0.15	0.20	0.25	0.30
2	0.25	1.362	1.303	1.248	1.206	0.305	0.262	0.253	0.253	0.270	0.239	0.241	0.243
	0.50	1.809	1.717	1.660	1.630	0.343	0.286	0.276	0.276	0.305	0.276	0.270	0.266
	0.75	2.726	2.619	2.567	2.540	0.373	0.316	0.303	0.304	0.330	0.294	0.293	0.295
	minutes	0.085	0.075	0.080	0.078	0.289	0.263	0.237	0.227	0.354	0.316	0.296	0.261
3	0.25	1.547	1.428	1.353	1.300	0.423	0.338	0.315	0.306	0.388	0.323	0.312	0.302
	0.50	1.992	1.856	1.798	1.767	0.449	0.376	0.349	0.337	0.410	0.346	0.329	0.328
	0.75	2.776	2.630	2.544	2.494	0.479	0.410	0.391	0.379	0.449	0.383	0.363	0.358
	minutes	0.091	0.082	0.082	0.081	0.384	0.326	0.282	0.273	0.413	0.354	0.327	0.296
4	0.25	1.806	1.604	1.480	1.412	0.501	0.405	0.375	0.372	0.447	0.381	0.361	0.358
	0.50	2.238	2.092	1.964	1.888	0.532	0.441	0.399	0.398	0.475	0.405	0.390	0.383
	0.75	3.105	2.898	2.754	2.654	0.569	0.479	0.448	0.429	0.509	0.439	0.414	0.407
	minutes	0.102	0.089	0.092	0.090	0.487	0.424	0.380	0.363	0.593	0.489	0.445	0.416
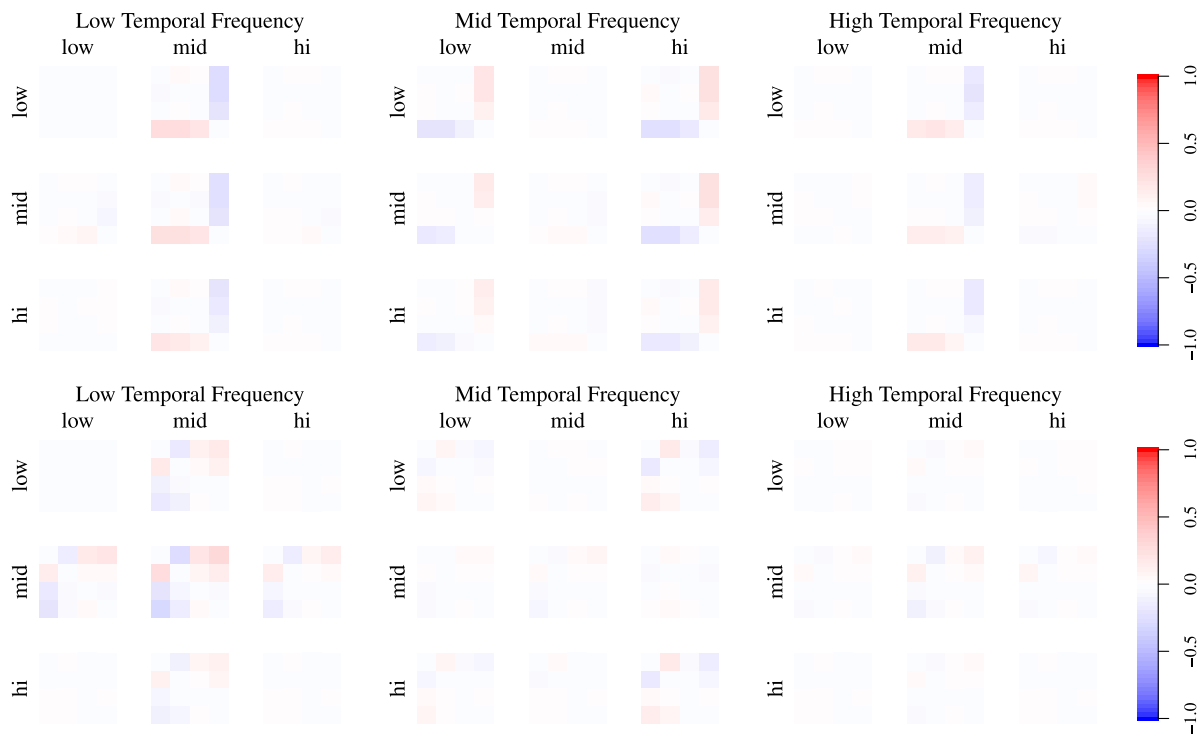
Fig. 7. Estimated imaginary part of coherences for ordinary storm (top row) and Hurricane Florence (bottom row) at various frequencies $\omega = (\omega_1, \omega_2, \omega_3)$. "Low" refers to $\omega_j = 0$, "mid" to $\omega_j = 1/4$, and "high" to $\omega_j = 1/2$. In each 3×3 group of coherence matrices, rows correspond to latitudinal frequencies, columns to longitudinal frequencies.

The paper describes a method for decomposing the estimated spectrum into a linear model of coregionalization plus a residual multivariate process. The decomposition is meant as an exploratory tool for understanding the variation in the data. Finding an optimal such decomposition proved to be an interesting topic, and we have provided some theoretical results that make the numerical search over possible decompositions feasible. The factor decomposition was applied to two storm datasets, where we found that two-component decompositions explained roughly half of the variation in the data.

Only one- and two-factor decompositions were pursued here. Higher-order decompositions are in principle computationally feasible using Theorem 1 and some tedious algebra, but we leave this problem for future work. We have also not explored the predictive capabilities of the fitted models. Prediction is an interesting and important topic in multivariate spatial–temporal models, see for example [28], but we have decided to focus instead on exploring the model fits and how they differ for two types of storms.
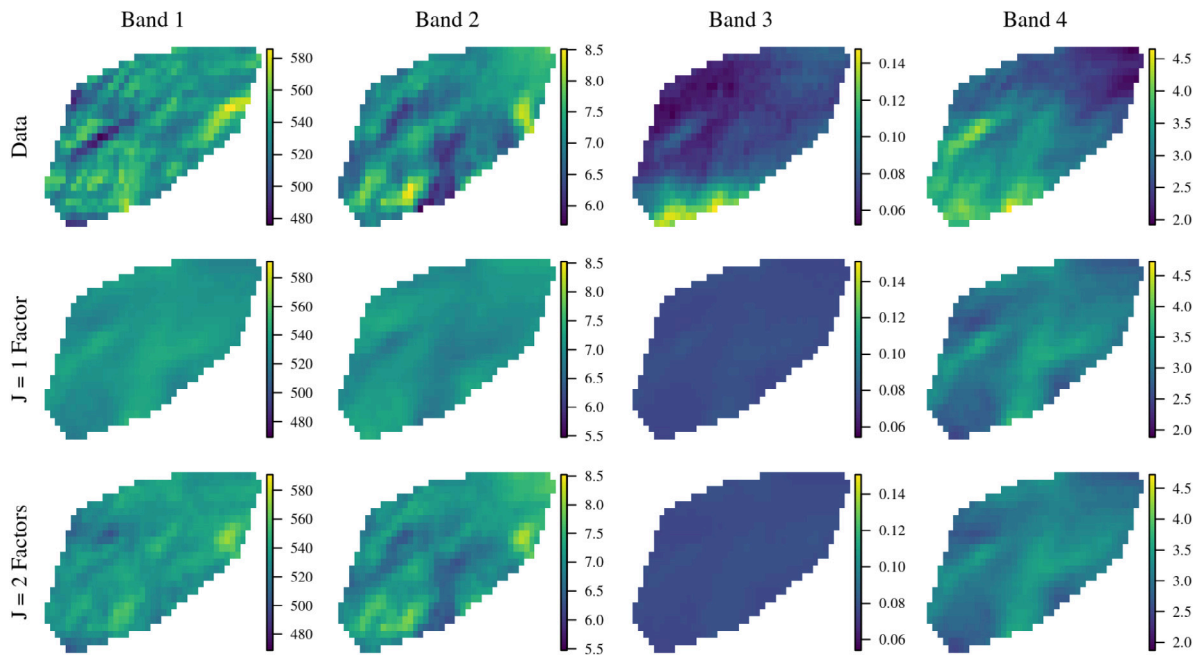
Fig. 8. Hurricane Florence data from time $t = 30$ (top row), one factor representation (middle row), and two factor representation (bottom row). Factor representations use the expected values of W_j given the data.

This paper explores only stationary models. We have argued that fitting stationary multivariate spatial and spatial temporals has remained a difficult task, despite significant advances in the univariate case. A look at the storm data indicates that nonstationary models are needed as well. We leave that topic for future work but note that it may be possible to partition the spatial–temporal domain into blocks, and use the methods proposed in this paper to fit separate stationary models to each of the blocks of data. It would be interesting to explore methods for stitching the stationary models together into a globally nonstationary model.

Finally, we have not addressed the important issue of how to obtain standard errors for the cross spectral densities, which is complicated by the missing values. This will become especially important as we move towards analyzing datasets with a large number of multivariate components; in this case, we often desire a procedure for selecting the non-zero components of the cross spectral density or its inverse.

Acknowledgments

Appendix. Supplementary table and figures

Table 5 contains simulation results for the periodic imputation methods for all four bandwidths. Fig. 7 contains estimated imaginary parts of the coherences for both storm datasets. Fig. 8 shows Hurricane Florence data and estimated one- and two-factor representations of the data.

References

[1] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D.W. Hogg, M. O'Neil, Fast direct methods for Gaussian processes, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 252–265.
[2] T.V. Apanasovich, M.G. Genton, Y. Sun, A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components, J. Amer. Statist. Assoc. 107 (497) (2012) 180–193.
[3] S. Banerjee, B.P. Carlin, A.E. Gelfand, Hierarchical Modeling and Analysis for Spatial Data, CRC Press, 2014.
[4] P. Brockwell, R. Davis, Time Series: Theory and Methods, vol. 2, Springer, 2006.
[5] Computational and Information Systems Laboratory, Cheyenne: SGI ICE XA System (NCAR Community Computing), National Center for Atmospheric Research Boulder, CO, 2017.
[6] H. Cramér, On the theory of stationary random processes, Ann. of Math. (1940) 215–230.
[7] R. Dahlhaus, H. Künsch, Edge effects and efficient parameter estimation for stationary random fields, Biometrika 74 (4) (1987) 877–882.

[8] A. Datta, S. Banerjee, A.O. Finley, A.E. Gelfand, Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets, J. Amer. Statist. Assoc. 111 (514) (2016) 800–812.

[9] D. Eddelbuettel, R. François, Rcpp: Seamless R and C++ Integration, J. Stat. Softw. 40 (8) (2011) 1–18, http://dx.doi.org/10.18637/jss.v040.i08, http://www.jstatsoft.org/v40/i08/.

[10] M.G. Genton, W. Kleiber, Cross-covariance functions for multivariate geostatistics, Statist. Sci. 30 (2) (2015) 147–163.

[11] T. Gneiting, Nonseparable, stationary covariance functions for space–time data, J. Amer. Statist. Assoc. 97 (458) (2002) 590–600.

[12] T. Gneiting, W. Kleiber, M. Schlather, Matérn cross-covariance functions for multivariate random fields, J. Amer. Statist. Assoc. 105 (491) (2010) 1167–1177.

[13] J. Guinness, Permutation and grouping methods for sharpening Gaussian process approximations, Technometrics 60 (4) (2018) 415–429.

[14] J. Guinness, Spectral density estimation for random fields via periodic embeddings, Biometrika 106 (2) (2019) 267–286.

[15] J. Guinness, M. Fuentes, Circulant embedding of approximate covariances for inference from Gaussian data on large lattices, J. Comput. Graph. Statist. 26 (1) (2017) 88–97.

[16] X. Guyon, Parameter estimation for a stationary process on a d-dimensional lattice, Biometrika 69 (1) (1982) 95–105.

[17] M. Katzfuss, J. Guinness, A general framework for vecchia approximations of Gaussian processes, 2017, arXiv preprint arXiv:1708.06302.

[18] W. Kleiber, Coherence for multivariate random fields, Statist. Sinica 27 (2017) 1675–1697.

[19] T. Lee, Z. Zhu, Nonparametric spectral density estimation with missing observations, in: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE Computer Society, 2009, pp. 3041–3044.

[20] C.Y. Lim, M. Stein, Properties of spatial cross-periodograms using fixed-domain asymptotics, J. Multivariate Anal. 99 (9) (2008) 1962–1984.

[21] A. Litvinenko, Y. Sun, M.G. Genton, D.E. Keyes, Likelihood approximation with hierarchical matrices for large spatial datasets, Comput. Statist. Data Anal. 137 (2019) 115–132.

[22] A. Majumdar, A.E. Gelfand, Multivariate spatial modeling for geostatistical data using convolved covariance functions, Math. Geol. 39 (2) (2007) 225–245.

[23] NOAA, GOES ABI Quick Information Guides, 2018, https://www.goes-r.gov/education/ABI-bands-quick-info.html. [Online; Accessed 3 November 2018].

[24] Y. Saad, Iterative Methods for Sparse Linear Systems, vol. 82, SIAM, 2003.

[25] A.K. Saibaba, S. Ambikasaran, J.Y. Li, P.K. Kitanidis, E.F. Darve, Application of hierarchical matrices to linear inverse problems in geostatistics, Oil Gas Sci. Technol.–Revue D'IFP Energies Nouvelles 67 (5) (2012) 857–875.

[26] A.V. Vecchia, Estimation and model identification for continuous spatial processes, J. R. Stat. Soc. Ser. B Stat. Methodol. 50 (2) (1988) 297–312.

[27] A.T. Wood, G. Chan, Simulation of stationary Gaussian processes in $[0, 1]^d$, J. Comput. Graph. Statist. 3 (4) (1994) 409–432.

[28] H. Zhang, W. Cai, When doesn't cokriging outperform kriging? Statist. Sci. 30 (2) (2015) 176–180.