Journal of the Korean Statistical Society (2022) 51:608–631
https://doi.org/10.1007/s42952-021-00156-y

Online ISSN 2005-2863
Print ISSN 1226-3192

RESEARCH ARTICLE

Spatial regression with non-parametric modeling of Fourier coefficients

Yoon Bae Jun[1] · Chae Young Lim[1]

Received: 16 May 2021 / Accepted: 7 November 2021 / Published online: 17 November 2021

Abstract
We consider modeling of Fourier coefficients, known as a spectral density function to represent spatial dependence of a stationary spatial random field and use it for spatial regression under a Bayesian framework. Especially, we switch from the space domain to the frequency domain and introduce a Gaussian process prior to the log spectral density. As we do not impose any further assumption on log spectral density, resulting covariance function is not of a parametric form and/or isotropic assumption. Simulation study supports that our approach is robust over various parametric covariance models. Also, our approach gives comparable or better prediction results over conventional spatial prediction under most parametric covariance models that we considered. Even though we need to estimate spectral density at all Fourier frequencies during the Bayesian procedure, our approach does not lose much computational efficiency compared to estimating only a few parameters in the parametric covariance models. We also compare our approach with some other existing spatial prediction approaches using two datasets of Korean ozone concentration. Our approach performs reasonably good in terms of mean absolute error and root mean squared error.

Keywords Spatial regression · Spectral density · Periodogram · Gaussian process prior

1 Introduction

When considering a spatial process, dependence of the process is typically modeled by its covariance as a function of spatial locations and stationarity is often further assumed, which indicates that the covariance is a function of the difference between two spatial locations. The dependence structure of the spatial process affects estimation in spatial regression and spatial prediction. Statistical methods to analyze

✉ Chae Young Lim
twinwood@snu.ac.kr

[1] Department of Statistics, Seoul National University, Seoul, South Korea

spatial data have enabled us not only to make statistical inference about the spatial distribution, but also to predict values of a variable of interest at unmonitored locations (Cressie 1993; Gelfand et al. 2010; Stein 1999), which requires a covariance estimate in either parametric or non-parametric way. An empirical covariance function at different lags by non-parametric estimates can be used to construct the covariance matrix, but positive definiteness could fail depending on an estimation method (Cressie 1993). Instead, one can consider a parametric structure of a covariance function such as spherical, Matérn, or powered exponential covariance functions and use the least square methods or maximum likelihood estimation for estimating their unknown parameters. However, this approach still has a possibility of model misspecification of the covariance structure, which may cause inaccurate inference for regression coefficients or poor prediction of a variable of interest. Also, it requires computation of a matrix inversion which causes great computational burden for high-dimensional spatial data as the spatial covariance matrix is a dense matrix (Aune et al. 2014; Gelfand et al. 2010). Furthermore, typical isotropy assumption is rather limited as it is frequently violated in many real world applications.

Recently developed methodologies have made great progress in fitting a complicated shape of spatial distribution and handling large spatial data. Low-rank approximation approach approximate the spatial process as a linear combination using a set of a priori designed basis functions that is fixed in number (Cressie and Johannesson 2008; Katzfuss and Cressie 2011; Stein 2008). Lattice Kriging considers multiresolution radial basis functions, which results in faster computation (Nychka et al. 2002, 2015). Predictive process approach (Banerjee et al. 2008; Finley et al. 2009) uses a set of knot locations on which the process is approximated in the form of basis functions representation under a Bayesian framework. Multiresolution approximations (Katzfuss 2017; Katzfuss and Gong 2017) also uses basis function representation with compactly supported basis functions at different resolutions, which can be adapted to any given covariance function. Stochastic partial differential equation approaches (Lindgren et al. 2011) approximate a Gaussian process with a Matérn covariance function with a Markov random field, which bring an efficient calculation of a likelihood.

Different from the approaches which make use of a basis function representation of a spatial process in some way, covariance tapering creates a sparse covariance matrix by multiplying compactly supported covariance function to increase computational efficiency and such approximation is theoretically investigated in Furrer et al. (2006), Kaufman et al. (2008) and Du et al. (2009). Spatial partitioning also creates a sparse covariance matrix by assuming independence between observations across partitioned subregions and various options for partitioning are suggested in Sang et al. (2011), Kim et al. (2005), Heaton et al. (2017) and Konomi et al. (2014). Nearest neighbor Gaussian process uses conditional specification of spatial processes to model a sparse structure of a covariance matrix which enables efficient computation (Datta et al. 2016).

On the other hand, there are algorithmic approaches in handling spatial data. Metakriging (Guhaniyogi and Banerjee 2018) is an approximate Bayesian method that introduces a combined posterior by subset posteriors from partitioned locations. The gapfill method (Gerber et al. 2018) is purely algorithmic and distribution-free.

The approach chooses a subset in a neighborhood and prediction is based on sorting algorithms and quantile regression. Local approximation Gaussian process approach (Gramacy and Apley 2015) focuses on prediction by training a Gaussian process predictor using the values nearby the prediction location based on a criterion related to mean squared prediction error. The algorithm allows adaptive selection of the number of the neighbors for training.

An alternative way to study spatially structured processes is the spectral representation approach. Note that any stationary process can be represented as a superposition of random harmonic oscillations, i.e. it can be represented by some modification of the conventional Fourier integral (Whittle 1954; Yaglom 1987). Spectral analysis is a study of the spectral measure or spectral density function, which is a Fourier coefficient for the sinusoidal component of a covariance (Gelfand et al. 2010). Once we define a spectral density function for a covariance function in a space domain, spatial dependency can be also modeled by spectral density because of one-to-one correspondence between them. In time series analysis, spectral methods are widely studied and the related theories are well-established (Brillinger 2001; Priestley 1981). Especially, several studies for spectral density estimation and its application in time series regression have been already presented (Carter and Kohn 1997; Choudhuri et al. 2004; Dey et al. 2018). Once we regard a temporal structure as a one-dimensional spatial structure, several aspects of spectral methods in time series analysis can be generalized into the process with more than one-dimension. Also, there are wide applications whose data are available on a grid so that a spectral method can be applied naturally.

Royle and Wikle (2005) and Paciorek (2007) consider representation of a spatial process using a spectral process so that the corresponding covariance matrix is decomposed into the orthogonal matrix with Fourier basis functions and the diagonal matrix with the values of the spectral density. This construction helps more efficient computation but it is under parametric modeling of spectral density. Reich and Fuentes (2012) used a Dirichlet process prior for spectral density so that the resulting covariance function is flexible. Guinness and Fuentes (2017) considers discrete spectral approximation of a covariance function so that the approximated covariance matrix has a nested block circulant structure which is computationally efficient and circulant embedding can be done in a smaller size compared to Stroud et al. (2017). However, the approach is under parametric modeling of the spectral density. Guinness (2019) proposes an iterative imputation approach to estimate spectral density non-parametrically from incomplete lattice data. This work has been extended to multivariate and spatial-temporal data (Guinness 2018).

We introduce non-parametric modeling of a spectral density under a Bayesian framework by considering a Gaussian process prior on log spectral density which leads estimation of a spatial covariance matrix more flexible. Prediction is made concurrently during Bayesian inference. Our work is an extension of Carter and Kohn (1997) and Dey et al. (2018) in that we consider a spatial process. In addition, we extend the method to handle incomplete lattice data. Given the Gaussian process prior, we expect that our approach produces robust prediction results regardless of a covariance structure as we do not assume any parametric nor isotropic model on the covariance function. The works by Guinness (2019, 2018) are comparable to our approach as the spectral density is estimated non-parametrically on incomplete lattice data but it has a different flavor as we handle estimation under the Bayesian framework. The work by Reich and Fuentes (2012) is a Bayesian approach and proposes a flexible modeling for spectral density but it can be computationally demanding due to the nature of posterior sampling with a Dirichlet process prior.

In the empirical results section, we compare our approach with a parametric Bayesian approach in the simulation study and found that our approach performs well for smooth processes. We then compare our approach with some other methods in terms of prediction using two real datasets. Our approach performs reasonably good in terms of mean absolute error and root mean squared error.

The rest of the paper is organized as follows: Sect. 2 introduces spectral methods and model description in detail. Section 3 provides simulation results for estimation and prediction for various scenarios, and real data analysis with two Korean ozone exposure studies. Section 4 provides a conclusion and some related discussion. The codes for implementing the proposed method is available at https://github.com/junpeea/NSBSR.

2 Models and methods

2.1 Preliminaries

A spatially distributed variable is typically modeled as a continuously indexed stochastic process, $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset R^d\}$, where D is a study region of interest, $\boldsymbol{s}$ represents a point of coordinate in D. Under a common regression structure, we consider $Y(\boldsymbol{s}) = \mu(\boldsymbol{s};X) + \epsilon(\boldsymbol{s})$, where $\mu(\boldsymbol{s};X)$ is a deterministic mean function including explanatory variables X with its popular choice $\mu(\boldsymbol{s};X) = X(s)\boldsymbol{\beta}$ and $\epsilon(\boldsymbol{s})$ is a zero mean stationary spatial process with a spatial dependence structure. We can further decompose $\epsilon(\boldsymbol{s}) = \sigma_\epsilon e(\boldsymbol{s})$, where $\sigma_\epsilon^2 = Var(\epsilon(\boldsymbol{s}))$ is a marginal variance and $e(\boldsymbol{s})$ is a normalized process characterized by a correlation function $c(\cdot)$ such that $Cov(e(\boldsymbol{s}), e(\boldsymbol{t})) = c(\boldsymbol{s} - \boldsymbol{t})$, which is a common assumption for spatial data. In other words, we consider the following model.

$$Y(\boldsymbol{s}) = X(\boldsymbol{s})\boldsymbol{\beta} + \sigma_\epsilon e(\boldsymbol{s};c),\ \boldsymbol{s} \in D \subset R^d. \tag{1}$$

In addition, we assume that Y is a Gaussian process which is well accepted for tractable modeling.

We defined $e(\cdot)$ in (1) as a zero mean stationary Gaussian process in R^d with a correlation function $c(\cdot)$. With additional assumption of mean squared continuity, the correlation function can be represented in the following Fourier integral form

$$c(\boldsymbol{s}) = \int_{R^d} \exp(\imath \boldsymbol{w}^t \boldsymbol{s}) F(d\boldsymbol{w}), \tag{2}$$

where F is a positive finite measure called a spectral measure. We further assume that F is absolutely continuous so that it has a Radon–Nikodym derivative with respect to Lebesgue measure, $f = \frac{dF}{dw}$, which is called a spectral density. The spectral density can be recovered by inverse Fourier transformation from $c(\cdot)$:

$$f(\boldsymbol{w}) = \frac{1}{(2\pi)^d} \int_{R^d} \exp(-\imath \boldsymbol{w}^t \boldsymbol{s}) c(\boldsymbol{s}) d\boldsymbol{s}. \tag{3}$$

The periodogram is a well-known non-parametric estimate of the spectral density using the data observed on regularly spaced lattice. For two dimensional space domain ($d = 2$), assume that observed data are located at $n_1 \times n_2$ regular grid over a rectangular study region $D \subset R^2$. Let $\Delta = (\delta_1, \delta_2)$ be the spacing between neighboring observations in each direction. Then, periodogram is defined as follows:

$$\mathcal{I}_{n_1 n_2}(w_1, w_2) = \frac{1}{4\pi^2 n_1 n_2} \left| \mathcal{D}_{n_1 n_2}(w_1, w_2) \right|^2, \tag{4}$$

where

$$\mathcal{D}_{n_1 n_2}(w_1, w_2) = \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} e(j\delta_1, k\delta_2) \exp[-\imath(w_1 j\delta_1 + w_2 k\delta_2)] \tag{5}$$

for $\boldsymbol{w} = (w_1, w_2) \in W_\Delta^2 = [-\pi/\delta_1, \pi/\delta_1) \times [-\pi/\delta_2, \pi/\delta_2)$.

$\mathcal{I}_{n_1 n_2}(\boldsymbol{w})$ are exponentially distributed with mean $f_{\delta_1 \delta_2}(\boldsymbol{w}) = \sum_{Q_1 \in \mathcal{Z}} \sum_{Q_2 \in \mathcal{Z}} f\left(w_1 + \frac{2\pi Q_1}{\delta_1}, w_2 + \frac{2\pi Q_2}{\delta_2}\right)$, where $\mathcal{Z}$ is the set of integers and they are asymptotically independent at distinct Fourier frequencies. These properties can be obtained by the same arguments used for a time series and Gaussian assumption (Brillinger 2001) since we consider a spatial process on a lattice, where Δ is fixed and the observation domain is increasing as the sample size is increasing. Similar results for a spatial lattice process when the spacing is decreasing while the observation domain is fixed are introduced in Lim and Stein (2008). Also, $\mathcal{I}_{n_1 n_2}$ is symmetric around the half of the Fourier frequencies, i.e. $\mathcal{I}_{n_1 n_2}(w_1, w_2) = \mathcal{I}_{n_1 n_2}\left(\frac{2\pi}{\delta_1} - w_1, \frac{2\pi}{\delta_2} - w_2\right)$ for $\boldsymbol{w} \in W_\Delta^2$.

2.2 Proposed model

Assuming $n_1 \times n_2$ regular grid over a rectangular study region $D \subset R^2$, let $|D_1|$ be the length of D in x-axis, $|D_2|$ be the length of D in y-axis, and $N = n_1 n_2$ be the sample size. We denote a complete set of regularly spaced locations $S_{com}^\Delta = \{\boldsymbol{s}_{jk} = (s_j, s_k) = (j\delta_1, k\delta_2); j = 0, 1, \ldots, (n_1 - 1), k = 0, 1, \ldots, (n_2 - 1)\}$, where $\delta_1 = \frac{|D_1|}{n_1}, \delta_2 = \frac{|D_2|}{n_2}$. We first consider completely observed samples $(\boldsymbol{Y}, \boldsymbol{X}) = \{(Y_{jk}, \boldsymbol{X}_{jk}) = (Y(\boldsymbol{s}_{jk}), X_1(\boldsymbol{s}_{jk}), \ldots, X_p(\boldsymbol{s}_{jk})); \forall \boldsymbol{s}_{jk} \in S_{com}^\Delta\}$, where p is the number of covariates. Then, the model (1) using the data becomes $Y_{jk} = \sum_{r=1}^p X_{rjk}\beta_r + \sigma_\epsilon e_{jk}$, for $j = 0, 1, \ldots, (n_1 - 1), k = 0, 1, \ldots, (n_2 - 1)$ and its matrix form is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma_\epsilon \boldsymbol{e},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^t$ and $\boldsymbol{e} = (e_1, \ldots, e_N)^t$.

Given $\boldsymbol{e}$, we can obtain the periodogram $\mathcal{I}_{n_1n_2}$ at the Fourier frequencies. Due to the symmetry, we only need the first half of them. Recall that $\mathcal{I}_{n_1n_2}$ are exponentially distributed and asymptotically independent at distinct Fourier frequencies. The exponential density expression for $\mathcal{I}_{n_1n_2}$ can be viewed as a Whittle likelihood by considering it as an approximation of the Gaussian density for $\boldsymbol{e}$ (Whittle 1954). Carter and Kohn (1997) introduced a five-component mixture Gaussian distribution as approximation of the distribution of the logarithm of an exponential distribution so that we use Carter and Kohn (1997)'s approximation for $\log \mathcal{I}_{n_1n_2}$. That is,

$$\log \mathcal{I}_{n_1n_2}(\boldsymbol{w}) = \log f_{\delta_1\delta_2}(\boldsymbol{w}) + \xi(\boldsymbol{w}) \tag{6}$$

with ξ having distribution $\pi(\xi)$ such that

$$\pi(\xi) = \sum_{l=1}^{5} p_l \phi_{v_l}(\xi - \kappa_l), \tag{7}$$

where $\phi_v(\cdot - \kappa)$ is a normal density function with mean κ and variance v^2. The weights (p_l), means (κ_l) and standard deviations (v_l) of the five components in the mixture Gaussian distribution to match the density of the logarithm of an exponential distribution are provided in Carter and Kohn (1997) and we also provide them in the Appendix.

Let ψ be a latent variable that indicates a component in (7), $\boldsymbol{\varphi}$ be a vector of $\log \mathcal{I}_{n_1n_2}(\boldsymbol{w})$ and $\boldsymbol{\theta}$ be a vector of $\log f_{\delta_1\delta_2}(\boldsymbol{w})$. We pursue a hierarchical model and Bayesian inference by considering a Gaussian process prior (GP) for $\log f_{\delta_1\delta_2}(\boldsymbol{w})$ with mean function $\nu(\cdot)$ and covariance function $\tau(\cdot,\cdot)$, and appropriate priors for hyper-parameters. The model and prior specifications are summarized as follows:

1. Data model:
   (a) $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma_\epsilon \boldsymbol{e}$ (space domain)
   (b) $\boldsymbol{\varphi} = \boldsymbol{\theta} + \boldsymbol{\xi}$ (frequency domain)
2. Process model:
   $\boldsymbol{\theta} \sim GP(\nu(\cdot), \tau(\cdot,\cdot))$ with $\nu(\boldsymbol{w}) \equiv 0$ and
   $\tau(\boldsymbol{w}_1, \boldsymbol{w}_2) = \tau_\theta^{-1} \exp(-\rho_{\theta_1}|w_{11} - w_{21}| - \rho_{\theta_2}|w_{12} - w_{22}|)$, $\boldsymbol{w}_i = (w_{i1}, w_{i2})$ for $i = 1, 2$.
3. Parameter models:
   $\boldsymbol{\beta} \sim N(\mu_\beta \mathbf{1}, \sigma_\beta^2 \boldsymbol{I})$,
   $P(\psi = l) = p_l$, for $l = 1, \ldots, 5$; $\rho_{\theta_1}, \rho_{\theta_2} \sim Unif(0, \rho_0)$ for some $\rho_0 > 0$, $\tau_\epsilon = 1/\sigma_\epsilon^2 \sim G(a, b)$, $\tau_\theta \sim G(c, d)$,
   where $G(a, b)$ is the Gamma distribution with mean ab.

We consider a Gibbs sampler from the above hierarchical structure by obtaining conditional posterior distribution of each parameter given the data and other parameters. The detail construction of conditional distributions is given in the Appendix.

Once we obtain R Gibbs samples, we predict Y over a given study region D at unmonitored locations. In Bayesian framework, prediction of Y is based on conditional expectation $E(Y|Y_{obs})$ given the observed data,

Y_{obs}. Given Gibbs samples, prediction of $Y(\boldsymbol{s}_0)$ at an unmonitored location $\boldsymbol{s}_0 \in D$ is given as $\hat{Y}(\boldsymbol{s}_0) = \frac{1}{R}\sum_{r=1}^{R} E(Y(\boldsymbol{s}_0)|Y_{obs};\hat{\boldsymbol{\beta}}^{(r)}, \hat{\sigma_\epsilon}^{(r)}, \hat{\boldsymbol{\theta}}^{(r)})$ with $E(Y(\boldsymbol{s}_0)|Y_{obs};\hat{\boldsymbol{\beta}}, \hat{\sigma_\epsilon}, \hat{\boldsymbol{\theta}}) = X(\boldsymbol{s}_0)\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{h}}^t\tilde{\boldsymbol{\Gamma}}^{-1}(Y_{obs} - X\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{h}} = \widehat{Cov}(\boldsymbol{e}, e(\boldsymbol{s}_0))$ and $\tilde{\boldsymbol{\Gamma}} = \widehat{Cov}(\boldsymbol{e})$ (Cressie 1993). The prediction error variance of Y is similarly obtained.

To obtain Gibbs samples and prediction results, we need to compute a matrix-vector multiplication involving $\tilde{\boldsymbol{\Gamma}}^{-1}$. If the sites at S^{Δ}_{com} are ordered from top to bottom and from left to right, the Eq. (5) leads to the covariance matrix $\tilde{\Gamma}$ being $n_2 \times n_2$ block circulant matrix, where each block is also circulant of the size $n_1 \times n_1$. We adopt the approach by Anitescu et al. (2012) which makes use of this block-circulant of circulant blocks (BCCB) structure for efficient computation of $\tilde{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{h}}$. The detail explanation is given in the Appendix.

The computation of $\hat{\boldsymbol{h}}$ requires an additional technique. The covariance estimates retrieved by the estimated spectral density are only at lags in the form of $(j\delta_1, k\delta_2)$ and this is not enough to reconstruct $\hat{\boldsymbol{h}}$ since it requires covariance estimates at a finer resolution. However, an interpolation of the estimated spectral density, analogous to the approach by Dey et al. (2018) is no longer applicable due to an aliasing effect of the spectrum of sample observations (Gelfand et al. 2010). To resolve this issue, we consider an interpolation of covariance estimates at a coarser resolution to get the covariance estimates at a finer resolution so that we can construct $\hat{\boldsymbol{h}}$. For example, $\hat{c}(0.5\delta_1, 0.5\delta_2)$ is obtained by bilinear interpolation with four neighboring values, $\hat{c}(0, 0)$, $\hat{c}(\delta_1, 0)$, $\hat{c}(0, \delta_2)$, $\hat{c}(\delta_1, \delta_2)$.

Our approach requires to sample θ, logarithm of the spectral density at the Fourier frequencies per Gibbs iteration, which can be time consuming. However, we argue that it does not lose much computational efficiency compared to a conventional Bayesian spatial regression method under a parametric set-up. This is partly due to the fact that we used discrete Fourier transform (DFT) by taking advantage of fast Fourier transform (FFT) algorithm (Bracewell 1986; Cooley and Tukey 1965), whose computation cost is $\mathcal{O}(n_1 n_2 \log(n_1 n_2))$. Also, due to symmetry of the spectral density and the periodogram about the origin, we only need to consider the first half of Fourier frequencies. If we permit to impose more restriction on the true spectral density such as isotropy, we can further improve computation speed and save memories by considering about one fourth of the frequencies.

2.3 Proposed model for observations on an incomplete grid

Let ζ_{jk} be a variable for indicating if Y is observed at $\boldsymbol{s}_{jk} \in S^{\Delta}_{com}$. That is, $\zeta_{jk} = 1$ if $Y(\boldsymbol{s}_{jk})$ is observed and $\zeta_{jk} = 0$ if it is missing. Now we consider a complete set which includes observations as well as missing values with indicators:

$$(\boldsymbol{Y}, \boldsymbol{\zeta}, \boldsymbol{X}) = \left\{(Y_{jk}, \zeta_{jk}, \boldsymbol{X}_{jk}) = (Y(\boldsymbol{s}_{jk}), \zeta(\boldsymbol{s}_{jk}), X_1(\boldsymbol{s}_{jk}), \ldots, X_p(\boldsymbol{s}_{jk}))\right\}.$$

Recall that the matrix form of the model using the data is

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \sigma_\epsilon \boldsymbol{e}.$$

Suppose that both $\boldsymbol{X}$ and $\boldsymbol{Y}_{obs}$ are observed but $\boldsymbol{Y}_{mis}$ is missing at random given observations, where $\boldsymbol{Y}_{obs}$ is an observed part and $\boldsymbol{Y}_{mis}$ is a missing part of $\boldsymbol{Y}$. Let $\boldsymbol{\Theta} = (\boldsymbol{\beta}^t, \sigma_\epsilon, \boldsymbol{\theta}^t, \psi, \tau_\theta, \rho_{\theta_1}, \rho_{\theta_2})^t$ be a vector of the entire model parameters. In Bayesian inference, it is common to treat $\boldsymbol{Y}_{mis}$ as a vector of latent variables. Then, we can augment missing observations by sampling from the conditional probabilities of missing observations, $\boldsymbol{h}_{mis} = P(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\zeta}, \boldsymbol{\Theta})$ in the MCMC procedure described in Sect. 2.2.

With Gaussian assumption of $\boldsymbol{Y}$, we can easily show that $\boldsymbol{h}_{mis}$ follows a multivariate normal distribution. Note that

$$\begin{pmatrix} \boldsymbol{Y}_{obs} \\ \boldsymbol{Y}_{mis} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{X}_{obs}\boldsymbol{\beta} \\ \boldsymbol{X}_{mis}\boldsymbol{\beta} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

so that the conditional distribution of $\boldsymbol{Y}_{mis}$ given $\boldsymbol{Y}_{obs}$ and $\boldsymbol{\Theta}$ is

$$\boldsymbol{Y}_{mis}|(\boldsymbol{Y}_{obs}, \boldsymbol{\Theta}) \sim N\big(\boldsymbol{\mu}_{mis|obs}, \boldsymbol{\Sigma}_{mis|obs}\big),$$

where $\boldsymbol{\mu}_{mis|obs} = \boldsymbol{X}_{mis}\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\big(\boldsymbol{Y}_{obs} - \boldsymbol{X}_{obs}\boldsymbol{\beta}\big)$ and $\boldsymbol{\Sigma}_{mis|obs} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$ are to be recovered from our model parameters. Then, Bayes formula gives

$$\boldsymbol{h}_{mis} = P(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\zeta}, \boldsymbol{\Theta}) = \frac{\pi(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\Theta})P(\boldsymbol{\zeta}|\boldsymbol{Y}, \boldsymbol{\Theta})}{\int \pi(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\Theta})P(\boldsymbol{\zeta}|\boldsymbol{Y}, \boldsymbol{\Theta})d\boldsymbol{Y}_{mis}}.$$

When we assume the missingness of $\boldsymbol{Y}$ occurs at random conditioning on both observed data and model parameters, i.e. $P(\boldsymbol{\zeta}|\boldsymbol{Y}, \boldsymbol{\Theta}) = P(\boldsymbol{\zeta}|\boldsymbol{Y}_{\boldsymbol{obs}}, \boldsymbol{\Theta})$, then the component is nothing but a constant with respect to the unknown quantities $\boldsymbol{Y}_{mis}$ given $\boldsymbol{Y}_{obs}$, and $\boldsymbol{\Theta}$ (Kim and Shao 2013). Therefore, we can get the samples from $\boldsymbol{h}_{mis}$ by sampling from $\pi(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\Theta})$ within the proposed Gibbs sampler.

3 Empirical results

3.1 Simulation study

In this section, we show performance of the proposed approach in terms of estimation and prediction. Then, we compare with a parametric Bayesian approach under various simulation settings. We consider a regular grid over a rectangular study region D, denoted by $S_{\delta,n}$, in which the distance between neighboring observations in each direction is δ and the length of each direction is n. In other words, $S_{\delta,n} = \{\boldsymbol{s}_{jk} = (s_j, s_k) = (j\delta, k\delta), j, k = 0, \dots, \lfloor (n-1)/\delta \rfloor\}$, where $\lfloor x \rfloor$ is the greatest integer less than equal to x. We consider two covariates X_1 and X_2 in addition to a constant term. X_1 is generated from a mixture of two normal distributions, i.e. $X_1 = p\xi_1 + (1-p)\sqrt{5}\xi_2; p \sim Ber(0.5), \xi_1, \xi_2 \sim N(0, 1)$, and X_2 is generated from a standard exponential distribution. The regression coefficients, $\boldsymbol{\beta}$, is set to $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^t = (0.01, 0.02, 0.03)^t$.

For Bayesian inference, we choose the values of hyper-parameters such that $\mu_\beta = 0$, $\sigma^2_\beta = 100$, $a = 100$, $b = 10$, $c = 100$, $d = 100$ and $\rho_0 = 0.001$. The small value of ρ_0 implies weak dependence of the covariance kernel, τ_{ω_1,ω_2}, for the prior of $\boldsymbol{\theta}$. The choices of a, b, c and d are to make variability of the prior distributions large. We also did some sensitivity analysis (not shown) for the choice of a, b, c, d and found that the prediction results are not much different. Three chains with 10,000 iterations each with 9000 burn-in are obtained. We call our proposed approach, the non-parametric spectral density Bayesian spatial regression as NSBSR and usual parametric Bayesian spatial regression as PBSR in short.

First, we consider simulated datasets with two different grid sizes ($S_{1,16}$ and $S_{1,32}$) and assuming an exponential covariance model, $\sigma^2 e^{-\|Ah\|/\phi}$ with $\sigma^2 = 1$ and $\phi = 10$ to investigate the estimation result of spectral density by comparing with the true spectral density. Two choices of A are considered: $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (isotropy) and $A = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$ (anisotropy). The anisotropic choice of A implies that the x-direction is stretched twice compared to the y-direction. Figure 1 shows the estimated log-scale spectral densities in three-dimensional visualization. The first row is the true spectral density and the second row is the NSBSR-estimated log-scale spectral density. Compared to the true spectral densities, estimated spectral densities tend to over-estimate at boundaries but they try to capture anisotropic patterns. Note that

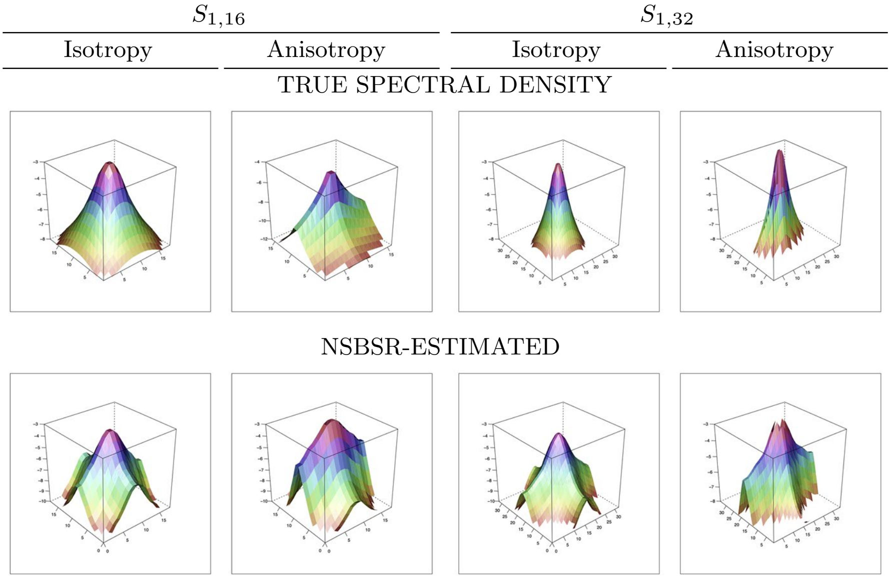

Fig. 1 Estimated log-scale spectral densities assuming an exponential covariance model with $\sigma^2 e^{-\|Ah\|/\phi}$, $\sigma^2 = 1$, $\phi = 10$, $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (isotropy), and $A = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$ (anisotropy). First row corresponds to the true log-scale spectral densities while the second row corresponds to the estimated log-scale spectral densities using the proposed method

this is one example dataset so that the result could vary by a different simulated dataset.

Next, we consider simulated datasets on $S_{1,32}$ and assuming a Matérn covariance model, $c(\boldsymbol{h};\sigma^2,\phi,\alpha)=\sigma^2\frac{2^{1-\alpha}}{\Gamma(\alpha)}\left(\frac{\|A\boldsymbol{h}\|}{\phi}\right)^{\alpha}\mathcal{K}_\alpha\left(\frac{\|A\boldsymbol{h}\|}{\phi}\right)$ at various smoothing levels α to investigate prediction performance. We set $\sigma^2=1$ and $\phi=10$ as before. For this simulation, we consider $A=\begin{pmatrix}1&0\\0&1\end{pmatrix}$(isotropy) and $A=\begin{pmatrix}1/\sqrt{2}&0\\0&1/(2\sqrt{2})\end{pmatrix}$(anisotropy). The anisotropy choice of A implies that x-direction is stretched twice compared to the y-direction while the norm is scaled to $1/\sqrt{2}$. We then fit the model using the data only on $S_{2,32}$, which is a subsample of the data on $S_{1,32}$ with neighboring distance in each direction being twice large. The prediction is made on $S_{1,32}$ and compared with the generated data on $S_{1,32}$. Figure 2 shows prediction results from our NSBSR approach with observed values (simulated values). We can see that our approach tries to capture observed patterns for both isotropic and anisotropic cases. Again, note that this is one example dataset so that the result could vary by a different simulated dataset.

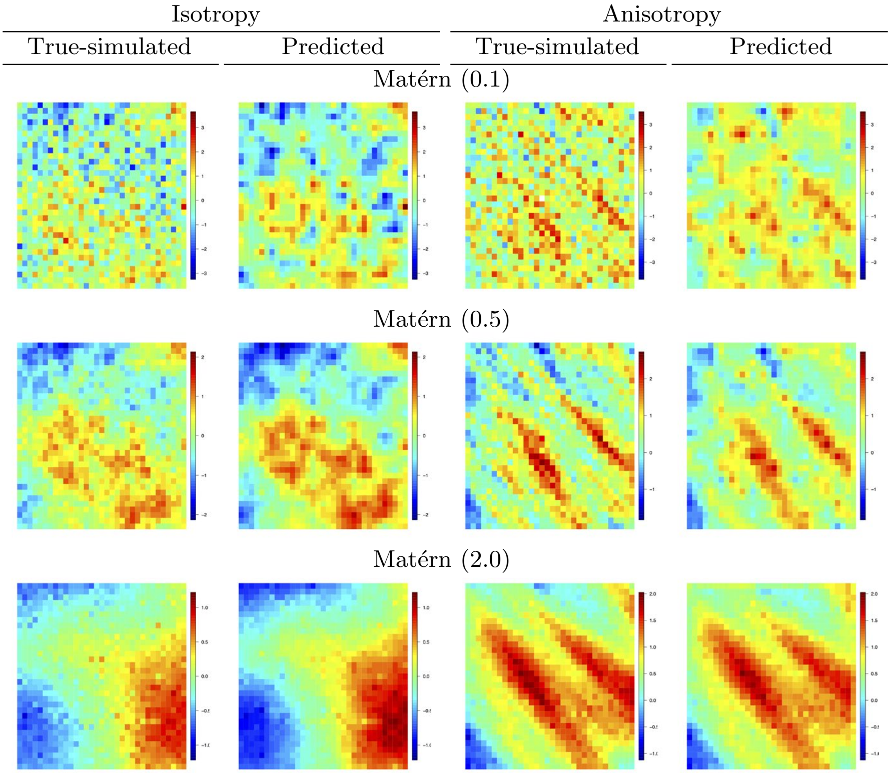

Fig. 2 True simulated Y (left) and NSBSR predicted Y (right) in isotropic and anisotropic Matérn covariance models. Each row corresponds to a different smoothness parameter, $\alpha=0.1, 0.5, 2.0$, respectively

Now, we investigate several covariance models: (1) pure nugget (i.i.d.), (2) Bumpy Matérn ($\alpha = 0.1$), (3) Matérn ($\alpha = 0.5$), (4) Smooth Matérn ($\alpha = 2.0$), (5) Bumpy Powered Exponential ($\alpha = 0.5$), (6) Smooth Powered Exponential ($\alpha = 1.5$) and (7) Gaussian. Note that the form of Matérn covariance function is introduced earlier and Matérn covariance model with $\alpha = 0.5$ corresponds to the exponential covariance model. The powered exponential covariance function is $c(\boldsymbol{h};\sigma^2, \alpha) = \sigma^2 \exp(-\|Ah\|^{\alpha})$. We set $\sigma^2 = 1$ and $\phi = 10$ as well. In addition to isotropic cases, we investigate anisotropic models with $A = \begin{pmatrix} 1 & 0 \\ 0 & 1/4 \end{pmatrix}$, which implies that the x-direction is stretched four times compared to the y-direction. We simulate 100 datasets for each covariance setting with isotropy and anisotropy. Similar to the second simulation case that produces Fig. 2, we first generate the data on $S_{1,32}$ and use the data on $S_{2,32}$ to fit the model. The prediction is made on $S_{1,32}$ and compared with the generated data on $S_{1,32}$.

Figure 3 shows prediction performance results of NSBSR (red) and PBSR with Matérn model with different degrees of fixed smoothness parameter $\alpha = 0.1$ (green), $\alpha = 0.5$ (blue), $\alpha = 2.0$ (purple), and unfixed smoothness parameter α (yellow) for each simulation setting. In addition, we also compared with universal kriging (gray) with maximum likelihood estimates, where the smoothness parameter α is also estimated. We call this approach UK. A fixed smoothness parameter means that we estimate other parameters except the smoothness parameter. An unfixed smoothness parameter means we estimate it as well. These are boxplots of root mean squared prediction errors (RMSPE) between observations and predicted values over 100 datasets. RMSPE is averaged over locations for each data set. From the left block (divided by the dotted vertical lines), the covariance models for data generation are, in turn, pure nugget (i.i.d.), Bumpy Matérn ($\alpha = 0.1$), exponential, Bumpy powered exponential ($\alpha = 0.5$), Smooth powered exponential ($\alpha = 1.5$), smooth Matérn and Gaussian. RMSPEs for NSBSR are overall comparable to those for PBSR and UK in the case of lower degree of smoothness (first four blocks) and lower in the case of higher degree of smoothness (last three blocks). Sample visualization results for NSBSR in Fig. 2 also imply that the predicted values for a smoother covariance model is relatively less biased than those for a bumpy covariance model. RMSPEs for NSBSR are quite robust compared to those for the PBSR and UK with various covariance models. Note that unfixed Matérn model results of PBSR are not impressive, although it is more flexible than the fixed Matérn model. The results by UK, non-Bayesian approach show less variability overall. Results for anisotropic cases are also similar, although the difference in prediction performance at smooth covariance models is reduced.

Table 1 shows additional prediction performance measure and estimation performance for each regression coefficient. The row with R^2 shows the average of coefficient of determination between observations and predicted values over 100 datasets. The rows with $\beta_0, \beta_1, \beta_2$ show root mean squared error (RMSE) of regression coefficients. These results are for the simulated data (isotropy case) used in Fig. 3. R^2s are not large enough for all approaches when the data are independent or less smooth processes while it is getting larger when the processes are getting

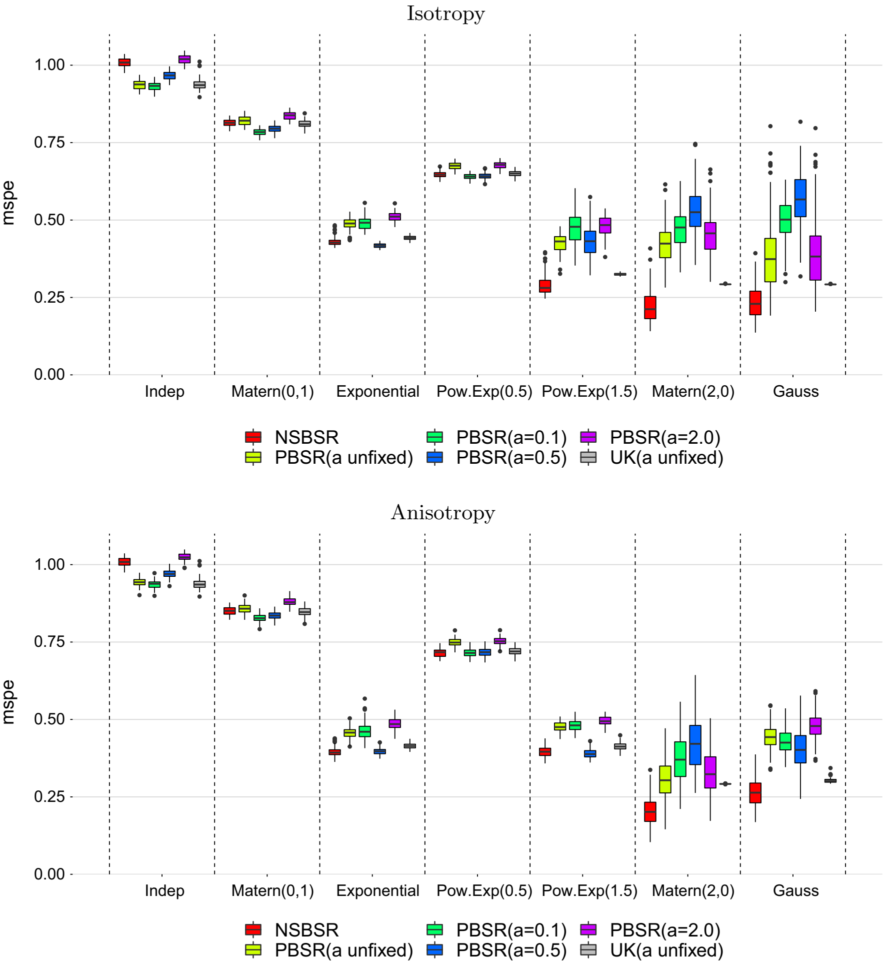

Fig. 3 Box plots of RMSPE: isotropic case is the first row and anisotropic case is the second row. Boxplots over 100 datasets using averages of root mean squared prediction errors over locations. Each block in the figure represents a different covariance model for data generation: pure nugget, Bumpy Matérn ($\alpha = 0.1$), exponential, Bumpy powered exponential ($\alpha = 0.5$), Smooth Powered exponential ($\alpha = 1.5$), smooth Matérn ($\alpha = 2.0$), and Gaussian. In each block, boxplots are ordered by estimation models: NSBSR (red), Bumpy PBSR (α fixed to 0.1; green), exponential PBSR (α fixed to 0.5; blue), smooth PBSR (α fixed to 2.0; purple), general PBSR with α unfixed (yellow), and UK with α unfixed (gray)

smoother. For both prediction and estimation, the proposed NSBSR approach and other parametric approaches are overall comparable in these measures.

We also compare NSBSR with PBSR when the data are on an incomplete grid. We generate three exemplary datasets from Matérn covariance models with $\alpha = 0.1, 0.5, 2.0$, respectively on the complete grid $S_{1,32}$. Then, we consider the data only on $S_{2,16}$ for fitting but we randomly remove grid points according to the missing ratio (MR, %). Then, our approach, NSBSR, is compared to PBSR

Table 1 Prediction and estimation results

True model	Fitted model				
	NSBSR	PBSR (α unfixed)	PBSR ($\alpha = 0.1$)	PBSR ($\alpha = 0.5$)	PBSR ($\alpha = 2.0$)
Independent					
R^2	0.125	0.241	0.255	0.180	0.100
β_0	0.090	0.091	0.089	0.090	0.142
β_1	0.016	0.016	0.016	0.016	0.020
β_2	0.059	0.058	0.059	0.059	0.071
Matérn(0.1)					
R^2	0.430	0.432	0.489	0.465	0.398
β_0	0.535	0.532	0.508	0.532	0.527
β_1	0.015	0.012	0.012	0.013	0.014
β_2	0.047	0.043	0.042	0.043	0.048
Exponential					
R^2	0.918	0.922	0.899	0.925	0.920
β_0	0.763	0.735	0.700	0.670	0.749
β_1	0.007	0.004	0.005	0.004	0.005
β_2	0.032	0.015	0.017	0.015	0.015
Pow.Exp(0.5)					
R^2	0.666	0.650	0.677	0.677	0.648
β_0	0.707	0.702	0.674	0.699	0.697
β_1	0.014	0.009	0.009	0.010	0.009
β_2	0.038	0.031	0.030	0.031	0.032
Pow.Exp(1.5)					
R^2	0.980	0.986	0.972	0.988	0.986
β_0	0.808	0.747	0.713	0.649	0.788
β_1	0.007	0.002	0.003	0.002	0.002
β_2	0.013	0.006	0.009	0.006	0.006
Matérn(2.0)					
R^2	0.992	0.998	0.994	0.996	0.998
β_0	0.866	0.701	0.783	0.718	0.711
β_1	0.006	0.000	0.001	0.000	0.000
β_2	0.011	0.000	0.004	0.001	0.000
Gaussian					
R^2	0.992	1.000	0.996	0.996	1.000
β_0	0.815	0.891	0.724	0.663	0.879
β_1	0.006	0.000	0.001	0.000	0.000
β_2	0.002	0.000	0.004	0.001	0.000

True model indicates a covariance model for data generation. NSBSR is the proposed approach for fitting. PBSR is a parametric Bayesian approach with a Matérn covariance function at various fixed smoothness parameters. The column with "(α unfixed)" is the result when a smoothness parameter is estimated as well. R^2 is the average of coefficient of determination between observations and predicted values over 100 datasets. The rows corresponding to $\beta_0, \beta_1, \beta_2$ show RMSE for regression coefficients with Matérn models with $\alpha = 0.1$ (P01), $\alpha = 0.5$ (P05), $\alpha = 2.0$ (P20) and unfixed α (P00). Table 2 shows mean squared prediction error (MSPE) and R^2 between observations and predicted values. Note that MSPE in this simulation study is the average over locations. The results in Table 2 show that MSPEs of NSBSR tend to get increased as MR increases but the increase is comparable to those of PBSR. Likewise, the R^2 of NSBSR get decreased as MR increases but the decrease is comparable to those of PBSR. For example, with the data from the Matérn covariance model with $\alpha = 0.1$, MSPE of NSBSR approach is 0.814 while PBSR with Matérn covariance model fixed at $\alpha = 0.1$ (P01) is 0.820 when MR is 10%. The missing ratio does not affect much on the estimation of regression coefficients for both NSBSR and PBSR as well (results are not shown for brevity). However, we need longer MCMC chains when the missing ratio gets higher for convergence. Although the impact of missing rates is not apparent for this particular simulation study, convergence can be an issue in conditional simulation for imputing the missing data as discussed in Guinness and Fuentes (2017).

For implementation, we used software R (www.r-project.org). When we fit the model using one dataset on $S_{2,16}$, NSBSR took only additional 0.25 min per 1000 iteration with three chains for one data set compared to PBSR with computer specification of CPU Intel(R) Core(TM) i5-4690 with RAM 8.00 GB.

Table 2 Mean squared prediction error (MSPE) and R squares (R^2) between observations and predicted values under various missing ratios (MR)

Matérn	MR(%)	MSPE					R^2				
		NSBSR	P00	P01	P05	P20	NSBSR	P00	P01	P05	P20
$\alpha = 0.1$	0	0.691	0.691	0.693	0.690	0.694	0.56	0.56	0.56	0.56	0.56
	10	0.814	0.815	0.820	0.813	0.816	0.54	0.54	0.54	0.54	0.53
	25	0.810	0.808	0.823	0.805	0.808	0.50	0.50	0.51	0.51	0.50
	50	0.778	0.775	0.775	0.777	0.773	0.44	0.45	0.45	0.45	0.45
$\alpha = 0.5$	0	0.187	0.187	0.187	0.187	0.186	0.96	0.96	0.96	0.96	0.96
	10	0.226	0.227	0228	0228	0.227	0.86	0.86	0.86	0.86	0.86
	25	0.230	0.229	0.233	0.232	0.229	0.84	0.84	0.84	0.84	0.84
	50	0.279	0.276	0.280	0.280	0.276	0.74	0.74	0.74	0.74	0.74
$\alpha = 2.0$	0	0.042	0.116	0.127	0.125	0.119	1.00	1.00	1.00	1.00	1.00
	10	0.133	0.172	0.166	0.210	0.170	1.00	1.00	1.00	1.00	1.00
	25	0.196	0.196	0.203	0.247	0.196	0.99	0.99	0.99	0.99	0.99
	50	0.252	0.209	0.220	0.265	0.209	0.80	0.82	0.81	0.82	0.82

The first column shows a smoothness parameter setting of Matérn covariance for data generation. NSBSR is the proposed approach. P00, P01, P05 and P20 are results from PBSR by considering Matérn covariance for model fitting. A smoothness parameter for PBSR is unfixed (P00), fixed at 0.1 (P01), fixed at 0.5 (P05), and fixed at 2.0 (P20), respectively

3.2 Real data analysis

In this section, we apply our approach to two ozone datasets. One is from Moderate Resolution Imaging Spectroradiometer (MODIS) Terra Level-3 Aerosol Cloud Water Vapor Ozone Daily Global product (MOD08D3) (https://ladsweb.modaps.eosdis.nasa.gov/search/). The other is from AURA (EOS CH-1) which is a multinational NASA scientific research satellite studying the Earth's ozone layer, air quality, and climate (https://disc.gsfc.nasa.gov/datasets?keywords=aura&page=1).

3.2.1 MODIS application

MOD08D3 contains daily-averaged values of atmospheric parameters related to aerosol particle properties, cloud optical and physical properties, atmospheric water vapor, atmospheric profile and stability indices, and total ozone burden on a 1° × 1° grid. Among them, we obtained quality controlled ozone exposure measurements and the missing values were left untreated. To properly impose a Gaussian assumption, log-transformed ozone exposure is used as a response variable Y. We focus on a neighborhood of the Korea peninsula, i.e. longitude ranged from 112 to 141°, latitude from 24 to 53°. The daily average on February 5, 2019 was used for analysis as an example to deal with a forecast of a short-term ambient exposure. For covariates, we used the world geodetic system (WGS 84) information so that X_1 refers to longitude, and X_2 refers to latitude.

MOD08D3 is a rectangular image pixels with 1° resolution as we mentioned above. We take a subset of the pixels with 2° resolution for model fitting. We then predict the values with 1° resolution. The missing rate is 13.0% for the original dataset. Both sample size and missing rate are moderate. As our approach assumes stationarity, we checked the data with a stationarity test introduced by Taylor et al. (2014), which is designed for testing stationarity of random fields on a regular lattice. The test is available as a `R` function (`TOS2D`) in `LS2Wstat` package. As it requires a complete set of data on a grid, we imputed the data by ordinary kriging with an exponential correlation function before applying the test. The p-value is 0.851, which indicates that we can not reject stationarity assumption of the data. Predicted values of ozone concentration can be used for exposure assessment to acquire valuable scientific meanings such as health effect estimation of ambient air pollution on mortality/morbidity in general epidemiological studies (Kim and Song 2017; Laden et al. 2006).

First, we compare prediction result with a parametric approach by prediction map and computing MSPE and R^2 between observations and predicted values. Figure 4 shows prediction results for the ozone data. The first plot shows the original dataset, which we can see some missing values. The second plot shows the data we used to fit the model. The third and fourth plots are prediction maps from our approach (NSBSR) and parametric approach (PBSR). For PBSR, we consider a Matérn covariance model with unfixed smoothness parameter for model fitting. We can see that the prediction map from our approach shows similarity to the original data. Compared with the result by PBSR, R^2 of NSBSR ($R^2 = 0.54$) is higher than

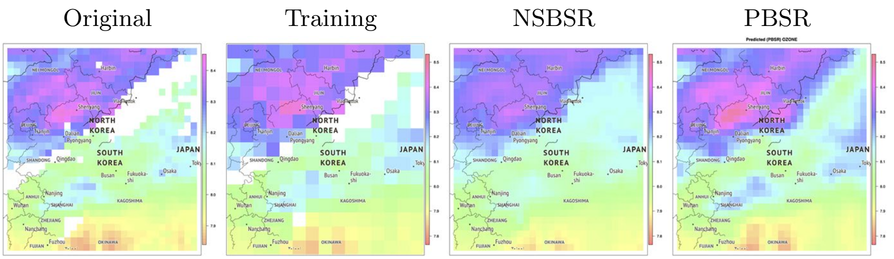

Fig. 4 MODIS result: the first plot is the MODIS dataset (1° resolution) (original). The second plot is a training set (2° resolution) (training). The third plot is prediction result from NSBSR and the fourth plot is prediction result from PBSR with a Matérn covariance function. The smoothness parameter in PBSR was estimated as well. (Longitude: 112° $\sim$ 141°; Latitude: 24° $\sim$ 53°; 30×30 pixels; Time: February 5, 2019)

R^2 for PBSR ($R^2 = 0.46$). MSPE of NSBSR (0.0793) is lower than MSPE of PBSR (0.0903).

Second, we compare prediction results with the methods reviewed in Heaton et al. (2019). Table 3 is the list of methods that we compare with and their abbreviations. A brief summary of each method is given in the Introduction section. For implementation, we basically used the codes available from Heaton et al. (2019) and some R packages when they are available. We used default settings if there are any tuning parameters. Thus, we would like to point out that the results in this section may not be the best for each method. Given the use of available codes, we omit two approaches: Metakriging and Multiresolution approximation in comparison. The computation time was rather long for Metakriging and we were not successful to implement the code for Multiresolution approximation.

Table 4 shows performance results for the MODIS data. We provide mean absolute error (MAE), root mean squared error (RMSE), the average length of 95% confidence intervals (for Non-Bayesian approaches)/credible interval (for Bayesian

Table 3 List of spatial prediction methods in Heaton et al. (2019) we compared with for prediction performance

Abbreviation	Method
NNGP	Nearest neighbor Gaussian process
PP	Predictive process
Tapering	Covariance tapering
Gapfill	Gapfill
Partition	Spatial partitioning
FRK	Fixed rank kriging
SPDE	Stochastic partial differential equations
Periodic	Periodic embedding
LK	Lattice kriging
LAGP	Local approximate Gaussian processes
NSBSR	The proposed Bayesian spectral method

Table 4 Prediction results for the MODIS data from various methods based on mean absolute error (MAE), root mean squared error (RMSE), confidence/credible interval length (INT), and prediction coverage (PC)

Name	MAE	RMSE	INT	PC
NNGP	0.2105	0.2392	0.8515	0.9656
NSBSR	0.2105	0.2395	1.3775	0.9512
LAGP	0.2135	0.2430	0.1614	0.9522
SPDE	0.2139	0.2439	0.0705	1.0000
Partition	0.2144	0.2436	0.2284	0.9389
LK	0.2146	0.2425	0.1257	0.9322
Gapfill	0.2147	0.2447	0.1347	0.7700
FRK	0.2150	0.2409	0.4133	0.8256
Periodic	0.2168	0.2457	0.4321	0.7578
PP	0.2505	0.3033	1.2204	0.7144
Tapering	0.2519	0.2926	2.8318	1.0000

The results are sorted in ascending order by MAE

approaches) (INT) and prediction coverage (PC) as a ratio of cases that 95% prediction intervals contain the observed values over the lattice grids of the study regions. The results are sorted in ascending order by MAE. Although most methods are comparable as the values of MAE and RMSE are similar, NSBSR provides the best result in terms of MAE (tied with NNGP) and the second best result in terms of RMSE. NSBSR has relatively large INT, though. It is interesting that values of INT are widely different among the methods compared to MAE and RMSE, which indicates that interval estimation is more challenging. For the prediction coverage, NSBSR shows good performance as it is the closest to 95%.

3.2.2 AURA application

The second dataset is total column ozone data from TOMS-Like Ozone and Radiative Cloud Fraction L3 1 day $0.25° \times 0.25°$ V3 (DOI: 10.5067/Aura/OMI/DATA3002) by the Ozone Monitoring Instrument (OMI) onboard the AURA satellite. We consider the following covariates which can affect the level of ozone concentration as the ozone is a secondary pollutant: (1) Radiative cloud fraction (DOI: 10.5067/Aura/OMI/DATA3002); (2) solar Zenith angle (DOI: 10.5067/Aura/OMI/DATA3002); (3) total column of nitrogen dioxide (DOI: 10.5067/Aura/OMI/DATA3007); (4) total column of formaldehyde (DOI: 10.5067/Aura/OMI/DATA2016); (5) ultra violet aerosol index (DOI: 10.5067/Aura/OMI/DATA2025); (6) total column of sulfar dioxide (DOI: 10.5067/Aura/OMI/DATA2025). (3) and (4) are log-transformed for better interpretability. The achieved OMI/AURA dataset has global coverage with $0.25° \times 0.25°$ resolution. We again focus on a neighborhood of the Korea peninsula, i.e. longitude ranged from 112 to 141°, latitude from 24 to 53°. For this time, we consider the averaged data between June 1 to August 31, 2019. Note that there were missing values in hourly data due to satellite's orbits and other random sources but we averaged the data wherever available. As a result, there is no missing value for ozone concentration while some covariates still have some missing values. In this case, we imputed those missing values by the ordinary kriging. Stationarity for the AURA data is also tested by the method used for the MODIS data. The corresponding p value is 0.672. So, we can not reject the non-stationarity for the AURA data based on this test as well. However, p value is smaller compared to the case of the MODIS data.

For the AURA data, we further randomly removed 20% of the data to see the impact of missing values. Then, we took a subset at 0.5° resolution for fitting each method and prediction is made at 0.25° resolution. Table 5 shows the prediction performance. Similar to Table 4, the results are sorted in ascending order by MAE. RMSE, INT and PC are provided as well. NSBSR provides the fifth best result in terms of MAE and RMSE at this time. The ranks among the methods are not the same compared to the results for the MODIS data. For example, NNGP which show the best result for the MODIS data places the eighth for the AURA data. The MAE and RMSE of first six methods are relatively small compared to those of the remaining methods. The results for our approach are not as good as the ones for the MODIS data. However, it shows reasonable performance given that our approach is under stationary assumption while several methods allow more flexible non-stationarity. The values of INT vary more among the methods compared to MAE and RMSE, which is same as the MODIS data. Note that LK is best in terms of MAE and RMASE but INT is largest. The results of prediction coverage also vary but our approach still shows reasonable performance.

4 Discussion

We have proposed Bayesian spatial regression with non-parametric modeling of spectral density derived from Fourier Transform. Our approach, NSBSR has achieved reasonable computational efficiency in terms of storage and speed by using the Whittle likelihood approximation and the Fast Fourier Transform algorithm, even though there are more parameters to estimate compared to parametric covariance models. Simulation studies show that NSBSR is relatively robust compared to parametric covariance

Table 5 Prediction results for the AURA data from various methods based on mean absolute error (MAE), root mean squared error (RMSE), confidence/credible interval length (INT), and prediction coverage(PC)

Name	MAE	RMSE	INT	PC
LK	0.937	1.303	34.868	0.9473
SPDE	1.059	1.720	6.784	0.9803
FRK	1.117	1.620	7.365	0.9875
LAGP	1.305	1.841	5.315	0.8303
NSBSR	1.502	1.901	13.344	0.9853
Partition	1.642	2.898	3.352	0.3964
Gapfill	2.454	4.192	4.419	0.5708
NNGP	5.330	6.743	26.258	0.9531
PP	5.883	7.130	7.431	0.7025
Periodic	5.894	8.100	26.393	0.5742
Tapering	7.713	9.560	6.348	0.4001

The results are sorted in ascending order by MAE models and/or isotropic assumption. Also, NSBSR shows better prediction results in a sense that RMSPE is lower than those of parametric covariance models for smoother processes. Our approach requires stationary assumption, which is rather limited given that several methods to handle non-stationary spatial data are available. However, comparison analysis (see Tables 4, 5) using two ozone concentration datasets show that our approach can provide reasonable prediction given the variation in prediction performance among methods for different datasets. Thus, NSBSR is a good alternative to the existing prediction approaches in spatial data analysis.

Our approach could be used as a baseline for capturing a more complicated spatial dependence structure than that of stationary Gaussian fields. We can consider the marginal variance σ_ϵ in the Eq. (1) to be $\sigma_\epsilon(s)$ so that it is spatially varying. The resulting process becomes non-stationary. We could apply our approach to a stationary error process component, $e(\boldsymbol{s})$, so that we can handle a class of non-stationary processes.

Estimated spectral densities are not as good as the prediction results. It could be due to the DFT approximation or Whittle likelihood approximation with further approximation using a five-component mixture Gaussian. Yaglom (1987) pointed out that DFT approximation rather than the exact Fourier transform could cause accuracy issue on the covariance estimation. In addition, insufficient sample size or truncated study region could possibly have a negative effect on the spectral density estimation. A possible remedy would be a different likelihood approximation than the Whittle likelihood so that we can avoid using periodogram itself but it requires theoretical justification. An empirical choice of hyperparameters such as ρ_0 for ρ_{θ_1}, ρ_{θ_2} might be beneficial to enhance prediction accuracy, as well. However, these are rather subjective and we tried usual practice of vague priors in our analysis.

The proposed method requires the observations on a spatial lattice. We introduced a way to handle when the observations are on an incomplete lattice, which can be viewed as irregularly spaced data as well. For completely random observation locations not on a spatial lattice, one can consider an idea by Fuentes (2007) when the sample size is large. Fuentes (2007) proposed to aggregate the data points within each grid and treated them as observations of an integrated process on a spatial lattice. This can be a future direction to extend our approach.

Appendix A: Conditional posterior distributions

Conditional posterior distributions of the parameters for Gibbs samplers are described below.

Posterior of $\boldsymbol{\beta}$ With a Gaussian prior $\boldsymbol{\beta} \sim \boldsymbol{N}(\mu_\beta \mathbf{1}, \sigma_\beta^2 \boldsymbol{I})$, we have

$$\begin{aligned}
\boldsymbol{\beta} \mid \dots &\sim \boldsymbol{N}(\boldsymbol{\mu}_\star, \boldsymbol{T}_\star),\\
\boldsymbol{T}_\star &= \boldsymbol{X}^t \tilde{\boldsymbol{\Gamma}}^{-1} \boldsymbol{X} + \sigma_\beta^{-2} \boldsymbol{I}_p,\\
\boldsymbol{\mu}_\star &= \boldsymbol{T}_\star^{-1} \left(\boldsymbol{X}^t \tilde{\boldsymbol{\Gamma}}^{-1} \boldsymbol{Y} + \sigma_\beta^{-2} \mu_\beta \boldsymbol{I} \right),
\end{aligned}$$

where $\tilde{\boldsymbol{\Gamma}}$ is the covariance matrix of the data and it is constructed by following. Let

$$\tilde{P}_{n_1n_2}(u_1,u_2)=\sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1} f_{\Delta}(w_j,w_k)\exp(\imath(w_ju_1+w_ku_2))$$

for $w_j=-\frac{\pi}{\delta_1}+\frac{2\pi}{\delta_1}\frac{j}{n_1}$, $w_k=-\frac{\pi}{\delta_2}+\frac{2\pi}{\delta_2}\frac{k}{n_2}$. We construct the circulant matrix $\tilde{B}^{(k)}_{n_1}=Circulant(\tilde{P}^{(k)}_{n_1n_2})$, where $\tilde{P}^{(k)}_{n_1n_2}$ is the kth column vector of $\tilde{P}_{n_1n_2}$. Then,

$$\tilde{\boldsymbol{\Gamma}}=\begin{bmatrix}\tilde{B}^{(1)}_{n_1} & \tilde{B}^{(2)}_{n_1} & \dots & \tilde{B}^{(n_2)}_{n_1}\\ \tilde{B}^{(n_2)}_{n_1} & \tilde{B}^{(1)}_{n_1} & \dots & \tilde{B}^{(n_2-1)}_{n_1}\\ \vdots & \ddots & \ddots & \vdots\\ \tilde{B}^{(2)}_{n_1} & \dots & \tilde{B}^{(n_2)}_{n_1} & \tilde{B}^{(1)}_{n_1}\end{bmatrix}.$$

Instead of computing $\tilde{\boldsymbol{\Gamma}}^{-1}$, we actually need to compute $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{X}$ and $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{Y}$. We adopt the method by Anitescu et al. (2012) so that we can reduce the computation of both $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{X}$ and $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{Y}$ to $\mathcal{O}(n_1n_2log(n_1n_2))$. The approach by Anitescu et al. (2012) is as follows. Since $\tilde{\boldsymbol{\Gamma}}$ is a $n_1n_2\times n_1n_2$ block circulant of circulant block (BCCB) matrix, it can be diagonalized by the Kronecker product of 2D FFT matrices of appropriate orders. Let each block $\tilde{B}^{(k)}_{n_1}$ be diagonalized as

$$F_{n_1}\tilde{B}^{(k)}_{n_1}F^{\star}_{n_1}=\Lambda^{(k)}_{n_1},\ \ k=0,\dots,n_2-1 \tag{8}$$

where F_{n_1} is a FFT-matrix such that $(F_{n_1})_{jk}=(w_{\Delta 1})^{jk}/\sqrt{n_1}$ with $w_{\Delta 1}=\pi/\delta_1$, $F^{\star}_{n_1}$ is its conjugate, and $\Lambda^{(k)}_{n_1}$ are its eigenvalues. F_{n2} is similarly defined. Then, the matrix inverse-vector multiplication $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{q}$ with generic vector $\boldsymbol{q}$ is computed by

$$\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{q}=(F_{n_2}\otimes F_{n_1})^{\star}\left[\Lambda^{-1}_{n_1n_2}\left((F_{n_2}\otimes F_{n_1})\boldsymbol{q}\right)\right], \tag{9}$$

where

$$\Lambda_{n_1n_2}=diag\left(\sum_{i=0}^{n_2-1}\Lambda^{(i)}_{n_1},\sum_{i=0}^{n_2-1}w^{i}_{\Delta 1}\Lambda^{(i)}_{n_1},\cdots,\sum_{i=0}^{n_2-1}w^{(n_2-1)i}_{\Delta 1}\Lambda^{(i)}_{n_1}\right).$$

Since the elements of $\Lambda_{n_1n_2}$ are modelled from the MCMC procedure, we can evaluate (9) during Gibbs iterations by first computing an 2D FFT on $\boldsymbol{q}$, then dividing the resulting vector by the eigenvalues, and performing an inverse 2D FFT on the resulting vector.

Posterior of τ_ϵ Let $\boldsymbol{r}=\boldsymbol{Y}-\boldsymbol{X\beta}$. With a Gamma prior for τ_ϵ we have

$$\tau_\epsilon\mid\dots\sim G(a_\star,b_\star),$$

where $a_\star=a+\frac{n_1n_2}{2}$, $b_\star=b+\frac{1}{2}\boldsymbol{r}^t\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{r}$.

Posterior of $\boldsymbol{\theta}$ Let the normalized residual $\boldsymbol{r}_\star=\tau_\epsilon^{1/2}(\boldsymbol{Y}-\boldsymbol{X\beta})$. Given $\boldsymbol{r}_\star$, we compute $\boldsymbol{\varphi}$ by (4). Let $\boldsymbol{Y}$ be a covariance kernel matrix of the process $\boldsymbol{\theta}$. With a Gaussian process prior of $\boldsymbol{\theta}$, we obtain

$$\begin{aligned}\boldsymbol{\theta} \mid \cdots &\sim \boldsymbol{N}(\boldsymbol{\nu}_\star, \boldsymbol{\Upsilon}_\star),\\ \boldsymbol{\Upsilon}_\star &= (\boldsymbol{\Upsilon}^{-1} + \boldsymbol{V}_\psi^{-1})^{-1},\\ \boldsymbol{\nu}_\star &= \boldsymbol{\Upsilon}_\star \boldsymbol{V}_\psi^{-1}(\boldsymbol{\varphi} - \boldsymbol{\kappa}_\psi - \boldsymbol{\nu}) + \boldsymbol{\nu}\end{aligned}$$

where $\boldsymbol{V}_\psi = diag\{v^2_{\psi_0}, \dots, v^2_{\psi_{n^h}}\}$, and $\boldsymbol{\kappa}_\psi = (\kappa_{\psi_0}, \dots, \kappa_{\psi_{n^h}})'$ for the assigned n^h Fourier frequencies. The dimension of $\boldsymbol{\Upsilon}_*$ is $n^h \times n^h$, where n^h is the total number of Fourier frequencies we considered. i.e. $n^h = \lceil \frac{N}{2} \rceil$, where $\lceil x \rceil$ is the smallest integer greater than or equal to x and N is the sample size in one direction. This is due to the symmetry of Periodogram $\mathcal{I} = e^{\theta}$. Thus the dimension is lower than that of the data covariance matrix, which is $N \times N$. Calculation of $\boldsymbol{\Upsilon}_\star = (\boldsymbol{\Upsilon}^{-1} + \boldsymbol{V}_\psi^{-1})^{-1}$ is done in a more effective way by the Woodbury's formula, i.e. $\boldsymbol{\Upsilon}_\star = (\boldsymbol{\Upsilon}^{-1} + \boldsymbol{V}_\psi^{-1})^{-1} = \boldsymbol{\Upsilon} - \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon} + \boldsymbol{V}_\psi)^{-1}\boldsymbol{\Upsilon}$. However, this part is still relatively expensive compared to the other parts of the algorithm.

Posterior of ψ Given the prior $P(\psi = l) = p_l$, for $l = 1, \dots, 5$,

$$P(\psi_s = l \mid \dots) = p_l \phi_{v_l}(\varphi_s - \theta_s - \kappa_l),$$

for $s = 1, 2, \dots, n^h$.

Posterior of τ_θ

$$\tau_\theta \mid \dots \sim G(c_\star, d_\star),$$

where $c_\star = c + \frac{n^h}{2}$ and $d_\star = d + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta} - \boldsymbol{\nu})$.

Posterior of $\rho_{\theta_1}, \rho_{\theta_2}$

We use a "grid-search" method to sample ρ_{θ_1} and ρ_{θ_2}, respectively. That is, for $j = 1, 2$, we sample ρ_{θ_j} among its candidates $\{\rho^{(1)}_{\theta_j}, \rho^{(2)}_{\theta_j}, \dots, \rho^{(M)}_{\theta_j}\}$ with probability weights $\pi(\rho^{(l)}_{\theta_j}) \big/ \sum_{m=1}^{M} \pi(\rho^{(m)}_{\theta_j})$, satisfying

$$\pi(\rho^{(l)}_{\theta_j}) \propto \exp\left(-\frac{1}{2} log \left| \boldsymbol{\Upsilon}^{-1}_{(\rho^{(l)}_{\theta_j}, \rho_{\theta_{-j}})} \right| - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}^{-1}_{(\rho^{(l)}_{\theta_j}, \rho_{\theta_{-j}})}(\boldsymbol{\theta} - \boldsymbol{\nu})\right),$$

where $|\cdot|$ is a determinant and $\boldsymbol{\Upsilon}^{-1}_{(\rho^{(l)}_{\theta_j}, \rho_{\theta_{-j}})}$ is a covariance kernel matrix with range parameters $\rho^{(l)}_{\theta_j}$ and $\rho_{\theta_{-j}}$ for $j = 1, 2$. We used $M (= 10)$ equi-spaced values in $(0, \rho_0)$ as the set of candidates.

Appendix B: Specification of a five-component mixture Gaussian distribution

A five-component mixture Gaussian distribution as approximation of the distribution of the logarithm of an exponential distribution in Carter and Kohn (1997) is given by

$$\pi(\xi) = \sum_{l=1}^{5} p_l \phi_{v_l}(\xi - \kappa_l),$$

where $\phi_v(\cdot - \kappa)$ is a normal density function with mean κ and variance v^2. The weights (p_l), means (κ_l) and standard deviations (v_l) of the five components in the mixture Gaussian distribution are as follows: If a Fourier frequency w is on the boundary, $(p_1,p_2,p_3,p_4,p_5) = (0.13, 0.16, 0.23, 0.22, 0.25)$; $(\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5) = (4.63, -2.87, -1.44, -0.33, 0.76)$; $(v_1, v_2, v_3, v_4, v_5) = (8.75, 1.95, 0.88, 0.45, 0.41)$. Otherwise, $(p_1,p_2,p_3,p_4,p_5) = (0.19, 0.11, 0.27, 0.25, 0.18)$; $(\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5) = (2.20, -0.80, -0.55, -0.035, 0.48)$; $(v_1, v_2, v_3, v_4, v_5) = (1.93, 1.01, 0.69, 0.60, 0.29)$.

Acknowledgements This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C1002213, No. 2020R1A4A1018207).

References

Anitescu, M., Chen, J., & Wang, L. (2012). A matrix-free approach for solving the parametric gaussian process maximum likelihood problem. *SIAM Journal on Scientific Computing, 34*(1), A240–A262.

Aune, E., Simpson, D. P., & Eidsvik, J. (2014). Parameter estimation in high dimensional gaussian distributions. *Statistics and Computing, 24*(2), 247–263.

Banerjee, S., Gelfand, A. E., Finley, A. O., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70*(4), 825–848.

Bracewell, R. N. (1986). *The Fourier transform and its applications* (2nd ed.). McGraw-Hill.

Brillinger, D. R. (2001). *Time series: Data analysis and theory*. SIAM.

Carter, C. K., & Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 59*(1), 255–268.

Choudhuri, N., Ghosal, S., & Roy, A. (2004). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association, 99*(468), 1050–1059.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation, 19*(90), 297–301.

Cressie, N. (1993). *Statistics for spatial data*. Wiley.

Cressie, N., & Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70*(1), 209–226.

Datta, A., Banerjee, S., Finely, A. O., & Gelfand, A. E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association, 111,* 800–812.

Dey, T., Kim, K. H., & Lim, C. Y. (2018). Bayesian time series regression with nonparametric modeling of autocorrelation. *Computational Statistics, 33,* 1715–1731.

Du, J., Zhang, H., & Mandrekar, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *The Annals of Statistics, 37,* 3330–3361.

Finley, A. O., Sang, H., Banerjee, S., & Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational Statistics & Data Analysis, 53,* 2873–2884.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association, 102,* 321–331.

Furrer, R., Genton, M. G., & Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics, 15,* 502–523.

Gelfand, A. E., Diggle, P., Fuentes, M., & Guttorp, P. (Eds.). (2010). *Handbook of spatial statistics*. CRC Press.

Gerber, F., Furrer, R., Schaepman-Strub, G., de Jong, R., & Schaepman, M. E. (2018). Predicting missing values in spatio-temporal satellite data. *IEEE Transactions on Geoscience and Remote Sensing, 56,* 2841–2853.

Gramacy, R., & Apley, D. (2015). Local gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics, 24,* 561–578.

Guhaniyogi, R., & Banerjee, S. (2018). Meta-kriging: Scalable bayesian modeling and inference for massive spatial datasets. *Technometrics, 60,* 430–444.

Guinness, J. (2018). Nonparametric spectral methods for multivariate spatial and spatial-temporal data. arXiv:1811.01280.

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

Guinness, J., & Fuentes, M. (2017). Circulant embedding of approximate covariances for inference from gaussian data on large lattices. *Journal of Computational and Graphical Statistics, 26*(1), 88–97.

Heaton, M. J., Christensen, W. F., & Terres, M. A. (2017). Nonstationary gaussian process models using spatial hierarchical clustering from finite differences. *Technometrics, 59,* 93–101.

Heaton, M. J., Datta, A., Finley, A. O., et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics, 24,* 398–425.

Katzfuss, M., & Gong, W. (2017). Multi-resolution approximations of gaussian processes for large spatial datasets. arXiv:1710.08976.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association, 112,* 201–214.

Katzfuss, M., & Cressie, N. (2011). Spatio-temporal smoothing and em estimation for massive remote-sensing data sets. *Journal of Time Series Analysis, 32,* 430–446.

Kaufman, C. G., Schervish, M. J., & Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association, 103,* 1545–1555.

Kim, H. M., Mallick, B. K., & Holmes, C. (2005). Analyzing nonstationary spatial data using piecewise gaussian processes. *Journal of the American Statistical Association, 100,* 653–668.

Kim, J. K., & Shao, J. (2013). *Statistical methods for handling incomplete data*. CRC Press.

Kim, S. Y., & Song, I. (2017). National-scale exposure prediction for long-term concentrations of particulate matter and nitrogen dioxide in South Korea. *Environmental Pollution, 226,* 21–29.

Konomi, B. A., Sang, H., & Mallick, B. K. (2014). Adaptive Bayesian nonstationary modeling for large spatial datasets using covariance approximations. *Journal of Computational and Graphical Statistics, 23,* 802–829.

Laden, F., Schwartz, J., Speizer, F. E., & Dockery, D. W. (2006). Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard six cities study. *American Journal of Respiratory and Critical Care Medicine, 173*(6), 667–672.

Lim, C., & Stein, M. L. (2008). Properties of spatial cross-periodograms using fixed-domain asymptotics. *Journal of Multivariate Analysis, 99,* 1962–1984.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(4), 423–498.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., & Sain, S. (2015). A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics, 24,* 579–599.

Nychka, D., Wikle, C., & Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling, 2*(4), 315–331.

Paciorek, C. J. (2007). Computational techniques for spatial logistic regression with large data sets. *Computational Statistics & Data Analysis, 51,* 3631–3653.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic Press.

Reich, B. J., & Fuentes, M. (2012). Nonparametric Bayesian models for a spatial covariance. *Statistical Methodology, 9*(1–2), 265–274.

Royle, A. J., & Wikle, C. K. (2005). Efficient statistical mapping of avian count data. *Environmental and Ecological Statistics, 12,* 225–243.

Sang, H., Jun, M., & Huang, J. Z. (2011). Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *The Annals of Applied Statistics, 5,* 2519–2548.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. Springer.

Stein, M. L. (2008). A modeling approach for large spatial datasets. *Journal of the Korean Statistical Society, 37*(1), 3–10.

Stroud, J. R., Stein, M. L., & Lysen, S. (2017). Bayesian and maximum likelihood estimation for gaussian processes on an incomplete lattice. *Journal of Computational and Graphical Statistics, 26*(1), 108–120.

Taylor, S. L., Eckley, I. A., & Nunes, M. A. (2014). A test of stationarity for textured images. *Technometrics, 56,* 291–301.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika, 41,* 434–449.

Yaglom, A. M. (1987). *Correlation theory of stationary and related random functions I: Basic results*. Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.