Statistics and Its Interface Volume 12 (2019) 549–559

Semiparametric Bayesian analysis of transformation spatial mixed models for large datasets*

Ying Wu, Dan Chen†, and Niansheng Tang

In *spatial mixed models* (SMMs), it is commonly assumed that stationary spatial process and random errors independently follow the Gassian distribution. However, in some applications, this assumption may be inappropriate. To this end, this paper proposes a *transformation spatial mixed models* (TSMMs) to accommodate large dataset that follows the non-Gaussian distribution. With the help of Gibbs sampler algorithm, a semiparametric Bayesian approach is developed to make inference on TSMMs by using Bayesian P-splines to approximate transformation function, and a fixed number of known but not necessarily orthogonal spatial basis functions with multi-resolution analysis method to approximate nonstationary spatial process. Instead of Wishart distribution assumption for the prior of precision matrix of random effects, we consider Cholesky decomposition of the precision matrix, and specify the priors for unknown components in low unit triangular matrix and diagonal matrix. Simulation studies and an example are used to illustrate the proposed methodologies.

AMS 2000 subject classifications: Primary 62F15; secondary 62J05.
Keywords and phrases: Gibbs sampler, Large dataset, Matrix decomposition, Semiparametric Bayesian analysis, Transformation spatial Mixed Models.

1. INTRODUCTION

In much economical, social, physical and epidemiological research, substantive theory and practice usually involve the analysis of spatial data (e.g., Cressie [3], Anselin and Bera [4], Anselin and Syabri [12]). Various statistical models, theories and methods have been developed to analyze spatial data over the past years (e.g., Bai [5], Cai [9], Chan [10]). In particular, SMMs are widely adopted to capture spatial structures of spatial and spatial-temporal datasets from very fine to very large scales by incorporating random effects in the analysis of spatial data. For example, see Cressie and Johannesson [12] for the analysis of very large datasets, Kang and Cressie [23] for Bayesian inference, Nychka, Wikle and Royle [32], Nychka et al. [31] and Katzfuss [25] for multi-resolution analysis based on the W-wavelet theories (e.g., Kwong and Tang [26]).

At present, there are more than a dozen spatial data analysis software packages to adapt to strong demand in various fields, for example, see SpaceStat package [1], GeoDa [2], spdep and DCluster on the open source Comprehensive R Archive Network, and ArcGIS 10.7. However, these packages cannot deal with non-normal spatial datasets. In particular, they cannot deal with binary/poisson data with a binomial/poisson distribution, which is often encountered in economical, social and epidemiological studies. For example, a remote sensing dataset, which was retrieved by the *Multi-angle Imaging SpectroRadiometer* (MISR) instrument on NASA's Terra and Aqua satellite, was collected monthly at a spatial resolution, which was converted to 3-degrees dataset (e.g., 1.0×1.0, 0.5×0.5, 0.25×0.25). The MISR instrument is one of the key instruments on board that were widely used to collect the global atmosphere information such as the *aerosol optical depth* (AOD), *Water Vapor* (WV), *Net Radiation* (NR) and *Carbon Monoxide* (CM). The histograms of the 2-degrees AOD, WV, NR and CM datasets during May 2016 between longitudes -179.75 and +179.75, and between latitudes -89.75 and +89.75 (e.g., Figure 6) show that the AOD, WV, NR and CM data are non-normal. Hence, the large scale dataset motivates considering the transformation spatial mixed models.

Considerable work has been focused on approaches to relax the normality assumption in spatial models over the past years. For example, see De Oliveira et al. [15], Genton and Zhang [19], Zhang and EI-Shaarawi [41], and Zareifard and Khahedi [40]. The aforementioned literature has been developed under the skewed distributional assumption of spatial data by utilizing the Box-Cox family of power transformations or a scale mixing of a unified skew Gaussian process. However, for transformation linear mixed models rather than spatial mixed models, Gurka [21] demonstrated that estimating transformation parameter in the Box-Cox family of power transformations may yield biased estimation

*This research is supported by the National Natural Science Foundation of China (No. 11671349) and the Key Projects of the National Natural Science Foundation of China (No. 11731101), the National Social Science Fund of China (No. 17BTJ038), Yunnan Applied Basic Research Projects (No. 2016FD087).
†Corresponding author.

of variance components, and maximum likelihood estimator of transformation parameter is not consistent. To address these issues, various transformation models have been developed in a Bayesian framework. In particular, Bayesian P-splines approach has been used to handle transformation models. For example, Song and Lu [36] presented a semiparametric transformation model with P-splines; Song and Liu [35] considered a transformation structure equation model with highly non-normal and incomplete data with P-splines; Tang, Wu and Chen [37] extended to a transformation linear mixed model based on the truncated centered Dirichlet Process prior approximation to the distribution of random effects. However, to our knowledge, little work has been done on TSMMs based on the idea of Bayesian P-splines in a Bayesian framework.

The main contributions of this paper include as follows. First, the P-splines is employed to approximate transformation functions because of its good properties, such as allowing for simultaneous estimation of smooth functions and smoothing parameters, allowing the smoothing parameters to be locally adaptive, and the flexibility of implementing Bayesian analysis via the R program. Second, the multi-resolution analysis technique with a fixed number of known but not necessarily orthogonal spatial basis functions is adopted to approximate nonstationary spatial process. Third, a *Markov chain Monte Carlo* (MCMC) algorithm is developed to to simultaneously obtain Bayesian estimations of unknown parameters, random effects and transformation functions by combining the Gibbs sampler and W-wavelet theories. Fourth, instead of traditional Wishart distribution assumption for the prior of precision matrix of random effects, we consider Cholesky decomposition of the precision matrix for random effects, and then specify the priors for unknown components in low unit triangular matrix and diagonal matrix, which largely reduce the number of unknown parameters and improve the convergence rate of MCMC algorithm. Fifth, we investigate the identifiability of the considered model. Sixth, we present a Bayesian model comparison approach by utilizing Bayes factor and Path sampling method, and discuss the goodness-of-fit assessment problem via the posterior predictive (PP) p-value.

The rest of this paper is organized as follows. Section 2 introduces TSMMs including the selection of the W-wavelet matrix, Bayesian P-splines approach to approximate the transformation function. Bayesian inference including parameter estimation, model comparison and goodness-of-fit statistic are given in Section 3. Simulation studies and an example are illustrated in Section 4. Technical details are presented in the Appendix.

2. TRANSFORMATION SPATIAL MIXED MODELS

2.1 Models and notation

Let $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{S} \subset \mathbb{R}^d\}$ be a real-valued spatial process. Our main interest is to make statistical inference on the Y-process on the basis of the actually observed non-normal dataset $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{S}\}$. To this end, we consider the following measurement error model

$$f(Z(\boldsymbol{s})) = Y(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{S}, \tag{1}$$

where $\{\varepsilon(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{S}\}$ is an independent Gaussian process that is independent of $Y(\cdot)$ and has mean zero and $\text{var}\{\varepsilon(\boldsymbol{s})\} = \sigma_\varepsilon^2$. It is assumed that the process $Z(\cdot)$ is known only at a finite number of spatial locations, e.g., $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$. Let $\boldsymbol{Z} \equiv \{Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_N)\}$ be the available data, $f(\cdot)$ be an unknown smooth transformation function for $Z(\boldsymbol{s})$, and denote $Z^*(\boldsymbol{s}) = f(Z(\boldsymbol{s}))$. Suppose that $f(\cdot)$ is strictly monotone and differentiable. Following Cressie and Johannesson [12] and Kang and Cressie [23], we consider the following *spatial mixed models* (SMMs):

$$\begin{cases} Y(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})^\top \boldsymbol{\gamma} + \nu(\boldsymbol{s}) + \xi(\boldsymbol{s}), \\ \nu(\boldsymbol{s}) = \boldsymbol{W}(\boldsymbol{s})^\top \boldsymbol{\eta}, \quad \boldsymbol{s} \in \mathbb{S}, \end{cases} \tag{2}$$

where $\boldsymbol{X}(\cdot) = (X_1(\cdot), \ldots, X_p(\cdot))^\top$ is a $p \times 1$ vector of covariate processes, $\boldsymbol{\gamma}$ is a $p \times 1$ vector of coefficients corresponding to the fixed effects, $\boldsymbol{X}(\cdot)^\top \boldsymbol{\gamma}$ describes the large-scale spatial variation, $\nu(\cdot)$ represents the smooth small-scale spatial variation, $\xi(\cdot)$ is the residual term, $\boldsymbol{\eta}$ is a $q \times 1$ vector of random effects, $\boldsymbol{\eta}$ and $\xi(\cdot)$ define an SMM that can handle spatial variability for small and fine scales, $\boldsymbol{W}(\cdot)$ is a $q \times 1$ vector of q deterministic, known and multi-resolutional spatial basis functions that are not necessarily orthogonal of each other, i.e., $\boldsymbol{W}(\cdot) = (\boldsymbol{w}_1(\cdot), \ldots, \boldsymbol{w}_q(\cdot))^\top$. It is assumed that $\xi(\cdot)$ is independent of $\boldsymbol{\eta}$, $\boldsymbol{\eta}$ is distributed as the multivariate Gaussian distribution with mean zero and $\text{cov}(\boldsymbol{\eta}) = \boldsymbol{\Sigma}$, i.e., $\boldsymbol{\eta} \sim \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{\Sigma})$, and $\xi(\cdot)$ follows the Gaussian distribution $\mathcal{N}(0, \sigma_\xi^2)$. Thus, Equations (1) and (2) define a *transformation spatial mixed models* (TSMMs). From the definition of the TSMMs, it is easily seen that $Z(\boldsymbol{s})$ is non-normal dataset, but $Z^*(\boldsymbol{s})$ is normally distributed utilizing the transformation function $f(\cdot)$.

2.2 Generating multiresolutional spatial basis functions via W wavelet

In model (2), $\boldsymbol{W}(\cdot)$ is usually unknown. To this end, following Kwong and Tang [26], Nychka, Wikle and Royle [32] and Katzfuss [25], we generate a spatial basis using the repeated translations and scalings of several fixed functions, it is called discrete wavelet transform. To generate W wavelet, we first choose a B-splines as the fixed function, and then use the repeated W translations and scalings of B-splines to obtain two templates (e.g., mother and father), and finally determine the number of W wavelets at different resolutions in the same domain. As an illustration, we consider wavelets defined in the interval $[0, 1]$, and a coarsest level of resolution (e.g., say L). The W wavelets used here are plotted in Figure 1, and a basis of 32 functions for $L = 3$ is displayed in Figure 2. The first 3 resolution basis functions are similar

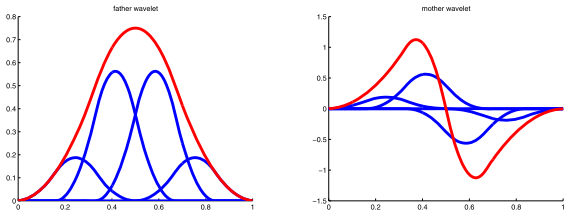

Figure 1. Continuous versions of the father and mother W-transform wavelets.

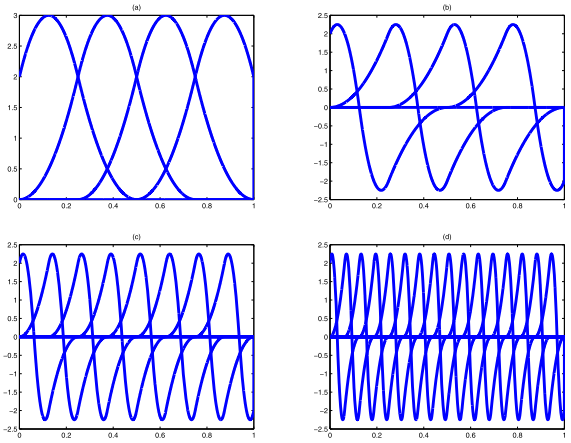

Figure 2. Family of 32 basis functions, (a) is the first 4 functions based on translations of the father wavelet; (b) is the same translated functions for the mother; (c) is 8 translations of the mother wavelet reduced by a factor of two; (d) is 16 translations of the reduced mother wavelet.

to the father wavelet translated in equally spaced locations (e.g., see Figure 2(a)). The first 3 resolution basis functions are the mother wavelet translated by the same manner, and are shown in Figure 2(b). The next generation of basis functions has twice as the resolution, and is similar to a scaling and translation of the mother wavelet. Figure 2(c) shows this generation. This cascade continues with the number of members in each subsequent generation, and the resolution increasing by a factor of two. Figure 2(d) completes the basis of size 32. The relationship between W-matrix and W-wavelet is given in Appendix.

2.3 P-splines approximation of transformation function

Due to the unknown form of function $f(\cdot)$, it is impossible to make statistical inference on $\boldsymbol{\gamma}$ and $\boldsymbol{\Sigma}$ via models (1) and (2). To this end, we here adopt the widely used Bayesian P-splines method to approximate $f(\cdot)$. Following Lang and Brezger [27], $f(Z(\boldsymbol{s}))$ can be approximated by the linear combination of the finite B-splines (e.g., De Boor [14]). That is, $f(Z(\boldsymbol{s})) \approx \sum_{k=1}^{K} \beta_k B_k(Z(\boldsymbol{s}))$, where K is the number of B-splines determined by the number of knots in the support of $Z(\boldsymbol{s})$ and the degrees of B-splines, β_k's are unknown coefficients, and $B_k(Z(\boldsymbol{s}))$ is the kth B-splines for $Z(\boldsymbol{s})$. Thus, based on the above approximation of $f(Z(\boldsymbol{s}))$, for any $\boldsymbol{s} \in \mathbb{S}$, (1) and (2) can be written as the following unified form

$$(3) \quad \sum_{k=1}^{K} \beta_k B_k(Z(\boldsymbol{s})) \approx \boldsymbol{X}(\boldsymbol{s})^\top \boldsymbol{\gamma} + \boldsymbol{W}(\boldsymbol{s})^\top \boldsymbol{\eta} + \xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}).$$

To control unknown function $f(\cdot)$ via penalizing the coefficients of the adjacent B-splines, we consider a random walk prior to β_k. In particular, we consider the following fist and second order random walks for β_k:

$$(4) \quad \begin{aligned} \beta_k &= \beta_{k-1} + u_k, && \text{for } k = 2, \ldots, K, \\ \beta_k &= 2\beta_{k-1} - \beta_{k-2} + u_k, && \text{for } k = 3, \ldots, K, \end{aligned}$$

respectively, where $u_k \sim \mathcal{N}(0, \tau^2/\varphi_k)$. It is assumed that β_1 follows a diffuse prior, i.e., $\beta_1 \propto$ constant for the fist-order random walk, and $\beta_1 \propto$ constant and $\beta_2 \propto$ constant for the second-order random walk. Here τ^2 can be regarded as an inverse smoothing parameter that controls the smoothness of the resulting function $f(\cdot)$, and φ_k's are treated as local smoothing parameters, which are employed to control the local smoothness of a function with significantly different curvatures at different $Z(\boldsymbol{s})$'s. The above defined prior distribution for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^\top$ can be written as

$$p(\boldsymbol{\beta}) \sim \left(\frac{1}{\sqrt{2\pi}\tau}\right)^{K-d} \exp\left\{-\frac{1}{2\tau^2}\boldsymbol{\beta}^\top \boldsymbol{M}(\boldsymbol{\varphi})\boldsymbol{\beta}\right\} \times I(\beta_1 < \ldots < \beta_K, \boldsymbol{Q}\boldsymbol{\beta} = 0),$$

where $\boldsymbol{M}(\boldsymbol{\varphi}) = \mathbb{D}^\top \mathrm{diag}(\boldsymbol{\varphi})\mathbb{D}$, $\mathbb{D} = \boldsymbol{D}_{d-1}\boldsymbol{D}_{d-2}\cdots\boldsymbol{D}_1\boldsymbol{D}_0$, $\boldsymbol{\varphi} = (\varphi_{d+1}, \ldots, \varphi_K)^T$, $\boldsymbol{D}_h$ is a $(K-h-1)\times(K-h)$ matrix and is defined as

$$\boldsymbol{D}_h = \begin{pmatrix} -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}$$

for $h = 0, \ldots, d-1$ with $d = 1$ and 2, and d is the order of the random walk.

There are two kinds of approaches to select K. The first one is to fix K at some fixed value, and the other one is to regard K as a random variable that is selected by using the reversible jump MCMC algorithm or Birth-and-Death method. Lang and Brezger [27] compared the above mentioned two methods, and found that the former performed better for functions with moderate curvature, whilst the latter performed better for highly oscillating functions. When the transformation function is monotonic, it is more likely to be a function with the moderate curvature, so we fix K in some applications. More discussions on the selection of K can refer to Lang and Brezger [27].

2.4 Identification

For the above considered TSMMs, it is assumed that the transformation function $f(Z(\boldsymbol{s})) = \sum_{k=1}^{K} \beta_k B_k(Z(\boldsymbol{s}))$ is strictly monotone increasing, and β_k's satisfy the following constraints $\beta_1 < \beta_2 < \ldots < \beta_K$, which is widely adopted to make the estimation of the monotone transformation function easier even if this condition is sufficient but not necessary.

Note that the above considered TSMMs is unidentifiable because (β_k, μ) and $(\beta_k + c, \mu + c)$ have the same likelihood function for any constant c due to the following fact that $\sum_{k=1}^{K} \beta_k B_k(Z(\boldsymbol{s})) - \mu = \sum_{k=1}^{K} (\beta_k + c) B_k(Z(\boldsymbol{s})) - (\mu + c)$ for any constant c, which is obtained by the property of the B-spline: $\sum_{k=1}^{K} B_k(\cdot) = 1$.

For identifiability, we assume that $\sigma_\varepsilon^2 = 1$, the function $f(\cdot)$ has zero mean, i.e., $\sum_{s \in \mathbb{S}} \sum_{k=1}^{K} \beta_k B_k(Z(\boldsymbol{s})) = 0$, which is equivalent to $\boldsymbol{Q}\boldsymbol{\beta} = 0$, where $\boldsymbol{Q} = (\boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_K^*)$ in which $\boldsymbol{B}_k^* = \sum_{s \in \mathbb{S}} B_k(Z(\boldsymbol{s}))$ for $k = 1, \ldots, K$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^\top$.

3. BAYESIAN INFERENCE

3.1 Prior distribution

Following Kang and Cressie [23] and Song and Lu [36], we consider the following priors for $\boldsymbol{\gamma}$, τ^{-2}, φ_k, σ_ξ and $\boldsymbol{\beta}$:

$$p(\boldsymbol{\gamma}) \sim \mathcal{N}_p(\boldsymbol{\mu}_{\gamma 0}, \sigma_{\gamma 0}^2 \boldsymbol{I}_p), \quad p(\tau^{-2}) \sim \Gamma(e_1, e_2),$$
$$p(\varphi_k) \sim \Gamma(b/2, b/2), \quad p(\sigma_\xi) \sim \mathrm{U}(0, \kappa_\xi),$$

where $\Gamma(\cdot, \cdot)$ represents the gamma distribution, $U(\cdot, \cdot)$ is the uniform distribution; $\boldsymbol{\mu}_{\gamma 0}$, $\sigma_{\gamma 0}^2$, e_1, e_2, b and κ_ξ are the pre-given hyperparameters.

To present the prior of $\boldsymbol{\Sigma}$ associated with $\boldsymbol{\eta}$, following Chan and Jeliazkov [10], we consider the following Cholesky decomposition of $\boldsymbol{\Sigma}^{-1}$:

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{A}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{A}, \tag{5}$$

where $\boldsymbol{A}$ is a lower unit triangular matrix, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_q)$ in which $\lambda_j > 0$ for $j = 1, \ldots, q$. For simplicity, let a_{ij} $(1 \le j < i \le q)$ denote the free elements of lower unitriangular matrix $\boldsymbol{A}$, i.e.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ a_{21} & 1 & 0 & \ldots & 0 & 0 \\ a_{31} & a_{32} & 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{q1} & a_{q2} & \ldots & \ldots & a_{q,q-1} & 1 \end{pmatrix}.$$

Also, denote $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)^\top$, $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{i,i-1})^\top$ for $i = 2, \ldots, q$, and $\boldsymbol{a} = (\boldsymbol{a}_2^\top, \ldots, \boldsymbol{a}_q^\top)^\top$. Based on the above parameterization, we consider the following priors for λ_i and $\boldsymbol{a}_j$ $(i = 1, \ldots, q,\ j = 2, \ldots, q)$:

$$\lambda_i^{-1} \sim \Gamma\left(\frac{v_{0i}}{2}, \frac{u_{0i}}{2}\right), \quad \boldsymbol{a}_j \sim \mathcal{N}_{j-1}(\boldsymbol{a}_{j0}, \boldsymbol{A}_{j0})$$

where v_{0i}, u_{0i}, $\boldsymbol{a}_{j0}$ and $\boldsymbol{A}_{j0}$ are the pre-given hyperparameters.

3.2 Gibbs sampler

To avoid the high-dimensional integral involved in making statistical inference via the marginal probability density function $p(\boldsymbol{Z})$, the Gibbs sample is adopted to draw a sequence of random observations from the joint conditional distribution $p(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma_\xi^{-2}, \tau^{-2}, \boldsymbol{\varphi}, \boldsymbol{a}, \boldsymbol{\lambda} | \mathcal{D})$, where $\mathcal{D} = \{\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{W}\}$. The Gibbs sampler is implemented by iteratively drawing observations from the following conditional distributions: (a) $p(\boldsymbol{\gamma} | \boldsymbol{\eta}, \sigma_\xi^{-2}, \boldsymbol{\beta}, \mathcal{D})$, (b) $p(\boldsymbol{\eta} | \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{\lambda}, \mathcal{D})$, (c) $p(\boldsymbol{\xi} | \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma_\xi^{-2}, \boldsymbol{\beta}, \mathcal{D})$, (d) $p(\boldsymbol{\beta} | \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma_\xi^{-2}, \boldsymbol{\varphi}, \mathcal{D})$, (e) $p(\sigma_\xi^{-2} | \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\beta}, \mathcal{D})$, (f) $p(\tau^{-2} | \boldsymbol{\beta}, \boldsymbol{\varphi})$, (g) $p(\boldsymbol{\varphi} | \boldsymbol{\beta}, \tau^{-2})$ (h) $p(\boldsymbol{a} | \boldsymbol{\eta}, \boldsymbol{\lambda})$, (i) $p(\boldsymbol{\lambda} | \boldsymbol{\eta}, \boldsymbol{a})$.

The conditional distributions required in implementing the aforementioned Gibbs sampler are presented in the Appendix. Convergence of the above presented Gibbs sampler algorithm can be monitored by the *estimated potential scale reduction* (EPSR) value evaluated from several parallel sequences of observations as the runs proceed [17]. Convergence is claimed if all the EPSR values of unknown parameters are less than 1.2.

3.3 Bayesian estimates

Let $\{(\boldsymbol{\gamma}^{(m)}, \boldsymbol{\eta}^{(m)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\theta}_\eta^{(m)}) : m = 1, \ldots, \mathcal{M}\}$ be the observations of $(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\theta}_\eta)$ simulated from the joint conditional distribution $p(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma_\zeta^{-2}, \tau^{-2}, \boldsymbol{\varphi}, \boldsymbol{\theta}_\eta | \mathcal{D})$ via the preceding presented MCMC algorithm after the MCMC algorithm converges, where $\boldsymbol{\theta}_\eta$ is a set of unknown parameters associated with the distribution of random effects. Thus, Bayesian estimates of $\boldsymbol{\theta}_\eta, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ can be evaluated by

$$\widehat{\boldsymbol{\theta}}_\eta = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \boldsymbol{\theta}_{\eta\xi}^{(m)}, \quad \widehat{\boldsymbol{\beta}} = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \boldsymbol{\beta}^{(m)},$$
$$\widehat{\boldsymbol{\gamma}} = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \boldsymbol{\gamma}^{(m)}, \quad \widehat{\boldsymbol{\eta}} = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} \boldsymbol{\eta}^{(m)},$$

respectively. The function $f(\boldsymbol{Z})$ can be estimated by $\widehat{f}(\boldsymbol{Z}) = \sum_{k=1}^{K} \widehat{\beta}_k B_k(\boldsymbol{Z})$, where $\widehat{\beta}_k$ is the kth component of $\widehat{\boldsymbol{\beta}}$. Similarly, the consistent estimates of covariance matrices of $\widehat{\boldsymbol{\theta}}_\eta$, $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\eta}}$ can be obtained by using their corresponding sample covariance matrices of the simulated observations $\{(\boldsymbol{\theta}_\eta^{(m)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)}, \boldsymbol{\eta}^{(m)}) : m = 1, \ldots, \mathcal{M}\}$. For example, $\widehat{\mathrm{var}}(\widehat{\boldsymbol{\beta}}) = \{\mathcal{M} - 1\}^{-1} \sum_{m=1}^{\mathcal{M}} (\boldsymbol{\beta}^{(m)} - \widehat{\boldsymbol{\beta}})(\boldsymbol{\beta}^{(m)} - \widehat{\boldsymbol{\beta}})^\top$. Then, the standard errors for the components of $\widehat{\boldsymbol{\beta}}$ can be obtained by using the square roots of their corresponding diagonal elements of $\widehat{\mathrm{var}}(\widehat{\boldsymbol{\beta}})$.

3.4 Bayesian model comparison

Bayes factor is an important statistic for comparing several competing models in a Bayesian framework, and is

widely used to make model comparison in various statistical models. But it is rather difficult to calculate Bayes factor because of the high-dimensional integral involved. To address the issue, some methods such as bridge sampling (e.g., Meng and Wong [30]) and path sampling (e.g., Gelman and Meng [18]) have been developed. Here, the path sampling method is used to calculate Bayes factor for comparing two competing models $\mathcal{H}_1$ and $\mathcal{H}_0$:

$$B_{10} = \frac{p(\boldsymbol{Z}^*, \boldsymbol{X}, \boldsymbol{W}|\mathcal{H}_1)}{p(\boldsymbol{Z}^*, \boldsymbol{X}, \boldsymbol{W}|\mathcal{H}_0)},$$

where $p(\boldsymbol{Z}^*, \boldsymbol{X}, \boldsymbol{W}|\mathcal{H}_t)$ is the marginal probability density of $(\boldsymbol{Z}^*, \boldsymbol{X}, \boldsymbol{W})$ under $\mathcal{H}_t$, $\boldsymbol{Z}^* = f(\boldsymbol{Z})$, $p(\boldsymbol{Z}^*, \boldsymbol{X}, \boldsymbol{W}|\mathcal{H}_t) = \int p(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{W}|\boldsymbol{\vartheta}_t, \mathcal{H}_t) p(\boldsymbol{\vartheta}_t|\mathcal{H}_t) d\boldsymbol{\eta} d\boldsymbol{\vartheta}_t$, $\boldsymbol{\vartheta}_t = \{\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_\zeta^{-2}, \boldsymbol{a}, \boldsymbol{\lambda}, \boldsymbol{\varphi}, \tau^{-2}\}$ is the set of parameters under $\mathcal{H}_t$ for $t = 0$ and 1. Similar to Lee and Song [28] and Lee and Tang [29], for a continuous parameter $t \in [0, 1]$, we consider the following class of probability density functions

$$F(t) = \int p(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{W}, t|\boldsymbol{\vartheta}_t) p(\boldsymbol{\vartheta}_t) d\boldsymbol{\eta} d\boldsymbol{\vartheta}_t,$$

where $p(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{W}, t|\boldsymbol{\vartheta}_t)$ is the joint probability density function of $(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{W})$ under $\mathcal{H}_t$ that links $\mathcal{H}_0$ and $\mathcal{H}_1$ with the continuous parameter t such that $\mathcal{H}_t = \mathcal{H}_0$ if $t = 0$, and $\mathcal{H}_t = \mathcal{H}_1$ if $t = 1$. Following Gelman and Meng [18], we have

$$\log B_{10} = \log \frac{F(1)}{F(0)} = \int_0^1 \mathbf{E}\{H(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{W}, \boldsymbol{\vartheta}_t, t)\} dt,$$

where $\mathbf{E}$ represents the expectation taken with respect to the joint conditional probability density $p(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}_t|\boldsymbol{X}, \boldsymbol{W}, t)$, $H(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{W}, \boldsymbol{\vartheta}_t, t) = d \log p(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{X}, \boldsymbol{W}, t|\boldsymbol{\vartheta}_t)/dt$. Then, $\log B_{10}$ can be estimated by

$$\widehat{\log B_{10}} = \frac{1}{2} \sum_{\ell=0}^{\hbar} (t_{(\ell+1)} - t_{(\ell)})(\bar{H}_{(\ell+1)} + \bar{H}_{(\ell)}),$$

where $0 = t_{(0)} < t_{(1)} < \cdots < t_{(\hbar)} < t_{(\hbar+1)} = 1$, $\bar{H}_{(\ell)} = \mathcal{M}^{-1} \sum_{m=1}^{\mathcal{M}} H(\boldsymbol{Z}^*, \boldsymbol{\eta}^{(m)}, \boldsymbol{\vartheta}_{t_{(\ell)}}^{(m)}, t_{(\ell)}, \boldsymbol{X}, \boldsymbol{W})$ and $\{(\boldsymbol{\eta}^{(m)}, \boldsymbol{\vartheta}_{t_{(\ell)}}^{(m)}) : m = 1, \ldots, \mathcal{M}\}$ are the observations generated from the joint conditional probability density $p(\boldsymbol{Z}^*, \boldsymbol{\eta}, \boldsymbol{\vartheta}_t|\boldsymbol{X}, \boldsymbol{W}, t_{(\ell)})$ via the preceding presented Gibbs sampler.

3.5 Goodness-of-fit statistic

To assess the plausibility of the posited model, the posterior predictive (PP) p-value [43] is here adopted. Following Gelman, Meng and Stern [43], the PP p-value is defined as

$$p_B = P\{D(\boldsymbol{Z}_{\text{rep}}|\boldsymbol{\Theta}) \geq D(\boldsymbol{Z}|\boldsymbol{\Theta})|\boldsymbol{X}, \boldsymbol{W}, H_0\},$$

where $\boldsymbol{Z}_{\text{rep}}$ denotes a replication of $\boldsymbol{Z}$, H_0 represents the plausibility of the above considered model, $D(\cdot|\cdot)$ is a discrepancy variable, and $\boldsymbol{\Theta} = \{\boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma_\zeta^{-2}, \tau^{-2}, \boldsymbol{\varphi}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{\lambda}\}$. For our considered model H_0, we take the discrepancy variable as

$$D(\boldsymbol{Z}_{\text{rep}}|\boldsymbol{\Theta}) = \sum_{s \in \mathbb{S}} \left\{ f(\boldsymbol{Z}_{\text{rep}}(\boldsymbol{s})) - \boldsymbol{X}(\boldsymbol{s})^\top \boldsymbol{\gamma} - \boldsymbol{W}(\boldsymbol{s})^\top \boldsymbol{\eta} - \xi(\boldsymbol{s}) \right\}^2 / \sigma_\varepsilon^2,$$

which is asymptotically distributed as the chi-squared distribution with N degrees of freedom. Thus, the PP p-value can be rewritten as

$$p_B = \int P(\chi^2(N) \geq D(\boldsymbol{Z}|\boldsymbol{\Theta})) p(\boldsymbol{\Theta}|\boldsymbol{X}, \boldsymbol{W}) d\boldsymbol{\Theta}.$$

It is rather challenging to evaluate the above integral. To solve the issue, the commonly used Monte Carlo method is adopted to approximate p_B, i.e.,

$$\hat{p}_B = \frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} P\left(\chi^2(N) \geq D(\boldsymbol{Z}|\boldsymbol{\Theta}^{(m)})\right),$$

where the observations $\{\boldsymbol{\Theta}^{(m)} : m = 1, \ldots, \mathcal{M}\}$ are generated from the joint probability density $p(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\beta}, \sigma_\zeta^{-2}, \tau^{-2}, \boldsymbol{\varphi}, \boldsymbol{a}, \boldsymbol{\lambda}|\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{W})$ via the preceding presented Gibbs sampler. Following Lee and Tang [29], the model is plausible if $\hat{p}_B$ is not far from 0.5 (e.g., within the interval (0.3, 0.7)).

4. NUMERICAL EXAMPLE

4.1 Simulation studies

To investigate the performance of the proposed Bayesian estimation procedure, we conducted the first simulation study. In this simulation study, the dataset $\{Z^*(\boldsymbol{s}), \boldsymbol{X}(\boldsymbol{s}), \boldsymbol{W}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{S}\}$ was generated from the TSMMs defined in Equations (3). Here, covariate $\boldsymbol{X} = (X_1, X_2, X_3)^\top$ was generated by sampling X_1 from the uniform distribution Uniform(-2,2), X_2 and X_3 from the standard normal distribution; we took $L = 2$ resolution W-wavelets $\boldsymbol{W}(\cdot)_{q=12}$ as the spatial basis functions; $\boldsymbol{\eta}$ was generated from the multivariate normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, ξ was sampled from the normal distribution $\mathcal{N}(0, \sigma_\xi^2)$; ε was sampled from the normal distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$; and we took $\mathbb{S} = \{s : s = 1, \ldots, N\}$ with $N = \text{longitude} \times \text{latitude}$, where longitude and latitude were taken as 20, respectively. The true values of parameters $\boldsymbol{\gamma}$, σ_ξ^2, σ_ε^2 and $\boldsymbol{\Sigma}$ were taken as $\boldsymbol{\gamma} = (-0.5, 0.5, 0.5)^\top$, $\sigma_\xi^2 = 0.001$, $\sigma_\varepsilon^2 = 1$ and $\boldsymbol{\Sigma} = g\boldsymbol{\Sigma}^0$, respectively, where $g = N(1 - \sigma_\xi^2)/\text{tr}(\boldsymbol{W}^\top \boldsymbol{\Sigma}^0 \boldsymbol{W})$, and $\boldsymbol{\Sigma}^0 = \boldsymbol{R}^{-1} \boldsymbol{Q}^\top \boldsymbol{\Delta}^0 \boldsymbol{Q} \boldsymbol{R}^{-\top}$ in which $\boldsymbol{\Delta}^0 = (\delta_{ij}^0)$ was a $N \times N$ stationary covariance matrix with $\delta_{ij}^0 = (1 - \sigma_\xi^2) \exp(-|i - j|/25)$, and $\boldsymbol{Q}$ and $\boldsymbol{R}$ were the QR decompositions of matrix $\boldsymbol{W}$ (i.e., $\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{R}$).

To investigate the effect of different transformation functions on Bayesian estimations of parameters of interest, we considered the following three types of functions for $f(Z(\boldsymbol{s}))$:

Type A. $f^{-1}(Z^*(\boldsymbol{s})) = \exp(0.5Z^*(\boldsymbol{s}) - 1.0)/(1.0 + \exp(0.5Z^*(\boldsymbol{s}) - 1.0))$ leading to highly skewed data;

Table 1. Performance of Bayesian estimates for different transformation functions (TFs) in the first simulation study

TF		γ_1		γ_2		γ_3	
		Bias	RMS	Bias	RMS	Bias	RMS
A	PS	-0.017	0.100	0.013	0.118	0.024	0.102
	BCO	0.367	0.370	-0.365	0.368	-0.365	0.369
	BC1	0.451	0.452	-0.451	0.452	-0.453	0.454
B	PS	0.002	0.074	0.000	0.084	0.002	0.082
	BCO	0.302	0.308	-0.303	0.311	-0.302	0.310
	BC1	0.403	0.403	-0.404	0.404	-0.407	0.405
C	PS	-0.017	0.078	0.019	0.095	0.013	0.080
	BCO	0.387	0.391	-0.386	0.392	-0.384	0.390
	BC1	0.387	0.391	-0.386	0.392	-0.383	0.390

Note: 'PS' represents estimation obtained with P-splines method, 'BCO' denotes estimation obtained with Box-Cox transformation at the optimal value of ρ, and 'BC1' is estimation obtained with Box-Cox transformation at $\rho = 1.0$.

Table 2. MSPEs of Y for different transformation functions (TFs) in the first simulation study

TF	P-splines	Box-Cox
A	0.113	1.056
B	0.066	0.812
C	0.079	1.221

Type B. $f^{-1}(Z^*(\boldsymbol{s})) = \arctan\{\exp(Z^*(\boldsymbol{s})) + 0.5\}$ yielding non-symmetrically U-shaped data;

Type C. $f^{-1}(Z^*(\boldsymbol{s})) = Z^*(\boldsymbol{s})/5 + \sin((Z^*(\boldsymbol{s}) - 5)/20) + 2$ leading to bimodal data.

Based on the above generated dataset, the preceding presented MCMC algorithm was adopted to calculate Bayesian estimates of parameters and $\mathcal{M} = 5000$ observations after 5000 burn-in iterations. For the P-splines approximation of $f(Z(\boldsymbol{s}))$, we fixed the total number of knots to be $n = n_0 + 2n_1 + 1 = 30$, where n_0 is the number of intervals for dividing the domain of $x_{\min}$ and $x_{\max}$ into n_0 intervals, $n_1 = 3$ is the degree of B-splines. Thus, the number of B-splines K was taken to be $K = n_0 + n_1 = 26$. We selected the second order random walk (i.e., $d = 2$) for specifying the prior of β_k's. The selection of the hyperparameters was given in Appendix. For comparison, we also calculated Bayesian estimates of parameters via the Box-Cox transformation, i.e., $f(Z) = (Z^\rho - 1)/\rho$ for $\rho \neq 0$ and $f(Z) = \log(Z)$ for $\rho = 0$, under the following two cases: (i) $\rho = 1.0$ and (ii) the optimal value of ρ, which is 0.303, 0.343 and 0.788 for Types A, B and C, respectively. Results for 100 replications were given in Table 1, where "Bias" was the difference between the true value and the mean of the estimates based on 100 replications, and "RMS" was the root mean square between the estimates based on 100 replications and its true value. *The average mean squared prediction errors* (MSPEs) of Y for 100 replications via P-splines and Box-Cox transformation with the optimal value of ρ were presented in Table 2.

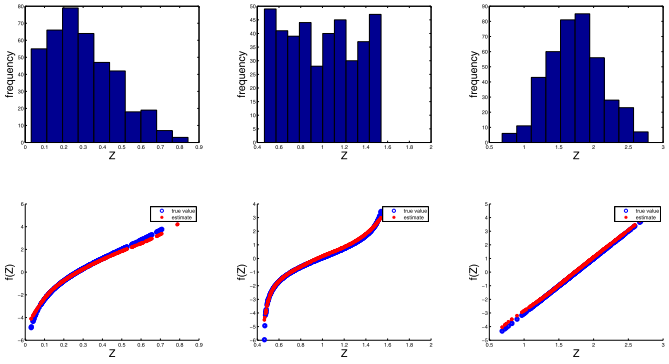

Figure 3. The histogram of $\boldsymbol{Z}$ (upper panel), estimated transformation function via P-spline of $f(Z)$ (lower panel) for type A (left), type B (middle) and type C (right) for a randomly selected replication in the first simulation study.

Examination of Tables 1 and 2 indicated that the above proposed Bayesian estimation procedure performed better than the Box-Cox transformation method, and Box-Cox transformation method with the optimal value of ρ was better than that with a particular value for ρ.

Figure 3 plotted the true and estimated curves of $f(Z)$, Figure 4 plotted the true process of Y and its predicted processes via the P-splines, and Figure 5 plotted the true and estimated covariance matrices of $\boldsymbol{\eta}$ for a randomly selected replication. Inspection of Figures 3 and 4 implied that the preceding proposed Bayesian P-splines method was an effective non-parameter method for estimating unknown transformation function. Examination of Figure 5 showed that the above presented Bayesian method behaved well in estimating covariance matrix of $\boldsymbol{\eta}$.

To illustrate the presented goodness-of-fit statistic, we calculated PP p-value for Types A, B and C, which were 0.435, 0.448 and 0.358 for a randomly selected replication, respectively, which showed that the considered model can fit the generated data well as expected.

To illustrate Bayes factor for comparing two competing spatial mixed models associated with random effects, we conducted the second simulation study. In this simulation study, we considered the following two competing models:

$\mathcal{H}_0 : f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top \boldsymbol{\gamma} + \nu(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$,

$\mathcal{H}_1 : f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top \boldsymbol{\gamma} + \nu(\boldsymbol{s}) + \xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$.

We simulated the dataset on the basis of the same setting as given in the first simulation study. Thus, $\mathcal{H}_1$ represented the true model. Similar to Lee and Tang [29], $\mathcal{H}_0$ and $\mathcal{H}_1$ was linked by $\mathcal{H}_{t01}$: $f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top \boldsymbol{\gamma} + \nu(\boldsymbol{s}) + t\xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$, where $t \in [0, 1]$. Clearly, $\mathcal{H}_{t01}$ was equal to $\mathcal{H}_0$ when $t = 0$, and $\mathcal{H}_{t01}$ was equal to $\mathcal{H}_1$ when $t = 1$. To calculate logarithm Bayes factor via the path sampling procedure, we took $\hbar = 10$ and $\mathcal{M} = 5000$ after 5000 burn-in iterations in evaluating $\bar{T}_{(\ell)}$ for $\ell = 0, 1, \ldots, \hbar + 1$. The estimated logarithm Bayes factors corresponding to three transformation functions (i.e., Types A, B and C) were $\widehat{\log B_{10}}$=10.186, 4.949, and 3.939,

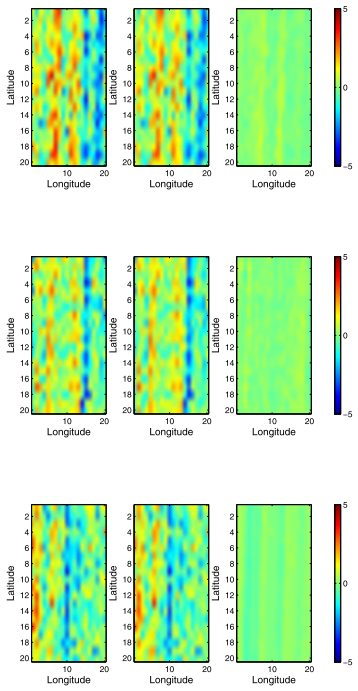

Figure 4. True process of Y (left), predicted process of $\widehat{Y}$ with P-spline (middle) and predicted bias of Y (right) for type A (upper), type B (middle) and type C (lower panel) for a randomly selected replication in the first simulation study.

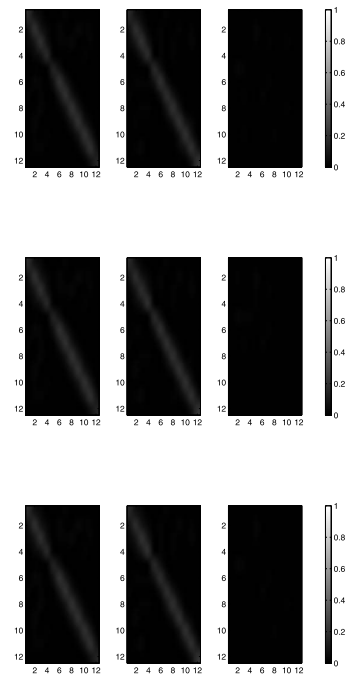

Figure 5. True covariance matrix K of $\boldsymbol{\eta}$ (left), estimated $\widehat{K}$ (middle) and predicted bias of K (right) for type A (upper), type B (middle) and type C (lower panel) in the first simulation study.

respectively. Following Kass and Raftery [24], the true model $\mathcal{H}_1$ was selected regardless of transformation functions as expected, which implied that the preceding proposed model comparison procedure is quite effective.

To illustrate Bayes factor in selecting the number of spatial basis functions or the number of resolution levels, we conducted the third simulation study. In this simulation study, we considered the following two competing models: $\mathcal{H}_{w0} : f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top\boldsymbol{\gamma} + \boldsymbol{W}^{(1)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(1)} + \boldsymbol{W}^{(2)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(2)} + \xi(\boldsymbol{s})+\varepsilon(\boldsymbol{s})$, $\mathcal{H}_{w1} : f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top\boldsymbol{\gamma}+\boldsymbol{W}^{(1)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(1)}+\xi(\boldsymbol{s})+\varepsilon(\boldsymbol{s})$, where $\boldsymbol{W}^{(1)}(\boldsymbol{s}) = (\boldsymbol{w}_1(\boldsymbol{s}),\ldots,\boldsymbol{w}_4(\boldsymbol{s}))^\top$, $\boldsymbol{W}^{(2)}(\boldsymbol{s}) = (\boldsymbol{w}_5(\boldsymbol{s}),\ldots,\boldsymbol{w}_{12}(\boldsymbol{s}))^\top$, $\boldsymbol{\eta}^{(1)} = (\eta_1,\ldots,\eta_4)^\top$, and $\boldsymbol{\eta}^{(2)} = (\eta_5,\ldots,\eta_{12})^\top$. Thus, $\mathcal{H}_{w0}$ was the true model. Similarly, $\mathcal{H}_{w0}$ and $\mathcal{H}_{w1}$ was linked by $\mathcal{H}_{tw01}$: $f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top\boldsymbol{\gamma} + \boldsymbol{W}^{(1)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(1)} + (1-t)\boldsymbol{W}^{(2)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(2)} + \xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$, where $t \in [0,1]$. Clearly, $\mathcal{H}_{tw01}$ was equal to $\mathcal{H}_{w0}$ when $t = 0$, and $\mathcal{H}_{tw01}$ was equal to $\mathcal{H}_{w1}$ when $t = 1$. We generated the dataset from the model $\mathcal{H}_{w0}$ on the basis of the same setting as given in the first simulation study. The preceding presented path sampling approach with $\hbar = 10$ and $\mathcal{M} = 5000$ after 5000 burn-in iterations was employed to calculate Bayes factors for each of three transformation functions: Types A, B and C. The estimated logarithm Bayes factors corresponding to Types A, B and C transformation functions were $\widehat{\log B_{10}}$=-202.892, -180.995 and -152.184, respectively, which showed that the true model $\mathcal{H}_{w0}$ was selected regardless of transformation functions as expected.

4.2 An example

In this subsection, the remote sensing dataset described in Introduction was used to illustrate the above presented TSMMs together with Bayesian method. As an illustration, we only fitted a small study region $\mathbb{S}$ between longitudes +30.25 and +15.75, and between latitudes -30.25 and -15.75 to the considered TSMMs. There were some missing data in $\mathbb{S}$, whose missing proportion was 2.56%. For simplicity, we deleted these missing data. We selected three covariates WV, NR and CM to construct a TSMMs for AOD, i.e., in the considered TSMMs, $Z(\boldsymbol{s}) = \mathrm{AOD}(\boldsymbol{s})$, and $\boldsymbol{X}(\boldsymbol{s}) = (\mathrm{CM}(\boldsymbol{s}), \mathrm{NR}(\boldsymbol{s}), \mathrm{WV}(\boldsymbol{s}))$. Because the units of covariates and response were inconsistent, the raw data were standardized on the basis of the fully observed data. In this case, $N = 30 \times 30 = 900$ in $\mathbb{S}$.

The above developed Bayes factor together with the path sampling method was used to select the number of the spatial basis functions. Thus, we considered the following two competing models: $\mathcal{H}_{w0} : f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top\boldsymbol{\gamma} + \boldsymbol{W}^{(1)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(1)} + \boldsymbol{W}^{(2)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(2)} + \xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$, $\mathcal{H}_{w1} : f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top\boldsymbol{\gamma} + \boldsymbol{W}^{(1)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(1)} + \xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$, where $\boldsymbol{\eta}^{(1)}$ and $\boldsymbol{\eta}^{(2)}$ are 4×1 and 8×1 vectors of random effects corresponding to multi-resolutional spatial basis functions $\boldsymbol{W}^{(1)}(\boldsymbol{s})$ and $\boldsymbol{W}^{(2)}(\boldsymbol{s})$, respectively. Similarly, $\mathcal{H}_{w0}$ and $\mathcal{H}_{w1}$ was linked by $\mathcal{H}_{tw01}$: $f(Z(\boldsymbol{s})) = \boldsymbol{X}(\boldsymbol{s})^\top\boldsymbol{\gamma} + \boldsymbol{W}^{(1)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(1)} + (1-t)\boldsymbol{W}^{(2)}(\boldsymbol{s})^\top\boldsymbol{\eta}^{(2)} + \xi(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$, where $t \in [0,1]$. In this case,

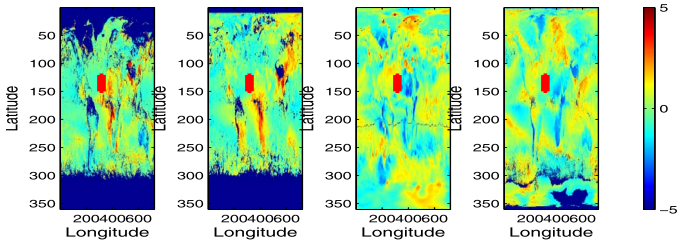

Figure 6. Images of AOD, CM, NR and WV datasets in which the studied datasets are showed in red region.

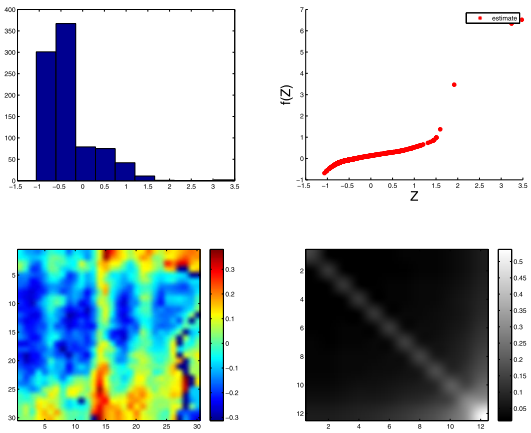

Figure 7. The hist of $\boldsymbol{Z}$ (first), estimated transformation function via P-spline of f (second),predicted process of $\widehat{Y}$ with P-spline (third), estimated covariance matrix K of $\boldsymbol{\eta}$ (fourth).

$\mathcal{H}_{tw01}$ reduced to $\mathcal{H}_{w0}$ when $t = 0$, and $\mathcal{H}_{tw01}$ was $\mathcal{H}_{w1}$ when $t = 1$. Similar to the third simulation study, we took $L = 2$, $\hbar = 10$ and $\mathcal{M} = 5000$ observations after 5000 burn-in iterations in computing $\overline{\top}_{(\ell)}$ for $\ell = 0, 1, \ldots, \hbar + 1$. The estimated logarithm Bayes factor was $\widehat{\log B_{10}} = -12.741$, which showed that we should select the model $\mathcal{H}_{w0}$ to fit the considered dataset, i.e., $q = 12$.

Based on the selected model $\mathcal{H}_{w0}$, we obtained Bayesian estimates of $\boldsymbol{\gamma}$: $\hat{\boldsymbol{\gamma}} = (-0.2006, -0.183, 0.248)$, which showed that CM and NR had a negative effect on AOD, while WV had a positive effect on AOD. Also, the PP p-values was 0.312, which showed that the selected model fitted the data well. Figure 7 showed that the histogram of the considered dataset, the estimated transformation function of $f(Z)$, the predicted process of $\widehat{Y}$, and the estimated covariance matrix K of $\boldsymbol{\eta}$.

APPENDIX A

A.1 Specification of hyperparameters

Generally, one takes $e_2 = 1$ and a small value of e_1 such as $e_1 = 0.005$ leading to an almost diffuse prior for τ^2. Also, we set $b = 1$. Following Kang and Cressie [23],we take $\boldsymbol{\mu}_{\gamma 0} = \mathbf{0}$ and $\sigma^2_{\gamma 0} = 0.25$, $\widehat{\sigma}^2_\zeta$ as an initial estimate of σ^2_ζ, $\kappa_\zeta = \varsigma \widehat{\sigma}^2_\zeta$ in which $\varsigma = 8$ in simulation studies and an example analysis. For the prior of $\boldsymbol{\Sigma}$, we choose $v_{0i} = 8, u_{0i} = 1, \boldsymbol{a}_{j0} = 1$, $\boldsymbol{A}_{j0} = 0.25\boldsymbol{I}$ in which $\boldsymbol{I}$ is an identity matrix.

A.2 MCMC algorithm

Step (a). The conditional distribution for $\boldsymbol{\gamma}$ is given by

$$p(\boldsymbol{\gamma}|\boldsymbol{\xi}, \boldsymbol{\eta}, \sigma_\xi^{-2}, \boldsymbol{\beta}, \mathcal{D}) \sim \mathcal{N}_p(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma),$$

where $\boldsymbol{\mu}_\gamma = \boldsymbol{\Sigma}_\gamma[\sigma_\varepsilon^{-2}\boldsymbol{X}^\top\{\sum_{k=1}^K \beta_k \boldsymbol{B}_k(\boldsymbol{Z}) - \boldsymbol{W}\boldsymbol{\eta} - \boldsymbol{\xi}\} + \sigma_{\gamma 0}^{-2}\boldsymbol{\mu}_{\gamma 0}]$, $\boldsymbol{\Sigma}_\gamma = (\boldsymbol{X}^\top\boldsymbol{X}/\sigma^2_\varepsilon + \sigma^{-2}_{\gamma 0}\boldsymbol{I}_p)^{-1}$, $\boldsymbol{X} = (\boldsymbol{X}(s_1), \ldots, \boldsymbol{X}(s_N))^\top$, $B_k(\boldsymbol{Z}) = (B_k(Z(s_1)), \ldots, B_k(Z(s_N)))^\top$, and $\boldsymbol{W} = (\boldsymbol{W}(s_1), \ldots, \boldsymbol{W}(s_N))^\top$.

Step (b). The conditional distribution for $\boldsymbol{\eta}$ is given by

$$p(\boldsymbol{\eta}|\boldsymbol{\gamma}, \boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{\lambda}, \mathcal{D}) \sim \mathcal{N}_q(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta),$$

where $\boldsymbol{\mu}_\eta = \boldsymbol{\Sigma}_\eta \boldsymbol{W}\{\sum_{k=1}^K \beta_k B_k(\boldsymbol{Z}) - \boldsymbol{X}\boldsymbol{\gamma} - \boldsymbol{\xi}\}/\sigma^2_\varepsilon$, and $\boldsymbol{\Sigma}_\eta = (\sigma_\varepsilon^{-2}\boldsymbol{W}\boldsymbol{W}^\top + \boldsymbol{\Sigma}^{-1})^{-1}$.

Step (c). The conditional distribution for $\boldsymbol{\xi}$ is given by

$$p(\boldsymbol{\xi}|\boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma_\xi^{-2}, \boldsymbol{\beta}, \mathcal{D}) \sim \mathcal{N}(\boldsymbol{\mu}_\xi, \boldsymbol{\Sigma}_\xi),$$

where $\boldsymbol{\mu}_\xi = \boldsymbol{\Sigma}_\xi\{\sum_{k=1}^K \beta_k B_k(\boldsymbol{Z}) - \boldsymbol{X}\boldsymbol{\gamma} - \boldsymbol{W}\boldsymbol{\eta}\}/\sigma^2_\varepsilon$, and $\boldsymbol{\Sigma}_\xi = (\sigma_\varepsilon^{-2} + \sigma_\xi^{-2})^{-1}\boldsymbol{I}$.

Step (d). It is easily shown from the prior distribution of $\boldsymbol{\beta}$ and the distribution of $\boldsymbol{Z}(s)$ that the full conditional distribution $\boldsymbol{\beta}$ can be expressed as

$$p(\boldsymbol{\beta}|\boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma_\zeta^{-2}, \boldsymbol{\varphi}, \mathcal{D}) \propto \exp\left\{-\tfrac{1}{2}(\boldsymbol{\beta} - \boldsymbol{m})^\top \boldsymbol{P}(\boldsymbol{\beta} - \boldsymbol{m})\right\} \times \prod_{k=2}^K I(\beta_k \geq \beta_{k-1}) I(\boldsymbol{Q}\boldsymbol{\beta} = 0),$$

where $\boldsymbol{P} = \boldsymbol{B}\boldsymbol{B}^\top/\sigma^2_\zeta + \boldsymbol{M}(\boldsymbol{\varphi})/\tau^2$ with $\sigma^2_\zeta = \sigma^2_\varepsilon + \sigma^2_\xi$, $\boldsymbol{B} = (B_1(\boldsymbol{Z}), \ldots, B_K(\boldsymbol{Z}))^\top$, and $\boldsymbol{m} = \boldsymbol{P}^{-1}\boldsymbol{B}(\boldsymbol{X}\boldsymbol{\gamma} + \boldsymbol{W}\boldsymbol{\eta})$. Clearly, the conditional distribution $p(\boldsymbol{\beta}|\cdot)$ is not a familiar distribution. Thus, it is impossible to directly sample observation from the conditional distribution $p(\boldsymbol{\beta}|\cdot)$. To this end, the Gibbs sampler is again adopted to sample each component β_k of $\boldsymbol{\beta}$ as follows.

For $t = 0, 1, 2, \ldots$, at the current value $\boldsymbol{\beta}^{(t)} = (\beta_1^{(t)}, \ldots, \beta_K^{(t)})^\top$ of $\boldsymbol{\beta}$ with the constrained conditions $\beta_k^{(t)} \geq \beta_{k-1}^{(t)}$ for $k = 2, \ldots, K$, the observations $\beta_1^{(t+1)}, \ldots, \beta_K^{(t+1)}$ are iteratively drawn by

(1) sampling $\beta_1^{(t+1)}$ from $\mathrm{TN}(\mu_{\beta_1}, \varphi^2_{\beta_1}, -\infty, \beta_2^{(t)})$,
(2) sampling $\beta_2^{(t+1)}$ from $\mathrm{TN}(\mu_{\beta_2}, \varphi^2_{\beta_2}, \beta_1^{(t+1)}, \beta_3^{(t)})$,
(3) sampling $\beta_3^{(t+1)}$ from $\mathrm{TN}(\mu_{\beta_3}, \varphi^2_{\beta_3}, \beta_2^{(t+1)}, \beta_4^{(t)})$,
$\vdots$
(K) sampling $\beta_K^{(t+1)}$ from $TN(\mu_{\beta_K}, \varphi^2_{\beta_K}, \beta_{K-1}^{(t+1)}, \infty)$,

where $\mathrm{TN}(\mu_d, \varphi^2_d, d_1, d_2)$ represents the one-dimensional

interval truncated Gaussian distribution in the interval (d_1, d_2) with mean μ_d and variance φ_d^2, $\mu_{\beta_k} = m_k + \boldsymbol{P}_{[k]}^\top \boldsymbol{P}_{(k)}^{-1}(\boldsymbol{\beta}_{[k]}^{(t)} - \boldsymbol{m}_{[k]})$ and $\varphi_{\beta_k}^2 = P_{kk} - \boldsymbol{P}_{[k]}^\top \boldsymbol{P}_{(k)}^{-1}\boldsymbol{P}_{[k]}$ for $k = 1, \ldots, K$, where P_{kk} is the kth diagonal element of matrix $\boldsymbol{P}$, $\boldsymbol{P}_{(k)}$ is a $(K-1)\times(K-1)$ submatrix of matrix $\boldsymbol{P}$ with the components corresponding to the kth row and the kth column of matrix $\boldsymbol{P}$ deleted, and $\boldsymbol{P}_{[k]}$ is a $(K-1)\times 1$ vector with the kth component in the kth column of matrix $\boldsymbol{P}$ deleted, $\boldsymbol{m}_{[k]}$ and $\boldsymbol{\beta}_{[k]}$ are the $(K-1)\times 1$ vectors with the kth components of vectors $\boldsymbol{m}$ and $\boldsymbol{\beta}$ deleted, respectively. For identifiability, the function $f(\cdot)$ satisfies the following restriction condition: $\boldsymbol{Q}\boldsymbol{\beta} = 0$, which can be implemented by centering the function $f(\cdot)$ on its mean at each of iterations of the Gibbs sampler when evaluating estimates of parameters via the above presented MCMC.

Step (e). The conditional distribution for σ_ξ^2 is given by

$$p(\sigma_\xi^2|\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\beta}, \mathcal{D}) \sim \mathrm{IG}(a_{\sigma_\xi^2}, b_{\sigma_\xi^2}) I\{\sigma_\xi \in (0, \kappa_\xi)\},$$

where $a_{\sigma_\xi^2} = (N-1)/2$, $b_{\sigma_\xi^2} = (\boldsymbol{B}^\top\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\gamma} - \boldsymbol{W}\boldsymbol{\eta})^\top(\boldsymbol{B}^\top\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\gamma} - \boldsymbol{W}\boldsymbol{\eta})/2$, and $\mathrm{IG}(\cdot,\cdot)$ represents the inverse-gamma distribution.

Step(f). The conditional distribution $p(\tau^{-2}|\boldsymbol{\beta}, \boldsymbol{\varphi})$ is given by

$$p(\tau^{-2}|\boldsymbol{\beta}, \boldsymbol{\varphi}) \sim \Gamma\left(e_1 + \frac{K-d}{2}, e_2 + \frac{1}{2}\boldsymbol{\beta}^\top \boldsymbol{M}(\varphi)\boldsymbol{\beta}\right),$$

where d is the order of random walk.

Step(g). The conditional distribution $p(\varphi_k|\boldsymbol{\beta}, \tau^{-2})$ is given by

$$p(\varphi_k|\boldsymbol{\beta}, \tau^{-2}) \sim \Gamma\left(\frac{b+1}{2}, \frac{b + u_k^2/\tau^2}{2}\right),$$

where u_k is equal to $\beta_k - \beta_{k-1}$ and $\beta_k - 2\beta_{k-1} + \beta_{k-2}$ for the fist and second order random walk, respectively.

Step (h). To obtain conditional distribution of $\boldsymbol{\lambda}$, it follows from $\boldsymbol{\omega} = \boldsymbol{A}\boldsymbol{\eta}$ and $|\boldsymbol{\Sigma}|^{-1} = |\boldsymbol{A}|^\top|\boldsymbol{\Lambda}|^{-1}|\boldsymbol{A}| = |\boldsymbol{\Lambda}^{-1}| = \prod_{i=1}^q \lambda_i^{-1}$ that the likelihood can be written as

$$\begin{aligned}
\ell(\boldsymbol{\eta}|\boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\eta}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}\right\} \\
&= \prod_{i=1}^q \lambda_i^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\omega}^\top\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega}\right\} \\
&= \prod_{i=1}^q \lambda_i^{-1/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega}\boldsymbol{\omega}^\top)\right\} \\
&= \prod_{i=1}^q \lambda_i^{-1/2}\exp\left\{-\frac{s_i}{2\lambda_i}\right\}
\end{aligned}$$

The conditional distribution for the ith component of $\boldsymbol{\lambda}$ is given by

$$p(\lambda_i^{-1}|\boldsymbol{\eta}, \boldsymbol{a}) \sim \Gamma\left(\frac{v_{0i}+1}{2}, \frac{u_{0i}+s_i}{2}\right),$$

where s_i is the ith diagonal element of matrix $\boldsymbol{\omega}\boldsymbol{\omega}^\top$ for $i = 1, \ldots, q$.

Step (i). Consider the conditional distribution of $\boldsymbol{a}$. Note that

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{\eta} &= \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ a_{21} & 1 & 0 & \ldots & 0 & 0 \\ a_{31} & a_{32} & 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{q1} & a_{q2} & \ldots & \ldots & a_{q,q-1} & 1 \end{pmatrix}\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_q \end{pmatrix} \\
&= \begin{pmatrix} \eta_1 \\ \eta_2 + a_{21}\eta_1 \\ \eta_3 + a_{31}\eta_1 + a_{32}\eta_2 \\ \vdots \\ \eta_q + \sum_{i=1}^{q-1} a_{qi}\eta_i \end{pmatrix}.
\end{aligned} \tag{6}$$

Thus, we can rewrite the likelihood function as

$$\begin{aligned}
\ell(\boldsymbol{\eta}|\boldsymbol{\Sigma}) &\propto \prod_{i=1}^q \lambda_i^{-1/2}\exp\{-\tfrac{1}{2}(\boldsymbol{A}\boldsymbol{\eta})^\top\boldsymbol{\Lambda}^{-1}(\boldsymbol{A}\boldsymbol{\eta})\} \\
&= \prod_{i=1}^q \lambda_i^{-1/2}\exp\{-\tfrac{\eta_1^2}{2\lambda_1}\}\exp\{-\tfrac{(\eta_2+a_{21}\eta_1)^2}{2\lambda_2}\} \\
&\quad \times \exp\{-\frac{(\eta_q + \sum_{i=1}^{q-1} a_{qi}\eta_i)^2}{2\lambda_q}\},
\end{aligned}$$

which leads to

$$\ell(\boldsymbol{\eta}|\boldsymbol{\Sigma}) \propto \prod_{i=1}^q \lambda_i^{-1/2}\exp\{-\frac{1}{2}(\boldsymbol{\eta} - \nabla\boldsymbol{a})^\top\boldsymbol{\Lambda}^{-1}(\boldsymbol{\eta} - \nabla\boldsymbol{a})\},$$

where

$$\boldsymbol{\nabla} = -\begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ \eta_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \eta_1 & \eta_2 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \eta_1 & \ldots & \eta_{q-1} \end{pmatrix}.$$

Therefore, the conditional distribution for $\boldsymbol{a}$ is given by

$$p(\boldsymbol{a}|\boldsymbol{\eta}, \boldsymbol{\Lambda}) \sim \mathcal{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a),$$

where $\boldsymbol{\Sigma}_a = (\boldsymbol{A}_0^{-1} + \boldsymbol{\nabla}\boldsymbol{\Lambda}^{-1}\boldsymbol{\nabla})^{-1}$, and $\boldsymbol{\mu}_a = \boldsymbol{\Sigma}_a(\boldsymbol{A}_0^{-1}\boldsymbol{a}_0 + \boldsymbol{\nabla}\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta})$. Because $\boldsymbol{\Lambda}$ is a diagonal matrix, when $\boldsymbol{A}_0$ is also a diagonal or block-diagonal matrix corresponding to rows of $\boldsymbol{A}$, the derivations are simplified and the elements of $\boldsymbol{a}$ can be updated in a series of independent steps

$$p(\boldsymbol{a}_i|\boldsymbol{\eta}, \boldsymbol{\Lambda}_i) \sim \mathcal{N}(\boldsymbol{\mu}_{ai}, \boldsymbol{\Sigma}_{ai}),$$

where $\boldsymbol{\Sigma}_{ai} = (\boldsymbol{A}_{i0}^{-1} + \lambda_i^{-1}\boldsymbol{\Omega}_i^\top\boldsymbol{\Omega}_i)^{-1}$, $\boldsymbol{\mu}_{ai} = \boldsymbol{\Sigma}_{ai}(\boldsymbol{A}_{i0}^{-1}\boldsymbol{a}_{i0} - \lambda_i^{-1}\boldsymbol{\Omega}_i^\top\eta_i)$, $\boldsymbol{\Omega}_i = (\eta_1, \ldots, \eta_{i-1})$ for $i = 2, \ldots, q$. Note that $\boldsymbol{A}_{i0}$ may depend on λ_i. Given the posterior draws $\boldsymbol{a}$ and $\boldsymbol{\lambda}$, a posterior draw of $\boldsymbol{\Sigma}$ can be obtained by $\boldsymbol{\Sigma} = \boldsymbol{A}^{-1}\boldsymbol{\Lambda}\boldsymbol{A}^{-\top}$.

The conditional distributions corresponding to steps (a)-(i) are some familiar distributions, thus sampling observations form these distributions is straightforward.

A.3 W-matrix

A W-matrix with even order has the form of

$$\boldsymbol{W} = \begin{pmatrix} 2 & 3 & -1 & & & & & & & \\ 2 & -3 & 1 & & & & & & & \\ & -1 & 3 & 3 & -1 & & & & & \\ & -1 & 3 & -3 & 1 & & & & & \\ & & & & \ddots & & & & & \\ & & & & & -1 & 3 & 3 & -1 & \\ & & & & & -1 & 3 & -3 & 1 & \\ & & & & & & & -1 & 3 & 2 \\ & & & & & & & -1 & 3 & -2 \end{pmatrix}.$$

A W-matrix with odd order has the form of

$$\boldsymbol{W} = \begin{pmatrix} 2 & 3 & -1 & & & & & & & \\ 2 & -3 & 1 & & & & & & & \\ & -1 & 3 & 3 & -1 & & & & & \\ & -1 & 3 & -3 & 1 & & & & & \\ & & & & \ddots & & & & & \\ & & & & & -1 & 3 & 3 & -1 & \\ & & & & & -1 & 3 & -3 & 1 & \\ & & & & & & & -1 & 5 & \end{pmatrix}.$$

Although the inverse matrix $\boldsymbol{W}^{-1}$ has a structure similar to that of $\boldsymbol{W}$, it is more appropriate to think of $\boldsymbol{W}^{-1}$ as being built by columns instead of rows. The building blocks of $\boldsymbol{W}^{-1}$ are the $\bar{g}$-vector $(1,3,3,1)$ and $\bar{h}$-vector $(1,3,-3,-1)$, which are dual to the g- and h-vectors of $\boldsymbol{W}$, respectively. The columns of the matrix come in pairs, made up of the transpose of the above two basic vectors; successive pairs of columns are shifted two positions downwards. Following Kwong and Tang [26], the W-matrix is associated with the wavelet theory. In particular, the following propositions hold.

Proposition 1. If there is a scaling function $\phi(\cdot)$ corresponding to a W-matrix, $\phi(\cdot)$ satisfies the dilation equation

$$\phi(x) = 2\sum_{i=1}^{n} \bar{\omega}_i \phi(2x - i + 1),$$

where $\bar{\omega}_i = \bar{g}_i / \sum_{j=1}^{n} \bar{g}_j$, $\bar{\boldsymbol{g}} = (\bar{g}_1, \ldots, \bar{g}_n)^\top$ is the first basis vector of the inverse W-matrix. In particular, ϕ can be taken as the classical quadratic spline.

Proposition 2. For $\bar{h}$-vector, we have

$$\phi(x) = C \sum \bar{h}_i \phi(2x - i + 1),$$

where $\bar{\boldsymbol{h}} = (\bar{h}_1, \ldots, \bar{h}_n)^\top$ is the second basis vector of the inverse W-matrix, and C is some scaling constant.

ACKNOWLEDGEMENTS

The authors are grateful to the Associate Editor and a referee for their valuable suggestions that greatly improved the paper.

SUPPLEMENTARY MATERIAL

Supplementary materials available include Matlab codes, http://intlpress.com/site/pub/files/_supp/sii/2019/0012/0004/SII-2019-0012-0004-s001.zip.

Received 15 August 2018

REFERENCES

[1] Anselin, L. (1992). *SpaceStat, a Software Program for Analysis of Spatial Data*. National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara, CA.

[2] Anselin, L. (2004). *GeoDa 0.95i Release Notes*. Spatial Analysis Laboratory (SAL). Department of Argicultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL.

[3] Anselin, L. and Bera, A. (1998). *Spatial dependenece in linear regression models with an introduction to spatial econometrics*. In: Ullah, A. and Giles, D. E. (eds), *Handbook of Applied Economic Statistics*, pages 237–289. Marcel Dekker, New York.

[4] Anselin, L., Syabri, I. and Kho, Y. (2005). GeoDa: an introduction to spatial data analysis. *Geographical Analysis* **38**, 5–22.

[5] Bai, Z. D. (2012). Large Dimensional Statistical Analysis. *Beijing: Higher Education Press of China*.

[6] 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 **70**, 825–848. MR2523906

[7] Box, G. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society* Series B **26**, 211–246. MR0192611

[8] Brezger, A. and Steiner, W. (2008). Monotonic regression based on Bayesian P-splines: an application to estimating price response functions from store-level scanner data. *Journal of Business Economic Statistics* **26**, 90–104. MR2422064

[9] Cai, T. T. and Zhang, A. (2016). Inference for high-dimensional differential correlation matrices. *Journal of Multivariate Analysis* **143**, 107–126. MR3431422

[10] Chan, J. C. C. and Jeliazkov, I. (2009). MCMC estimation of restricted covariance matrix. *Journal of Computational and Graphical Statistics* **18**, 457–480. MR2749841

[11] Cressie, N. (1993). *Statistics for Spatial Data (Revised Ed.)*. New York, Wiley. MR1239641

[12] Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large data sets. *Journal of the Royal Statistical Society* Series B **70**, 209–226. MR2412639

[13] Cressie, N. and Kang, E. (2010). *High-Resolution Digital Soil Mapping: Kriging for very large datasets*. In: Viscarra Rossel, R., McBratney, A., and Minasny, B. (eds), Proximal Soil Sensing. Springer, Dordrecht, NL. chapter 4, 49–63.

[14] De Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer-Verlag. MR0507062

[15] De Oliveira, V., Kedem, B. and Short, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association* **92**, 1422–1433. MR1615252

[16] Eilers, P. H. C. and Marx, B. D. (1996). Flexible soomthing using B-splines and penalized. *Statistical Science* **11**, 89–121. MR1435485

[17] Gelman, A. (1996). *Inference and Monitoring Convergence*. In: Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (eds), Markov Chain Monte Carlo in Practice, Chapman and Hall. MR1397966

[18] Gelman, A. and Meng, X. L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science* **13**, 163–185. MR1647507

[19] Genton, M. G. and Zhang, H. (2012). Identifiability problems in some non-Gaussian spatial random fields. *Chilean Journal of Statistics* **3**, 171–179. MR2982069

[20] Guinness, J. (2017). Spectral density estimation for random fields via periodic embeddings. *Journal of Computational and Graphical Statistics* **24**, 579–599.
[21] Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *American Statistician* **60**, 19–26. MR2224133
[22] Kang, E. L. (2009). *Reduced-Dimension Hierarchical Statistical Models for Spatial and Spatio-Temporal Data*. Ph.D. thesis, The Ohio State University, Columbus, OH.
[23] Kang, E. L. and Cressie, N. (2011). Bayesian Inference for the Spatial Random Effects Model. *Journal of the American Statistical Association* **106**, 972–983. MR2894757
[24] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795. MR3363402
[25] Katzfuss, M. (2015). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association* **112**, 201–204. MR3646566
[26] Kwong, M. K. and Tang, P. T. P. (1994). W-matrices, nonorthogonal multiresolution analysis, and finite signals of arbitrary length. Technical report, Argonne National Laboratory, 449-0794.
[27] Lang, S. and Brezger, S. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13**, 183–212. MR2044877
[28] Lee, S. Y. and Song, X. Y. (2003). Model comparison of nonlinear structural equation models with fixed covariates. *Psychometrika* **68**, 27–47. MR2272368
[29] Lee, S. Y. and Tang, N. S. (2006). Bayesian analysis of nonlinear structural equation models with nonignorable missing data. *Psychometrika* **71**, 541–564. MR2272542
[30] Meng, X. L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* **6**, 831–860. MR1422406
[31] Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F. and Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics* **24**, 579–599. MR3357396
[32] Nychka, D., Wikle, C. and Royle, J. A. (2002). Multiresolution Models for Nonstationary Spatial Covariance Functions. *Statistical Modeling* **2**, 315–331. MR1951588
[33] Robert, C. P., (1995). Simulation of truncated normal variables. *Statistics and Computing* **5**, 121–125.
[34] Shi, T. and Cressie, N. (2007). Global statistical analysis of MISR aerosol data: a massive data product from NASA's terra satellite. *Environmetrics* **18**, 665–680. MR2408937
[35] Song, X. Y. and Liu, P. F. (2015). Transformation structural equation models with highly non-normal and incomplete Data. *Structural Equation Modeling: A Multidisciplinary Journal* **22**, 401–415. MR3357141
[36] Song, X. Y. and Lu, Z. H. (2012). Semi-parametric transformation models with Bayesian P-splines. *Statistics and Computing* **22**, 1085–1098. MR2950087
[37] Tang, N., Wu, Y. and Chen, D. (2018). Semiparametric Bayesian analysis of transformation linear mixed models. *Journal of Multivariate Analysis* **166**, 225–240. MR3799645
[38] Wikle, C. K., Milliff, R., Nychka, D. W. and Berliner, L. M. (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. Technical report, Argonne National Laboratory, 449-0794. MR1939342
[39] Xu, B., Wikle, C. K. and Fox, N. (2005). A kernel-based spatiotemporal dynamical model for nowcasting radar precipitation. *Journal of the American Statistical Association* **100**, 1133–1144. MR2236929
[40] Zareifard, H. and Khaledi, M. J. (2013). Non-Gaussian modeling of spatial data using scale mixing of a unified skew Gaussian process. *Journal of Multivariate Analysis* **114**, 16–28. MR2993870
[41] Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics* **21**, 33–47. MR2842222
[42] Zhu, H. T., Ibrahim, J. G., Chi, Y. Y. and Tang, N. S. (2012). Bayesian influence measures for joint models for longitudinal and survival data. *Biometrics* **68**, 954–964. MR3055200
[43] Gelman, A., Meng, X.L., Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistical Sinical* **6**, 733–807. MR1422404

Ying Wu
Yunnan Key Laboratory of Statistical Modeling and Data Analysis
Yunnan University, Kunming 650091
P. R. of China
E-mail address: wy@cxtc.edu.cn

Dan Chen
Yunnan Key Laboratory of Statistical Modeling and Data Analysis
Yunnan University, Kunming 650091
P. R. of China
E-mail address: danchen@ynu.edu.cn

Niansheng Tang
Yunnan Key Laboratory of Statistical Modeling and Data Analysis
Yunnan University, Kunming 650091
P. R. of China
E-mail address: nstang@ynu.edu.cn