

Fast Bayesian Functional Regression for Non-Gaussian Spatial Data*

Hyun Bin Kang[†], Yeo Jin Jung^{‡,¶} and Jaewoo Park^{§,||}

Abstract. Functional generalized linear models (FGLM) have been widely used to study the relations between non-Gaussian response and functional covariates. However, most existing works for FGLM assume independence among observations and therefore they are of limited applicability for correlated data. A particularly important example is spatial functional data, where we observe functions over spatial domains, such as the age population curve or temperature curve at each areal unit. In this paper, we extend FGLM by incorporating spatial random effects. Especially, we study the relationship between the non-Gaussian response variable and functional covariates that are spatially observed. However, such model has computational and inferential challenges. The high-dimensional spatial random effects cause the slow mixing of Markov chain Monte Carlo (MCMC) algorithms. Furthermore, spatial confounding can lead to bias in parameter estimates and inflate their variances. To address these issues, we propose an efficient Bayesian method using a sparse reparameterization of high-dimensional random effects. We also study the average coverage probabilities of the credible intervals of functional parameters. We apply our methods to simulated and real data examples, including malaria incidence data and US COVID-19 data. The proposed method is fast while providing accurate functional estimates.

Keywords: functional regression, non-Gaussian spatial data, Markov chain Monte Carlo, dimension reduction, Gaussian Markov random fields.

1 Introduction

With the rapid development of data collection technology, researchers in various disciplines face the challenge of extracting information from complex data, such as data that vary over a continuum (e.g., time and frequency). Examples include growth curves (Chen and Müller, 2012), temperature curves (Zhang and Chen, 2007), and credit card transaction volumes over time (Kokoszka and Reimherr, 2012). In order to understand such data that can be naturally viewed as curves or functions and are inherently infinite-dimensional, a field of statistics called *functional data analysis* (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012; Wang et al., 2016; Kokoszka and Reimherr, 2017) has been developed. Although standard functional data analysis methods are based on

*Jaewoo Park and Yeo Jin Jung were supported by the National Research Foundation of Korea (NRF-2020R1C1C1A0100386811).

[†]Department of Statistics, Western Michigan University, h.kang@wmich.edu

[‡]Department of Statistics, The University of Chicago, yeojinjung@uchicago.edu

[§]Department of Statistics and Data Science, Yonsei University.

[¶]Department of Applied Statistics, Yonsei University.

^{||}Department of Applied Statistics, Yonsei University, jwpark88@yonsei.ac.kr

the independence assumption between functions, there is an increasing interest in analyzing dependent functions. An important example is *spatial functional data analysis* (Delicado et al., 2010; Guillas and Lai, 2010; Ruiz-Medina, 2012; Martínez-Hernández et al., 2020), where we observe functional data over spatial locations such as spatio-temporal data.

In this manuscript, we provide a new methodology to model the relationship between non-Gaussian scalar response and functional covariates when each variable is observed on spatial locations. For example, we regress coronavirus disease 2019 (COVID-19) incidence (scalar response) on the age population curve (functional covariates) over US counties. Furthermore, our model can easily incorporate scalar covariates, such as the proportion of males in the same counties. It is an extension of previous works in both modeling and computational aspects. In terms of modeling, we provide a flexible regression framework for studying non-Gaussian spatial responses with functional and scalar covariates simultaneously. For computational aspects, we adapt dimension-reduced Markov chain Monte Carlo (MCMC) algorithm (Hughes and Haran, 2013) to carry out Bayesian inference for this new model quickly.

We use functional regression models when one of the variables of interest is functional. Functional regression models can be divided into the categories of 1) function-on-scalar models, which take the functional response and scalar covariate, 2) scalar-on-function models, which take the scalar response and functional covariate, and 3) function-on-function models, which take the functional response and functional covariate. Although there are ample studies on scalar-on-function models (Cardot et al., 1999; Cai et al., 2006; Hall et al., 2007; Crambes et al., 2009; James et al., 2009; Goldsmith and Scheipl, 2014; Morris, 2015; Reiss et al., 2017), the attempts to incorporate random effects in the framework of functional mixed models lie in function-on-scalar framework (Guo, 2002; Morris and Carroll, 2006; Antoniadis and Sapatinas, 2007; Zhu et al., 2011; Chen and Wang, 2011; Ma et al., 2021). Our proposed model, however, will be one of a few studies that include random effects in scalar-on-function models. Furthermore, our response variable is non-Gaussian. Functional generalized linear models (James, 2002; Müller and Stadtmüller, 2005; McLean et al., 2014) have been developed to study a non-Gaussian response variable in scalar-on-function framework. However, most studies concerning functional generalized regressions assume independent random errors. Examples include a generalized functional regression with image predictors (Reiss and Ogden, 2010), variable selection method for generalized functional regression (Gertheiss et al., 2013), functional generalized additive models (McLean et al., 2014), and estimation methods for functional logistic regressions (Mousavi and Sørensen, 2018).

In this manuscript, we provide functional generalized linear models with spatial random effects. Our Bayesian method is flexible in that it can apply to a wide variety of exponential family models, including Bernoulli, Poisson, zero-inflated Poisson, negative binomial, which is an important contribution of this paper. Note that there is abundant literature on the applications of functional logistic regression; however, there are relatively few studies on count responses. To the best of our knowledge, James (2002), and Müller and Stadtmüller (2005) provide theoretical foundations on the estimation and inference for functional generalized linear models with responses following exponential-

family distributions, but their examples only concern Bernoulli responses. More examples of functional logistic regressions include Alzheimer’s disease classification using PET images and Haar wavelet functions (Wang et al., 2017), variable selection for logistic regression by embedding the logistic model in RKHS (Bueno-Larraz et al., 2018), and comparison of the performance of three logistic regressions with different penalties (Mousavi and Sørensen, 2018), to name a few. But those models do not account for spatial correlations as our proposed models do.

There have been several recent proposals to account for spatial correlations in the context of functional regression. Zhang et al. (2016) propose a conditional autoregressive (CAR) model for spacial correlations in function-on-scalar regression. In scalar-on-function framework, Pineda-Ríos et al. (2019) propose functional simultaneous autoregressive (SAR) models to analyze an econometric data set. Huang et al. (2020) propose a robust functional SAR model based on error terms’ t-distribution assumption. Aw and Cabral (2020) develop a Bayesian inference for functional spatial linear models. We note that all these works have focused on the Gaussian response variable. It is challenging to specify a closed-form of maximum likelihood estimators by marginalizing out spatial random effects for a non-Gaussian response. Furthermore, most previous literature have studied small size spatial data sets ($n \approx 100$). This motivates the development of computationally efficient methods that allow researchers to study non-Gaussian spatial data sets with functional covariates.

Spatial generalized linear mixed models (SGLMM) (Besag, 1974; Besag et al., 1991) are popular for studying areal spatial data sets by adjusting spatial dependence within a generalized linear model framework. In this manuscript, we introduce a functional SGLMM (FSGLMM) which extends the areal SGLMM by incorporating functional covariates. The Bayesian approach is useful for these models; it can easily provide uncertainties of estimates by constructing joint posterior distributions of spatial random effects and model parameters. However, similar to SGLMM, the Bayesian inference for FSGLMM may suffer from computational and inferential challenges. (1) With increasing observations, the number of spatial random effects grows, which results in high-dimensional posterior distributions. (2) Spatial confounding (Reich et al., 2006; Hanks et al., 2015) can inflate the variance of estimates, and make it difficult to interpret fixed effects. To address these issues, we adopt recently developed sparse reparameterization (Hughes and Haran, 2013) within a SoFR framework. We propose functional sparse SGLMM (FSSGLMM) that uses reparameterization of high-dimensional random effects. Our method can alleviate spatial confounding and is faster than FSGLMM as it reduces the dimension of spatial random effects.

The outline of the remainder of this manuscript is as follows. In Section 2, we introduce FSGLMM and introduce relevant notation. In Section 3, we propose functional sparse SGLMM (FSSGLMM) and describe Bayesian inference via MCMC. Furthermore, we provide theoretical justification for our method. In Section 4, we provide simulation studies to investigate the performance of our methods. In Section 5, we apply our approach to two non-Gaussian spatial data sets and conclude with a discussion and summary in Section 6.

2 Functional Spatial Generalized Linear Mixed Models

Non-Gaussian spatial data sets frequently arise in many disciplines, including climate science, epidemiology, and social science. Examples include Poisson data on malaria incidence (Gopal et al., 2019) and satellite image data sets on the amount of water vapor (Gao and Kaufman, 2015). Spatial generalized linear mixed models (SGLMM) are useful for studying the relationship between scalar variables or understanding overall spatial patterns from fitted results. In this section, we extend SGLMM to take functional covariates.

Let $\mathcal{S} \subset \mathbb{R}^2$ be the spatial domain of interest. For a spatial location $s \in \mathcal{S}$, let Y_s be a real-valued non-Gaussian random variable; let \mathbf{z}_s be a p -dimensional vector with finite second moments; and let $\{X_s(t), t \in \mathcal{T}\}$ be a zero mean, second-order stochastic process in $L^2(\mathcal{T})$, the set of all square integrable functions on \mathcal{T} . The space $L^2(\mathcal{T})$ is equipped with the inner product $\langle X_1, X_2 \rangle = \int_{\mathcal{T}} X_1(t)X_2(t)dt$. These are observed at spatial locations s_1, \dots, s_n . Consider $\mathbf{Y} = (Y_{s_1}, \dots, Y_{s_n})^\top$, the observed response over spatial locations; $\mathbf{Z} = (\mathbf{z}_{s_1}, \mathbf{z}_{s_2}, \dots, \mathbf{z}_{s_n})^\top \in \mathbb{R}^{n \times p}$, the observed scalar covariates over spatial locations; and $\mathbf{X}(t) = (X_{s_1}(t), \dots, X_{s_n}(t))^\top$, the functional covariates that depend on spatial locations.

At each spatial location we can define the spatially correlated random variable $\{W_s : s \in \mathcal{S}\}$. Here, $\mathbf{W} = (W_{s_1}, \dots, W_{s_n})^\top$ is assumed to follow the Gaussian Markov random field (Besag, 1974) with zero-mean. With an invertible link function $g(\cdot)$ and strictly positive variance function $\sigma^2(\cdot)$, we can define FSGLMM as

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{Z}\boldsymbol{\gamma} + \int \mathbf{X}(t)\beta(t)dt + \mathbf{W} \\ f(\mathbf{W}|\tau) &\propto \tau^{n/2} \exp\left(-\frac{\tau}{2}\mathbf{W}^\top \mathbf{Q}\mathbf{W}\right), \end{aligned} \quad (1)$$

where $g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu} = \mathbb{E}(\mathbf{Y}|\mathbf{W}, \mathbf{Z}, \mathbf{X})$ and $\mathbb{V}(\mathbf{Y}|\mathbf{W}, \mathbf{Z}, \mathbf{X}) = \sigma^2(\boldsymbol{\mu}) = \sigma^2(g(\boldsymbol{\eta}))$. The functional parameter β is a square integrable function on \mathcal{T} . Here, $\mathbf{Q} = \text{diag}(\mathbf{D}\mathbf{1}) - \mathbf{D}$ is the precision matrix where $\mathbf{1}$ is the n -dimensional vector of 1's and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is the spatial adjacency matrix. $D_{ij} = 1$, if i th location and j th location are neighbors, and otherwise $D_{ij} = 0$. The parameter τ measures the smoothness of the spatial field. In (1), estimating the functional parameter $\beta(t)$ is our main interest. In this manuscript, we propose a Bayesian approach via MCMC to estimate model parameters.

Following Pineda-Ríos et al. (2019), we expand the functional covariate and parameter with respect to orthonormal basis functions $\{\phi_j\}_{j \in \mathbb{N}} \in L^2(\mathcal{T})$. Then we have

$$X_{s_i}(t) = \sum_{j=1}^{\infty} a_{ij}\phi_j(t), \quad \beta(t) = \sum_{j=1}^{\infty} b_j\phi_j(t). \quad (2)$$

In (2), $\{a_{ij}\}_{j \in \mathbb{N}}$ are random variables associated with a spatial location s_i . Without loss of generality, we assume $\mathbb{E}(a_{ij}) = 0$ and $\mathbb{E}(a_{ij}^2) = \sigma_j^2$, for all $i = 1, \dots, n$. Under the spatial homoscedasticity assumption (Pineda-Ríos et al., 2019) we have $\int \mathbb{E}(X_{s_i}^2(t))dt = \sum_{j=1}^{\infty} \sigma_j^2 < \infty$. Since $\{\phi_j\}_{j \in \mathbb{N}}$ are orthonormal basis functions, using

dominated convergence we have

$$\begin{aligned} \int X_{s_i}(t)\beta(t)dt &= \int \left(\sum_{l=1}^{\infty} a_{il}\phi_l(t)\right)\left(\sum_{j=1}^{\infty} b_j\phi_j(t)\right)dt \\ &= \sum_{l=1}^{\infty} \sum_{j=1}^{\infty} a_{il}b_j \left(\int \phi_l(t)\phi_j(t)dt\right) = \sum_{j=1}^{\infty} a_{ij}b_j \end{aligned} \tag{3}$$

for each spatial location s_i . We approximate this infinite sum by the finite sum $\sum_{j=1}^k a_{ij}b_j$ where $k = k_n$ increases as $n \rightarrow \infty$. We assume that the sum $\sum_{j=k+1}^{\infty} a_{ij}b_j$ becomes negligible. See Müller and Stadtmüller (2005) for more details. Let $\mathbf{A} = (a_{ij})_{n \times k}$ be the design matrix with the coefficients $\mathbf{b} = (b_1, \dots, b_k)^\top$. From this we can represent (1) as

$$\boldsymbol{\eta} \approx \mathbf{Z}\boldsymbol{\gamma} + \mathbf{A}\mathbf{b} + \mathbf{W} = \tilde{\mathbf{A}}\tilde{\mathbf{b}} + \mathbf{W}. \tag{4}$$

where $\tilde{\mathbf{A}} = (\mathbf{Z}, \mathbf{A})$ and $\tilde{\mathbf{b}} = (\boldsymbol{\gamma}^\top, \mathbf{b}^\top)^\top$.

Let $f(\mathbf{Y}|\tilde{\mathbf{b}}, \mathbf{W})$ be the conditional distribution of the response, which is from the exponential family. Once we specify priors $p(\tilde{\mathbf{b}})$, $p(\tau)$, we can define the joint posterior distribution as $\pi(\mathbf{W}, \tilde{\mathbf{b}}, \tau|\mathbf{Y}) \propto f(\mathbf{Y}|\tilde{\mathbf{b}}, \mathbf{W})f(\mathbf{W}|\tau)p(\tilde{\mathbf{b}})p(\tau)$. However, Bayesian inference for such models is computationally challenging due to the high-dimensional spatial random effects $\mathbf{W} \in \mathbb{R}^n$. The dimension of the joint posterior increases with the increasing number of observations n . Such high-dimensional random effects are spatially dependent, which results in slow mixing of the MCMC algorithm.

Furthermore, there can be serious inferential issues in FSGLMM due to spatial confounding between fixed and random effects as in a classical multivariate statistics (Reich et al., 2006; Hodges and Reich, 2010). Spatial confounding arises frequently when the spatial covariates \mathbf{A} are collinear with the spatial random effects \mathbf{W} . Consider the projection matrix $\mathbf{P} = \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}$ and its complement $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$. Then we can represent (4) as

$$\boldsymbol{\eta} \approx \tilde{\mathbf{A}}\tilde{\mathbf{b}} + \mathbf{W} = \tilde{\mathbf{A}}\tilde{\mathbf{b}} + \mathbf{P}\mathbf{W} + \mathbf{P}^\perp\mathbf{W}. \tag{5}$$

Here, $\tilde{\mathbf{A}}$ is confounded with $\mathbf{P}\mathbf{W}$ because of their linear relationship. To address such multicollinearity issues, Reich et al. (2006) propose a restricted spatial regression model by removing $\mathbf{P}\mathbf{W}$ in (5). However, the Bayesian credible intervals obtained from a restricted spatial regression model can be narrow; Hanks et al. (2015) suggest adding a posteriori adjustments to the MCMC samples from this model. Hughes and Haran (2013) develop sparse SGLMM to alleviate the spatial confounding and gain computational efficiency. Based on this approach, we develop a SoFR for non-Gaussian spatial data that is fast, while providing accurate functional estimates.

3 Functional Sparse Spatial Generalized Linear Mixed Models

Hughes and Haran (2013) develop a reduced rank approach that uses Moran eigenvectors as a set of basis functions for spatial random effects. These natural basis functions

are orthogonal to the fixed effects. Spatial filtering methods (cf. Griffith, 2000; Getis and Griffith, 2002; Griffith, 2004) have been widely used for geographical and environmental science problems. By extracting eigenvectors from spatial configuration matrices, filtering methods account for spatial dependencies or missing spatial covariates; therefore they are useful for addressing confounding issues. The eigenvectors are used as basis functions to describe distinct spatial associations. Especially, Griffith (2000) uses eigenvectors of $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{D}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ as additional predictors to filter out spatial correlations (see Griffith (2002, 2004), as cases for non-Gaussian spatial data). This operator is called as the Moran operator (Moran, 1950) that can measure dependencies corresponding to spatial adjacency matrix \mathbf{D} . By adapting these approaches, Hughes and Haran (2013) develop a reduced rank approach that uses eigenvectors of the generalized Moran operator, $\mathbf{P}^\perp\mathbf{D}\mathbf{P}^\perp$. These natural basis functions are orthogonal to the fixed effect and account for spatial dependencies. Such Moran basis functions have also been used in many spatiotemporal applications (e.g., Bradley et al. (2015); Musgrove et al. (2016); Heaton et al. (2020)). Based on this we propose the functional sparse spatial generalized linear mixed models (FSSGLMM) as follows.

$$\begin{aligned}\boldsymbol{\eta} &= \mathbf{Z}\boldsymbol{\gamma} + \int \mathbf{X}(t)\beta(t)dt + \mathbf{M}\boldsymbol{\delta} \\ f(\boldsymbol{\delta}|\tau) &\propto \tau^{m/2} \exp\left(-\frac{\tau}{2}\boldsymbol{\delta}'\mathbf{M}'\mathbf{Q}\mathbf{M}\boldsymbol{\delta}\right),\end{aligned}\tag{6}$$

where $\mathbf{M} \in \mathbb{R}^{n \times m}$ is the projection matrix. To obtain the projection matrix \mathbf{M} , we use the following procedure. (1) We calculate \mathbf{P}^\perp , a complement of the projection matrix $\mathbf{P} = \tilde{\mathbf{A}}(\tilde{\mathbf{A}}^\top\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}$. Here, $\tilde{\mathbf{A}}$ is a matrix consisting of the observed scalar covariates $\mathbf{Z} \in \mathbb{R}^{n \times p}$ and the design matrix from the functional basis expansion $\mathbf{A} \in \mathbb{R}^{n \times k}$ (i.e., $\tilde{\mathbf{A}} = (\mathbf{Z}, \mathbf{A}) \in \mathbb{R}^{n \times (p+k)}$). (2) Then we compute the Moran operator $\mathbf{P}^\perp\mathbf{D}\mathbf{P}^\perp$, where $\mathbf{D} \in \mathbb{R}^{n \times n}$ is the spatial adjacency matrix defined in Section 2. (3) Lastly, we obtain the $\mathbf{M} \in \mathbb{R}^{n \times m}$ by taking the first m principal components of the Moran operator. Note that this requires a spectral decomposition of the $n \times n$ matrix, which becomes computationally expensive with increasing n , but this computation needs to be done only once. For instance, for the COVID-19 data example (Section 5.2), we have $n = 3,108$, and a spectral decomposition of $\mathbf{P}^\perp\mathbf{D}\mathbf{P}^\perp$ takes about 150 seconds. However, we only need to compute a spectral decomposition a single time before implementing our MCMC algorithm. In the MCMC updates, \mathbf{M} is not a model parameter but is a fixed design matrix consisting of Moran's eigenvectors. Therefore, we don't need to repeat the spectral decomposition of a high-dimensional matrix. Incorporating \mathbf{M} in the model allows us to capture distinct clustering patterns of residual to $\tilde{\mathbf{A}}$, which contains the information of both scalar and functional covariates. Furthermore, it also allows us to consider the underlying neighborhood structure among observations (Boots and Tiefelsdorf, 2000). The positive eigencomponents of the Moran operator correspond to the spatial clustering and the negative ones correspond to spatial repulsion. Since modeling spatial clustering is of interest in many cases, we consider the positive eigencomponents in our model as in Hughes and Haran (2013). With this projection approach, we can represent (4) as

$$\boldsymbol{\eta} \approx \tilde{\mathbf{A}}\tilde{\mathbf{b}} + \mathbf{M}\boldsymbol{\delta}.\tag{7}$$

Following recommendation in Hughes and Haran (2013), we use approximately 10% of the eigencomponents to select m . For example, in our simulation studies we found that $m = 90$ for $n = 900$ can provide accurate functional estimates.

3.1 Markov Chain Monte Carlo Methods for FSSGLMM

With the conditional distribution $f(\mathbf{Y}|\tilde{\mathbf{b}}, \mathbf{M}, \boldsymbol{\delta})$ and priors $p(\tilde{\mathbf{b}})$, $p(\tau)$, we can define the joint posterior distribution for the FSSGLMM as $\pi(\boldsymbol{\delta}, \tilde{\mathbf{b}}, \tau|\mathbf{Y}) \propto f(\mathbf{Y}|\tilde{\mathbf{b}}, \mathbf{M}, \boldsymbol{\delta})f(\boldsymbol{\delta}|\tau)p(\tilde{\mathbf{b}})p(\tau)$. Compared to the original posterior $\pi(\mathbf{W}, \tilde{\mathbf{b}}, \tau|\mathbf{Y})$, the dimension of $\pi(\boldsymbol{\delta}, \tilde{\mathbf{b}}, \tau|\mathbf{Y})$ is much smaller ($m + k + 1 \ll n + k + 1$). Here, we use a diffuse multivariate normal prior on $\tilde{\mathbf{b}} \sim \mathcal{N}(0, 1000\mathbf{I})$ and an approximated reference prior $1/(0.5 + \tau)^2$ proposed by Ferreira et al. (2021). Note that the application of a naive uniform prior for τ would lead to an improper distribution and the reference prior provides adequate uncertainty quantification especially for Gaussian Markov random field (Keefe et al., 2019; Ferreira et al., 2021). We observe that our method performs well across the different choices of priors in practice. We provide prior sensitivity analysis in the supplementary material (Kang et al., 2023). From this we can obtain full conditionals as

$$\begin{aligned}\pi(\tilde{\mathbf{b}}|\boldsymbol{\delta}, \tau) &\propto \prod_{i=1}^n f(Y_i|\tilde{\mathbf{b}}, \mathbf{M}, \boldsymbol{\delta}) \times p(\tilde{\mathbf{b}}) \\ \pi(\tau|\boldsymbol{\delta}, \tilde{\mathbf{b}}) &\propto \tau^{m/2} \exp\left(-\frac{\tau}{2}\boldsymbol{\delta}'\mathbf{M}'\mathbf{Q}\mathbf{M}\boldsymbol{\delta}\right) \times p(\tau) \\ \pi(\boldsymbol{\delta}|\tilde{\mathbf{b}}, \tau) &\propto \prod_{i=1}^n f(Y_i|\tilde{\mathbf{b}}, \mathbf{M}, \boldsymbol{\delta}) \times \exp\left(-\frac{\tau}{2}\boldsymbol{\delta}'\mathbf{M}'\mathbf{Q}\mathbf{M}\boldsymbol{\delta}\right),\end{aligned}\tag{8}$$

which can be easily sampled using Metropolis-Hasting random walk updates. Starting with an arbitrary initial value $(\boldsymbol{\delta}^{(0)}, \tilde{\mathbf{b}}^{(0)}, \tau^{(0)})$ (e.g., sampling from the prior distributions), we successively update the parameters. Following Hughes and Haran (2013), we update $\tilde{\mathbf{b}}$ with a normal proposal with variance $\hat{\mathbf{V}}$, where $\hat{\mathbf{V}}$ is the estimated asymptotic covariance matrix from the standard GLM fit. We update $\boldsymbol{\delta}$ by using a multivariate random walk with normal proposals. We use the probabilistic programming language `nimble` for implementing this, which is popular for Bayesian inferences. Then the stationary distribution from this MCMC algorithm converges towards the joint posterior distribution $\pi(\boldsymbol{\delta}, \tilde{\mathbf{b}}, \tau|\mathbf{Y})$.

Computationally efficient Bayesian approaches for latent variable models have been widely developed. For example, Christensen and Waagepetersen (2002); Christensen (2004); Christensen et al. (2006) proposed a fast mixing Langevin Hastings algorithm for SGLMM that utilizes gradient information of the log posterior in proposal distribution. The integrated nested Laplace approximation (Rue et al., 2009; Lindgren et al., 2011) can also quickly estimate marginal posteriors from latent Gaussian Markov random fields. Polson et al. (2013) develop a data augmentation strategy based on the Pólya-Gamma distributions, resulting in efficient Gibbs-sampling algorithms. Recently, Bradley et al. (2018, 2020) introduce latent conjugate multivariate distributions for dependent non-Gaussian data that allows a fast simulation from full conditional distribution. Adapting some of these computational methods to estimate our posterior

$\pi(\boldsymbol{\delta}, \tilde{\mathbf{b}}, \tau | \mathbf{Y})$ is possible. For the sake of concreteness, we focus here on the projection-based MCMC algorithm (Hughes and Haran, 2013). They use the projection-based basis functions, which can reduce the dimension of spatial random effects and make them less correlated. Then a classic MCMC algorithm is implemented for this reduced-dimensional representation. This is quite simple but effective for spatially correlated datasets, including our cases. Due to such practical advantages, the reduced-rank approach using Moran eigenvectors is applied in many complex Bayesian hierarchical models. Examples include the Bayesian spatio-temporal model for functional magnetic resonance imaging data (Musgrove et al., 2016) and the Bayesian change point model for estimating epidemic curves of respiratory syncytial virus (Heaton et al., 2020).

3.2 Theoretical Justifications

Consider the model in (8) and the following representations.

1. Let $\beta \in L^2(\mathcal{T})$ be the true regression function that can be represented by $\beta = \sum_{j=1}^{\infty} \langle \beta, \phi_j \rangle \phi_j \equiv \sum_{j=1}^{\infty} b_j \phi_j$.
2. Let $\beta_{k_n} = \sum_{j=1}^{k_n} b_j \phi_j$ be the truncated regression function with k_n large enough so that $\sum_{j>k_n} b_j \phi_j$ is negligible. And let $\mathbf{b}_{k_n} = (b_1, \dots, b_{k_n})^\top$.
3. Let $\mathbf{b}_{k_n}^{(u)} = (b_1^{(u)}, \dots, b_{k_n}^{(u)})^\top$ be the posterior sample from u th iteration of MCMC (Section 3.1).
4. Let $\beta_{k_n}^{(u)} = \sum_{j=1}^{k_n} b_j^{(u)} \phi_j \in L^2(\mathcal{T})$ be a function constructed using MCMC samples.
5. Let $\hat{\beta}_{k_n}^{(u)} = \sum_{j=1}^{k_n} \bar{b}_j^{(u)} \phi_j$ be the estimated regression function where $\bar{b}_j^{(u)} = \frac{1}{u} \sum_{l=1}^u b_j^{(l)}$. Let $\bar{\mathbf{b}}_{k_n}^{(u)} = (\bar{b}_1^{(u)}, \dots, \bar{b}_{k_n}^{(u)})^\top$.

Assumption 3.1 summarizes our presumptions. Overall, these assumptions are general and not restrictive. The first assumption is that we are working under a given fixed basis system. In the second assumption, k_n depends on n , and both go to infinity, which is commonly assumed. The third assumption allows us to do any computation on these functions. Note that $\|\cdot\|$ is the norm in $L^2(\mathcal{T})$, and $\|\cdot\|_{\mathbb{R}^p}$ denotes the Euclidean norm.

Assumption 3.1. *We make the following assumptions.*

1. *The basis system $\{\phi_j\}_{j=1}^{\infty}$ is a fixed orthonormal system of $L^2(\mathcal{T})$.*
2. *The truncation $k_n \rightarrow \infty$ as $n \rightarrow \infty$.*
3. *All of β , β_{k_n} , $\beta_{k_n}^{(u)}$, and $\hat{\beta}_{k_n}^{(u)}$ are defined on the same probability space.*

Under Assumption 3.1, we show that the functions constructed from the MCMC sample converge in distribution to the true β in Theorem 3.1. Since the functions from the MCMC sample follow the distribution of the true β , we can create the corresponding credible intervals as in Section 3.3.

Theorem 3.1. Consider the sequence of random functions $\{\beta_{k_n}^{(u)}\}_{n,u} \in L^2(\mathcal{T})$. Then $\beta_{k_n}^{(u)} \xrightarrow{\mathcal{D}} \beta$ as $n \rightarrow \infty$ and $u \rightarrow \infty$.

Proof. An ergodic Markov chain $\{\mathbf{b}_k^{(u)}\}_{u \in \mathbb{N}}$ will converge to the stationary distribution $\pi(\mathbf{b}|\mathbf{Y})$ in terms of total variation distance. Therefore, we have $\mathbf{b}_{k_n}^{(u)} \xrightarrow{\mathcal{D}} \mathbf{b}_{k_n}$ as $u \rightarrow \infty$ for every $k_n \in \mathbb{N}$. Then $\beta_{k_n}^{(u)} \xrightarrow{\mathcal{D}} \beta_{k_n}$ for every k_n by the continuous mapping theorem. Since $\beta_{k_n} \xrightarrow{\mathcal{D}} \beta$ as $n \rightarrow \infty$, we have $\beta_{k_n}^{(u)} \xrightarrow{\mathcal{D}} \beta$ as $u \rightarrow \infty$ and $n \rightarrow \infty$. \square

In Theorem 3.2, we show that our estimated regression function $\hat{\beta}_{k_n}$ converges in probability to the true regression function β as both n (sample size) and u (MCMC sample size) go to infinity.

Theorem 3.2. Consider the estimated regression function $\hat{\beta}_{k_n}^{(u)}$. Under Assumption 3.1, we have $\hat{\beta}_{k_n}^{(u)} \xrightarrow{\mathcal{P}} \beta$ as $n \rightarrow \infty$ and $u \rightarrow \infty$.

Proof. First we fix k_n . Since $\{\mathbf{b}_k^{(u)}\}_{u \in \mathbb{N}}$ is an ergodic and stationary Markov chain, we have $\bar{\mathbf{b}}_{k_n}^{(u)} \xrightarrow{\mathcal{P}} \mathbf{b}_{k_n}$ as $u \rightarrow \infty$ by the law of large numbers. As

$$P\left(\|\hat{\beta}_{k_n}^{(u)} - \beta_{k_n}\|^2 \geq \frac{\epsilon}{2}\right) = P\left(\|\bar{\mathbf{b}}_{k_n}^{(u)} - \mathbf{b}_{k_n}\|_{\mathbb{R}^{k_n}}^2 \geq \frac{\epsilon}{2}\right) \rightarrow 0,$$

we have $\hat{\beta}_{k_n}^{(u)} \xrightarrow{\mathcal{P}} \beta_{k_n}$ as $u \rightarrow \infty$ for every $k_n \in \mathbb{N}$.

Since β_{k_n} is the truncated version of β , we have $\beta_{k_n} \xrightarrow{\mathcal{P}} \beta$ as $n \rightarrow \infty$ (i.e. $k_n \rightarrow \infty$) and this gives $P\left(\|\beta_{k_n} - \beta\|^2 \geq \frac{\epsilon}{2}\right) \rightarrow 0$.

$$\begin{aligned} P\left(\|\beta_{k_n}^{(u)} - \beta\|^2 \geq \epsilon\right) &\leq P\left(\|\beta_{k_n}^{(u)} - \beta_{k_n}\|^2 + \|\beta_{k_n} - \beta\|^2 \geq \epsilon\right) \\ &\leq P\left(\|\beta_{k_n}^{(u)} - \beta_{k_n}\|^2 \geq \frac{\epsilon}{2}\right) + P\left(\|\beta_{k_n} - \beta\|^2 \geq \frac{\epsilon}{2}\right) \\ &\rightarrow 0 + 0 \end{aligned}$$

as $u \rightarrow \infty$ and $n \rightarrow \infty$. \square

Investigating the ergodicity and stability of the Markov chain is crucial to guarantee convergence to the stationary distribution. For instance, Tong and Van Handel (2012) studied the ergodic theory of nonlinear filters (i.e., the conditional distribution of unobserved process for given observed process). Please also see seminal papers for the convergence of Markov chain (Mengersen and Tweedie, 1996; Tierney, 1994). For SGLMM, Christensen et al. (2001) showed that the Metropolis-Hastings algorithm with a random walk proposal produces an ergodic Markov chain. This result is based on Lemmas 1.1, 1.2 in Mengersen and Tweedie (1996) and Corollary 2 in Tierney (1994), which implies that an aperiodic, irreducible, positive Harris recurrent Markov chain is ergodic.

Here, we also used a random walk proposal, which is a default setting in `nimble` to update model parameters. Since Theorems 1, 2 do not require more than the standard SGLMM settings in Christensen et al. (2001), we can apply this result to guarantee convergence of the MCMC algorithm. However, the more challenging question is the rate of convergence. For instance, Christensen et al. (2001) showed the geometric ergodicity for both random walk Metropolis-Hastings and truncated Langevin-Hastings algorithms in the standard SGLMM. Cowles (2002) also provide a convergence rate for a spatial linear model with a pairwise-differences prior based on a drift condition and a minorization condition (Rosenthal, 1995). However, these are quite MCMC sampler specific. Deriving such conditions for the projection-based MCMC sampler is challenging even without functional covariates. This is an interesting avenue for future research that is out of the scope of this paper.

3.3 Credible Intervals for Functional Parameter

The credible intervals for functional parameter can be constructed using the posterior sample of MCMC, $\{\mathbf{b}_k^{(l)}\}_{l=1}^u = \{(b_1^{(l)}, \dots, b_k^{(l)})^\top\}_{l=1}^u$. We can reconstruct $\{\beta_k^{(l)}(t) = \sum_{j=1}^k b_j^{(l)} \phi_j(t)\}_{l=1}^u$. We evaluate the $\beta_k^{(l)}(t_h)$ for each point t_h , $h = 1, \dots, H$, where $t_1 < t_2 < \dots < t_H$, and $l = 1, \dots, u$. We calculate the simultaneous credible bounds following Crainiceanu et al. (2007). Let $\hat{\beta}_k(t_h)$ and $SD(\beta_k(t_h))$ be the sample mean and sample standard deviation of $\{\beta_k^{(l)}(t_h)\}_{l=1}^u$. Let M_α be the $(1 - \alpha)$ sample quantile of

$$\max_{1 \leq h \leq H} \left| \frac{\beta_k^{(l)}(t_h) - \hat{\beta}_k(t_h)}{SD(\beta_k(t_h))} \right|.$$

Then we find the $(1 - \alpha)$ simultaneous credible intervals become

$$I(\alpha, t_h) = \hat{\beta}_k(t_h) \pm M_\alpha SD(\beta_k(t_h)), \quad h = 1, \dots, H.$$

We summarize the procedure as the following.

1. We reconstruct $\beta_k^{(l)}(t) = \sum_{j=1}^k b_j^{(l)} \phi_j(t)$ from $\mathbf{b}_k^{(l)}$, for $l = 1, \dots, u$.
2. We evaluate $\beta_k^{(l)}(t_h)$ for grid points $t_h \in \mathcal{T}$, $h = 1, \dots, H$.
3. We find M_α by first calculating $\max_{1 \leq h \leq H} \left| \frac{\beta_k^{(l)}(t_h) - \hat{\beta}_k(t_h)}{SD(\beta_k(t_h))} \right|$ for each $l = 1, \dots, u$ and then computing the $(1 - \alpha)$ sample quantile of them.
4. We construct the simultaneous confidence intervals by $I(\alpha, t_h) = \hat{\beta}_k(t_h) \pm M_\alpha SD(\beta_k(t_h))$, $h = 1, \dots, H$.

4 Simulated Data Examples

We apply our approach to Poisson and Bernoulli data over the discrete spatial domain. To illustrate the performance of FSSGLMM, we compare it with standard functional GLM (FGLM) and FSGLMM. Our method is flexible in that we can construct a wide variety of Bayesian hierarchical models with a minor change in `nimble` code. We also provide simulation studies for zero-inflated Poisson and negative binomial in the supplementary material. MCMC algorithms are run until the Monte Carlo standard errors (Jones et al., 2006; Flegal et al., 2008) for FSSGLMM are at or below 0.01 (300,000 iterations). The simulation steps are as follows.

1. We construct the spatial adjacency matrix $\mathbf{D} \in \mathbb{R}^{900 \times 900}$ using the neighbor structure on a 30×30 lattice over $[0, 1]^2$ domain. We set a scalar covariate \mathbf{Z} as the coordinates of the vertices to the unit square with the scalar regression parameter $\boldsymbol{\gamma} = (-1, 1)^\top$.
2. To consider spatial correlations among functional observations, we generate functional covariate coefficients $\mathbf{A} = (a_{ij})_{n \times k}$ from the Matérn class (Stein, 2012) covariance function, obtained from the distance matrix of \mathbf{Z} . For each j , we simulate $\mathbf{a}_{\cdot j} = (a_{ij})_{i=1}^n$ from a Gaussian process with mean 1 and Matérn covariance with variance 1, range 0.3, and smoothness as 1.5. By using 7 Fourier basis functions over $\mathcal{T} = [0, 1]$, we create $X_{s_i}(t) = \sum_{j=1}^7 a_{ij} \phi_j(t)$ for $i = 1, \dots, 900$.
3. We calculate $\int_{\mathcal{T}} \mathbf{X}(t) \beta(t) dt$ numerically for given the true functional regression parameter $\beta(t)$. Especially, we investigate three different types of $\beta(t)$ with zero function, low frequency function, and high frequency function as follows:
 - $\beta_{zero}(t) = 0$
 - $\beta_{low}(t) = 0.2 \cos(11t - 2) + 0.1 \sin(8t - 1)$
 - $\beta_{high}(t) = 0.5(2(2t - 1)^5 + 3(2t - 1)^2 + \cos(3\pi(2t - 1)))$.

For brevity, we focus on visualizing $\beta_{high}(t)$ in the manuscript. We provide results for $\beta_{zero}(t)$ and $\beta_{low}(t)$ in the supplementary material.

4. Following Hughes and Haran (2013), we construct a projection matrix $\mathbf{M} \in \mathbb{R}^{900 \times 90}$ by taking positive eigencomponents of the Moran operator; this allows positive spatial dependence (i.e., clustering) among observations. Then we simulate random effects $\boldsymbol{\delta}$ from $N(0, \tau \mathbf{M}' \mathbf{Q} \mathbf{M})$ with $\tau = 0.3$.
5. Given the generated random effects and covariates, non-Gaussian observations are simulated from a link function $g(\cdot)$.
 - We used a logit link $\log(p/(1-p)) = \mathbf{Z}\boldsymbol{\gamma} + \int \mathbf{X}(t)\beta(t)dt + \mathbf{M}\boldsymbol{\delta}$ for Bernoulli observations.
 - We used a log link $\log(\lambda) = \mathbf{Z}\boldsymbol{\gamma} + \int \mathbf{X}(t)\beta(t)dt + \mathbf{M}\boldsymbol{\delta}$ for count observations.

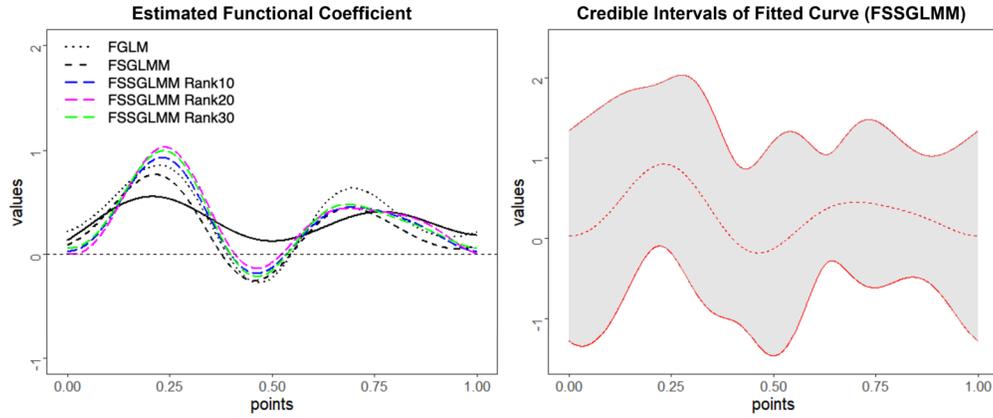


Figure 1: Regression function estimates for a simulated Bernoulli dataset. The left panel shows the posterior mean estimates of $\beta_{high}(t)$ from FGLM, FSGLMM, and FSSGLMM with varying ranks. The right panel shows the posterior mean estimates and corresponding 95% credible intervals from FSSGLMM with 10% of n (i.e., $m = 90$).

4.1 Bernoulli Data

Figure 1 compares $\hat{\beta}_{high}(t)$ estimates from different methods. We fit FSSGLMM at $m = 90, 180, 270$, which correspond to 10%, 20%, 30% of n (sample size). We observe that $\hat{\beta}_{high}(t)$ obtained from FGLM, FSGLMM, and FSSGLMM with different ranks are all reasonably close to the simulated truth. We also observe that the credible intervals obtained from FSSGLMM can cover the true $\beta_{high}(t)$ well. Table 1 shows that FSSGLMM can recover the true parameter values well compared to the other two methods. To measure the spatial correlation among residuals, we calculate the Moran’s I (Moran, 1950), a simple and popular nonparametric statistic. We note that this is more useful for exploratory data analysis rather than rigorous statistical inference. We observe that residuals from FGLM and FSGLMM are spatially correlated. On the other hand, FSSGLMM account for such correlations and result in errors that are not spatially correlated. Furthermore, we observe that FSSGLMM is faster than FSGLMM due to the projection step.

We repeat the simulation 100 times to assess the performance of our method. Figure 2 shows that the estimated curves from FSSGLMM are densely distributed around the true curve. In general, the estimated curves from FSSGLMM are accurate with different rank sizes. Figure 3 compares the mean square error (MSE) for all parameters. We obtain MSE of $\hat{\beta}_{high}(t)$ numerically as

$$\frac{1}{1000} \sum_{h=1}^{1000} (\hat{\beta}_{high}(t_h) - \beta_{high}(t_h))^2,$$

where t_h is an evaluated point for $h = 1, \dots, 1, 200$. We observe that FSSGLMM shows the smallest MSE values for all model parameters. As described in Section 3.3, we inves-

FGLM			
	γ_1	γ_2	τ
Mean	-0.936	0.961	NA
95%HPD	(-1.826,-0.079)	(0.398,1.513)	NA
Moran's I (p-value)	0.168 (0.000)		
Time (min)	1.22		
FSGLMM			
	γ_1	γ_2	τ
Mean	-0.788	0.919	0.260
95%HPD	(-1.833,0.057)	(0.282,1.512)	(0.237,0.284)
Moran's I (p-value)	0.149 (0.000)		
Time (min)	18.92		
FSSGLMM (10% of n)			
	γ_1	γ_2	τ
Mean	-1.145	1.179	0.238
95%HPD	(-2.135,-0.290)	(0.591,1.834)	(0.124,0.386)
Moran's I (p-value)	-0.029 (0.877)		
Time (min)	3.31		

Table 1: Inference results for a simulated Bernoulli dataset. The true values are $\gamma_1 = -1$, $\gamma_2 = 1$, $\tau = 0.3$. FGLM and FSGLMM show high Moran's I value, which means there can be spatial correlations not accounted for by the model.

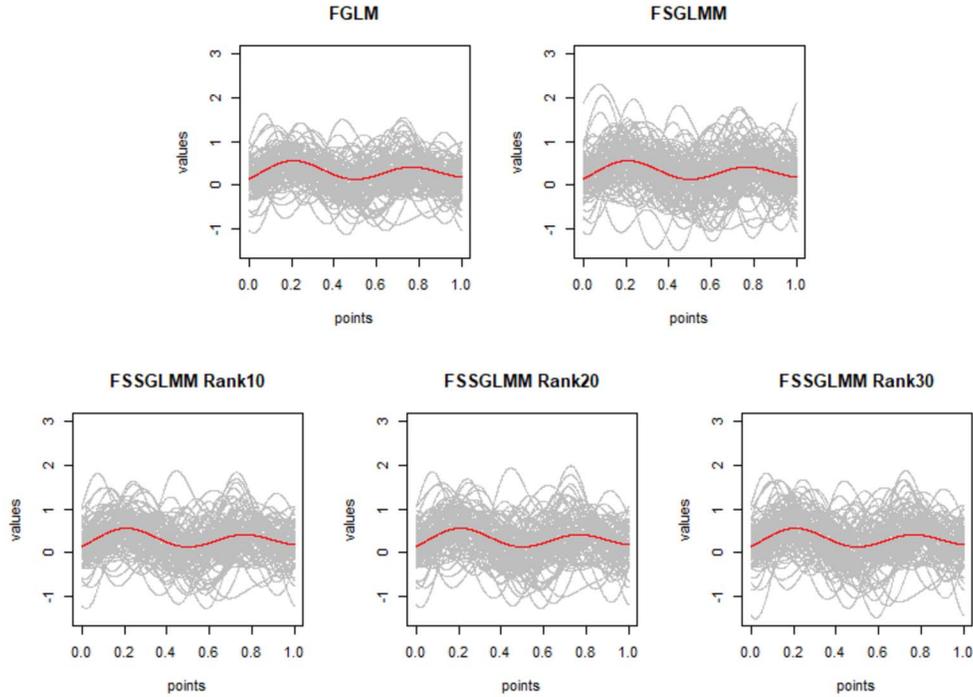


Figure 2: The posterior mean estimates of $\beta_{high}(t)$ for the 100 simulated Bernoulli datasets. Red lines indicate true function. We fit FSSGLMM with varying ranks (10%, 20%, 30% of n).

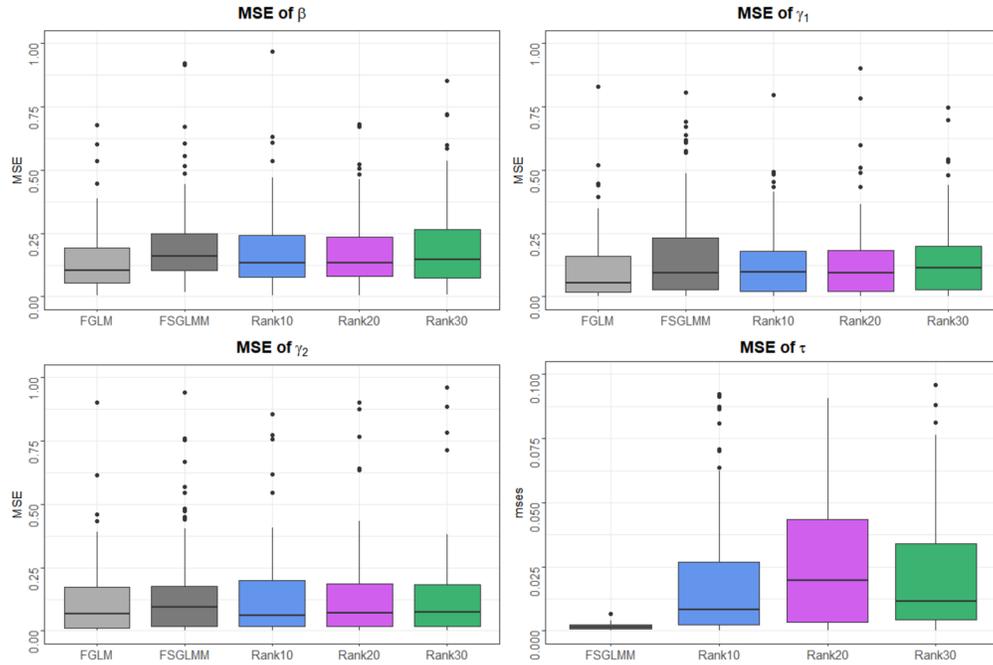


Figure 3: MSE obtained from the 100 simulated Bernoulli datasets under the high frequency function. Overall the performance is robust across the different choice of ranks for FSSGLMM.

tigate the coverages for 100 simulated datasets for β_{zero} , β_{low} and β_{high} in Table 2. Here, we report the mean coverages from the simulated data sets. The mean coverages from FSSGLMM are close to the 95% nominal rate. We also observe that the performance of FSSGLMM is robust across different choices of ranks.

	β	FGLM	FSGLMM	FSSGLMM Rank10	FSSGLMM Rank20	FSSGLMM Rank30
Bernoulli	β_{zero}	0.999	0.989	0.998	0.997	0.996
	β_{low}	0.995	0.988	0.994	0.993	0.993
	β_{high}	0.998	0.985	0.995	0.994	0.994
Poisson	β_{zero}	0.718	0.683	0.993	0.941	0.884
	β_{low}	0.692	0.645	0.992	0.944	0.867
	β_{high}	0.698	0.657	0.987	0.933	0.848

Table 2: Coverage obtained from 1,200 grid points for Bernoulli and Poisson datasets. Mean coverage of 100 simulated datasets is calculated.

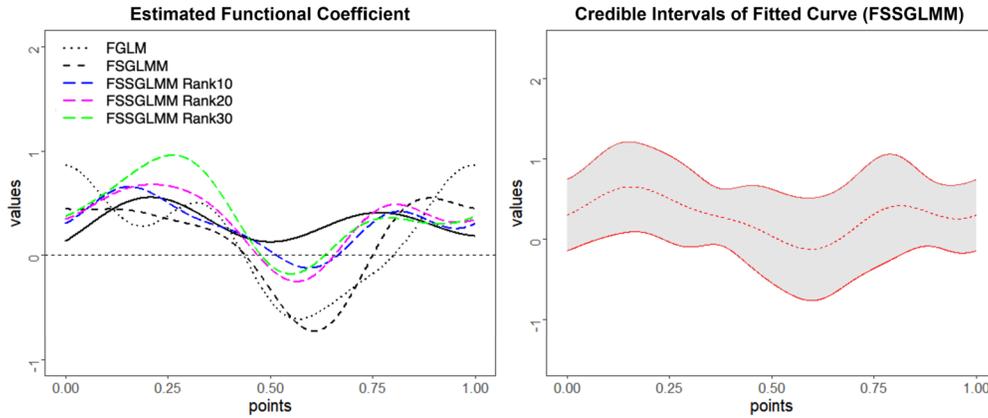


Figure 4: Regression function estimates for a simulated Poisson dataset. The left panel shows the posterior mean estimates of $\beta_{high}(t)$ from FGLM, FSGLMM, and FSSGLMM with varying ranks. The right panel shows the posterior mean estimates and corresponding 95% credible intervals from FSSGLMM with 10% of n (i.e., $m = 90$).

4.2 Poisson Data

We now repeat the same procedure with Poisson data. Figure 4 shows the estimated functional parameters $\hat{\beta}_{high}(t)$ of the three methods along with a credible intervals of FSSGLMM. Compared to the estimates from FSGLMM and FGLM, $\hat{\beta}_{high}(t)$ from FSSGLMM are close to the truth. One noticeable difference with the Bernoulli case is that the credible intervals are much narrower for the Poisson dataset. Table 3 shows that FSSGLMM estimates can recover the truth. The point estimates from FSSGLMM are much closer to the true parameters than the estimates from the other two models. Furthermore, parameter estimates in FSGLMM are biased due to spatial confounding. FSSGLMM can account for the spatial correlation in the data better than other methods, signified by the higher p-values of Moran’s I. We note that Moran’s I test is an exploratory rather than a confirmatory tool.

We conduct the simulation 100 times. Figure 5 shows that the estimated curves from FSSGLMM are densely distributed around the true curve compared to those from the other two models. In the Poisson examples, the estimated curves become more accurate with smaller m . In general, FSSGLMM shows the smallest MSE values across model parameters except for τ (Figure 6). When we study the coverages for 100 simulated data sets in Table 2, we observe that coverages from FSSGLMM are improved with smaller rank, becoming close to the 95% nominal rate. Table 2 indicates that Poisson responses are more affected by the frequency of curves than Bernoulli responses; coverages become lower for a higher frequency curve. In the Bernoulli case, we only have two possible outcomes (0 or 1); therefore, extreme values do not occur even with the high or low peak of the curves. On the other hand, Poisson responses are more likely to be affected by the frequency of curves because the peak values can lead to extreme count values. Therefore, achieving the nominal rate for Poisson cases is more challenging when we

FGLM			
	γ_1	γ_2	τ
Mean	-0.888	1.244	NA
95%HPD	(-1.207,-0.575)	(1.069,1.423)	NA
Moran's I (p-value)	0.497 (0.000)		
Time (min)	2.03		
FSGLMM			
	γ_1	γ_2	τ
Mean	-1.052	0.414	0.258
95%HPD	(-1.333,-0.732)	(0.244,0.572)	(0.237,0.282)
Moran's I (p-value)	0.289 (0.000)		
Time (min)	19.71		
FSSGLMM (10% of n)			
	γ_1	γ_2	τ
Mean	-1.165	0.965	0.224
95%HPD	(-1.684,-0.642)	(0.705,1.253)	(0.137,0.289)
Moran's I (p-value)	-0.110 (0.999)		
Time (min)	3.53		

Table 3: Inference results for a simulated Poisson dataset. Estimates of FSSGLMM are closest to the true values $\gamma_1 = -1$, $\gamma_2 = 1$, $\tau = 0.3$. FGLM and FSGLMM show high Moran's I values which mean there can be spatial correlations not accounted by the models.

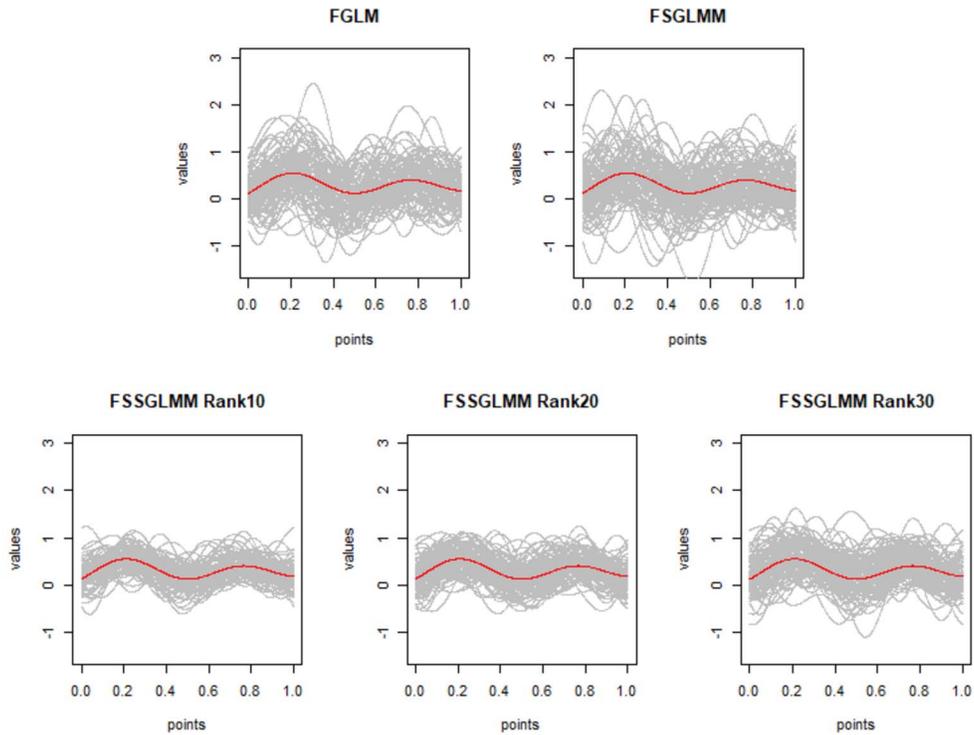


Figure 5: The posterior mean estimates of $\beta_{high}(t)$ for the 100 simulated Poisson datasets. Red lines indicate true function. We fit FSSGLMM with varying ranks (10%, 20%, 30% of n).

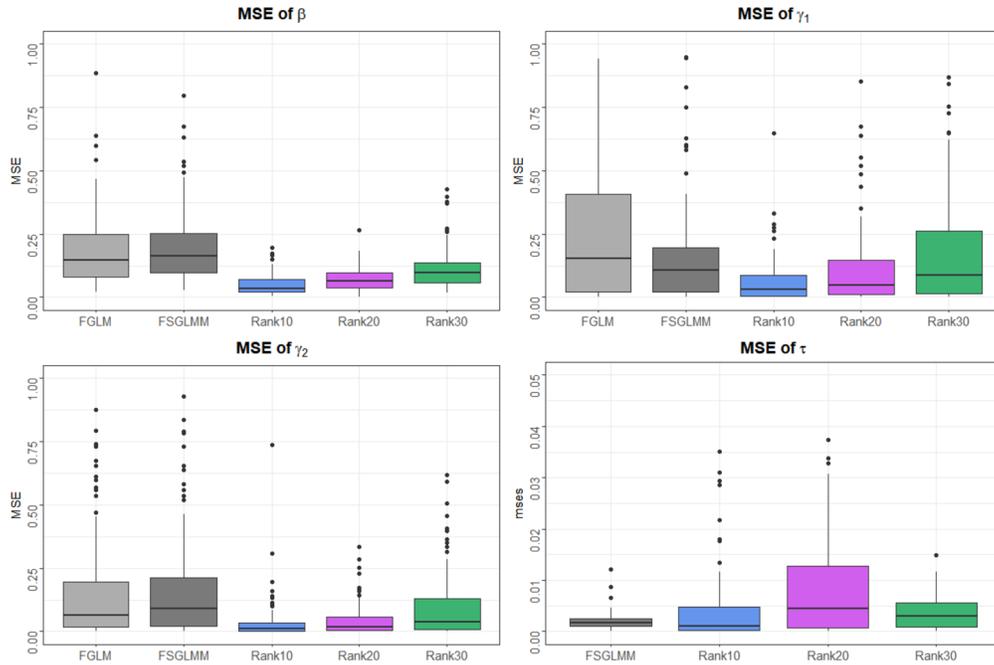


Figure 6: MSE obtained from the 100 simulated Poisson datasets under the high frequency function. In general, FSSGLMM shows the lowest MSE except for τ . MSE of FSSGLMM becomes bigger as m increases.

have high-frequency functions. Note that Poisson data have not been studied much, especially in the functional GLM context, and adapting extreme value models into our framework could also be a potential extension.

5 Real Data Examples

We apply our method to two real data examples: (1) malaria incidence data and (2) presence-absence confirmed cases of COVID-19 data. For both cases, FSSGLMM can provide comparable functional estimates with the non-spatial model (FGLM) and account for spatial correlation. FSSGLMM can conduct statistical inference much faster than FSGLMM. We also provide fitted results using FPCA in the supplementary material.

5.1 Malaria Incidence Data

The malaria incidence data set has been collected from Ethiopia Demographic and Health Surveys of 2015 (ICF, 2004-2017 (Accessed July, 1, 2020)). The data set contains malaria incidence (the Poisson response) from 557 GPS clusters in Ethiopia. Malaria is

one of the leading causes of death worldwide and hence is of significant public health interest. Investigating the relationship between malaria incidence and environmental variables is important. We use the vegetation index of the region, the proximity to water, and average annual rainfall as scalar covariates. The vegetation index can be used to quantify the greenness of the GPS cluster; the higher value indicates the region with dense vegetation. These variables are used in Gopal et al. (2019) for analyzing malaria incidence in Keyna under a spatial regression framework. Furthermore, we use the average monthly temperature (1970-2000) as a functional covariate. We expand monthly temperature with 7 Fourier basis functions (Figure 7). We chose the number of basis functions by minimizing generalized cross-validation (GCV) error. Then, we use the population of the region as an offset. We construct the adjacency matrix using 5-nearest neighbors based on the centroid of GPS clusters for spatial models. Algorithms are run until the Monte Carlo standard errors (Jones et al., 2006; Flegal et al., 2008) for FSSGLMM are at or below 0.001 (210,000 iterations).

Figure 7 shows the estimates of $\beta(t)$ from different methods. Functional estimates from FGLM and FSSGLMM show a similar trend compared to that from FSGLMM. This is because parameter estimates from FSGLMM can be confounded. For FSSGLMM estimates, we observe double peaks, which confirms high malaria incidence in March and August. Table 4 provides estimates for scalar covariates and τ . We observe that the vegetation index shows a negative relationship with malaria incidence. This suggests that increased urbanization may result in low malaria transmission intensity. In FSSGLMM, water and rainfall variables have a positive relationship with the response variable. The results coincide with a study in Zhou et al. (2012), which shows that the malaria risk was significantly higher in regions surrounding water bodies. The credible intervals obtained from FSGLMM are wider than those from the other two methods; spatial confounding can inflate variance estimates. As an exploratory data analysis, we apply the Moran's I statistics to check whether the residuals are spatially correlated. A significant spatial correlation exists among residuals from FGLM estimates, while the other two models (FSSGLMM and FSGLMM) can account for spatial correlation. FSSGLMM is faster than FSGLMM for this example. Figure 8 compares the observed malaria incidence with Poisson intensity estimates from FSSGLMM; spatial patterns are similar between them.

5.2 US COVID-19 Data

In this section, we aim to model the COVID-19 incidences in 3108 US counties. We use the percentage of male residents and the percentage of black residents as scalar covariates. We also use the population pyramid of each county as a functional covariate; we have 13 age groups (≤ 5 , 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-59, 60-64, 65-74, 75-84, ≥ 85). We expand the age population with 7 Fourier basis functions (Figure 9), with the number of basis selected to minimize GCV error. The data on confirmed COVID-19 cases in US counties as of May 23, 2020, is obtained from Kaggle (<https://www.kaggle.com/c/covid19-global-forecasting-week-5/>), and is turned into Bernoulli data based on whether the county had at least one confirmed case or not. The age pyramid and demographic variables are obtained from the US Census. US

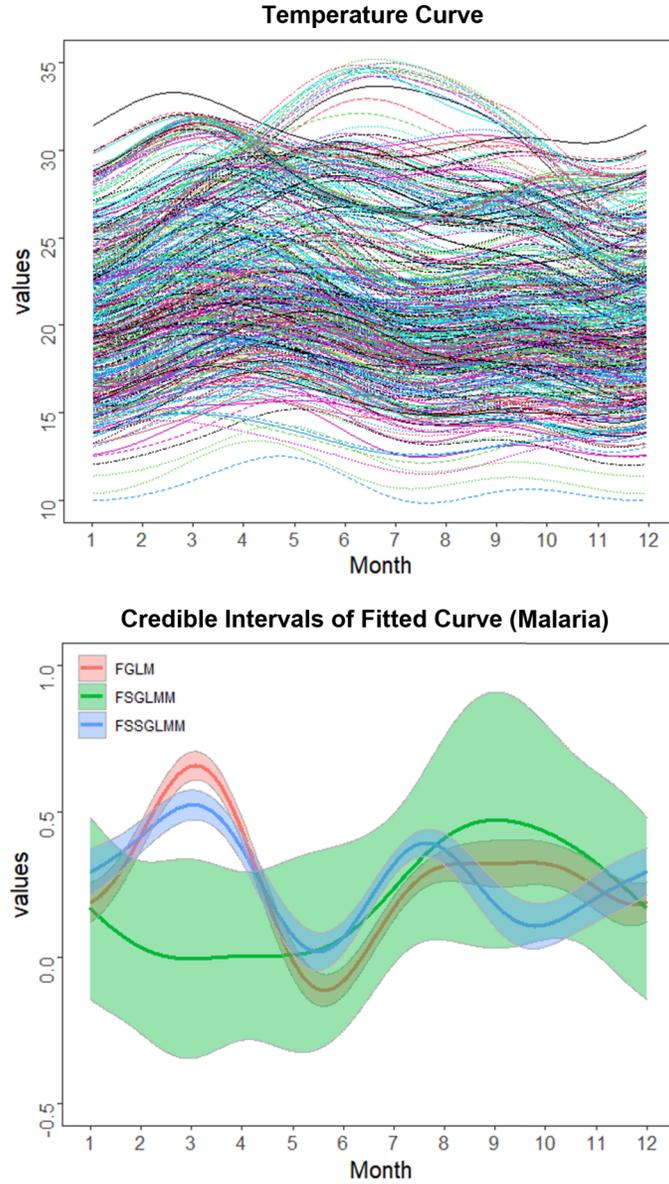


Figure 7: The top panel shows curves of monthly temperature across the 557 GPS clusters of Ethiopia. The bottom panel shows the estimated functional parameter and corresponding 95% credible intervals of each method.

Census provides county adjacency information for identifying which counties are neighbors (<https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html>). We use that file to construct the spatial adjacency matrix in our

FGLM	Vegetation	Water	Rainfall	τ
Mean	-0.427	-0.037	0.044	NA
95%HPD	(-0.483, -0.364)	(-0.081, 0.008)	(-0.027, 0.119)	NA
Moran's I (p-value)	0.488 (0.000)			
Time (min)	0.552			
FSGLMM	Vegetation	Water	Rainfall	τ
Mean	-0.384	0.386	-0.195	0.193
95%HPD	(-0.688, -0.090)	(-0.052, 0.792)	(-0.646, 0.215)	(0.166, 0.222)
Moran's I (p-value)	-0.204 (1.000)			
Time (min)	5.31			
FSSGLMM	Vegetation	Water	Rainfall	τ
Mean	-0.458	0.154	0.183	0.209
95%HPD	(-0.526, -0.393)	(0.110, 0.198)	(0.100, 0.260)	(0.160, 0.258)
Moran's I (p-value)	-0.146 (1.000)			
Time (min)	3.40			

Table 4: Inference results for malaria incidence data using Fourier basis functions. 210,000 posterior samples are generated from each method.

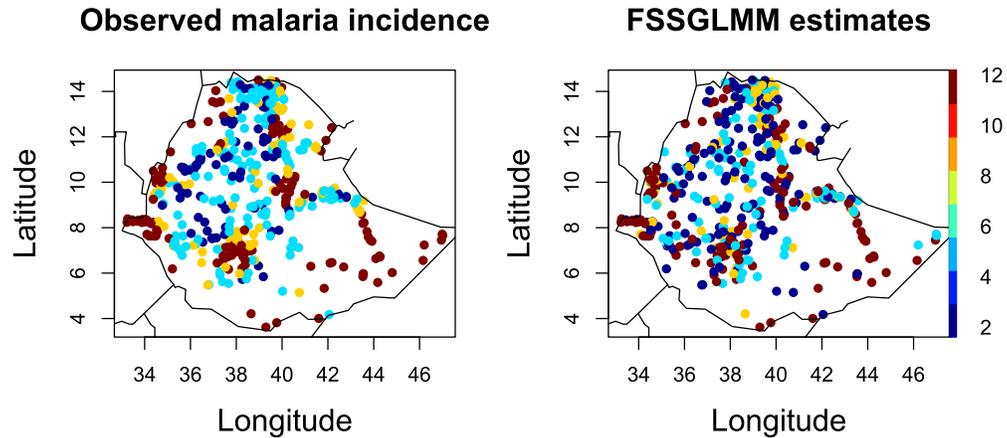


Figure 8: The left panel shows the observed malaria incidence. The right panel shows Poisson intensity estimates from FSSGLMM. Spatial pattern are similar between the estimated and the observed data.

models. MCMC algorithms were run until the Monte Carlo standard errors (Jones et al., 2006; Flegal et al., 2008) for FSSGLMM are at or below 0.01 (330,000 iterations).

Figure 9 illustrates the estimates of $\beta(t)$ from different models. Functional estimates from FSSGLMM and FGLM are similar, while the functional estimate of FSGLMM has a slightly different curve. For estimates from FSSGLMM, we observe high (age group 45-54) and low (age group 75-84) peaks; people in the 45-54 (75-84) age group have a high (low) risk for contracting COVID-19. We note that this is a county-level

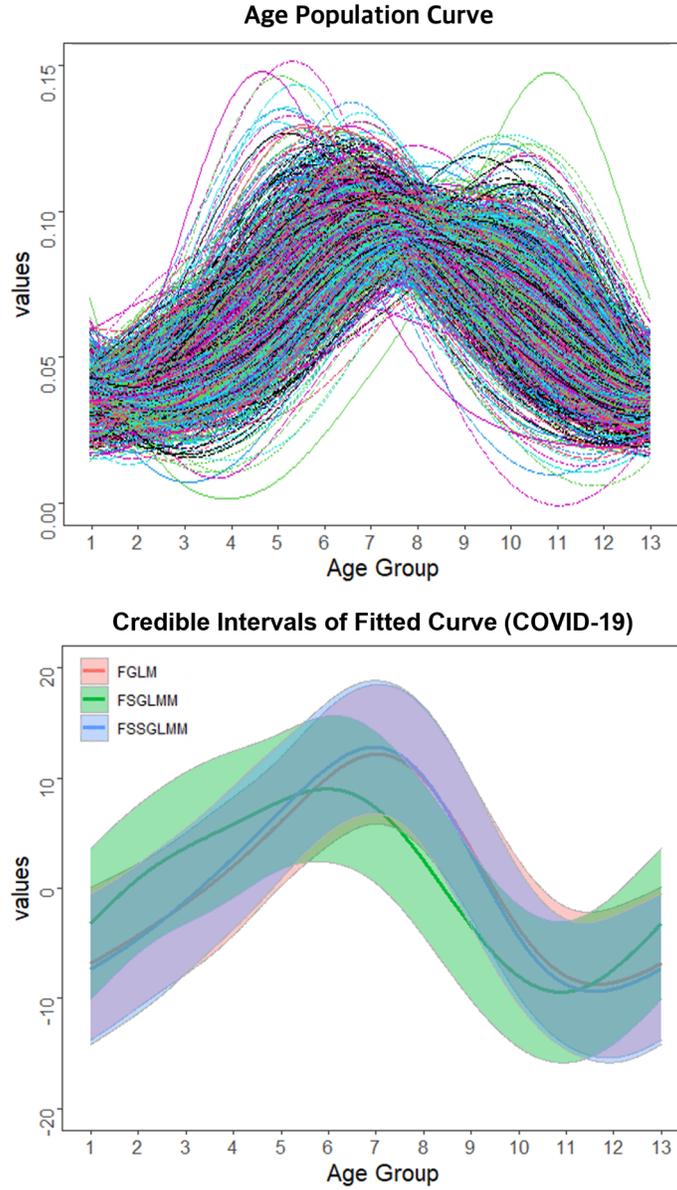


Figure 9: The top panel shows age population curves across the 3108 US counties. The bottom panel shows the estimated functional parameter and corresponding 95% credible intervals of each method.

study; therefore, we cannot interpret this result at the individual level. For instance, we do not imply that an individual in the 75-84 age group has a low risk, but rather that the county with a high population of individuals in the 75-84 age group has a

FGLM	Male	Black	τ
Mean	-13.558	2.383	NA
95%HPD	(-16.700, -10.685)	(1.820, 2.942)	NA
Moran's I (p-value)	0.399 (0.000)		
Time (min)	3.260		
FSGLMM	Male	Black	τ
Mean	-6.279	13.851	0.086
95%HPD	(-10.401, -2.156)	(2.181, 5.568)	(0.052, 0.123)
Moran's I (p-value)	-0.067 (1.000)		
Time (min)	132.073		
FSSGLMM	Male	Black	τ
Mean	-14.957	3.109	0.097
95%HPD	(-18.224, -11.859)	(2.437, 3.769)	(0.067, 0.127)
Moran's I (p-value)	0.069 (0.000)		
Time (min)	22.814		

Table 5: Inference results for US COVID-19 data using Fourier basis functions. 330,000 posterior samples are generated from each method.

low COVID-19 risk. From our functional estimates, we can conclude that having more people in 45-54 (75-84) age group would increase (decrease) the spread of COVID-19. This is reasonable, because people in 45-54 (75-84) age group are likely (less likely) to move for economic/daily activities. Table 5 provides estimates for scalar covariates and τ . We observe that the proportion of males shows a negative relationship, while the proportion of black shows a positive relationship with the risk of contracting COVID-19. Millett et al. (2020) also report that US counties with a higher proportion of blacks had more COVID-19 diagnoses. Due to spatial confounding, the credible intervals from FSGLMM are wider than those from the other two models. Moran's I statistics indicate a significant spatial correlation among residuals from the non-spatial model (FGLM). The Moran's I statistics of FSSGLMM also implies significant spatial correlation. However, we note that the value has greatly decreased compared to Moran's I statistics of FGLM. For this large non-Gaussian spatial data set, FSSGLMM takes about 20 minutes, while FSGLMM takes about 2.2 hours. Figure 10 indicates that there are similar spatial patterns between observed and estimated presence from FSSGLMM.

6 Discussion

In this manuscript, we propose a fast Bayesian functional regression for a non-Gaussian spatial response. Based on a sparse reparameterization (Hughes and Haran, 2013), FSSGLMM is computationally efficient and can alleviate spatial confounding. We have studied the convergence in distribution of functional parameters with increasing n (sample size) and u (MCMC iteration). Furthermore, we show that the estimated regression function obtained from the posterior mean converges to the true function. Through simulated and real data applications, we show that our method provides accurate estimates and can account for spatial correlations. To the best of our knowledge, this is the

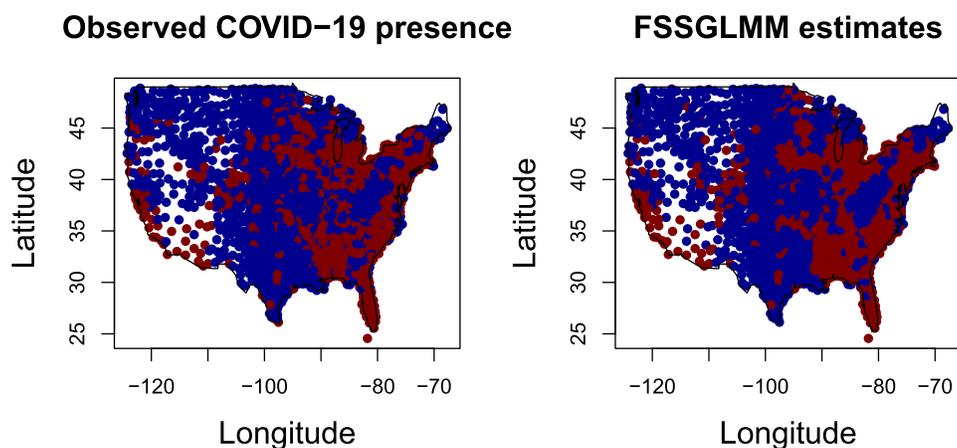


Figure 10: The left panel shows the observed COVID-19 presence at the observation. The right panel shows the estimated presence. Red points denote the COVID-19 presence and blue points denote absence. Spatial patterns are similar between FSSGLM estimates and the observed data.

first attempt to study SoFR with spatially correlated errors and with response variables following under-studied exponential families such as Poisson responses.

We have studied the performance of simultaneous credible intervals of functional parameters. We note that these simultaneous credible intervals are different from simultaneous credible band. Simultaneous credible intervals, often used in spline regression, achieve the average coverage probability of 0.95 whereas simultaneous credible band is expected to achieve coverage probability of 0.95. The frequentist methodologies for confidence bands in functional data analysis are underdeveloped (Liebl and Reimherr, 2019), let alone the studies on utilizing MCMC samples. Simultaneous confidence bands are mostly found using either simulation-based methods (e.g. Bunea et al., 2011; Degras, 2011; Cao et al., 2012) or bootstrap methods (e.g. Cuevas et al., 2006; Chang et al., 2017). Most simulation-based methods often assume the asymptotic normality of functional data in the Banach space and focus on finding the simultaneous confidence band for the mean function whereas the bootstrap methods allow finding confidence bands of other functional estimators such as sample variance and trimmed mean (Cuevas et al., 2006) or coefficient function in function-on-scalar regression (Chang et al., 2017). Here, we provided the simultaneous credible intervals based on the MCMC-sampled functions as in Crainiceanu et al. (2007). We observe that the average coverage probability obtained from FSSGLMM can achieve a nominal rate, while Poisson cases are more affected by the frequency of curves. As we pointed out, peak values of the curves can lead to extreme Poisson responses, resulting in lower coverage.

Although the simultaneous credible intervals can provide useful insights for the true function, they may not be directly applicable to determine the significance of function (i.e., testing whether $\beta(t) = 0$ or not). In order to come up with credible bands that

achieve average probability of $1-\alpha$, we may need to consider a bootstrap-based methods, but it is computationally challenging and requires further studies. We also plan to develop a hypothesis testing method for FSSGLMM to actually determine whether the functional parameter is nonzero or not. Testing for the functional mixed models are studied in Morris and Carroll (2006) and Antoniadis and Sapatinas (2007) but they concern function-on-scalar models with functional responses and scalar covariates, whereas our model stems from scalar-on-function models. Also, as our response is non-Gaussian and our random effects are spatially correlated for FSSGLMM, these further complicate the development of an inference method for FSSGLMM. In addition, since the interest in many applications lies not on testing whether the function is identically 0 or not, but rather on determining whether specific regions or features of the curves are different from 0 (Morris and Carroll, 2006), we plan to develop a hypothesis testing for a given domain of the curve.

In the classical linear models, introducing spatially correlated random effects poses challenges in statistical inference. Unlike the i.i.d. case, the evaluation of likelihood function requires high-dimensional integration with respect to spatial random effects. Especially for non-Gaussian responses we consider in this manuscript, such integration is intractable; therefore, we need to construct high-dimensional joint posterior with model parameters and random effects. The same issues arise in functional spatial models because we represent functional terms with a finite basis expansion, which turn into the classical linear model form. Therefore, the slow mixing of MCMC for high-dimensional posteriors and spatial confounding can lead to unreliable basis coefficient estimates, resulting in difficulties in recovering true curves compared to the i.i.d case. Our method can alleviate such issues by adapting sparse reparameterization (Hughes and Haran, 2013) and show improved performance compared to other competitors. Furthermore, the complexity of the functional parameter adds another layer of difficulty in estimation. If the functional parameter is of high frequency, it can require more basis functions, resulting in higher dimensional posteriors. In our method, we choose the best number of basis based on cross-validation for our functional predictor $X(t)$ and use the same basis functions in estimation for functional parameter $\beta(t)$. If the functional predictors are of lower frequency or of less complexity, then the resulting expansion using these bases may not lead to a close estimation.

Our method can incorporate functional variables into a spatial regression framework without information loss. Variables such as temperatures or growth curves are innately functional data. Instead of using a numeric summary (e.g., mean, quantiles) of these functional data, our model uses the full information. This can provide a time-varying (functional domain-specific) interpretation of spatial covariates. Furthermore, analyzing such variables within the multivariate statistics framework cannot adjust for the correlation within each sample (function).

A great deal of spatial functional data has been collected and analyzed in scientific studies. We focus on studying epidemiological data sets, but the method we have developed can be broadly applicable to data sets in different disciplines. Examples include the brain signals (frequency curves) which are observed over brain voxels (Musgrove et al., 2016), predicting the risk of drought (Escabias et al., 2005) using atmospheric

functional variables, and identifying the gene and time-varying environment exposure interaction of human disease (Wei et al., 2014).

Supplementary Material

Supplementary Material of “Fast Bayesian Functional Regression for Non-Gaussian Spatial Data” (DOI: [0.1214/22-BA1354SUPP](https://doi.org/10.1214/22-BA1354SUPP); .pdf). Supplementary material available online contains the results of simulation studies under different basis functions, coverage for simulated datasets with varying ranks, different frequencies of the true functional parameter, zero-inflated Poisson and negative binomial responses, and weak spatial dependence setting. It also provides MCMC diagnostics for real data examples and sensitivity analysis of different priors.

References

- Antoniadis, A. and Sapatinas, T. (2007). “Estimation and inference in functional mixed-effects models.” *Computational Statistics & Data Analysis*, 51(10): 4793–4813. MR2364541. doi: <https://doi.org/10.1016/j.csda.2006.09.038>. 408, 430
- Aw, A. and Cabral, E. N. (2020). “Bayesian estimation of the functional spatial lag model.” *Journal of Time Series Econometrics*, 1(ahead-of-print). MR4137589. doi: <https://doi.org/10.1515/jtse-2019-0047>. 409
- Besag, J. (1974). “Spatial interaction and the statistical analysis of lattice systems.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 36: 192–236. MR0373208. 409, 410
- Besag, J., York, J., and Mollié, A. (1991). “Bayesian image restoration, with two applications in spatial statistics.” *Annals of the Institute of Statistical Mathematics*, 43(1): 1–20. MR1105822. doi: <https://doi.org/10.1007/BF00116466>. 409
- Boots, B. and Tiefelsdorf, M. (2000). “Global and local spatial autocorrelation in bounded regular tessellations.” *Journal of Geographical Systems*, 2(4): 319–348. 412
- Bradley, J. R., Holan, S. H., and Wikle, C. K. (2015). “Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics.” *The Annals of Applied Statistics*, 9(4): 1761–1791. MR3456353. doi: <https://doi.org/10.1214/15-AOAS862>. 412
- Bradley, J. R., Holan, S. H., and Wikle, C. K. (2020). “Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family.” *Journal of the American Statistical Association*, 115(532): 2037–2052. MR4189775. doi: <https://doi.org/10.1080/01621459.2019.1677471>. 413
- Bradley, J. R., Holan, S. H., Wikle, C. K., et al. (2018). “Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion).” *Bayesian Analysis*, 13(1): 253–310. MR3773410. doi: <https://doi.org/10.1214/17-BA1069>. 413

- Bueno-Larraz, B., Berrendero, J. R., and Cuevas, A. (2018). “On functional logistic regression via RKHS’s.” *arXiv preprint arXiv:1812.00721*. 409
- Bunea, F., Ivanescu, A. E., and Wegkamp, M. H. (2011). “Adaptive inference for the mean of a Gaussian process in functional data.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4): 531–558. MR2853729. doi: <https://doi.org/10.1111/j.1467-9868.2010.00768.x>. 429
- Cai, T. T., Hall, P., et al. (2006). “Prediction in functional linear regression.” *The Annals of Statistics*, 34(5): 2159–2179. MR2291496. doi: <https://doi.org/10.1214/009053606000000830>. 408
- Cao, G., Yang, L., and Todem, D. (2012). “Simultaneous inference for the mean function based on dense functional data.” *Journal of Nonparametric Statistics*, 24(2): 359–377. MR2921141. doi: <https://doi.org/10.1080/10485252.2011.638071>. 429
- Cardot, H., Ferraty, F., and Sarda, P. (1999). “Functional linear model.” *Statistics & Probability Letters*, 45(1): 11–22. MR1718346. doi: [https://doi.org/10.1016/S0167-7152\(99\)00036-X](https://doi.org/10.1016/S0167-7152(99)00036-X). 408
- Chang, C., Lin, X., and Ogden, R. T. (2017). “Simultaneous confidence bands for functional regression models.” *Journal of Statistical Planning and Inference*, 188: 67–81. MR3648318. doi: <https://doi.org/10.1016/j.jspi.2017.03.002>. 429
- Chen, H. and Wang, Y. (2011). “A penalized spline approach to functional mixed effects model analysis.” *Biometrics*, 67(3): 861–870. MR2829260. doi: <https://doi.org/10.1111/j.1541-0420.2010.01524.x>. 408
- Chen, K. and Müller, H. (2012). “Conditional quantile analysis when covariates are functions, with application to growth data.” *Journal of the Royal Statistical Society, Series B*, 74: 67–89. MR2885840. doi: <https://doi.org/10.1111/j.1467-9868.2011.01008.x>. 407
- Christensen, O. F. (2004). “Monte Carlo maximum likelihood in model-based geostatistics.” *Journal of Computational and Graphical Statistics*, 13(3): 702–718. MR2087723. doi: <https://doi.org/10.1198/106186004X2525>. 413
- Christensen, O. F., Møller, J., and Waagepetersen, R. P. (2001). “Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models.” *Methodology and Computing in Applied Probability*, 3(3): 309–327. MR1891114. doi: <https://doi.org/10.1023/A:1013779208892>. 415, 416
- Christensen, O. F., Roberts, G. O., and Sköld, M. (2006). “Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models.” *Journal of Computational and Graphical Statistics*, 15(1): 1–17. MR2269360. doi: <https://doi.org/10.1198/106186006X100470>. 413
- Christensen, O. F. and Waagepetersen, R. (2002). “Bayesian prediction of spatial count data using generalized linear mixed models.” *Biometrics*, 58(2): 280–286. MR1908167. doi: <https://doi.org/10.1111/j.0006-341X.2002.00280.x>. 413
- Cowles, M. K. (2002). “MCMC sampler convergence rates for hierarchical normal

- linear models: A simulation approach.” *Statistics and Computing*, 12(4): 377–389. MR1951710. doi: <https://doi.org/10.1023/A:1020700515208>. 416
- Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). “Spatially adaptive Bayesian penalized splines with heteroscedastic errors.” *Journal of Computational and Graphical Statistics*, 16(2): 265–288. MR2370943. doi: <https://doi.org/10.1198/106186007X208768>. 416, 429
- Crambes, C., Kneip, A., Sarda, P., et al. (2009). “Smoothing splines estimators for functional linear regression.” *The Annals of Statistics*, 37(1): 35–72. MR2488344. doi: <https://doi.org/10.1214/07-AOS563>. 408
- Cuevas, A., Febrero, M., and Fraiman, R. (2006). “On the use of the bootstrap for estimating functions with functional data.” *Computational Statistics & Data Analysis*, 51(2): 1063–1074. MR2297507. doi: <https://doi.org/10.1016/j.csda.2005.10.012>. 429
- Degras, D. A. (2011). “Simultaneous confidence bands for nonparametric regression with functional data.” *Statistica Sinica*, 1735–1765. MR2895997. doi: <https://doi.org/10.5705/ss.2009.207>. 429
- Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). “Statistics for spatial functional data: some recent contributions.” *Environmetrics: The Official Journal of the International Environmetrics Society*, 21(3-4): 224–239. MR2842240. doi: <https://doi.org/10.1002/env.1003>. 408
- Escabias, M., Aguilera, A., and Valderrama, M. (2005). “Modeling environmental data by functional principal Component Logistic Regression.” *Environmetrics: The Official Journal of the International Environmetrics Society*, 16(1): 95–107. MR2146901. doi: <https://doi.org/10.1002/env.696>. 430
- Ferreira, M. A., Porter, E. M., and Franck, C. T. (2021). “Fast and scalable computations for Gaussian hierarchical models with intrinsic conditional autoregressive spatial random effects.” *Computational Statistics & Data Analysis*, 162: 107264. MR4256170. doi: <https://doi.org/10.1016/j.csda.2021.107264>. 413
- Flegal, J. M., Haran, M., and Jones, G. L. (2008). “Markov chain Monte Carlo: Can we trust the third significant figure?” *Statistical Science*, 23: 250–260. MR2516823. doi: <https://doi.org/10.1214/08-STS257>. 417, 424, 426
- Gao, B. and Kaufman, Y. (2015). “MODIS atmosphere L2 water vapor product.” *NASA MODIS adaptive processing system, Goddard Space Flight Center, USA*. 410
- Gertheiss, J., Maity, A., and Staicu, A.-M. (2013). “Variable selection in generalized functional linear models.” *Stat*, 2(1): 86–101. MR4027303. doi: <https://doi.org/10.1002/sta4.20>. 408
- Getis, A. and Griffith, D. A. (2002). “Comparative spatial filtering in regression analysis.” *Geographical Analysis*, 34(2): 130–140. 412
- Goldsmith, J. and Scheipl, F. (2014). “Estimator selection and combination in scalar-

- on-function regression." *Computational Statistics & Data Analysis*, 70: 362–372. MR3125499. doi: <https://doi.org/10.1016/j.csda.2013.10.009>. 408
- Gopal, S., Ma, Y., Xin, C., Pitts, J., and Were, L. (2019). "Characterizing the spatial determinants and prevention of malaria in Kenya." *International Journal of Environmental Research and Public Health*, 16(24): 5078. 410, 424
- Griffith, D. A. (2000). "A linear regression solution to the spatial autocorrelation problem." *Journal of Geographical Systems*, 2(2): 141–156. 412
- Griffith, D. A. (2002). "A spatial filtering specification for the auto-Poisson model." *Statistics & Probability Letters*, 58(3): 245–251. MR1920751. doi: [https://doi.org/10.1016/S0167-7152\(02\)00099-8](https://doi.org/10.1016/S0167-7152(02)00099-8). 412
- Griffith, D. A. (2004). "A spatial filtering specification for the autologistic model." *Environment and Planning A*, 36(10): 1791–1811. 412
- Guillas, S. and Lai, M.-J. (2010). "Bivariate splines for spatial functional regression models." *Journal of Nonparametric Statistics*, 22(4): 477–497. MR2662608. doi: <https://doi.org/10.1080/10485250903323180>. 408
- Guo, W. (2002). "Functional mixed effects models." *Biometrics*, 58(1): 121–128. MR1891050. doi: <https://doi.org/10.1111/j.0006-341X.2002.00121.x>. 408
- Hall, P., Horowitz, J. L., et al. (2007). "Methodology and convergence rates for functional linear regression." *The Annals of Statistics*, 35(1): 70–91. MR2332269. doi: <https://doi.org/10.1214/009053606000000957>. 408
- Hanks, E. M., Schliep, E. M., Hooten, M. B., and Hoeting, J. A. (2015). "Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification." *Environmetrics*, 26(4): 243–254. MR3340961. doi: <https://doi.org/10.1002/env.2331>. 409, 411
- Heaton, M. J., Ingersoll, C., Berrett, C., and Hartman, B. (2020). "A Bayesian Approach to Real-time Spatiotemporal Prediction Systems for Respiratory Syncytial Virus." *Biometrics*, 1: 16. 412, 414
- Hodges, J. S. and Reich, B. J. (2010). "Adding spatially-correlated errors can mess up the fixed effect you love." *The American Statistician*, 64(4): 325–334. MR2758564. doi: <https://doi.org/10.1198/tast.2010.10052>. 411
- Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. New York: Springer. MR2920735. doi: <https://doi.org/10.1007/978-1-4614-3655-3>. 407
- Huang, T., Saporta, G., Wang, H., and Wang, S. (2020). "A robust spatial autoregressive scalar-on-function regression with t-distribution." *Advances in Data Analysis and Classification*, 1–25. MR4242479. doi: <https://doi.org/10.1007/s11634-020-00384-w>. 409
- Hughes, J. and Haran, M. (2013). "Dimension reduction and alleviation of confounding for spatial generalized linear mixed models." *Journal of the Royal Statistical Society:*

- Series B (Statistical Methodology)*, 75(1): 139–159. MR3008275. doi: <https://doi.org/10.1111/j.1467-9868.2012.01041.x>. 408, 409, 411, 412, 413, 414, 417, 428, 430
- ICF (2004–2017 (Accessed July, 1, 2020)). “Demographic and Health Surveys (various [Datasets].” Funded by USAID. Data retrieved from, <http://dhsprogram.com/data/available-datasets.cfm>. 423
- James, G. M. (2002). “Generalized linear models with functional predictors.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3): 411–432. MR1924298. doi: <https://doi.org/10.1111/1467-9868.00342>. 408
- James, G. M., Wang, J., Zhu, J., et al. (2009). “Functional linear regression that’s interpretable.” *The Annals of Statistics*, 37(5A): 2083–2108. MR2543686. doi: <https://doi.org/10.1214/08-AOS641>. 408
- Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). “Fixed-width output analysis for Markov chain Monte Carlo.” *Journal of the American Statistical Association*, 101(476): 1537–1547. MR2279478. doi: <https://doi.org/10.1198/016214506000000492>. 417, 424, 426
- Kang, H. B., Jung, Y. J., and Park, J. (2023). “Supplementary Material for “Fast Bayesian functional regression for non-Gaussian spatial data.”” *Bayesian Analysis*. doi: <https://doi.org/10.1214/22-BA1354SUPP>. 413
- Keefe, M. J., Ferreira, M. A., and Franck, C. T. (2019). “Objective Bayesian analysis for Gaussian hierarchical models with intrinsic conditional autoregressive priors.” *Bayesian Analysis*, 14(1): 181–209. MR3910043. doi: <https://doi.org/10.1214/18-BA1107>. 413
- Kokoszka, P. and Reimherr, M. (2012). “Determining the order of the functional autoregressive model.” *Journal of Time Series Analysis*, 34: 116–129. MR3008019. doi: <https://doi.org/10.1111/j.1467-9892.2012.00816.x>. 407
- Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press. 407
- Liebl, D. and Reimherr, M. (2019). “Fast and Fair Simultaneous Confidence Bands for Functional Parameters.” *arXiv preprint arXiv:1910.00131*. MR4365792. doi: https://doi.org/10.1007/978-3-030-47756-1_21. 429
- Lindgren, F., Rue, H., and 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. MR2853727. doi: <https://doi.org/10.1111/j.1467-9868.2011.00777.x>. 413
- Ma, W., Xiao, L., Liu, B., and Lindquist, M. A. (2021). “A functional mixed model for scalar on function regression with application to a functional MRI study.” *Biostatistics*, 22(3): 439–454. MR4287162. doi: <https://doi.org/10.1093/biostatistics/kxz046>. 408

- Martínez-Hernández, I., Genton, M. G., et al. (2020). “Recent developments in complex and spatially correlated functional data.” *Brazilian Journal of Probability and Statistics*, 34(2): 204–229. MR4093256. doi: <https://doi.org/10.1214/20-BJPS466>. 408
- McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., and Ruppert, D. (2014). “Functional generalized additive models.” *Journal of Computational and Graphical Statistics*, 23(1): 249–269. MR3173770. doi: <https://doi.org/10.1080/10618600.2012.729985>. 408
- Mengersen, K. L. and Tweedie, R. L. (1996). “Rates of convergence of the Hastings and Metropolis algorithms.” *The Annals of Statistics*, 24(1): 101–121. MR1389882. doi: <https://doi.org/10.1214/aos/1033066201>. 415
- Millett, G. A., Jones, A. T., Benkeser, D., Baral, S., Mercer, L., Beyrer, C., Honermann, B., Lankiewicz, E., Mena, L., Crowley, J. S., et al. (2020). “Assessing differential impacts of COVID-19 on Black communities.” *Annals of Epidemiology*. 428
- Moran, P. A. (1950). “Notes on continuous stochastic phenomena.” *Biometrika*, 37(1/2): 17–23. MR0035933. doi: <https://doi.org/10.1093/biomet/37.1-2.17>. 412, 418
- Morris, J. S. (2015). “Functional regression.” *Annual Review of Statistics and Its Application*, 2: 321–359. 408
- Morris, J. S. and Carroll, R. J. (2006). “Wavelet-based functional mixed models.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2): 179–199. MR2188981. doi: <https://doi.org/10.1111/j.1467-9868.2006.00539.x>. 408, 430
- Mousavi, S. N. and Sørensen, H. (2018). “Functional logistic regression: a comparison of three methods.” *Journal of Statistical Computation and Simulation*, 88(2): 250–268. MR3740728. doi: <https://doi.org/10.1080/00949655.2017.1386664>. 408, 409
- Müller, H.-G. and Stadtmüller, U. (2005). “Generalized functional linear models.” *The Annals of Statistics*, 33(2): 774–805. MR2163159. doi: <https://doi.org/10.1214/009053604000001156>. 408, 411
- Musgrove, D. R., Hughes, J., and Eberly, L. E. (2016). “Fast, fully Bayesian spatiotemporal inference for fMRI data.” *Biostatistics*, 17(2): 291–303. MR3516001. doi: <https://doi.org/10.1093/biostatistics/kxv044>. 412, 414, 430
- Pineda-Ríos, W., Giraldo, R., and Porcu, E. (2019). “Functional SAR models: With application to spatial econometrics.” *Spatial Statistics*, 29: 145–159. MR3895110. doi: <https://doi.org/10.1016/j.spasta.2018.12.002>. 409, 410
- Polson, N. G., Scott, J. G., and Windle, J. (2013). “Bayesian inference for logistic models using Pólya–Gamma latent variables.” *Journal of the American Statistical Association*, 108(504): 1339–1349. MR3174712. doi: <https://doi.org/10.1080/01621459.2013.829001>. 413
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer. MR2168993. 407

- Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). “Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models.” *Biometrics*, 62(4): 1197–1206. MR2307445. doi: <https://doi.org/10.1111/j.1541-0420.2006.00617.x>. 409, 411
- Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017). “Methods for scalar-on-function regression.” *International Statistical Review*, 85(2): 228–249. MR3686566. doi: <https://doi.org/10.1111/insr.12163>. 408
- Reiss, P. T. and Ogden, R. T. (2010). “Functional generalized linear models with images as predictors.” *Biometrics*, 66(1): 61–69. MR2756691. doi: <https://doi.org/10.1111/j.1541-0420.2009.01233.x>. 408
- Rosenthal, J. S. (1995). “Minorization conditions and convergence rates for Markov chain Monte Carlo.” *Journal of the American Statistical Association*, 90(430): 558–566. MR1340509. 416
- Rue, H., Martino, S., and Chopin, N. (2009). “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.” *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2): 319–392. MR2649602. doi: <https://doi.org/10.1111/j.1467-9868.2008.00700.x>. 413
- Ruiz-Medina, M. (2012). “Spatial functional prediction from spatial autoregressive Hilbertian processes.” *Environmetrics*, 23(1): 119–128. MR2873789. doi: <https://doi.org/10.1002/env.1143>. 408
- Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media. MR1697409. doi: <https://doi.org/10.1007/978-1-4612-1494-6>. 417
- Tierney, L. (1994). “Markov chains for exploring posterior distributions.” *the Annals of Statistics*, 1701–1728. MR1329166. doi: <https://doi.org/10.1214/aos/1176325750>. 415
- Tong, X. T. and Van Handel, R. (2012). “Ergodicity and stability of the conditional distributions of nondegenerate Markov chains.” *The Annals of Applied Probability*, 22(4): 1495–1540. MR2985168. doi: <https://doi.org/10.1214/11-AAP800>. 415
- Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). “Functional data analysis.” *Annual Review of Statistics and Its Application*, 3: 257–295. 407
- Wang, X., Nan, B., Zhu, J., Koeppel, R., and Frey, K. (2017). “Classification of ADNI PET images via regularized 3D functional data analysis.” *Biostatistics & epidemiology*, 1(1): 3–19. 409
- Wei, P., Tang, H., and Li, D. (2014). “Functional logistic regression approach to detecting gene by longitudinal environmental exposure interaction in a case-control study.” *Genetic epidemiology*, 38(7): 638–651. 431
- Zhang, J. T. and Chen, J. (2007). “Statistical inferences for functional data.” *The Annals of Statistics*, 35(3): 1052–1079. MR2341698. doi: <https://doi.org/10.1214/009053606000001505>. 407

- Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A., and Morris, J. S. (2016). “Functional CAR models for large spatially correlated functional datasets.” *Journal of the American Statistical Association*, 111(514): 772–786. MR3538704. doi: <https://doi.org/10.1080/01621459.2015.1042581>. 409
- Zhou, S.-s., Zhang, S.-s., Wang, J.-j., Zheng, X., Huang, F., Li, W.-d., Xu, X., and Zhang, H.-w. (2012). “Spatial correlation between malaria cases and water-bodies in Anopheles sinensis dominated areas of Huang-Huai plain, China.” *Parasites & vectors*, 5(1): 1–7. 424
- Zhu, H., Brown, P. J., and Morris, J. S. (2011). “Robust, adaptive functional regression in functional mixed model framework.” *Journal of the American Statistical Association*, 106(495): 1167–1179. MR2894772. doi: <https://doi.org/10.1198/jasa.2011.tm10370>. 408

Acknowledgments

The authors are grateful to anonymous reviewers for their careful reading and valuable comments.