A Projection-based Laplace Approximation for Spatial Latent Variable Models

Jaewoo Park[1,2] and Sangwan Lee[1]

[1]Department of Statistics and Data Science, Yonsei University

[2]Department of Applied Statistics, Yonsei University

June 24, 2021

Abstract

Laplace method is a practical tool for obtaining maximum likelihood estimators for a broad class of latent variable models. The main idea is to approximate the integrand using a Gaussian distribution. However, with increasing observations, the Laplace approximation becomes infeasible because the dimension of the correlated latent variables grows, which results in the high-dimensional optimization problem. One important example is spatial latent variable models, which are widely used in many fields, such as ecology, epidemiology, and sociology. Spatial latent variable models are useful for investigating the relationship between spatial covariates or predicting the unobserved area. Here, we propose a fast Laplace approximation based on the dimension reduction of the latent variables. Our methods are faster and have fewer components to be tuned than simulation-based meth-

ods such as Markov chain Monte Carlo maximum likelihood and Monte Carlo expectation-maximization. Our approach can be applied to the large non-Gaussian spatial data sets, commonly used in modern environmental sciences. Especially, we show how we may understand spatial patterns of non-Gaussian responses for two case studies: confirmed COVID-19 cases in the United States and thickness of the Antarctic ice sheet. Through simulation studies under different scenarios, we investigate that our method can provide accurate maximum likelihood estimations and predictions quickly. Our study can be broadly applicable for practical maximum likelihood inference for high-dimensional random effect models.

Keywords: Laplace method; numerical optimization; dimension reduction; spatial interpolation; spatial generalized linear mixed models

1 Introduction

Latent variable models are useful when the underlying structures of interest are unobservable. Laplace methods (Tierney and Kadane, 1986) are practical tools for maximum likelihood estimation for such a class of models, such as longitudinal models (cf. Molenberghs and Verbeke, 2005) and generalized linear latent variable models (cf. Bianconcini and Cagnone, 2012). The likelihood evaluations for such models require integration with respect to latent variables. Laplace methods approximate the integrals through the Gaussian distribution by finding the optimized values of the latent variables for the given model parameters. However, when the dimension of the latent variables grows, they may be computationally infeasible. One such case is spatial latent variable models (or spatial generalized linear mixed models, SGLMMs), which are widely used to study observed data that have associated spatial locations (cf. Banerjee et al., 2014, Lawson et al., 2016). The maximum likelihood estimation of such models is challenging because the evaluation of the likelihood function requires high-dimensional integration as well as spatially correlated latent variables. These facts result in a high-dimensional optimization problem for the Laplace approximation, which is computationally infeasible. Here, we develop a practical Laplace approximation for high-dimensional spatial latent variable models. We replace high-dimensional integration with a reduced-dimensional representation based on the projection and provide a nested optimization method to find the maximum likelihood estimators (MLEs). We show that our approach is efficient at producing accurate estimations and predictions. We apply our method to two important environmental science problems and draw meaningful conclusions based on inference results.

The Bayesian framework is a popular approach for spatial latent variable models and

a large literature addresses the computational challenges in the Bayesian context. For instance, the predictive process approach (Banerjee et al., 2008) projects the original high-dimensional process onto a low-dimensional subspace at a specified set of locations. Nearest-neighbor Gaussian process approaches (Datta et al., 2016, Finley et al., 2019) can acheive scalability with sparse representation of precision matrices. (see Zhang et al. (2021) for multivariate extensions). To improve the mixing of Markov chain Monte Carlo (MCMC), several reparameterizations (cf. Rue and Held, 2005, Christensen et al., 2006) have been suggested. Combining such reparameterizations with an efficient subsampling strategy, Entezari et al. (2020) improve predictions of SGLMMs. Rue et al. (2009) develops the integrated nested laplace approximation (INLA), which allows practical Bayesian inference for non-Gaussian spatial data using a stochastic partial differential equation approximation to the latent Gaussian model. INLA has been widely used in many spatial and spatio-temporal applications with large number of observations (cf. Clifford et al., 2019, Forlani et al., 2020). Hughes and Haran (2013), Guan and Haran (2018) propose efficient projection-based approaches to address computational and inferential challenges under the Bayesian framework.

In this manuscript, we focus on developing a practical maximum likelihood approach for high-dimensional spatial latent variable models based on the Laplace method. Bayes approaches are convenient for constructing hierarchical spatial models. For instance, we can consider flexible link functions (Li et al., 2019) or incorporate spatially varying coefficient (Berrett et al., 2020) for spatial latent variable models. On the contrary, frequentist methods may provide faster alternatives and can be convenient by avoiding the need to tune MCMC algorithms. Several maximum likelihood approaches have been developed for spatial latent variable models. Zhang (2002) developed Monte Carlo expectation maximization (MCEM), which uses Monte the Carlo version of the

conditional expectation in the E-step. Recently, Tadayon and Torabi (2019) propose MCEM for non-Gaussian spatial data with covariatemeasurement error. Christensen (2004) proposed Markov chain Monte Carlo maximum likelihood (MCML), which maximizes the Monte Carlo approximated log-likelihood function. However, these approaches are computationally infeasible for large data sets because Monte Carlo approximations require simulating high-dimensional latent variables. Instead of simulation-based methods, Evangelou et al. (2011), Bonat and Ribeiro (2016) provide the Laplace approximation for the conditional distribution of the latent variables. These approaches approximate the integration in the likelihood function through a Gaussian distribution, which require optimized values of the latent variables. For large data sets, such optimization becomes computationally demanding or even unstable. Recently, to address such challenges, Guan and Haran (2019), Park and Haran (2020) develop the projection-based MCEM and MCML, respectively. Although their approaches are computationally feasible for large non-Gaussian spatial data sets, both require several components to be tuned. For instance, users need to decide the number of Monte Carlo samples as well as to check the convergence of their iterative approaches manually. Furthermore, obtaining accurate standard errors is challenging in practice.

The main contribution of this manuscript is to develop a projection-based Laplace method that is computationally efficient as well as easy to implement for practitioners. Our algorithms can be applied to both continuous and lattice spatial domains. Especially, we focus on analyzing two scientific data sets: (1) *US COVID-19 data* contain a number of confirmed cases from US counties. Analyzing the spatial patterns of confirmed cases is important for detecting disease hotspots, which is useful for issuing public health advisories. (2) *Antarctic ice sheet data* represent binary ice thickness (ice-no ice) patterns. Understanding ice thickness distribution is crucial for informative projec-

tions of global sea-level rise. For both data sets, we have thousands of non-Gaussian spatial observations; inference for such data sets is computationally challenging. The proposed method allows us to carry out accurate maximum likelihood inference quickly. The outline of the rest of the manuscript is as follows. In Section 2, we introduce the Laplace approximations for spatial latent variable models and discuss their computational challenges. In Section 3, we propose the projection-based Laplace approximation to maximum likelihood inference and describe the implementation details. In Section 4, we present the estimation and prediction performance for our method using simulation studies. In Section 5, we show the application of our methods to real data sets with large observations. We conclude with a discussion and summary in Section 6.

2 Spatial Latent Variable Models

The Spatial generalized linear mixed models (SGLMMs) consist of the fixed effect for explaining the mean trend and random effect for capturing the spatial correlation. SGLMMs have been developed both for point-referenced data in a continuous domain (Diggle et al., 1998) and for areal data in a lattice domain (Besag, 1974). Let $\mathbf{Z}$ be the n-dimensional observations at locations $\mathbf{s} \in R^{n\times 2}$. $\mathbf{X} \in R^{n\times p}$ is the matrix of spatial covariates depending on the locations. At each coordinate we can consider spatially correlated random effects $\mathbf{W} \in R^n$. Depending on the spatial domain of interest, random effects are modeled differently. For a continuous domain, $\mathbf{W}$ is often modeled as a second-order stationary Gaussian process with mean zero and covariance kernel $\mathbf{K}_{\boldsymbol{\theta}}(\mathbf{h}) = \mathbf{K}_{\boldsymbol{\theta}}(W(\mathbf{s}+\mathbf{h}), W(\mathbf{s}))$. For instance, with the parameter $\boldsymbol{\theta} = (\sigma^2, \phi)$, the Matérn class (Stein, 2012) covariance kernel is widely used, where σ^2 controls the overall variance of the process, and ϕ determines the range of spatial correlation.

Let $\mathbf{C}_\theta = \sigma^2\mathbf{R}_\phi$ be the $n \times n$ covariance matrix obtained from the kernel $\mathbf{K}_\theta(\mathbf{h})$ defined above. Here $\mathbf{R}_\phi$ is a positive definite correlation matrix. Then the distribution of $\mathbf{W}$ is $f_W(\mathbf{W}|\boldsymbol{\theta}) \propto |\mathbf{C}_\theta|^{-1/2} \exp(-\frac{1}{2}\mathbf{W}^\top\mathbf{C}_\theta^{-1}\mathbf{W})$. To account for measurement errors, the independent nugget effect can be added into this model. For a lattice domain, $\mathbf{W}$ follows a Gaussian Markov random field (GMRF) (Besag, 1974) with mean zero. GMRF is defined by its neighbor structure using the $n \times n$ adjacency matrix $\mathbf{A}$: $A_{i,j} = 1$ if the ith location and jth location are neighbors, and 0 otherwise. Then, $f_W(\mathbf{W}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\text{rank}(\mathbf{Q})/2} \exp(-\frac{\theta}{2}\mathbf{W}^\top\mathbf{Q}\mathbf{W})$, where $\mathbf{Q}$ =diag$(\mathbf{A}\mathbf{1}) - \mathbf{A}$ and $\boldsymbol{\theta}$ determines the smoothness of the process.

With the spatial random effects defined above, we can write a link function as

$$g\{E[\mathbf{Z}|\mathbf{W}, \boldsymbol{\beta}]\} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}. \tag{1}$$

Here, the observed response $\mathbf{Z}$ is conditionally independent for the given $\mathbf{W}$ and $\boldsymbol{\beta}$. We assume that the conditional distribution of $\mathbf{Z}$ is from the exponential family as

$$f_{\mathbf{Z}|\mathbf{W}}(\mathbf{Z}|\mathbf{W}, \boldsymbol{\beta}) = \exp(\mathbf{Z}^\top(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}) - \mathbf{1}^\top b(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}) + \mathbf{1}^\top c(\mathbf{Z})), \tag{2}$$

where $b(\cdot), c(\cdot)$ are known. With the model parameters $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$, the likelihood function can be written as

$$L(\boldsymbol{\psi}|\mathbf{Z}) = \int_{R^n} f_{\mathbf{Z}|\mathbf{W}}(\mathbf{Z}|\mathbf{W}, \boldsymbol{\beta}) f_{\mathbf{W}}(\mathbf{W}|\boldsymbol{\theta}) d\mathbf{W}. \tag{3}$$

The number of random effects grows with the size of the data, which makes evaluating the likelihood function difficult. Therefore, the direct maximization of (3) is infeasible.

2.1 Laplace Approximation and Computational Challenges

Laplace approximations have been studied for maximum likelihood inference of SGLMMs. Evangelou et al. (2011), Bonat and Ribeiro (2016) use the Laplace method, which approximates the integrand using a Gaussian distribution. These works are closely related to the idea of INLA (Rue et al., 2009) in Bayes approaches. Let $Q(\mathbf{W})$ be the sum of the log-likelihood from exponential families and the multivariate normal log-likelihood for the latent variable $\mathbf{W}$. $Q(\mathbf{W})$ is a unimodal and bounded function of n-dimensional random effects $\mathbf{W}$. Then, using the Laplace method, the likelihood function in (3) can be approximated as

$$L(\psi|\mathbf{Z}) = \int_{R^n} \exp\left(Q(\mathbf{W})\right) d\mathbf{W} \approx (2\pi)^{n/2} \left| -\frac{\partial^2 Q(\widehat{\mathbf{W}})}{\partial \mathbf{W} \partial \mathbf{W}^\intercal} \right|^{-1/2} \exp\left(Q(\widehat{W})\right), \tag{4}$$

where $\widehat{\mathbf{W}}$ is the value that maximizes $Q(\mathbf{W})$. The Laplace approximation in (4) requires optimizing the integrand and calculating the second derivative of $Q(\mathbf{W})$ (Hessian matrix) for a fixed parameter value ψ. However, since the dimension of the random effects grows with the number of observed locations, such optimizations do not generally work well for large data sets. To address these issues, we develop a projection-based Laplace method, as shown in the following section.

3 Fast Laplace Approximation

3.1 Dimension Reduction through Projection

To address the computational and inferential challenges of such models, projection methods have been developed for both continuous (Guan and Haran, 2018) and lattice

(Hughes and Haran, 2013) spatial domains. The main idea is to reduce the dimension of the latent variables from n to m based on $n \times m$ projection matrix $\mathbf{M}$. We note that m is much lower than n; for instance, in US COVID-19 data example (Section 5.1) we use $m = 207$ for $n = 3,108$.

In the continuous domain, the projection matrix is based on the covariance of the latent variable. For the zero-mean Gaussian process $\mathbf{W}$, the covariance matrix is $\mathbf{C}_{\boldsymbol{\theta}} = \sigma^2\mathbf{R}_\phi$, where σ^2 is the overall variance and ϕ is the range parameter. Taking the first m principal components of $\mathbf{R}_\phi$, the projection matrix $\mathbf{M} = \mathbf{U}_\phi\mathbf{D}_\phi^{1/2}$ is computed. Here, $\mathbf{U}_\phi$ is the $n \times m$ matrix of the first m eigenvectors of $\mathbf{R}_\phi$ and, $\mathbf{D}_\phi$ is the diagonal matrix with corresponding eigenvalues. Then the reduced-dimensional model is

$$\begin{aligned} g\{E[\mathbf{Z}|\boldsymbol{\beta}, \mathbf{M}, \boldsymbol{\delta}]\} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta} \\ \boldsymbol{\delta} &\sim N(0, \sigma^2\mathbf{I}). \end{aligned} \tag{5}$$

However, this requires the eigendecomposition of the $n \times n$ correlation matrix $\mathbf{R}_\phi$ for each ϕ update, which is computationally demanding for large spatial data. Therefore, Guan and Haran (2018) use a stochastic version of this for their implementation. We provide a summary of this in the supplementary material.

In the lattice domain, $\mathbf{M}$ is constructed based on the underlying graph structure as well as the covariates of interest. With the adjacency matrix $\mathbf{A}$, the Moran operator is $\mathbf{P}^\perp\mathbf{A}\mathbf{P}^\perp$, where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$. By taking the largest m principal components of $\mathbf{P}^\perp\mathbf{A}\mathbf{P}^\perp$, the projection matrix $\mathbf{M}$ can be constructed. Then the reduced-dimensional model in Hughes and Haran (2013) can be written as

$$\begin{aligned} g\{E[\mathbf{Z}|\boldsymbol{\beta}, \mathbf{M}, \boldsymbol{\delta}]\} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta} \\ f_\delta(\boldsymbol{\delta}|\theta) &\propto \theta^{m/2}\exp(-\frac{\theta}{2}\boldsymbol{\delta}^\intercal\mathbf{M}^\intercal\mathbf{Q}\mathbf{M}\boldsymbol{\delta}). \end{aligned} \tag{6}$$

In what follows, we provide projection-based Laplace methods that can provide accurate estimation and prediction quickly. Compared with simulation-based methods such as MCEM (Guan and Haran, 2019) and MCML (Park and Haran, 2020), our methods can be an easier alternative in that users do not have to tune numerous components manually as in convergence analysis for Monte Carlo samples. Furthermore, our methods can provide accurate standard errors for covariance parameters.

3.2 Projection-based Laplace Method

Here, we describe a fast Laplace method for SGLMMs. We apply our method to both continuous and lattice domains. Based on the projection approaches described in the previous section, a reduced-dimensional representation of the likelihood function is

$$L(\boldsymbol{\psi}|\mathbf{Z}) = \int_{R^m} f_{\mathbf{Z}|\delta}(\mathbf{Z}|\boldsymbol{\beta}, \mathbf{M}, \boldsymbol{\delta}) f_{\delta}(\boldsymbol{\delta}|\boldsymbol{\theta}) d\boldsymbol{\delta} = \int_{R^m} \exp\left(Q(\boldsymbol{\delta})\right) d\boldsymbol{\delta}, \tag{7}$$

where

$$Q(\boldsymbol{\delta}) = \mathbf{Z}^{\top}(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta}) - \mathbf{1}^{\top} b(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta}) + \mathbf{1}^{\top} c(\mathbf{Z}) - \frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\boldsymbol{\delta}^{\top}\boldsymbol{\delta}. \tag{8}$$

$Q(\boldsymbol{\delta})$ is the sum of the log-likelihood from exponential families and the multivariate normal log-likelihood for the reduced dimensional random effects $\boldsymbol{\delta}$. For the Laplace approximation in (4), we need the optimized $\widehat{\boldsymbol{\delta}}$ from $Q(\boldsymbol{\delta})$ for the given model parameter $\boldsymbol{\psi}$. The projection-based Laplace method can dramatically reduce computational costs. This is because, compared with $Q(\mathbf{W})$ in Bonat and Ribeiro (2016), solving $Q(\boldsymbol{\delta})$ is a much lower-dimensional ($m << n$) maximization problem. Even with long computing time, the estimates from the original method (Bonat and Ribeiro, 2016) are unstable

for the problems considered in our manuscript ($n > 1,000$), because this requires high-dimensional optimization with each iteration. To obtain $\widehat{\boldsymbol{\delta}}$, we use the following iterative Newton-Raphson method as

$$\boldsymbol{\delta}^{+} = \boldsymbol{\delta} - \Big(\frac{\partial^2 Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^{\intercal}}\Big)^{-1} \frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}, \tag{9}$$

where the first and second derivatives are

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^{\intercal}} &= \mathbf{Z}^{\intercal}\mathbf{M} - b'(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta})^{\intercal}\mathbf{M} - \frac{1}{\sigma^2}\boldsymbol{\delta}^{\intercal} \\ \frac{\partial^2 Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^{\intercal}} &= -\mathbf{M}^{\intercal}\mathrm{diag}(b''(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta}))\mathbf{M} - \frac{1}{\sigma^2}\mathbf{I}. \end{aligned} \tag{10}$$

This is repeated until convergence (inner optimization).

Then, the Laplace approximation for the reduced-dimensional representation is

$$L(\boldsymbol{\psi}|\mathbf{Z}) = \int_{R^m} \exp\left(Q(\boldsymbol{\delta})\right) d\boldsymbol{\delta} \approx (2\pi)^{m/2} \Big| - \frac{\partial^2 Q(\widehat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^{\intercal}} \Big|^{-1/2} \exp\Big(Q(\widehat{\boldsymbol{\delta}})\Big), \tag{11}$$

where the reduced-dimensional log-likelihood is

$$\begin{aligned} \widehat{l}(\boldsymbol{\psi}|\mathbf{Z}) &= \frac{m}{2}\log(2\pi) - \frac{1}{2}\log\Big| - \frac{\partial^2 Q(\widehat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}^{\intercal}} \Big| + Q(\widehat{\boldsymbol{\delta}}) \\ &= \frac{m}{2}\log(2\pi) - \frac{1}{2}\log\Big|\mathbf{M}^{\intercal}\mathrm{diag}(b''(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\widehat{\boldsymbol{\delta}}))\mathbf{M} + \frac{1}{\sigma^2}\mathbf{I}\Big| \\ &\quad + \mathbf{Z}^{\intercal}(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\widehat{\boldsymbol{\delta}}) - \mathbf{1}^{\intercal} b(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\widehat{\boldsymbol{\delta}}) + \mathbf{1}^{\intercal} c(\mathbf{Z}) - \frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\widehat{\boldsymbol{\delta}}^{\intercal}\widehat{\boldsymbol{\delta}}. \end{aligned} \tag{12}$$

This function is maximized over the model parameter value $\boldsymbol{\psi}$ using a numerical optimization (outer optimization). Then, the asymptotic distribution of the MLE $\widehat{\boldsymbol{\psi}}$ is $N(\boldsymbol{\psi}^*, \mathbf{I}^{-1}(\widehat{\boldsymbol{\psi}}))$, where $\boldsymbol{\psi}^*$ is the true value and $\mathbf{I}^{-1}(\widehat{\boldsymbol{\psi}})$ is the observed Fisher information.

In the lattice spatial domain, $\mathbf{W}$ follows the GMRF. Then, the $Q(\boldsymbol{\delta})$ function can be defined as

$$Q(\boldsymbol{\delta}) = \mathbf{Z}^\top(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta}) - \mathbf{1}^\top b(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta}) + \mathbf{1}^\top c(\mathbf{Z}) + \frac{m}{2}\log(\theta) - \frac{\theta}{2}\boldsymbol{\delta}^\top\mathbf{M}^\top\mathbf{Q}\mathbf{M}\boldsymbol{\delta}, \quad (13)$$

and their derivatives are

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^\top} &= \mathbf{Z}^\top\mathbf{M} - b'(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta})^\top\mathbf{M} - \theta\boldsymbol{\delta}^\top\mathbf{M}^\top\mathbf{Q}\mathbf{M} \\ \frac{\partial^2 Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}\partial\boldsymbol{\delta}^\top} &= -\mathbf{M}^\top\mathrm{diag}(b''(\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta}))\mathbf{M} - \theta\mathbf{M}^\top\mathbf{Q}\mathbf{M}. \end{aligned} \quad (14)$$

Similar to the continuous domain case, we can construct the Laplace approximation for the reduced-dimensional likelihood function. We optimize the logarithm of the variance parameters ($\log(\sigma^2)$, $\log(\phi)$) for numerical stability. Algorithm 1 summarizes our fast Laplace method.

Algorithm 1 A projection-based Laplace method

For $(t+1)$st update, given $\boldsymbol{\psi}_t$

Step 1. Inner optimization:

Obtain $\widehat{\boldsymbol{\delta}}$ using an iterative Newton-Raphson method as

$$\boldsymbol{\delta}^+ = \boldsymbol{\delta} - \Big(\frac{\partial^2 Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}\partial\boldsymbol{\delta}^\top}\Big)^{-1}\frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}$$

Step 2. Outer optimization:

Obtain $\boldsymbol{\psi}_{t+1}$ using a numerical optimization of

$$\widehat{l}(\boldsymbol{\psi}|\mathbf{Z}) = \frac{m}{2}\log(2\pi) - \frac{1}{2}\log\Big|-\frac{\partial^2 Q(\widehat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}\partial\boldsymbol{\delta}^\top}\Big| + Q(\widehat{\boldsymbol{\delta}})$$

Repeat Steps 1 and 2 until convergence, which gives MLEs $\widehat{\boldsymbol{\psi}}$.

3.3 Rank Selection

For areal data sets, Hughes and Haran (2013) suggest choosing the rank from the desired proportion of variation (e.g., 90% using the first several principal components). For point-referenced data sets over continuous domain, Guan and Haran (2019), Lee and Haran (2019) select the rank based on fitting the non-spatial GLMs as a preliminary step. Here, we adopt some of these ideas and recommend the following steps:

1. We split data into training and validation sets. For a given initial value of $\phi^{(0)}$ (e.g., first quartile of distance matrix), we can construct a synthetic spatial variable $\mathbf{U}_{\phi_{(0)}}\mathbf{D}^{1/2}_{\phi_{(0)}}$ from the eigencomponents of $\mathbf{R}_{\phi_{(0)}}$.

2. We use maximum likelihood approaches (`glm` function in `R`) to fit the standard GLM on a covariate $[\mathbf{X}, \mathbf{U}_{\phi_{(0)}}\mathbf{D}^{1/2}_{\phi_{(0)}}]$.

3. We can repeat Step 1-2 across varying ranks. Then we select the rank m that has the lowest out-of-sample cross-validated mean squared prediction error (CVMSPE).

4 Simulated Data Examples

To validate our methods, we conduct simulation studies to count and binary data for both the continuous and the lattice spatial domains. The code for this is implemented in `R`. For our method, we only need to obtain the first m eigencomponents quickly. To implement this efficiently, we use the `RSpectra` package. The source code can be downloaded from the following repository (https://github.com/jwpark88/projLaplace). In the supplementary material, we also provide a Poisson example in a continuous domain.

4.1 Negative Binomial Example in a Continuous Domain

For the fixed model parameters $(\beta_1, \beta_2, \sigma^2, \phi) = (1, 1, 1, 0.2)$, we simulate a negative binomial data set with $n = 1,400$ observations in the $[0, 1]^2$ domain. Random effects $\mathbf{W}$ are generated from the Matérn class (Stein, 2012) covariance kernel with a smoothing parameter of 2.5. Then we use the log-link function $\log(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}$, where $\mathbf{X}$ is the coordinate for the random effects $\mathbf{W}$. We set the true dispersion parameter ζ for the negative binomial distribution to 2. We use the first $1,000$ observations for the model fitting and 400 for the validation. We obtain the initial values for the Laplace approximation from the GLM estimates. We chose $m = 38$ by following the procedure in Section 3.3 (Figure 1).

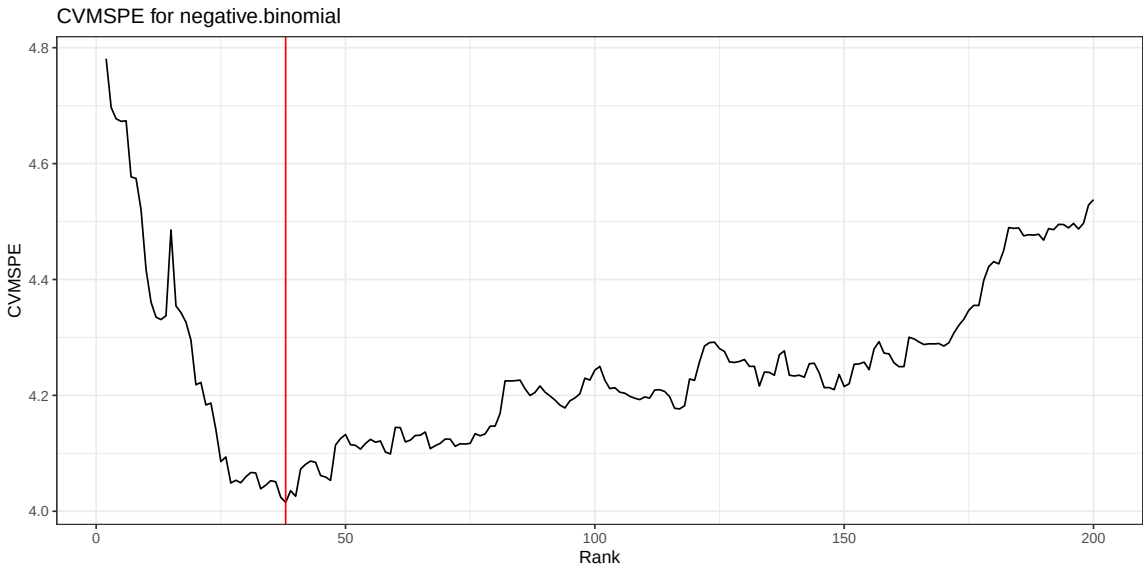

Figure 1: CVMSPE plot for ranks from 2 to 200. The vertical red line denotes the lowest CVMPSE (rank = 38)

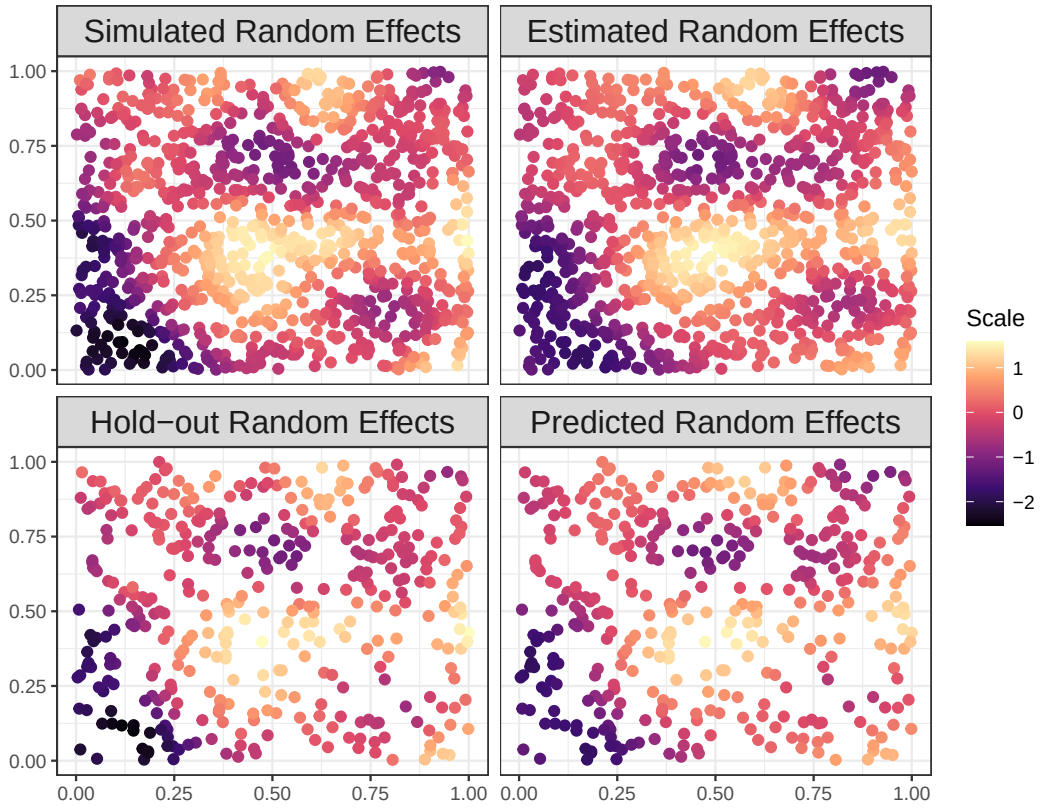

Figure 2: The first row represents the simulated and estimated random effects at the observed locations. The second row represents the out-of-sample cross-validated random effects and predicted random effects at the unobserved locations.

Figure 2 compares the simulated and the predicted random effects. The results suggest that they have similar spatial patterns. To investigate the performance of our method, we conduct simulation studies with varying n (data size) and m (rank) values. Especially, we consider $n = 1,400, 2,800, 5,600$, where $1,000, 2,000, 4,000$ observations are used for model fitting respectively. For each n, we consider three different choices of m: (1) 2% of training data, (2) recommended rank from Section 3.3, and (3) 10% of training data. This results in 9 combinations of n and m; simulation is repeated 100 times for each combination. To measure the parameter estimation accuracy, we cal-

culate the mean square error (MSE), mean of estimates, coverage of each parameter. To assess prediction accuracy, we calculate the mean square prediction error (MSPE) from test data sets. Since the recommended rank is different for each simulated data set, we report the average chosen rank from 100 simulations. Since the ratio between the covariance parameters is often of interest under fixed-domain asymptotics (Zhang, 2004), we provide results for σ^2/ϕ. Note that we can easily obtain confidence intervals for covariance parameters, which is challenging under other simulation-based methods such as MCML (Park and Haran, 2020) and MCEM (Guan and Haran, 2019). For the given ϕ value, they need Monte Carlo samples of the random effects to approximate the log-likelihood (MCML) or the conditional expectation (MCEM); this makes it challenging to calculate a gradient of ϕ to obtain standard errors. Table 1 summarize results.

When m is at or larger than a recommended rank, our projection-based Laplace methods provide estimates accurately, and the coverages for the parameters are close to the nominal rate (95%). Furthermore, prediction accuracies from validation data sets are reasonably small. With increasing m, the amount of information loss due to the dimension reduction becomes decreased at the expense of computing time. However, diminishing returns are observed based on previous works (Hughes and Haran, 2013, Guan and Haran, 2018); increasing m beyond certain points has little benefits but increases computing time. We also observe that the proposed method provides slightly more accurate results (estimation/prediction) with increasing m. However, such differences are marginal compared to results obtained from the "recommended" rank. Therefore, the "recommended" rank can be a practical choice for fast inference. Our approach performs well in terms of estimation and prediction aspects even for large n. We can conduct accurate maximum likelihood inference within a reasonable amount of time.

4.2 Binomial Example in a Continuous Domain

We use the model parameters $(\beta_1, \beta_2, \sigma^2, \phi) = (1, 1, 1, 0.2)$ to generate a binary data set with $n = 1,400$ in the unit domain. Random effects are simulated from the Matérn covariance with a smoothing parameter of 2.5. For the given simulated $\mathbf{W}$, we use a logit link function $\text{logit}(p) = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}$, where $\mathbf{X}$ is the coordinate for the random effects $\mathbf{W}$. For the model fitting, we use the first $1,000$ observations, and 400 for the prediction. Similar to the negative binomial example, we chose a rank $m = 21$ by following the procedure in Section 3.3 (Figure 3). We obtain the initial values of the Laplace method from the GLM estimates.

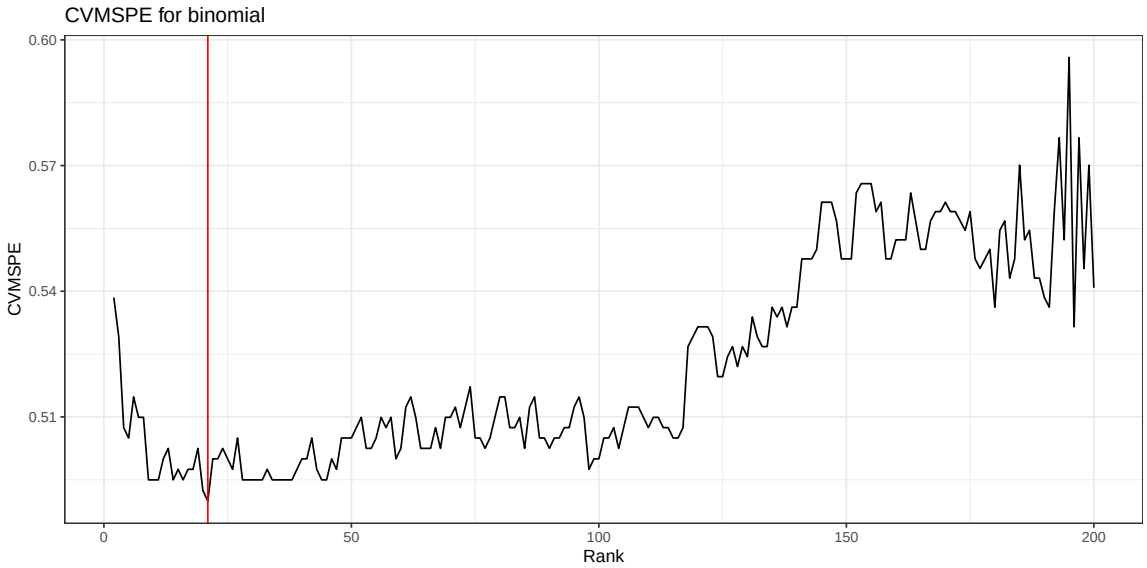

Figure 3: CVMSPE plot for ranks from 2 to 200. The vertical red line denotes the lowest CVMPSE (rank = 21)

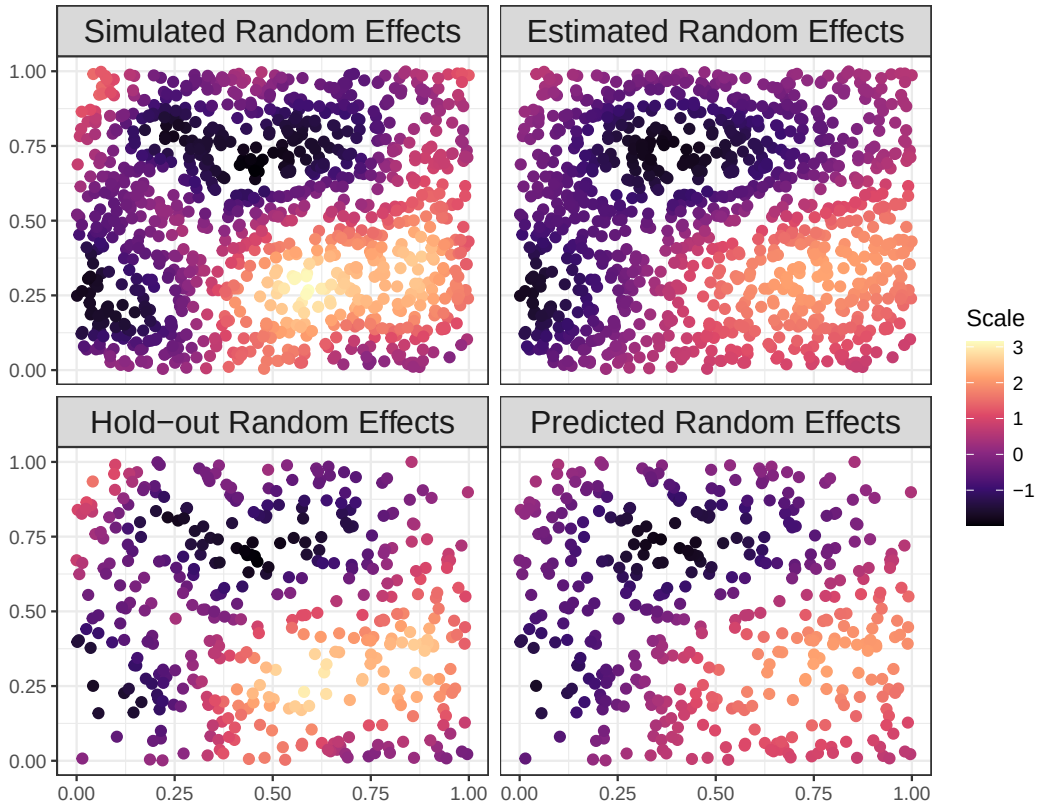

Figure 4: The first row represents the simulated and estimated random effects at the observed locations. The second row represents the out-of-sample cross-validated random effects and predicted random effects at the unobserved locations.

The simulated and predicted spatial random effects have similar patterns (Figure 4). To validate our method, we conduct simulation studies with varying n and m as in the previous section (Table 2). We observe that our projection-based Laplace methods provide accurate estimation, and the coverages are close to the nominal rate (95%) even with the smallest m (i.e., 2% of training data size). We observe that a recommended ranks are chosen as smaller values compared to the negative binomial cases. MSPEs from test data sets are also small for different choices of m. Similar to the negative binomial cases, we can conduct accurate maximum likelihood inference quickly for n.

Especially when we use a recommended rank for $n = 5,600$, the fast Laplace method provides accurate maximum likelihood inference within an hour.

4.3 Poisson Example in a Lattice Discrete Domain

Following the simulation settings in Hughes and Haran (2013), we show the use of our approach in a lattice domain. As in the continuous domain examples, we test the sensitivity of our method by comparing the performance across different n, m values. With the fixed parameter values $(\beta_1, \beta_2, \theta) = (1, 1, 6)$ we simulate Poisson data sets with $n = 900$ (30×30 lattice), $2,500$ (50×50 lattice), and $4,900$ (70×70 lattice). We simulate $\mathbf{W}$ from $N(0, (\theta\mathbf{M}'\mathbf{Q}\mathbf{M})^{-1})$, where $\mathbf{M}$ is obtained by taking the first k eigenvectors of the Moran operator (i.e., $\mathbf{M} \in R^{n\times k}$). We use $k = 400, 1200, 2400$ for $n = 900, 2,500, 4,900$ repectively. Then, we use the log-link function to generate a count data set. Similar to the previous simulation studies, we consider m as (1) 2% of training data, (2) recommended rank from Section 3.3 and (3) 10% of training data. Simulation is repeated 100 times for each n, m combination. Table 3 summarize results.

In general, our projection-based Laplace methods provide reasonably accurate estimates for regression coefficients when m is at or larger than a recommended rank. We observe that estimates are close to the simulated truth with small MSEs. However, with increasing sample size ($n \geq 2,500$), it becomes more challenging to achieve the nominal rate (95% coverage), even with m as 10% of training data. Therefore, we conduct additional experiments with m as 20% of training data and observe the improved coverages. This implies that compared to the point-referenced data examples, we need larger m to achieve the nominal rate. For researchers who want to have conservative results, we suggest choosing m as 10%-20% of training data size at the expense of computing

time. Furthermore, we observe that the covariance parameter θ is slightly biased even with large m; this results in low coverage ($< 95\%$). Note that the correct estimation of θ is challenging for conditional autoregressive models, as reported by Hughes and Haran (2013).

4.4 Comparison with the Standard Laplace Method

Here, we compare the performance of our method with that of the standard Laplace method (Bonat and Ribeiro, 2016) under different numbers of data points. We use the same simulation setting as in the previous examples. Figure 5 shows that computing time increases dramatically for the standard Laplace algorithm (Bonat and Ribeiro, 2016), whereas our method can obtain the maximum likelihood estimates quickly.

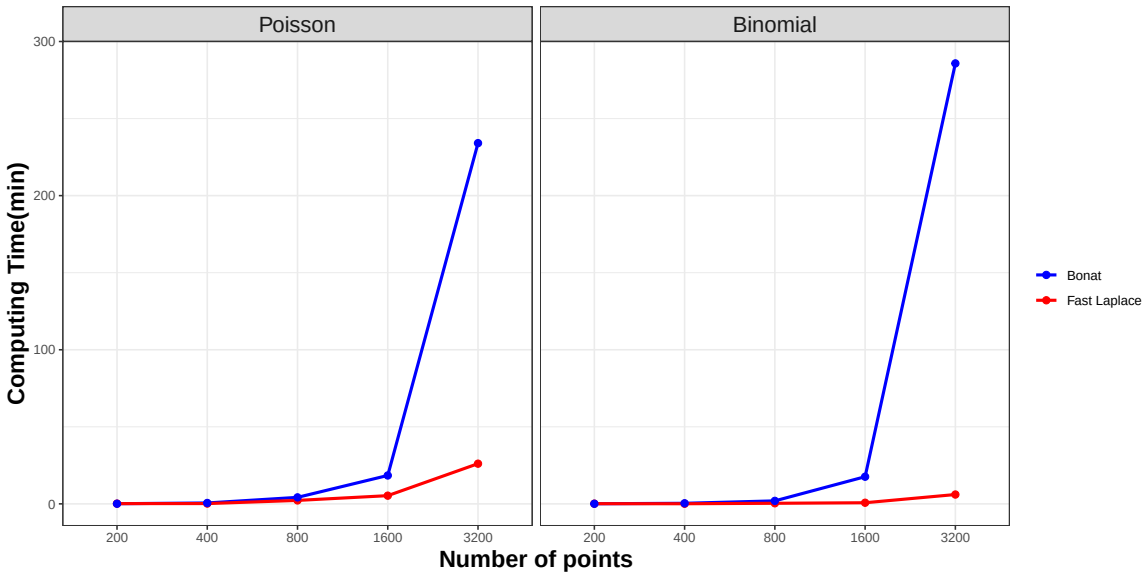

Figure 5: The observed computing time for the algorithms for the Poisson and binomial examples.

Furthermore, we compare our method with competitors (standard Laplace (Bonat and Ribeiro, 2016), projection-based MCML (Park and Haran, 2020), MCMC (Guan and Haran, 2018)) in the supplementary material (Section D). We compare different algorithms in large binomial data sets with $n = 2,800, 5,600$ using the same simulation settings in Section 4.2. In summary, our algorithm provides comparable estimations and predictions to competitors within a shorter time. We observe that the projection-based Laplace method is about 6 times faster than the standard Laplace method. In addition, our method can achieve the nominal rate (95% coverage) for covariance parameters as well, which is challenging to projection-based MCML (see the supplementary material for more details).

Note that the reported computing time is obtained from our simulation settings; therefore, it is not absolute but relative scales. Both Laplace methods update model parameters from outer optimization (Step 2 in Algorithm 1), which has order p complexity. Therefore, both algorithms will take longer with an increasing number of spatial covariates.

5 Real Data Examples

We apply our methods to two real data examples with a large number of observations: (1) COVID-19 count data from US counties and (2) binary patterns of Antarctic ice sheets over the continuous domain. In both cases, our approach provides MLEs and predictions quickly.

5.1 US COVID-19 Data

Caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) coronavirus disease 2019 (COVID-19) was first reported in Wuhan, Hubei providence, China in December 2019 (WHO, 2020). As of March 2021, there have been more than 100 million confirmed cases, with more than 2 million deaths worldwide. The disease has spread rapidly through contact with infected people (CDC, 2020), which has led to significant societal challenges. Understanding the spatial patterns of confirmed cases is valuable not only for scientific reasons but also for public health, such as developing vaccination and intervention strategies to control epidemics. Furthermore, investigating the relationship between spatial covariates can provide useful insights into the mechanism of disease spread. The estimated regression coefficients tell us whether there is a positive (or negative) relation between each spatial covariate and the confirmed cases (count response). Our method provides accurate maximum likelihood inference for thousands of observations quickly.

Here, we study COVID-19 confirmed cases data set collected from the Centers for Disease Control and Prevention(CDC) and the Census Bureau. The data set provides a cumulative number of confirmed cases from 3,108 US counties between January 1, 2020, and July 11, 2020. As covariates, we use the percentage of white residents, the percentage of black residents, the percentage of male, and the percentage of older people ($\geq$ 65). We use the population of each county as an offset. We obtain the initial values of the model parameters from the Poisson GLM estimates. We use a reduced rank $m = 207$, which has the lowest CVMSPE (Figure 7). The projection-based Laplace method takes about 45 minutes. Regular Laplace methods (Bonat and Ribeiro, 2016) are computationally expensive because they have to deal with a 3,108-dimensional op-

timization to find a mode of the Laplace approximation for the fixed model parameter value.

Table 4 shows the parameter estimates and their confidence intervals. We observe that the black proportion and the male proportion show a positive relationship, while the white proportion and the old people proportion have a negative relationship with the COVID-19 risk. Note that our study is county-level; we cannot interpret this result at the individual level. For instance, we do not imply that older people have a low COVID-19 risk, but rather that the county with a high population of older people has a low risk. This is reasonable because counties with a higher proportion of older people would have less economic/daily activities; this results in a low COVID-19 risk. Figure 6 shows that there are similar spatial patterns between the observed and fitted number of confirmed cases. We detect three high-intensity regions near California, Florida, and eastern United States (New York City, New Jersey), suggesting these areas could be disease hotspots. The prediction standard errors are larger at larger predicted values, which is natural as more extreme values tend to be more difficult to predict.

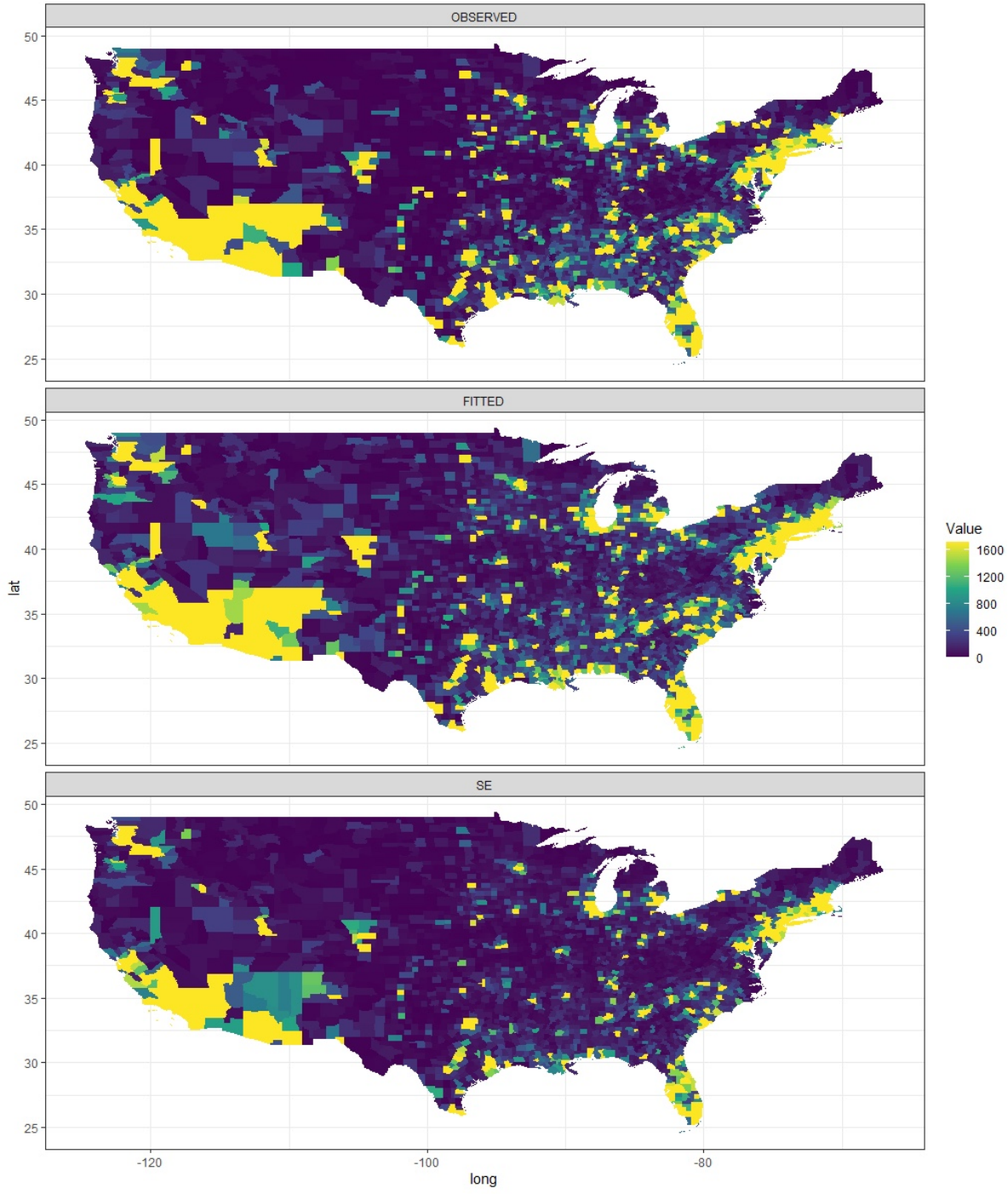

Figure 6: The upper panel shows the observed number of confirmed cases by county. The middle panel shows the estimated mean of confirmed cases by county. The lower panel is the corresponding standard error of estimates.

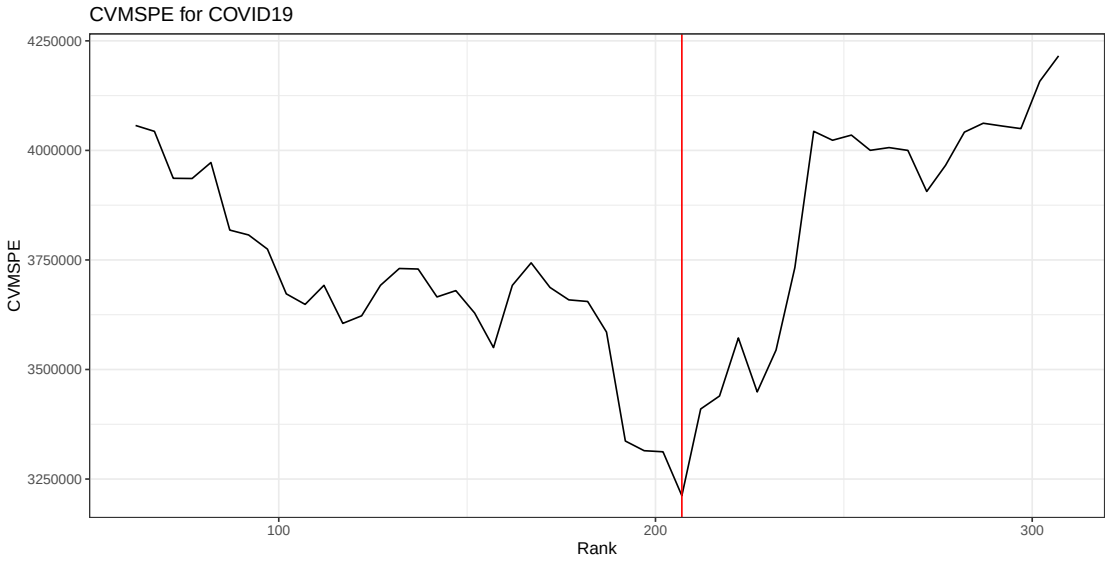

Figure 7: CVMSPE plot for US COVID-19 data. The vertical red line denotes the lowest CVMPSE (rank = 207)

5.2 Antarctic Ice Sheet Data

The mass loss from ice sheets may cause a global sea-level rise (Deschamps et al., 2012). Especially, the Antarctic ice sheet contains 30 million cubic kilometers of ice; this is sufficient ice to contribute up to 60 meters to sea-level rise if they were to melt completely (White et al., 2019). Several recent studies (Zhang, 2007, Serreze and Barry, 2011) indicate that some fraction of this ice may melt within the next few hundred years due to global climate change. A significant proportion of the world's population lives within a low-elevation coastal zone (Greve et al., 2011), and sea-level rise poses increased risk for their life. Considering that ice thickness is an important indicator of climate change, understanding the spatial pattern of ice thickness is crucial for future sea-level rise. Despite its importance, providing accurate estimation/prediction of ice thickness is challenging because it is a high-dimensional and non-Gaussian spatial vari-

able. For this study, our method provides accurate interpolation of ice thickness within a reasonable amount of time.

We apply our method to the Antarctic ice sheet data set (Pollard et al., 2015). This data set represents a binary spatial pattern of ice thickness (the binary ice-no ice patterns are derived from these data). The ice sheet data set is observed with a fine 20 km resolution over 171 $\times$ 171 grid points, centered on the South Pole, and it spans the whole Antarctic continent (see Pollard et al. (2015) for more details). We randomly select 6,000 locations. Then, 4,900 locations are used for training and the remaining 1,100 locations are used for model prediction. As in the simulated examples, we use the initial values for the model parameters taken from the binomial GLM estimates. We choose a reduced rank of 50, which has the lowest CVMSPE.

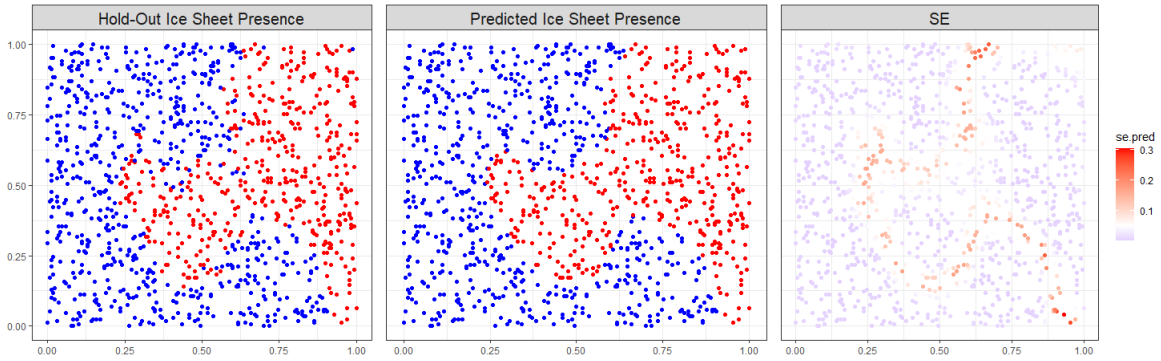

Figure 8: The left panel shows the observed ice thickness and the middle panel shows the predicted ice thickness. Red and blue points indicate ice-no ice patterns. The right panel shows the corresponding standard error of estimates.

Our fast Laplace approximation only takes about 50 minutes. Because of the high-dimensional 4,900 optimization for the latent variables, the regular Laplace method (Bonat and Ribeiro, 2016) is infeasible. Projection-based MCEM approaches (Guan and Haran, 2019) and MCML approaches (Park and Haran, 2020) can provide similar

computing costs to our method. However, both require numerous components to be tuned, which can lead to practical implementation issues for users. Furthermore, we can obtain standard errors for the range parameter easily from the numerical optimization. Table 5 reports the parameter estimates and their confidence intervals. Based on the regression coefficient estimates, we verify that, on average, ice becomes thicker in northeastern directions, which can also be checked from observed ice sheet presence (Figure 8). The predictions of the ice thickness patterns are the main interest of this analysis. To classify ice-no ice pattern, we dichotomize the predicted probabilities at 0.5. Figure 8 shows that the similar binary spatial patterns between the observed and predicted locations with accuracy 0.96. We observe that the prediction standard errors are larger at the boundaries of the ice sheet. Predicting the boundary regions are more difficult because of the mixed binary patterns.

6 Discussion

In this manuscript, we develop a projection-based Laplace method for high-dimensional spatial latent variable models. Based on novel projection approaches, we replace a high-dimensional optimization problem with a much lower-dimensional optimization problem to obtain the Laplace approximation of the integrand. We show that our approaches provide accurate maximum likelihood estimations and predictions quickly for both continuous and lattice domains. Compared with simulation-based methods such as MCEM (Guan and Haran, 2019) and MCML (Park and Haran, 2020), our method is automatic in that users can avoid tuning of the algorithm, including the convergence diagnostic of MCMC chains. Therefore, our fast Laplace method may be an useful alternative for maximum likelihood inference. Furthermore, our algorithm can provide accurate

standard errors for the estimates, which can be challenging for the MCEM and MCML approaches.

We apply our methods to environmental and epidemiological data sets, providing accurate interpolation. Furthermore, we could obtain scientific insights into the spatial patterns of non-Gaussian responses from regression coefficients. For instance, we observe that the percentage of black residents and male residents have positive relationships with COVID-19 cases. On the other hand, the percentage of white residents and old residents have negative relationships with COVID-19 cases. For the Antarctic ice sheet example, we found that, on average, ice becomes thicker in northeastern directions.

The main contribution of this study is to develop a practical Laplace method for maximum likelihood inference for large non-Gaussian spatial data sets, which frequently arise in environmental sciences. Developing extensions of the projection-based Laplace approach to the high-dimensional spatio-temporal model may be an interesting direction for future research. Considering that the Laplace approximation is convenient for maximum likelihood inference for latent variable models, the methods and ideas described here in could be applicable to a broader class of problems. Examples include generalized linear latent variable models (Bianconcini and Cagnone, 2012), longitudinal data analysis (Molenberghs and Verbeke, 2005), and stochastic volatility models in finance (Jacquier et al., 2007).

Supplementary Materials

The supplementary material describes the random projection method and comparison with competing algorithms. It also contains a simulation study for the Poisson example

in a continuous domain.

Acknowledgement

We are grateful to Won Chang for providing the Antarctic ice sheet data set. We are grateful to Won Chang and Yawen Guan for their helpful discussions. Jaewoo Park and Sangwan Lee were partially supported by the Yonsei University Research Fund of 2020-22-0501 and the National Research Foundation of Korea (NRF-2020R1C1C1A0100386812). The authors are grateful to the anonymous reviewers for their careful reading and valuable comments.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press.

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

Berrett, C., Christensen, W. F., Sain, S. R., Sandholtz, N., Coats, D. W., Tebaldi, C., and Lopes, H. F. (2020). Modeling sea-level processes on the US Atlantic Coast. *Environmetrics*, 31(4):e2609.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36:192–236.

Bianconcini, S. and Cagnone, S. (2012). Estimation of generalized linear latent variable models via fully exponential laplace approximation. *Journal of Multivariate Analysis*, 112:183–193.

Bonat, W. H. and Ribeiro, P. J. (2016). Practical likelihood analysis for spatial generalized linear mixed models. *Environmetrics*, 27(2):83–89.

CDC (2020). How COVID-19 spreads. *https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html*.

Christensen, O. F. (2004). Monte Carlo maximum likelihood in model-based geostatistics. *Journal of computational and graphical statistics*, 13(3):702–718.

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.

Clifford, S., Low-Choy, S., Mazaheri, M., Salimi, F., Morawska, L., and Mengersen, K. (2019). A Bayesian spatiotemporal model of panel design data: Airborne particle number concentration in Brisbane, Australia. *Environmetrics*, 30(7):e2597.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

Deschamps, P., Durand, N., Bard, E., Hamelin, B., Camoin, G., Thomas, A. L., Henderson, G. M., Okuno, J., and Yokoyama, Y. (2012). Ice-sheet collapse and sea-level rise at the bølling warming 14,600 years ago. *Nature*, 483(7391):559–564.

Diggle, P. J., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Entezari, R., Brown, P. E., and Rosenthal, J. S. (2020). Bayesian spatial analysis of hardwood tree counts in forests via MCMC. *Environmetrics*, 31(4):e2608.

Evangelou, E., Zhu, Z., and Smith, R. L. (2011). Asymptotic inferente for spatial GLMM using high order Laplace approximation. *Journal of Statistical Planning and Inference*, 141(11):3564–3577.

Finley, A. O., Datta, A., Cook, B. D., Morton, D. C., Andersen, H. E., and Banerjee, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *Journal of Computational and Graphical Statistics*, 28(2):401–414.

Forlani, C., Bhatt, S., Cameletti, M., Krainski, E., and Blangiardo, M. (2020). A joint Bayesian space–time model to integrate spatially misaligned air pollution data in R-INLA. *Environmetrics*, 31(8):e2644.

Greve, R., Saito, F., and Abe-Ouchi, A. (2011). Initial results of the SeaRISE numerical experiments with the models SICOPOLIS and IcIES for the greenland ice sheet. *Annals of Glaciology*, 52(58):23–30.

Guan, Y. and Haran, M. (2018). A computationally efficient projection-based approach for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 27(4):701–714.

Guan, Y. and Haran, M. (2019). Fast expectation-maximization algorithms for spatial generalized linear mixed models. *arXiv preprint arXiv:1909.05440*.

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.

Jacquier, E., Johannes, M., and Polson, N. (2007). MCMC maximum likelihood for latent state models. *Journal of Econometrics*, 137(2):615–640.

Lawson, A. B., Banerjee, S., Haining, R. P., and Ugarte, M. D. (2016). *Handbook of spatial epidemiology*. CRC Press.

Lee, B. S. and Haran, M. (2019). Picar: An efficient extendable approach for fitting hierarchical spatial models. *arXiv preprint arXiv:1912.02382*.

Li, D., Wang, X., and Dey, D. (2019). Power link functions in an ordinal regression model with Gaussian process priors. *Environmetrics*, 30(6):e2564.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

Park, J. and Haran, M. (2020). Reduced-dimensional monte carlo maximum likelihood for latent gaussian random field models. *Journal of Computational and Graphical Statistics*, (just-accepted):1–35.

Pollard, D., Chang, W., Haran, M., Applegate, P., and DeConto, R. (2015). Large ensemble modeling of last deglacial retreat of the west antarctic ice sheet: Comparison of simple and advanced statistical techniques. *Geoscientific Model Development Discussions*, 8:9925–9963.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

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.

Serreze, M. C. and Barry, R. G. (2011). Processes and impacts of Arctic amplification: A research synthesis. *Global and planetary change*, 77(1-2):85–96.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

Tadayon, V. and Torabi, M. (2019). Spatial models for non-Gaussian data with covariate measurement error. *Environmetrics*, 30(3):e2545.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments

and marginal densities. *Journal of the american statistical association*, 81(393):82–86.

White, P. A., Reese, C. S., Christensen, W. F., and Rupper, S. (2019). A model for Antarctic surface mass balance and ice core site selection. *Environmetrics*, 30(8):e2579.

WHO (2020). Novel Coronavirus - China. *https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/*.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58(1):129–136.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.

Zhang, J. (2007). Increasing Antarctic sea ice under warming atmospheric and oceanic conditions. *Journal of Climate*, 20(11):2515–2529.

Zhang, L., Banerjee, S., and Finley, A. O. (2021). High-dimensional multivariate geostatistics: A Bayesian matrix-normal approach. *Environmetrics*, page e2675.

Table 1: Inference results for simulated negative binomial data sets in a continuous domain. For different combinations of n and m, average computing time (minutes), MSPE, mean of estimates, MSE, and coverage are calculated from 100 simulations. Asterisk symbols indicate a recommended rank.

n	m	Time	MSPE	Parameter	True	Mean	MSE	Coverage
1,400	20	1	0.154	β_1	1	1.065	0.057	0.66
				β_2	1	1.043	0.057	0.71
				$\log \zeta$	0.693	0.505	0.062	0.37
				$\log(\sigma^2/\phi)$	1.609	1.727	0.314	0.87
	56$*$	4	0.075	β_1	1	1.005	0.013	0.94
				β_2	1	1.001	0.017	0.89
				$\log \zeta$	0.693	0.664	0.010	0.91
				$\log(\sigma^2/\phi)$	1.609	1.585	0.096	0.95
	100	7	0.063	β_1	1	1.003	0.011	0.95
				β_2	1	1.001	0.013	0.91
				$\log \zeta$	0.693	0.699	0.008	0.97
				$\log(\sigma^2/\phi)$	1.609	1.552	0.100	0.93
2,800	40	18	0.051	β_1	1	1.018	0.007	0.91
				β_2	1	1.001	0.006	0.97
				$\log \zeta$	0.693	0.649	0.006	0.82
				$\log(\sigma^2/\phi)$	1.609	1.555	0.066	0.97
	82$*$	22	0.040	β_1	1	1.016	0.005	0.97
				β_2	1	1.003	0.005	0.98
				$\log \zeta$	0.693	0.683	0.004	0.95
				$\log(\sigma^2/\phi)$	1.609	1.531	0.055	0.95
	200	76	0.039	β_1	1	1.019	0.007	0.94
				β_2	1	1.007	0.007	0.98
				$\log \zeta$	0.693	0.686	0.032	0.96
				$\log(\sigma^2/\phi)$	1.609	1.514	0.056	0.94
5,600	80	58	0.024	β_1	1	1.006	0.003	0.95
				β_2	1	1.001	0.004	0.91
				$\log \zeta$	0.693	0.680	0.002	0.93
				$\log(\sigma^2/\phi)$	1.609	1.492	0.052	0.97
	120$*$	97	0.023	β_1	1	1.007	0.003	0.95
				β_2	1	1.001	0.004	0.93
				$\log \zeta$	0.693	0.686	0.002	0.93
				$\log(\sigma^2/\phi)$	1.609	1.496	0.053	0.96
	400	430	0.023	β_1	1	1.006	0.003	0.96
				β_2	1	0.999	0.004	0.94
				$\log \zeta$	0.693	0.695	0.002	0.93
				$\log(\sigma^2/\phi)$	1.609	1.486	0.053	0.95

Table 2: Inference results for simulated binary data sets in a continuous domain. For different combinations of n and m, average computing time (minutes), MSPE, mean of estimates, MSE, and coverage are calculated from 100 simulations. Asterisk symbols indicate a recommended rank.

n	m	Time	MSPE	Parameter	True	Mean	MSE	Coverage
1,400	20	1	0.254	β_1	1	0.993	0.093	0.91
				β_2	1	0.994	0.089	0.90
				$\log(\sigma^2/\phi)$	1.609	1.307	0.534	0.94
	30$*$	1	0.223	β_1	1	1.009	0.090	0.92
				β_2	1	1.012	0.075	0.91
				$\log(\sigma^2/\phi)$	1.609	1.411	0.432	0.91
	100	4	0.211	β_1	1	1.021	0.087	0.96
				β_2	1	1.034	0.075	0.94
				$\log(\sigma^2/\phi)$	1.609	1.520	0.206	0.97
2,800	40	6	0.142	β_1	1	1.008	0.047	0.89
				β_2	1	0.991	0.036	0.95
				$\log(\sigma^2/\phi)$	1.609	1.490	0.170	0.94
	47$*$	7	0.139	β_1	1	1.000	0.046	0.93
				β_2	1	0.985	0.032	0.96
				$\log(\sigma^2/\phi)$	1.609	1.494	0.188	0.92
	200	37	0.136	β_1	1	1.003	0.044	0.91
				β_2	1	0.991	0.032	0.97
				$\log(\sigma^2/\phi)$	1.609	1.503	0.159	0.93
5,600	80	55	0.084	β_1	1	0.990	0.018	0.95
				β_2	1	0.988	0.018	0.97
				$\log(\sigma^2/\phi)$	1.609	1.538	0.108	0.94
	85$*$	56	0.084	β_1	1	0.990	0.018	0.96
				β_2	1	0.987	0.018	0.97
				$\log(\sigma^2/\phi)$	1.609	1.535	0.111	0.93
	400	445	0.084	β_1	1	0.992	0.018	0.95
				β_2	1	0.989	0.018	0.97
				$\log(\sigma^2/\phi)$	1.609	1.540	0.111	0.94

Table 3: Inference results for simulated Poisson data sets in a lattice domain. For different combinations of n and m, average computing time (minutes), mean of estimates, MSE, and coverage are calculated from 100 simulations. Asterisk symbols indicate a recommended rank.

n	m	Time	Parameter	True	Mean	MSE	Coverage
900	18	<1	β_1	1	1.026	0.003	0.90
			β_2	1	1.025	0.003	0.94
			$\log\theta$	1.792	1.451	0.341	0.81
	45$*$	<1	β_1	1	1.018	0.003	0.94
			β_2	1	1.017	0.002	0.96
			$\log\theta$	1.792	1.438	0.265	0.70
	90	1	β_1	1	1.014	0.003	0.96
			β_2	1	1.012	0.002	0.99
			$\log\theta$	1.792	1.577	0.138	0.81
2,500	50	2	β_1	1	1.028	0.002	0.80
			β_2	1	1.028	0.002	0.85
			$\log\theta$	1.792	1.430	0.210	0.65
	110$*$	5	β_1	1	1.021	0.001	0.89
			β_2	1	1.019	0.001	0.94
			$\log\theta$	1.792	1.440	0.164	0.52
	250	7	β_1	1	1.016	0.001	0.91
			β_2	1	1.013	0.001	0.96
			$\log\theta$	1.792	1.570	0.074	0.68
	500	34	β_1	1	1.011	0.001	0.94
			β_2	1	1.007	0.001	0.98
			$\log\theta$	1.792	1.663	0.040	0.78
4,900	98	16	β_1	1	1.025	0.001	0.74
			β_2	1	1.025	0.001	0.73
			$\log\theta$	1.792	1.451	0.153	0.47
	203$*$	38	β_1	1	1.018	0.001	0.81
			β_2	1	1.013	0.001	0.80
			$\log\theta$	1.792	1.487	0.118	0.38
	490	123	β_1	1	1.012	0.001	0.87
			β_2	1	1.013	0.001	0.85
			$\log\theta$	1.792	1.607	0.049	0.57
	980	649	β_1	1	1.006	0.001	0.91
			β_2	1	1.007	0.001	0.89
			$\log\theta$	1.792	1.681	0.023	0.81

Table 4: Inference results for the US COVID-19 data.

Parameter	Estimate	95% CI
Intercept	-6.996	(-7.011,-6.081)
White	-1.485	(-1.503,-1.466)
Black	0.178	(0.159,0.197)
Male	7.363	(7.327,7.399)
Old	-2.705	(-2.752,-2.658)
$\log \theta$	-0.751	(-0.944,-0.558)

Table 5: Inference results for the Antarctic ice sheet data.

Parameter	Estimate	95% CI
Intercept	-4.661	(-7.102,-2.219)
x-coordinate	1.535	(-1.793,4.862)
y-coordinate	5.812	(2.050 ,9.574)
$\log \sigma^2$	1.824	(1.409 ,2.239)
$\log \phi$	-2.238	(-2.436,-2.040)