Spatial Prediction for Multivariate Non-Gaussian Data

XUTONG LIU, ebay Inc
FENG CHEN, University at Albany, SUNY
YEN-CHENG LU and CHANG-TIEN LU, Virginia Tech

With the ever increasing volume of geo-referenced datasets, there is a real need for better statistical estimation and prediction techniques for spatial analysis. Most existing approaches focus on predicting multivariate Gaussian spatial processes, but as the data may consist of non-Gaussian (or mixed type) variables, this creates two challenges: (1) how to accurately capture the dependencies among different data types, both Gaussian and non-Gaussian; and (2) how to efficiently predict multivariate non-Gaussian spatial processes. In this article, we propose a generic approach for predicting multiple response variables of mixed types. The proposed approach accurately captures cross-spatial dependencies among response variables and reduces the computational burden by projecting the spatial process to a lower dimensional space with knot-based techniques. Efficient approximations are provided to estimate posterior marginals of latent variables for the predictive process, and extensive experimental evaluations based on both simulation and real-life datasets are provided to demonstrate the effectiveness and efficiency of this new approach.

CCS Concepts: • **Machine Learning** → **Learning in probabilistic graphical models**; *Supervised Learning by regression*; • **Probability and statistics** → Probabilistic inference problems;

Additional Key Words and Phrases: Approximate bayesian inference, computational statistics, Gaussian and non-Gaussian processes, geostatistics, Laplace approximation, predictive process model

ACM Reference Format:
Xutong Liu, Feng Chen, Yen-Cheng Lu, and Chang-tien Lu. 2017. Spatial prediction for multivariate non-Gaussian data. ACM Trans. Knowl. Discov. Data 11, 3, Article 36 (March 2017), 27 pages.
DOI: http://dx.doi.org/10.1145/3022669

1. INTRODUCTION

Increasing public sensitivity and concern regarding environmental issues have led to huge amounts of spatial data being collected, and this volume continues to increase at an even faster pace. As one of today's major research issues, the prediction of multivariate spatial observations has attracted significant attention from many researchers, particularly those working in areas such as biology [McBratney et al. 2005], epidemiology [Kibria et al. 2002], geography [Gelfand et al. 2004], and economics [Chica-Olmo 2007]. For example, weather forecasting is an important area of investigation with serious implications for many aspects of human life.

Spatial prediction is the process of estimating values of a target quantity at unvisited locations, based on the observed measures at sampled ones. In the univariate case, spatial prediction has been well studied for different data types, including continuous

Authors' addresses: X. Liu, One Bellevue Center, 411-108th Avenue NE, Suite 400, Bellevue, WA 98004; email: xutongl@vt.edu; F. Chen, UAB 426, 1215 Western Ave, Albany, NY 12222; email: chenf@vt.edu; Y. Lu, 295 E Evelyn Ave 301, Sunnyvale, CA 94086; email: kevinlu@vt.edu; C.T. Lu, 7054 Haycock Road, Room 310, Falls Church, VA 22043; email: ctlu@vt.edu.

DOI: http://dx.doi.org/10.1145/3022669

[Cressie 1991; Wackernagel 2003; Zhuo et al. 2011; Torgo and Ohashi 2011], discrete [Oliveira 2000; Webster and Oliver 1990], and Poisson spatial processes [Wolpert and Ickstadt 1997]. Recently, a number of methods have been proposed for processing large univariate spatial datasets of different types, including fixed rank kriging [Cressie and Johannesson 2008], the knot-based spatial process [Banerjee et al. 2008], and the integrated nested Laplace approximation (INLA)-based predictive process [Rue et al. 2009]. In many cases, geo-referenced datasets are multivariate. Multivariate spatial analysis refers to those situations wherein there is an explicit vector of responses at each spatial location of interest and focuses on capturing potential relationships among spatial locations and multiple responses. The prediction of multivariate spatial processes at unsampled locations plays an important role when there are cross-spatial dependencies between multiple response variables of interest and there is therefore a considerable literature on the modeling and prediction of multivariate spatial processes [Cressie 1991; Ohashi and Torgo 2012]. Most related works focus on multivariate Gaussian processes (GPs), the key component of which is the modeling of cross-covariance functions between attributes at different spatial locations.

Multivariate spatial interpolation has attracted a great of research due to its usefulness in a wide range of applications, such as monitoring environmental pollution, managing responses to natural disasters, and public health-related activities. Observations are often of mixed types, including numerical, nominal, and count, each of which contains interesting information. In geological studies, it is often desirable to predict related properties of different types, such as moisture content (numeric), granularity (count), and coloration (categorical) for pedological data [Chagneau et al. 2010]. These mixed type responses and cross-variable dependencies further complicate the spatial inference process. The prediction of multivariate non-Gaussian (or mixed type) data has been identified as one of the 10 most important challenges in data mining for the next decade [Piatetsky-Shapiro et al. 2006]. Consider the example of weather monitoring. Weather forecasts can benefit from the more accurate estimation of real-time precipitation amounts, storm surges, and flood warnings for the protection of life and property, but this requires predictive models that can extract all the critical information simultaneously. The general objective here is to learn more about the relationships between response variables (flood warnings, chance of rain, and wind speed) and predictor variables (location, elevation, season, and plant areas). This is a typical application of spatial prediction involving mixed type observations, which creates multiple issues: *Challenge 1) Modeling mixed type observations.* What is the best way to model non-Gaussian multivariate spatial data that is as analytically intractable as Gaussian data in spatial prediction? For example, modeling the mixed-type response variables involved in weather forecasting, including precipitation amounts (numeric), flood warnings (binary: yes/no), and cloud levels (ordinal: mostly or partly cloudy). *Challenge 2) Capturing correlations among dependent variables.* Multivariate spatial analysis potentially involves two confounded dimensions of dependencies—between different responses, and between different spatial locations. Among these multivariate response types, there exist statistical relationships that need capturing. For example, the effects of wind speed on cloud level, and the possibility of flooding caused by the precipitation amounts. Such dependencies need integrating into the predictive model to improve the final estimation. *Challenge 3) Improving scalability and availability.* When training the predictive model, large amounts of weather data across multiple locations are being collected and estimating the forecast model thus involves expensive matrix decompositions whose computational complexity increases in cubic order with the number of spatial locations, rendering the task infeasible for large spatial datasets. *Challenge 4) Analytically intractable posterior inferences.* As an additional challenge, the likelihood related to non-Gaussian observations yields a distribution that is nonstandard and analytically intractable by nature. Therefore, approximate methods are needed for the particular likelihood function of multivariate mixed-type responses.

Traditional spatial models often assume the outcomes follow normal distributions, which is difficult to verify empirically and overly restrictive. Only a limited amount of research has been proposed to support non-Gaussian multivariate spatial processes. Markov Chain Monte Carlo (MCMC) methods [Schmidt and Rodriguez 2011] are popular ways for addressing problems that are analytically intractable [Bandyopadhyay 2005; Ridgeway and Madigan 2002; Li et al. 2008; Zhang and Sheng 2004; Boley and Grosskreutz 2008]. However, it becomes prohibitively expensive for large spatial datasets . To the best of our knowledge, there is no previous work that addresses all the above challenging problems. This article proposes a flexible hierarchical Bayesian framework that supports simultaneous modeling and prediction of mixed-type observations. A joint distribution for multivariate spatial responses is specified indirectly through specific link functions and the complicated dependencies among them are captured by the“cross-covariances,” which are easily parameterized. Efficient approximations are integrated to estimate the posterior marginals of latent variables. In order to reduce the computational burden, we project the spatial process to a lower dimension space by utilizing knot-based techniques [Banerjee et al. 2008]. Our generic approach model can also be applied to several data mining problems, including spatial outlier detection [Wu et al. 2009, 2010], spatial temporal outlier detection [Liu et al. 2011; Wu et al. 2008], spatial-temporal scan [Mohammadi et al. 2009], and spatial anomaly cluster identification [Neill and Moore 2004]. The major contributions of this work can be summarized as follows.

—*Constructing novel multivariate non-Gaussian hierarchical framework*. The spatial model is based on a hierarchical framework and designed to take account of mixed type random variables. Specifically, the mixed-type attributes are mapped to latent numerical random variables via corresponding link functions, such as the logit function for binary attributes and the log function for count attributes.
—*Capturing dependencies among mixed-type responses*. The dependency among mixed type attributes is mapped to the relationship between their latent variables using a conditional variance covariance matrix. This enables the complicated correlations to be captured more easily by parameterizing them in an analytically tractable way.
—*Modeling a multivariate non-Gaussian reduced-rank predictive process.* The knot-based technique is utilized to model the predictive process as a reduced-rank spatial process, which projects the process realizations of the spatial model onto a lower dimensional subspace. This projection significantly reduces the computational cost.
—*Designing an enhanced statistical approximation.* By integrating the link functions into spatial references, the likelihood models involved are no longer analytically tractable. Gaussian approximation and iterative Laplace approximation (LA) can then be utilized to make approximations to the posterior marginal of latent variables for the predictive processes.
—*Conducting extensive experiments for performance evaluations.* Theoretical analysis and extensive experiments on both simulations and real datasets have been conducted for this study. The results clearly demonstrated the superior performance of the proposed hierarchical mixed model compared to existing state-of-the art comparison approaches, with comparable prediction accuracy and computational efficiency.

The remainder of this article is organized as follows. Section 2 reviews related works. Section 3 presents a generic multivariate non-Gaussian model that can not only model mixed-type data (Challenge 1), but also capture correlations among them (Challenge 2). This section also introduces the reduced rank spatial predictive process (Challenge 3). Section 4 proposes an approximate inference for multivariate spatial prediction to solve the issue of analytical intractable inference (Challenge 4). Experiments on both simulated and real datasets are presented in Section 5, and the article concludes with a summary of the research in Section 6.

2. RELATED WORK

This section summarizes the current status of research achievements on spatial references, including spatial multivariate prediction for numeric data, spatial temporal multivariate prediction, and spatial multivariate prediction for mixed-Type Data.

Spatial Multivariate Analysis for Numeric Data. Most research works [Bailey and Krzanowski 2000; Wang and Wall 2003] on multivariate spatial analysis have focused on capturing potential relationships among spatial locations and multiple responses. The book by Wackernagel [2003] and the review by Gelfand and Banerjee [2010] provide comprehensive surveys of a wide range of different spatial Gaussian multivariate modeling and prediction techniques. For example, cokriging [Goovaerts 1997] exploits the spatial dependencies within the variables as well as the cross-spatial dependencies. Bailey and Krzanowski [2000] proposed two approaches that are concerned with the identification of linear components and identifying factors. Wang and Wall [2003] designed a generalized common spatial factor model in which the parameters are estimated using the Bayesian method and MCMC techniques, and Ren and Banerjee [2013] discussed how to capture associations among responses by reducing the dimensions of both the length of the response vectors and the very large number of spatial locations. Christensen and Amemiya [2001] developed a generalized shifted-factor model that allows asymmetric spatial dependencies, and then they [Christensen and Amemiya 2002] went on to propose a latent variable-based approach to fitting model and estimating parameter. Bonilla et al. [2008] described multi-task learning based on a GP prior that has inter-task dependencies. The model utilized a convariance function on multiple features under the assumption of noise-free observations. Kanevski [2012] proposed a generic non-linear multivariate modelling by using the best MTL (Multitask Learning) group. The model was based on the criterion of nonlinear predictability of each dependent variable by analyzing all possible models composed from the rest of the variables. Finally, Minozzo and Ferracuti [2012] provided some valid constructions of stationary stochastic processes that are capable of modeling multivariate skew-normal data.

Multivariate Spatial-Temporal Data Analysis. Various approaches have been proposed for analyzing multivariate space-time data. Calder [2007] introduced a Bayesian convolution model that provides a descriptive parametrization of the cross-covariance structure of space-time processes and dimension reduction features. Zhu et al. [2005] extended a multivariate space-only model to space–time data by utilizing an adjustment of the Monte Carlo Expectation-Maximization algorithm. Reich and Fuentes [2007] designed a new Bayesian multivariate spatial statistical framework that builds on the stick breaking prior to handling sudden changed data in time or space. Choi et al. [2009] introduced a Bayesian hierarchical framework wherein a linear model of coregionalization was developed to account for spatial and temporal dependency for each observation as well as the correlations among them. Grzebyk and Wackernagel [1994] presented the Bilinear Model that is suitable for modeling a coregionalization in space or along the time axis.

Spatial Multivariate Prediction for Mixed-Type Data. A number of related works have focused on non-Gaussian multivariate domains. Wibrin et al. [2006] explored the Bayesian Maximum Entropy (BME) approach in which both continuous and categorical values are considered using a “cross-covariance” function. Schmidt and Rodriguez [2011] proposed MCMC methods for modeling multivariate counts, while Chagneau et al. proposed a hierarchical Bayesian model for the modeling of Gaussian, count, and ordinal variables, and designed MCMC methods using the Gibbs sampler with Metropolis–Hastings (M-H) steps. Minozzo and Fruttini [2004] proposed a model based on a generalized linear mixed multivariate framework. By integrating Monte Carlo Expectation-Maximization, Minozzo and Ferrari [2013] designed another hierarchical model in which non-Gaussian variables of different kinds can be processed simultaneously. However, most of these methods are unable to provide a flexible framework that supports multivariate mixed-type data inferences simultaneously, including binomial, count, nominal, ordinal, and numeric data types.

3. THEORETICAL BACKGROUND

This section introduces the exponential family, the framework for the knot-based spatial process and the approximation Bayesian inference using INLA.

3.1. The Exponential Family

Let $Y(\mathbf{s})$ be a response variable at the location $\mathbf{s} \in \mathcal{D} \subset \mathcal{R}^2$. It is assumed that $Y(\mathbf{s})$ follows an exponential family distribution with the probability density:

$$f(Y(s)|\theta(\mathbf{s}), \tau) = \exp\left(\frac{Y(\mathbf{s})\theta(\mathbf{s}) - a(\theta(\mathbf{s}))}{d(\tau)} + h(Y(\mathbf{s}), \tau)\right), \tag{1}$$

where $\theta(\mathbf{s})$ and τ are the model parameters. $\theta(\mathbf{s})$ is related to the mean of the distribution that varies by location, and τ is a dispersion parameter related to the variance of the distribution. The functions $h(y(\mathbf{s}), \tau)$, $a(\theta(\mathbf{s}))$, and $d(\tau)$ are known. $Y(\mathbf{s})$ has mean and variance

$$E(Y(\mathbf{s})) := \mu(\mathbf{s}) = a'(\theta(\mathbf{s})), \tag{2}$$

$$Var(Y(\mathbf{s})) := \sigma(\mathbf{s})^2 = a''(\theta(\mathbf{s}))d(\tau), \tag{3}$$

where $a'(\theta(\mathbf{s}))$ and $a''(\theta(\mathbf{s}))$ are the first and second derivatives of $a(\theta(\mathbf{s}))$. Many popular distributions belong to this family, including the Gaussian, exponential, Binomial, Poisson, gamma, Inverse Gaussian, Dirichlet, and Chi-Squared Beta distributions.

For example, the Binomial distribution $B(n(\mathbf{s}), \pi(\mathbf{s}))$ has the density

$$p(Y(\mathbf{s})) = \binom{n(\mathbf{s})}{Y(\mathbf{s})}\pi(\mathbf{s})^{Y(\mathbf{s})}(1 - \pi(\mathbf{s}))^{n(\mathbf{s})-Y(\mathbf{s})}. \tag{4}$$

Taking log, we can rewrite the density function as

$$\log p(Y(\mathbf{s})) = Y(\mathbf{s})\log\left(\frac{\pi(\mathbf{s})}{1 - \pi(\mathbf{s})}\right) + n(\mathbf{s})\log(1 - \pi(\mathbf{s})) + \log\binom{n(\mathbf{s})}{Y(\mathbf{s})}. \tag{5}$$

This shows that $\theta(\mathbf{s}) = \log(\frac{\pi(\mathbf{s})}{1-\pi(\mathbf{s})})$, $a(\theta(\mathbf{s})) = n(\mathbf{s})\log(1 + \exp\theta(\mathbf{s}))$, and $h(Y(\mathbf{s}), \tau) = \log(\binom{n(\mathbf{s})}{Y(\mathbf{s})})$, where the second term in the density function is rewritten as $\log(1 - \pi(\mathbf{s})) = -\log(1 + \exp\theta(\mathbf{s}))$.

3.2. Knot-Based Spatial Process Model

Estimation and prediction in spatial process models often involve a high computational complexity, which is cubic order with the number of spatial locations. To facilitate the spatial process, Banerjee et al. [2008] proposed a knot-based spatial predictive model to reduce the computational cost through lower dimensional process observations.

Let us define a numerical random field $Y(s)$ on a domain $D \subseteq \mathcal{R}^2$, and let $Y = (Y(s_1), \dots, Y(s_n))'$ be the $n \times 1$ vector of observed responses, each of which is accompanied by a $p \times 1$ vector of spatially referenced predictors, $x(s)$. The associated spatial regression model can be represented as

$$Y(s) = x^T(s)\beta + \omega(s) + \epsilon(s). \tag{6}$$

The spatial process $\omega(s)$ captures spatial correlations and is a GP with zero mean and a covariance function $C(s, s'; \theta)$. Spatial prediction requires matrix factorizations involving the dense $n \times n$ covariance matrix that may become prohibitively expensive for a large n. Instead, knot-based models consider a fixed set of "knots" $S^* = (s_1^*, \dots, s_{n^*}^*)$ with $n^* \ll n$. The GP $\omega^*(s)$ yields an n^*-vector of realizations over the knots, that is, $\omega^* = (\omega(s_1^*), \dots, \omega(s_{n^*}^*))'$, which follows a $GP\{0, C(s_i^*, s_j^*; \theta)\}$. Spatial estimation at a generic site s is operated through

$$\tilde{\omega}(s) = E\{\omega(s)|\omega^*\} = c^T(s;\theta)C^{*-1}(\theta)\omega^*, \tag{7}$$

where $c(s;\theta) = [C(s, s_j^*; \theta)_{j=1}^{n^*}]$. As shown in Equation (2), the *predictive process* $\tilde{w}(s)$ is derived from the *parent process* $\omega(s)$. The realizations of $\tilde{\omega}(s)$ are referred to as the predictions that are conditional on a realization of $\omega^*(s)$. Replacing $\omega(s)$ in model (6) with $\tilde{\omega}(s)$, we obtain the predictive process model

$$Y(s) = x^T(s)\beta + \tilde{\omega}(s) + \epsilon(s), \tag{8}$$

where $\tilde{\omega}(s)$ is defined as a spatially varying linear transformation of ω^*. The dimension reduction is reduced from the original n to n^*; thus, the spatial interpolation process involves only $n^* \times n^*$ matrices.

It is important to select an appropriate number of knots as well as their spatial locations. This is related to the problem of spatial design. There are two popular knots selection strategies. One is to draw a uniform grid to cover the study region and each grid is considered as a knot. Another is to place knots such that each covers a local domain and the regions with dense data have more knots. In practice, it is feasible to validate models by using different numbers of knots and different choices of knots to obtain a reliable and robust configuration.

3.3. Approximate Bayesian Inference Using INLA

The INLA [Rue et al. 2009] is a computational approach that is proposed as an alternative of the time-consuming MCMC method. It approximates the marginal posteriors of latent variables

$$\pi(v_i|Y) = \int \pi(v_i|\theta, Y)\pi(\theta|Y)d\theta. \tag{9}$$

This approximation is an efficient combination of LAs to the full conditionals $\pi(\theta|Y)$ and $\pi(v_i|\theta, Y)$, and finally executes numerical integration routines by integrating out the parameter θ.

The INLA approach consists of three main approximations to obtain the marginal posterior for each latent variable. The first step is to approximate the full posterior $\pi(\theta|Y)$, which is executed using the LA

$$\tilde{\pi}(\theta|Y) \propto \frac{\pi(v, \theta, Y)}{\tilde{\pi}_G(v|\theta, Y)}\Big|_{v=v^*(\theta)}. \tag{10}$$

As shown above, we need to approximate the full conditional distribution of $\pi(v|Y, \theta)$, which can be achieved by a multivariate Gaussian density $\tilde{\pi}_G(v|Y, \theta)$ [Rue and Held 2005]. $v^*(\theta)$ is the mode of the full conditional distribution of v for a given θ and can be estimated using $\tilde{\pi}_G(v|Y, \theta)$. The posterior $\tilde{\pi}(\theta|Y)$ will be used later to integrate out the uncertainty with respect to θ when approximating $\pi(v_i|Y)$.

The second step executes the LA of the full conditionals $\pi(v_i|\theta, Y)$ for specified θ values. The density $\pi(v_i|\theta, Y)$ is approximated using the LA defined by

$$\tilde{\pi}_{LA}(v_i|\theta, Y) \propto \frac{\pi(v, \theta, Y)}{\tilde{\pi}_G(v_{-i}|v_i, \theta, Y)}\Big|_{v_{-i}=v^*(v_i,\theta)}, \tag{11}$$

where $\tilde{\pi}_G(v_{-i}|v_i, \theta, Y)$ refers to the Gaussian approximation of $\pi(v_{-i}|v_i, \theta, Y)$ that takes v_i as a fixed value. $v^*(v_i, \theta)$ is the mode of $\pi(v_{-i}|v_i, \theta, Y)$.

Finally, we can approximate the marginal posterior density of v_i by combining the full posteriors obtained in the previous steps. The approximation expression is shown as follows:

$$\pi(v_i|Y) \approx \sum_k \tilde{\pi}(v_i|\theta_k, Y)\tilde{\pi}(\theta_k|Y)\triangle_k. \tag{12}$$

It is a numerical summation on a representative set of θ_k, with the area weight, $\triangle_k$ for $k = 1, \ldots, K$. Note that a good choice of the set of θ_k is crucial to the accuracy of the above numerical integration.

4. SPATIAL MULTIVARIATE NON-GAUSSIAN MODEL

This section presents a spatial process model for random variables that is capable of dealing with Gaussian and non-Gaussian variables. The new multivariate model proposed here is designed based on a Bayesian hierarchical framework that allows any number of mixed-type response variables (Challenges 1 and 2). The computational burden of modeling large mixed-type spatial datasets is addressed by integrating the knot-based predictive process (Challenge 3). Table I summarizes the key notations used in this article.

4.1. Model Formulation

The spatial multivariate predictive model is specifically designed to deal with responses of different types, which are assumed to follow an exponential family distribution. Here, we consider two different types: Gaussian and non-Gaussian variables (e.g., Poisson).

Let $s_1, \ldots, s_n$ be the n sampled locations, $Y(s_i)$ be the Gaussian variable at location s_i, and $Z(s_i)$ be the non-Gaussian variable, such as the Poisson variable. Let $Y = (Y(s_1), \ldots, Y(s_n))'$ and $Z = (Z(s_1), \ldots, Z(s_n))'$. Geostatistics typically assumes that the Gaussian response variable $Y(s)$ is modeled as a spatial regression model with a $p \times 1$ vector of spatially referenced predictors, $x(s)$, such as

$$Y(s) = x(s)^T \beta_y + \omega(s) + \epsilon(s). \tag{13}$$

The residual includes the spatial random effect, $\omega(s)$, and the independent process $\epsilon(s)$, known as the measurement error. Usually, $\epsilon(s) \sim \mathcal{N}(0, \tau^2)$. $\omega(s)$ provides a local adjustment to the mean, interpreted as the effect of unmeasured covariates on the spatial pattern.

Let the first stage of Z be the non-GP. Essentially, we assume that the function of the expected value of $Z(s_i)$ is linear on a transformed scale, such as

$$\eta_z(s) \equiv g(E[Z(s)|\theta_Z]) = x(s)^T \beta_z + \gamma(s), \tag{14}$$

where $g(\cdot)$ is a suitable link function, and θ_Z is the parameter set of process $Z(s)$.

The Gaussian variable $Y(s_i)$ and the non-Gaussian variable $Z(s_i)$ depend on the latent variables $\omega(s_i)$ and $\gamma(s_i)$, respectively, which together are responsible for the spatial dependences. Given $\omega(s_i)$ and $\gamma(s_i)$, the variables $Y(s_i)$ and $Z(s_i)$ are conditionally independent. The customary process specification for $(\omega', \gamma')'$ is a mean zero GP with covariance function C, denoted as $GP(0, C)$. The most obvious specification of a valid

Table I. Description of Major Symbols

Symbol	Description
S	$S = \{s_1, \ldots, s_n\}$, a set of n training locations, where $s_i \in \mathbb{R}^2$;
S^*	$S^* = \{s_1^*, \ldots, s_m^*\}$, a set of m knot locations, where $s_i^* \in \mathbb{R}^2$;
Y	A given set of observations with a numerical attribute that follows a Gaussian distribution. $Y = \{Y(s_i)\}_{i=1}^n$;
Z	A given set of observations with a discrete attribute. If a count dataset (Z_c), it follows a Poisson distribution; if a binary dataset (Z_b), it follows a Binomial distribution. $Z = \{Z(s_i)\}_{i=1}^n$;
X	A set of explanation variables. $\{X(s_i)\}_{i=1}^n$ is a $p \times 1$ vector at location s_i. $X = \{X(s_i)\}_{i=1}^n$;
ω, γ	Spatial random effects of the observations, which provide local adjustments to the means, and is interpreted as the effects of unmeasured covariates with spatial patterns. $\omega = \{\omega(s_i)\}_{i=1}^n$,$\gamma = \{\gamma(s_i)\}_{i=1}^n$;
ω^*, γ^*	Spatial random effects of the knots. $\omega^* = \{\omega^*(s_i)\}_{i=1}^m$, $\gamma^* = \{\gamma^*(s_i)\}_{i=1}^m$;
$\tilde{\omega}, \tilde{\gamma}$	The predicted values of ω, γ by ω^*, γ^*. $\tilde{\omega} = \{\tilde{\omega}(s_i)\}_{i=1}^n$,$\tilde{\gamma} = \{\tilde{\gamma}(s_i)\}_{i=1}^n$;
ϵ_y	The estimation bias values for $\tilde{\omega}$. $\epsilon_y(s) = \{\epsilon_y(s_i)\}_{i=1}^n$, $\{\tilde{\omega}_\epsilon(s_i) = \tilde{\omega}(s_i) + \epsilon_y(s_i)\}_{i=1}^n$
ϵ	$\{\epsilon(s_i)\}_{i=1}^n$ is the measurement error for $\{Y(s_i)\}_{i=1}^n$. $\epsilon = \{\epsilon(s_i)\}_{i=1}^n$;
v^*	$v^* = ((\omega^{*'}, \gamma^{*'}), (\beta_y', \beta_z'))'$, it is a $(2m + 2p) \times 1$ vector comprising the realizations of the spatial multivariate predictive process and the regression parameters;
ϕ	The decay and smoothness parameter;
$F(\phi)$	A transformation matrix that defines $\{\tilde{\omega}, \tilde{\gamma}\}$ as a spatially varying linear transformation of $\{\omega^*, \gamma^*\}$.
η_z	The expected value of Z which is linear on a transformed scale. $\eta_z = \{\eta_z(s_i)\}_{i=1}^n$. $\eta_z = H_z^* v^*$.
H_y^*	$H_y^* = [F_y(\phi), [X\, 0_{n \times p}]]$. $F_y(\phi)$ consists of the first n rows of matrix $F(\phi)$;
H_z^*	$H_z^* = [F_z(\phi), [0_{n \times p}\ X]]$. $F_z(\phi)$ consists of the last n rows of matrix $F(\phi)$;
S_θ	The set of sample locations of θ based on the mode and Hessian at it of $\hat{\pi}(\theta \mid Y, Z)$. $S_\theta = \{\theta_k\}_{k=1}^K$;
w	The set of weighted values of sample θ, which are computed by their corresponding posterior distributions. $w = \{w_{\theta_k}\}_{k=1}^K$;

cross-covariance function C for $(\omega', \gamma')'$ is to let ρ be a valid correlation function for a univariate spatial process. Let T be a $d \times d$ (here $d = 2$ refers to the dimension of the dataset) positive definite matrix $T = (\begin{smallmatrix} \sigma_y^2 & \sigma_{yz}^2 \\ \sigma_{yz}^2 & \sigma_z^2 \end{smallmatrix})$, which is interpreted as the covariance matrix associated with $(\omega', \gamma')'$. T follows an Inverse Wishart distribution, denoted as $T \sim \mathcal{W}^{-1}(\Psi, m)$, and $\rho(s_i, s_j; \phi)$ attenuates the association as s_i and s_j become farther apart. The covariance matrix for $(\omega', \gamma')'$ can be easily shown to be

$$\Sigma_{(\omega', \gamma')'} = T \otimes R(\phi), \tag{15}$$

where $R(\phi)_{i,j} = \rho(s_i, s_j; \phi)$ is a correlation function, like Exponential, Gaussian and Spherical, and so on. ϕ includes both decay and smoothness parameters, yielding constant process variances, and $\otimes$ denotes the Kronecker product.

The prior distributions of the remaining parameters construct the third level of the hierarchical model. Customarily, the regression parameters β_y and β_z are assigned multivariate Gaussian priors, i.e., $\beta_y \sim \mathcal{N}(\mu_{\beta_y}, \Sigma_{\beta_y})$, $\beta_z \sim \mathcal{N}(\mu_{\beta_z}, \Sigma_{\beta_z})$, while the latent variance components σ_y, σ_z, and σ_{yz} are assigned $\mathcal{W}^{-1}$ as described above. The measurement error variance τ^2 is assigned an $\mathcal{G}^{-1}(a_\tau, b_\tau)$ prior (Inverse Gamma). The process correlation parameter ϕ is usually assigned an informative prior (e.g., uniform over a finite range) based on the underlying spatial domain.

Let Y and Z be two $n \times 1$ vectors of observed responses. The mixed data likelihood can be obtained by combining their hierarchical specifications, as shown in Figure 1,

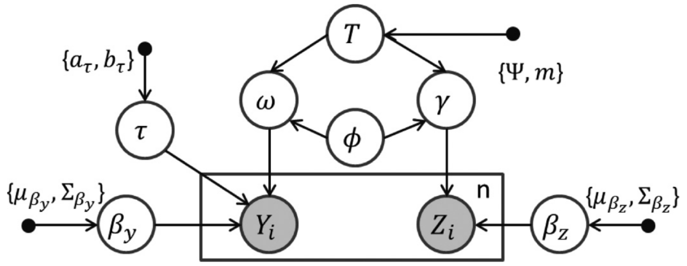


Fig. 1. Graphical model representation.

to derive a posterior distribution $\pi(\beta_y, \beta_z, \omega, \gamma, T, \tau^2, \phi|Y, Z)$ that is proportional to

$$\begin{aligned}&\pi(\phi) \times \mathcal{G}^{-1}(\tau^2|a_\tau, b_\tau) \times \mathcal{W}^{-1}(T|\Psi, m) \times \mathcal{N}\left(\begin{pmatrix}\omega\\ \gamma\end{pmatrix}|0, \Sigma_{(\omega',\gamma')'}\right)\\ &\times \mathcal{N}(\beta_y|\mu_{\beta_y}, \Sigma_{\beta_y}) \times \prod_{i=1}^{n} \mathcal{N}(Y(s_i)|x(s_i)^T\beta_y + \omega(s_i), \tau^2)\\ &\times \mathcal{N}(\beta_z|\mu_{\beta_z}, \Sigma_{\beta_z}) \times \prod_{i=1}^{n} \pi(Z(s_i)|x(s_i)^T\beta_z + \gamma(s_i))).\end{aligned} \tag{16}$$

4.2. Reduced-Rank Spatial Multivariate Non-Gaussian Process

For the spatial multivariate non-GP model, both the estimation and prediction steps require the $(d * n) \times (d * n)$ covariance matrix to be evaluated for the d dependent response variables. Unfortunately, fitting hierarchical mixed models often involves expensive matrix decompositions whose computational cost is $O((d*n)^3)$, thus rendering such models not scalable for large spatial datasets. To facilitate the spatial process, a knot-based technique [Banerjee et al. 2008] can be utilized to reduce the computational cost by lowering dimensional process, as this only requires a fixed set of "knots" over which spatial estimation is operated to be considered. In this section, the spatial multivariate predictive model, referred to here as the reduced-rank spatial multivariate non-GP, is constructed by projecting the full process into a subspace generated by a specified set of representative locations. To generate the knots, a uniform grid is plotted across the whole study region. Each grid is considered as a knot.

Consider a set of "knots," $S^* = \{s_1^*, \ldots, s_m^*\}$, representing the vector of corresponding centroids of the m spatial clusters generated by the spatial attributes of the dataset. The latent variables $(\omega^{*'}, \gamma^{*'})'$ follow a mean zero Gaussian distribution with the covariance function C^*, denoted as $GP(0, C^*)$. The covariance matrix for $(\omega^{*'}, \gamma^{*'})'$ is $\Sigma_{(\omega^{*'},\gamma^{*'})'} = T \otimes R^*(\phi)$. $R^*(\phi)$ is the corresponding $m \times m$ covariance matrix, where $R^*(\phi)_{i,j} = \rho(s_i^*, s_j^*; \phi)_{i,j=1,\ldots,m}$. The spatial interpolant at a site s_0 is estimated by

$$\begin{aligned}\begin{pmatrix}\tilde{\omega}(s_0)\\ \tilde{\gamma}(s_0)\end{pmatrix} &= E\left\{\begin{pmatrix}\omega(s_0)\\ \gamma(s_0)\end{pmatrix}\middle|\begin{pmatrix}\omega^*\\ \gamma^*\end{pmatrix}\right\} = \Upsilon(s_0)\Sigma^{-1}_{(\omega^{*'},\gamma^{*'})'}\begin{pmatrix}\omega^*\\ \gamma^*\end{pmatrix}\\ &= \begin{pmatrix}f_\omega^\omega(s_0) & f_\omega^\gamma(s_0)\\ f_\gamma^\omega(s_0) & f_\gamma^\gamma(s_0)\end{pmatrix}\begin{pmatrix}\omega^*\\ \gamma^*\end{pmatrix}.\end{aligned} \tag{17}$$

Here, $\Upsilon(s_0) = T \otimes r(s_0; \phi)'$, and $r(s_0; \phi)$ is an $m \times 1$ vector whose jth element is given by $\rho(s_0, s_j^*; \phi)$. The f series represent four $1 \times m$ matrices. This yields a spatial GP $(\tilde{\omega}', \tilde{\gamma}')' \sim \mathcal{N}(0, T \otimes \tilde{\rho})$, where $\tilde{\rho}(s_i, s_j; \phi) = \Upsilon(s_i)\Sigma^{-1}_{(\omega^{*'},\gamma^{*'})'}\Upsilon'(s_j)$, and $(\tilde{\omega}', \tilde{\gamma}')'$ is referred to as the *predictive process* derived from the *parent process* $(\omega', \gamma')'$. As shown in Equation (17), $(\tilde{\omega}(s)', \tilde{\gamma}(s)')'$ is a spatially adaptive linear transformation of the realizations of $(\omega(s)', \gamma(s)')'$ over S^* with $\Upsilon(s_0)\Sigma^{-1}_{(\omega^{*'}, \gamma^{*'})'}$ comprising the coefficients of the transformation.

Replacing $\omega(s)$ and $\gamma(s)$ in Equations (13) and (14) with $\tilde{\omega}$ and $\tilde{\gamma}$, we obtain the reduced-rank predictive model,

$$Y(s) = x(s)^T \beta_y + \tilde{\omega}(s) + \epsilon(s), \tag{18}$$

$$\eta_z(s) \equiv g(E[Z(s)|\theta_Z]) = x(s)^T \beta_z + \tilde{\gamma}(s). \tag{19}$$

Using Equations (38) and (39) as the likelihood, we derive a posterior distribution $\pi(\beta_y, \beta_z, \omega^*, \gamma^*, T, \tau^2, \phi | Y, Z)$ that is proportional to

$$\begin{aligned} &\pi(\phi) \times \mathcal{G}^{-1}(\tau^2 | a_\tau, b_\tau) \times \mathcal{W}^{-1}(T|\Psi, m) \times \mathcal{N}(\beta_y | \mu_{\beta_y}, \Sigma_{\beta_y}) \times \mathcal{N}(\beta_z | \mu_{\beta_z}, \Sigma_{\beta_z}) \\ &\quad \times \mathcal{N}\left(\begin{pmatrix} \omega^* \\ \gamma^* \end{pmatrix} | 0, \Sigma_{(\omega^{*'}, \gamma^{*'})'}\right) \times \prod_{i=1}^{n} \mathcal{N}(Y(s_i) | x(s_i)^T \beta_y + \tilde{\omega}(s_i), \tau^2) \\ &\quad \times \prod_{i=1}^{n} \pi(Z(s_i) | x(s_i)^T \beta_z + \tilde{\gamma}(s_i))). \end{aligned} \tag{20}$$

The reduced variability in $\tilde{\omega}$ often incurs an overestimation of the measurement error variance τ^2. Banerjee et al. [2008] explained these biases. The predictive process systematically underestimates the variance of $(\omega', \gamma')'$ at any location s. It has $0 \leq var((\omega'(s), \gamma(s)')' | (\omega^{*'}(s), \gamma(s)^{*'})') = T - \Upsilon(s)\Sigma^{-1}_{(\omega^{*'}, \gamma^{*'})'}\Upsilon'(s)$, which denotes the bias underestimation over the observed locations. With regard to this issue, Finley et al. [2009] proposed replacing $\tilde{\omega}(s)$ and $\tilde{\gamma}(s)$ in Equations (38) and (39) with $\tilde{\omega}_\epsilon(s) = \tilde{\omega}(s) + \tilde{\epsilon}_y(s)$ and $\tilde{\gamma}_\epsilon(s) = \tilde{\gamma}(s) + \tilde{\epsilon}_z(s)$. $\left(\begin{smallmatrix} \tilde{\epsilon}_y \\ \tilde{\epsilon}_z \end{smallmatrix}\right)$ represents a process involving independent variables with spatially adaptive variances. Using $\tilde{\omega}_\epsilon(s)$ and $\tilde{\gamma}_\epsilon(s)$ in place of $\tilde{\omega}(s)$ and $\tilde{\gamma}(s)$ for the spatial process yields

$$\begin{aligned} &\pi(\phi) \times \mathcal{G}^{-1}(\tau^2 | a_\tau, b_\tau) \times \mathcal{W}^{-1}(T|\Psi, m) \times \mathcal{N}(\beta_y | \mu_{\beta_y}, \Sigma_{\beta_y}) \times \mathcal{N}(\beta_z | \mu_{\beta_z}, \Sigma_{\beta_z}) \\ &\quad \times \mathcal{N}\left(\begin{pmatrix} \omega^* \\ \gamma^* \end{pmatrix} | 0, \Sigma_{(\omega^{*'}, \gamma^{*'})'}\right) \times \mathcal{N}\left(\begin{pmatrix} \tilde{\omega}_\epsilon \\ \tilde{\gamma}_\epsilon \end{pmatrix} \middle| F(\phi)\begin{pmatrix} \omega^* \\ \gamma^* \end{pmatrix}, \Sigma_{(\tilde{\epsilon}'_y, \tilde{\epsilon}'_z)'}\right) \\ &\quad \times \prod_{i=1}^{n} \mathcal{N}(Y(s_i) | x(s_i)^T \beta_y + \tilde{\omega}_\epsilon(s_i), \tau^2) \times \prod_{i=1}^{n} \pi(Z(s_i) | x(s_i)^T \beta_z + \tilde{\gamma}_\epsilon(s_i))). \end{aligned} \tag{21}$$

$F(\phi)$ is a transformation matrix that defines $\{\tilde{\omega}, \tilde{\gamma}\}$ as a spatially varying linear transformation of $\{\omega^*, \gamma^*\}$. According to Equation (17), $\left(\begin{smallmatrix} \tilde{\omega}(s_0) \\ \tilde{\gamma}(s_0) \end{smallmatrix}\right) = \Upsilon(s_0)\Sigma^{-1}_{(\omega^{*'}, \gamma^{*'})'}\left(\begin{smallmatrix} \omega^* \\ \gamma^* \end{smallmatrix}\right)$ and $\Upsilon(s_0) = T \otimes r(s_0; \phi)'$, and $r(s_0; \phi)$ is an $m \times 1$ vector whose jth element is given by $\rho(s_0, s_j^*; \phi)$. Therefore, $F(\phi) = (T \otimes \mathcal{R}(\phi)')\Sigma^{-1}_{(\omega^{*'}, \gamma^{*'})'}$, where $\mathcal{R}(\phi)'$ is an $n \times m$ matrix whose ith row is given by $r(s_i; \phi)'$, and $r(s_i; \phi)$ is an $m \times 1$ vector whose jth element is given by $\rho(s_i, s_j^*; \phi)$, for $i = 1, \ldots, n$, $j = 1, \ldots, m$. $\left(\begin{smallmatrix} \tilde{\epsilon}_y \\ \tilde{\epsilon}_z \end{smallmatrix}\right) \sim \mathcal{N}(0, \Sigma_{(\tilde{\epsilon}'_y, \tilde{\epsilon}'_z)'})$ and $\Sigma_{(\tilde{\epsilon}'_y, \tilde{\epsilon}'_z)'}$ is a $2n \times 2n$ matrix that consists of four diagonal matrices $(n \times n)$ in which the following four specified diagonal elements $\left(\begin{smallmatrix} (i,i)\text{th} & (i+n,i)\text{th} \\ (i,i+n)\text{th} & (i+n,i+n)\text{th} \end{smallmatrix}\right)$ are computed as $T - \Upsilon(s_i)\Sigma^{-1}_{(\omega^{*'}, \gamma^{*'})'}\Upsilon'(s_i)$, where $\Upsilon(s_i) = T \otimes r(s_i; \phi)'$. Finley et al. [2009] detailed the estimation of the modified predictive process.

Let $v^* = ((\omega^{*'}, \gamma^{*'}), (\beta_y', \beta_z'))'$ be a $(2m + 2p) \times 1$ vector comprising the realizations of the spatial multivariate predictive process and the regression parameters. Since Z is related to the discrete variables, we assume there is no estimation bias [Rue et al. 2009]. The posterior $\pi(v^*, T, \phi, \tau^2|Y, Z)$ is proportional to

$$\begin{aligned}&\pi(\phi) \times \mathcal{G}^{-1}(\tau^2|a_\tau, b_\tau) \times \mathcal{W}^{-1}(T|\Psi, m) \times \mathcal{N}(v^*|\mu_{v^*}, \Sigma_{v^*})\\ &\times \prod_{i=1}^{n} \mathcal{N}(Y(s_i)|x(s_i)^T \beta_y + f_\omega^\omega(s)\omega^* + f_\omega^\gamma(s)\gamma^* + \epsilon_y(s), \tau^2)\\ &\times \prod_{i=1}^{n} \pi(Z(s_i)|x(s_i)^T \beta_z + f_\gamma^\omega(s)\omega^* + f_\gamma^\gamma(s)\gamma^*), \end{aligned} \tag{22}$$

where $\mu_{v^*} = (0_{1\times 2m}, \mu_{\beta_y}', \mu_{\beta_z}')'$ and the $(2m + 2p) \times (2m + 2p)$ covariance matrix is

$$\Sigma_{v^*} = \begin{bmatrix} \Sigma_{(\omega^{*'}, \gamma^{*'})'} & 0_{2m\times p} & 0_{2m\times p} \\ 0_{p\times 2m} & \Sigma_{\beta_y} & 0_{p\times p} \\ 0_{p\times 2m} & 0_{p\times p} & \Sigma_{\beta_z} \end{bmatrix}. \tag{23}$$

Under Gaussian likelihood assumptions,

$$\mathcal{N}(Y|H_y^* v^*, \tau_y^2 I_n + \epsilon_y I_n), H_y^* = [F_y(\phi), [\, X \;\, 0_{n\times p} \,]]. \tag{24}$$

The generalized linear model (GLM) likelihood model of Z can thus be defined by

$$\eta_z = H_z^* v^*, H_z^* = [F_z(\phi), [\, 0_{n\times p} \;\, X]]. \tag{25}$$

Here, $F_y(\phi)$ consists of the first n rows of matrix $F(\phi)$, and $F_z(\phi)$ consists of the last n rows of the matrix.

5. APPROXIMATE BAYESIAN INFERENCE

Since the likelihood model of the spatial multivariate observations is non-Gaussian, this renders the predictive process no longer analytically available (Challenge 4). To address this issue, we can formalize the multivariate predictive process by applying approximate Bayesian inference methods.

5.1. Gaussian Approximation to the Posterior Distribution of v^*

First, we need to approximate $\pi(\upsilon^*|Y, Z, \theta)$. For the predictive process model, the covariance parameters would be $\theta = (T, \phi, \tau^2)$. The simplest approximation to $\pi(\upsilon^*|Y, Z, \theta)$ is the Gaussian approximation. We have

$$\pi(\upsilon^*|Y, Z, \theta) \propto \pi(Y|\upsilon^*, \theta)\pi(Z|\upsilon^*, \theta)\pi(\upsilon^*|\theta), \tag{26}$$

where $\pi(Y, Z|\upsilon^*, \theta) = \pi(Y|\upsilon^*, \theta)\pi(Z|\upsilon^*, \theta)$ is derived based on the D-separation rules in the graphic model theory (see Figure 1). As discussed in Section (3.2), $\pi(Y|\upsilon^*, \theta)$ follows a Gaussian distribution, but $\pi(Z|\upsilon^*, \theta)$ does not. We therefore need to conduct a Gaussian approximation on $\pi(Z|\upsilon^*, \theta)$, and then on $\pi(Y, Z|\upsilon^*, \theta)$. Under the Gaussian distribution assumption $\mathcal{N}(Y|H_y^* v^*, \tilde{\epsilon}_y I_n + \tau^2 I_n)$ and the prior $v^* \sim \mathcal{N}(\mu^*, \Sigma^*)$, the full conditional distribution of υ^* conditional to $\{Y, \theta\}$ is thus

$$\begin{aligned}\pi(v^*|Y, \theta) &\propto \mathcal{N}(Y|H_y^* v^*, U)\mathcal{N}(\mu_v^*, \Sigma_v^*)\\ &\propto exp\left\{ \left[-\frac{1}{2}(Y - H_y^* v^*)' U^{-1}(Y - H_y^* v^*) \right] \right.\\ &\left. - \frac{1}{2}(v^* - \mu_v^*)' \Sigma_v^{*-1}(v^* - \mu_v^*) \right\} \propto exp\left(-\frac{1}{2} v^{*'} Q_y v^* + v^{*'} b_y \right),\end{aligned} \tag{27}$$

where $U = \tilde{\epsilon}_y I_n + \tau^2 I_n$, the full conditional precision matrix $Q_y = H_y^{*'} U^{-1} H_y^* + \Sigma_v^{*-1}$, and the canonical parameter $b_y = H_y^{*'} U^{-1} Y + \Sigma_v^{*-1} \mu_v^*$.

The likelihood model of Z is non-Gaussian, so we need to expand the likelihood in a quadratic form utilizing the Gaussian approximation. The GLM likelihood of Z is $\prod_i \pi(Z(s_i)|\eta_z(s_i))$, where the GLM parameter $\eta_z = H_z^* v^* = [F(\phi_z), [0_{n\times p} X]]v^*$.

The distributions in a natural exponential family take the form

$$\pi(Z|\eta_z) = exp\{\eta_z Z - f(\eta_z)\}h(Z). \tag{28}$$

For example, for a binomial distribution, $Binomial(1, \pi)$, $\eta_z = \log(\frac{\pi}{1-\pi})$, $f(\eta_z) = \log(1 + exp(\eta_z))$, and $h(Z) = 1$, while for the Poisson case, $Poisson(\lambda)$, $\eta_z = \log(\lambda)$, $f(\eta_z) = exp(\eta_z)$, and $h(Z) = \frac{1}{Z!}$.

By performing a Taylor expansion of $f(\eta_z) = f(H_z^* v^*)$ to the second order, we obtain the quadratic form of v^*,

$$\begin{aligned} \pi(Z|\eta_z) &\propto exp\left\{-\frac{1}{2} v^{*'} Q_z v^* + v^{*'} b_z\right\}, \\ Q_z &= H_z^{*'} \nabla^2 f(H_z^* \hat{v}^*) H_z^*, \\ b_z &= H_z^{*'} (Z - \nabla f(H_z^* \hat{v}^*) + \nabla^2 f(H_z^* \hat{v}^*) H_z^* \hat{v}^*). \end{aligned} \tag{29}$$

Combining Equations (26), (27), and (29) gives

$$\pi(v^*|Y, Z, \theta) \propto exp\left[-\frac{1}{2} v^{*'} (Q_y + Q_z) v^* + v^{*'} (b_y + b_z)\right]. \tag{30}$$

Finally, the full conditional precision matrix $Q = Q_y + Q_z$, and the canonical parameter $b = b_y + b_z$. Thus, the full conditional distribution is $\pi(v^*|Y, Z, \theta) \sim \mathcal{N}(Q^{-1}b, Q^{-1})$. We can compute the required inverse and determinant of the size $(2m + 2p) \times (2m + 2p)$ matrix Q by utilizing the structure of H_z^*, H_y^*, and Σ_v^*. Assuming $m \gg p$, the main cost of the matrix inversion is thus $O(m^3)$, since the number of knots is m. The supplementary material provides further details of the Taylor expansion for the Binomial and Poisson distributions.

5.2. Laplace Approximation for the Posterior Distribution of θ

Unlike $\pi(v^*|Y, Z)$, the posterior $\pi(\theta|Y, Z)$ is usually highly skewed, and its approximation as a Gaussian distribution is thus inappropriate [Rue et al. 2009]. The posterior $\pi(\theta|Y, Z)$ plays an important role in the inference of the marginal posterior of latent variables. Taking v^* as an example, we can estimate the marginal posterior $\pi(v^*|Y, Z)$, which takes the form of

$$\pi(v^*|Y, Z) = \int \pi(v^*|Y, Z, \theta)\pi(\theta|Y, Z) d\theta. \tag{31}$$

It is possible to obtain a sample set $\{\theta_1, \ldots, \theta_K\}$ from the input space of θ that represents an approximate discrete form of the posterior $p(\theta|Y, Z)$. We can estimate the approximate $\hat{p}(v^*|Y, Z)$ by

$$\hat{\pi}(v^*|Y, Z) = \sum_{k=1}^{K} \pi(v^*|Y, Z, \theta_k)\pi(\theta_k|Y, Z) w_{\theta_k}, \tag{32}$$

where w_{θ_k} is the weight of the sample point θ_k that can be measured by its normalized probability density. The critical step is to efficiently identify a representative sample set $\{\theta_1, \ldots, \theta_K\}$, as well as the corresponding set of weights $\{w_{\theta_1}, \ldots, w_{\theta_K}\}$.

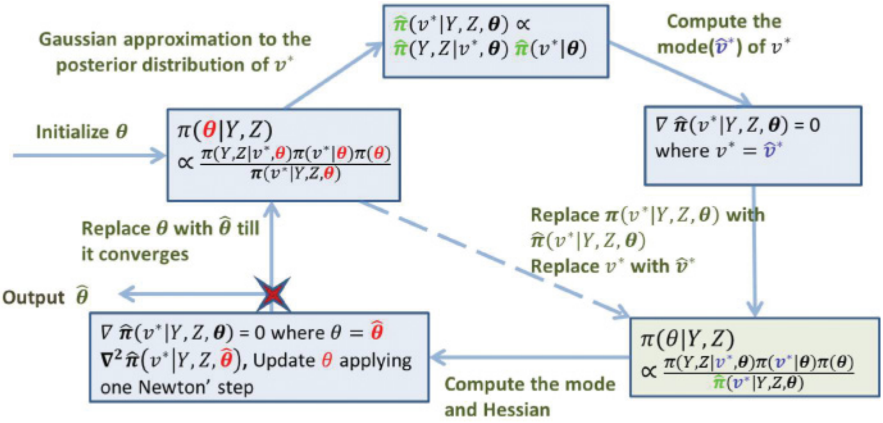

Gaussian approximation to the posterior distribution of v^*
$\hat{\pi}(v^*|Y,Z,\boldsymbol{\theta}) \propto \hat{\pi}(Y,Z|v^*,\boldsymbol{\theta})\,\hat{\pi}(v^*|\boldsymbol{\theta})$
Compute the mode($\hat{v}^*$) of v^*
Initialize θ
$\pi(\boldsymbol{\theta}|Y,Z) \propto \frac{\pi(Y,Z|v^*,\boldsymbol{\theta})\pi(v^*|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(v^*|Y,Z,\boldsymbol{\theta})}$
$\nabla\,\hat{\pi}(v^*|Y,Z,\boldsymbol{\theta}) = 0$ where $v^* = \hat{v}^*$
Replace θ with $\hat{\theta}$ till it converges
Replace $\pi(v^*|Y,Z,\theta)$ with $\hat{\pi}(v^*|Y,Z,\theta)$
Replace v^* with $\hat{v}^*$
Output $\hat{\theta}$
$\nabla\,\hat{\pi}(v^*|Y,Z,\boldsymbol{\theta}) = 0$ where $\theta = \hat{\theta}$
$\nabla^2\hat{\pi}(v^*|Y,Z,\hat{\boldsymbol{\theta}})$, Update θ applying one Newton' step
Compute the mode and Hessian
$\pi(\theta|Y,Z) \propto \frac{\pi(Y,Z|v^*,\theta)\pi(v^*|\theta)\pi(\theta)}{\hat{\pi}(v^*|Y,Z,\theta)}$

Fig. 2. Laplace approximation to posterior distribution.

The posterior $\pi(\theta^*|Y,Z)$ can be re-formalized, and the LA can be applied to approximate the denominator $\pi(\upsilon^*|Y,Z,\theta)$ as a Gaussian distribution. The LA method uses a similar approach to that utilized for Bayesian spatial inference:

$$\hat{\pi}(\theta|Y,Z) \propto \left.\frac{\pi(Y,Z|\upsilon^*,\theta)\pi(\upsilon^*|\theta)\pi(\theta)}{\hat{\pi}(\upsilon^*|Y,Z,\theta)}\right|_{\upsilon^*=\hat{\upsilon}^*}, \tag{33}$$

where $\hat{\pi}(\upsilon^*|Y,Z,\theta)$ is a Gaussian approximation.

Utilizing the above approximation yields the mode $\hat{\upsilon}^*$ and the curvature at the mode of this full conditional expression. In our framework, we apply the GLM to capture the distributions of non-Gaussian variables. The preceding Gaussian approximation can be efficiently conducted using the popular Iterated Re-weighted Least Squares (IRLS) algorithm. The detailed procedures are presented in Algorithm 1.

ALGORITHM 1: Exploring the Posterior Distribution of $\pi(\theta|Y,Z)$

```
Input: $S, S^*, S^0, Y, Z, X$
Output: $\hat{\theta}$
1: Choose an initial value $\theta = \{\tau^2, T, \phi\}$;
2: repeat
3:    Construct $\mu_{\upsilon^*}$, $\Sigma_{\upsilon^*}$ with $\theta$ (see Equation (23)).
4:    Calculate the transformation matrix $F(\phi)$.
5:    Calculate the likelihood of Y for Gaussian variables (see Equation (24)) and the GLM
      likelihoodof Z for exponential variables (see Equation (25)).
6:    Apply IRLS to find the mode $\hat{\upsilon}^*$ and Hessian matrix at the modeof $\hat{\pi}(\upsilon^*|Y,Z,\theta)$, then
      make a Gaussian approximation by applying Equation (30).
7:    Compute the gradient and Hessian matrix of $\hat{\pi}(\theta^*|Y,Z)$ andapply one Newton step to
      update $\theta$ as $\hat{\theta}$.
8: until Convergence
9: Output optimal $\hat{\theta}$.
```

As shown in Figure 2, Algorithm 1 describes the steps of exploring the posterior distribution of θ in terms of the following steps.

—*Step 1* (*line:* 1) *Initialize* θ. An initial value θ randomly is chose.
—*Step 2* (*lines:* 2–5) *Gaussian approximate the posterior distribution of* v^*. A Gaussian approximation on $\pi(Z|v^*,\theta)$ is conducted by expanding its likelihood in a quadratic form and further on $\pi(v^*|Y,Z,\theta)$.

—*Step 3* (*line: 6*) *Update* $\pi(\theta|Y, Z)$. The mode of v^* as $\hat{v}^*$ is calculated and $\pi(\theta|Y, Z)$ is updated by replacing v^* with $\hat{v}^*$ as $\tilde{\pi}(\theta|Y, Z) \propto \frac{\pi(Y,Z|\hat{v}^*,\theta)\pi(\hat{v}^*|\theta)\pi(\theta)}{\tilde{\pi}(\hat{v}^*|Y,Z,\theta)}$.

—*Step 4* (*line: 7*) *Update* $\tilde{\pi}(\theta|Y, Z)$. The gradient and Hessian matrix of $\tilde{\pi}(\theta|Y, Z)$ are computed and $\tilde{\pi}(\hat{\theta}|Y, Z) \propto \frac{\pi(Y,Z|\hat{v}^*,\hat{\theta})\pi(\hat{v}^*|\hat{\theta})\pi(\hat{\theta})}{\tilde{\pi}(\hat{v}^*|Y,Z,\hat{\theta})}$ is updated.

—*Step 5* (*lines: 8–9*) *Output optimal* θ. Steps 2–5 are repeated till the value of θ converges. Finally, θ is output.

Among these steps, Step 3 has the highest time cost. Because the solution is analytically intractable, numerical optimization techniques need to be applied (see Appendix).

Computational Complexity: In Algorithm 1, suppose that l_2 iterations are required to find the mode $\hat{v}^*$ and the Hessian matrix at the mode of $\tilde{\pi}(v^*|Y, Z, \theta)$, and the time cost of Step 6 is $O(l_2 * (n * m^2 + m^3))$. For Step 5, the Gaussian approximation of $\tilde{\pi}(v^*|Y, Z, \theta)$ takes $O(n * m)$. Overall, Steps 2–8, which generate the converged gradient and the Hessian matrix of $\pi(\theta|v^*)$, take $O(l_1 * l_2 * (n * m^2 + m^3) + l_1 * n * m)$. Finally, sampling the θ set and computing their corresponding weighted values take $O(K)$. The overall framework is designed based on Newton's method, whose convergence is generally rapid. The performance on problems in $\mathcal{R}^{10,000}$ is thus similar to that on problems in $\mathcal{R}^{10}$, and the required number of Newton steps (l_1) only increases modestly [Boyd and Vandenberghe 2004]. Step 6 applies IRLS to capture the mode of $\tilde{\pi}(v^*|Y, Z, \theta)$, and in practice five iterations ($l_2 = 5$) are sufficient. Assuming $m \gg K$, $m \gg l_1$, and $m \gg l_2$, the total computational complexity of parameter estimation is therefore $O(n * m^2)$.

5.3. Spatial Prediction via Laplace Approximation

Given a set of unsampled locations $\{s_1^0, \ldots, s_{N_{te}}^0\}$, we are interested in predicting the Y and Z attribute values at these locations, denoted as $Y^0 = (Y(s_1^0), \ldots, Y(s_{N_{te}}^0))'$ and $Z^0 = (Z(s_1^0), \ldots, Z(s_{N_{te}}^0))'$. The first step is to estimate the posterior distributions of the corresponding latent variables $\pi(\omega^0|Y, Z)$ and $\pi(\gamma^0|Y, Z)$, where $\omega^0 = (\omega(s_1^0), \ldots, \omega(s_{N_{te}}^0))'$ and $\gamma^0 = (\gamma(s_1^0), \ldots, \gamma(s_{N_{te}}^0))'$. The posterior distributions of Y^0 and Z^0 can then be obtained as

$$\pi(Y^0|Y, Z) = \int \pi(Y^0|\omega^0)\pi(\omega^0|Y, Z)d\omega^0, \tag{34}$$

$$\pi(Z^0|Y, Z) = \int \pi(Z^0|\gamma^0)\pi(\gamma^0|Y, Z)d\gamma^0. \tag{35}$$

We denote $v^0 = (\omega^{0\prime}, \gamma^{0\prime})'$. Given the approximated $\tilde{\pi}(v^*|Y, Z, \theta)$ and $\tilde{\pi}(\theta|Y, Z)$ obtained in Sections 4.1 and 4.2, the posterior distribution $\pi(v^0|Y, Z)$ can be estimated at each θ sample by

$$\tilde{\pi}(v^0|Y, Z, \theta) \approx \left. \frac{\pi(Y, Z|v^0, v^*)\pi(v^0, v^*|\theta)}{\tilde{\pi}(v^*|v^0, Y, Z, \theta)} \right|_{v^*=\hat{v}^*}. \tag{36}$$

Furthermore, it can be computed as

$$\pi(v^0|Y, Z) \approx \sum_{k=1}^{K} \tilde{\pi}(v^0|Y, Z, \theta_k)\tilde{\pi}(\theta_k|Y, Z)w_{\theta_k}. \tag{37}$$

Based on the above theoretical analysis, the main procedures involved in predicting multivariate non-Gaussian variables are described by Algorithm 2. As shown in Figure 3, Algorithm 2 introduces spatial prediction of multivariate non-Gaussian variables via LA as the following steps.

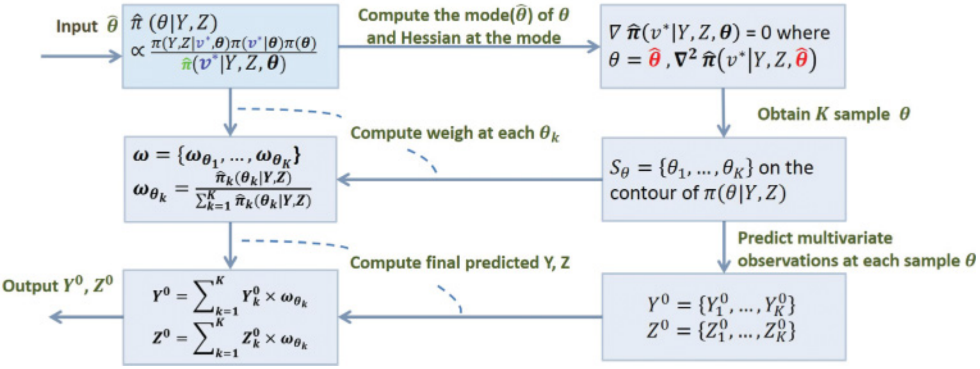

Input $\hat{\theta}$
$\hat{\pi}(\theta|Y,Z) \propto \frac{\pi(Y,Z|v^*,\theta)\pi(v^*|\theta)\pi(\theta)}{\hat{\pi}(v^*|Y,Z,\theta)}$
Compute the mode($\hat{\theta}$) of θ and Hessian at the mode
$\nabla \hat{\pi}(v^*|Y,Z,\theta) = 0$ where $\theta = \hat{\theta}$, $\nabla^2 \hat{\pi}(v^*|Y,Z,\hat{\theta})$
Obtain K sample θ
Compute weigh at each θ_k
$\omega = \{\omega_{\theta_1}, \dots, \omega_{\theta_K}\}$
$\omega_{\theta_k} = \frac{\hat{\pi}_k(\theta_k|Y,Z)}{\sum_{k=1}^K \hat{\pi}_k(\theta_k|Y,Z)}$
$S_\theta = \{\theta_1, \dots, \theta_K\}$ on the contour of $\pi(\theta|Y,Z)$
Predict multivariate observations at each sample θ
Compute final predicted Y, Z
Output Y^0, Z^0
$Y^0 = \sum_{k=1}^K Y_k^0 \times \omega_{\theta_k}$
$Z^0 = \sum_{k=1}^K Z_k^0 \times \omega_{\theta_k}$
$Y^0 = \{Y_1^0, \dots, Y_K^0\}$
$Z^0 = \{Z_1^0, \dots, Z_K^0\}$

Fig. 3. Spatial prediction via Laplace approximation.

ALGORITHM 2: Spatial Multivariate Non-Gaussian Prediction

Input: S, S^*, S^0, Y, Z, X, X^0, $\hat{\theta}$
Output: Y^0, Z^0

```
1: Explore the contour of $\hat{\pi}(\theta|Y, Z)$ based on its mode and Hessian matrix at the mode, obtain
   $K$ sample locations, $S_\theta = \{\theta_1, \dots, \theta_K\}$.
2: Compute and normalize $\{\hat{\pi}(\theta_1|Y, Z), \dots, \hat{\pi}(\theta_K|Y, Z)\}$ to obtain the set of weights
   $w = \{w_{\theta_1}, \dots, w_{\theta_K}\}$ as $w_{\theta_k} = \frac{\hat{\pi}_k(\theta_k|Y,Z)}{\sum_{k=1}^K \hat{\pi}_k(\theta_k|Y,Z)}$.
3: for $k = 1$ to $K$ do
4:     Construct $\mu_{\upsilon^*}$, $\Sigma_{\upsilon^*}$ with $\theta_k$ and $S^*$ (see Equation (23)).
5:     Calculate the transformation matrix $F(\phi)$ with $\theta_k$, $S^*$, $S$, $X$.
6:     Calculate the likelihood of Y for Gaussian variables (see Equation (24)) and the GLM
       likelihoodof Z for exponential ones (see Equation (25)).
7:     Calculate the mode, the Hessian matrix at the mode of $\hat{\pi}(\upsilon^*|Y, Z, \theta_k)$, and its Gaussian
       approximation (see Equation (30)).
8:     Predict $Y_k^0$, $Z_k^0$ for new locations $S^0$. (see Equations (34) and (35))
9: end for
10: Calculate the final $Y^0$, $Z^0$ values as $Y^0 = \sum_{k=1}^K Y_k^0 \times w_{\theta_k}$, $Z^0 = \sum_{k=1}^K Z_k^0 \times w_{\theta_k}$
```

—*Step 1* (*line: 1*) *Generate sample set of* S_θ. First, the contour of $\hat{\pi}(\theta|Y, Z)$ is explored based on its mode and Hessian matrix at the mode, and then its K sample values are generated.

—*Step 2* (*line: 2*) *Compute the weighted set of* ω_{θ_K}. The weighted values of θ samples are computed as $w_{\theta_k} = \frac{\hat{\pi}_k(\theta_k|Y,Z)}{\sum_{k=1}^K \hat{\pi}_k(\theta_k|Y,Z)}$.

—*Step 3* (*lines: 3–9*) *Predict* Y_0^k *and* Z_0^k *at each sample* θ_k. Each $\theta_k (k = 1, \dots, K)$ is utilized to perform a Gaussian approximation on the posterior distribution of v^* and then the mode of $\hat{\pi}(\upsilon^*|Y, Z, \theta)$ is calculated, which contributes to predict the multivariate observations Y_k^0 and Z_k^0.

—Step 4 (*line: 10*) *Obtain the final predicted* Y^0 *and* Z^0. Finally, the predicted Y and Z are calculated as $Y^0 = \sum_{k=1}^K Y_k^0 \times w_{\theta_k}$, $Z^0 = \sum_{k=1}^K Z_k^0 \times w_{\theta_k}$.

Computational complexity: Step 6 dominates the computational costs here, because it is analytically intractable. With the numerical optimization discussed in Sections 4.1 and 4.2, it takes $O(n*m)$ to operate a Gaussian approximation of $\hat{\pi}(\upsilon^*|Y, Z, \theta_k)$ for each sample θ_k. Computing the mode and Hessian matrix for $\hat{\pi}(\upsilon^*|Y, Z, \theta_k)$ costs $O(l_2*(m^3+n*m^2))$. Repeating Steps 1–7 for K sample θs therefore takes $O(K*(n*m+l_2*(m^3+n*m^2)))$.

Table II. Parameter Settings in Simulations

Variable	Setting description
Data type	Gaussian(Y)+Binomial(Z), Gaussian(Y)+Poisson(Z), Binomial(Y)+Poisson(Z)
N_{tr}, N_{te}	$N_{tr} = 1,000, N_{te} = 400, 500$. Training data were randomly generated at N_{tr} spatial locations $\{s_i\}_{i=1}^{N_{tr}}$ for the range [0,50]×[0,50] units. Test data were generated at N_{te} spatial locations $\{s_i\}_{i=1}^{N_{te}}$ over the same range.
β_y, β_z	The regression coefficient $\beta_y = [2, 2]', \beta_z = [2, 1]'$ in G+P; $\beta_y = [0.5, 0.5]', \beta_z = [0.1, 0.1]'$ in G+B; $\beta_y = [0.1, 0.1]', \beta_z = [2, 1]'$ in B+G.
$\sigma_y, \sigma_z, \sigma_{yz}$	$\sigma_y^2 = 4, \sigma_z^2 = 3.24, \sigma_{yz}^2 = 2.52$ in all types of simulations.
ϕ	$\phi = 25$ in all types of simulations.
τ	The measurement error variance, τ^2, was set to 1 in both G+B and G+P simulations.
Correlation model	An exponential spatial correlation function $C(h, \phi) = \sigma^2 exp(-\frac{h}{\phi})$ was used in all types of simulations.

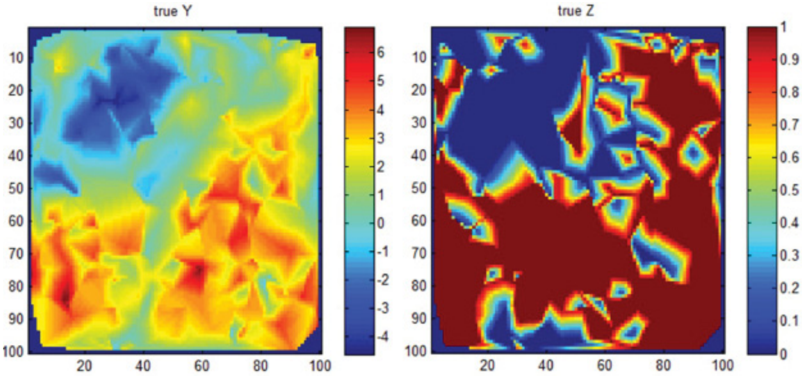


Fig. 4. Density maps of a typical G+B simulation.

The total computational complexity of the Spatial Multivariate Non-Gaussian Prediction algorithm is thus $O(n * m^2)$, assuming $m \gg K$ and $m \gg l_2$.

6. EXPERIMENTAL RESULTS AND ANALYSIS

This section evaluates the effectiveness and efficiency of our proposed framework based on experiments on simulations and four real-life datasets. We focus on three bivariate scenarios: (1) the response variables consist of one Gaussian and one binomial, G+B; (2) the response variables consist of one Gaussian and one Poisson, G+P; (3) the response variables consist of one binomial and one Poisson, B+P. All the experiments were conducted on a PC with Intel(R) Core(TM) I5-2400, CPU 3.1GHz, and 8.00GB memory. The development tool was MATLAB 2011.

6.1. Simulation Study

6.1.1. Simulation Settings.

Dataset: We utilized a similar simulation model in Chagneau et al. [2011]. The parameter settings used in our experiments are shown in Table II. We also evaluated different combinations of parameters and observed similar patterns. Figure 4 depicts density maps of the numerical(Y) and binary(Z) responses from a typical G+B simulation, revealing the complicated distributions involved and clearly illustrating why a higher processing ability is required for the predictive models.

Seven state-of-the-art competing methods: Based on our literature survey, there only exist two methods proposed for predicting multivariate non-Gaussian spatial data. One is the BME method by Wibrin et al. [2006] that only supports the mixture of one numerical and one categorical value. Another method is the MCMC designed by Chagneau et al. [2011] that is based on the Gibb sampler with M-H steps. We observed that the BME method is only restricted to bivariate data with one Gaussian and one Categorical, and MCMC method is flexible for a variety of mixture types. Hence, we implemented the MCMC method using the same framework (Gibbs sampler with M-H steps) and denoted this method as Spa-Multi-MCMC.

We also implemented an R toolbox function named "MCMCglmm" using the same MCMC framework, denoted as Multi-MCMC. Spa-Multi-MCMC and Multi-MCMC do not scale well to large datasets. Therefore, we designed two approximate versions of them, namely, Spa-Multi-MCMC-K and Multi-MCMC-K. CART (Classification and Regression Trees) [Breiman et al. 1984], MARS (Mulivariate Adaptive Regression Splines) [Friedman 1991], and Treenet (also known as MART, Multiple Additive Regression Trees) [Friedman 2000] are popular techniques for non-spatial predictive modeling. They have been implemented into the Salford Systems [2017]. We used them to make predictions for Y and Z separately. The model proposed in this article is identified as Spa-Multi-INLA.

Performance metric: We ran the experiments with 20 realizations of each parameter combination and then calculated the mean and standard deviation of every parameter combination in the multivariate process model. For each observation, we computed the Mean Absolute Error (MAE) for numerical and count observations, and the prediction error for binary ones based on their corresponding predicted and true values. To validate the new model's effectiveness and efficiency, we compared the results of estimations, predictions, and response times for the Spa-Multi-INLA and MCMC-based approaches, as well as CART, MARS, and Treenet. All parameters used in CART, MARS, and Treenet were tuned using cross-validation (10-folder), like the MinLeafSize, TreeSize, and NumofTrees, and so on. For both simulation and real datasets, the corresponding parameters were selected by the optimal points given a visual representation of the cross-validation results in CART, MARS, and Treenet models. Finally, we utilized Moran's I-statistic to capture the spatial dependency of the numerical observations.

6.1.2. Simulation Results.

Model parameter estimates: Tables III, IV, and V show the estimation results for the model parameters for datasets of size 1,000 for the G+B, G+P, and B+P simulations, respectively. "Spa-Multi-INLA(64)" refers to our approach with a knot size equal to 64. No results are shown for Multi-MCMC and Spa-Multi-MCMC because they became very slow when the data size exceeded 1,000. The iterations of Spa-Multi-MCMC-K and Multi-MCMC-K were set to 3,000 iterations, and K equal to 170 blocks (clusters). Also, there are no results for CART, MARS, and Treenet in these tables. This is because these models do not include these parameters. Instead, these ran on the Salford tool, which is a powerful well-developed and optimized tool, and it was not considered reasonable to directly compare their running times with those of the LA and MCMC-based approaches, both of which ran on Matlab. However, we did compare the prediction performances of Y and Z among all of these approaches and the results are shown in Figure 5. By comparing the estimated parameters with true values, we observed our method was able to accurately estimate most of the model parameters with only small deviations compared to the other two MCMC-based methods for all simulations. The true range parameter ϕ is 25, but both the LA- and MCMC-based approaches underestimated the range parameter at around 11. This indicates the difficulty of capturing the degree of spatial autocorrelation over the spatial distance.

Table III. Comparisons of the Parameter Estimation and Computational Cost in G+B.(Spa-Multi-MCMC and Multi-MCMC are Unable to Process Datasets with Data Sizes Greater than 1,000)

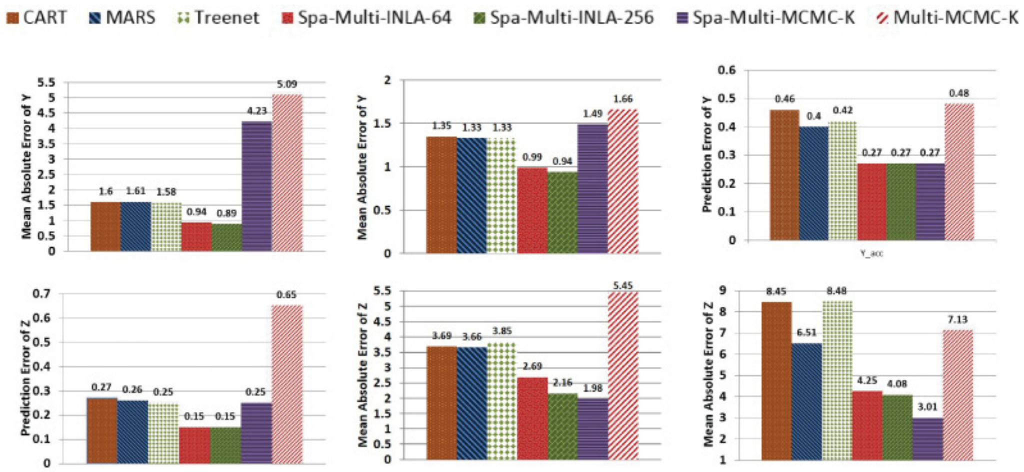
Approach	β_y	β_z	ϕ	σ_y^2	σ_z^2	σ_{yz}^2	τ^2	Time(m)
True values	$\begin{bmatrix} 0.50 \\ 0.50 \end{bmatrix}$	$\begin{bmatrix} 0.10 \\ 0.10 \end{bmatrix}$	25	4	3.24	2.52	1.00	—
Spa-Multi-INLA(64)	$\begin{bmatrix} 0.60(0.05) \\ 0.63(0.05) \end{bmatrix}$	$\begin{bmatrix} 0.18(0.07) \\ 0.15(0.07) \end{bmatrix}$	11.44 (2.64)	5.65 (1.86)	3.20 (1.02)	2.47 (0.12)	1.31 (0.06)	1.3
Spa-Multi-INLA(256)	$\begin{bmatrix} 0.60(0.03) \\ 0.63(0.03) \end{bmatrix}$	$\begin{bmatrix} 0.21(0.07) \\ 0.19(0.07) \end{bmatrix}$	10.64 (2.25)	4.73 (1.21)	3.02 (0.86)	2.40 (0.19)	1.12 (0.06)	1.5
Spa-Multi-MCMC-K	$\begin{bmatrix} 0.18(6.89) \\ -0.13(6.77) \end{bmatrix}$	$\begin{bmatrix} 0.22(0.47) \\ -0.09(0.43) \end{bmatrix}$	12 (1.05)	22.48 (24.39)	51.41 (58.19)	4578 (1054)	0.12 (0.15)	171.93
Multi-MCMC-K	$\begin{bmatrix} 0.29(7.11) \\ 0.06(7.29) \end{bmatrix}$	$\begin{bmatrix} -2.84(0.48) \\ 0.28(1.11) \end{bmatrix}$	−−	−−	−−	−−	0.13 (0.16)	234.27

Table IV. Comparisons of the Parameter Estimation and Computational Cost in G+P

Para Approach	β_y	β_z	ϕ	σ_y^2	σ_z^2	σ_{yz}^2	τ^2	Time(m)
True values	$\begin{bmatrix} 2.00 \\ 2.00 \end{bmatrix}$	$\begin{bmatrix} 2.00 \\ 1.00 \end{bmatrix}$	25	4	3.24	2.52	1.00	–
Spa-Multi-INLA(64)	$\begin{bmatrix} 2.51(0.22) \\ 2.01(0.02) \end{bmatrix}$	$\begin{bmatrix} 1.38(0.17) \\ 1.03(0.01) \end{bmatrix}$	13.87 (3.29)	4.90 (1.27)	3.02 (0.77)	3.31 (0.67)	1.28 (0.06)	1.83
Spa-Multi-INLA(256)	$\begin{bmatrix} 0.92(0.15) \\ 1.94(0.01) \end{bmatrix}$	$\begin{bmatrix} 0.44(0.12) \\ 0.94(0.01) \end{bmatrix}$	9.42 (2.29)	3.37 (0.94)	2.08 (0.48)	2.03 (0.35)	1.07 (0.05)	3.20
Spa-Multi-MCMC-K	$\begin{bmatrix} -0.07(7.01) \\ 0.13(6.94) \end{bmatrix}$	$\begin{bmatrix} 0.94(0.05) \\ 0.89(0.10) \end{bmatrix}$	2.97 (1.75)	36.31 (35.36)	2.82 (2.64)	1.84 (0.48)	0.26 (0.29)	72
Multi-MCMC-K	$\begin{bmatrix} 0.39(6.95) \\ -0.03(7.12) \end{bmatrix}$	$\begin{bmatrix} 1.39(0.05) \\ 0.99(0.07) \end{bmatrix}$	−−	−−	−−	−−	0.14 (0.15)	27

Table V. Comparisons of the Parameter Estimation and Computational Cost in B+P

Para Approach	β_y	β_z	ϕ	σ_y^2	σ_z^2	σ_{yz}^2	Time(m)
True values	$\begin{bmatrix} 0.10 \\ 0.10 \end{bmatrix}$	$\begin{bmatrix} 2.00 \\ 1.00 \end{bmatrix}$	25	4	3.24	2.52	–
Spa-Multi-INLA(64)	$\begin{bmatrix} 0.16(0.03) \\ 0.23(0.03) \end{bmatrix}$	$\begin{bmatrix} 2.09(0.01) \\ 1.04(0.01) \end{bmatrix}$	20.05 (3.58)	2.38 (0.87)	2.89 (0.54)	1.47 (0.03)	0.95
Spa-Multi-INLA(256)	$\begin{bmatrix} 0.12(0.03) \\ 0.35(0.03) \end{bmatrix}$	$\begin{bmatrix} 1.97(0.01) \\ 1.08(0.01) \end{bmatrix}$	20.05 (3.58)	3.56 (1.21)	3.87 (0.88)	2.60 (0.32)	9.17
Spa-Multi-MCMC-K	$\begin{bmatrix} 0.17(0.10) \\ 0.26(0.10) \end{bmatrix}$	$\begin{bmatrix} 1.91(0.03) \\ 1.01(0.03) \end{bmatrix}$	6.37 (0.61)	1.34 (0.06)	1.51 (0.09)	1.04 (0.06)	101
Multi-MCMC-K	$\begin{bmatrix} 0.15(0.08) \\ 0.20(0.08) \end{bmatrix}$	$\begin{bmatrix} 3.15(0.02) \\ 1.36(0.02) \end{bmatrix}$	−−	−−	−−	−−	95



(b) Gaussian+Binomial (c) Gaussian+Possion (d) Binomial+Possion

Fig. 5. Comparison of the performances for six approaches on simulation datasets.

In addition, we found it was more hard to estimate β_z than β_y in G+B and G+P, and β_y than β_z in B+P. This is reasonable, since it is usually more difficult to estimate the corresponding β for binary data than for count and numerical data. Compared with numerical data, count data are more difficult to model.

Prediction error: Figure 5 provides the prediction results of different approaches for the G+B, G+P, and B+P simulations. Applying Moran's I-statistic, we computed the spatial dependencies for Y numerical attributes in G+B and G+P, as 0.7006 and 0.7222, which indicates that the existing high spatial auto-correlation needs to be considered during the estimation and prediction processes. As we can see in Figure 5, for the case of G+B, Spa-Multi-INLA(256) has the lowest MAE(0.89) for Y and the lowest error (0.25) for Z. In contrast, CART, MARS, and Treenet have higher MAEs (1.60, 1.61, and 1.58) and higher errors (0.27, 0.26, and 0.25) since they are unable to capture the spatial dependencies. Spa-Multi-MCMC-K has the worse performance (MAE: 4.23, Error: 0.25) at the cost of large computational iterations. The Multi-MCMC-K approach failed to accurately execute spatial predictions since it was unable to learn the spatial dependency and operated without sufficient iterations. Spa-Multi-MCMC and Multi-MCMC cannot process mixed type datasets whose data sizes are greater than 1,000 because in MCMC-based approaches the un-marginalized models used to fit the Binomial+Poisson outcome data required more MCMC iterations.

Figure 5(b) and (c) also provides the prediction comparisons of Spa-Multi-INLA models with 64 and 256 knots against other approaches for the G+P and B+P simulations. These exhibit the same estimation patterns as those in G+B. By comparing different knot intensities, we see the predictive process with 64 knots has quite a close performance to that with 256 that indicates that the parameter effects can be accurately estimated with the proper knot selections.

Evaluation between approximating and true posterior distribution: Spa-Multi-INLA predictive model integrates *reduced-rank methodology* with *iterative LA* to achieve accurate and much faster inference. Iterative LA (including the Gaussian approximation involved) is utilized to solve analytically intractable issues when modeling non-Gaussian response variables and capturing correlations among them, while the reduced-rank technology helps improve the scalability and availability when large

Table VI. Comparisons of Approximated and True Posterior Distribution in G+B

Para Approach	β_y	β_z	ϕ	σ_y^2	σ_z^2	σ_{yz}^2	τ^2	Time(s)	MAE_y	Acc_z
True values	$\begin{bmatrix} 0.60 \\ 0.60 \end{bmatrix}$	$\begin{bmatrix} 0.15 \\ 0.15 \end{bmatrix}$	25	5.76	3.24	3.02	1.00	–	–	–
Spa-Multi-INLA(64)	$\begin{bmatrix} 0.40(0.06) \\ 0.50(0.06) \end{bmatrix}$	$\begin{bmatrix} -0.49(0.07) \\ -0.07(0.01) \end{bmatrix}$	8.38 (2.42)	7.65 (3.87)	2.60 (0.79)	1.56 (1.10)	1.49 (0.06)	38.62	1.2495	0.7525
Spa-Multi-INLA-Full	$\begin{bmatrix} -2.13(0.15) \\ -0.70(0.14) \end{bmatrix}$	$\begin{bmatrix} -1.41(0.14) \\ -4.38(0.13) \end{bmatrix}$	19.59 (4.20)	6.17 (1.65)	4.81 (1.67)	1.91 (0.55)	0.99 (0.07)	7082.41	1.1765	0.77
Spa-Multi-MCMC-Knots	$\begin{bmatrix} 0.04(6.98) \\ -0.08(7.02) \end{bmatrix}$	$\begin{bmatrix} -0.44(0.13) \\ -0.19(0.13) \end{bmatrix}$	18.73 (3.04)	9.71 (4.69)	0.01 (0.29)	0.90 (0.12)	0.11 (0.14)	321.39	1.584	0.6375

amounts of data are being collected. Table VI evaluates the performances for parameter estimation and prediction error of the Spa-Multi-INLA-Full (LA-based spatial multivariate predictive model with full dataset), Spa-Multi-INLA(64) (LA-based spatial multivariate predictive model with 64 knots) and Spa-Mulit-MCMC-Knots (MCMC-based spatial multivariate predictive model with 64 knots) models for one Gaussion+Binary simulation.

The best prediction results were generated by Spa-Multi-INLA-Full, which achieved performances around 5% and 25% better than those of Spa-Multi-INLA(64) and Spa-Multi-MCMC-Knots(64), respectively. When estimating parameters, Spa-Multi-INLA-Full was able to capture spatial random effects across different locations more accurately. This can be verified by comparing the estimated values of σ_y, σ_z, and σ_{yz}, which exhibited improvements of 25% and more better than the knot-based approaches. Meanwhile, the Spa-Multi-INLA(64) was better modeling relationships between dependent and explanatory variables; its estimated values of β_y and β_z were around 20%–50% better than those generated by Spa-Multi-INLA-Full and Spa-Multi-MCMC-Knots. Theoretically, Spa-Multi-INLA-Full should have the best performance for spatial predictive inference, because it models the predictive process with a far larger training dataset (1,000), while the knot-based approaches (Spa-Multi-INLA(64) and Spa-Multi-MCMC-knots) use only a relatively small number of points (64 of 1,000). However, in practice this is not always the case. Since Spa-Multi-INLA is a complex model where many parameters are involved, the existence of outliers or noises requires the inference process to incorporate these into the model, which on occasion can mean that the spatial statistical model describes the random error as an underlying relationship. This overfitting issue can adversely affect the estimation accuracy of several parameters.

Spa-Multi-INLA(64) has a similar parameter estimation and predictive capability to that of the full predictive process. But there was a clear reduction in the computation cost when using the reduced-rank technique with the time required for the prediction process dropping from 7082.41 s (full process) to 38.62 s (knot-based process). If there is an appropriate selection of knots that covers most of the domain knowledge, Gaussian and LA techniques can clearly provide accurate parameter estimation much faster. This approach has the added advantage of avoiding mistaking random errors as underlying relationships, the well-known overfitting issue described above. As shown in Table VI, when estimating β_y and β_z, Spa-Multi-INLA(64) better captures the relationship between the observed responses (Y and Z) and spatially referenced predictors (X).

In order to measure the closeness of the approximated posterior distribution achieved by the new approach proposed here to the true posterior distribution, we calculated the root mean squared error (RMSE) values between the MAP estimation of model

Table VII. Comparison of RMSE on Estimated Parameters of Spa-Multi-INLA Among Different Data Sizes in G+B

Para. / Size	β_y	β_z	ϕ	σ_y^2	σ_z^2	σ_{yz}^2	τ^2	Avg
200	$\left[0.27; 0.25\right]$	$\left[0.54; 0.57\right]$	17.6	11.82	1.57	1.31	0.26	3.8
400	$\left[0.21; 0.18\right]$	$\left[0.35; 0.35\right]$	17.66	9.73	1.29	1.25	0.36	3.48
600	$\left[0.14; 0.14\right]$	$\left[0.29; 0.3\right]$	16.98	6.24	1.25	1.43	0.39	3.02
800	$\left[0.12; 0.12\right]$	$\left[0.22; 0.23\right]$	16.83	4.4	1.27	1.54	0.41	2.79
1,000	$\left[0.11; 0.1\right]$	$\left[0.22; 0.2\right]$	15.55	4.47	1.26	1.59	0.43	2.66

Table VIII. Settings in the Four Real Datasets

Dataset	Size	N_{tr}	N_{te}	Y	Z	Spatial dependence on Y
BEF spBayes [2012]	437	337	100	BE basal area	EH basal area	0.1672
Lake Varin et al. [2005]	371	271	100	Trout abundance	Lake acidity	0.0072
MLST Dubin [1992]	211	150	61	House price	If located in county	0.1753
House Pace and Barry [1997]	20,640	2,000 5,000	200 500	House price	House age	0.2529

parameters based on our approximated posterior and the true posterior. Because the true posterior is analytically intractable and it is difficult to evaluate the closeness to our approximated posterior at different sample sizes, we approximated the MAP estimation of the true posterior distribution by adjusting the parameters used to generate the simulations. The comparison results shown in Table VII indicate that as the sample size increases, the MAP estimation of model parameters obtained using our new approach becomes closer to the true model parameters.

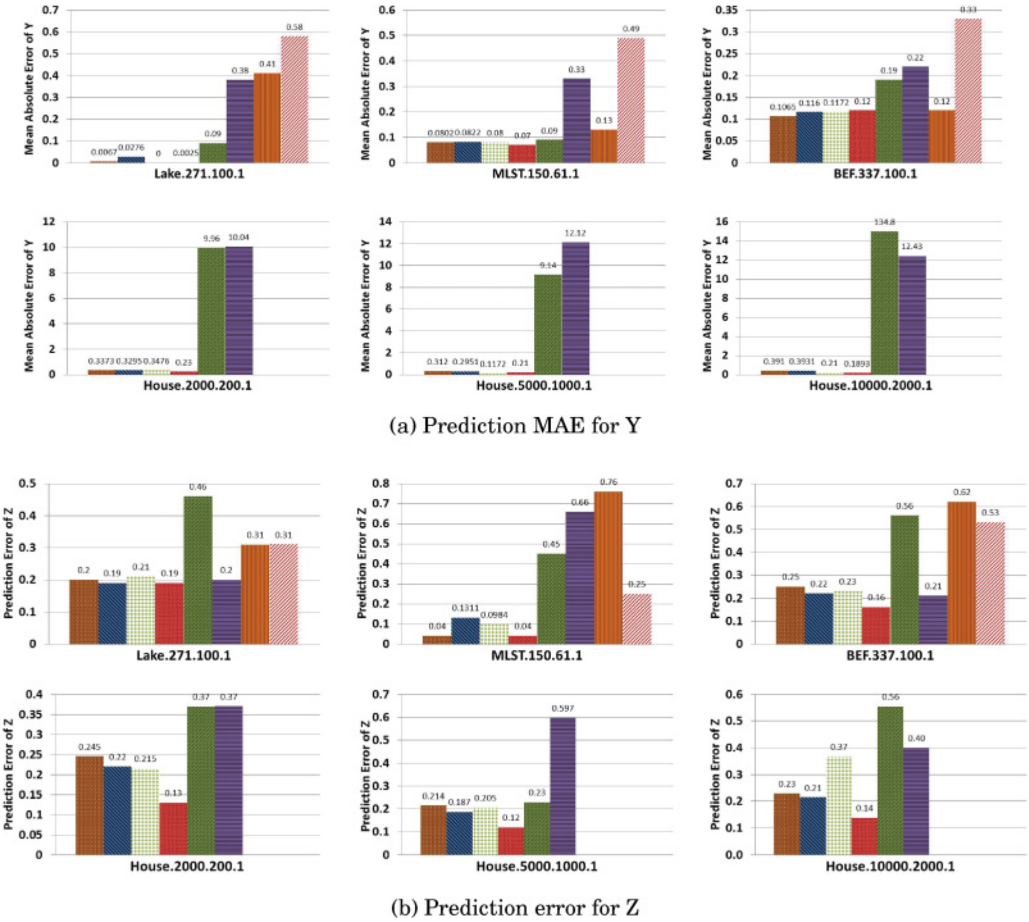
Computational Cost: The last column in Tables III, IV, and V shows the computing times required to deliver the estimation and prediction results for each of the simulations. For the MCMC-based approaches, the main evaluation cost is the matrix inversion at $O((2*1000)^3)$. For the Spa-Multi-INLA model, the main cost is $O(2*n*(2*m)^2)(m = 64 \text{ or } 256)$, which is the cost of building the required inverse and determinant of the size $(2m+2p)\times(2m+2p)$ matrix Q as shown in Equation (30), which assumes $m \gg p$. As shown in Tables III, IV and V, there is a clear reduction in the computational cost when using the Spa-Multi-INLA approach and the predictive process with 64 knots has a similar prediction capability but lower computational burden compared to that 256 knots. Integrating the LA into the spatial multivariate predictive model clearly helps achieve sufficiently five results in a moderate time.

6.2. Real-Life Datasets

We validated our approach using four real datasets, which are all G+B datasets. Table VIII summarizes the main information used in our experiment. The spatial dependencies were computed using Moran's I-statistic function.

6.2.1. Experimental Results.

Prediction error: Figure 6 summarizes the comparisons among Spa-Multi-INLA(64), the four MCMC-based approaches, CART, MARS, and Treenet. The data name "Lake.271.100.1" indicates that it is the first realization generated from the original



(a) Prediction MAE for Y

(b) Prediction error for Z

Fig. 6. Comparison of the performances for eight approaches on real-life datasets.

Lake data, with 271 training data and 100 test data points. By learning their spatial dependencies, we determined that most of the real datasets have lower spatial autocorrelations, which suggests that non-spatial attributes will contribute substantially to the prediction of the outcome variables. For the predicted Y (numerical observations), the MAEs were computed to demonstrate the prediction performance. Neither Multi-MCMC nor Spa-Multi-MCMC could process the *House* data because of the large data sizes (2,000 and 5,000 points) involved. The MAE values from Spa-Multi-MCMC-K and Multi-MCMC-K were also much higher (around 10 times) than those (0.21–0.33) of the others. When plotting the performance comparisons among Spa-Multi-INLA, CART, MARS, and Treenet, we did not include the MCMC-based plots for *House* as all the MCMC-based approaches generated poor results due to the large datasets involved (2,000 and 5,000), which incurred excessive computation times of around 2 days. In our experiments, the iteration values for all the datasets were set to 3,000, although this still cost around 1–3.5 hours with Multi-MCMC-K, and 2.5–4.5 hours with Spa-Multi-MCMC-K for the *House* datasets. As shown in Figure 6(a), Spa-Multi-INLA achieved an average 10% improvement over CART, MARA, and Treenet, 40–50% over Spa-Multi-MCMC-K and Spa-Multi-MCMC, and 60–70% over Multi-MCMC-K and Multi-MCMC.

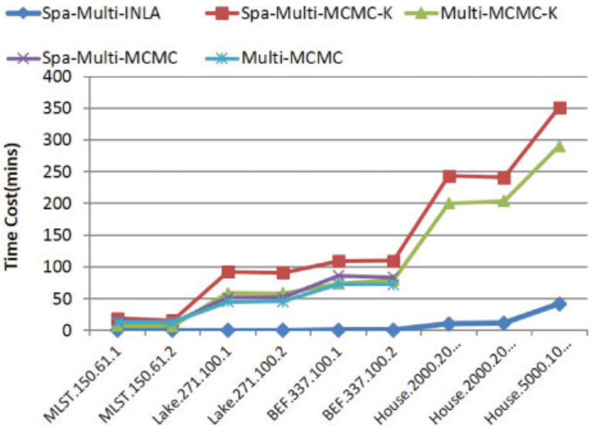


Fig. 7. Total response time by varying data size.

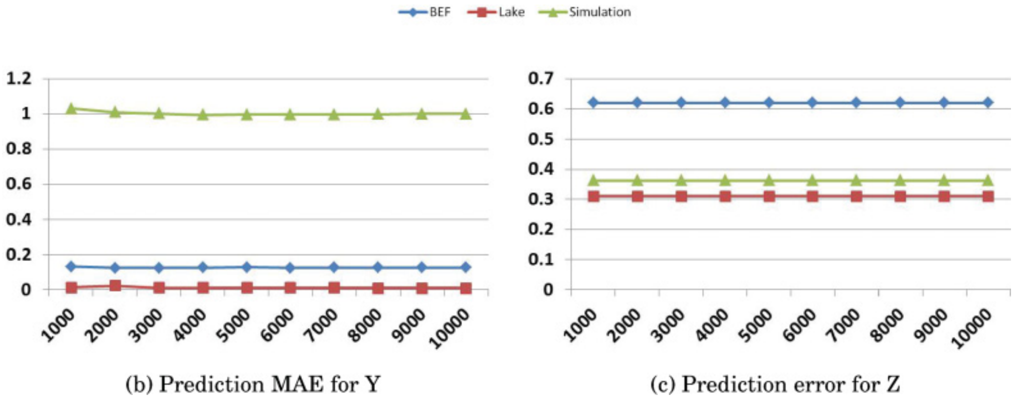
For the predicted Z(binary observations), the accuracies were computed and are shown in Figure 6(b). Spa-Multi-INLA achieved average improvements of 10% over CART, MARS, and Treenet, 40–50% over Spa-Multi-MCMC-K and Spa-Multi-MCMC, and 60–70% over Multi-MCMC-K and Multi-MCMC.

It is worth noting that CART, MARS and Treenet all generated impressive prediction result, and their overall performance was much better than those of the MCMC-based approaches. This is because the degrees of spatial auto-correlations of the four real datasets are not obvious. The predictions of outcome variables are mainly controlled by the non-spatial predictors, and these have less relationship with the spatial distances among the objects. For *Lake*, which has the lowest spatial dependency (0.0072), their performance is close to that of the LA-based approach. For *MLST* and *BEF*, the spatial dependencies increase a little but are still lower and the performances of CART, MARS, and Treenet are all a little worse than that of the LA-based approach. For *House*, the degree of spatial auto-correlation is more obvious and clearly demonstrates the effectiveness of the LA-based approach, since it takes spatial dependency into consideration during the predictive process. The MCMC-based approaches have similar estimation patterns to these in simulations and cannot perform well.

Computational cost: The real datasets used in these experiments showcase the speed and associated scalability achieved by the approaches that we evaluated. Figure 7 compares the runtime performance of these algorithms in the various datasets for varying numbers of training and testing points. For example, for *BEF.337.100.2*, Spa-Multi-INLA completed its run in around 0.63 min, while Spa-Multi-MCMC-K took around 15.19 min and Multi-MCMC-K, Spa-Multi-MCMC, and Multi-MCMC took from 11.43 to 32.64 min. In particular, for *House.2000.200.2*, our methods finished running in 10 min, while Spa-Multi-MCMC-K and Multi-MCMC-K took several hours and the other two approaches were not able to execute this dataset since it exceeds 2,000 points. As noted above, the main cost incurred in the Spa-Multi-INLA model is that needed to build the required inverse and determinant of the size $(2m + 2p) \times (2m + 2p)$ matrix. When $m \ll n$, the main cost $O(nm^2)$ can be approximated as $O(n)$. As shown by the bold red curve in Figure 7, the Spa-Multi-INLA model clearly reduced the computational burden, which makes the predictive process nearly linear in complexity.

6.3. Analysis of the Results

The above experimental results demonstrate that Spa-Multi-INLA is both effective and efficient in estimating the parameters and predicting different types of variables. Its



(b) Prediction MAE for Y

(c) Prediction error for Z

Fig. 8. Prediction Performance of Spa-Multi-MCMC-K by varying iterations.

identification quality is clearly superior to that of existing techniques, achieving around 10–30% improvement over CART, MARs, and Treetnet, and 40–50% over MCMC-based approaches. The experimental results verified three observations. (1) *Appropriate Knot Selection*. If there is an appropriate selection of knots that covers most of the domain of interest, the cost of the predictive process will be significantly reduced to a linear order. For the Spa-Multi-INLA model, the main cost is to compute the mode and Hessian matrix of the posterior distribution of latent variables, which costs $O(n{*}m^2)$. When $m \ll n$, the predictive process becomes linear in complexity. (2) *Efficient Approximation Process.* When combined with numerical routines, Gaussian and LA techniques can provide much faster and more accurate parameter estimation than MCMC-based algorithms for spatial multivariate non-Gaussian prediction. (3) *Effectiveness for Large Spatial Data Analysis*. When processing more complicated datasets, such as the simulation data shown in Figure 4, MCMC-based approaches need a very high number of iterations to achieve acceptable results, thus incurring unacceptably high computational costs, and CART MARS and Treenet cannot handle data with high spatial dependencies, but the new approach proposed here can complete the prediction computation in moderate times with no loss of accuracy.

Finally, there was an interesting result for the MCMC-based methods. MCMC approaches are sampling-based, and require sufficient iterations provided. They depend on the sample selections of the latent variables. In some cases, they performed well, as in the B+P simulation, where both Spa-Multi-MCMC-K and Multi-MCMC-K did approximately estimate β_y and β_z. However, they sometimes failed to make good estimations, as in the G+P and G+B simulations, where all of the estimated parameters deviated substantially from their true values. Executing MCMC approaches with appropriate iterations provided comparable results but were very time iterations. To make sure MCMC-based approaches performed with sufficient iterations, we evaluated the prediction errors of MCMC-based methods by varying iteration numbers in both simulation and real-life datasets. Figure 8 depicts such analysis on Spa-Multi-MCMC-K approach. And we can determined 2,000–3,000 is the optimal range of the iterations.

7. CONCLUSIONS

This article proposes a novel framework for estimating multivariate predictive process models that is designed to take into account mixed type response variables. It integrates multivariate predictive process models with approximate Bayesian inference using iterative LA. The predictive model consists of a representative selection of knot locations that projects the spatial process to a lower dimensional subspace. The approximation process provides more accurate and much faster inference for spatial multivariate predictive models. Experimental results for synthetic and real datasets conclusively demonstrated that our proposed non-Gaussian prediction model is capable of achieving a much higher processing capability in terms of prediction accuracy and computation time.

REFERENCES

T. C. Bailey and W. J. Krzanowski. 2000. Extensions to spatial factor methods with an illustration in geochemistry. *Mathematical Geology* 32, 6 (2000), 657–682. DOI:http://dx.doi.org/10.1023/A:1007589505425

S. Bandyopadhyay. 2005. Simulated annealing using a reversible jump Markov Chain Monte Carlo algorithm for fuzzy clustering. *IEEE Transactions on Knowledge and Data Engineering* 17, 4 (2005), 479–490.

Sudipto Banerjee, Alan E. Gelfand, Andrew O. Finley, and Huiyan Sang. 2008. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70, 4 (2008), 825–848. DOI:http://dx.doi.org/10.1111/j.1467-9868.2008.00663.x

Mario Boley and Henrik Grosskreutz. 2008. A randomized approach for approximating the number of frequent sets. In *Proceedings of the 2008 8th IEEE International Conference on Data Mining*. 43–52. DOI:http://dx.doi.org/10.1109/ICDM.2008.85

Edwin V. Bonilla, Kian M. Chai, and Christopher Williams. 2008. Multi-task Gaussian process prediction. *Advances in Neural Information Processing Systems*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis (Eds.). Vol. 20. Curran Associates, Inc., 153–160. http://papers.nips.cc/paper/3189-multi-task-gaussian-process-prediction.pdf.

Stephen Boyd and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, New York, NY.

Leo Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Wadsworth.

Catherine A. Calder. 2007. Dynamic factor process convolution models for multivariate space time data with application to air quality assessment. *Environmental and Ecological Statistics* 14, 3 (2007), 229–247.

Pierrette Chagneau, Frederic Mortier, Nicolas Picard, and Jean-NoÃńl Bacro. 2010. Hierarchical Bayesian model for Gaussian, Poisson and ordinal random fields. In *Quantitative Geology and Geostatistics*, P. M. Atkinson and C. D. Lloyd (Eds.), Vol. 16. Springer, The Netherlands, 333–344. DOI:http://dx.doi.org/10.1007/978-90-481-2322-3_29

Pierrette Chagneau, Fradaric Mortier, Nicolas Picard, and Jean-Nol Bacro. 2011. A hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields. *Biometrics* 67, 1 (2011), 97–105. http://dx.doi.org/10.1111/j.1541-0420.2010.01415.x

Jorge Chica-Olmo. 2007. Prediction of housing location price by a multivariate spatial method: Cokriging. *Journal of Real Estate Research* 29, 1 (2007), 95–114.

Jungsoon Choi, Brian J. Reich, Montserrat Fuentes, and Jerry M. Davis. 2009. Multivariate spatial-temporal modeling and prediction of speciated fine particles 3, 2 (2009), 407–418. DOI:http://dx.doi.org/10.1080/15598608.2009.10411933

William F. Christensen and Yasuo Amemiya. 2001. Generalized shifted-factor analysis method for multivariate geo-referenced data. *Mathematical Geology* 33, 7 (2001), 801–824. DOI:http://dx.doi.org/10.1023/A:1010998730645

William F. Christensen and Yasuo Amemiya. 2002. Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association* 97 (2002), 302–317.

Noel Cressie. 1991. *Statistics for Spatial Data*. Wiley-Interscience.

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

Robin A. Dubin. 1992. Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics* 22, 3 (September 1992), 433–452.

Andrew O. Finley, Huiyan Sang, Sudipto Banerjee, and Alan E. Gelfand. 2009. Improving the performance of predictive process modeling for large datasets. *Computational Statistics & Data Analysis* 53, 8 (2009), 2873–2884.

Jerome H. Friedman. 1991. Rejoinder: Multivariate adaptive regression splines. *The Annals of Statistics* 19, 1 (1991), 123–141.

Jerome H. Friedman. 2000. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29 (2000), 1189–1232.

A. E. Gelfand and S Banerjee. 2010. Multivariate spatial process models. In *Handbook of Spatial Statistics*. Chapman & Hall/CRC Press, 495–515.

A. E. Gelfand, A. M. Schmidt, S. Banerjee, and C. F. Sirmans. 2004. Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). *Test* 13, 2 (2004), 1–50.

P. Goovaerts. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, USA.

Michel Grzebyk and Hans Wackernagel. 1994. Multivariateanalysis and spatial/temporal scales: Real and complex models. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.3893.

Mikhail Kanevski. 2012. Multitask learning of environmental spatial data. In *6th International Congress on Environmental Modelling and Software Society (iEMSs)*. Leipzig, Germany.

B. M. Golam Kibria, Li Sun, James V. Zidek, and Nhu D. Le. 2002. Bayesian spatial prediction of random space-time fields with application to mapping PM2.5 exposure. *Journal of the American Statistical Association* 97 (2002), 112–124.

Guichong Li, Nathalie Japkowicz, Trevor J. Stocki, and R. Kurt Ungar. 2008. Border sampling through coupling Markov Chain Monte Carlo. (2008), 393–402. DOI:http://dx.doi.org/10.1109/ICDM.2008.52

Wei Liu, Yu Zheng, Sanjay Chawla, Jing Yuan, and Xie Xing. 2011. Discovering spatio-temporal causal interactions in traffic data streams. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*. ACM, New York, NY, 1010–1018. DOI:http://dx.doi.org/10.1145/2020408.2020571

A. B. McBratney, I. O. A. Odeh, T. F. A. Bishop, M. S. Dunbar, and T. M. Shatar. 2005. An overview of pedometric techniques for use in soil survey. *Geoderma* 97, 3–4 (2005), 293–327.

Marco Minozzo and Clarissa Ferrari. 2013. Multivariate geostatistical mapping of radioactive contamination in the maddalena archipelago (Sardinia, italy): Spatial special issue. *AStA Advances in Statistical Analysis* 97, 2 (2013), 195–213. DOI:http://dx.doi.org/10.1007/s10182-012-0201-x

Marco Minozzo and Daniela Fruttini. 2004. Loglinear spatial factor analysis: An application to diabetes mellitus complications. *Environmetrics* 15, 5 (2004), 423–434. DOI:http://dx.doi.org/10.1002/env.675

Marco Minozzo and Laura Ferracuti. 2012. On the existence of some skew-normal stationary processes. *Chilean Journal of Statistics* 3 (2012), 157–170.

Seyed H. Mohammadi, Vandana Pursnani Janeja, and Aryya Gangopadhyay. 2009. Discretized spatio-temporal scan window. In *Proceedings of the SIAM International Conference on Data Mining, SDM 2009*. 1195–1206. DOI:http://dx.doi.org/proceedings/datamining/2009/dm09_109_mohammadis.pdf

Daniel B. Neill and Andrew W. Moore. 2004. Rapid detection of significant spatial clusters. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 256–265. DOI:http://dx.doi.org/10.1145/1014052.1014082

Orlando Ohashi and Luís Torgo. 2012. Spatial interpolation using multiple regression. In *Proceedings of ICDM*. 1044–1049.

Victor De Oliveira. 2000. Bayesian prediction of clipped Gaussian random fields. *Computational Statistics and Data Analysis* 34, 3 (2000), 299–314.

Kelley Pace and Ronald Barry. 1997. Sparse spatial autoregressions. *Statistics & Probability Letters* 33, 3 (1997), 291–297.

Gregory Piatetsky-Shapiro, Chabane Djeraba, Lise Getoor, Robert Grossman, Ronen Feldman, and Mohammed Zaki. 2006. What are the grand challenges for data mining? KDD-2006 panel report. *SIGKDD Explorations Newsletter* 8, 2 (December 2006), 70–77. DOI:http://dx.doi.org/10.1145/1233321.1233330

Brian J. Reich and Montserrat Fuentes. 2007. A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *The Annals of Applied Statistics* 1, 1 (2007), 249–264.

Qian Ren and Sudipto Banerjee. 2013. Hierarchical factor models for large spatially misaligned data: A low-rank predictive process approach. *Biometrics* 69, 1 (2013), 19–30.

Greg Ridgeway and David Madigan. 2002. Bayesian analysis of massive datasets via particle filters. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02)*. ACM, New York, NY, USA, 5–13.

Håvard Rue and Leonhard Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*, Monographs on Statistics and Applied Probability, Vol. 104. Chapman & Hall, London.

Håvard Rue, Sara Martino, and Nicolas Chopin. 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 (2009), 319–392. DOI:http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x

A. Schmidt and M. Rodriguez. 2011. Modelling multivariate counts varying continuously in space. *Bayesian Statistic* (2011).

Salford Systems. 2017. Homepage. Retrieved from http://www.salford-systems.com/.

spBayes. 2012. spBayes: Univariate and multivariate spatial modeling. Retrieved from http://cran.r-project.org/web/packages/spBayes/.

Luis Torgo and Orlando Ohashi. 2011. 2D-interval predictions for time series. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*. ACM, New York, NY, 787–794. DOI:http://dx.doi.org/10.1145/2020408.2020546

Cristiano Varin, Gudmund Host, and Oivind Skare. 2005. Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics & Data Analysis* 49, 4 (June 2005), 1173–1191.

H. Wackernagel. 2003. *Multivariate Geostatistics: An Introduction with Applications* (2nd ed.). Springer-Verlag. 381.

F. Wang and M. M. Wall. 2003. Generalized common spatial factor model. *Biostatistics (Oxford, England)* 4, 4 (October 2003), 569–582.

R. Webster and M. A. Oliver. 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press, 316.

M. A. Wibrin, P. Bogaert, and D. Fasbender. 2006. Combining categorical and continuous spatial information within the Bayesian Maximum Entropy paradigm. *Stochastic Environmental Research and Risk Assessment* 20, 6 (2006), 423–433.

Robert L. Wolpert and Katja Ickstadt. 1997. Poisson/gamma random field models for spatial statistics. *Biometrika* 85, 2 (1997), 251–267.

Elizabeth Wu, Wei Liu, and Sanjay Chawla. 2008. Spatio-temporal outlier detection in precipitation data. In *Proceedings of the KDD Workshop on Knowledge Discovery from Sensor Data.* 115–133.

Mingxi Wu, Chris Jermaine, Sanjay Ranka, Xiuyao Song, and John Gums. 2010. A model-agnostic framework for fast spatial anomaly detection. *TKDD* 4, 4 (2010), 20.

Mingxi Wu, Xiuyao Song, Chris Jermaine, Sanjay Ranka, and John Gums. 2009. A LRT framework for fast spatial anomaly detection. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Paris, France, 887–896. DOI:http://doi.acm.org/10.1145/1557019.1557116

Harry Zhang and Shengli Sheng. 2004. Learning weighted naive Bayes with accurate ranking. In *Proceedings of the Fourth IEEE International Conference on Data Mining*. 567–570.

J. Zhu, J. C. Eickhoff, and P. Yan. 2005. Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* 61, 3 (2005), 674–683. DOI:http://dx.doi.org/10.1111/j.1541-0420.2005.00343.x

Wei Zhuo, Prabhat, Chris Paciorek, Cari Kaufman, and Wes Bethel. 2011. Parallel kriging analysis for large spatial datasets. In *Proceedings of ICDM Workshops'11*. 38–44.

Received November 2014; revised May 2015; accepted November 2016