

Likelihood Inference in A Multivariate Spatial GLMM with Skew Gaussian Random Effects Using the Slice-SAB Algorithm

Firoozeh Rivaz*

*Department of Statistics, Shahid Beheshti University,
Tehran, Iran
f_rivaz@sbu.ac.ir*

Majid Jafari Khaledi

*Department of Statistics, Tarbiat Modares University,
Tehran, Iran
jafari-m@modares.ac.ir*

Received 1 May 2015

Accepted 6 June 2016

This paper introduces a multivariate skew Gaussian process and uses it to extend the family of multivariate spatial generalized linear mixed models to include skew Gaussian random effects. In this setting, the parameter estimation encounters problems because the likelihood function involves high dimensional integrations which are computationally intensive. For estimating parameters of the complicated model structure, this article proposes an algorithm which is a combination of boosting with a variant of stochastic approximation. This algorithm which known as stochastic approximation boosting (SAB) algorithm, uses the Markov chain Monte Carlo method based on slice sampling to obtain simulations from full conditional distribution of random effects. A simulation study is conducted to assess the performance of our method. The proposed methodology is further illustrated through an application to a data set of soil pollution in a province of Iran.

Keywords: Multivariate spatial data; Random effects; Maximum likelihood estimates; Boosting; Stochastic approximation; Skew normal.

2000 Mathematics Subject Classification: 62H11, ,62F15

1. Introduction

Spatial generalized linear mixed models (SGLMMs) as a special class of spatial random effects models have been a popular method for analyzing the spatial variables. In fact, they are described

*Corresponding author.

in terms of unobservable spatial random effect which is linked to the observed variable through a link function. In the univariate response case, Bayesian analysis for SGLMMs includes [7], [4], [6] and [8]. Apart from the Bayesian method, [17] determined the maximum likelihood estimations of model parameters using a Monte Carlo EM (MCEM) algorithm. [16] also proposed a pair-wise likelihood approach which is useful to deal with estimation and inference in SGLMMs with high-dimensional data. Considering an integrated likelihood function, [5] provided the Monte Carlo approximation to the likelihood.

In all aforementioned works, normality of the underlying random effects is a basic assumption. Moreover, spatial data collection increasingly turns to vector valued measurements at spatial locations. In this paper, we focus on the situation in which multivariate responses are observed simultaneously on each locations. Thus in particular, a multivariate spatial generalized linear mixed model is used to analyze the spatial variables. In order to relax unrealistic normality assumption, this article also considers a multivariate skew Gaussian random field for random effects. [11] defined a skew Gaussian random field based on the multivariate skew normal distribution ([2] and [3]). [9] remarked identifiability problems associated with this model. In a different way, [18] recently introduced a class of stationary processes that have skewed marginal distributions. A key feature of this model is that it is formulated as a linear combination of a Gaussian process and absolute of further Gaussian process, which is useful for random number generation and for theoretical derivations. By introducing the multivariate case of this skew Gaussian process and modeling random effects based on it, we investigate how to obtain maximum likelihood estimations of parameters. Within the classical likelihood framework, fitting these complex models via the Monte Carlo EM algorithm can be computationally challenging since this method involves high dimensional optimizations. In addition, because of difficulties in determining integrated likelihood function, the Monte Carlo approximation of the likelihood as in [5] is so hard to provide. Moreover, this method requires a proper reference point. Updating the reference point (as proposed by [5]) is only an ad hoc solution and convergence to the true MLE cannot be proven theoretically.

To overcome complicated model-structure, we explore another algorithmic avenue. In fact, a stochastic approximation boosting (SAB) algorithm of [?] is proposed for the maximum likelihood inference in the multivariate SGLMMs. This algorithm combines boosting which is a very flexible and powerful tool, with stochastic approximation method (was originally introduced by [14] and extended by [10]) which is similar to the SAEM method ([12]). Generally, the Robbins-Monro procedure is a stochastic root-finding method and the component-wise boosting approach makes the technique straightforward to implement. With regard to problems remain with convergence and mixing properties of Metropolis-Hastings algorithm, as well as there is no routine to choose an appropriate candidate distribution for this algorithm, we will introduce auxiliary variables and employ the slice sampling method ([13]) to simulate from full conditional distribution of random effects. Actually, this is a technique of generating from an arbitrary variables by introducing an auxiliary variable and sampling from two or more uniform distributions.

The article is organized as follows. Section 2 introduces the multivariate SGLMM with skew Gaussian random effects and outlines the details of method for parameter estimation. Then, it gives the proposed method for spatial prediction. Section 3 explores a simulation experiment to assess the effectiveness of the proposed method. In Section 4, we apply this method in order to analyze a data set related to soil pollution. The paper ends with a brief concluding remark.

2. Statistical Model

The considered multivariate SGLMM is set up by the following components. Suppose we observe p random variables in the set of locations $\mathbf{x} = \{x_1, \dots, x_n\}$. Let $\mathbf{y}_j = (y_{1j}, \dots, y_{nj})'$, $j = 1, \dots, p$, as a realization of $\mathbf{Y}_j = (Y_{1j}, \dots, Y_{nj})'$, represents the data measured at the sampling locations corresponding to the j -th response. Let $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_p)'$ be the vector of all variables at all sampling locations. By extending the skew Gaussian model of [18] to the multivariate case, we define the p -dimensional random process $\mathbf{S}(x) = (S_1(x), \dots, S_p(x))'$ such that:

$$S_j(x) = \alpha_j |V_j(x)| + \sigma_j U_j(x), \quad j = 1, \dots, p,$$

as a multivariate skew Gaussian process. In above equation, $\alpha_j \in \mathbb{R}$, $\sigma_j > 0$, $\mathbf{V}(x) = (V_1(x), \dots, V_p(x))'$ and $\mathbf{U}(x) = (U_1(x), \dots, U_p(x))'$ are independent multivariate stationary Gaussian processes with zero means and covariance matrices $C_V(\cdot; \theta_1)$ and $C_U(\cdot; \theta_2)$, respectively. Notice that $\frac{S_j(x)}{(\alpha_j^2 + \sigma_j^2)^{\frac{1}{2}}}$ has a skew normal distribution with the shape parameter $\frac{\alpha_j}{\sigma_j}$. We have $\mathbf{V} = (\mathbf{V}'(x_1), \dots, \mathbf{V}'(x_n))' \sim N_{np}(\mathbf{0}, \Sigma_{\theta_1})$ and $\mathbf{U} = (\mathbf{U}'(x_1), \dots, \mathbf{U}'(x_n))' \sim N_{np}(\mathbf{0}, \Sigma_{\theta_2})$, respectively, where Σ_{θ_1} and Σ_{θ_2} are the covariance matrix of dimension $np \times np$. These matrices are specified through parametric structures describing the spatial dependency as well as the correlation between the responses. We could use a linear model of coregionalization for \mathbf{V} and \mathbf{U} . Although, for convenience, we define $\Sigma_{\theta_1} = B_1(\phi_1) \otimes C_1(\gamma_1)$, where $B_1(\phi_1)$ is the matrix specified through the correlation function $\rho_1(\cdot; \phi_1)$ and the matrix $C_1(\gamma_1)$ defines the correlation between responses. The parameter vector $\theta_1 = (\phi_1, \gamma_1)$ denotes the whole parameters included in matrices $B_1(\phi_1)$ and $C_1(\gamma_1)$. Similarly, the covariance matrix Σ_{θ_2} is defined.

Conditional on the realized values of latent random variables \mathbf{V} and \mathbf{U} , we suppose that Y_{ij} are mutually independent with distributions $f_j(y|U_j(x_i), V_j(x_i)) = f(y; M_j(x_i))$ of an exponential family, specified by the values of the conditional expectations $M_j(x_i) = E(Y_{ij}|U_j(x_i), V_j(x_i))$;

$$M_j(x_i) = g_j^{-1}(\mathbf{d}_j(x_i)^T \beta_j + \alpha_j |V_j(x_i)| + \sigma_j U_j(x_i)),$$

for some known link function g_j . The $q_j \times 1$ vector β_j shows regression coefficients. Then, the likelihood function of the model parameters $\eta = (\beta_1, \dots, \beta_p, \alpha_1, \dots, \alpha_p, \sigma_1, \dots, \sigma_p, \theta_1, \theta_2)$ by marginalizing can be written as:

$$\begin{aligned} L(\eta; \mathbf{y}) &= \int f(\mathbf{y}, \mathbf{u}, \mathbf{v} | \eta) d\mathbf{u} d\mathbf{v} \\ &= \int f(\mathbf{y} | \mathbf{u}, \mathbf{v}, \eta) \pi(\mathbf{u}, \mathbf{v} | \eta) d\mathbf{u} d\mathbf{v}. \end{aligned} \tag{2.1}$$

The ML estimate of parameters vector $\hat{\eta}$ is the value of η which maximizes the likelihood function (2.1). Because $L(\eta; \mathbf{y})$ involves a high-dimensional intractable integral, it is infeasible to maximize the likelihood function of observed data directly. The maximum likelihood estimate $\hat{\eta}$ can be obtained as a solution to $h(\hat{\eta}; \mathbf{y}) = 0$, where the function h is

$$\begin{aligned} h(\eta; \mathbf{y}) &= \frac{\partial \ln L(\eta; \mathbf{y})}{\partial \eta} = E\left(\frac{\partial \ln f(\mathbf{y}, \mathbf{u}, \mathbf{v} | \eta)}{\partial \eta} | \mathbf{y}\right) \\ &= \int \frac{\partial \ln f(\mathbf{y}, \mathbf{u}, \mathbf{v} | \eta)}{\partial \eta} \pi(\mathbf{u}, \mathbf{v} | \eta, \mathbf{y}) d\mathbf{u} d\mathbf{v}. \end{aligned}$$

2.1. Monte Carlo approximation of function h

Suppose we know the values of all parameters in the vector η . As the distribution $\pi(\mathbf{u}, \mathbf{v} | \mathbf{y}, \eta)$ can not be analytically solved, then one can apply Markov chain Monte Carlo methods to sample from it and approximate the function h . To implement this, a Gibbs sampling algorithm could be applied. The algorithm iteratively proceeds through update the latent variables \mathbf{u} and \mathbf{v} based on sampling from their complete conditional distributions:

$$\begin{aligned} \pi(\mathbf{u} | \mathbf{y}, \mathbf{v}, \eta) &\propto f(\mathbf{y} | \mathbf{u}, \mathbf{v}, \eta) \pi(\mathbf{u}) \\ &\propto \prod_{j=1}^p \prod_{i=1}^n f(y_{ij} | U_j(x_i), V_j(x_i), \eta) \exp\left(-\frac{1}{2}(\mathbf{u}^T \Sigma_{\theta_1}^{-1} \mathbf{u})\right), \\ \pi(\mathbf{v} | \mathbf{y}, \mathbf{u}, \eta) &\propto f(\mathbf{y} | \mathbf{u}, \mathbf{v}, \eta) \pi(\mathbf{v}) \\ &\propto \prod_{j=1}^p \prod_{i=1}^n f(y_{ij} | U_j(x_i), V_j(x_i), \eta) \exp\left(-\frac{1}{2}(\mathbf{v}^T \Sigma_{\theta_2}^{-1} \mathbf{v})\right). \end{aligned}$$

These full conditionals do not define standard probability distributions. So drawings from them are not that trivial to conduct. To solve this problem, the Metropolis-Hastings algorithm could be used. Although this algorithm requires effective proposal density with careful tuning which are difficult to provide in our problem. Recently, auxiliary variable methods based on slice sampler is found to provide an attractive strategy hence, are receiving utmost attention by those who used the Markov chain Monte Carlo (MCMC) algorithms to simulate from complex nonnormalized multivariate densities ([13]). In this paper, for sampling from the full conditional $\pi(\mathbf{u} | \mathbf{y}, \mathbf{u}, \eta)$, we implement slice sampling algorithm based on two auxiliary variables ([1]). Similarly, draws could be obtained from $\pi(\mathbf{v} | \mathbf{y}, \mathbf{u}, \eta)$. Specifically, if $Z_1 | \mathbf{y}, \mathbf{u}, \mathbf{v}, \eta$ and $Z_2 | \mathbf{u}$ have the uniform distribution on the intervals $[0, f(\mathbf{y} | \mathbf{u}, \mathbf{v}, \eta)]$ and $[0, \pi(\mathbf{u})]$, respectively, then

$$\pi(Z_1, Z_2, \mathbf{u} | \mathbf{y}, \mathbf{v}, \eta) \propto I_{\{Z_1 < f(\mathbf{y} | \mathbf{u}, \mathbf{v}, \eta)\}} I_{\{Z_2 < \pi(\mathbf{u})\}}, \tag{2.2}$$

where I denotes the indicator function. Thus,

$$\pi(\mathbf{u} | \mathbf{y}, \mathbf{v}, Z_1, Z_2) \propto I_{\{Z_1 < f(\mathbf{y} | \mathbf{u}, \mathbf{v}, \eta)\}} I_{\{Z_2 < \pi(\mathbf{u})\}}. \tag{2.3}$$

Now, if e_1 and e_2 represent exponential distribution with mean 1, say $\exp(1)$, then $\log Z_1 = \log f(\mathbf{y} | \mathbf{u}, \mathbf{v}, \eta) - e_1$ and $\log Z_2 = \log \pi(\mathbf{u}) - e_2$, given $\mathbf{y}, \mathbf{v}, \eta$. Based on these assumptions, we introduce an algorithm for sampling of $\pi(\mathbf{u} | \mathbf{y}, \mathbf{u}, \eta)$. Indeed, given $\mathbf{y}, \mathbf{v}^{(t)}, \eta$, we can summarize the main steps in iteration $(t + 1)$ th of the slice sampling algorithm as:

- (1) Draw $e_1^{(t+1)}$ and $e_2^{(t+1)}$ of $\exp(1)$, and let $a_t = \log f(\mathbf{y} | \mathbf{u}^{(t)}, \mathbf{v}^{(t)}, \eta) - e_1^{(t+1)}$, $b_t = \log \pi(\mathbf{u}^{(t)}) - e_2^{(t+1)}$.
- (2) Draw $\mathbf{u}^{(t+1)}$ from a uniform distribution on $\{a_t < \log f(\mathbf{y} | \mathbf{u}, \mathbf{v}^{(t)}, \eta)\} \cap \{b_t < \log \pi(\mathbf{u}) - e_2^{(t+1)}\}$.

At iteration $t + 1$ of the algorithm, let $\Pi_\eta(\mathbf{u}, \mathbf{v}, A) = P_\eta[(\mathbf{U}^{(t+1)}, \mathbf{V}^{(t+1)}) \in A | \mathbf{u}^{(t)} = \mathbf{u}, \mathbf{V}^{(t)} = \mathbf{v}]$ be the Markov transition with $\pi(\mathbf{u}, \mathbf{v} | \mathbf{y}, \eta)$ as its unique stationary distribution.

2.2. Maximum Likelihood Estimation

The likelihood function (2.1) does not have a simple form and direct maximization of it seems intractable. Here, we propose using the Stochastic Approximation Boosting algorithm, proposed by [?], to obtain the ML estimates in the context of the spatial generalized linear mixed models. Really, the component-wise boosting approach makes the algorithm straightforward to implement. We now consider and implement a SAB algorithm. For this purpose, we use the notation $\eta = (\eta_1, \eta_2, \eta_3)'$ for model parameters where $\eta_1 = (\beta_1, \dots, \beta_p, \sigma_1, \dots, \sigma_p, \alpha_1, \dots, \alpha_p)$, $\eta_2 = \theta_1$ and $\eta_3 = \theta_2$. For $j = 1, 2, 3$, we also set $H_j(\eta, \mathbf{U}, \mathbf{V}) = \frac{\partial \ln f(\mathbf{y}|\mathbf{U}, \mathbf{V}|\eta)}{\partial \eta_j}$, i.e

$$\begin{aligned} H_1(\eta, \mathbf{U}, \mathbf{V}) &= \frac{\partial \ln f(\mathbf{y}|\mathbf{U}, \mathbf{V}, \eta_1)}{\partial \eta_1}, \\ H_2(\eta, \mathbf{U}, \mathbf{V}) &= \frac{\partial \ln \pi(\mathbf{U}|\eta_2)}{\partial \eta_2}, \\ H_3(\eta, \mathbf{U}, \mathbf{V}) &= \frac{\partial \ln \pi(\mathbf{V}|\eta_3)}{\partial \eta_3}, \end{aligned}$$

and $I_j(\eta, \mathbf{U}, \mathbf{V}) = -\frac{\partial H_j(\eta, \mathbf{U}, \mathbf{V})}{\partial \eta_j} - H_j^2(\eta, \mathbf{U}, \mathbf{V})$. Differentiations with respect to parameters could be simply determined, so we omit the details. In the SAB algorithm, instead of updating the whole parameters vector $\eta = (\eta_1, \eta_2, \eta_3)'$, a component of η is chosen and only this component is updated with other component being kept fixed. Specifically, the algorithm goes as follows:

Slice-SAB algorithm

Step 0: Choose initial values $\Gamma_j^{(0)}$ and $\eta_j^{(0)}$, $j = 1, 2, 3$, $\mathbf{U}_m^{(0)}$ and $\mathbf{V}_m^{(0)}$ and set $t = 1$.

Step 1: Given t : Set $\mathbf{U}_0^{(t)} = \mathbf{U}_m^{(t-1)}$ and $\mathbf{V}_0^{(t)} = \mathbf{V}_m^{(t-1)}$, and for $k = 1 \dots, m$, generate $\mathbf{U}_k^{(t)}$ and $\mathbf{V}_k^{(t)}$ from the Markov transition probability $\Pi_{\eta^{(t-1)}}(\mathbf{u}_{k-1}^{(t)}, \mathbf{v}_{k-1}^{(t)}, \cdot)$ based on the proposed MCMC algorithm.

Step 2: For $j = 1, \dots, r$, update the estimates as follows:

$$\begin{aligned} \Gamma_j^{(t)} &= \Gamma_j^{(t-1)} + \gamma \{ \bar{I}_j(\eta^{(t-1)}, \mathbf{U}^{(t)}, \mathbf{V}^{(t)}) - \Gamma_j^{(t-1)} \} \\ \delta_j^{(t)} &= \gamma (\Gamma_j^{(t)})^{-1} \bar{H}_j(\eta^{(t-1)}, \mathbf{U}^{(t)}, \mathbf{V}^{(t)}) \end{aligned}$$

where $\mathbf{U}^{(t)} = (\mathbf{U}_1^{(t)}, \dots, \mathbf{U}_m^{(t)})$ and $\mathbf{V}^{(t)} = (\mathbf{V}_1^{(t)}, \dots, \mathbf{V}_m^{(t)})$ and

$$\begin{aligned} \bar{I}_j(\eta^{(t-1)}, \mathbf{U}^{(t)}, \mathbf{V}^{(t)}) &= \frac{1}{m} \sum_{k=1}^m I_j(\eta^{(t-1)}, \mathbf{U}_k^{(t)}, \mathbf{V}_k^{(t)}) \\ \bar{H}_j(\eta^{(t-1)}, \mathbf{U}^{(t)}, \mathbf{V}^{(t)}) &= \frac{1}{m} \sum_{k=1}^m H_j(\eta^{(t-1)}, \mathbf{U}_k^{(t)}, \mathbf{V}_k^{(t)}). \end{aligned}$$

Step 3: Select component $S_t = \arg \max_j \delta_j^{(t)} \bar{H}_j(\eta^{(t-1)}, \mathbf{U}^{(t)}, \mathbf{V}^{(t)})$.

Step 4: Update $\eta_{S_t}^{(t)} = \eta_{S_t}^{(t-1)} + \delta_{S_t}^{(t)}$ and $\eta_j^{(t)} = \eta_j^{(t-1)}$ for $j \neq S_t$. Hence $\eta^{(t)}$ is defined.

Step 5: Set $t = t + 1$ and go step 1 until the sequence $\{\eta^{(t)}\}$ converges.

It must be noted that in the algorithm, the parameter m represents the total number of Monte Carlo simulations in each iteration. To ensure against a local, rather than global maximum, the proposed algorithm should be executed from multiple initial values.

2.3. Prediction at new locations

An important goal of the SGLMM is prediction of random effects at new locations. In what follows, we consider the prediction of random effects at location x_0 , say $\mathbf{S}_0 = (S_1(x_0), \dots, S_p(x_0))'$, based on the predictive distribution

$$\begin{aligned} f(\mathbf{S}_0|\mathbf{y}, \hat{\boldsymbol{\eta}}) &= \int \int \int f(\mathbf{S}_0, \mathbf{u}, \mathbf{v}, \mathbf{u}_0|\mathbf{y}, \hat{\boldsymbol{\eta}}) d\mathbf{u} d\mathbf{v} d\mathbf{u}_0 \\ &= \int \int \int f(\mathbf{S}_0|\mathbf{u}_0, \mathbf{u}, \mathbf{v}, \hat{\boldsymbol{\eta}}) f(\mathbf{u}_0|\mathbf{u}, \hat{\boldsymbol{\eta}}) \pi(\mathbf{u}, \mathbf{v}|\mathbf{y}, \hat{\boldsymbol{\eta}}) d\mathbf{u} d\mathbf{v} d\mathbf{u}_0, \end{aligned} \quad (2.4)$$

where \mathbf{u}_0 denotes the unknown latent values at new location and $f(\mathbf{S}_0|\mathbf{u}_0, \mathbf{u}, \mathbf{v}, \hat{\boldsymbol{\eta}})$ and $f(\mathbf{u}_0|\mathbf{u}, \hat{\boldsymbol{\eta}})$ are conveniently normal distributions. To estimate the predictive distribution (2.4), we first use the described algorithm to obtain J draws after burn-in period from $\pi(\mathbf{u}, \mathbf{v}|\hat{\boldsymbol{\eta}})$. Simulating from $f(\mathbf{S}_0|\mathbf{u}_0, \mathbf{u}, \mathbf{v}, \hat{\boldsymbol{\eta}})$, in which $\mathbf{u}_0^{(j)}$, $j = 1, \dots, J$ are generated from $f(\mathbf{u}_0|\mathbf{u}^{(j)}, \hat{\boldsymbol{\eta}})$, we can sample from $f(\mathbf{S}_0|\mathbf{y}, \hat{\boldsymbol{\eta}})$, say $[\mathbf{S}_0]^{(j)}$, $j = 1, \dots, J$. Then, a prediction of \mathbf{S}_0 could simply made by these values.

3. Simulation Study

To examine the performance of the proposed method in estimation of parameters and compare the results of skew normal and normal prior models for multivariate spatial GLMMs, a simulation example is conducted. We also compare the obtained estimates by univariate and multivariate procedures in estimating of regression parameters which have the same interpretation in assumed models. In this example, we simulate 50 data sets from bivariate spatial Poisson model with skew normal random effects on two small and medium equally spaced regular grids of locations, 10×10 and 30×30 in the region of the unit square $D = [0, 1] \times [0, 1]$. Specifically, we consider the conditionally mean $M_j(x_1, x_2) = \exp\{\beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + S_j(x_1, x_2)\}$, $j = 1, 2$, with $\beta_1 = (0, 0.1, -0.1)$ and $\beta_2 = (0, -0.2, 0.2)$. Let $S_1(\cdot)$ and $S_2(\cdot)$ be skew Gaussian random fields with $\alpha_1 = 5$, $\alpha_2 = 10$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho_1(h; \phi_1) = \rho_2(h; \phi_2) = \exp(-||h||)$ and $\gamma_1 = \gamma_2 = 0.7$. Here, the spatial correlation function for two latent processes $V(\cdot)$ and $U(\cdot)$ are exponential with $\phi_1 = \phi_2 = 1$. Each simulated data set is analyzed by the bivariate spatial GLMM under the assumptions of skew normal and normal random effects as well as univariate and bivariate cases. By setting the true parameters as starting values, we applied the slice-SAB algorithm for finding the maximum likelihood estimates of the model parameters. The slice-SAB algorithm is implemented with the total number of Monte Carlo simulations performed at each iteration as $m = 100$.

Now, in order to illustrate the performance of the method, we obtain two measures, the average of bias and the root of mean-square error of regression parameters estimate, i.e. $\hat{\beta}_{11}$, $\hat{\beta}_{12}$ and $\hat{\beta}_{21}$, $\hat{\beta}_{22}$. The results are summarized in Table 1. As seen, the bivariate model with skew Gaussian random effects (SN-BSGLG) has the best performance, in the sense that the estimates under this model are most appropriate than those of the bivariate model with Gaussian random effects (N-BSGLG) and the univariate models with skew Gaussian random effects (SN-USGLG). Beside that, for $n = 30 \times 30$, the bias and MSE of the SN-USGLG model are lower than those of N-BSGLG in most cases. For more comparisons, we fitted two multivariate models for 50 simulated data sets and then calculated the Akaike information criterion (AIC). The *mean* $\pm 1.96 \times$ *standard error* of AIC values based on 50 replications were obtained as (708.805, 815.481) and (919.146, 1087.427), respectively.

We can imply as a result of this table that the proposed method gives satisfactory results and brought valid inference.

Table 1 Bias and mean square errors of regression parameters estimate under various models

	True Value	Normal BSGLMM		Skew-Normal BSGLMM		Skew-Normal USGLMM	
		Bias	MSE	Bias	MSE	Bias	MSE
Grid 10×10	$\beta_{11} = 0.1$	-0.4262	0.0402	0.1097	0.0105	0.3058	0.0784
	$\beta_{21} = -0.1$	0.3175	0.0306	-0.0716	0.0100	0.3296	0.0523
	$\beta_{21} = -0.2$	-0.3690	0.0505	-0.1893	0.0085	0.1641	0.0337
	$\beta_{22} = 0.2$	-0.2070	0.0370	-0.0980	0.0098	-0.3472	0.0693
Grid 30×30	$\beta_{11} = 0.1$	-0.2701	0.0358	0.0878	0.0095	-0.0812	0.0091
	$\beta_{21} = -0.1$	0.1922	0.0260	-0.0521	0.0059	0.2019	0.0322
	$\beta_{21} = -0.2$	-0.1789	0.0144	-0.1384	0.0078	0.1272	0.0206
	$\beta_{22} = 0.2$	0.1967	0.0400	0.0259	0.0039	-0.1365	0.0081

4. Numerical Example

An application of the proposed model to soil pollutant data in Golestan province, located in north of Iran, is presented. Soil pollution is a widespread problem and can be harmful to plants, animals and humans. Specially, soil influences human health directly because the different elements in the soil can easily be absorbed by plants. The analyzed data set comprises a total of 215 locations that were collected for mapping the beryllium (Be) and lead (Pb) contaminated areas in soil of Golestan province. In this data set which measured by a university research, the values of two heavy metals have been reported as a binary variable by clipping them via permissible thresholds. Sampling locations and the responses values are shown in Figure 1. As observed two variables behave similarly in the study region which indicates strong correlation between them. Consider now two response variables $Y(x) = (Y_1(x), Y_2(x))$, where $Y_1(x)$ and $Y_2(x)$ are one if the measurements of Br and Pb exceed safety or standard limits at location x . Hence, we use a bivariate SGLMM model with bernoulli responses and logit link. In the absence of any regression covariates, the mean function is assumed to be constant. To determine proper initial values, the slice-SAB algorithm was first run in the univariate case for each variable, and the estimates thus obtained were used as initial values for the bivariate procedure. Note that the weighted least squares fits provided starting values for single response variables. In the slice-SAB algorithm, the total number of Monte Carlo simulations performed at each iteration was chosen $m = 100$.

Based on the slice-SAB algorithm, the parameter estimates are obtained as: $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\phi}_1, \hat{\phi}_2, \hat{\gamma}) = (-1.278, -1.699, 3.921, 4.413, 2.532, 1.982, 0.360, 0.124, 0.591)$. To make a comparison, we also apply the proposed algorithm for the bivariate SGLMM with Gaussian random effects. Now, we compare the performance of the two models Based on AIC criterion. The values of this criterion were obtained as 426.891 and 773.257, respectively, which shows the advantage of working with the proposed model in this application. Finally, the prediction maps of two heavy metals were plotted in Figure 2 to identify contaminated areas and better understand soil pollution condition of the region.

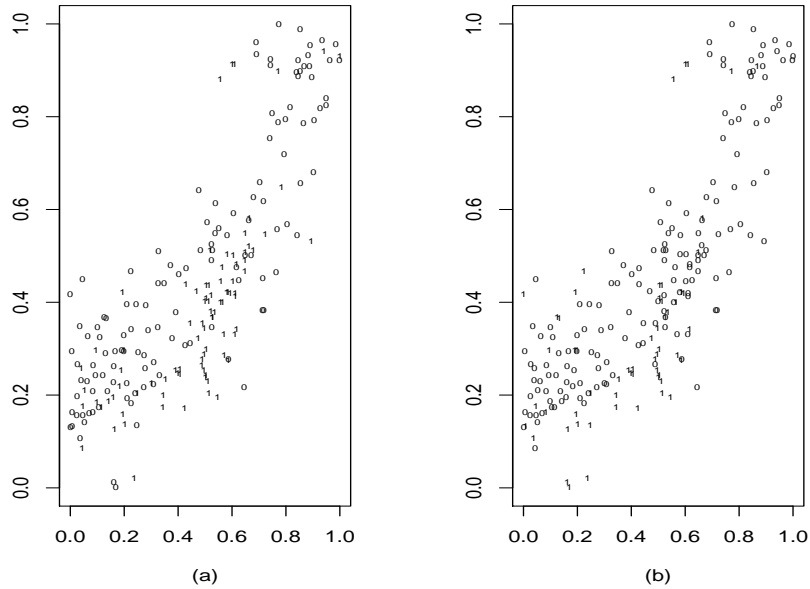


Fig. 1. Sampled locations and values of (a) Pb and (b) Be.

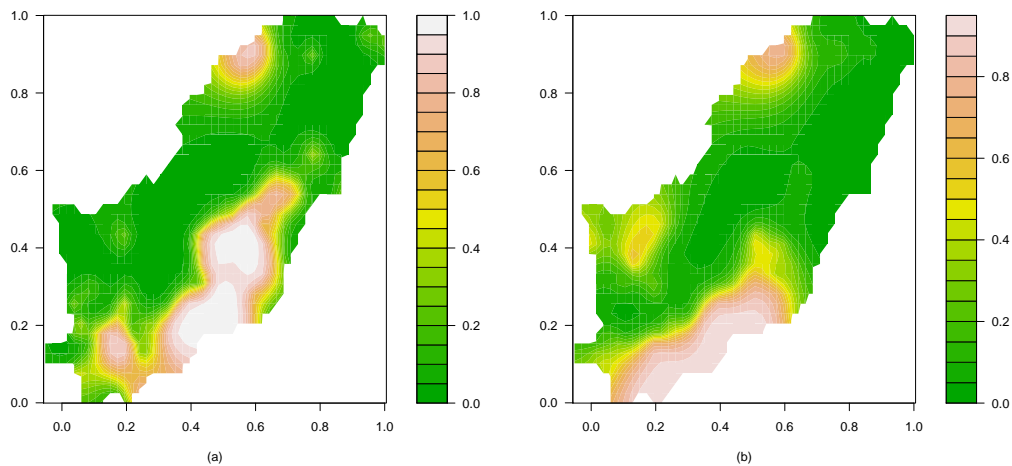


Fig. 2. The prediction map corresponding to probability of pollution of (a) Pb, (b) Be.

5. Concluding Remarks

The main aim of this paper was to introduce a multivariate skew Gaussian random process and use it in the structure of spatial generalized linear mixed models. Further, we showed how ML estimates of this complex and high dimensional model can be computed using a stochastic approximation boosting algorithm. This method eliminates the costs of computing numerical approximation. In

addition, our results showed that the slice-SAB algorithm gives reasonable parameter estimates. As a result from the simulation and numerical examples, we came to the conclusion that the bivariate skew Gaussian random effects model outperforms the bivariate Gaussian random effects model. The special case considered in this paper includes a skew Gaussian random field for the random effects, though the proposed algorithm could be generalized to more complex model for random effects instead of the skew Gaussian random field. In this setting, the computational burden might increase. This is an interesting area to investigate in further research.

References

- [1] D.K. Agarwal and A.E. Gelfand, Slice sampling for simulation based fitting of spatial data models, *Stat. Comput.* **15** (2005) 61–69.
- [2] A. Azzalini and A. Capitanio, Statistical applications of the multivariate skew normal distribution, *J. R. Stat. Soc., series B.* **61** (1999) 579–602.
- [3] A. Azzalini and A. Dalla Valle, The multivariate skew-normal distribution, *Biometrika* **83** (1996) 715–726.
- [4] O.F. Christensen and R.P. Waagepetersen, Bayesian prediction of spatial count data using generalized linear mixed models, *Biometrics* **58** (2002) 280–286.
- [5] O.F. Christensen, Monte Carlo maximum likelihood in model-based geostatistics, *J. Comp. Graph. Stat.* **13** (2004) 702–718.
- [6] O.F. Christensen, G.O. Roberts and M. Sköld, Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *J. Comp. Graph. Stat.* **15** (2006) 1–17.
- [7] P.J. Diggle, J.A. Tawn and R.A. Moeved, Model based geostatistics (with discussion), *Appl. Stat.* **47** (1998) 299–350.
- [8] J. Eidsvik, S. Martino and H. Rue, Approximate Bayesian inference in spatial generalized linear mixed models, *Scand. J. Stat.* **36** (2009) 1–22.
- [9] M.G. Genton and H. Zhang, Identifiability problems in some non-Gaussian spatial random fields, *Chil. J. Stat.* **3** (2012) 171–179.
- [10] M.G. Gu and F.H. Kong, A stochastic approximation algorithm with Markov chain Monte Carlo method for incomplete data estimation problems, *Proc. Natl. Acad. Sci. U.S.A.* **95** (1998) 7270–7274.
- [11] H. Kim and B.K. Mallick, A Bayesian prediction using the skew-gaussian processes, *J. Stat. Plan. Inference* **120** (2004) 85–101.
- [12] E. Kuhn and M. Lavielle, Coupling a stochastic approximation version of EM with a MCMC procedure, *ESAIM, Probab. stat.* **8** (2004) 115–131.
- [13] R. Neal, Slice sampling, *Ann. Stat.* **31** (2003) 705–767.
- [14] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Stat.* **22** (1951) 400–407.
- [15] J. Sexton and P. Laake, Stochastic approximation boosting for incomplete data problem, *Biometrics* **65** (2009) 1156–1163.
- [16] C. Varin, G. Host and O. Skare, Pairwise likelihood inference in spatial generalized linear mixed models, *Comput. Stat. Data Anal.* **49** (2005) 1173–1191.
- [17] H. Zhang, On estimation and prediction for spatial generalized linear mixed models, *Biometrics* **58** (2002) 129–136.
- [18] H. Zhang and A. El-Shaarawi, On spatial skew-Gaussian processes and applications, *Environmetrics* **21** (2010) 33–47.