Environ Ecol Stat (2016) 23:205–217
DOI 10.1007/s10651-015-0335-6

A new algorithm for estimating the parameters of the spatial generalized linear mixed models

Fatemeh Hosseini[1]

Received: 5 February 2015 / Revised: 27 November 2015 / Published online: 12 January 2016

Abstract Non-Gaussian spatial responses are usually modeled using a spatial generalized linear mixed model with location specific latent variables. The likelihood function of this model cannot usually be given in a closed form, thus the maximum likelihood approach is very challenging. So far, several numerical algorithms to solve the problem of calculating maximum likelihood estimates of this model have been presented. In this paper to estimate the parameters an approximate method is considered and a new algorithm is introduced that is much faster than existing algorithms but just as accurate. This is called the Approximate Expectation Maximization Gradient algorithm. The performance of the proposed algorithm and is illustrated with a simulation study and on a real data set.

Keywords Approximate Expectation Maximization Gradient · Latent variables · Spatial generalized linear mixed models

1 Introduction

The spatial generalized linear mixed model (SGLMM) is commonly used for count or proportion data acquired over a continuous spatial domain, see e.g. Diggle et al. (1998). Spatial correlation of the data is usually modeled by latent variables. In biomedical studies, Breslow and Clayton (1993) considered two approximate methods,

Handling Editor: Bryan F. J. Manly.

✉ Fatemeh Hosseini
fatemeh.hoseini@semnan.ac.ir

[1] Department of Statistics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, Semnan 35131-19111, Iran

the Penalised Quasi-Likelihood and the Marginal Quasi-Likelihood, to infer a generalized linear mixed model.

Usually, in this model, the likelihood function cannot be given in a closed form and maximum likelihood estimation generally involves numerical integration of a high-dimensional integral. The integral may be computed by Monte Carlo integration. McCulloch (1997) reviews several Monte Carlo techniques for ML estimation within GLMMs. Zhang (2002) used a ML approach together with a Monte Carlo Expectation Maximization (MCEM) algorithm to estimate parameters of a general spatial GLMM and Baghishani et al. (2011, 2012) also adapted the data cloning algorithm for computing the ML estimate of the SGLM model. As an application of data cloning, Torabi (2013) conducted a frequentist analysis of GLMM with covariates subject to the measurement error model. Torabi (2015) proposed a frequentist approach based on data cloning to predict spatial latent variables and kriging. Evangelou et al. (2011) proposed an approximate likelihood method based on the asymptotic expansion of the log-likelihood using the modified Laplace approximation and applied this method for estimation and prediction for SGLMM.

Monte Carlo sampling can be time-consuming if the Markov chain mixes slowly, and may be computationally very slow or even prohibitive. An alternative approach is to use approximate inference based on deterministic and numerical methods. In the Bayesian framework, Rue and Martino (2007), Rue et al. (2009) and Eidsvik et al. (2009) argue that this type of approximate inference takes seconds of computation time, while Monte Carlo sampling can take days.

The main methodological idea is to use their approximation to obtain ML estimations and predictions. With considering normal distribution for the spatial latent variables in SGLMM, we introduced a new approximate algorithm based on likelihood inference. Our presentation is written for the typical geostatistical application with discrete spatial data acquired on a non-lattice continuous domain. The methods generalize to other settings such as lattice data or observations obtained on a map of regions.

This paper is organized as follows: in Sect. 2, the SGLMM with the normal distribution for the latent variables is defined. The proposed method of approximate inference and prediction is described in Sect. 3. In Sect. 4, an algorithm is introduced to obtain the ML estimations of the model parameters. In Sect. 5, the performance of the proposed algorithm and the approximate prediction method are illustrated on a simulation study and on an application to freeze/frost data from Semnan region of Iran. Finally, closing remarks are given in Sect. 5.

2 Spatial GLMM

In order to define a SGLMM, let $\boldsymbol{x} = (x_1, \ldots, x_n)'$ be spatial latent variables at n sites $\{s_1, \ldots, s_n\}$ in a domain $D \subseteq R^n$ with density $N_n(H\boldsymbol{\beta}, \Sigma_{\boldsymbol{\theta}})$. Here, the location parameter $H\boldsymbol{\beta}$ consists of $n \times (p+1)$ matrix H of covariates, and $p+1$ regression parameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$. The spatial interaction matrix $\Sigma_{\boldsymbol{\theta}}$ is a positive definite $n \times n$ matrix, with two dimensional parameter $\boldsymbol{\theta} = (\sigma, \varphi)'$ indicative of the scale and spatial correlation length, respectively. In the example below we use an isotropic exponential correlation structure for the entries in this matrix. This entails that $\Sigma_{\boldsymbol{\theta}}(i, j) = \sigma^2 \exp(-\| s_i - s_j \|/\varphi)$, where$\| s_i - s_j \|$ is the Euclidean distance between sites s_i and s_j. Therefore, the vector of the model parameters is defined as $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\theta}')'$.

We consider the situation where sites $\{s_1, \ldots, s_k\}$ are observation sites, while one of our goals is to predict the spatial latent variables at the unobserved sites $\{s_{k+1}, \ldots, s_n\}$. The latent variables at the k observed sites are denoted by $\boldsymbol{x}^{obs} = A\boldsymbol{x}$, where $A = [I_{k\times k}|\mathbf{0}_{k\times n-k}]$. Thus, the matrix A decomposes $\boldsymbol{x}$ into $\boldsymbol{x} = (\boldsymbol{x}^{obs'}, \boldsymbol{x}^{pred'})'$, where the elements of $\boldsymbol{x}^{\boldsymbol{pred}}$ are latent variables at $(n-k)$ prediction sites. The likelihood part of the model links the data to the latent variables at the k observation sites. Let $\boldsymbol{y}' = (y_1, \ldots, y_k)$ represent the discrete spatial response variables at the observation sites $\{s_1, \ldots, s_k\}$. We assume that the measurements are conditionally independent of an exponential family (McCullagh and Nelder 1989), given by

$$f(y_i|x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}, \quad i = 1, \ldots, k, \tag{1}$$

where $b(\cdot)$ and $c(\cdot)$ are known functions. The mean $\mathrm{E}(y_i|x_i)$ and x_i are in general related by a known link function g, i.e. $\mathrm{E}(y_i|x_i) = g^{-1}(x_i)$. To summarize, the model has the following components:

$$\begin{aligned} f(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{\eta}) &= f\left(\boldsymbol{y}|\boldsymbol{x}^{obs}\right) f(\boldsymbol{x}|\boldsymbol{\eta}) \propto |\Sigma_\theta|^{-1/2} \\ &\quad \times \exp\left\{\sum_{i=1}^{k}[y_i x_i - b(x_i) + c(y_i)] - \frac{1}{2}(\boldsymbol{x} - H\boldsymbol{\beta})'\Sigma_\theta{}^{-1}(\boldsymbol{x} - H\boldsymbol{\beta})\right\}. \end{aligned}$$

3 Approximate inference and prediction

An important problem in the application of an SGLM model is to predict the values of the spatial latent variables at $n-k$ sites $\{s_{k+1}, \ldots, s_n\}$, using response observations $\boldsymbol{y}$ and H. First, we study the prediction of latent variables when the model parameters are known and the regression parameters are considered in the distribution of the latent variables. The approach is based on the following theorem.

Theorem 1 *Let $\boldsymbol{x}$ be Gaussian as $N(H\boldsymbol{\beta}, \Sigma_{\boldsymbol{\theta}})$. If conditionally on $\boldsymbol{x}$, $\boldsymbol{y}$ is an independent random vector of an exponential family, then the minimum mean squared error (MMSE) spatial prediction of $\boldsymbol{x}^{pred}$ at $n-k$ sites $\{s_{k+1}, \ldots, s_n\}$ is given by*

$$\mathrm{E}\left(\boldsymbol{x}^{pred}|\boldsymbol{y}\right) = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\mathrm{E}(\boldsymbol{x}^{obs}|\boldsymbol{y}) - \boldsymbol{\mu}_1\right), \tag{2}$$

Proof Considering $\boldsymbol{x}' = (\boldsymbol{x}^{obs'}, \boldsymbol{x}^{pred'})$, we can write

$$H\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \quad \Sigma_\theta = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},$$

On the other hand, when the joint distribution of $\boldsymbol{x}$ is multivariate normal, the conditional distribution of $(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs})$, is also within the multivariate normal class. Therefore,

$$\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs} \sim N_{n-k}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\boldsymbol{x}^{obs} - \boldsymbol{\mu}_1\right), \Sigma_{22.1}\right),$$

where, $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Using the conditional expectation of the multivariate normal distribution, we get

$$\mathrm{E}\left(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs}\right) = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\boldsymbol{x}^{obs} - \boldsymbol{\mu}_1\right).$$

Due to the model formation, the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ is the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}^{obs}$, this implies

$$\begin{aligned} f\left(\boldsymbol{x}^{obs}, \boldsymbol{x}^{pred}, \boldsymbol{y}\right) &= f\left(\boldsymbol{y}|\boldsymbol{x}^{pred}, \boldsymbol{x}^{obs}\right) f\left(\boldsymbol{x}^{pred}, \boldsymbol{x}^{obs}\right) \\ &= f\left(\boldsymbol{y}|\boldsymbol{x}^{obs}\right) f\left(\boldsymbol{x}^{pred}, \boldsymbol{x}^{obs}\right) \\ &= f\left(\boldsymbol{y}, \boldsymbol{x}^{obs}\right) f\left(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs}\right). \end{aligned}$$

Also, $f(\boldsymbol{x}^{pred}, \boldsymbol{x}^{obs}, \boldsymbol{y}) = f(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs}, \boldsymbol{y})f(\boldsymbol{x}^{obs}, \boldsymbol{y})$, therfore $\mathrm{E}(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs}) = \mathrm{E}(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs}, \boldsymbol{y})$. Now, using the conditional expectation formula $\mathrm{E}(\boldsymbol{x}^{pred}|\boldsymbol{y}) = \mathrm{E}(\mathrm{E}(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs}, \boldsymbol{y})|\boldsymbol{y})$, we get

$$\begin{aligned} \mathrm{E}\left(\boldsymbol{x}^{pred}|\boldsymbol{y}\right) &= \mathrm{E}\left(\mathrm{E}\left(\boldsymbol{x}^{pred}|\boldsymbol{x}^{obs}\right)|\boldsymbol{y}\right) \\ &= \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\mathrm{E}\left(\boldsymbol{x}^{obs}|\boldsymbol{y}\right) - \boldsymbol{\mu}_1\right). \end{aligned}$$

where, to compute the expectation $\mathrm{E}(\boldsymbol{x}^{obs}|\boldsymbol{y})$, we need to find $f(\boldsymbol{x}^{obs}|\boldsymbol{y})$. Since $f(\boldsymbol{x}|\boldsymbol{y}) \propto f(\boldsymbol{x})f(\boldsymbol{y}|\boldsymbol{x})$, we can write

$$f(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{x}'\Sigma_\theta^{-1}\boldsymbol{x} + \boldsymbol{x}'\Sigma_\theta^{-1}H\boldsymbol{\beta} + \sum_{i=1}^{k}[y_i x_i - b(x_i)]\right\}.$$

Thus, $f(\boldsymbol{x}|\boldsymbol{y})$ has a complicated form and $E(\boldsymbol{x}|\boldsymbol{y})$ can not be given in closed form, but it is possible to generate the Monte Carlo random draws $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$ from the conditional distribution $f(\boldsymbol{x}|\boldsymbol{y})$. Then one can compute $\mathrm{E}(\boldsymbol{x}^{obs}|\boldsymbol{y}) \approx \frac{1}{N}\sum_{m=1}^{N}\boldsymbol{x}^{obs\,(m)}$ by the Metropolis–Hastings algorithm. □

MCMC sampling can be time-consuming if the Markov chain mixes slowly, and an alternative approach is to use approximate inference based on deterministic and numerical methods.

3.1 Approximate MMSE prediction

The main methodological idea is to use a normal approximation for the conditional density $f(\boldsymbol{x}|\boldsymbol{y})$, and next use this to obtain approximate MMSE prediction, when the model parameters are known. Eidsvik et al. (2009) used the closedness property of the normal distribution to construct an approximation for the density $f(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$ as following

$$\hat{f}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta}) \approx N_n\left(\hat{\boldsymbol{\mu}}_{x|y,\eta}\left(\boldsymbol{y}, \boldsymbol{x}^0\right), \hat{\Sigma}_{x|y,\eta}\left(\boldsymbol{x}^0\right)\right), \tag{3}$$

where $\hat{\boldsymbol{\mu}}_{x|y}(\boldsymbol{y}, \boldsymbol{x}^0) = H\boldsymbol{\beta} + \Sigma_\theta A' R^{-1}(z(\boldsymbol{y}, \boldsymbol{x}^{obs}) - AH\boldsymbol{\beta})$, and $z_i(y_i, x_i^0) = [y_i - b'(x_i^0) + x_i b''(x_i^0)]/b''(x_i^0)$, $i = 1, \ldots, k$, is a linearization of the exponential family likelihood part of $f(\boldsymbol{y}|\boldsymbol{x}^{obs}) f(\boldsymbol{x}|\boldsymbol{\eta})$ at a fixed value of $\boldsymbol{x}$. Moreover, $R = A\Sigma_\theta A' + P$ P is a diagonal matrix with entries element $P(i, i) = 1/b''(x_i)$, $i = 1, \ldots, k$. Finally, $\hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^0) = \Sigma_\theta - \Sigma_\theta A' R^{-1} A \Sigma_\theta$, see Appendix (Eidsvik et al. 2009) for further explanation.

Now, we assume the model parameters are known and obtain the approximate MMSE prediction of latent variables based on the following methodology: Let $\boldsymbol{x} = (\boldsymbol{x}^{obs'}, \boldsymbol{x}^{pred'})'$, and from (3), we can write

$$\hat{\boldsymbol{\mu}}_{x|y}(\boldsymbol{y}, \boldsymbol{x}^0) = \begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu}_p \end{pmatrix}, \quad \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^0) = \begin{pmatrix} \Sigma_{oo} & \Sigma_{op} \\ \Sigma_{po} & \Sigma_{pp} \end{pmatrix}.$$

By using the property of the multivariate normal distribution, $(\boldsymbol{x}^{pred}|\boldsymbol{y})$ has an approximate normal distribution as $N(\boldsymbol{\mu}_p, \Sigma_{pp})$, therefore,

$$\mathrm{E}\left(\boldsymbol{x}^{pred}|\boldsymbol{y}\right) = \boldsymbol{\mu}_p, \tag{4}$$

is the approximate MMSE prediction.

4 Maximum likelihood estimation

In this section an algorithm is proposed to obtain estimate of model parameters based on the EM gradient (EMG) algorithm and the approximate likelihood approach when the latent variables have a normal distribution. Let $\boldsymbol{x} = (x_1, \ldots, x_n)'$ with density $N_n(H\boldsymbol{\beta}, \Sigma_{\boldsymbol{\theta}})$ and $f(y_i|x_i) = \exp(y_i x_i - b(x_i) + c(y_i))$, then the likelihood function is obtained as

$$L(\boldsymbol{\eta}; \boldsymbol{y}) = \int \prod_{i=1}^{k} f(y_i|x_i) f(\boldsymbol{x}|\boldsymbol{\eta}) d\boldsymbol{x}, \tag{5}$$

where, $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\theta}')'$. Likelihood function (5) is a complicated function. It cannot be given in a closed form, because the integration dimension in (5) is equal to the number of latent variables, and consequently it is intractable to find MLE by directly maximizing L.

The EM iterative algorithm is a broadly applicable statistical technique for maximizing complex likelihoods and handling the incomplete data problem or the likelihood function involves latent variables. In many practical applications, unfortunately, convergence of the EM algorithm can be extremely slow. Lange (1995) proposed the EMG algorithm to accelerate convergence of the EM algorithm. One of the difficulties of general EM is in finding the expected log-likelihood, Wei and Tanner (1990) proposed Monte Carlo approach in finding expected log-likelihood. Zhang (2002) proposed a Monte Carlo EM gradient (MCEMG) algorithm for computing the maximum likelihood estimate. In this algorithm, we have

$$\boldsymbol{\eta}^{(m+1)} = \boldsymbol{\eta}^{(m)} - \left[E\left(\frac{\partial^2 \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}\Big| \boldsymbol{y}\right)\right]^{-1}_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}} \left[E\left(\frac{\partial \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\Big| \boldsymbol{y}\right)\right]_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}}, \tag{6}$$

where, $\ell(\boldsymbol{\eta}) = \ln L(\boldsymbol{\eta}; \boldsymbol{y})$. The conditional expectations in (6) cannot be calculated in closed form but can be approximated using the Monte Carlo samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$. Thus, the MCEMG algorithm iterates the following steps until convergence. The derivatives in (7) can be given in closed form since the distribution of $\boldsymbol{x}$ is multivariate normal as follows:

Algorithm 1 : MCEMG algorithm

1: Choose a starting value $\boldsymbol{\eta}^{(0)}$, such that $L(\boldsymbol{\eta}^{(0)}|\boldsymbol{y}) > 0$ and set $m = 0$.
2: Draw the Monte Carlo samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$ from $f_{x|y}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta}^{(m)})$ using the Metropolis–Hastings algorithm.
3: Calculate the Monte Carlo estimates $\mathrm{E}\{\frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}|\boldsymbol{y}\}$ and $\mathrm{E}\{\frac{\partial \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}|\boldsymbol{y}\}$, using $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$, as $\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial^2 \ln f(\boldsymbol{x}^{(j)}|\boldsymbol{\eta}^{(m)})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\right)$ and $\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\partial \ln f(\boldsymbol{x}^{(j)}|\boldsymbol{\eta}^{(m)})}{\partial \boldsymbol{\eta}}\right)$.
4: Let $\boldsymbol{\eta}^{(m+1)}$ as following from (6) and step 3:

$$\boldsymbol{\eta}^{(m+1)} = \boldsymbol{\eta}^{(m)} - \left[\mathrm{E}\left\{\frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\Big|\boldsymbol{y}\right\}\right]^{-1}_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}} \left[\mathrm{E}\left\{\frac{\partial \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\Big|\boldsymbol{y}\right\}\right]_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}}, \tag{7}$$

5: Set $m = m + 1$. Go to step (2) until convergence is reached.

$$\ln f(\boldsymbol{x}|\boldsymbol{\eta}) = -\frac{1}{2}\ln|\Sigma_\theta| - \frac{1}{2}(\boldsymbol{x} - H\boldsymbol{\beta})'\Sigma_\theta^{-1}(\boldsymbol{x} - H\boldsymbol{\beta}), \tag{8}$$

for example

$$\frac{\partial \ln f(\boldsymbol{x}|\boldsymbol{\beta})}{\partial \beta} = H'\Sigma_\theta^{-1}(\boldsymbol{x} - H\boldsymbol{\beta})$$
$$\frac{\partial \ln f(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2}\mathrm{tr}\left(\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_k}\right) + \frac{1}{2}\boldsymbol{x}'\left(\Sigma_\theta^{-1}\frac{\partial \Sigma_\theta}{\partial \theta_k}\Sigma_\theta^{-1}\right)\boldsymbol{x}$$

$$\frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\theta_k\theta_\ell} = -\frac{1}{2}\mathrm{tr}\Big(\Sigma_\theta^{-1}\frac{\partial^2\Sigma_\theta}{\partial\theta_k\partial\theta_\ell} - \Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_k}\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_\ell}\Big) - \frac{1}{2}\boldsymbol{x}'\Lambda\boldsymbol{x},$$

where, $\Lambda = \Sigma_\theta^{-1}\big(\frac{\partial\Sigma_\theta}{\partial\theta_k}\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_\ell} + \frac{\partial\Sigma_\theta}{\partial\theta_\ell}\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_k} - \frac{\partial^2\Sigma_\theta}{\partial\theta_k\partial\theta_\ell}\big)\Sigma_\theta^{-1}$, (for more detail see, Mardia and Marshall 1984). The conditional expectations in (7) cannot be calculated in closed form, but can be approximated using the Monte Carlo samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$.

4.1 Approximate EMG algorithm

MCEMG algorithm can be applied but may in such cases be computationally very slow or even prohibitive. Here, we introduce a new approximate algorithm to estimate the model parameters and call it the Approximate EMG (AEMG) algorithm. The AEMG algorithm can estimate the model parameters with equivalent accuracy, but is much faster than MCEMG algorithms. The main idea is to use the approximate conditional distribution (3) for the latent variables, therefore

$$\mathrm{E}\left[\{\ln f(\boldsymbol{x}|\boldsymbol{\eta})\}\,|\boldsymbol{y}\right] = \int \ln f(\boldsymbol{x}|\boldsymbol{\eta})\hat{f}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})d\boldsymbol{x}.$$

Now, we can compute the conditional expections in EMG algorithm without MCMC sampling. The algorithm goes as follows:

Algorithm 2 : AEMG algorithm

1: Choose a starting value $\boldsymbol{\eta}^{(0)}$, such that $L(\boldsymbol{\eta}^{(0)}|\boldsymbol{y}) > 0$ and set $m = 0$.
2: Approximation step:
(a) Choose a starting value $\boldsymbol{x}^{(0)}$. For instance set $\boldsymbol{x}^{(0)}$ be the mode of $f(\boldsymbol{x}|\boldsymbol{\eta})$. Set $d = 0$.
(b) From (3), calculate $\hat{f}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}^{(m)}) = N(\hat{\boldsymbol{\mu}}_{x|y,\eta^{(m)}}(\boldsymbol{y},\boldsymbol{x}^{(d)}), \hat{\Sigma}_{x|y,\eta^{(m)}}(\boldsymbol{x}^{(d)}))$
(c) Let $\boldsymbol{x}^{(d+1)}$ be the mode of $\hat{f}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}^{(m)})$.
(d) Set d=d+1. Go to (b). Convergence is obtained after a few iterations.
3: Expectation Maximization Gradient step:
Choose $\boldsymbol{\eta}^{(m+1)}$ such that

$$\boldsymbol{\eta}^{(m+1)} = \boldsymbol{\eta}^{(m)} - \left[\hat{\mathrm{E}}\Big\{\frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\Big|\boldsymbol{y}\Big\}\right]^{-1}_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}}\left[\hat{\mathrm{E}}\Big\{\frac{\partial \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big|\boldsymbol{y}\Big\}\right]_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}} \tag{9}$$

4: Let $m = m + 1$. Go to step (3) until convergence is reached.

To simply and calculate the Eq. (9) in AEMG algorithm, we have

$$\begin{aligned}\boldsymbol{\eta}^{(m+1)} &= \boldsymbol{\eta}^{(m)} - \left[\hat{\mathrm{E}}\Big\{\frac{\partial^2 \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\Big|\boldsymbol{y}\Big\}\right]^{-1}_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}}\left[\hat{\mathrm{E}}\Big\{\frac{\partial \ln f(\boldsymbol{x}|\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\Big|\boldsymbol{y}\Big\}\right]_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}}\\ &= \boldsymbol{\eta}^{(m)} - \left[\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\hat{\mathrm{E}}\Big\{(\ln f(\boldsymbol{x}|\boldsymbol{\eta}))|\boldsymbol{y}\Big\}\right]^{-1}_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}}\left[\frac{\partial}{\partial\boldsymbol{\eta}}\hat{\mathrm{E}}\Big\{(\ln f(\boldsymbol{x}|\boldsymbol{\eta}))\Big|\boldsymbol{y}\Big\}\right]_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}}.\end{aligned} \tag{10}$$

The conditional expectations in (10) can be calculated in closed form, we can write

$$\hat{\mathrm{E}}\left[\{\ln f(\boldsymbol{x}|\boldsymbol{\eta})\}\,|\boldsymbol{y}\right] = -\frac{1}{2}\ln|\Sigma_{\boldsymbol{\theta}}| - \frac{n}{2}\ln 2\pi$$
$$-\frac{1}{2}\mathrm{E}\left\{(\boldsymbol{x}-H\boldsymbol{\beta})'\Sigma_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x}-H\boldsymbol{\beta})|\boldsymbol{y}\right\}. \quad (11)$$

Let $\boldsymbol{W} = [(\boldsymbol{x} - H\boldsymbol{\beta})'|\boldsymbol{y}]$, from (3), $\boldsymbol{W}$ has a multivariate normal distribution as $N(\hat{\boldsymbol{\mu}}_w, \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^0))$, where $\hat{\boldsymbol{\mu}}_w = (\hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{y}, \boldsymbol{x}^0) - H\boldsymbol{\beta})$, thus

$$\hat{\boldsymbol{\mu}}_w = \Sigma_\theta A' R^{-1}(z(\boldsymbol{y}, \boldsymbol{x}^0) - AH\boldsymbol{\beta}).$$

Since $\left\{(\boldsymbol{x} - H\boldsymbol{\beta})'\Sigma_{\boldsymbol{\theta}}^{-1}(\boldsymbol{x} - H\boldsymbol{\beta})|\boldsymbol{y}\right\}$, in Eq. (11) is a quadratic form as $\boldsymbol{W}'\Sigma_{\boldsymbol{\theta}}^{-1}\boldsymbol{W}$, by using the expectation of the quadratic form

$$\hat{\mathrm{E}}\left[\{\ln f(\boldsymbol{x}|\boldsymbol{\eta})\}\,|\boldsymbol{y}\right] = -\frac{1}{2}\ln|\Sigma_{\boldsymbol{\theta}}| - \ln 2\pi - \frac{1}{2}\mathrm{tr}\left(\Sigma_{\boldsymbol{\theta}}^{-1}\hat{\Sigma}_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}}\right)$$
$$-\frac{1}{2}\hat{\boldsymbol{\mu}}_w'\Sigma_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\mu}}_w, \quad (12)$$

and by replacing Eq. (12) in (10) the derivatives can be given in closed form.

5 Examples

In this section, we present a simulation study and a real example with discrete spatial data to evaluate the performance of the proposed algorithm. The real example concerns the frost data, using a binomial likelihood with 179 trials.

5.1 Simulation 1

Here, we present a simulation to study the performance of our proposed algorithm. The main goal of this simulation study is to compare the two algorithms; AEMG and MCEMG, based on running time and estimation accuracy of the model parameters.

First, $n = 85$ random locations are generated inside a irregular grid of Iran region. We fix the parameters of the model and draw latent variables from the distribution $N_{85}(\beta_0 + \beta_1 z, \Sigma_\theta)$, where, the covariate vector h was standard of log(latitude). In the simulation, we set parameters $\beta_0 = -2$, $\beta_1 = 0.7$. We consider an isotropic exponential covariance function as $\sigma^2\exp(-\|h\|/\phi)$, $h = \|s_i - s_j\|$, to build the elements of Σ_θ and set parameters $\boldsymbol{\theta} = (\sigma^2, \phi) = (1, 2)$. Conditional on the latent variables, binomial responses y_i, $i = 1, \ldots, n$ were generated according to $y_i \sim Bin(u_j, p_j)$, where $p_i = \frac{\exp(x_i)}{1+\exp(x_i)}$, $u_i = 50$. One example of generated data are shown in Fig. 1. Keeping the spatial design fixed, the above data generation scheme was carried out for 100 data sets in each case. We run MCEMG algorithm for 1000, 5000 and 10,000 number of iterations in the Monte Carlo step. In most data sets, convergence was achieved in less 10 iterations for AEMG algorithm and 15 iterations

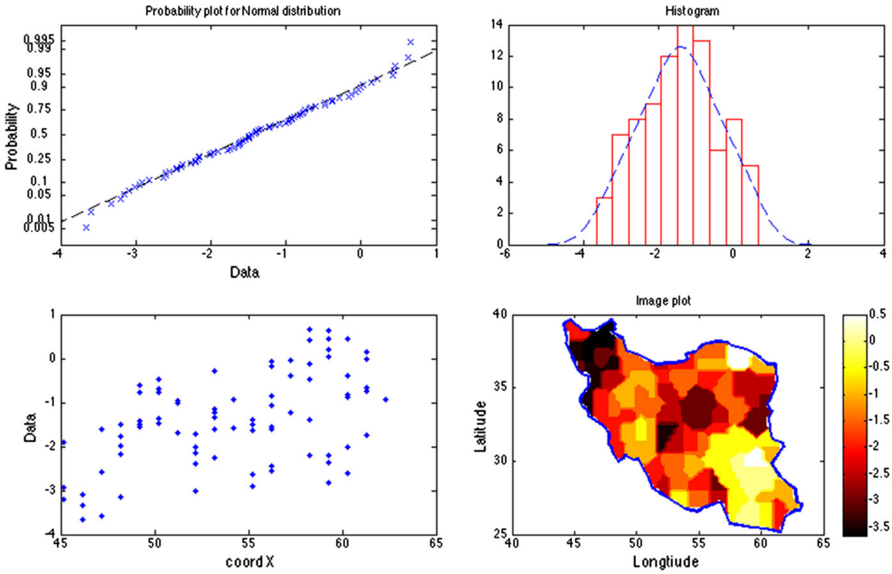

Fig. 1 Realization from the spatial binomial model. From *top-left* to *bottom-right* normal QQ plot, histogram, scatter plot and map of the simulated latent variables

for MCEMG algorithm. By using the MCEMG and AEMG, we obtained ML estimates for 100 data sets and calculated the mean square error (MSE), average estimates and their standard deviations. Each simulated data set was analyzed by the SGLMM under the assumptions of normal latent variables, results are given in Table 1. It seemed that a Monte Carlo sample size 10,000 is sufficient, by this sample size the estimate of the parameters based on the AEMG algorithm are nearly identical to that of the MCEMG algorithm. The running time for AEMG on a data set using a computer 2.4 GHz intel core i5 with 4GB RAM was about 150 s and it was about 3700 s for MCEMG.

6 Frost data

From a meteorological point of view, frost or freezing occurs when the temperature of air falls below the freezing point of water (0 C, 32 F, 273.15 K). This is usually measured at the height of 1.2 m above the ground surface. Frost is an important meteorological parameter and precise knowledge is useful for the researchers in many disciplines. In this paper we studied the data frost in the Semnan province of Iran. Semnan province is located in the central northern portion of the Iran with an area of 96,816 km^2. This province lies at latitudes between 34°40′ and 37°10′ north and longitude between 51°59′ and 57°4′east. It has a changing climate due to its expanse, local and natural factors. The main reason for climate changes of this region is the interaction between different columns of air which enter the region from the southwest and northeast of this province.

Table 1 Simulation results from 100 data set: average of estimates, standard deviation and MSE of the estimates

Par.	Real. V	MCEMG, $N = 1000$			MCEMG, $N = 5000$		
		Ave. Est.	SD	MSE	Ave. Est.	SD	MSE
σ^2	1.0	0.6775	0.0385	0.1173	0.6797	0.0427	0.1139
ϕ	2.0	2.8701	0.3780	1.9595	2.8208	0.2873	1.6647
β_0	−1.0	−0.8013	0.1276	0.1860	−0.9066	0.1262	0.1464
β_1	0.5	0.3935	0.1124	0.1319	0.4230	0.0974	0.1021
Par.	**Real. V**	**MCEMG, N = 10,000**			**AEMG**		
		Ave. Est.	SD	MSE	Ave. Est.	SD	MSE
σ^2	1.0	0.7322	0.0355	0.0837	0.7369	0.0357	0.0807
ϕ	2.0	2.7047	0.4105	1.3802	2.5925	0.2423	0.8796
β_0	−1.0	−0.9460	0.0632	0.0672	−0.9434	0.0760	0.0766
β_1	0.5	0.5466	0.0910	0.0967	0.5070	0.1110	0.0840

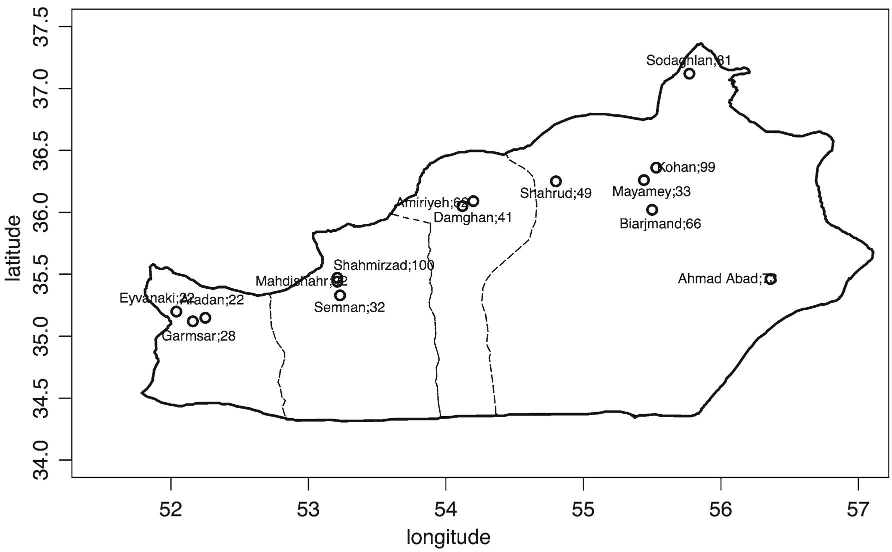

Fig. 2 Frost data shown in a map of Semnan. *Circles* are 14 registration sites of frost

The data are number of days with frost of the number of days in operation ($u_i = 179$), observed for $i = 1, \ldots, 14$ observation sites. This dataset was obtained from Semnan Meteorological Organization. Figure 2 shows the frost data on the map of Semnan. The counts are assumed to be conditionally independent binomial variables, hence $f(y_i|x_i) = \exp\{y_i x_i - u_i \log(1 + \exp(x_i))\}$, where $u_i = 179$, $i = 1, \ldots, 14$. An isotropic exponential correlation structure for $\boldsymbol{\Sigma_\theta}$ is used with the parameter

Table 2 MMSE estimations of the latent variables

Station name	y	$\boldsymbol{x}^{obs}(MCEMG)$	$\boldsymbol{x}^{obs}(AEMG)$
Mehdishahr	92	0.0336	0.0321
Ahmadabad	73	−0.3820	−0.3917
Sodaghlan	81	−0.3064	−0.3097
Kohan	99	0.1627	0.1630
Aradan	22	−1.6762	−1.6715
Eyvanki	22	−1.7776	−1.7818
Amiriyeh	62	−0.6768	−0.6916
Semnan	32	−1.5154	−1.5273
Shahrud	49	−1.1031	−1.1184
Damghan	41	−1.1716	−1.1730
Shahmirzad	100	0.1062	0.0952
Garmsar	28	−1.5233	−1.5168
Meyami	33	−1.4620	−1.4700
Beyarjomand	66	−0.5470	−0.5494

$\boldsymbol{\theta} = (\sigma, \varphi)'$. A normal distribution is used for the latent variables, i.e. $\boldsymbol{x} \sim N(\beta_0 \mathbf{1}', \Sigma_{\boldsymbol{\theta}})$, There are no covariates in the model.

The MCEMG and AEMG algorithms are used to find the ML estimations of the model parameters, the constructed results after convergence of the algorithms are as $\hat{\boldsymbol{\eta}}'_{MCEMG} = (0.654, 4.122, 20.289)$, $\hat{\boldsymbol{\eta}}'_{AEMG} = (0.641, 4.276, 20.599)$. The parameter estimates of the MCEMG algorithm are nearly identical to that of the AEMG. The estimations of the latent variables $\boldsymbol{x}^{obs} = \{x_i; i = 1, \ldots, 14\}$, are fitted by the two algorithms, the results are presented in Table 2. There is not significant difference between the results of the two algorithms and the mean absolute difference between two columns of the Table 2 was 6.7×10^{-3}. MMSE and Approximate MMSE are used to find the prediction of the latent variable, Fig. 3 shows predictive maps and six unobserved sites Lasjerd, Dehnamak, Dehmola, Chashm, Dizej, Abbas Abad is specified in northern provinces. They are drawn on a defined grid of size 150×150, points out the map is removed. 10971 sites on the map is intended to predict. These diagrams show the correlation between the different areas of the frost. No significant difference is seen between the two maps. Mean absolute difference between the predictions in the two methods was 2.4×10^{-3}.

7 Conclusion

The contribution of our paper is to propose methods for approximate frequentist inference and prediction for SGLMM. A new algorithm was used to obtain maximum likelihood estimates of model parameters based on EMG algorithm and an approximate approach. Lange (1995) introduced EMG algorithm that substitutes a one-step Newton–Raphson algorithm for the M-step to speed up convergence of EM and showed this algorithm is locally equivalent to the EM algorithm. Our methods for the inference

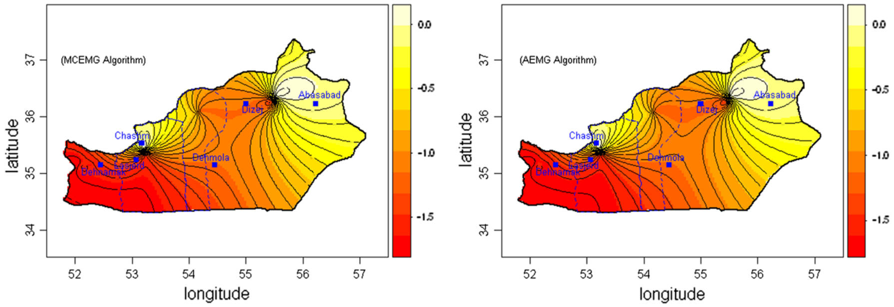

Fig. 3 Prediction maps of frost data obtained by AEMG and MCEMG algorithms for the spatial latent variables

and spatial prediction extend the approach of Eidsvik et al. (2009) and Zhang (2002) to the SGLM models. For AEMG algorithm the estimation and prediction results are nearly identical to that of MCEMG algorithm. The AEMG algorithm and the approximation MMSE predictions took less than 5 min of computer time, while the execution time for 10,000 iterations of the MCEMG algorithm and MMSE predictions were several hours.

Hosseini et al. (2011) and Hosseini and Mohammadzadeh (2012) showed that misspecification of spatial latent variables in a SGLM model affects both the estimation of the parameters and the spatial prediction of the spatial latent variables. Although normal distribution is often used for latent variables in an SGLM model, they used a closed skew normal distribution and proposed methods for approximate Bayesian inference and prediction for SGLMM. Their model is more general than the usual one based on a normal model, since the closed skew normal distribution is fully parametric, and contains several closed form solutions, facilitating efficient inference of model parameters and prediction of the latent process. Also, Varin et al. (2005) considered a pairwise likelihood, which is the product of likelihoods of subsets of data and introduced a new EM algorithm which uses numerical quadrature. This suggests that the closed skew normal is a more robust model and the pairwise likelihood is computationally efficient. The computational savings from using the approximate approach and the pairwise likelihood are great compared to Monte carlo approach and likelihood inference.

References

Baghishani H, Rue H, Mohammadzadeh M (2011) A data cloning algorithm for computing maximum likelihood estimates in spatial generalized linear mixed models. Comput Stat Data Anal 55:1748–1759

Baghishani H, Rue H, Mohammadzadeh M (2012) On a hybrid data cloning method and its application in generalized linear mixed models. Stat Comput 22:597–613

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88:9–25

Diggle P, Tawn JA, Moyeed RA (1998) Model-based geostatistic. J R Stat Soc Ser C Appl Stat 47:299–350

Eidsvik J, Martino S, Rue H (2009) Approximate Bayesian inference in spatial generalized linear mixed models. Scand J Stat 36:1–22

Evangelou E, Zhu Z, Smith RL (2011) Estimation and prediction for spatial generalized linear mixed models using high order Laplace approximation. J Stat Plan Inference 141(11):3564–3577

Hosseini F, Eidsvik J, Mohammadzadeh M (2011) Approximate Bayesian inference in spatial GLMM with skew normal latent variables. Comput Stat Data Anal 55:1791–1806

Hosseini F, Mohammadzadeh M (2012) Bayesian prediction for spatial GLMMs with closed skew normal latent variables. Aust N Z J Stat 54:43–62

Lange K (1995) A gradient algorithm locally equivalent to the EM algorithm. J R Stat Soc 57:425–437

Mardia KV, Marshall RJ (1984) Maximum likelihood estimation of models for residual covariance in spatial statistics. Biometrika 71:135–146

McCullagh P, Nelder JA (1989) Generalized linear models. Chapman and Hall, London

McCulloch CE (1997) Maximum likelihood algorithms for generalized linear mixed models. J Am Stat Assoc 92(437):162–170

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B 71:319–392

Rue H, Martino S (2007) Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. J Stat Plan Inference 137:3177–3192

Torabi M (2013) Likelihood inference in generalized linear mixed measurement error models. Comput Stat Data Anal 57:549–557

Torabi M (2015) Likelihood inference for spatial generalized linear mixed models. Commun Stat Simul Comput 44:1692–1701

Varin C, Høst G, Skare Ø (2005) Pairwise likelihood inference in spatial generalized linear mixed models. Comput Stat Data Anal 49:1173–1191

Wei GCG, Tanner M (1990) A Monte Carlo implementation of the EM algorithm and the poor mans data augmentation. J Am Stat Assoc 85:699–704

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

Fatemeh Hosseini is an assistant professor of Statistics at the Semnan University since 2011 to present. His research focuses on spatial statistics and spatial generalized linear models. He received Ph.D. from the Department of Statistics at the Tarbiat Modares University of Tehran in 2011.