Journal of Sciences, Islamic Republic of Iran 35(2): 135 - 145 (2024) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Penalized Composite Likelihood Estimation for Spatial Generalized Linear Mixed Models

L. Salehi and M. Mohammadzadeh*

Department of Statistics, Tarbiat Modares University, Tehran, Islamic Republic of Iran

Received: 15 October 2024 / Revised: 1 January 2025 / Accepted: 13 January 2025

Abstract

When discussing non-Gaussian spatially correlated variables, generalized linear mixed models have enough flexibility for modeling various data types. However, the maximum likelihood methods are plagued with substantial calculations for large data sets, resulting in long waiting times for estimating the model parameters. To alleviate this drawback, composite likelihood functions obtained from the product of the likelihoods of subsets of observations are used. The current paper uses the pairwise likelihood method to study the parameter estimations of spatial generalized linear mixed models. Then, we use the weighted pairwise and penalized likelihood functions to estimate the parameters of the mentioned models. The accuracy of estimates based on these likelihood functions is evaluated and compared with full likelihood function-based estimation using simulation studies. Based on our results, the penalized likelihood function improved parameter estimation. Prediction using penalized likelihood functions is applied. Ultimately, pairwise and penalized pairwise likelihood methods are applied to analyze count real data sets.

Keywords: Spatial generalized linear model; Composite likelihood; Penalized pairwise likelihood; Weighted pairwise likelihood.

Introduction

Generalized linear models were first introduced by (1), while (2) employed these models to discrete response variables. These models assume independence of observations to establish a linear relationship between the mean of observations and the explanatory variables. An improved class model, the Generalized Linear Mixed (GLMM), is employed for correlated observations. In these models, the assumption of independence of observations is adjusted as conditional independence, and the correlation between the variables is established by introducing random effects through latent variables to the model. For cases in which a model can be established to represent the correlations of spatial responses, the Kriging method can be used for spatial analysis and predict the unknown values through a random field; otherwise, a Spatial Generalized Linear Mixed (SGLMM) can be used. Unlike linear models, the likelihood functions in SGLMMs do not offer a closed form owing to the non-Gaussian nature of the response variable; the parameters cannot, hence, be estimated using the maximum likelihood. Therefore, most articles accept the assumption of latent variables' normality and provide a solution to estimate model parameters and latent variables by maximizing likelihood functions, Penalized quasi-likelihood, or

* Corresponding Author: Tel: +989122066712; Fax: +982182883483; Email: mohsen_m@modares.ac.ir

hierarchical likelihood using numerical methods. Among others (3), utilized maximum likelihood algorithms for GLMMs with non-spatial random effects using numerical methods such as Monte Carlo Expectation Maximization (MCEM) (4). Also employed the MCEM algorithm to estimate the maximum likelihood of the model parameters with the assumption of closed skew normal of the latent variables (5, 6).

Presented another method based on a data cloning algorithm. Recently, approximate Bayesian inference methods, which are less computationally burdensome and hence faster than simulation-based, have garnered much attention among the literature contributors (7).

Introduced approximate Bayesian inference for Gaussian Markov random processes (8) applied a similar method to SGLMs.

Recent studies have studied other approximate methods, which are not based on the complete likelihood of observations. Compared to the approximate likelihood methods, the advantage of these methods is the presumed lack of need for simultaneously modeling all the observations. As such, (9, 10) have employed quasi-likelihood functions, a subclass of composite likelihood methods (11) applied the pairwise composite likelihood approach to binary spatial data using a probit link function for the first time. The evaluation of pairwise likelihood is computationally efficient when the bivariate density function can be computed quickly compared to the joint density of n observations. In spatial modeling, accurately estimating covariance functions is crucial. Pairwise likelihood methods facilitate this by allowing for the estimation of covariance structures based on pairs of observations. Pairwise likelihood methods specifically target bivariate relationships among observations, often more straightforward to model and validate than higher-dimensional dependencies. This focus allows researchers to effectively capture the essential structure of spatial data without needing a complete joint distribution (see 12, 13). A general discussion of pairwise likelihood can be found in (14), which, due to its simplicity, has been applied in many fields of statistics (15) used the pairwise composite likelihood method for SGLMMs, in which a novel EM algorithm that uses numerical quadrature was introduced (16) used the weighted likelihood function for the spatiotemporal data and proved it is a good approximation of maximum likelihood (17) showed that weighting the likelihood function increases the asymptotic relative estimation efficiency in categorical data. We employed the pairwise likelihood function for SGLMMs. We were looking to increase the accuracy of estimates compared to other methods proposed so far, so the weighted pairwise likelihood function was developed and used to estimate the parameters of the models. Moreover, the penalized pairwise likelihood function was used to increase the accuracy of model parameter estimation. Incorporating penalization into composite likelihood methods not only mitigates issues related to overfitting and computational complexity but also enhances model interpretability and performance in high-dimensional settings (see 18). In a simulation study, the accuracy of model parameter estimation in the pairwise, weighted, and penalized pairwise likelihood was evaluated and compared using the Mean Squared Error (MSE) and Standard Error (SE) criteria. Then, we estimated parameters based on the full likelihood function of the data with the Laplace approximation method introduced by (19), and the results were compared with three pairwise likelihood functions. Finally, penalized pairwise likelihood was used to predict and estimate two real data sets.

The article's structure is as follows: In the next section, we delve into the details of SGLMMs and explain their formulation. Then, it discusses these likelihood functions' theoretical foundations and practical implementations. We introduce the EM algorithm, detailing its steps and how it is applied within the context of SGLMMs. The following Section presents the results of various simulation studies conducted to evaluate the performance of the proposed methods. We discuss the setup, results, and implications of these simulations. Then, we apply the proposed methods to real data sets, providing a comprehensive analysis and interpretation of the results. The final Section summarizes the study's key findings and discusses their implications.

Materials and Methods

In this section, maximum likelihood estimation method for SGLMMs is investigated. First, we introduce the SGLMMs, then we discuss the composite likelihood and penalized likelihood functions used in this models.

1. Spatial Generalized Linear Mixed Models

Let $Y(\mathbf{s})$ be a discrete spatial response variable, $Z_1(\mathbf{s}), \ldots, Z_p(\mathbf{s})$ be p covariates, and $\{X(\mathbf{s}): \mathbf{s} \in \mathbb{R}^2\}$ is a latent spatial random field, where $X(\mathbf{s})$ denotes a random effect at location s. (8) have defined the SGLMMs as follows: Let $\{X(\mathbf{s}): \mathbf{s} \in \mathbb{R}^2\}$ denotes a zero mean stationary Gaussian Random Field (GRF) with spatial covariance function $\mathrm{Cov}\big(X(\mathbf{s}), X(\mathbf{s}')\big) = \sigma^2 \rho(\mathbf{s} - \mathbf{s}'; \boldsymbol{\theta})$. Here, $\rho(\cdot; \boldsymbol{\theta})$ is a positive definite function, and $\boldsymbol{\theta} \in \mathbb{R}^q$ is a vector of correlation parameters. Given $\{X(\mathbf{s}): \mathbf{s} \in \mathbb{R}^2\}$, the random field $Y(\mathbf{s})$ denotes a set of independent random variables whose distribution is characterized by conditional mean $E[Y(\mathbf{s}) \mid X(\mathbf{s})]$. For each link function g and regression parameters $\beta_1, \dots, \beta_p$, we have $g\{E[Y(\mathbf{s}) \mid X(\mathbf{s})]\} = \sum_{j=1}^{p} Z_j(\mathbf{s})\beta_j + X(\mathbf{s})$. Now the conditional distribution of $[Y(\mathbf{s}) \mid X(\mathbf{s})]$ belongs to the exponential family.

2. Weighted Pairwise Composite Likelihood Function

The composite likelihood function is obtained by multiplying a set of likelihood components, in which each likelihood component represents a subset of observations. There are two main logical reasons for using composite likelihood. The first is to reduce the computational load in modeling the joint distribution of a high-dimensional random vector. The second is robustness under uncertainty of high-dimensional distributions, as composite likelihood requires assumptions about low-dimensional marginal and conditional densities, and having the details of the simultaneous distribution would not be necessary. Inference based on the likelihood function for particularly voluminous data is associated with integrating and inverting high-dimensional matrices, which may be difficult to solve even for potent computers. These multiple integrals can be converted into the sum of integrals with lower dimensions using the weighted composite likelihood function.

Consider the n-dimensional random vector $\mathbf{y} = (y_1, \dots, y_n)$ with the probability density function $f(\mathbf{y};\boldsymbol{\theta})$ for the q-dimension parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q) \in \Theta$. Let $\{\mathcal{A}_1, \dots, \mathcal{A}_K\}$ be a set of conditional or marginal events with the likelihood functions $L_k(\boldsymbol{\theta};\mathbf{y}) \propto f(\mathbf{y} \in \mathcal{A}_k;\boldsymbol{\theta}), k = 1, \dots, K$, in which case the composite likelihood function for the parameter $\boldsymbol{\theta}$ would be defined as follows

$$L_C(\boldsymbol{\theta};\mathbf{y}) = \prod_{k=1}^{K} L_k(\boldsymbol{\theta},\mathbf{y})^{w_k},$$

where $w_1, \dots, w_K$ are non-negative weights (see (20)). An example of the composite likelihood function is the weighted pairwise composite likelihood function. (20) applied the weighted likelihood function to spatio-temporal data. They showed that weighted composite likelihood estimators are consistent and asymptotically Gaussian with a variance equal to the inverse of the Godambe information. They also showed with simulation studies that this estimation approximates maximum likelihood estimation and requires far fewer calculation overloads than the maximum likelihood and composite likelihood estimations (13) applied the weighted composite likelihood function to categorize data with a variable number of categories and examined the asymptotic relative efficiency measure for different weights. The results indicate that the composite likelihood function increases the relative asymptotic efficiency. In this research, we used a Weighted pairwise composite likelihood function for SGLMMs, and a simulation study showed that it is more efficient and has better accuracy for such models than a full likelihood function. Weighted pairwise composite likelihood function can be represented as follows

$$L_W(\boldsymbol{\theta};\mathbf{y}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} f(y_i, y_j;\boldsymbol{\theta})^{w_{ij}}.$$

Therefore, the logarithm of the pairwise weighted likelihood function can be restructured as

$$\ell_W(\boldsymbol{\theta};\mathbf{y}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \log f(y_i, y_j;\boldsymbol{\theta}). \qquad (1)$$

For the SGLMMs

$$w_{ij} = \begin{cases} 1 & \|s_i - s_j\| \le d_s \\ 0 & \text{o.w.} \end{cases}$$

where d_s represents the distance between spatial points. In this paper, we introduced new values for d_s discussed in detail in the simulation section .As such, the weighted pairwise composite maximum likelihood estimation of $\boldsymbol{\theta}$ that maximizes the function in Equation (1) under regularity conditions is equal to the unique solution of the equation $u(\boldsymbol{\theta};\mathbf{y}) = \nabla_{\boldsymbol{\theta}} \ell_w = 0$. The pairwise likelihood function for the SGLMMs is as follows

$$L_W(\boldsymbol{\eta};\mathbf{y}) = \prod_{(i,j)\in\chi} L(\boldsymbol{\eta} \mid y_i, y_j) \propto$$

$$\prod_{(i,j)\in\chi} \iint f(y_i \mid x_i) f(y_j \mid x_j) f(x_i, x_j \mid \boldsymbol{\eta}) dx_i dx_j,$$

where $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ is the vector of model parameters, and χ is the pairwise neighborhood set of (y_i, y_j).

3. Penalized Pairwise Likelihood Function

The maximum likelihood method in parameter estimation in various problems is sometimes plagued with overfitting, low accuracy, or high variance of the estimators. Penalization of the likelihood function is a solution to alleviate the behaviors of estimators mentioned above using the usual maximum likelihood method, called the penalized maximum likelihood method (see (21)). Using the approximation of the likelihood function instead of the usual likelihood due to its approximate nature can reduce the accuracy of the estimators. The composite likelihood function can be employed to estimate the parameters for extensive spatial data, where obtaining the likelihood function is analytically extremely difficult. The composite likelihood function approximates the likelihood function of observation, and hence, estimators based on this function may have low accuracy. The penalized likelihood function can be, as such, used to improve the estimation accuracy in the SGLMMs, in which a penalty function is embedded in the logarithm of the likelihood of observations. Penalized likelihood methods provide a robust framework for estimating correlation parameters in spatial models by enhancing stability, reducing variability, and directly incorporating spatial dependence. These advantages make them particularly effective compared to traditional maximum likelihood estimation techniques. In spatial models, accounting for spatial autocorrelation is vital. Penalized likelihood methods allow for the direct integration of spatial dependence into the likelihood function, leading to more efficient estimates than methods that only indirectly consider spatial relationships. This direct approach enhances model fit and improves inference regarding correlation parameters (see 22). For the SGLMs, the penalized pairwise likelihood function is as follows

$$\ell(\boldsymbol{\eta};\mathbf{y}) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \log f\big(y_i, y_j; \boldsymbol{\eta}\big) - \lambda J(\boldsymbol{\eta}),$$

where λ is the smoothing parameter and $J(\boldsymbol{\eta})$ is the penalty function, which can be selected using various methods. For example, (23) defined the penalty function as $\lambda \frac{1}{2}\log|I(\boldsymbol{\eta})|$ for the exponential class of functions, where $I(\boldsymbol{\eta})$ is the Jeffreys prior of $\boldsymbol{\eta}$. (24) presented the Lasso penalty as $J_\lambda(\boldsymbol{\eta}) = \lambda\boldsymbol{\eta}$ for estimation in linear models. Here, the Lasso and Green (see (25)) function $2\boldsymbol{\eta}\boldsymbol{\eta}^T$(developed for the penalization of SGLMMs likelihood function and based on MSE criterion showed that have better results compared to other models). When using the Lasso penalty, the regularization parameter lambda controls the degree of shrinkage applied to the parameters. The optimal value of lambda can be chosen using cross-validation, which involves splitting the data into training and validation sets and selecting the value of lambda that minimizes the mean squared error on the validation set.

4. Expectation Maximization Algorithm

(15) presented the pairwise Expectation Maximization (EM) algorithm for maximizing the likelihood. Based on this algorithm, the value of $\boldsymbol{\eta}^{(0)}$ should be selected in such a way that $PL\big(\boldsymbol{\eta}^{(0)};\mathbf{y}\big) > 0$. Set $m = 0$. In the E step, the value of conditional expectation is selected as follows

$$Q(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(m)}) = \sum_{i,j\in\chi} \iint \log\{f(x_i, x_j, y_i, y_j; \boldsymbol{\eta})\} f(x_i, x_j \mid y_i, y_j; \boldsymbol{\eta}^{(m)}) dx_i dx_j \,. \quad (2)$$

Then, in the M step, the value $\boldsymbol{\eta}^{(m+1)}$ is selected such that

$$Q\big(\boldsymbol{\eta}^{(m+1)} \mid \boldsymbol{\eta}^{(m)}\big) \geq Q\big(\boldsymbol{\eta}^{(m)} \mid \boldsymbol{\eta}^{(m)}\big).$$

Then, by setting $m = m + 1$, the algorithm goes to the next iteration and continues until convergence. If the conditional expectation of step M cannot be expressed in a closed form, it should be approximated numerically. (15) defined an approximate EM algorithm as replacing Q with an approximate $\hat{Q}$ value. In the approximate EM algorithm, the initial value of $\boldsymbol{\eta}^{(0)}$ must be chosen such that $PL\big(\boldsymbol{\eta}^{(0)};\mathbf{y}\big) > 0$ and $m = 0$. The approximate EM algorithm repeats the following steps until convergence.

Algorithm 1:

Step 1: Approximate E step: the conditional expectation value in Equation (2) is approximated with $\hat{Q}\big(\boldsymbol{\eta};\boldsymbol{\eta}^{(m)}\big)$.

Step 2: Generalized M step: the value of $\boldsymbol{\eta}^{(m+1)}$ is chosen such that $\hat{Q}\big(\boldsymbol{\eta}^{(m+1)} \mid \boldsymbol{\eta}^{(m)}\big) \geq \hat{Q}\big(\boldsymbol{\eta}^{(m)} \mid \boldsymbol{\eta}^{(m)}\big)$.

Step 3: Reiterate steps 1 to 2 of the algorithm until convergence.

The convergence criterion in this algorithm is $\max_i\big|\boldsymbol{\eta}^{(m+1)} - \boldsymbol{\eta}^{(m)}\big|/\big|\boldsymbol{\eta}^{(m)}\big| < 0.0005$. (3,26) used Monte Carlo integration in the expectation step for SGLMMs. In pairwise likelihood maximization, the expectation step involves the sum of double integrals. Results indicate that the Gauss Hermite quadrature is more efficient than Monte Carlo integration. Therefore, (15) presented the Quadrature Pairwise EM (QPEM) algorithm for SGLMMs and showed that its speed is faster than the MCEMG algorithm. In the E step of this algorithm, Equation (2) is approximated using the Gauss-Hermite quadrature. Moreover, our results showed that the QPEM algorithm is faster than the ML method with place approximation.

Gauss-Hermite quadrature is developed to approximate integrals that involve distributions close to the normal distribution. In this approximation, the sub-integral function $f(t)$ is divided into two parts: the Gaussian part (the envelope) and the remaining part, the latter of which becomes $\tilde{f}(\mathbf{t}) = e^{\|t\|^2/2} f(\mathbf{t})$ after changing the variables. Gauss-Hermite quadrature decreases the integral of a function into a weighted sum of sub-integral functions computed at K nodes. Adaptive Gauss-Hermite quadrature uses Gaussian approximation $f\big(x_i, x_j \mid y_i, y_j; \boldsymbol{\eta}^m\big)$ to good accuracy with low values of K but requires the mode and the second derivative of those mentioned above. (15) considered the distribution of $f\big(x_i, x_j\big)$ as the envelope function and the likelihoods of $f(y_i \mid x_i)$ as the remaining part to approximate the likelihood function of the SGLMM. To solve the double integral in Equation (2), the vector $\big(x_i, x_j\big)^T$ is transformed into the standardized components of $\big(v_i, v_j\big)$, where $v_i = \frac{x_i}{\sigma}$ and

$\nu_j = \frac{x_j - \rho_{ij} x_i}{\sigma\sqrt{1-\rho_{ij}^2}}$ and $\rho_{ij} = \rho(s_i - s_j; \boldsymbol{\theta})$. After changing the variables, by solving for $x_i(\nu_i)$ and $x_j(\nu_i, \nu_j)$ the approximate value of the function $Q(\boldsymbol{\eta}; \boldsymbol{\eta}^{(m)})$ is as follows

$$\hat{Q}(\boldsymbol{\eta}; \boldsymbol{\eta}^{(m)}) = \sum_{i,j=1}^{n} \sum_{k_1,k_2=1}^{K} \log f\left(x_i(h(k_1)), x_j(h(k_1), h(k_2)), y_i, y_j; \boldsymbol{\eta}\right) w_{ij}(k_1, k_2; \boldsymbol{\eta}^{(m)})$$

where

$$w_{ij}(k_1, k_2; \boldsymbol{\eta}^{(m)}) = \frac{f(y_i | x_i(h(k_1); \boldsymbol{\eta}^{(m)}) f(y_j | x_j(h(k_1), h(k_2)); \boldsymbol{\eta}^{(m)}) \ell(k_1)\ell(k_2)}{\sum_{k_1,k_2} f(y_i | x_i(h(k_1); \boldsymbol{\eta}^{(m)}) f(y_j | x_j(h(k_1), h(k_2)); \boldsymbol{\eta}^{(m)}) \ell(k_1)\ell(k_2)}.$$

Here, $h(k)$ s and $\ell(k)$ s denote, respectively, the nodes and the weights.

5. Simulation Study

In this section, we use the QPEM algorithm to estimate the parameters of spatial generalized linear mixed models through simulation. In this study, pairwise likelihood, weighted pairwise likelihood, and penalized pairwise likelihood functions are used, and the accuracy of each of these functions is checked through the MSE and SE criteria. A neighborhood with a radius of 4 is used for each observation. If we want to use all 48 neighbors for each point in the model, there are $48n = 10800$ pairs, which is far less than all possible ordered pairs, i.e., $n(n-1)/2 = 25200$.

The QPEM algorithm with $M = 4 \times 4$ nodes of Gauss-Hermite quadrature was used to estimate the parameters. Based on the method proposed by (15), 15 pairs from a radius of four neighbors are randomly selected for each observation, as shown in Figure 1. This method would, in turn, reduce the number of observations to $15n = 3375$. The parameters were estimated using pairwise likelihood and weighted pairwise likelihood function, and the accuracy was compared through the MSE and SE criteria. Results were obtained for 100 datasets. Different values of d_s were inputted in the Equation (1) to obtain the optimal value of the weight function. Consider a matrix of Euclidean distance between spatial points and sort them from smallest to largest; then, d_s 's are a function of the quantiles of these values. For this research, we consider $d_s = q(0.4), q(0.6), q(0.8), q(0.9)$. The results of the simulations are presented in Table 1. The current research also examined the effect of the number of Gauss-Hermite quadrature nodes and the corresponding neighborhood's radius. (15) revealed that reducing the number of nodes to 3×3 leads to lower performance, while increasing the number of nodes to 5×5 significantly does not improve the results. Increasing the neighborhood radius, and thus increasing the number of pairs, did not have much effect on the results of the simulations.

The present research also studied the effect of weighting the pairwise likelihood function on the accuracy of the SGLMM parameter estimation. The result was compared with estimation based on the full likelihood function of the data, and an ML algorithm with Laplace approximation was used for evaluation. The examined model was SGLM with Poisson response and logarithm link function. The data is generated from a 25×25 regular grid with nodes $\{(s_1, s_2): s_1, s_2 = 0, 0.04, \dots, 1\}$. To generate the spatial latent variables, X, the normal distribution of $N(0, \Sigma_\theta)$, the isotropic exponential covariance function of $C(h; \boldsymbol{\theta}) = \sigma^2 \exp(-3h/\phi), h > 0, \boldsymbol{\theta} = (\sigma^2, \phi)$ and values $\sigma^2 = 1.5, \phi = 6$ and $\boldsymbol{\beta} = (\beta_0, \beta_1) = (1, 0.5)$ are considered. The explanatory variable in each position $s = (s_1, s_2)$ is considered as $z_s = s_1$. The response variable, Y_s, is also generated by conditioning on spatial latent variables from the distribution $Y_s \sim \text{Poisson}\left(n, \exp(\beta_0 + \beta_1 z_s + x_{s,})\right)$ where $n = 25 \times 25$ is the number of samples. The graphs of these data are shown in Figure 2. Logarithm

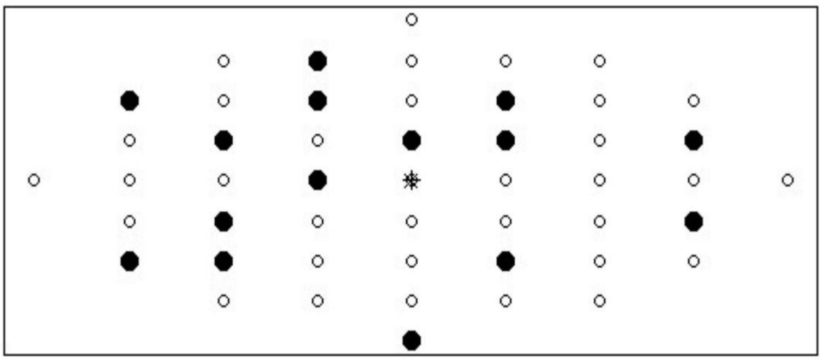

Figure 1. Sampling pairs within a neighborhood of radius 4. Here, $*$ is the observation location, and the filled circles are 15 neighbors sampled randomly without replacement. The contributing pairs consist of $*$ and each of the 15 sampled neighbors.

Table 1. Estimation of the SGLMM based on full, pairwise and weighted pairwise likelihoods.

Likelihood	Weight	Parameter	Estimate	MSE	SE
Full	1	β_0	0.892	0.0930	0.0401
		β_1	0.485	0.0852	0.0331
		σ^2	1.340	0.1245	0.0185
		ϕ	6.0820	4.0980	0.3633
Pairwise	1	β_0	0.842	0.1052	0.0380
		β_1	0.565	0.0934	0.0305
		σ^2	1.230	0.1042	0.0177
		ϕ	5.640	3.4111	0.3820
Weighted	$q(0.4)$	β_0	0.901	0.8563	0.0374
		β_1	0.327	0.5946	0.0307
		σ^2	1.196	0.9116	0.0176
		ϕ	6.105	14.054	0.3756
	$q(0.6)$	β_0	0.807	0.1702	0.0336
		β_1	0.474	0.1170	0.0312
		σ^2	1.470	0.1042	0.0187
		ϕ	5.830	6.8114	0.3756
	$q(0.8)$	β_0	0.883	0.1180	0.0377
		β_1	0.469	0.1067	0.0300
		σ^2	1.554	0.1111	0.0160
		ϕ	5.907	3.5143	0.3677
	$q(0.9)$	β_0	0.922	0.1120	0.0327
		β_1	0.519	0.0900	0.0243
		σ^2	1.511	0.1137	0.0166
		ϕ	6.041	3.4819	0.2849

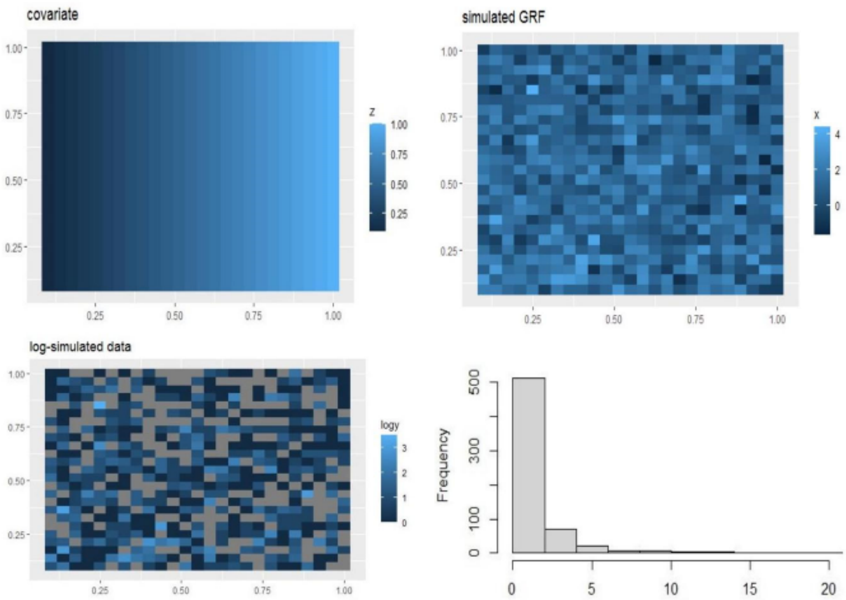

Figure 2. Realization from the spatial Poisson model. From top-left to bottom-right: covariate $z(s) = s_1$, simulated mixed effects field $x(s)$, (log) simulated data, histogram of the data.

transformation was used to maximize the parameters of random effects to avoid singularity. That is, the parameters were $\sigma^{2,}$ and ϕ inputted in the model as $\log \sigma^2$ and $\log \phi$. The relation mentioned in Section 5 was used as the criterion of convergence.

The regression parameters β_0 and β_1 are estimated without considering the random variable and using a simple GLMM to obtain the initial values. Then, using the link function, the observed values are converted, and the remaining values are estimated as $\hat{r}(s_i) = g(y_i) -$

Table 2. Estimation of SGLMMs by weighted pairwise and penalized pairwise likelihoods.

Likelihood	Parameter	Estimate	MSE	SE
Weighted pairwise	σ^2	1.230	0.1042	0.0177
	ϕ	5.640	3.411	0.3820
Penalized with Lasso	σ^2	1.506	0.071	0.0169
	ϕ	1.895	1.925	0.1504
Penalized with Green	σ^2	1.489	0.093	0.0169
	ϕ	1.895	2.252	0.1504

$x_i\hat{\boldsymbol{\beta}}, i = 1, \ldots, n$. The appropriate covariance function is fitted to the experimental variogram of the residuals, and the estimation of σ^2 and ϕ are considered the initial values.

The results from Table 1 indicated that the weighted pairwise composite likelihood outperforms the pairwise composite likelihood function when d_s is equal to $q(0.8)$ or $q(0.9)$. The lower SE criterion values of weighted pairwise likelihood estimation show this function is more efficient than pairwise and full likelihood function-based methods. For example, the SE values of β_0, β_1, σ, and ϕ with the full likelihood function are 0.0401, 0.0331, 0.185, and 0.3633, respectively. In weighted pairwise likelihood, when d_s is equal to $q(0.9)$, these values are 0.0327, 0.0243, 0.0166, and 0.02849, respectively. Moreover, the output values for these two inputs were similar, implying that excluding the outlying pairs from the likelihood function would not result in substantial information loss, and the model parameters can be estimated with acceptable accuracy even with fewer pairs. Results show that for β_0 and β_1, the MSE of the full likelihood function is slightly smaller than pairwise and weighted pairwise likelihood functions, and for σ and ϕ, the results are the opposite. The R program was used for this research. The computing time for QPEM on a typical data set using a 2.40 GHz computer with 4 GB RAM was 250 seconds. The ML estimates with Laplace approximation for the same data set were computed in 1200 seconds.

6. *Penalized Composite Likelihood Function*

This section examines the effect of penalizing the pairwise likelihood function on the accuracy of SGLMM parameter estimation. For this purpose, two penalization functions, Lasso and $2\boldsymbol{\eta}\boldsymbol{\eta}^T$ (see (25)), were used where $\boldsymbol{\eta} = (\sigma^2, \phi)$ is the vector of model parameters. We used the cross-validation method mentioned in Section 3.1 to select optimal λ in the Lasso penalty, and for our simulated data, $\lambda = 2$ was selected. The QPEM algorithm estimates the parameters, maximizing $Q(\boldsymbol{\eta} \mid \boldsymbol{\eta}^m) - \lambda J(\boldsymbol{\eta})$ in the M step. In order to examine the effect of penalizing the pairwise likelihood function, the estimation accuracy based on the MSE criterion has been compared for two functions: pairwise likelihood and penalized pairwise likelihood. Findings from Table 2 indicate that MSE and SE criteria in estimating the model parameters using the penalized pairwise likelihood function for both penalty functions are lower than those of the pairwise likelihood function. The penalized pairwise likelihood function can significantly increase parameter estimation accuracy. Also, the results indicate that the lasso penalty function outperforms that of (25).

To compare two penalty functions in prediction accuracy, we compute prediction error measures by MSPE at $(0.35, .55)$ location for 100 data sets. This measure for the Lasso and Green penalty was 1.5134 and 1.5245, respectively. So, two penalty functions have the same operation.

Results

Real Binomial and count datasets were employed to check the performance of composite likelihood functions based on the QPEM algorithm: Rhizoctonia root rot data and Swedish weed data.

1. *Binomial Example*

Rhizoctonia root and stem rot is a disease that affects plant roots and prevents water and nutrient absorption. This dataset contains counts of Rhizoctonia root rot disease in barley collected at 100 sampling sites at Cunningham Farm in the northwestern United States. For each sampling site, 15 plants were pulled out from the ground for examination. This dataset contains five columns and 100 rows. The first two columns contain the coordinates of the sampling sites, the third column contains the total number of crown roots in the extracted plants, the fourth column represents the total number of infected roots in the plants, and the last column contains the yield of barley in the sampling sites, see Figure 3. Using the findings from (16,19,27), a Binomial SGLMM with a logit link function, a constant mean, β_0, and an exponential correlation function of

$$\rho(h; \boldsymbol{\theta}) = (1 - \tau^2)\sigma^2 \exp\left(-\frac{h}{\phi}\right), \boldsymbol{\theta} = (\tau, \sigma^2, \phi),$$

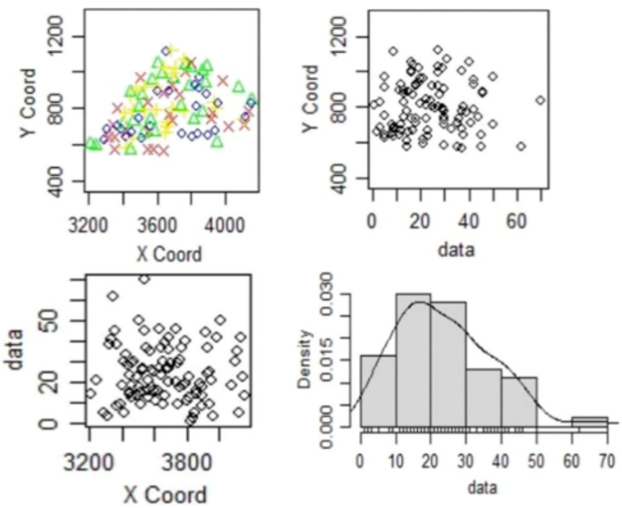

Figure 3. Plot of Rhizoctonia data set

Table 3. Estimation by weighted pairwise and penalized pairwise likelihoods

Parameter	Method: Lasso penalty	Method: Green penalty	Weighted pairwise likelihood
β_0	-1.69	-1.75	-1.73
σ^2	0.15	0.09	0.18
ϕ	149.08	152.65	148.4
τ^2	0.58	0.61	0.46

was used for the spatial random effect that τ^2 is the nugget effect.

Furthermore, the penalized composite likelihood function with the Lasso and Green penalty functions was implemented to estimate the SGLMM. We used the cross-validation method mentioned in Section 3.1 for selecting the optimal λ in the Lasso penalty. For our data, $\lambda = (0.2, 0.5, 0.7, 1, 1.5, 1.75, 2, 4)$ was examined, and $\lambda = 2$ was selected. The method mentioned in the simulation study was also used to obtain the initial values shown in Table 3. The results were consistent with those obtained by (19). We used the penalized likelihood function with the Lasso penalty to predict new locations. Based on estimated parameters and latent variables on 100 observations, we predict the responses for 36 new locations, and the results are shown in Figure 4. We use the minimum mean-squared error (MMSE) prediction method proposed by (27) with the Metropolis-Hastings algorithm.

Results show that prediction values are similar to

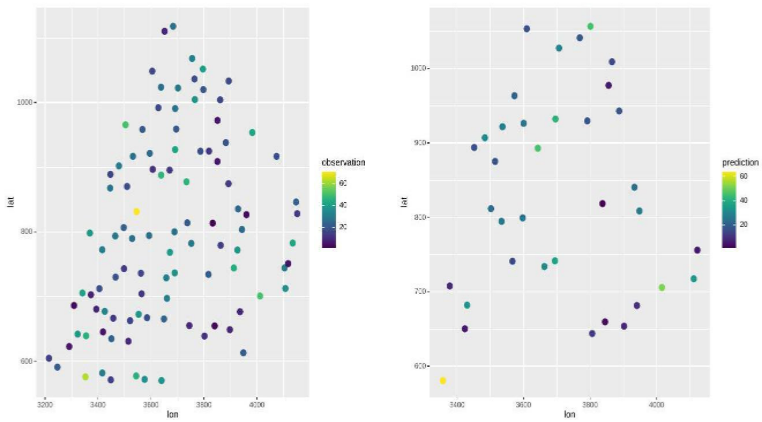

Figure 4. Observed proportion of infected roots and prediction of Binomial probability for Rhizoctonia example

observation values. The Cross-Validation MSE for predicting new data based on the MMSE method was 0.30548.

2. Poisson Example

The weed dataset was first published by (28) and was modeled using a spatial Poisson-lognormal model. The data pertained to the number of weeds of non-crop plants measured in more than 100 frames of 0.5×0.7 meter dimensions at Gertrup Farm in southern Sweden. This dataset contains four columns and 100 rows. The first two columns are the coordinates of the sampling locations. The third column includes the number of observations, and the last column estimates the number of photos obtained. The areas of the frames on which the number of weeds is collected are all equal. Therefore, they can be considered equal to 1, see Figure 5. The Poisson SGLMM with a lognormal link function fits this data set. The covariance function of latent variables is exponential with a nugget effect, and the experimental variogram of data in Figure 6 shows the suitability of this model. There is no explanatory variable for this data, but we consider an intercept parameter for the model. Penalized pairwise likelihood was used to estimate parameters. For Lasso, the penalty $\lambda = 0.7$ was optimal. The results of the parameter estimation for the weed data set are in Table 4.

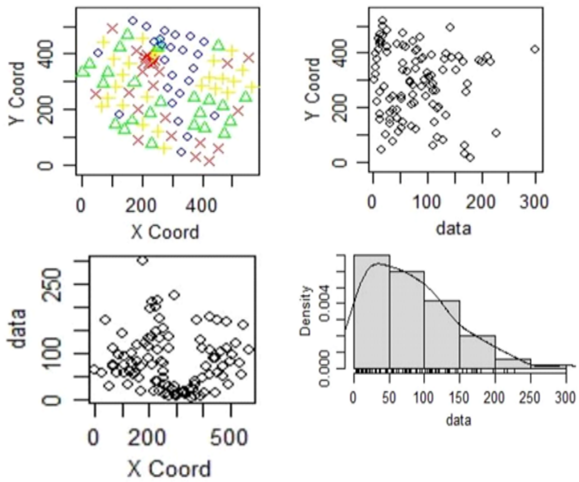

Figure 5. Plot of weed data set

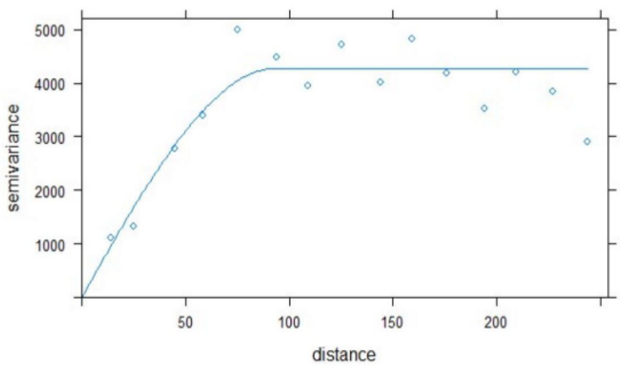

Figure 6. Experimental variogram of the weed data set

Table 4. Parameter estimation by weighted pairwise and penalized pairwise likelihood.

	Method		
Parameter	**Lasso penalty**	**Green penalty**	**Weighted pairwise likelihood**
β_0	2.14	2.07	1.98
σ^2	0.18	0.35	0.26
ϕ	56.6	53.7	54.9
τ^2	15.6	17.2	18.8

Discussion

This research used full likelihood, pairwise likelihood, weighted pairwise likelihood, and penalized pairwise likelihood functions for spatial generalized linear mixed models. The QPEM algorithm was used to maximize pairwise likelihood functions, and for the full likelihood function method, the ML algorithm with Laplace approximation was used. During a simulation study, the accuracy of model parameter estimation based on these functions was evaluated and compared using the mean squared error (MSE) parameter. The simulation results showed that the weighted pairwise likelihood function and the penalized pairwise likelihood function outperformed the pairwise likelihood function in estimating the parameters of the SGLMM.

Also, we showed that the full likelihood function has better results for regression parameters, and pairwise and weighted pairwise likelihood functions have better applications for correlation parameters. The findings further established that the penalized pairwise likelihood function was more accurate than others in estimating the correlation parameters σ^2 and ϕ. Also, the Lasso penalty function exhibited better results than the Green penalty function among penalized composite likelihood functions. The results were analyzed and compared with those of other studies for two real data sets with Binomial and Poisson distributions. Overall, the results of the current study showed that the proposed functions outperformed other functions in estimating the parameters of the SGLMM.

Acknowledgment

The authors thank the editorial team and the anonymous reviewers for their comments, helpful suggestions, and encouragement, which helped improve the final version of this paper. Receiving support from the Center of Excellence in Analysis of Spatio-Temporal Correlated Data at Tarbiat Modares University is acknowledged.

References

1. Nelder, JA and Wedderburn, RWM. Generalized Linear Models: Journal of the Royal Statistical Association, Series A. 1972;135: 370-384.
2. McCullagh, CE and Nelder, JA. Generalized Linear Models: Chapman and Hall, London 1989.
3. McCulloch, C. Maximum Likelihood Algorithms for Generalized Linear Mixed Models: Journal of the American Statistical Association. 1997; 92: 162-170.
4. Mohammadzadeh, N. and Hosseini, F. Maximum-Likelihood Estimation for Spatial GLMMs: Procedia Environmental Sciences. 2011; 3: 63-68.
5. Baghishani, H and Mohammadzadeh, M. A Data Cloning Algorithm for Computing Maximum Likelihood Estimates in Spatial Generalized Linear Mixed Models: Computational Statistics and Data Analysis. 2011; 55: 1748-1759.
6. Torabi, M. Likelihood Inference for Spatial Generalized Linear Mixed Models: Communications in Statistics-Simulation and Computation. 2015; 44: 1692-1701.
7. Rue, H and Martino, S. Approximate Bayesian Inference for Hierarchical Gaussian Markov Random Field Models: Journal of Statistical Planning and Inference. 2007; 137: 3177-3192.
8. Eidsvik, J, Finley, AO, Banerjee, S and Rue, H. Approximate Bayesian Inference for Large Spatial Datasets Using Predictive Process Models: Computational Statistics and Data Analysis. 2012; 56: 1362-1380.
9. McCullagh, P. Quasi-Likelihood Functions: Annals of Statistics. 1983; 11: 59-67.
10. Wedderburn, R. Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method: Biometrica. 1974; 61: 973-981.
11. Heagerty, P. and Lele, S. A Composite Likelihood Approach to Binary Spatial Data: Journal of American Statistical Association. 1998; 93: 1099-1111.
12. Bevilacqua, M and Gaetan, C. Comparing composite likelihood methods based on pairs for spatial Gaussian random fields: Statistics and Computing. 2015; 25: 877-892.
13. Mazo, G, Karlis, D and Rau, A. A randomized pairwise likelihood method for complex statistical inferences: Journal of the American Statistical Association. 2024; 119(547): 2317-2327.
14. Cox, D and Reid, NA. Note on Pseudo Likelihood Constructed from Marginal Densities: Biometrika. 2004; 91: 729-737.
15. Varin, C, Host, G. and Skare, O. Pairwise Likelihood Inference in Spatial Generalized Linear Mixed Models: Computational Statistics and Data Analysis. 2005; 49: 1173-1191.
16. Bevilacqua, M, Mateu, J, Porcu, E, Zhang, H and Zini, A. Weighted Composite Likelihood-Based Tests for Space-Time Separability of Covariance Functions: Statistics and Computing. 2010; 20: 283-293.
17. Joe, H and Lee, Y. On Weighting of Bivariate Margins in Pairwise Likelihood: Journal of Multivariate Analysis. 2009; 100: 670-685.
18. Li, Q, Sun, X, Wang, N and Gao, X. Penalized composite likelihood for colored graphical Gaussian models: Statistical Analysis and Data Mining: The ASA Data Science Journal. 2021;14(4): 366-378.
19. Bonat, WH and Ribeiro, PJ. Practical Likelihood Analysis for Spatial Generalized Linear Mixed Models: Environmetrics. 2016; 27: 83-89.
20. Lindsay, B. Composite Likelihood Methods: Contemporary Mathematics. 1988; 80: 220-239.
21. Azzalini, A and Arellano-Valle, RB. Maximum Penalized Likelihood Estimation for Skew-Normal and Skew-t Distributions: Journal of Statistical Planning and Inference. 2013; 143:419-433.
22. Smith, B, Wang, S, Wong, A and Zhou, X. A penalized likelihood approach to parameter estimation with integral reliability constraints: Entropy. 2015; 17(6): 4040-4063.

23. Firth, D. Bias Reduction of Maximum Likelihood Estimates: Biometrika. 1993; 80: 27-38.
24. Tibshirani, R. Regression Shrinkage and Selection via the Lasso: Journal of the Royal Statistical Society, Series B(Methodological). 1996; 267-288.
25. Green, PJ. On Use of the EM for Penalized Likelihood Estimation: Journal of the Royal Statistical Society, Series B (Methodological). 1990; 443-452.
26. Booth, JG and Hobert, JP. Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm: Royal Statistical Society, Series B. 1999; 61: 265-285.
27. Zhang, H. On Estimation and Prediction for Spatial Generalized Linear Mixed Models, Biometrics. 2002; 58: 129-136.
28. Guillot, G, Loren, N, Rudemo, M. Spatial Prediction of Weed Intensities from Exact Count Data and Image-Based Estimates, Journal of the Royal Statistical Society. 2009; 58: 525-542.