Research Article Environmetrics

Received: 18 May 2015, Revised: 11 November 2015, Accepted: 12 November 2015, Published online in Wiley Online Library: 20 December 2015

(wileyonlinelibrary.com) DOI: 10.1002/env.2375

Practical likelihood analysis for spatial generalized linear mixed models

Wagner Hugo Bonat[a,b]* and Paulo Justiniano Ribeiro Jr[b]

We investigate an algorithm for maximum likelihood estimation of spatial generalized linear mixed models based on the Laplace approximation. We compare our algorithm with a set of alternative approaches for two datasets from the literature. The *Rhizoctonia root rot* and the *Rongelap* are, respectively, examples of binomial and count datasets modeled by spatial generalized linear mixed models. Our results show that the Laplace approximation provides similar estimates to Markov Chain Monte Carlo likelihood, Monte Carlo expectation maximization, and modified Laplace approximation. Some advantages of Laplace approximation include the computation of the maximized log-likelihood value, which can be used for model selection and tests, and the possibility to obtain realistic confidence intervals for model parameters based on profile likelihoods. The Laplace approximation also avoids the tuning of algorithms and convergence analysis, commonly required by simulation-based methods. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: spatial data; likelihood inference; Laplace approximation; random effects

1. INTRODUCTION

Evaluating the likelihood function for spatial generalized linear mixed models (SGLMMs) (Diggle *et al.*, 1998) requires the solution of a high-dimensional integral for the spatially correlated random effects. For the standard setup, where the observed data are regarded as a single partial realization of the process within an area, the dimension of the integral equals the number of observations. Solving this intractable integral is the main computational problem for fitting such models.

In the literature, SGLMMs are frequently fitted by using Monte Carlo Markov Chain (MCMC) methods under the Bayesian paradigm (Møller, 2003; Banerjee *et al.*, 2004; Diggle and Ribeiro Jr, 2007). Diggle *et al.* (1998) presented Bayesian inference for SGLMMs. A more efficient Langevin-Hastings MCMC algorithm was proposed by Christensen *et al.* (2000), and a robust version was introduced by Christensen *et al.* (2006). Christensen and Waagepetersen (2002) discussed the general framework for Bayesian prediction in the particular context of count data. In practice, MCMC methods for SGLMMs have various problems in terms of convergence and computational time. Moreover, the implementation itself can be problematic, especially for end-users who might not be expert in programming.

Recently, some authors have turned their attention to approximation methods within the Bayesian framework. Rue and Martino (2007) proposed approximate Bayesian inference in the context of Gaussian Markov random fields. Eidsvik *et al.* (2012) presented a similar approach for SGLMMs, and Hosseini *et al.* (2011) proposed approximate Bayesian inference for SGLMMs using skew Gaussian latent variables. Both approaches use the Laplace approximation (Tierney and Kadane, 1986) to deal with the high dimensional integral. Rue *et al.* (2009) consolidated the application of the Laplace approximation and showed that the approximation is quite accurate for practical data analysis.

Likelihood analysis for SGLMMs was proposed by Christensen (2004) but still relied on MCMC algorithms. Both MCMC likelihood and Bayesian algorithms are implemented for Poisson and binomial models in the `R` (R Core Team, 2013) package `geoRglm` (Christensen and Ribeiro Jr, 2002) and described in detail by Diggle and Ribeiro Jr (2007). Alternative methods were proposed by Zhang (2002) adopting a spatial expectation maximization algorithm. Recently, Baghishani and Mohammadzadeh (2011) and Torabi (2015) proposed a different approach based on the data clone algorithm (Lele *et al.*, 2010).

These algorithms provide likelihood analysis for SGLMMs, but as MCMC methods, they still carry the computational inconvenience and burden of being simulation based. In many cases, simulation-based methods do not provide the value of the maximized log-likelihood, and it is not clear how to obtain profile likelihoods, two useful tools in the context of likelihood inference for SGLMMs. Alternatively, Heagerty and Lele (1998) and Varin *et al.* (2005) proposed the use of pairwise likelihood for estimation of SGLMMs. The approach is computationally convenient, because it only requires the solution of two dimensional integrals, although the problem is the need to choose the pairs of observations. There are various proposals, but it is difficult to assess the impact of different choices on the final result.

* *Correspondence to: Wagner Hugo Bonat, Department of Statistics, Paraná Federal University, Curitiba, Brazil. E-mail: wbonat@ufpr.br*

a *Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark*

b *Department of Statistics, Paraná Federal University, Curitiba, Brazil*

Likelihood analysis for SGLMMs has been the subject of a great deal of research over the past decades. Early references include the work of Breslow and Clayton (1993) on penalized quasi-likelihood and marginal quasi-likelihood. It is a well-known result in the literature that Breslow and Clayton's method yields a non-negligible bias when applied to binary clustered data (Breslow and Lin, 1995), (Lin and Breslow, 1996). Evangelou *et al.* (2011) present a modified Laplace approximation and apply this method for estimation and prediction of SGLMMs obtaining better performance than penalized quasi-likelihood and MCMC likelihood. Millar (2011) argues that, in the context of generalized linear mixed models, penalized quasi-likelihood has much in common with Laplace approximation . The author also points out that the weaknesses of this ad hoc likelihood method are the additional approximations needed to avoid the optimisation and second derivative calculations required by the Laplace approximation.

The main goal of this paper is to investigate and implement a general algorithm for maximum likelihood estimation of SGLMM based on the Laplace approximation (Tierney and Kadane, 1986). Under this approach, the computational burden of integration when evaluating the likelihood is replaced by a maximization to find optimized values for the random effects and the evaluation of the Hessian around such a maximum. In other words, we have an optimisation algorithm for the random effects within each step of the optimisation for the model parameters. We describe the algorithm and present analysis of the Rhizoctonia root rot data, comparing results with the previous analysis by Zhang (2002) and Evangelou *et al.* (2011). Using the Rongelap data, which motivated Diggle *et al.* (1998), we compare our approach with MCMC likelihood implemented in the package `geoRglm` , and we extend the analysis to consider the negative binomial distribution. Our procedure can be immediately extended to other GLM families and beta distribution models. In the supporting information, we provide `R` code to fit binomial, Poisson, negative binomial, gamma, and beta SGLMMs.

The Laplace approximation is frequently applied in the context of longitudinal data analysis (Molenberghs and Verbeke, 2005); however, it is seldom applied in the context of SGLMMs. The seminal paper on model-based geostatistics (Diggle *et al.*, 1998) deals with SGLMMs in the context of non-Gaussian data, but comparison, or mention, of the Laplace approximation was not the object of this paper nor is it found in the subsequent related literature. The same applies for key books on model-based geostatistics, such as Diggle and Ribeiro Jr (2007). Also, there does not appear to be an `R` package able to fit SGLMMs using the Laplace approximation for standard likelihood analysis, nor are there any applications or references to Laplace approximation in the context of likelihood analysis for SGLMMs in the recent special issues of the Journal Statistical Software (Vol 63(1)) on spatial data analysis. The package `geoRglm` is probably the earliest referenced package in `R` to fit SGLMMs, and it uses Monte Carlo-based likelihood methods. Similar implementations can be found in the package `geoCount` (Jing and De Oliveira, 2015). Only in (Evangelou *et al.*, 2011) is a Laplace approximation used in a similar fashion.

2. SPATIAL GENERALIZED LINEAR MIXED MODELS

The SGLMM assumes that observations $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ at locations $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ are a partial realization of a processes described by the hierarchical model as follows:

$$\begin{aligned} \boldsymbol{Y}(\boldsymbol{x})|\boldsymbol{S}(\boldsymbol{x}) &\sim f(\cdot; \boldsymbol{\mu}(\boldsymbol{x}), \psi) \\ g(\boldsymbol{\mu}(\boldsymbol{x})) &= \boldsymbol{D}\boldsymbol{\beta} + \boldsymbol{S}(\boldsymbol{x}) \end{aligned}$$

It is assumed that the components of $\boldsymbol{Y}(\boldsymbol{x})$ are conditionally independent given a Gaussian spatial process $\boldsymbol{S}(\boldsymbol{x}) = \sigma \boldsymbol{U}(\boldsymbol{x}; \phi) + \tau \boldsymbol{Z}$ and distributed as $f(\cdot; \boldsymbol{\mu}(\boldsymbol{x}), \psi)$. The linear predictor is linked to the mean by a link function g and consists of the sum of fixed effects $\boldsymbol{D}\boldsymbol{\beta}$, spatially correlated random effects $\sigma \boldsymbol{U}(\boldsymbol{x}; \phi)$ and uncorrelated random effects $\tau \boldsymbol{Z} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{I})$. The parameter τ^2 is sometimes called the nugget effect. The $n \times p$ design matrix $\boldsymbol{D}$ contains values of p covariates and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters. Finally, ψ is a nuisance parameter that depends on the assumed form for f.

In the geostatistical context, $\boldsymbol{U}(\boldsymbol{x}; \phi)$ is a unit variance Gaussian random field with correlation function $\rho(u, \phi)$, where $\rho \in \Re^d$ is a valid correlation function parametrized by ϕ with d being the dimension of the spatial domain. In particular, we assume a correlation function for a spatially continuous process that depends only on the Euclidean distance $u = ||\boldsymbol{x}_i - \boldsymbol{x}_j||$ between pair of points. Popular choices for the correlation function are the exponential, Matérn, and spherical models. Based on these assumptions, the random part of the linear predictor is Gaussian with covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{R}(\boldsymbol{x}; \phi) + \tau^2 \boldsymbol{I}$.

2.1. Estimation and Laplace approximation

The main objective is to estimate the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \tau^2, \phi, \psi)$ by maximizing the marginal likelihood function, obtained by integrating out the random effects $\boldsymbol{S}(\boldsymbol{x})$ from the joint distribution of the random terms in the model, that is, as follows:

$$\mathrm{L}(\boldsymbol{\theta}; \boldsymbol{y}(\boldsymbol{x})) = \int_{\Re^n} f(\boldsymbol{y}(\boldsymbol{x})|\boldsymbol{S}(\boldsymbol{x})) f(\boldsymbol{S}(\boldsymbol{x})) \mathrm{d}\boldsymbol{S}(\boldsymbol{x}) \tag{1}$$

The marginal likelihood function (1) is a product of two distributions and, in general, analytically intractable. The first term is the sampling distribution for the observed variable $\boldsymbol{y}(\boldsymbol{x})$ given the vector of random effects $\boldsymbol{S}(\boldsymbol{x})$. The second term is assumed to be a multivariate Gaussian distribution. A notable special case is when $f(y; \boldsymbol{\mu}(\boldsymbol{x}), \psi)$ is also a Gaussian distribution in which case the integral has a closed form expression and the parameters τ and ψ becomes confounded. Numerical integration of this function is challenging because the dimension of the integral equals the number of observed values. Methods such as numerical quadrature, Gauss-Hermite, or adaptive Gauss-Hermite (Pinheiro and Bates, 1995) are impracticable. Monte Carlo integration is slow and difficult to monitor for convergence and accuracy (McCulloch, 1997).

An alternative and attractive method is provided by the Laplace approximation (Tierney and Kadane, 1986), frequently used in the analysis of longitudinal data (Molenberghs and Verbeke, 2005). The idea is to approximate the integrand to obtain a tractable closed-form expression

of the integra, allowing the numerical maximization of the marginal likelihood. The Laplace method has been designed to approximate integrals in the form as follows:

$$\int_{\Re^n} \exp\{\boldsymbol{Q}(\boldsymbol{u})\}\mathrm{d}\boldsymbol{u} \approx (2\pi)^{n/2} \left|-\boldsymbol{Q}''(\hat{\boldsymbol{u}})\right|^{-1/2} \exp\{\boldsymbol{Q}(\hat{\boldsymbol{u}})\} \tag{2}$$

where $\boldsymbol{Q}(\boldsymbol{u})$ is a known, uni-modal, and bounded function of a n-dimensional variable $\boldsymbol{u}$ and $\hat{\boldsymbol{u}}$ is the value for which $\boldsymbol{Q}(\boldsymbol{u})$ is maximized. The method requires obtaining the maximum of the integrand and the Hessian $\boldsymbol{Q}''(\hat{\boldsymbol{u}})$, the matrix of second derivatives, either analytically or numerically. In what follows, we discuss how to use the Laplace approximation in the context of SGLMMs.

To simplify the discussion, we assume that the distribution $f(\boldsymbol{y}(\boldsymbol{x})|\boldsymbol{S}(\boldsymbol{x}))$ can be written in the form of the one-parameter exponential family, that is, as follows:

$$f(\boldsymbol{y}(\boldsymbol{x})|\boldsymbol{S}(\boldsymbol{x});\boldsymbol{\beta}) = \exp\left\{\boldsymbol{y}(\boldsymbol{x})^\top(\boldsymbol{D}\boldsymbol{\beta}+\boldsymbol{S}(\boldsymbol{x})) - \mathbf{1}^\top \mathrm{b}(\boldsymbol{D}\boldsymbol{\beta}+\boldsymbol{S}(\boldsymbol{x})) + \mathbf{1}^\top \mathrm{c}(\boldsymbol{y}(\boldsymbol{x}))\right\} \tag{3}$$

where $\mathrm{b}(\cdot)$ and $\mathrm{c}(\cdot)$ are known functions. Commonly used distributions are the Poisson and the Bernoulli/binomial. The multivariate Gaussian density function is given by the following:

$$f(\boldsymbol{S}(\boldsymbol{x});\boldsymbol{\Sigma}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{S}(\boldsymbol{x})^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\boldsymbol{x})\right\} \tag{4}$$

The integrand in (1) is the product of (3) and (4). The likelihood function is in a suitable form for the Laplace approximation with

$$\mathrm{L}(\boldsymbol{\theta};\boldsymbol{y}(\boldsymbol{x})) = \int_{\Re^n} \exp\{\boldsymbol{Q}(\boldsymbol{S}(\boldsymbol{x}))\}\mathrm{d}\boldsymbol{S}(\boldsymbol{x})$$

where

$$\begin{aligned}\boldsymbol{Q}(\boldsymbol{S}(\boldsymbol{x})) = {} & \boldsymbol{y}(\boldsymbol{x})^\top(\boldsymbol{D}\boldsymbol{\beta}+\boldsymbol{S}(\boldsymbol{x})) - \mathbf{1}^\top \mathrm{b}(\boldsymbol{D}\boldsymbol{\beta}+\boldsymbol{S}(\boldsymbol{x})) + \mathbf{1}^\top \mathrm{c}(\boldsymbol{y}(\boldsymbol{x})) \\ & - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\boldsymbol{S}(\boldsymbol{x})^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{S}(\boldsymbol{x})\end{aligned} \tag{5}$$

Equation (5) highlights the convenience of Laplace method for estimation of SGLMMs. The function $\boldsymbol{Q}(\boldsymbol{S}(\boldsymbol{x}))$ is the sum of a GLM log-likelihood for the observed variable and a multivariate Gaussian log-likelihood for the latent variable. The approximation in (2) requires the maximum $\hat{\mathbf{s}}$ of the function $\boldsymbol{Q}(\boldsymbol{S}(\boldsymbol{x}))$ for a particular set of model parameters, a high n-dimensional maximization problem. We adopt an efficient Newthon–Raphson (NR) algorithm to find $\hat{\boldsymbol{s}}$, although other numerical maximization methods can be used. The NR algorithm consists of an iterative scheme as follows:

$$\boldsymbol{s}_{i+1} = \boldsymbol{s}_i - \boldsymbol{Q}''(\boldsymbol{s}_i)^{-1}\boldsymbol{Q}'(\boldsymbol{s}_i)$$

until convergence, which gives $\hat{\boldsymbol{s}}$. Note that up to this stage, all parameters are considered known. After convergence, the value of the integral is computed to evaluate the log-likelihood. The generic expressions for the derivatives required by the NR algorithm are given by the following:

$$\boldsymbol{Q}'(\boldsymbol{s}) = \{\boldsymbol{y}(\boldsymbol{x}) - \mathrm{b}'(\boldsymbol{D}\boldsymbol{\beta}+\boldsymbol{s})\}^\top - \boldsymbol{s}^\top\boldsymbol{\Sigma}^{-1}$$

and

$$\boldsymbol{Q}''(\boldsymbol{s}) = -\mathrm{diag}\{\mathrm{b}''(\boldsymbol{D}\boldsymbol{\beta}+\boldsymbol{s})\} - \boldsymbol{\Sigma}^{-1}$$

Finally, the Laplace approximation for the log-likelihood is as follows:

$$\begin{aligned}\mathrm{l}(\boldsymbol{\theta};\boldsymbol{y}(\boldsymbol{x})) = {} & \frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left|\mathrm{diag}\left\{\mathrm{b}''(\boldsymbol{D}\boldsymbol{\beta}+\hat{\boldsymbol{s}}(\boldsymbol{\theta}))\right\} + \boldsymbol{\Sigma}^{-1}\right| + \boldsymbol{y}(\boldsymbol{x})^\top(\boldsymbol{D}\boldsymbol{\beta}+\hat{\boldsymbol{s}}(\boldsymbol{\theta})) \\ & - \mathbf{1}^\top\mathrm{b}(\boldsymbol{D}\boldsymbol{\beta}+\hat{\boldsymbol{s}}(\boldsymbol{\theta})) + \mathbf{1}^\top\mathrm{c}(\boldsymbol{y}(\boldsymbol{x})) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\hat{\boldsymbol{s}}(\boldsymbol{\theta})^\top\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{s}}(\boldsymbol{\theta})\end{aligned}$$

that can now be numerically maximized over the model parameters. Here, we use the notation $\hat{\boldsymbol{s}}(\boldsymbol{\theta})$ to emphasize that $\hat{\boldsymbol{s}}$ is a function of the model parameters $\boldsymbol{\theta}$. For the maximization step, we typically use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, implemented by the function `optim()` in `R` and parametrize the model as $\boldsymbol{\theta} = (\boldsymbol{\beta}, \log(\sigma^2), \log(\phi), \log(\tau^2), \log(\psi))$. Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$. Then, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is as follows:

$$\hat{\boldsymbol{\theta}} \sim N\left(\boldsymbol{\theta}, \boldsymbol{I}_o^{-1}(\hat{\boldsymbol{\theta}})\right) \tag{6}$$

where $\boldsymbol{I}_o(\hat{\boldsymbol{\theta}})$ denotes the observed information matrix. Often, this type of asymptotic approximation does not work well for inference on covariance parameters. However, in the context of SGLMMs, the main interest lies in the covariance parameters σ^2, τ^2, and ϕ. Thus, we propose to use the profile likelihood approach to compute confidence intervals for the covariance parameters, especially when analyzing small or medium sized data sets. Details on how to implement profile likelihoods in `R` can be found in Bolker and Team (2014).

The maximization step over the model parameters requires reasonable starting values for fast convergence. We propose the following simple strategy: fit a generalized linear model to obtain initial values for $\boldsymbol{\beta}$, based on these values compute $\hat{\boldsymbol{\mu}}$, and then compute the raw residuals $\hat{\boldsymbol{r}} = (\mathbf{y} - \hat{\boldsymbol{\mu}})$. The sample variance of $\hat{\boldsymbol{r}}$ can be used as an initial value for σ^2. If the model has a nugget effect τ^2, we use some percentage of the initial value of σ^2, for example 10%, as an initial value for τ^2. Finally, for ϕ, we recommend 10% of the largest distance between two observed points. The NR algorithm also requires good initial values, and we use the expected values of the random effects.

The `R` code in the supplementary material provides the generic fit function `sglmm()`, which works similarly to the standard `R` function `glm()` with additional arguments to specify the correlation function. Initial values, as decribed earlier, are obtained by the function `start.values.sglmm()`.

3. DATA ANALYSES

To evaluate the performance of our algorithm, we compare our estimates with the results for two examples from the literature. The Rhizoctonia root rot data and the Rongelap data are examples of binomial and count SGLMMs, respectively. Additionally, we also fit a negative binomial SGLMM for count data and reanalyse the Rongelap data.

3.1. Binomial example

Rhizoctonia root rot is a disease that affects the roots of the plants and hinders the process of absorbing water and nutrients. In this experiment, 15 plants were taken from each of 100 randomly chosen locations on a farm, and the number of crown roots and infected crown roots was counted.

Following Zhang (2002) and Evangelou *et al.* (2011), we adopted a binomial SGLMM along with a logit link function, a constant mean, β_0, and a spherical correlation function with a nugget effect for the spatial random effects. Evangelou *et al.* (2011) compare five methods for estimation of SGLMMs for these data. Here, `trans` denotes a simple linear mixed model on the logit observed proportions, while `LA1`, `LA2a`, and `LA2b` are different modifications of the modified Laplace approximation. Monte Carlo likelihood `(MCL)` is the MCMC likelihood implemented in the `R` package `geoRglm` with a burn-in of 10,000 and followed by 10,000 iterations with a thinning of rate of 50. Here, we include in the comparison the Monte Carlo expectation maximization algorithm proposed by Zhang (2002) and the `LAPLACE` approximation from this paper.

The results in Table 1 show that the `Laplace` approximation gives similar estimates to `LA2a`, `LA2b`, and `MCL`. Zhang (2002) and Evangelou *et al.* (2011) reported standard errors for the parameter estimates shown in Table 1. We argue that in general, inferences based on the standard errors do not make sense when analysing small data sets, such as the Rhizoctonia data, especially for the covariance parameters associated with the spatial random effect. Instead, we use profile likelihoods to quantify the uncertainty associated with these estimates. Figure 1 shows the profile likelihoods expressed in terms of the square root of the profile deviances for β_0 and the logarithm of the covariance parameters.

Table 1. Parameter estimates by different methods for the Rhizoctonia root rot data

	Methods						
Estimates	`trans`	`LA1`	`LA2a`	`LA2b`	`MCL`	`MCEM`	`Laplace`
$\hat{\beta}_0$	−1.76	−1.72	−1.72	−1.72	−1.72	−1.62	−1.72
$\hat{\sigma}^2$	0.13	0.08	0.10	0.11	0.10	0.18	0.11
$\hat{\tau}^2$	0.64	0.49	0.47	0.47	0.47	0.35	0.47
$\hat{\phi}$	151.20	149.10	148.60	148.80	148.30	145.11	148.66

MCL, Monte Carlo likelihood; MCEM, Monte Carlo expectation maximization.

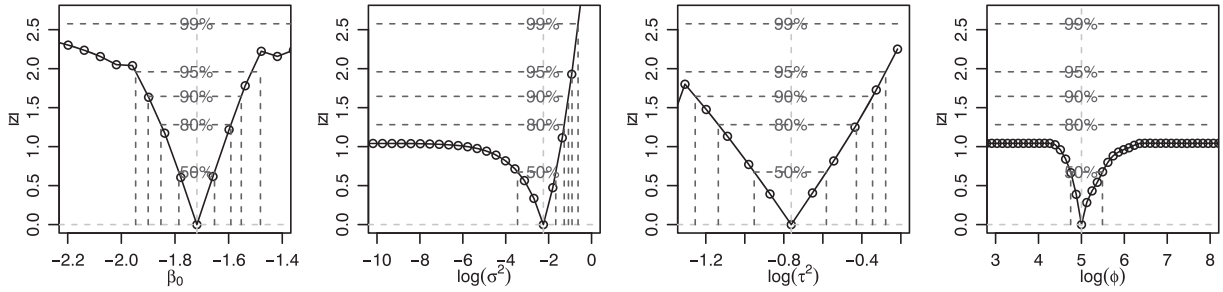

Figure 1. Profile likelihoods for parameters in the binomial spatial generalized linear mixed model fitted for the Rhizoctonia root rot data

Table 2. Parameter estimates using `MCL` and `Laplace` methods by models for the Rongelap data

Methods	MCL				Laplace				
Model	$\hat{\beta}_0$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\phi}$	$\hat{\beta}_0$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\phi}$	logLik
$\kappa = 0.5$	1.83	0.30	—	103.13	1.83	0.30	—	103.27	−1317.99
$\kappa = 0.5 + \tau^2$	1.82	0.26	0.04	152.00	1.82	0.26	0.04	151.80	−1317.19
$\kappa = 1.5$	1.94	0.22	—	0.00	1.94	0.22	—	2.47	−1337.25
$\kappa = 1.5 + \tau^2$	1.82	0.24	0.07	75.53	1.82	0.24	0.07	75.49	−1315.75
$\kappa = 2.5$	1.94	0.22	—	0.00	1.94	0.22	—	0.00	−1337.25
$\kappa = 2.5 + \tau^2$	1.82	0.24	0.07	53.61	1.83	0.24	0.07	53.58	−1315.08
Spherical	1.84	0.33	—	212.37	1.84	0.33	—	212.43	−1318.02
Spherical + τ^2	1.84	0.24	0.04	252.28	1.84	0.24	0.04	252.08	−1315.91

MCL, Monte Carlo likelihood; logLik, log-likelihood.

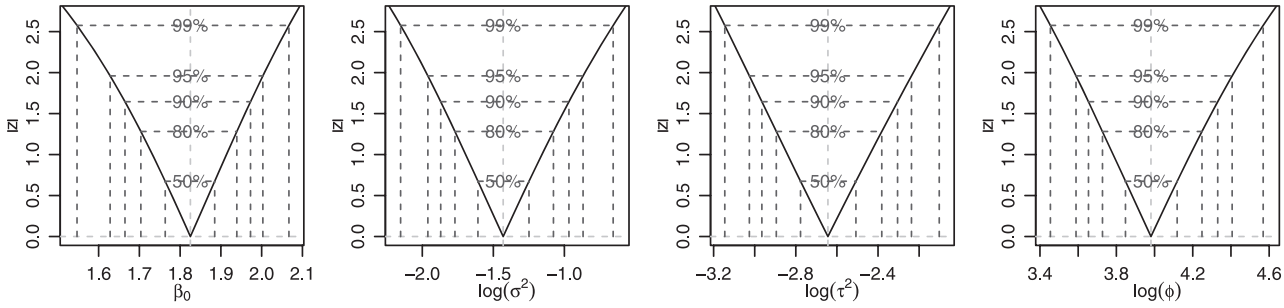

Figure 2. Profile likelihoods for the Poisson spatial generalized linear mixed model fitted with Matérn correlation function and $\kappa = 2.5$ for the Rongelap data

The plots for β_0 and τ^2 are compatible with a quadratic profile likelihood. In such cases, the asymptotic result (6) can be used to compute confidence intervals. For the parameters in the spatial structure, the profile likelihoods are highly asymmetric, showing that these data do not provide enough information for standard inferences on these parameters.

The attraction of the Laplace method is that it provides the maximized log-likelihood, which can be used for model comparison. For example, the maximized log-likelihood values for a binomial GLM and SGLMM fitted for the Rhizoctonia root rot data are −726.89 and −400.27, respectively. Based on these values, we can calculate the *Akaike* information criterion giving 1455.79 and 808.53, respectively, indicating the clearly superior fit of the binomial SGLMM over the basic GLM.

3.2. Poisson example

The Rongelap dataset consists of 157 measurements of radionuclide counts for various time durations. We adopted a Poisson SGLMM along with a log link function, constant mean, β_0, and two correlation functions, Matérn, and spherical. The models are fitted with and without a nugget effect. Furthermore, for the Matérn correlation function, we fix the smooth parameter κ at three different values, 0.5, 1.5, and 2.5. We fitted the Poisson SGLMM using `Laplace` and `MCL` methods. For the last method, we considered a burn-in size of 20,000 and a subsequent iteration size of 50,000 with thinning rate of 50 using the `geoRglm` package. The strategy described in Section 2 for initial values was applied for both methods. Table 2 presents parameter estimates from `MCL` and `Laplace` methods, along with the maximized log-likelihood value obtained by Laplace approximation.

The results in Table 2 show that both methods provide similar estimates. In general, the fitted models support the inclusion of a nugget effect. Overall, the best fitting model is the one with a Matérn correlation function and $\kappa = 2.5$, although there is little difference with Matérn ($\kappa = 1.5$) and spherical correlation models both with a nugget effect. Figure 2 shows the profile likelihoods for the parameters in the best fitting model.

The plots in Figure 2 show that for all model parameters, the profile likelihoods are nearly quadratic in behavior. The asymptotic result (6) can be used to compute confidence intervals on the logarithmic scale, and the invariance of the maximum likelihood estimator can be used to obtain (asymmetric) confidence intervals on the original scale.

3.3. Negative binomial example

In this section, we revisit the analysis of the Rongelap data by now fitting a negative binomial SGLMM for the counts. As would be expected, all negative binomial models provide a better fit than the Poisson counterparts, based on the maximized log-likelihood values. Table 3 presents the parameter estimates and the maximized log-likelihood values for the same models as fitted in Section 3.2.

Table 3. Parameter estimates for the negative binomial spatial generalized linear mixed model fitted by Laplace approximation for the Rongelap data

Model	$\hat{\beta}_0$	$\hat{\sigma}^2$	$\hat{\tau}^2$	$\hat{\phi}$	$\hat{\psi}$	logLik
$\kappa = 0.5$	1.98	0.03	—	663.84	7.24	−1310.08
$\kappa = 0.5 + \tau^2$	1.98	0.03	0.00	664.60	7.24	−1310.08
$\kappa = 1.5$	1.97	0.03	—	269.48	7.23	−1309.71
$\kappa = 1.5 + \tau^2$	1.97	0.03	0.00	269.44	7.23	−1309.72
$\kappa = 2.5$	1.97	0.03	—	198.80	7.21	−1309.66
$\kappa = 2.5 + \tau^2$	1.97	0.03	0.00	198.73	7.21	−1309.66
Spherical	1.97	0.03	—	1265.15	7.26	−1309.83
Spherical + τ^2	1.97	0.03	0.00	1265.60	7.26	−1309.83

logLik, log-likelihood.

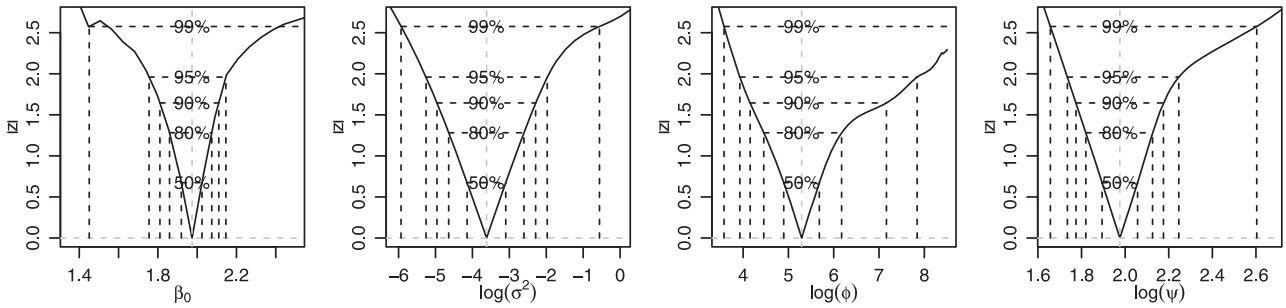

Figure 3. Profile likelihoods for the negative binomial spatial generalized linear mixed model fitted with Matérn correlation function and $\kappa = 2.5$ for the Rongelap data

Results in Table 3 show that the extra variability captured by nugget effect is now redundant with the inclusion of the additional dispersion parameter in the negative binomial model. This possibly emerges from the fact that the negative binomial distribution corresponds to the marginal distribution of a hierarchical model with Poisson responses conditional unstructured gamma distributed random effects (Molenberghs *et al.*, 2007).

As with the Poisson model, the best fit is achieved using a Matérn correlation function with $\kappa = 2.5$. The partial sill from the negative binomial model (0.03) is smaller than that from the Poisson (0.24), and the estimate of the range parameter is also larger. Results indicate that the gamma random effect implicit in the negative binomial distribution captures much of the variability in the data, and the spatial effect is now weak and smooth. Figure 3 presents the profile likelihoods for the parameters in the best fitted model.

In this case, the profile likelihoods are slightly right skewed, and so are the corresponding confidence intervals.

4. DISCUSSION

We investigated the performance of a simple algorithm to fit SGLMMs based on the Laplace approximation. Through the analysis of two data sets, we showed that our approach provides results similar to other methods, such as MCMC likelihood, Monte Carlo expectation maximization, and modified Laplace approximation. The advantages of Laplace approximation are its simplicity and the possibility of computing the maximized log-likelihood value and obtaining profile likelihoods, two valuable tools for inference.

We argue that the Laplace method is easier than the alternatives based on simulation, because it avoids the additional burdens of tuning and convergence checks inherent in MCMC methods. The choice of initial values may be critical; however, the computation of the profile likelihoods works as an algorithm check, because they explore the log-likelihood function in greater detail and may warn against local maxima.

The computational overheads are the computations with a dense covariance matrix, but these also affect alternative approaches. This may be alleviated using, for example, covariance tapering (Furrer *et al.*, 2006) (Kaufman *et al.*, 2008), predictive process (Eidsvik *et al.*, 2012), low rank kriging (Cressie and Johannesson, 2008), and Stochastic Partial Differential Equations (SPDE) models (Lindgren *et al.*, 2011). Each method has its own cost replacing the complexity of the integral by procedures, which also depend on choices and tunings, which may be tricky to apply or influential on the final results.

We showed through examples that the Laplace method is well suited and a safe option for standard applications. Despite developments of new and sometimes sophisticated algorithms and methods, it is desirable to have a straightforward implementation of inference and prediction for SGLMMs such that they can be safely used in routine applications that do not demand more sophisticated and specialized algorithms.

Possible topics for further investigation and extensions include analysing other datasets, performing simulations studies, improving the `R` code, and incorporating this method into general software such as the `geoR` package. The Laplace approximation may be extended to deal with multivariate SGLMMs, as well as space-time GLMMs. One more ambitious extension is to specify SGLMMs using the larger dispersion families of Tweedie and Poisson–Tweedie distributions, which are a generalization of the distributions discussed in this paper.

Acknowledgements

The authors thank the editor, anonymous reviewers, and Professor John Hinde for their comments and suggestions that substantially improve the article.

REFERENCES

Baghishani H, Mohammadzadeh M. 2011. A data cloning algorithm for computing maximum likelihood estimates in spatial generalized linear mixed models. *Computational Statistics and Data Analysis* **55**(4):1748–1759.
Banerjee S, Carlin BP, Gelfand AE. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Taylor & Francis: Boca Raton.
Bolker B, Team R Development Core: bbmle: Tools for general maximum likelihood estimation. R package version 1.0.16
Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**(421):9–25.
Breslow NE, Lin X. 1995. Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82**(1):81–91.
Christensen OF. 2004. Monte Carlo maximum likelihood in model-based geostatistics. *Journal of Computational and Graphical Statistics* **13**(1):702–718.
Christensen OF, Møller J, Waagepetersen R. 2000. Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo, Technical Report, Department of Mathematical Sciences, University of Aarhus.
Christensen OF, Ribeiro Jr PJ. 2002. geoRglm: A package for generalised linear spatial models. *R News* **2**(2):26–28.
Christensen OF, Roberts GO, Sköld M. 2006. Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics* **15**(1):1–17.
Christensen OF, Waagepetersen R. 2002. Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics* **58**(2):280–286.
Cressie N, Johannesson G. 2008. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B* **70**(1):209–226.
Diggle PJ, Ribeiro Jr PJ. 2007. *Model Based Geostatistics*. Springer: New York.
Diggle PJ, Tawn JA, Moyeed RA. 1998. Model based geostatistics. *Journal of Royal Statistical Society: Series C* **47**(3):299–350.
Eidsvik J, Finley AO, Banerjee S, Rue H. 2012. Approximate Bayesian inference for large spatial datasets using predictive process models. *Computational Statistics and Data Analysis* **56**(6):1362–1380.
Evangelou E, Zhu Z, Smith RL. 2011. Asymptotic inferente for spatial GLMM using high order Laplace approximation. *Journal of Statistical Planning and Inference* **141**(11):3564–3577.
Furrer R, Genton MG, Nychka D. 2006. Covariance tapering for interpolation of large spatial datasets. *Journal of Computation and Graphical Statistics* **15**(3):502–523.
Heagerty PJ, Lele SR. 1998. A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association* **93**(443):1099–1111.
Hosseini F, Eidsvik J, Mohammadzadeh M. 2011. Approximate Bayesian inference in spatial GLMM with skew normal latent variables. *Computational Statistics and Data Analysis* **55**(4):1791–1806.
Jing L, De Oliveira V. 2015. geoCount: An R package for the analysis of geostatistical count data. *Journal of Statistical Software* **63**(11):1–33.
Kaufman CG, Schervish CG, Nychka DW. 2008. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association* **103**(484):1545–1555.
Lele SR, Nadeem K, Schmuland B. 2010. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association* **105**(492):1617–1625.
Lin X, Breslow NE. 1996. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91**(435):1007–1016.
Lindgren F, Rue H, Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B* **73**(4):423–498.
McCulloch CE. 1997. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**(437): 162–170.
Millar RB. 2011. *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB*, Statistics in Practice. Wiley: Chichester.
Molenberghs G, Verbeke G. 2005. *Models for Discrete Longitudinal Data*. Springer: New York.
Molenberghs G, Verbeke G, Demétrio CGB. 2007. An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis* **13**(4):513–531.
Møller J. 2003. *Spatial Statistics and Computational Methods*. Springer-Verlag: New York.
Pinheiro JC, Bates DM. 1995. Approximations to the log likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics* **4**(1):12–35.
R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria
Rue H, Martino S. 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference* **137**(10):3177–3192.
Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B* **71**(2):319–392.
Tierney L, Kadane J. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**(393):82–86.
Torabi M. 2015. Likelihood inference for spatial generalized linear mixed models. *Communications in Statistics - Simulation and Computation* **44**(7): 1692–1701.
Varin C, Høst G, Skare Ø. 2005. Pairwise likelihood inference in spatial generalized linear mixed models. *Computational Statistics and Data Analysis* **49**(4):1173–1191.
Zhang H. 2002. On estimation and prediction for spatial generalized linear mixed models. *Biometrics* **58**(1):129–136.

SUPPORTING INFORMATION

Datasets and R code for the analysis are available at the paper companion page at http://www.leg.ufpr.br/sglmm.

