Practical likelihood analysis for Spatial Generalized Linear Mixed Models

Wagner Hugo Bonat *

LEG/DEST - Paraná Federal University

Abstract

We propose a standard approach to make inference for spatial generalized linear mixed models using Laplace approximation. Based on analysis of two datasets previous analysed in literature, we compare our approach with different approaches. The first the rhizoctonia root rot dataset is an example of Binomial SGLMM and the second rongelap dataset is an example of Poisson or Negative Binomial SGLMM. Our results show that Laplace approximation provides point estimate really similar to MCMC likelihood, MCEM and modified Laplace approximation. The advantege to use Laplace approximation is to avoid tuning and convergence analysis when using based simulation method. Furthermore, using Laplace approximation we can compute the maximum log-likelihood value and realistic confidence interval based on profile likelihood. We provide `R` code to use our approach on the supplement material webpage.

keywords: Laplace approximation, likelihood inference, spatial data, generalized linear mixed models

*Corresponding author: wbonat@ufpr.br, Dept. Estatística-UFPR, CP 19.081, Curitiba, PR Brazil, 81.531-990

1 Introduction

Evaluating the likelihood function for spatial generalised linear mixed models (SGLMM) (Diggle et al., 1998) requires the solution of a high dimensional integral for the spatially correlated random effects. E.g., for the standard setup with observed data regarded as a single partial realisation of a process within an area, the dimension of the integral is equals the number of data. Solve such intractable integral is the main problem in these models.

In the literature, such models are most commonly fitted by Monte Carlo Markov chain (MCMC) methods under the Bayesian paradigm. Diggle et al. (1998) presents Bayesian inference for SGLMM, a more efficient Langevin-Hastings MCMC algorithm was proposed by Christensen et al. (2000). A robust version of their method was presented by Christensen et al. (2006). Christensen and Waagepetersen (2002) presents the general ideas about Bayesian prediction in the context of count data. There is a rich literature about Bayesian inference using MCMC methods for SGLMM, an extended review may be found in Banerjee et al. (2004).

In practice, MCMC methods for SGLMM comes with a wide range of problems in terms of convergence and computational time. Moreover, the implementation itself can be problematic, especially for end users who might not be experts in programming. There are several software platforms for fitting generic models with random effects via MCMC, including JAGS (Plummer, 2003), BayesX (Belitz et al., 2012), WinBUGS (Lunn et al., 2000) and others.

Recently, some authors have dedicated attention for approximate methods inside the Bayesian framework. Rue and Martino (2007) proposed approximate Bayesian inference in the context of Gaussian Markov Random Fields. Eidsvik et al. (2012) presents a similar approach for SGLMM and Hosseini et al. (2011) proposes approximate Bayesian inference for SGLMM but with skew Normal latent variables. All these approaches have in common the fact of using Laplace approximation (Tierney and Kadane, 1986) to solve a high dimensional integral. Rue et al. (2009) consolidates the Laplace approximation use, and show that this type of approximation is accurate enough for practical data analysis. Although, computa-

tionally convenient the most serious criticism of the Bayesian approach is its dependence on the choice of the prior distributions.

Christensen (2004) presents the likelihood analysis still relying 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, 2002) and described in details in Diggle and Ribeiro Jr (2007). Alternative methods are proposed by Zhang (2002) adopting a spatial EM algorithm (still with Monte Carlo method).Baghishani and Mohammadzadeh (2011) and Torabi (2014) proposes a different approach based on data clone algorithm Lele et al. (2010).

Although, these algorithms made frequentist analysis of SGLMM they are equally computationally inconvenient that MCMC algorithms in Bayesian analysis. Once that, they make inference based on simulation. In general these methods do not provide maximum log-likelihood value and it is not clear how to use these methods to compute profile likelihood, two important tools in the context of likelihood analysis for SGLMM. Varin et al. (2005) proposes to use pairwise likelihood to make inference in SGLMM, this approach is computationally convenient, because its solve just bidimensional integrals, the overhead is that we need to choose these pair of observations. There are some approaches to do it, but is difficult to evaluate the impact of different choices in the final result.

Likelihood analysis for SGLMM has been the subject of a great deal of research over the past decades. Perhaps the work of Breslow and Clayton (1993) about penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) was the initial point of likelihood analysis for SGLMM. However, it seems to work away the interest of researchers in the use of Laplace aproximation in the context of likelihood analysis. 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) presents a modified Laplace approximation and applied this method for estimation and prediction for SGLMM. The authors compare their results with other approach as PQL

and MCMC likelihood and conclude that their method has better performance.

In the context of generalized linear mixed models Millar (2011) argument that penalized quasi-likelihood have much in common with Laplace approximation, but the author states that the weakness this ad hoc likelihood method is that it use additional approximations to avoid the optimization and second derivative calculation required by the Laplace approximation.

Computational capabilities have increased since the proposal of the SGLMM. This motivates us to revisit standard methods for evaluating the likelihood function. We adopt the Laplace approximation (Tierney and Kadane, 1986) for integrating for the random effects. Under this approach, for each evaluation of the likelihood function, the computational burden for integration is replaced a maximization to find optimised values for the random effects and the Hessian around such maximum. In other words we have an optimisation algorithm for the random effect within the optimisation for the model parameters.

In what follows we describe the algorithm and present analysis for the rhizoctonia root rot data previous analysed by Zhang (2002) and Evangelou et al. (2011) and compare our results with the results from these papers. We reanalyse the Rongelap data which motivates Diggle et al. (1998). Using rongelap data we compare our approach with MCMC likelihood, such that implementated in the package **geoRglm** and we propose an extended analysis using Negative Binomial distribution. Our procedure can be immediately extended to other GLM families and Beta models.

There is no free lunch each method "charges its price" replacing the complexity of the integral by some procedures which also depends on choices and tunings, which may be trick or influential in the results. We argue that, despite the developments of new and sometimes sophisticated algorithms and methods, there is a need to have an straightforward implementation of inference and prediction for SGLMM such that they can be widely used in routine simple applications which may be safer for general usage than specialised algorithms.

2 Spatial Generalized Linear Mixed Models

The spatial generalized linear mixed models (SGLMM), assume that we have recorded observations $\mathbf{y} = (y_1, \ldots, y_n)^T$ at locations $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $\mathbf{y}$ is a realization of $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$. The hierarquical description for the SGLMM is the following:

$$\begin{aligned} \mathbf{Y}(\mathbf{x})|S(\mathbf{x}) &\sim f(\cdot; \boldsymbol{\mu}(\mathbf{x}), \psi) \\ g(\boldsymbol{\mu}(\mathbf{x})) &= D\boldsymbol{\beta} + S(\mathbf{x}) = D\boldsymbol{\beta} + \sigma R(\mathbf{x}; \phi) + \tau z \quad \text{and} \\ S(\mathbf{x}) &\sim N(\mathbf{0}, \Sigma). \end{aligned} \tag{1}$$

The model assumes that $Y_1, \ldots, Y_n$ are conditionally independent given a (Gaussian) spatial process $S(\mathbf{x})$ and distributed as $f(\cdot; \boldsymbol{\mu}(\mathbf{x}), \psi)$. This distribution has two set of parameters. The first denoted by $\boldsymbol{\mu}(\mathbf{x})$ is related to the linear predictor through the link function g. The second denoted by ψ is a precision or dispersion parameter which must be included in the density or probability function f and here is treated only as an extra parameter in the likelihood function. The spatial process is decomposed on spatially dependent ($R(\mathbf{x}; \phi)$) and independent (Z) processes both with unit variance and scaled by the parameters σ and τ, respectively.

The linear predictor consist of the sum of fixed effects $D\boldsymbol{\beta}$, spatially correlated random effects $R(\mathbf{x}; \phi)$ and uncorrelated $\tau z = \epsilon \sim N(0, \tau^2 I)$. D is a design matrix with potential covariates and $\boldsymbol{\beta}$ is a vector of regression parameters with dimension $p \times 1$.

In geostatistical context $R(\mathbf{x}; \phi)$ is a unit variance Gaussian random field (GRF) with correlation function $\rho(u, \phi)$ where u denotes the distance between a pair of spatial locations and ϕ a parameter that measures the spatial correlation. For a given choice of a valid correlation function $\rho \in \Re^d$ parametrized by ϕ with d being the dimension of the spatial domain. Here we assume in particular correlation functions for spatially continuous process depending only on Euclidean distance $u = ||x_i - x_j||$ between pair of points. The common choices for the correlation functions are the exponential, Matérn and spherical. The random

part of the linear predictor has variance-covariance matrix $\Sigma = \sigma^2 R(\mathbf{x}; \phi) + \tau^2 I$.

2.1 Estimation and Laplace approximation

The main interest here is to make inference with respect the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \tau^2, \phi, \psi)$ whose estimates are given by the maximization of the marginal likelihood function, obtained by integrating out the random effects $S(\mathbf{x})$ from the joint distribution of the random terms of the model:

$$L(\boldsymbol{\theta}; \mathbf{y}) = \int_{\Re^n} [\mathbf{Y}(\mathbf{x})|S(\mathbf{x})][S(\mathbf{x})] \mathrm{d}S(\mathbf{x}). \tag{2}$$

The marginal likelihood function (10) is a product between two distributions and analytically intractable in general. The first is the sampling distribution for the observed variable $\mathbf{Y}$ and the second is typically assumed to be a multivariate Gaussian distribution. A notable exception is the case where a Gaussian distribution is also assumed for $\mathbf{Y}$, in this case the integral has a closed form expression and notice that the parameters τ and ψ becomes indistinguishable.

Numerical integrate this function may be challenge because its dimension n is equal the number of observed values. Methods such as numerical quadrature, Gauss-Hermite or adaptive Gauss-Hermite (Pinheiro and Bates, 1995) are impracticable here. Monte Carlo integration is slow and difficult to monitor for convergence and accuracy (McCulloch, 1997).

An alternative is provided by the Laplace approximation of the integral (Tierney and Kadane, 1986) which is largely adopted in longitudinal data analysis (Molenberghs and Verbeke, 2005). Overall, the Laplace approach is not widely explored or adopted for spatial generalized linear mixed models. The idea is to approximate the integrand to obtain a tractable expression for which a closed-form expression of the integral can be obtained, turning feasible the numerical maximization of the marginal likelihood. The Laplace method has been designed to approximate integrals as follows

$$I = \int_{\Re^n} \exp\{Q(\mathbf{s})\} \mathrm{d}\mathbf{s} \approx (2\pi)^{n/2} |-Q''(\hat{\mathbf{s}})|^{-1/2} \exp\{Q(\hat{\mathbf{s}})\} \tag{3}$$

where $Q(\mathbf{s})$ is a known, unimodal, and bounded function of a n-dimensional variable ($\mathbf{s}$). Let $\hat{\mathbf{s}}$ be the value for which Q is maximized. The method requires obtaining the maximum of the integrand and the Hessian ($Q''(\hat{\mathbf{s}})$), the matrix of second derivatives either analytically or numerically.

Such approximation may be used for likelihood based inference for generalized linear mixed models in general and in particular for spatial generalized linear mixed models as follows. For simplicity, assume that the distribution f can be written in the form of the one parameter exponential family,

$$f(\mathbf{y};\boldsymbol{\beta}) = \exp\{\mathbf{y}^T(D\boldsymbol{\beta} + S(\mathbf{x})) - \mathbf{1}^T b(D\boldsymbol{\beta} + S(\mathbf{x})) + \mathbf{1}^T c(\mathbf{y})\} \tag{4}$$

where b is a specific function. Common used distribution are the Poisson and the Bernoulli/Binomial. The multivariate Gaussian density function is given by,

$$f(S(\mathbf{x});\Sigma) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\{-\frac{1}{2}S(\mathbf{x})^T\Sigma^{-1}S(\mathbf{x})\}. \tag{5}$$

The integrand in (10) is the product of (4) and (5). The likelihood function written in a suitable format for the Laplace approximation is

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

where

$$Q(S(\mathbf{x})) = \mathbf{y}^T(D\boldsymbol{\beta}+S(\mathbf{x}))-\mathbf{1}^T b(D\boldsymbol{\beta}+S(\mathbf{x}))+\mathbf{1}^T c(\mathbf{y})-\frac{n}{2}\log(2\pi)-\frac{1}{2}\log|\Sigma|-\frac{1}{2}S(\mathbf{x})^T\Sigma^{-1}S(\mathbf{x}). \tag{6}$$

The equation (6) highlights the convenience of adopting the Laplace method to fit SGLMM. The function $Q(S(\mathbf{x}))$ is the sum of a GLM log-likelihood for the observed variable and a multivariate Gaussian log-likelihood for the latent variable. To use the approximation at (3)

we need the maximum $\hat{\mathbf{s}}$ of the function $Q(S(\mathbf{x}))$ for a particular set of model parameters. This is a high n-dimensional maximization. We propose to use an efficient Newthon-Raphson (NR) algorithm to find $\hat{\mathbf{s}}$ although other numerical maximisation methods can be used. The NR algorithm consists in computing:

$$\mathbf{s}_{i+1} = \mathbf{s}_i - Q''(\mathbf{s}_i)^{-1} Q'(\mathbf{s}_i) \tag{7}$$

up to convergence to get $\hat{\mathbf{s}}$. Note that at this stage all parameter are considered known. After convergence for the NR algorithm the value of the integral is computed to evaluate the likelihood. The generic derivatives are given by:

$$Q'(\mathbf{s}) = \{\mathbf{y} - b'(D\boldsymbol{\beta} + \mathbf{s}\}^T - \mathbf{s}^T \Sigma^{-1}, \tag{8}$$

and

$$Q''(\mathbf{s}) = -diag\{b''(X\boldsymbol{\beta} + \mathbf{s}\} - \Sigma^{-1}. \tag{9}$$

The log-likelihood approximate by Laplace method is given by:

$$l(\boldsymbol{\theta}; \mathbf{y}) = \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|-diag\{b''(D\boldsymbol{\beta} + \hat{\mathbf{s}})\} - \Sigma^{-1}| + \mathbf{y}^T(D\boldsymbol{\beta} + \hat{\mathbf{s}}) - \mathbf{1}^T b(D\boldsymbol{\beta} + \hat{\mathbf{s}}) + \mathbf{1}^T c(\mathbf{y}) \\ -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\hat{\mathbf{s}}^T\Sigma^{-1}\hat{\mathbf{s}}$$

which now is maximised numerically over the model parameters. For the outer iterative step we typically use the BFGS algorithm, implemented by the function `optim()` in `R` and parametrised the model as $\boldsymbol{\theta} = (\boldsymbol{\beta}, \log(\sigma^2), \log(\phi), \log(\tau^2), \log(\psi))$. Others algorithms can be used in the inner maximisation, however our tests showed a best performance with NR.

To make inference about $\hat{\theta}$ we can use the well known asymptotic distribution of the maximum likelihood estimator,

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, I_o(\hat{\boldsymbol{\theta}})^{-1})$$

where $I_o(\boldsymbol{\theta})$ denote the observed information of $\boldsymbol{\theta}$. Note that, in the SGLMM models we cannot compute the Fisher information, because the second derivatives of log-likelihood function are not available in a closed form. In this way, we use the observed information matrix computed numerically on the point $\hat{\boldsymbol{\theta}}$.

In general this type of asympotic approximation does not work well for parameters of variance or correlation. In the context of SGLMM we have interest to make inference with respect the covariance parameters (σ^2,τ^2 e ϕ), so we propose to use the profile likelihood approach to build confidence interval for covariance parameters, when analysing small or medium data sets. Details about how implement profile likelihood in `R` can be found in (Bolker and Team, 2014).

2.2 Practical issues

Our procedure is compost of a lot of steps, in the following we try to summary these steps.

1. Choose a set of initial values ($\boldsymbol{\theta}_i$) for the model parameters $\boldsymbol{\theta}$;
2. build the covariance matrix Σ and its inverse Σ^{-1};
3. estimate $\hat{\mathbf{s}}$ by the following maximisation;
   (a) choose a set of initial values for $\mathbf{s}$;
   (b) compute $Q'(\mathbf{s})$ by (8) and $Q''(\mathbf{s})$ by (9);
   (c) solve the system $Q''(\mathbf{s})\mathbf{s}^* = Q'(\mathbf{s})$;
   (d) update $\mathbf{s} = \mathbf{s} + \mathbf{s}^*$;
   (e) iterate up to convergence to obtain $\hat{\mathbf{s}}$.
4. replace $S(\mathbf{x})$ by $\hat{\mathbf{s}}$ and compute $Q(\hat{\mathbf{s}})$ in (6);
5. compute the approximated value of the integral by (3) to obtain the value of the marginal likelihood (10);

6. use for example the $BFGS$ algorithm to get the next value $\boldsymbol{\theta}_{i+1}$;

7. iterate up to convergence to obtain parameter estimates $\hat{\boldsymbol{\theta}}$.

The Newton-Raphson algorithm is efficient, but its require a good initial values. In the case of SGLMM, we have a really good initial value, once that our model suppose that the expected value of the random effects is equal zero, we can use the expected value of $S(\mathbf{x})$ to start our Newton-Raphson algorithm. This step is computationally fast, because everything that we need is solve a system of equations.

Reasonable starting values are important for fast convergence to the correct maximum of the approximate likelihood. We propose a simple strategy to get initial values for the outer iterative step. First, fit a simple generalized linear model to get initial values for $\boldsymbol{\beta}$, based on these values compute $\hat{\boldsymbol{\mu}}$ (predict values), then compute the ordinary residuals $\hat{r} = (\hat{\boldsymbol{\mu}} - \mathbf{y})$ the variance of $\hat{r}$ may be used as a initial value of σ^2, if the model has nugget effect, we can used some percent of the initial value of σ^2 for example 10% as an initial value for τ^2. Finally, the initial value for ϕ depends on the correlation function. In general we recomend 10% of the bigger distance between two observed points. It is a safe habit to try different initial values for ϕ, to be sure that found the correct mode.

The procedure described here to get initial values is implemented in `R` through the function `start.values.sglmm()`. The generic function `sglmm()` implements the algorithm described and its use is similar the function `glm()` with additional arguments to specify the correlation function.

3 Results

3.1 Binomial example

To evaluate the quality of our procedure we opted to compare our results with previous results on literature. The first example, we analyse the rhizoctonia root rot data set, that

was analysed by (Zhang, 2002) and (Evangelou et al., 2011). The rhizoctonia root rot is a disease that attaches on the roots of plants and hinders their process of absorbing water and nutrients. In this study 15 plants were pulled from each of 100 randomly chosen locations in a farm and the number of crown roots and infected crown roots were counted. In the same way that (Zhang, 2002) and (Evangelou et al., 2011) we assume constant mean ($Intercept$), spherical covariance structure plus nugget effect for the underlying Gaussian random field. We suppose binomial distribution for the response variable, here the number of infected crown roots.

(Evangelou et al., 2011) compares five methods to estimation of SGLMM for the rhizoctonia root rot data set. The first is a simple logit transformation of the observed probabilities with additional supposition of normal distribution for the transformed response variable, denote this approach by trans. The next three approaches are different combinations of modified Laplace approximation proposed by (Evangelou et al., 2011) denoted by LA1, LA2a and LA2b, for details see (Evangelou et al., 2011), Section 6.1. The fourth is the MCMC likelihood (MCMCLik) implemented in the `R` package `geoRglm`. The burn-in size was 10000 and the subsequent iteration size 10000 with thinning 50. The fifth method is the MCEM approach proposed by (Zhang, 2002). We reproduce here the Table 5 from (Evangelou et al., 2011) with some modifications and include our estimates in the column called LAPLACE in Table 1.

	trans	LA1	LA2a	LA2b	MCMCLik	MCEM	LAPLACE
Intercept	-1.76	-1.72	-1.72	-1.72	-1.72	-1.62	-1.72
Partiall sill	0.13	0.08	0.10	0.11	0.10	0.18	0.11
Nugget	0.64	0.49	0.47	0.47	0.47	0.35	0.47
Range	151.20	149.10	148.60	148.80	148.30	145.11	148.66

Table 1: Point estimates from different methods using the rhizoctonia root rot data set.

The results in Table 1 show that LAPLACE approach presents point estimates really similar to LA2a, LA2b and MCMCLik. It is important to emphasize that any method outside LAPLACE provides the maximum likelihood value. (Zhang, 2002) and (Evangelou et al.,

2011) reported standard errors of the parameter estimates showed in Table 1. In this paper we argument that sometimes such standard error does not make sense for covariance parameters ans small sample size, as rhizoctonia data set. Then, we propose to use profile likelihood to acess the uncertainty of these estimates. The Figure 1 presents the profile likelihood for the covariance parameters reparametrised on the logarithmic scale for computational convenience.

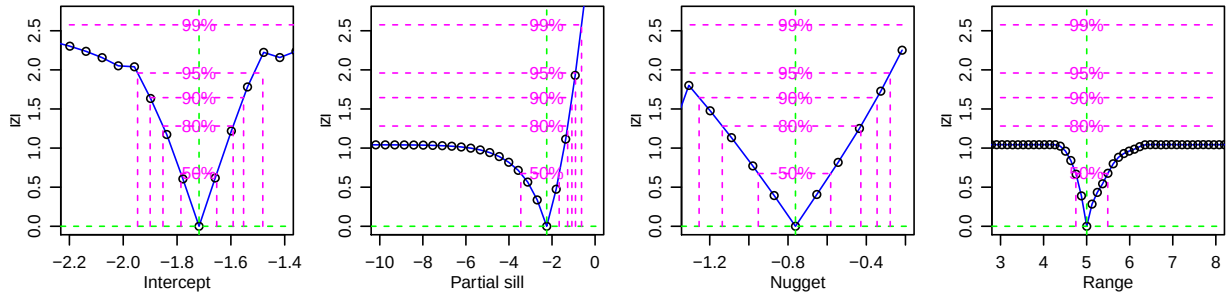

Figure 1: Profile likelihood contours for SGLMM fitted for rhizoctonia root rot data set.

The profile likelihood for the intercept and nugget effect are similar to a quadratic function, so in this case make sense to use the asymptotic result and to build confidence interval based on asymptotic normality of maximum likelihood estimator. However, for covariance parameters that describes the spatial structure the behavior of the profile likelihood is high asymmetric and show clearly that based in this data set we do not have enough information to make inference with for example 95% of confidence.

The second important contribution form our approach is that we can compute the maximum log-likelihood value. For example, the maximum log-likelihood value for a simple GLM model fitted for rhizoctonia root rot data set is -584.50 the maximum log-likelihood value for the SGLMM fitted using spherical correlation function is -400.26, based on these values we can compute for example the *Akaike* criterion (AIC) and choice the best fit. For the rhizoctonia root rot data set the AIC for GLM is 1171.00 and for the SGLMM is 808.53, so no doubts about the best fit.

3.2 Poisson example

The rongelap dataset was analysed by (Diggle et al., 1998) and others. The dataset consist of $n = 157$ measurements of y_i radionuclide counts for various time durations m_i. The data are modeled as a SGLMM with a Poisson distribution. We fit eight models for this dataset, we consider a global mean (Intercept) and two correlation functions (Matérn and spherical). For the Matérn correlation function we consider three values for the smooth parameter ($\kappa = 0.5$, 1.5 and 2.5). We consider too models with and without nugget effect and fit all models using Laplace approximation and MCMC likelihood, using the `R` package `geoRglm`. The burn-in size was 20000 and the subsequent iteration size 50000 with thinning 50. We use the same strategy for initial values for both approaches. The idea here is first to compare the point estimates from Laplace with the benchmarking MCMC likelihood method and after to use the maximum log-likelihood value to choice the best among these eight models.

Table 2 presents the point estimates from Laplace and MCMC likelihood method together the maximum log-likelihood value obtained from Laplace approximation for the eight models described.

Models	Intercept		Partial Sill		Nugget		Range		logLik
	LA	MCL	LA	MCL	LA	MCL	LA	MCL	LA
Matérn $\kappa = 0.5$	1.83	1.83	0.30	0.30	–	–	103.27	103.13	-1317.99
Matérn $\kappa = 0.5$ + nugget	1.82	1.82	0.26	0.26	0.04	0.04	151.80	152.00	-1317.19
Matérn $\kappa = 1.5$	1.94	1.94	0.22	0.22	–	–	2.47	0.00	-1337.25
Matérn $\kappa = 1.5$ + nugget	1.82	1.82	0.24	0.24	0.07	0.07	75.49	75.53	-1315.75
Matérn $\kappa = 2.5$	1.94	1.94	0.22	0.22	–	–	0.00	0.00	-1337.25
Matérn $\kappa = 2.5$ + nugget	1.83	1.82	0.24	0.24	0.07	0.07	53.58	53.61	-1315.08
Spherical	1.84	1.84	0.33	0.33	–	–	212.43	212.37	-1318.02
Spherical + nugget	1.84	1.84	0.24	0.24	0.04	0.04	252.08	252.28	-1315.91

Table 2: Point estimates using Laplace approximation (LA) and MCMC likelihood (MCL) for the rongelap dataset.

We note that both approaches show results really similars. In general the fitted models to indicate that the nugget effect is important. The maximum log-likelihood value indicates that the best model is using Matérn correlation function with $\kappa = 2.5$ which indicate a

smooth spatial process. But the difference between Matérn with $\kappa = 1.5$ and spherical both with nugget effect is really small. The Figure 2 shows the profile likelihood contour for each model parameter.

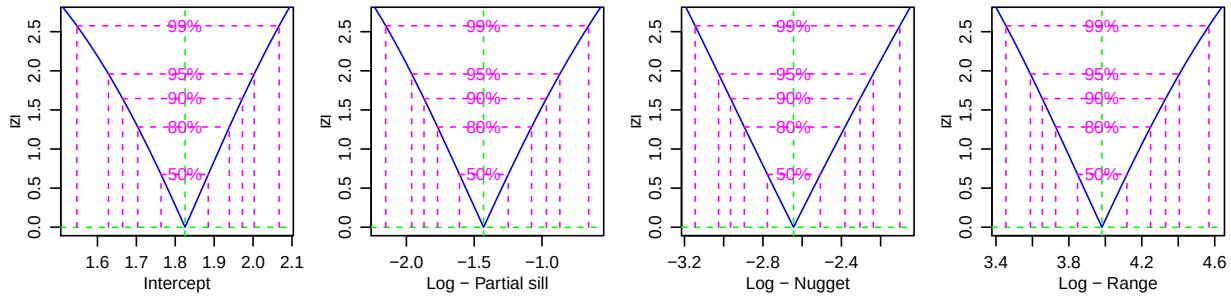

Figure 2: Profile likelihood contours for SGLMM fitted with Matérn correlation function ($\kappa = 2.5$) for rongelap dataset.

The profile likelihood contours presented in the Figure 2 are well-behaved, in the sense that are near a quadratic form. In this way the asymptotic results works well for this data set. We can compute the confidence interval based on the reparametrized model and use the invariance property of the maximum likelihood estimator to obtain confidence interval in the original scale.

3.3 Negative binomial example

In previous section we presented an likelihood analysis for the rongelap dataset. The response variable is the radionuclide counts, so we adopted the Poisson distribution, but we can use some alternative model to compare with Poisson, for example the Negative Binomial. Once that our method provides the maximum log-likelihood value we can compare different supposition about the sampling distribution in a natural way. Table 3 presents the point estimates and maximum log-likelihood values for the same eight models considered in Section 3.2, but now we use the Negative Binomial distribution instead of Poisson.

The results in Table 3 show in general that the nugget effect is not important when using Negative Binomial. The nugget effect captures the non structured random effects, but when

Models	Intercept	Partial sill	Nugget	Range	Dispersion	logLik
Matérn $\kappa = 0.5$	1.98	0.03	–	663.84	7.24	-1310.08
Matérn $\kappa = 0.5$+ nugget	1.98	0.03	0.00	664.60	7.24	-1310.08
Matérn $\kappa = 1.5$	1.97	0.03	–	269.48	7.23	-1309.71
Matérn $\kappa = 1.5$+ nugget	1.97	0.03	0.00	269.44	7.23	-1309.72
Matérn $\kappa = 2.5$	1.97	0.03	–	198.80	7.21	-1309.66
Matérn $\kappa = 2.5$+ nugget	1.97	0.03	0.00	198.73	7.21	-1309.66
Spherical	1.97	0.03	–	1265.15	7.26	-1309.83
Spherical + nugget	1.97	0.03	0.00	1265.60	7.26	-1309.83

Table 3: Point estimates for the Negative Binomial SGLMM using Laplace approximation for the rongelap dataset.

using Negative Binomial distribution the extra dispersion parameter captures these effects. Remember that Negative Binomial is Poisson plus Gamma random effects, so the Gamma effect captures the non structured variation. In this model is difficult to know if is possible to estimate both nugget effect and extra dispersion parameter. We fit these type of model to be sure that our procedure is able to identify non significant nugget effect.

Comparing the maximum log-likelihood values we can see that the best fit is using Matérn correlation function with $\kappa = 2.5$, this result agrees with previous result using Poisson distribution. Now, comparing Negative Binomial and Poisson the maximum log-likelihood value indicate that Negavite Binomial provides a better fit than Poisson. The partial sill from Negative Binomial fit (0.03) is smaller than from Poisson (0.24) and the range is much bigger, these results show that the Gamma random effect in the Negative Binomial distribution capture much of the variability in the data and the spatial effect is weak and really smooth. To finish this likelihoof analysis for the rongelap dataset, we present in Figure 3 the profile likelihood contour for the model parameter in the Negative Binomial SGLMM using Matérn correlation function with $\kappa = 2.5$.

The profile likelihood contours show a slightly right asymmetric, that should be to take into account to compute confidence interval. The general way our procedure is able to estimate SGLMM, compute the maximum log-likelihood value and provide profile likelihood contour, which provides a complete tool for likelihood analysis of the spatial generalized

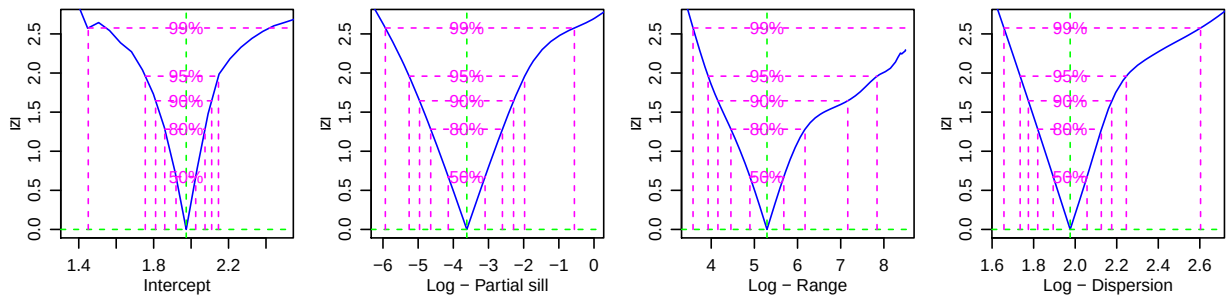

Figure 3: Profile likelihood contours for Negative Binomial SGLMM fitted with Matérn correlation function ($\kappa = 2.5$) for rongelap dataset.

linear mixed models.

4 Conclusion

In this paper we showed how to use Laplace approximation to estimate parameter for spatial generalized linear mixed models. Through a couple of datasets we show that our approach provides results really similar with other method such as MCMC likelihood, MCEM and modified Laplace approximation. The advantege to use Laplace approximation is that we can compute the maximum log-likelihood value and profile likelihood contour, two important tools for likelihood analysis for SGLMM.

The Laplace approximation avoids subjective analysis, such as tuning and convergence in MCMC methods used in MCMC likelihood and MCEM methods. Tuning such algorithm no always is a easy task, mainly for not specialized users. In this sense, Laplace approach is easier to use than methods based on simulation.

The computational overhead from our approach is the inversion of a dense variance-covariance matrix, but this overhead is found on other approaches too. This overhead 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 SPDE models Lindgren et al. (2011).

We provide `R` function to fit SGLMM using Laplace approximation. Our goal here is

provide a standard method to make likelihood analysis for small and medium datasets using spatial generalized linear mixed models. On the supplement material webpage we make available our `R` code and examples using the generic function `sglmm()`.

Possible topics for further investigation and extensions include analyse other real datasets, make simulations studies to evaluate the quality of our estimates, improve the `R` code and include this method in the `geoR` package. The Laplace approximation may be extended to deal with multivariate SGLMM as well as space-time GLMM.

Acknowledgements

References

Baghishani, H. and Mohammadzadeh, M. (2011). A data cloning algorithm for computing maximum likelihood estimates in spatial generalized linear mixed models, *Computational Statistics Data Analysis* **55**(4): 1748 – 1759.

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis.

Belitz, C., Brezger, A., Kneib, T., Lang, S. and Umlauf, N. (2012). *BayesX: Software for Bayesian Inference in Structured Additive Regression Models*. Version 2.1.

Bolker, B. and Team, R. D. C. (2014). *bbmle: Tools for general maximum likelihood estimation*. R package version 1.0.16.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**(421): 9–25.

Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion, *Biometrika* **82**(1): 81–91.

Christensen, O. F. (2004). Monte Carlo maximum likelihood in model-based geostatistics, *Journal of Computational and Graphical Statistics* **13**(1): 702–718.

Christensen, O. F., Mller, J. and Waagepetersen, R. (2000). Analysis of spatial data using generalized linear mixed models and langevin-type markov chain monte carlo, *Technical report*.

Christensen, O. F. and Ribeiro, P. J. (2002). geoRglm: A package for generalised linear spatial models, *R News* **2**(2): 26–28.

Christensen, O. F., Roberts, G. O. and Skld, 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, O. F. and Waagepetersen, R. (2002). Bayesian prediction of spatial count data using generalized linear mixed models, *Biometrics* **58**(2): 280–286.

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

Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model based geostatistics (with discussion), *Applied Statistics* **47**(3): 299–350.

Diggle, P. and Ribeiro Jr, P. (2007). *Model Based Geostatistics*, Springer, New York.

Eidsvik, J., Finley, A. O., Banerjee, S. and Rue, H. (2012). Approximate bayesian inference for large spatial datasets using predictive process models, *Computational Statistics & Data Analysis* **56**(6): 1362 – 1380.

Evangelou, E., Zhu, Z. and Smith, R. L. (2011). Asymptotic inferente for spatial glmm using high order laplace approximation, *Journal of Statistical Planning and Inference* **141**(11): 35643577.

Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets, *J. Comput. Graph. Stat.* **15**(3): 502–523.

Hosseini, F., Eidsvik, J. and Mohammadzadeh, M. (2011). Approximate bayesian inference in spatial {GLMM} with skew normal latent variables, *Computational Statistics Data Analysis* **55**(4): 1791 – 1806.

Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets, *JASA* **103**(484): 1545–1555.

Lele, S. R., Nadeem, K. and 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. and Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion, *Journal of the American Statistical Association* **91**(435): pp. 1007–1016.

Lindgren, F., Rue, H. and 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 (Statistical Methodology)* **73**(4): 423498.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). Winbugs: A bayesian modelling framework: Concepts, structure, and extensibility, *Statistics and Computing* **10**(4): 325–337.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models, *Journal of the American Statistical Association* **92**(437): pp. 162–170.

Millar, R. (2011). *Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB*, Statistics in Practice, Wiley.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*, Springer.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the loglikelihood function in the nonlinear mixed effects model, *Journal of Computational and Graphical Statistics* **4**: 12–35.

Plummer, M. (2003). JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rue, H. and Martino, S. (2007). Approximate bayesian inference for hierarchical gaussian markov random field models, *Journal of Statistical Planning and Inference* **137**(10): 3177 – 3192. Special Issue: Bayesian Inference for Stochastic Processes.

Rue, H., Martino, S. and 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. and Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* **81**(393): 82–86.

Torabi, M. (2014). Likelihood inference for spatial generalized linear mixed models, *Communications in Statistics - Simulation and Computation* **0**(ja): null.

Varin, C., Hst, G. and ivind Skare (2005). Pairwise likelihood inference in spatial generalized linear mixed models, *Computational Statistics Data Analysis* **49**(4): 1173 – 1191.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models, *Biometrics* **58**(1): 129136.