Likelihood inference for spatial generalized linear mixed models

Mahmoud Torabi *†

Spatial modeling is widely used in environmental sciences, biology and epidemiology. Generalized linear mixed models are employed to account for spatial variations of point-referenced data called spatial generalized linear mixed models (SGLMMs). Frequentist analysis of these type of data is computationally difficult. On the other hand, the advent of the Markov chain Monte Carlo algorithm has made the Bayesian analysis of SGLMM computationally convenient. Recent introduction of the method of data cloning, which leads to maximum likelihood estimate, has made frequentist analysis of mixed models also equally computationally convenient. Recently, the data cloning was employed to estimate model parameters in SGLMMs, however, the perdiction of spatial random effects and kriging are also very important. In this paper, we propose a frequentist approach based on data cloning to predict (and provide prediction intervals) spatial random effects and kriging. We illustrate this approach using a real dataset and also by a simulation study.

Keywords: Bayesian computation; generalized linear mixed model; kriging; point-referenced data; spatial statistics

*Correspondence to: Mahmoud Torabi, Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, R3E 0W3, Canada. E-mail: torabi@cc.umanitoba.ca

†Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

1 Introduction

linear mixed models (GLMMs) (Breslow and Clayton [3]) are obtained from generalized linear models (GLMs) (McCullagh and Nelder [14]) by incorporating random effects into the linear predictors, and include the well-known linear mixed models (LMMs) for normal responses (Laird and Ware [11]) as a special case. These models are useful for modeling the dependence among response variables inherent in longitudinal or repeated measures studies, for accommodating overdispersion among binomial or Poisson responses, and for producing shrinkage estimators in multiparameter problems, such as the construction of maps of small area disease rates (Breslow and Clayton [3]; Rao [17]; Torabi [21, 23, 24]; Torabi and Rosychuk [26, 27, 28]).

many practical applications such as biology, environmental sciences and health sciences, conventional geostatistical (also called point-referenced) methods such as kriging may be used in making inference in estimation and prediction for a random field which is Gaussian. One may also encounter discrete or continuous data for these methods. The GLMMs with spatially correlated random effects are then employed. The random effects for the geostatistical data are usually considered to be a spatial process with a zero mean Gaussian random field (GRF).

In making inference for such spatial models which fall in the class of GLMMs, there have been many attempts in both frequentist and Bayesian approaches. A potential difficulty in making inference in GLMMs is that a full-likelihood analysis is burdened by often intractable numerical integration (McCulloch [15]). Parameter estimation under GLMM is then extremely difficult using the frequentist approach. Hence, many approximate methods have been developed in past two decades. Breslow and Clayton [3] considered an approximate approach called penalized quasi-likelihood (PQL) for inference in SGLMM. To analyze a binary spatial data in a spatial probit model, Heagerty and Lele [9] used a

pairwise likelihood approach. Zhang [32] used an expected-maximization (EM) algorithm using Markov chain Monte Carlo (MCMC) to obtain maximum likelihood estimation (MLE) for SGLMMs. Varin *et al.* [30] used a pairwise likelihood approach using a new EM-type algorithm which utilizes numerical quadrature to making inference in SGLMMs.

The Bayesian approach, especially the non-informative Bayesian approach, has become quite popular because of its computational convenience. Diggle *et al.* [5] used a Bayesian approach using MCMC algorithm to making inference in SGLMMs. Christensen and Waagepetersen [4] used an alternate MCMC algorithm using Langevin-Hastings approach in making inference based on Bayesian method. A Bayesian approach in fitting a general design of SGLMM was also introduced by Zhao *et al.* [33]. Implementation of the non-informative Bayesian approach, however, requires substantial care. The inferences may also depend on the choice of the prior (Efron [6]; Jonson [10]).

Recently, Lele *et al.* [12] introduced an alternative approach, called data cloning (DC), to compute the ML estimates and their standard errors for general hierarchical models. Similar to the Bayesian approach, data cloning avoids high dimensional numerical integration and requires neither maximization nor differentiation of a function. Extending this work to the GLMM situation, Lele *et al.* [13] described an approach to compute prediction and prediction intervals for the random effects. Torabi [22] also extended the DC approach to the GLMM with two components of dispersion. Torabi [25] used the DC approach in the context of GLMM with measurement error in covariates. The data cloning approach was also extended to cross-sectional and time-series data in the context of small area estimation (Torabi and Shokoohi [29]). The method of DC was applied to spatial Poisson mixed model with areal data (Torabi [23]). Torabi [24] also used the DC approach for cluster detection. Baghishani and Mohammazadeh [1] extended the DC approach to SGLMMs with point-referenced data to estimate the model parameters. However, the main issues in SGLMMs are also to predict the spatial random effects and kriging which

are extremely difficult from the frequentist perspective. The DC approach, thus, is well suited to offer a frequentist approach in the context of SGLMMs.

In this paper, we use DC in the context of GLMMs with spatially random effects for geostatistical data. We first describe the SGLMM in general form (Section 2). We, then, describe how DC can be used to analyze SGLMMs in terms of estimation of model parameters and also prediction of random effects and kriging (Section 3). In Section 4, the performance of DC, which yields to maximum likelihood estimation, is evaluated by using a real dataset and also by a simulation study. Finally, some concluding remarks are given in Section 5.

2 Spatial generalized linear mixed model

The basic model in spatial statistics can be described as follows. Let $Y(s_i)$ be the variable of interest at location s $(i = 1, ..., n)$. Assume the observations $Y(s_i)$ are conditionally independent, given β and $u(s_i)$, with exponential family p.d.f.

$$f\big(y(s_i)|\beta, u(s_i)\big) = \exp\Big\{\big(y(s_i)\eta(s_i) - a(\eta(s_i))\big)/\phi(s_i) + c\big(y(s_i), \phi(s_i)\big)\Big\}. \tag{1}$$

The density (1) is parameterized with respect to the canonical parameters $\eta(s_i)$, known scale parameters $\phi(s_i)$ and functions $a(\cdot)$ and $c(\cdot)$. The exponential family (1) covers well-known distributions including Gaussian, binomial and Poisson. We can then model

$$g(\eta(s_i)) = X'(s_i)\beta + u(s_i), \tag{2}$$

for some link function g, where $X(s_i)$ and β are observed covariates and unknown regression coefficients with both as vectors with dimension p, respectively. We presume the $u(s_i)$ to be spatial random effects coming from a Gaussian process. That is, the

$\boldsymbol{u}(\boldsymbol{s}) = (u(s_1), ..., u(s_n))'$ are assumed to be realizations from a zero-centered stationary Gaussian spatial process. Thus, the $\boldsymbol{u}(\boldsymbol{s})$ introduce the partial sill (σ^2) and range (ϕ) parameters. The second-stage specification is then $\boldsymbol{u}(\boldsymbol{s}) \sim N(0, \sigma^2 \boldsymbol{H}(\phi))$ where $\boldsymbol{H}$ is a correlation matrix with $\boldsymbol{H}_{ij} = \rho(s_i - s_j; \phi)$ and $\rho(\cdot)$ is a valid isotropic correlation function on $\Re^2$ indexed by a parameter ϕ.

The joint distribution $f\big(y(s_1), ..., y(s_n)\big)$ of the SGLMM is then given by

$$L(\boldsymbol{\alpha}; \boldsymbol{y}) := \int \prod_{i=1}^{n} f\Big(y(s_i)|\beta, u(s_i)\Big) h\big(\boldsymbol{u}(\boldsymbol{s})|\sigma^2, \phi\big) d\boldsymbol{u}(\boldsymbol{s}), \tag{3}$$

where $\boldsymbol{\alpha} = (\beta, \sigma^2, \phi)$ and $h(\cdot)$ is a Gaussian distribution. To estimate the model parameters $\boldsymbol{\alpha}$, we need to maximize the likelihood function in (3), however, it is clear from (3) that the marginal likelihood function involves intractable integrals. The computational burden also increases with the number of observations as the dimension of the random field is equal to the number of observations.

3 Frequentist inference using data cloning

For simplicity, we use y_i and $\boldsymbol{u}$ instead of $y(s_i)$ and $\boldsymbol{u}(\boldsymbol{s})$. Let $\boldsymbol{y} = (y_1, ..., y_n)'$ be the observed data vector and, conditionally on the random effects, $\boldsymbol{u}$, assume that the elements of $\boldsymbol{y}$ are independent and drawn from a distribution belongs to exponential family with parameter $\boldsymbol{\alpha}_1 = \beta$. It is also assumed that distribution for $\boldsymbol{u}$ depends on parameters $\boldsymbol{\alpha}_2 = (\sigma^2, \phi)$:

$$\boldsymbol{y}_i|\boldsymbol{u} \sim f_{\boldsymbol{y}_i|\boldsymbol{u}}(\boldsymbol{y}_i|\boldsymbol{u}, \boldsymbol{\alpha}_1)$$

$$\boldsymbol{u} \sim h_{\boldsymbol{u}}(\boldsymbol{u}|\boldsymbol{\alpha}_2), \tag{4}$$

where $f(\cdot)$ and $h(\cdot)$ are appropriate distributions. The goal of the analysis is to estimate the model parameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)'$ and predict the random effects $\boldsymbol{u}$ or its variant.

To illustrate the DC approach, we start with standard Bayesian approach to inference for hierarchical models. Denote $\pi(\boldsymbol{\alpha})$ as prior distribution on the parameter space. The posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ is given by

$$\pi(\boldsymbol{\alpha}|\boldsymbol{y}) = \frac{L(\boldsymbol{\alpha};\boldsymbol{y})\pi(\boldsymbol{\alpha})}{C(\boldsymbol{y})}, \tag{5}$$

where $C(\boldsymbol{y}) = \int L(\boldsymbol{\alpha};\boldsymbol{y})\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. There are computational tools, MCMC algorithms, that facilitate generation of random variates from the posterior distribution $\pi(\boldsymbol{\alpha}|\boldsymbol{y})$ without computing the integrals in the numerator or the denominator of (5)(Gilks *et al.* [7]; Spiegelhalter *et al.* [20]).

The DC method uses the Bayesian computational approach for frequentist purposes. To understand the idea in DC, imagine a hypothetical situation where the observations $\boldsymbol{y}$ are repeated independently by K different individuals, and all these individuals happen to result in exactly the same set of observations $\boldsymbol{y}$ called $\boldsymbol{y}^{(K)} = (\boldsymbol{y}, \boldsymbol{y}, ..., \boldsymbol{y})$. The posterior distribution of $\boldsymbol{\alpha}$ conditional on the data $\boldsymbol{y}^{(K)}$ is then given by

$$\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)}) = \frac{\{L(\boldsymbol{\alpha};\boldsymbol{y})\}^K\pi(\boldsymbol{\alpha})}{C(\boldsymbol{y}^{(K)})}, \tag{6}$$

where $C(\boldsymbol{y}^{(K)}) = \int\{L(\boldsymbol{\alpha};\boldsymbol{y})\}^K\pi(\boldsymbol{\alpha})d\boldsymbol{\alpha}$ is the normalizing constant. The expression $\{L(\boldsymbol{\alpha};\boldsymbol{y})\}^K$ is the likelihood for K copies of the original data. Walker [31] and Lele *et al.* [12, 13] showed that, for K large enough, $\pi_K(\boldsymbol{\alpha}|\boldsymbol{y}^{(K)})$ converges to a multivariate Gaussian distribution with mean equal to the MLE of the model parameters and variance-covariance matrix equal to $1/K$ times the inverse of the Fisher information matrix for the MLE. Hence, this distribution is nearly degenerated at the MLE $\boldsymbol{\alpha}$ for large K. Moreover, the

sample mean vector of the generated random numbers provides the MLE of the model parameters, and K times their sample variance-covariance matrix is an estimate of the asymptotic variance-covariance matrix for the MLE $\hat{\boldsymbol{\alpha}}$.

Lele *et al.* [13] also provided various checks to determine the adequate number of clones. For instance, one may plot the largest eigenvalue of the posterior variance as a function of the number of clones K to determine if the posterior distribution has become nearly degenerate (say < 0.05). As another criterion, it is approximately true that as we increase the number of clones K,

$$(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})' \boldsymbol{V}^{-1} (\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}) \sim \chi^2_q, \tag{7}$$

where $\boldsymbol{V}$ is the variance of the posterior distribution and q is the dimension of $\boldsymbol{\alpha}$. One may compute the following two statistics: a) $\zeta = \frac{1}{B} \sum_{b=1}^{B} (O_b - Q_b)^2$, where O_b and Q_b are observed and estimated quantiles for χ^2_q random variable, and b) $\tilde{r}^2 = 1 - \gamma^2$, where γ is the correlation between (O, Q). If these statistics are close to zero (say < 0.01), it indicates that the approximation (7) is reasonable.

3.1 Prediction of random effects and kriging

Prediction of random effects, particularly from the frequentist viewpoint, is usually problematic. If the parameters $\boldsymbol{\alpha}$ are known, then one can clearly use the conditional distribution of $\boldsymbol{u}$, the latent variables, given the observed data. That is, one can use $\pi(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\alpha}^*)$ where $\boldsymbol{\alpha}^*$ is the true value of the parameter. A naive approach, when $\boldsymbol{\alpha}$ is estimated using the data, is to use $\pi(\boldsymbol{u}|\boldsymbol{y}, \hat{\boldsymbol{\alpha}})$. However, this approach does not take into account the variability introduced by the model parameters estimate. An approach that has been suggested in the literature (e.g., Hamilton [8], Lele *et al.* [13]) to take into account the

variation of the estimator is to use the density:

$$\pi(\boldsymbol{u}|\boldsymbol{y}) = \frac{\int f(\boldsymbol{y}|\boldsymbol{u},\boldsymbol{\alpha}_1)h(\boldsymbol{u}|\boldsymbol{\alpha}_2)\zeta(\boldsymbol{\alpha},\hat{\boldsymbol{\alpha}},I^{-1}(\hat{\boldsymbol{\alpha}}))d\boldsymbol{\alpha}}{C(\boldsymbol{y})}, \tag{8}$$

where $\zeta(.,\mu,\sigma^2)$ denotes Gaussian density with mean μ and variance σ^2, which are equal to the MLE and the inverse of the Fisher information matrix here. In this paper, we obtain prediction (and prediction interval which is equally tailed) of the $\boldsymbol{u}$ using the density in equation (8) along with MCMC sampling. We can use the same approach for prediction of some function of $\boldsymbol{u}$, e.g., $g(\eta(\boldsymbol{s}))$, or $\eta(\boldsymbol{s})$ for some new locations within the study region which is called *kriging*.

In this paper, for the DC analysis, the proper priors are used for fixed effects and variance component. In particular, the independent Gaussian distribution is assigned for fixed effects with zero mean and variance 10^6, and gamma distribution for the inverse of variance component with shape and scale parameter 0.001, and a uniform (0,10) prior for ϕ. Since the DC is invariant to the priors, one may use different priors, however, if the appropriate priors are used, the convergence will be achieved sooner. To monitor the convergence of the model parameters, we used several diagnostic methods such as Gelmen and Rubin convergence diagnostic, Geweke convergence diagnostic, and Heidelberger and Welch convergence diagnostic, implemented in the Bayesian output analysis (BOA) program (Smith [18]), a freely available package created for R [16]. We also used diagnostic methods implemented in the dclone package (Sólymos [19]), which were described in Section 3, to monitor the convergence of the model parameters in terms of number of clones K (Lele *et al.* [13]). The program is available from the author upon request.

Table 1. Parameters estimate and standard errors for the binary spatial mixed model using MLE approach via data cloning.

Parameter	Estimate	Standard error
Intercept	-0.027	0.137
Living area	0.016	0.072
Age	-0.0001	0.003
σ^2	0.029	0.102
ϕ	5.204	26.03

4 Application

4.1 Data analysis

We now evaluate the method of DC by applying to a real dataset. We consider a real estate dataset with observations at 50 locations in Baton Rouge, LA (Banerjee *et al.* [2]). The variable of interest is a binary variable, with $Y(s) = 1$ indicating that the price of the property at location s is "high" (above the median price for the region), and $Y(s) = 0$ indicating that the price is "low". The model (1)-(2) with g as a logit function, is used for our dataset where observed covariates are total living area and house's age. Table 1 reports the model parameters estimate and corresponding standard errors. The covariates effects are generally uninteresting.

For this application, the number of clones was $K = 100$ to obtain MLE, and the number of iterations for convergence of the model parameters in DC was about 30,000. As mentioned in Section 3, if scaled variances are decreased at a $1/K$ rate and have reached a lower bound (say < 0.05), the DC approach has converged (Figure 1); noting that we also checked the other criteria and observed no convergence issue.

One of the main features of DC is the ability to predict the random effects. Figure 2 shows the image plot with overlaid contour lines for the surface estimate of the latent spatial process $\boldsymbol{u}(\boldsymbol{s})$. The image plot shows lower prices in the northern region, and

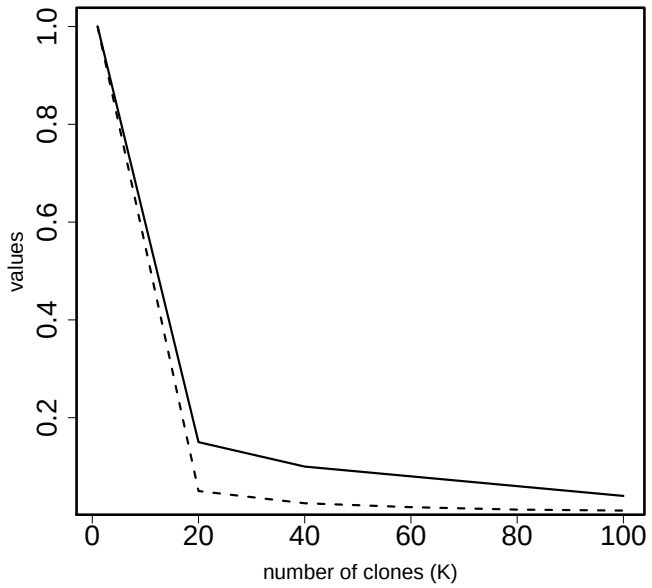

Figure 1: Data cloning convergence diagnostics for real estate dataset. The standardized eigenvalues (solid line) converge to zero at the expected rate (dashed line).

generally higher prices in the south-central region, although the southeast region shows some lower price zones as well. The distribution of the contour lines indicate smooth flat stretches across the central parts, with downward slopes toward the north and southeast.

We can also use DC for kriging of the values of random fields in our study region. One may be interested to know how expensive would be in some regions? For instance, our interest it to predict the expensiveness of two locations in Baton Rouge, $x_0 = (4.7, 4.3)$ and $y_0 = (-1.0, 0.8)$ as longitude and latitude, with living areas and houses' ages as Living area= $(1600, 1000)$ and Age= $(7, 13)$. Our prediction with corresponding standard errors are then $\hat{p}_0 = (0.506, 0.494)$ and $se(\hat{p}_0) = (0.062, 0.061)$, respectively.

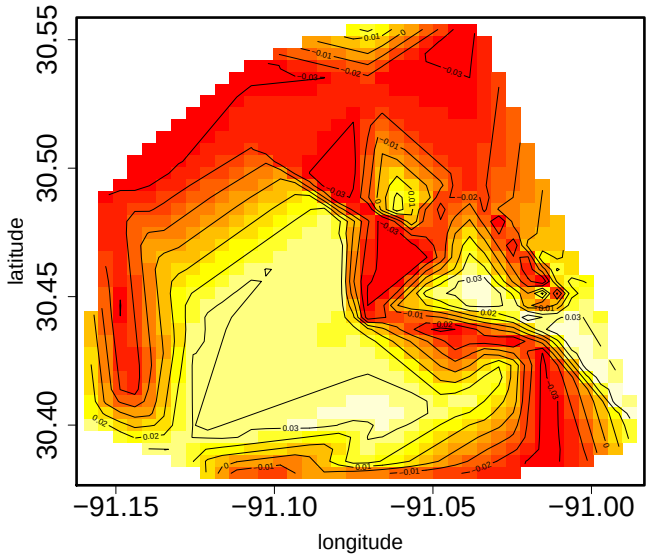

Figure 2: Image plot of the prediction of the latent spatial process $\boldsymbol{u}(\boldsymbol{s})$ for binary spatial mixed model.

4.2 Simulation study

We also conduct a simulation study to evaluate the performance of ML estimates, via DC approach, using a scenario as in Baghishani and Mohammadzadeh [1]. In particular, we consider a spatial Poisson count data with equally spaced regular grids of locations of 10×10. To implement the simulation study, we first generate $R = 1,000$ independent samples from a Poisson distribution:

$$y_i^{(r)} \sim Poisson(\mu_i^{(r)})$$

$$\log(\mu_i^{(r)}) = \beta x_i + u_i^{(r)} \qquad (i = 1, ..., n; r = 1, ..., R), \tag{9}$$

where x_i is the first component of ith location s_i and $u_i^{(r)}$ is generated from a zero mean GRF with isotropic exponential covariogram $C(s - s'; \sigma^2, \phi) = \sigma^2 \exp(-||s - s'||/\phi)$, where $||s - s'||$ refers to the Euclidean distance between s and s'. We set $(\beta, \sigma^2, \phi) =$

$(0.5, 1.25, 3.0)$.

Using the simulated datasets $\{y_i^{(r)}; i = 1, ..., n\}$, we compute the model parameters estimate from (9) for each simulation run r. Table 2 presents the mean of the parameters estimators. We also want to explore if the reported standard errors properly represent the true standard deviations. Towards this goal, we report the true standard error of the estimated parameters and mean values of the estimated standard errors. It seems that the estimators of β and σ^2 are unbiased and their standard errors are also estimated very well. However, the parameter ϕ shows some departure from its true value. Overall, it seems that DC approach performs well in terms of point estimates and standard errors in our simulation set-up.

Surprisingly, our results are different from Baghishani and Mohammadzadeh [1]. In particular, the model parameters (β, σ^2, ϕ) estimates and corresponding (standard errors) for our approach are $0.498(0.020), 1.234(0.205)$ and $3.684(1.568)$ while these values in [1] are $0.400(0.067), 1.563(0.477)$ and $3.723(0.460)$; noting that the results of [1] are based on 800 clones with priors $U(0, 9), U(0.1, 4)$ and $U(0.1, 5)$ for β, σ^2, and ϕ, respectively. In our algorithm, unlike Baghishani and Mohammadzadeh [1], we increased the number of clones (K) for each simulation run r until convergence is achieved. For this simulation set-up, the average number of clones was $K = 10$ to obtain MLE, and the average number of iterations for convergence of the model parameters was about $25,000$. Note that Baghishani and Mohammazadeh [1] reported the results for 100, 200, 400, and 800 clones, however, they did not have any criteria to evaluate the convergence of DC to see how many clones they need to reach MLE. On the other hand, we evaluated the convergence of DC for our simulation study based on the criteria described in Section 3.

We now turn to evaluate the performance of prediction of spatial random effects. To that end, for each simulation run r, the posterior mean $\hat{\mu}_i^{(r)}$ is computed using the model

Table 2. Mean values of parameters estimates and corresponding standard errors, and simulated standard errors of MLE based on the spatial Poisson mixed model (9).

Parameter	Estimate	Standard error	
		Simulated	MLE
$\beta = 0.50$	0.498	0.019	0.020
$\sigma^2 = 1.25$	1.234	0.205	0.205
$\phi = 3.0$	3.684	0.834	1.568

parameters estimated by DC approach. The true mean squared error (TMSE) of $\hat{\mu}_i$ is approximated as $\text{TMSE}(\hat{\mu}_i) = R^{-1}\sum_{r=1}^{R}(\hat{\mu}_i^{(r)} - \mu_i^{(r)})^2, (i = 1, ..., n)$. We then calculate the relative bias (RB) of an estimator of the MSE, say mse, as

$$\text{RB}[\text{mse}(\hat{\mu}_i)] = \Big\{\frac{1}{R}\sum_{r=1}^{R}\text{mse}^{(r)}(\hat{\mu}_i) - \text{TMSE}(\hat{\mu}_i)\Big\}/\text{TMSE}(\hat{\mu}_i),$$

where $\text{mse}^{(r)}(\hat{\mu}_i)$ is the value of posterior variance of $\hat{\mu}_i$ for the rth simulation study. The average absolute RB values over locations for the DC method is 6.8% which shows that the DC approach accounts the variability of spatial random effects very well.

We also study the coverage probabilities and average lengths of prediction intervals of the $\hat{\mu}_i$ for different confidence coefficients. We compute the average coverage probabilities and average lengths of the estimates of the spatial means $(\hat{\mu}_i)$ over location i (Table 3). The DC method performs very well in terms of coverage probabilities of the spatial means for different confidence coefficients.

We now use the method of DC for kriging of the values of random fields with also providing corresponding standard errors (or prediction intervals). For instance, our interest is to predict random fields $\mu_j = \beta x_j + u(s_j), (j = 1, ..., 5)$, in our simulation study, for five locations, $x_0 = (4.7, 3.3, 1.9, 7.5, 4.3)$ and $y_0 = (1.0, 1.6, 2.2, 3.1, 4.8)$ as longitude and latitude, respectively. Table 4 shows the average prediction of μ_j with corresponding standard errors over 1000 simulation runs.

Table 3. Average coverage probabilities and average lengths of the spatial means ($\hat{\mu}_i$) over locations with different confidence coefficients in the spatial Poisson mixed model (9).

Confidence coefficient	Average coverage probabilities	Average lengths
0.90	0.897	18.877
0.95	0.948	22.546
0.98	0.979	26.811
0.99	0.989	29.713

Table 4. Prediction of spatial means ($\hat{\mu}_i$) and corresponding standard errors for five new locations in the spatial Poisson mixed model (9).

Location	Prediction	Standard error
1	10.381	3.043
2	5.134	3.307
3	2.580	2.907
4	42.224	3.251
5	8.593	3.168

5 Conclusion

Often, for fitting complex models in the context of spatial statistics, Bayesian methods are advocated because they are computationally more convenient than the likelihood-based methods. Analysis based on data cloning (DC) overcomes the computational difficulties of the maximum likelihood method.

Using DC, we have presented and illustrated spatial generalized linear mixed models (SGLMMs). Baghishani and Mohammazadeh [1] also used the DC approach in SGLMMs by only providing model parameters estimate and corresponding standard errors. However, it is also very important to provide prediction of spatial random effects and kriging in the SGLMMs. We used the method of DC, which yields to maximum likelihood estimation, to provide prediction of spatial random effects and kriging in the SGLMMs. We evaluated the performance of DC in the SGLMMs, using a real dataset for binary spatial mixed model and also by a simulation study for spatial Poisson mixed model, in terms

of model parameters estimates, prediction of spatial random effects, and kriging.

Acknowledgements

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

[1] H. Baghishani, and M. Mohammadzadeh, *A data cloning algorithm for computing maximum likelihood estimates in spatial generalized linear mixed models*, Comput. Stat. Data An. 55 (2011), pp. 1748–1759.

[2] S. Banerjee, B.P. Carlin, and A.E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, London/ Boca Raton, FL: Chapman and Hall/CRC, 2004.

[3] N.E. Breslow, and D.G. Clayton, *Approximate inference in generalized linear mixed models*, J. Am. Statist. Assoc. 88 (1993), pp. 9–25.

[4] O.F. Christensen, and R.P. Waagepetersen, *Bayesian prediction of spatial count data using generalized linear mixed models*, Biometrics 58 (2002), pp. 280–286.

[5] P. Diggle, J.A. Tawn, and R.A. Moyeed, *Model-based geostatistic*, J. Appl. Statist. 47 (1998), pp. 299–350.

[6] B. Efron, *Why isn't everyone a Bayesian?*, Am. Statist. 40 (1986), pp. 1–5.

[7] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice*, New York: Springer - Verlag, 1996.

[8] J.D. Hamilton, *A standard error for the estimated state vector of a state-space model*, J. Economet. 33 (1986), pp. 387–397.

[9] P. Heagerty, and S.R. Lele, *A composite likelihood approach to binary data in space*, J. Am. Statist. Assoc. 93 (1998), pp. 1099–1111.

[10] N.E.G. Jonson, *Everyday diagnostics - a critique of the Bayesian model*, Med. Hypotheses 34 (1991), pp. 289–295.

[11] N.M. Laird, and J.H. Ware, *Random-effects models for longitudinal data*, Biometrics 38 (1982), pp. 963–974.

[12] S.R. Lele, B. Dennis, and F. Lutscher, *Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods*, Ecol. Lett. 10 (2007), pp. 551–563.

[13] S.R. Lele, K. Nadeem, and B. Schmuland, *Estimability and likelihood inference for generalized linear mixed models using data cloning*, J. Am. Statist. Assoc. 105 (2010), pp. 1617–1625.

[14] P. McCullagh, and J.A. Nelder, *Generalized Linear Models*, 2nd ed., London: Chapman and Hall, 1989.

[15] C.E. McCulloch, *Maximum likelihood algorithms for generalized linear mixed models*, J. Am. Statist. Assoc. 92 (1997), pp. 162–170.

[16] R Development Core Team: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.R-project.org, 2012.

[17] J.N.K. Rao, *Small Area Estimation*, New Jersey: Wiley, 2003.

[18] B.J. Smith, *BOA User Manual (version 1.1.7)*, Department of Biostatistics, College of Public Health, University of Iowa, Ames, 2007.

[19] P. Sólymos, *dclone: data cloning in R*, R J. 2 (2010), pp. 29–37.

[20] D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn, *WinBUGS version 1.4 User Manual.* MRC Biostatistics unit, Institute of Public Health, London, 2004.

[21] M. Torabi, *Hierarchical Bayes estimation of spatial statistics for rates*, J. Statist. Plann. Infer. 142 (2012a), pp. 358–365.

[22] M. Torabi, *Likelihood inference in generalized linear mixed models with two components of dispersion using data cloning*, J. Comput. Statist. Data An. 56 (2012b), pp. 4259–4265.

[23] M. Torabi, *Spatial modeling using frequentist approach for disease mapping*, J. Appl. Statist. 39 (2012c), pp. 2431–2439.

[24] M. Torabi, *Spatial disease cluster detection: an application to childhood asthma in Manitoba, Canada*, J. Biometrics and Biostatist. (2012d), S7:010. doi:10.4172/2155-6180.S7-010.

[25] M. Torabi, *Likelihood inference in generalized linear mixed measurement error models*, J. Comput. Statist. Data An. 57 (2013), pp. 549–557.

[26] M. Torabi, and R.J. Rosychuk, *Spatio-temporal modelling of disease mapping of rates*, Canad. J. Statist. 38 (2010), pp. 698–715.

[27] M. Torabi, and R.J. Rosychuk, *Spatio-temporal modeling using spline for disease mapping: analysis of childhood cancer trends*, J. Appl. Statist. 38 (2011), pp. 1769–1781.

[28] M. Torabi, and R.J. Rosychuk, *Hierarchical Bayesian spatio-temporal analysis of childhood cancer trends*, J. Geogr. An. 44 (2012), pp. 109–120.

[29] M. Torabi, and F. Shokoohi, *Likelihood inference in small area estimation by combining time-series and cross-sectional data*, J. Mult. An. 111 (2012), pp. 213–221.

[30] C. Varin, H. Gudmund, and O. Skare, *Pairwise likelihood inference in spatial generalized linear mixed models*, Comput. Statist. Data An. 49 (2005), pp. 1173–1191.

[31] A.M. Walker, *On the asymptotic behavior of posterior distributions*, J. Roy. Statist. Soc. Ser. B, 31 (1969), pp. 80–88.

[32] H. Zhang, *On estimation and prediction for spatial generalized linear mixed models*, Biometrics 58 (2002), pp. 129–136.

[33] Y. Zhao, J. Staudenmayer, B.A. Coull, and M.P. Wand, *General design Bayesian generalized linear mixed models with applications to spatial statistics*, Stat. Sci. 21 (2006), pp. 35–51.