11-1-2003

Fitting Generalized Linear Mixed Models For Point-Referenced Spatial Data

Armin Gemperli
Swiss Tropical Institute

Penelope Vounatsou
Swiss Tropical Institute

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Fitting Generalized Linear Mixed Models For Point-Referenced Spatial Data

Fitting Generalized Linear Mixed Models For Point-Referenced Spatial Data

Armin Gemperli   Penelope Vounatsou
Swiss Tropical Institute
Basel, Switzerland

Non-Gaussian point-referenced spatial data are frequently modeled using generalized linear mixed models (GLMM) with location-specific random effects. Spatial dependence can be introduced in the covariance matrix of the random effects. Maximum likelihood-based or Bayesian estimation implemented via Markov chain Monte Carlo (MCMC) for such models is computationally demanding especially for large sample sizes because of the large number of random effects and the inversion of the covariance matrix involved in the likelihood. We review three fitting procedures, the Penalized Quasi Likelihood method, the MCMC, and the Sampling-Importance-Resampling method. They are assessed in terms of estimation accuracy, ease of implementation, and computational efficiency using a spatially structured dataset on infant mortality from Mali.

Key words: Geostatistics, infant mortality, kriging, Markov chain Monte Carlo (MCMC), penalized quasi likelihood (PQL), risk mapping, sampling-importance-resampling (SIR)

Introduction

Point referenced spatial data arise from observations collected at geographical locations over a fixed continuous space. Proximity in space introduces correlations between the observations rendering the independence assumption of standard statistical methods invalid. Ignoring spatial correlation will result in underestimation of the standard error of the parameter estimates, and therefore liberal inference as the null hypothesis is rejected too often. A wide range of analytical tools within the field of geostatistics have been developed concerning with the description and estimation of spatial patterns, the modeling of data in the presence of spatial correlation and the kriging, that is the spatial prediction at unobserved locations.

Statistical inference of point referenced data often assumes that the observations arise from a Gaussian spatial stochastic process and introduce covariate information and possibly trend surface specification on the mean structure while spatial correlation on the variance-covariance matrix Σ of the process. Under second order stationarity, Σ determines the well-known variogram. When isotropy is also assumed, the elements of Σ are modeled by parametric functions of the separation between the corresponding locations. For non-Gaussian data, the spatial correlation is modeled on the covariance structure of location-specific random effects introduced into the model and assumed to arise from a Gaussian stationary spatial process.

For Gaussian data, the generalized least squares (GLS) approach can be used iteratively to obtain estimates $\hat{\boldsymbol{\beta}}$ of the regression coefficients conditional on the covariance parameters. The covariance parameters $\boldsymbol{\theta}$ can be estimated conditional on $\hat{\boldsymbol{\beta}}$ by fitting the semivariogram empirically or by maximum likelihood or restricted maximum likelihood methods (Zimmerman and Zimmerman, 1991).

Statistical estimation for non-Gaussian data is based on the theory of generalized linear mixed models (GLMM). A common approach is

to integrate out the random effects and proceed with maximum likelihood based approaches for estimating the covariate and covariogram parameters. This integration can be implemented numerically (Anderson and Hinde, 1998; Preisler, 1988; Lesaffre and Spiessens, 2001) when dimensionality is low or via approximations. Breslow and Clayton (1993) show, that for known covariance parameters, the Laplace approximation leads to the same estimator for the fixed and random effects parameters as the one arising by maximizing the penalized quasi-likelihood (PQL). Implementation of this approach requires iterating between iterated weighted least squares for estimating the fixed and random effects parameters and maximizing the profile likelihood for estimating the covariance parameters. An extension of the PQL procedure is discussed by Wolfinger and O'Connell (1993). The PQL approach is implemented in some statistical packages due to its relative simplicity, however it provides biased estimates when the number of random effects increases (McCulloch, 1997; Booth and Hobert, 1999) or when the data are far from normal.

The generalized estimating equation methods developed by Liang and Zeger (1986) and Zeger and Liang (1986) estimate covariate effects under the assumption of independence, but correct their standard error to account for the spatial dependence. The method is unable to estimate the spatial random effects. The EM algorithm (Dempster, Laird and Rubin, 1977) has been implemented in model fit by treating the spatial random effects as "missing" data. The intractable integration of the random effects which is required in the E-step is overcome by simulation, such as Metropolis-Hastings algorithm (McCulloch, 1997) or importance sampling/rejection sampling method (Booth and Hobert, 1999). For spatial settings, particular Pseudo-Likelihood approaches have been established which capture solely the site to site variation between pairs or groups of observations (Besag, 1974). For the special case of a binary outcome, Heagerty and Lele (1998) have proposed a thresholding model using a composite likelihood approach.

A drawback of the maximum likelihood-based methods employed in geostatistical modeling is the large sample asymptotic inference. For a spatial stochastic process $\{\mathbf{Y}(\mathbf{u}); \mathbf{u} \in D\}$, with $D \subset R^2$ the asymptotic concept can be applied either to the sample size within a fixed space D (infill asymptotics) or to the space D (increasing domain asymptotics). In the latter, observations are spaced far enough to be considered uncorrelated. The results can differ, depending on the type of asymptotics used (see e.g. Tubilla, 1975).

Bayesian hierarchical geostatistical models implemented via Monte Carlo methods avoid asymptotic inference as well as many computational problems in model fitting and prediction. Diggle et al. (1998) suggest inference on the posterior density via Markov chain Monte Carlo (MCMC). This iterative approach requires repeated inversions of the covariance matrix of the spatial process, which is involved in the likelihood. The size of this matrix increases with the number of locations. Inversion of large matrices can drastically slow down the running time of the algorithm and cause numerical instabilities affecting the accuracy of the estimates. To overcome this problem Gelfand et al. (1999) suggest non-iterative simulation via the Sampling-Importance-Resampling (SIR) algorithm (Rubin, 1987). The quality of SIR hinge on the ability to formulate an easy-to-draw-from importance-density, which comes as close as possible to the true joint posterior distribution of the parameters.

In this article, we review three fitting procedures; the maximum likelihood-based PQL method, the MCMC simulation and the SIR. We assess these methods in terms of estimation accuracy, ease of implementation and computational efficiency using a spatially structured dataset on infant mortality from Mali collected over 181 locations. A description of the dataset and the applied questions which motivated this work are given in the next section. Then we describe the model as well as the three fitting approaches. A discussion on the ease of implementation of each approach and a comparison of the inferences obtained is given in the conclusion section.

Data

The data that motivated this work were collected under the Demographic and Health Surveys (DHS) program. The aim of the program is to collect and analyze reliable demographic and health data for regional and national family and health planning. Data are commonly collected in developing countries. DHS is funded by the U.S. Agency for International Development (USAID) and implemented by Macro International Inc. The standard DHS methodology involves collecting complete birth histories from women of childbearing age, from which a record of age and survival can be computed for each child. The data are available to researchers via the internet (www.measureDHS.com).

Birth histories corresponding to 35,906 children were extracted from the data of the DHS-III 1995/96 household survey carried out in Mali. Additional relevant covariates extracted were the year of birth, residence, mothers education, infant's sex, birth order, preceding birth interval and mothers age at birth. Using location information provided by Macro International, we were able to geo-locate 181 distinct sites by using digital maps and databases, such as the African data sampler (1995) and the Geoname Gazetteer (1995). The objective of data analysis was to assess the effect of birth and socio-economic parameters on infant mortality and produce smooth maps of mortality risk in Mali. These maps will help identifying areas of high mortality risk and assist child mortality intervention programs.

Methodology

Let Y_{ij} be a binary response corresponding to the mortality risk of child j at site s_i, $i = 1, \ldots, n$ taking value 1 if the child survived the first year of life and 0 otherwise, and let $\mathbf{X}_{ij}$ be the vector of associated covariates. Within the generalized linear model framework (GLM), we assume Y_{ij} are i.i.d. Bernoulli random variables with $E(Y_{ij}) = \pi_{ij}$ and model predictors as $g(\pi_{ij}) = \mathbf{X}_{ij}^t \boldsymbol{\beta}$ where $g(\cdot)$ is a link function

such as logit in our mortality risk application. However the spatial structure of the data renders the independence assumption of Y_{ij} invalid, leading to narrower confidence intervals for $\boldsymbol{\beta}$ and thus to overestimation of the significance of the predictors.

One approach to take into account spatial dependence is via the generalized linear mixed model (GLMM) reviewed by Breslow and Clayton (1993). In particular, we introduce the unobserved spatial variation by a latent stationary, isotropic Gaussian process $\mathbf{U}$ over our study region D, such that $\mathbf{U} = (U_1, U_2, \ldots, U_n) \sim N(0, \boldsymbol{\Sigma})$, where Σ_{ij} is a parametric function of the distance d_{ij} between locations $\mathbf{s}_i$ and $\mathbf{s}_j$. Conditional on the random term U_i, we assume that Y_{ij} are independent with $E(Y_{ij} | U_i) = \pi_{ij}$. The U_i enters the model on the same scale as the predictors, that is

$$g(\pi_{ij}) = \mathbf{X}_{ij}^t \boldsymbol{\beta} + U_i \qquad (1)$$

and captures unmeasured geographical heterogeneity (small scale variation).

A commonly used parameterization for the covariance $\boldsymbol{\Sigma}$ is $\Sigma_{ij} = \sigma^2 \rho(\phi; d_{ij})$ where σ^2 is the variance of the spatial process and $\rho(\phi; d_{ij})$ a valid correlation function with a scale parameter ϕ which controls the rate of correlation decay with increasing distance. In most applications a monotonic correlation function is chosen i.e. the exponential function which has the form $\rho(\phi; d_{ij}) = \exp(-\phi d_{ij})$. Ecker and Gelfand (1997) propose several other parametric correlation forms, such as the Gaussian, Cauchy, spherical and the Bessel.

A separate set of location-specific random effects, $\mathbf{W} = (W_1, W_2, \ldots, W_n)^t$ is often added in Equation 1 to account for unexplained non-spatial variation (Diggle et al., 1998), where W_i, $i = 1, \ldots, n$ are considered to be independent, arising from a Normal distribution, $W_i \sim N(0, \tau^2)$. The τ^2 is known in

geostatistics as the nugget effect and introduces a discontinuity at the origin of the covariance function:

$$\Sigma_{ij} = \tau^2 1(i = j) + \sigma^2 \rho(\phi; d_{ij}).$$

A large number of repeated samples at the same location make the nugget identifiable, otherwise its use in the model is not justifiable because the extra binomial variation is already accounted for by the spatial random effect.

Parameter estimation

The above GLMM is highly parameterized and maximum likelihood methods can fail to estimate all parameters simultaneously. The estimation approach starts by integrating out the random effects and estimating the other parameters using the marginal likelihood

$$\int p(\mathbf{Y} \mid \mathbf{U}, \boldsymbol{\beta}, \sigma^2, \phi) \; p(\mathbf{U} \mid \sigma^2, \phi) d\mathbf{U}.$$

However, this integral has analytical solution only for Gaussian data. For non-Gaussian data the integrand can be approximated using a first-order Taylor series expansion around its maximizing value, after which the integration is feasible. This approach, known as the Laplace approximation, results in the penalized quasi-likelihood (PQL) estimator (Breslow and Clayton, 1993), which was shown in various simulation studies to produce biased results (Browne and Draper, 2000; Neuhaus and Segal, 1997). Breslow and Lin (1995) determined the asymptotic bias in variance component problems for first- and second-order approximations in comparison to McLaurin approximations.

Following the Bayesian modeling specification, we need to adopt prior distributions for all model parameters. We chose non-informative Uniform priors for the regression coefficients, i.e. $p(\boldsymbol{\beta}) \propto \mathbf{1}$, and vague inverse Gamma priors for the σ^2 and ϕ parameters: $p(\phi) = IG(a_1, b_1)$ and $p(\sigma^2) = IG(a_2, b_2)$. Bayesian inference is based on the joint posterior distribution

$$p(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi \mid \mathbf{Y}) \propto L(\boldsymbol{\beta}, \mathbf{U}; \mathbf{Y}) \times$$
$$p(\boldsymbol{\beta}) p(\mathbf{U} \mid \sigma^2, \phi) p(\sigma^2) p(\phi),$$

where $p(\mathbf{U} \mid \sigma^2, \phi)$ is the distribution of the spatial random effects, that is $p(\mathbf{U} \mid \sigma^2, \phi) \equiv N(0, \boldsymbol{\Sigma})$.

Markov chain Monte Carlo estimation

Diggle et al. (1998) suggest Markov chain Monte Carlo and in particular Gibbs sampling for fitting GLMM for point-referenced data. The standard implementation of the Gibbs algorithm requires sampling from the full conditional posterior distributions which in our application have the following forms:

$$p(\beta_k, \mid \boldsymbol{\beta}_{-k}, \mathbf{U}, \mathbf{Y}) \propto$$
$$\prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{\exp(X_{ijk}\beta_k Y_{ij})}{1 + \exp(\mathbf{X}_{ij}^t \boldsymbol{\beta} + U_i)} \qquad (2)$$

$$p(U_i \mid \mathbf{U}_{-i}, \sigma^2, \phi, \mathbf{Y}) \propto$$
$$\prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{\exp(U_i Y_{ij})}{1 + \exp(\mathbf{X}_{ij}^t \boldsymbol{\beta} + U_i)} \times$$
$$\mid \sigma^2 - \boldsymbol{\Sigma}_{i,-i} \boldsymbol{\Sigma}_{-i}^{-1} \boldsymbol{\Sigma}_{-i,i} \mid^{-1/2} \times$$
$$\exp(-\frac{1}{2}(U_i - \boldsymbol{\Sigma}_{i,-i} \boldsymbol{\Sigma}_{-i}^{-1} \mathbf{U}_{-i})^2 \times$$
$$(\sigma^2 - \boldsymbol{\Sigma}_{i,-i} \boldsymbol{\Sigma}_{-i}^{-1} \boldsymbol{\Sigma}_{-i,i})^{-1}) \qquad (3)$$

$$p(\phi \mid \mathbf{U}, \sigma^2) \propto \mid \boldsymbol{\Sigma} \mid^{-1/2} \times$$
$$\exp(-\frac{1}{2}(\mathbf{U}^t \boldsymbol{\Sigma}^{-1} \mathbf{U} + b_1/\phi))\phi^{-(a_1+1)} \qquad (4)$$

$$p(\sigma^2 \mid \mathbf{U}, \phi) \sim$$
$$\text{InverseGamma}(a_2 + n/2,$$
$$b_2 + \frac{1}{2}\mathbf{U}^t \mathbf{R}^{-1} \mathbf{U}),$$
$$R_{kl} = \rho(\phi; d_{kl}) \qquad (5)$$

where

$$\boldsymbol{\beta}_{-k} = (\beta_1, \ldots, \beta_{k-1}, \beta_{k+1}, \ldots, \beta_K)^t,$$
$$\mathbf{U}_{-i} = (U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_n)^t,$$
$$\boldsymbol{\Sigma}_{-i,i} = \boldsymbol{\Sigma}_{i,-i}^t = Cov(\mathbf{U}_{-i}, U_i) \text{ and}$$

$$\Sigma_{-i} = Cov(\mathbf{U}_{-i}, \mathbf{U}_{-i}^{t}) \,.$$

Samples from $p(\sigma^2 \mid \mathbf{U}, \phi)$ can be drawn easily as this is a known distribution. The conditionals of the other parameters do not have standard forms and a random walk Metropolis algorithm with a Gaussian proposal density having mean equal to the estimate from the previous iteration and variance derived from the inverse second derivative of the log-posterior could be employed for simulation.

The likelihood calculations in Equations 3, 4, and 5 require inversions of the $(n-1) \times (n-1)$ matrices Σ_{-i}, $i = 1, \ldots, n$ and the $n \times n$ matrix Σ, respectively. Matrix inversion is an order 3 operation, which has to be repeated for evaluating the conditional distribution of all n random effects U_i and that of the ϕ parameter, within each Gibbs sampling iteration. This leads to an enormous demand of computing capacity and makes implementation of the algorithm extremely slow (or possibly infeasible), especially for large number of locations.

Sampling-Importance-Resampling

Gelfand et al. (1999) propose Bayesian inference for point-referenced data using non-iterative Sampling-Importance-Resampling (SIR) simulation. They replace matrix inversion with simulation by introducing a suitable importance sampling density $g(\cdot)$ and re-write the joint posterior as

$$p^{*}(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi \mid \mathbf{Y}) = \frac{p(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi \mid \mathbf{Y})}{g(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi; \mathbf{Y})} g(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi; \mathbf{Y}). \quad (6)$$

They construct the importance sampling density (ISD) by

$$g(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi; \mathbf{Y}) = g_s(\boldsymbol{\beta} \mid \mathbf{U}; \mathbf{Y}) g_s(\mathbf{U} \mid \sigma^2, \phi) g_s(\sigma^2, \phi) \quad (7)$$

which is easy to simulate from and then re-sample from $g(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi; \mathbf{Y})$ according to the importance weights

$$w(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi) = \frac{p(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi \mid \mathbf{Y})}{g(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi; \mathbf{Y})}. \quad (8)$$

The density $g_s(\sigma^2, \phi)$ of the ISD could be taken as a product of independent inverse Gamma distributions $g_s(\sigma^2) g_s(\phi)$. It is however preferable to adopt a bivariate distribution which accounts for interrelations between the two parameters and thus it approximates closer the $p(\sigma^2, \phi \mid \mathbf{Y})$. We considered a bivariate t-distribution on $\log(\sigma^2)$ and $\log(\phi)$ with low degrees of freedom and mean around the maximum likelihood estimates of $\log(\sigma^2)$ and $\log(\phi)$. The spatial random effects can be simulated from a multivariate normal distribution,

$$g_s(\mathbf{U} \mid \sigma^2, \phi) \equiv N(0, \sigma^2 \rho(\phi, \cdot)) \,.$$

This step requires matrix decomposition of $\sigma^2 \rho(\phi, \cdot)$, repeatedly at every iteration. This is an operation of order 2 and the most expensive numerical part of the simulation from the ISD. The density $g_s(\boldsymbol{\beta} \mid \mathbf{U}; \mathbf{Y})$ can be a Normal distribution, $g_s(\boldsymbol{\beta} \mid \mathbf{U}; \mathbf{Y}) \equiv N(\hat{\boldsymbol{\beta}}_{\mathbf{U}}, \hat{\Sigma}_{\boldsymbol{\beta}})$, with $\hat{\boldsymbol{\beta}}_{\mathbf{U}}$ equal to the regression coefficients estimated from an ordinary logistic regression with offset $\mathbf{U}$ and $\hat{\Sigma}_{\boldsymbol{\beta}}$ equal to the covariance matrix of $\hat{\boldsymbol{\beta}}_{\mathbf{U}}$.

When the ISD approximates well the posterior distribution, one expects that the standardized importance weights are Uniformly distributed. When this is not the case, the ISD would give rise to very few dominant weights leading to an inefficient and wrong sampler. A possible remedy would be to embed the Sampling-Importance-Resampling simulation in an iterative scheme which refines the initial

guesses of the ISD and allows after few iterations more uniform weights.

Point estimates of the parameters should preferably be calculated from the importance weights using all sampled values, rather than from the re-sampled values, what leads to smaller bias. For example the mean and variance of β_i is estimated by $\bar{\beta}_i = \sum_k w_k \beta_i^{(k)} / \sum_k w_k$ and $\sum_k w_k (\beta_i^{(k)} - \bar{\beta}_i)^2 / \sum_k w_k$ respectively, where $\beta_i^{(k)}$ is the kth sampled value of β_i from the ISD.

Spatial Prediction

Modeling point-referenced data is not only useful for identifying significant covariates but for producing smooth maps of the outcome by predicting it at unsampled locations. Spatial prediction is usually refereed as kriging.

Let $\mathbf{Y}_0$ be a vector of the binary response at new, unobserved locations s_{0i}, $i = 1, \ldots, n_0$. Following the maximum likelihood approach, the distribution of $\mathbf{Y}_0$ is given by:

$$P(\mathbf{Y}_0 \mid \hat{\boldsymbol{\beta}}, \hat{\mathbf{U}}, \hat{\sigma}^2, \hat{\phi}) = \int P(\mathbf{Y}_0 \mid \hat{\boldsymbol{\beta}}, \mathbf{U}_0) P(\mathbf{U}_0 \mid \hat{\mathbf{U}}, \hat{\sigma}^2, \hat{\phi}) d\mathbf{U}_0 \qquad (9)$$

where $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\hat{\phi}$ are the maximum likelihood estimates of the corresponding parameters. In PQL, $\hat{\mathbf{U}}$ is derived as part of the iterative estimation process (Breslow and Clayton, 1993). $P(\mathbf{Y}_0 \mid \hat{\boldsymbol{\beta}}, \mathbf{U}_0)$ is the Bernoulli-likelihood at new locations and $P(\mathbf{U}_0 \mid \hat{\mathbf{U}}, \hat{\sigma}^2, \hat{\phi})$ is the distribution of the spatial random effects $\mathbf{U}_0$ at new sites, given $\hat{\mathbf{U}}$ at observed sites and is Normal

$$P(\mathbf{U}_0 \mid \hat{\mathbf{U}}, \hat{\sigma}^2, \hat{\phi}) = N(\boldsymbol{\Sigma}_{01} \boldsymbol{\Sigma}_{11}^{-1} \hat{\mathbf{U}}, \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{10}) \qquad (10)$$

with $\boldsymbol{\Sigma}_{11} = E(\mathbf{U}\mathbf{U}^t)$, $\boldsymbol{\Sigma}_{00} = E(\mathbf{U}_0 \mathbf{U}_0^t)$ and $\boldsymbol{\Sigma}_{01} = \boldsymbol{\Sigma}_{10}^t = E(\mathbf{U}_0 \mathbf{U}^t)$. The mean of the Gaussian distribution in (10) is the classical kriging estimator (Matheron, 1963).

The Bayesian predictive distribution of $\mathbf{Y}_0$ is given by:

$$P(\mathbf{Y}_0 \mid \mathbf{Y}) = \int P(\mathbf{Y}_0 \mid \boldsymbol{\beta}, \mathbf{U}_0) P(\mathbf{U}_0 \mid \mathbf{U}, \sigma^2, \phi) \times P(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi \mid \mathbf{Y}) d\boldsymbol{\beta} d\mathbf{U}_0 d\mathbf{U} d\sigma^2 d\phi \qquad (11)$$

$P(\boldsymbol{\beta}, \mathbf{U}, \sigma^2, \phi \mid \mathbf{Y})$ is the posterior distribution of the parameters and obtained by the Gibbs sampler or the SIR approach. Simulation-based Bayesian spatial prediction is performed by consecutive drawing samples from the posterior distribution, the distribution of the spatial random effects at new locations and the Bernoulli-distributed predicted outcome. In SIR, drawing is performed from the set of all sampled parameters with weighting given in Equation (8).

The maximum-likelihood predictor (Equation 9) can be interpreted as the Bayesian predictor (Equation 11), with parameters fixed at their maximum-likelihood estimates. In contrast to Bayesian kriging, classical kriging does not account for uncertainty in estimation of $\boldsymbol{\beta}$ and the covariance parameters.

Results

A generalized linear mixed model was fitted to the infant mortality data in Mali using the three estimation approaches discussed in the methodology-section, PQL, MCMC and SIR together with an ordinary logistic regression (GLM) which did not account for spatial dependence. The purpose of the analysis was to assess the effect of maternal and socio-economic factors on infant mortality, produce a smooth map of mortality risk in Mali and compare the results obtained from the above procedures. Univariate analysis based on the ordinary logistic regression revealed that the following variables should be included in the model: child's birthday, region type, mother's degree of education, sex, birth order, preceding birth interval and mother's age at birth.

We fitted the non-spatial logistic model (GLM) in SAS (SAS Institute Inc., Cary, NC, USA) using Proc Logistic. The spatial model with the PQL estimation method was also fitted in SAS using the `%GLIMMIX`-macro. This macro is based on the approach of Wolfinger and O'Connell (1993) and does subsequent calls of Proc Mixed to iteratively estimate mixed models for non-normal data. It is supported by a collection of spatial correlation functions, such as the exponential, Gaussian, linear, power and spherical. In our application, we have chosen the exponential function. MCMC and SIR estimation were implemented in software written by the authors in FORTRAN 95 (Compaq Visual Fortran v6.6) and run on an Unix AlphaServer 8400. For small number of locations the freeware software WinBUGS (www.mrc-bsu.cam.ac.uk/bugs) can also be used to obtain MCMC simulation-based estimates. Proc Mixed for normal data supports Bayesian modeling by allowing specification of prior distributions for the parameters and MCMC simulation. However, this possibility is currently available only for variance component models and not for spatial covariances, which holds for the `%GLIMMIX` macro, too.

Table 1: Comparison of the computational costs for the Bayesian, simulation based approaches.

Model	Initial sample size	Final sample from posterior	No. of batches and size	Iterations to convergence	Thinning*	Time per 1,000 iterations
MCMC	50,000	1,720	-	7,000	25	7 hrs 14 min
SIR	400,000	1,600	800 batches with 500 values (2 batches per draw)	0	0	1 hr 23 min

*Minimum lag at which autocorrelation was not significant.

Table 2: Comparison of parameter estimates from the binary spatial model using different estimation strategies. The binary outcome is the survival of the first year of life.

Model	Estimate	σ^2	ϕ	Intercept	Birth year 1966-71	1972-77	1978-83	1984-89
GLM	MLE	-	-	1.81	-0.18	0.04	0.09	0.12
	95% CI	-	-	1.43,2.11	-0.44,0.09	-0.22,0.29	-0.16,0.34	-0.13,0.37
PQL	MLE	1.05	2.07	2.59	-0.19	0.03	0.09	0.12
	95% CI	0.72,1.81	0.54,4.63	1.43,3.74	-0.48,0.11	-0.26,0.31	-0.19,0.37	-0.17,0.40
MCMC	Mean	1.32	0.07	1.76	-0.20	0.01	0.07	0.10
	Median	0.91	0.04	1.75	-0.21	0.01	0.07	0.09
	95% CI	0.22,3.89	0.008,0.24	1.47,2.09	-0.46,0.08	-0.25,0.27	-0.19,0.33	-0.16,0.36
SIR	Mean	0.91	0.005	1.77	-0.19	0.03	0.08	0.11
	Median	0.61	0.03	1.73	-0.18	0.03	0.08	0.11
	95% CI	0.22,2.62	0.0004,0.015	0.34,3.25	-0.44,0.06	-0.21,0.27	-0.16,0.31	-0.13,0.34

Model	Estimate	Birth year 1990-96	Residency Urban	Education		Sex Male	Birth order	
				No	Primary		2nd or 3rd	4th to 6th
GLM	MLE	0.17	0.32	-0.56	-0.66	-0.14	-1.90	-1.97
	95% CI	-0.08,0.42	0.22,0.36	-0.75,-0.31	-0.83,-0.43	-0.16,-0.05	-2.40,-1.32	-2.48,-1.38
PQL	MLE	0.16	0.29	-0.54	-0.58	-0.14	-1.88	-1.95
	95% CI	-0.12,0.44	0.19,0.39	-0.75,-0.32	-0.78,-0.38	-0.20,-0.09	-2.42,-1.34	-2.50,-1.40
MCMC	Mean	0.15	0.3	-0.55	-0.6	-0.14	-1.90	-2.00
	Median	0.14	0.3	-0.54	-0.59	-0.14	-1.95	-2.02
	95% CI	-0.11,0.40	0.23,0.38	-0.74,-0.36	-0.78,-0.42	-0.16,-0.10	-2.39,-1.44	-2.48,-1.51
SIR	Mean	0.16	0.33	-0.50	-0.57	-0.14	-1.88	-1.96
	Median	0.16	0.34	-0.50	-0.57	-0.14	-1.88	-1.96
	95% CI	-0.08,0.39	0.25,0.41	-0.68,0.32	-0.75,-0.40	-0.19,-0.09	-2.34,-1.42	-2.43,-1.49

Model	Estimate	Birth order 7th or higher	Preceding birth interval		Mothers age at birth		
			2-4	> 4	20-29	30-39	39-49
GLM	MLE	-2.10	2.34	2.71	0.24	0.31	0.19
	95% CI	-2.62,-1.51	1.76,2.84	2.11,3.22	0.13,0.29	0.15,0.40	-0.02,0.42
PQL	MLE	-2.07	2.31	2.67	0.25	0.32	0.19
	95% CI	-2.63,-1.52	1.77,2.85	2.12,3.22	0.17,0.33	0.19,0.44	-0.07,0.44
MCMC	Mean	-2.10	2.37	2.73	0.26	0.33	0.20
	Median	-2.15	2.38	2.74	0.26	0.33	0.20
	95% CI	-2.16,-1.63	1.87,2.82	2.20,3.22	0.19,0.32	0.23,0.43	-0.009,0.43
SIR	Mean	-2.09	2.31	2.65	0.25	0.32	0.21
	Median	-2.09	2.30	2.65	0.25	0.32	0.21
	95% CI	-2.56,-1.62	1.82,2.77	2.16,3.13	0.18,0.31	0.22,0.43	0.01,0.42

Convergence of the PQL approach to the global mode of the likelihood was highly dependent on the starting values. We suggest to compare the results by running the procedure with several starting values. Computationally, the PQL is fast in comparison to the simulation-based procedures, MCMC and SIR, but it runs quickly out of workspace for larger dataset. A comparison of the computational time required for the MCMC and SIR algorithms is given in table 1. MCMC estimation was applied using a single chain. Convergence was assessed using Geweke's (1992) criterion. The algorithm converged after 7,000 iterations. A final sample from the posterior distribution of size 1,720 was obtained by sampling every 25th iterations after convergence was reached. The SIR algorithm required extensive fine tuning in order to derive good estimates. We ran the sampler several times and adjusted the degrees of freedom and mean parameter in the bivariate t-distribution $g_s(\sigma^2, \phi)$, according to those values leading to large weights. Instead of resampling from the whole sequence of parameters according to their weights, we obtained better results by dividing the generated parameters into batches and drawing an equal number of samples with replacement from every batch. The implementation of the SIR algorithm was found to be difficult. Despite the effort applied to improve the SIR estimator, the derived weights show a highly skewed distribution, with a few dominating values (Figure 1).

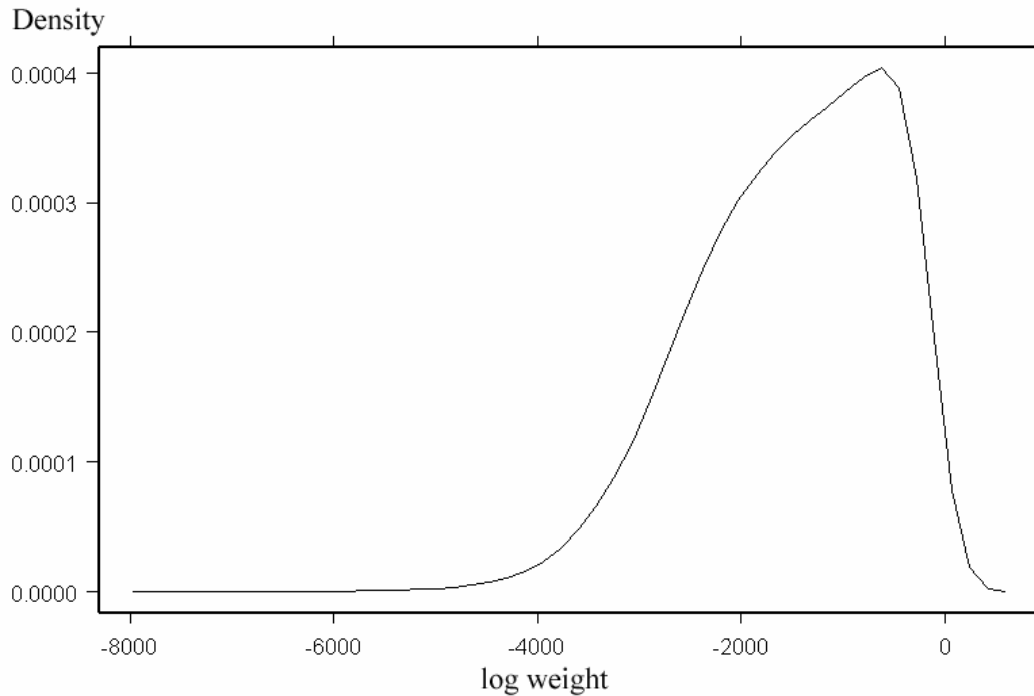Figure 1: Distribution of the weights in the Sampling-Importance-Resampling (SIR) procedure.

Table 2 gives the parameter estimates obtained by the four approaches. The fixed effect coefficients **β** show no fundamental difference in their point estimates between the competitive models, with the exception of the intercept coefficient. The PQL estimate of the intercept is higher than from the other estimators. The standard error of **β** estimated from GLM is narrower than in the spatial models, as we were expecting. Discrepancies between the fitting approaches are observed in the estimates of the covariance parameters σ^2 and ϕ. The posterior density of σ^2 obtained from MCMC simulation was found to be highly skewed to the left. PQL overestimates ϕ suggesting a lower spatial variation than the Bayesian approaches. This confirms known results about bias in the PQL estimates especially for the covariance parameters σ^2 and ϕ due to the bad quality of the first-order approximation of the integrand. The SIR estimates are similar to those obtained from MCMC.

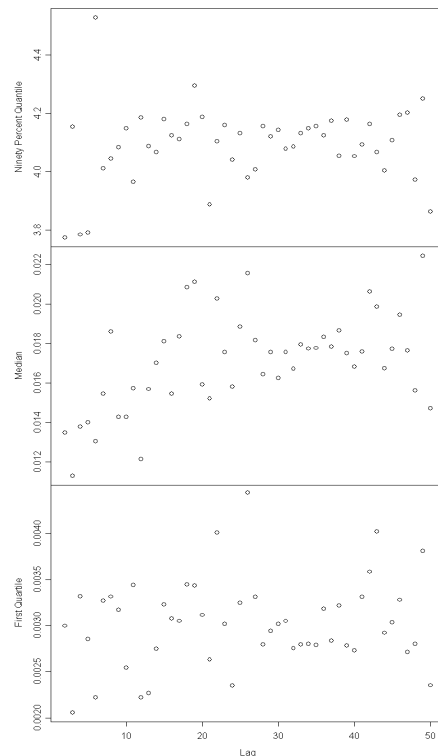Figure 2: Variogram cloud of the residuals in a non-spatial model.

Figure 2 shows three plots of the semivariogram cloud based on the Anscombe residuals obtained after fitting the GLM model. The semivariogram cloud is a plot of half the squared difference of the residuals versus the distance between their sample locations. The mean of the squared differences at each lag gives an estimator of the semivariogram. The three plots correspond to the 5%, 50% and 95% quartile of the squared difference of the residuals. The semivariogram cloud shows high variability and an increasing trend from the origin indicating lag-dependent variation. For a stationary spatial process, the semivariogram relates to the covariance of the random effects. Therefore we expect high variability in the covariance parameters.

Figure 3 depicts different semivariogram estimators. The classical estimator by Matheron (1963) was calculated by

$$\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2 \,,$$

where $Z(\mathbf{s}_i)$ is the Anscombe residual at location $\mathbf{s}_i$,

$$N(h) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| = h \pm \varepsilon\}$$

and $|N(h)|$ is its cardinality. This estimator is sensitive to outliers and a robust version was proposed by Cressie and Hawkins (1980), which is displayed in Figure 3, too. The MCMC, SIR and PQL based estimators were calculated by replacing the estimates of σ^2 and ϕ obtained from the three approaches in

$$\gamma(h) = \sigma^2(1 - \exp(-\phi \cdot h)) \,.$$

The MCMC and SIR estimators appear to be between the two other empirical semivariogram estimators. Because we have omitted the nugget term, they pass through the origin. Nevertheless, their values fit nicely into the graph. The PQL estimate does not capture the correlation present at large lags. It represents the classical semivariogram estimator well, but it is far off the robust version.

Regarding our application, Figure 4 displays the locations of the DHS surveys and the observed infant mortality risk in Mali. The risk factors which were found to be statistically significant related to infant mortality (table 2) confirm findings made by other authors. The negative association between maternal education and mortality has been described by Farah and Preston (1982) and Cleland and Ginneken (1989). Higher education may result in higher health awareness, better utilization of health facilities (Jain, 1988), higher income and ability to purchase goods and services which improves infants health (Schultz, 1979).

Figure 3: Semivariogram estimators: Classical semivariogram estimator by Matheron (circles), Robust version by Cressie and Hawkins (triangles), MCMC (long dashed line), SIR (short dashed line) and PQL (line) fit.
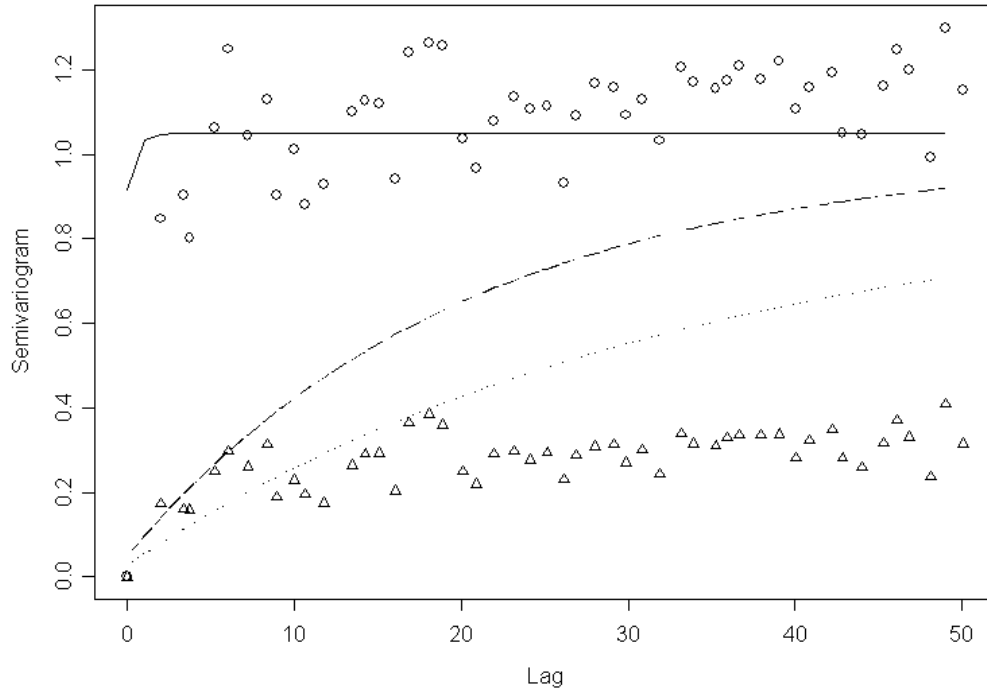
The observed time trend, with higher infant survival for more recent years, was found not statistically significant. Longer birth intervals and low birth order reduce the risk of infant death. Mortality was related to the residency and sex of the infant with girls and urbanites being at lower risk of dying during the first year of life. The impact mothers age has on infant mortality shows the typical J-shape (Kalipeni, 1993) with lowest risk for age around thirty. The higher risk in young women may be explained by not fully developed maternal resources and that in older women by the effect of ageing. The MCMC-based estimate of the ϕ parameter revealed strong spatial correlation which reduces to less than 5% for distances longer than 75km.

Figure 4: Observed mortality in 36,906 infants from the DHS surveys conducted in the years 1995 and 1996 at 181 distinct locations in Mali.
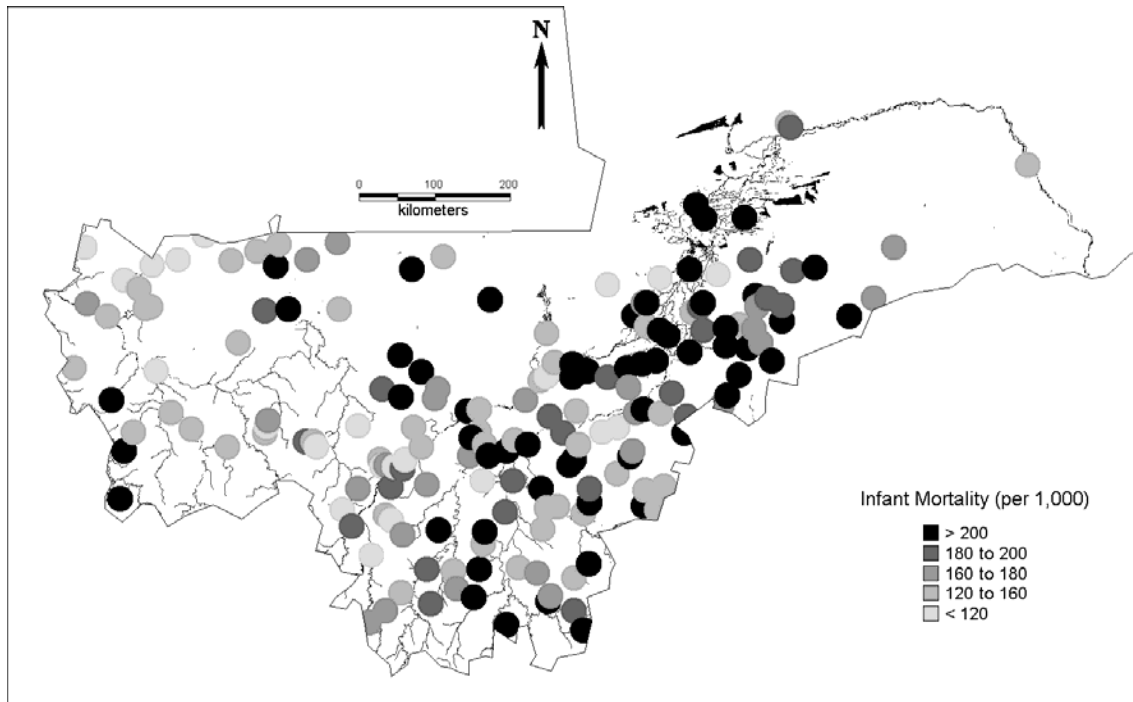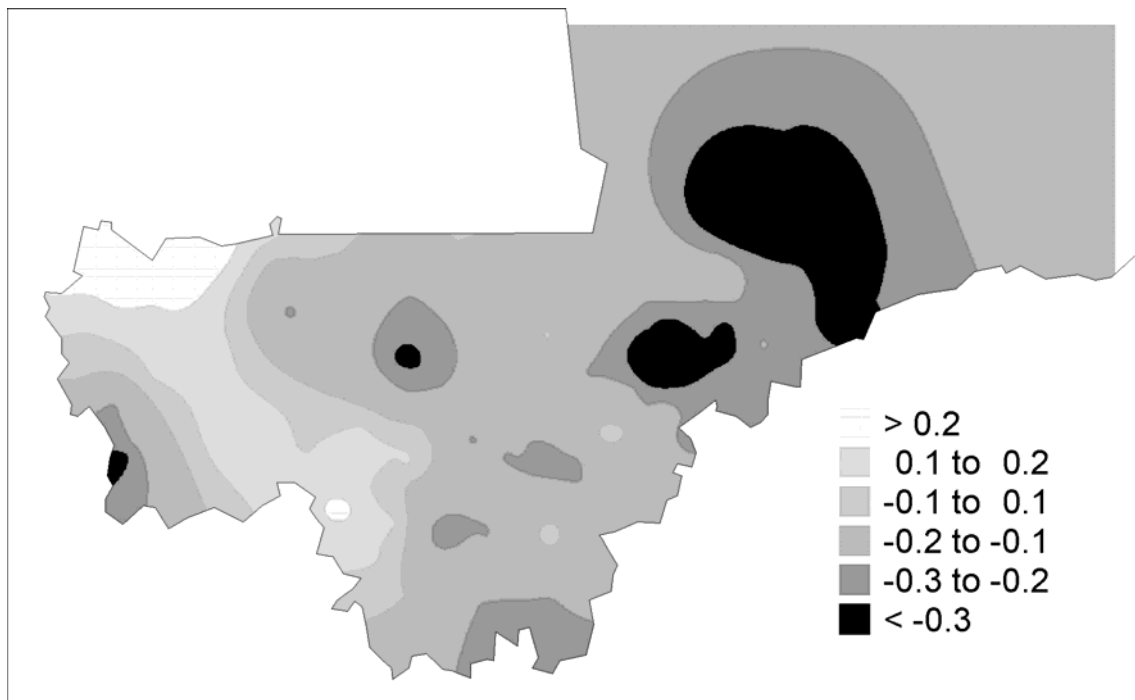

Figure 5: Predicted spatial random effects from the infant mortality model using MCMC. The darker the shading, the lower the survival.

Predictions of the child mortality risk using the MCMC approach were made at 600,000 new locations on a regular grid, covering the whole area of Mali south of 18 degrees latitude north. Because the covariates are infant-specific and can not be extrapolated for the new locations, we predict the random effects only. The map with prediction is displayed in Figure 5. The map indicates a higher infant mortality risk mainly in the Northern part of the Niger delta. This region has low population density and water availability is seasonal. The many lakes in this region are preferred breeding site for the malaria mosquito. Low mortality is predicted in North-Western Mali at the border to Mauritania and Senegal. In this region, the population is more active in migrating to other countries for business purposes, bringing money to the region. Health facility coverage is also reflected in the predictive map, where the coverage is low in the Northern Niger delta and high in the North-East.

Conclusion

Generalized linear mixed models for large point-referenced spatial data are highly parameterized and their estimation is hampered by computational problems. Reliable estimation methods that can be applied in standard software or algorithms that can accurately estimate the model parameters within practical time constraints do not exist. In this paper we compared a few recent developments using a real dataset on infant mortality in Mali.

The advantage of the PQL method is that it can be applied in standard statistical software package. However estimates are biased especially those for the covariance parameters. The algorithm depends highly on the starting values and can easily converge to a local mode. For medium to large number of locations implementations of this algorithm is impeded by computer memory problems.

Bayesian methods can provide flexible ways of modeling point-referenced data, give unbiased estimates of the parameters and their standard error and have computational advantages for problems larger than the ones the maximum likelihood methods can handle. However, for very large number of locations, an implementation may be infeasible due to long computing time. The SIR runs considerably faster than MCMC, but it requires tedious tuning. Finding an ISD which approximates well the posterior distribution is difficult to develop and application-specific. Rigorous methods for evaluating the suitability of the ISD do not exist. This increases the possibility of drawing misleading inference.

MCMC is the most practical and, when it comes to prediction, accurate approach to date for fitting geostatistical problems. However, it is computationally intensive, especially for dataset with large number of locations. More research is required in ways of improving the convergence of the algorithm and the inversion of large matrices. Gilks and Roberts (1996), Mira and Sargent (2000) and Haran et al. (2001) have proposed general MCMC algorithms for improving convergence. Rue (2000) and Pace and Barry (1997) have applied innovative numerical methods using sparse matrix solvers for fitting areal data. In future, similar approaches need to be adapted and assessed for modeling point-referenced spatial data.

References

Anderson, D. A., & Hinde, J. P. (1988). Random effects in generalized linear models and the EM algorithm. *Communication in Statistics: Theory and Methods*, *17*, 3847–3856.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36, 192–236.

Booth, J. G., & Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, *61*, 265–285.

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

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

Browne, W. J., & Draper, D. (2000). A comparison of Bayesian and likelihood methods for fitting multilevel models. Submitted.

Cleland, J., & van Ginneken, J. K. (1989). Maternal education and child survival in developing countries: The search for pathways of influence. *Social Science and Medicine*, *27*, 1357–1368.

Cressie, N., & Hawkins, D. M. (1980). Robust estimation of the variogram. *Mathematical Geology*, *12*, 115–125.

Dempster, A. P., Laird, N. M, & Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Diggle, P. J.,Tawn, J. A., & Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society, Series C*, *47*, 299-350.

Ecker, M., & Gelfand, A. E. (1997). Bayesian variogram modeling for an isotropic spatial process. *Journal of Agricultural, Biological and Environmental Statistics*, *2,* 347–369.

Farah, A. A., & Preston, S. H. (1982). Child mortality differentials in Sudan. *Population and Development Review*, *8*, 365–383.

GDE Systems Inc. (1995). *Geoname Digital Gazetteer*, *Version I*, CD-ROM.

Gelfand. A. E.,Ravishanker. N., & Ecker, M. (1999). Modeling and inference for point-referenced binary spatial data. In D. Dey,S. Ghosh & B. Mallick (Eds.), *Generalized linear models: a bayesian perspective*. 373–386. Marcel Dekker Inc.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger,A.P. Dawid & A. F. M. Smith (Eds.), *Bayesian statistics, 4*, 169–193. Oxford University Press.

Gilks, W. R., & Roberts, G. O. (1996). Strategies for improving MCMC. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice*. 89–114. London: Chapman and Hall.

Haran, M., Hodges, J. S., & Carlin, B. P. (2001). Accelerating computation in Markov random field models for spatial data via structured MCMC. *Journal of Computaional and Graphical Statistics*, *12*, 249-264.

Heagerty, P. J., & Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, *93*, 1099–1111.

Jain, A. (1988). Determinants of regional variation in infant mortality in rural India. In A. Jain, & L. Visaria (Eds.), *Infant mortality in India: Differentials and determinants*, 127–167. Sage Publications.

Kalipeni, E. (1993). Determinants of infant mortality in Malawi: a spatial perspective. *Social Science and Medicine*, *37*, 183–198.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: An example. *Journal of the Royal Statistical Society, Series C*, *50*, 325–335.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13–22.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*, 1246–1266.

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

Mira, A., & Sargent, D. J. (2000). Strategies for speeding Markov chain Monte Carlo algorithms. *Technical Report*, University of Insubria, Varese.

Neuhaus, J. N., & Segal, M. R. (1997). An assessment of approximate maximum likelihood estimators in generalized linear mixed models. In T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren, & R. D. Wolfinger (Eds.), *Modeling longitudinal and spatially correlated data*: *Lecture notes in statistics*, *122,* 11–22.

Pace, K. R., & Barry, R. (1997). Quick computation of the regressions with spatially autoregressive dependent variable. *Geographical Analysis*, *29*, 232–247.

Preisler. H. K. (1988). Maximum likelihood estimates for binary data with random effects. *Biometrical Journal*, *3*, 339–350.

Rubin, D. B. (1987). Comment on: The calculation of posterior distributions by data augmentation. by M. A. Tanner and W. H. Wong. *Journal of the American Statistical Association*, *82*, 543–546.

Rue, H. (2000). Fast sampling of gaussian markov random fields. *Journal of the Royal Statistical Society, Series B*, *48*, 233–243.

Schultz, T. P. (1979, June 19-25). Interpretation of relations among mortality, economics of the household and the health environment. *Proceedings of the meeting on socio-economic determinants and consequences of mortality*, Mexico City. Geneva: World Health Organization.

Tubilla, A. (1975). Error convergence rates for estimates of multidimensional integrals of random functions. *Technical Report* No. 72, Department of Statistics, Stanford University, Stanford, CA.

Wolfinger, R., & O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, *48*, 233–243.

World Resources Institute (1995) *Africa Data Sampler*. Edition I, CD-ROM.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrika*, *42*, 121–130.

Zimmerman, D. L., & Zimmerman, M. B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, *33*, 77–91.