Abstract. Linear mixed models are able to handle an extraordinary range of complications in regression-type analyses. Their most common use is to account for within-subject correlation in longitudinal data analysis (e.g. Laird and Ware, 1982). They are also the standard vehicle for smoothing spatial count data (e.g. Wakefield, Best and Waller, 2000). However, when treated in full generality, mixed models can also handle spline-type smoothing and closely approximate kriging (e.g. Robinson, 1991; Speed, 1991). This allows for nonparametric regression models (e.g. additive models, varying coefficient models) to be handled within the mixed model framework. The key is to allow the random effects design matrix to have general structure; hence our label general design. For continuous response data, particularly when Gaussianity of the response is reasonably assumed, computation is now quite mature and supported by the SAS, S-PLUS and R packages. Such is not the case for binary and count responses where generalized linear mixed models (GLMMs) are required, but are hindered by the presence of intractable multivariate integrals. Software known to to us supports special cases of the GLMM (e.g. PROC NLMIXED in SAS or glmML in R) or relies on the sometimes crude Laplace-type approximation of integrals (e.g. the SAS macro glimmix or glmmPQL in R). This paper describes the fitting of general design generalized linear mixed models. A Bayesian approach is taken and Markov Chain Monte Carlo (MCMC) is used for estimation and inference. In this generalized setting MCMC requires sampling from non-standard distributions. In this article, we demonstrate that the MCMC package WinBUGS facilitates sound fitting of general design Bayesian generalized linear mixed models in practice.

Key words and phrases: Generalized additive models; Hierarchical centering; Kriging; Markov chain Monte Carlo; Nonparametric regression; Penalized splines; Spatial count data; WinBUGS.
1 INTRODUCTION

The generalized linear mixed model (GLMM) is one of the most useful structures in modern Statistics, allowing many complications to be handled within the familiar linear model framework. The fitting of such models has been the subject of a great deal of research over the past decade. Early contributions to fitting various forms of the GLMM include Stiratelli, Laird and Ware (1984), Anderson and Aitkin (1985), Gilmour, Anderson and Rae (1985), Schall (1991), Breslow and Clayton (1993) and Wolfinger and O’Connell (1993). A summary is provided in McCulloch and Searle (2000, Chapter 10).

Most of the literature on fitting GLMMs is geared towards grouped data. Examples include repeated binary responses on a set of subjects and standardized mortality ratios in geographical sub-regions. However, GLMMs are much richer than the sub-class needed for these situations. The key to full generality is the use of general design matrices, for both the fixed and random components. Once again, we refer to McCulloch and Searle (2000, Chapter 8) for an overview of general design GLMMs. An excellent synopsis of general design linear mixed models is provided by Robinson (1991) and the ensuing discussion. One of the biggest payoffs from the general design framework is the incorporation of nonparametric regression, or smoothing, through penalized regression splines (e.g. Wahba, 1990; Speed, 1991; Verbyla, 1994; Brumback, Ruppert and Wand, 1999). Higher-dimensional extensions essentially correspond to generalized kriging (Diggle, Tawn and Moyeed, 1998). This allows for smoothing-type models such as generalized additive models to be fit as a GLMM, and combined with the more traditional grouped data uses. This is the main thrust of the recent book by Ruppert, Wand and Carroll (2003) and a summary is provided by Wand (2003). General designs also permit the handling of crossed random effects (e.g. Shun, 1997) and multilevel models (e.g. Goldstein, 1995; Kreft and de Leeuw, 1998).

The simplest method for fitting general design GLMMs involves Laplace approximation of integrals (Breslow and Clayton, 1993; Wolfinger and O’Connell, 1993) and is commonly referred to as penalized quasi-likelihood (PQL). However, the approximation can be quite inaccurate in certain circumstances. Breslow and Lin (1995) and Lin and Breslow (1996) show that PQL leads to estimators that are asymptotically biased. For situations such as paired binary data the PQL approximation is particularly poor. In their summary of PQL McCulloch and Searle (2000, Chapter 10, pp. 283-284) conclude by stating that they “cannot recommend the use of simple PQL methods in practice”. In this article we take a Bayesian approach and explore the Markov Chain Monte Carlo (MCMC) fitting of general design GLMMs. One advantage of a Bayesian approach over its frequentist counterpart include the fact that uncertainty in variance components is more easily taken into account (e.g. Handcock and Stein, 1993; Diggle et al., 1998). As summarized in Section 9.6 of McCulloch and Searle (2000) the fre-
quentist approach to this problem is thwarted by largely intractable distribution theory. Under a Bayesian approach, posterior distributions of parameters of interest take this variability into account. The hierarchical structure of the Bayesian GLMMs lends itself to Gibbs sampling schemes, albeit with some non-conjugate full conditionals, to sample from these posteriors. In addition, it is computationally simpler to obtain variance estimates of the predictions of the random effects. Booth and Hobert (1998) showed that, in a frequentist framework, second-order estimation of the conditional standard error of prediction for the random effects requires bootstrapping the maximum likelihood estimates of the fixed effects and variance components. For complicated random effects structures, computation of a single maximum likelihood fit can be expensive, making the bootstrap computationally prohibitive. However, in the Bayesian framework, interest focuses on the posterior variance of the random effects given the data, which is a by-product of the MCMC output.

There have been a few other contributions to Bayesian formulations of GLMMs in the literature. Those known to us are Zeger and Karim (1993), Clayton (1996), Diggle, Tawn and Moyeed (1998), and Fahrmeir and Lang (2001). However, each of these articles are geared towards special cases of GLMMs. The GLMMs described in this article are much more general and allow for random effects models for longitudinal data, crossed random effects, smoothing of spatial count data, generalized additive models, generalized geostatistical models, additive models with interactions, varying coefficient models and various combinations of these (Wand, 2003).

Section 2 lays out notation for general design GLMMs and gives several important examples. MCMC implementation is described in Section 3, with a focus on the WinBUGS package. Section 4 provides three illustratory data analyses. We close with some discussion in Section 5.

## 2 MODEL FORMULATION

GLMMs for canonical one-parameter exponential families (e.g. Poisson, logistic) and Gaussian random effects take the general form

\[ [y | \beta, u] = \exp\{ y^T (X \beta + Zu) - 1^T b(X \beta + Zu) + 1^T c(y) \}, \]  

\[ [u | G] \sim N(0, G) \]  

where here, and throughout, the distribution of a random vector \( x \) is denoted by \( [x] \) and the conditional distribution of \( y \) given \( x \) is denoted by \( [y|x] \).

In the Poisson case \( b(x) = e^x \), while in the logistic case \( b(x) = \log(1 + e^x) \).

A few other models (e.g. gamma, inverse Gaussian) also fit into this structure (McCullagh and Nelder, 1989). A number of extensions and modifications are
possible. One is to allow for overdispersion, especially in the Poisson case. In this paper we will restrict attention to the canonical one-parameter exponential family structure.

In most situations, the main parameters of interest are contained in $\beta$ and $G$. Throughout we will take the prior distribution of $\beta$ to be of the form $[\beta] \sim N(0, F)$ for some covariance matrix $F$. In practice it is common to take $F$ to be diagonal with very large entries, corresponding to non-informative priors on the entries of $\beta$. The components of $G$ are discussed in Section 2.1.

It is important to separate out random effects structure for handling grouping. One reason is that it allows for the possibility of hierarchical centering in the MCMC implementations (Section 2.2). It also recognizes the different covariance structures used in longitudinal data modeling, smoothing and spatial statistics. Such considerations suggest the breakdown

$$X \beta + Zu = X^R \beta^R + Z^Ru^R + X^G \beta^G + \sum_{\ell=1}^{L} Z_{\ell}^G u^G_\ell + Z^C u^C,$$

where

$$X^R \equiv \begin{bmatrix} X^R_1 \\ \vdots \\ X^R_m \end{bmatrix}, \quad Z^R \equiv \text{blockdiag}(X^R_i)_{1 \leq i \leq m}$$

and

$$\text{Cov}(u^R) \equiv \text{blockdiag}(\Sigma^R) = I_m \otimes \Sigma^R$$

correspond to random intercepts and slopes, as typically used for repeated measures data on $m$ groups with sample sizes $n_1, \ldots, n_m$. Here $X^R_i$ is a $n_i \times q^R$ matrix for the random design corresponding to the $i$th group, $\Sigma^R$ is an unstructured $q^R \times q^R$ covariance matrix and $\otimes$ denotes Kronecker product.

Next, $Z^G$ is a general design matrix, usually of different form than that arising in random effects models. In many of our examples, $Z^G$ contains spline basis functions and may be further decomposed as $Z^G = [Z^G_1 \ldots Z^G_L]$, with each $Z^G_\ell$ corresponding to a smooth term in an additive model. Also, in keeping with spline penalization, we only consider

$$\text{Cov}(u^G) \equiv \text{blockdiag}(\sigma^2_{u\ell} I)_{1 \leq \ell \leq L}$$

for some integer $L$.

We note that the decomposition of the random effect vector $u$ into $u^R$ and $u^G$ may not be unique. For instance, in the crossed random effects model given in the following Example 3, we present two ways of decomposition.

4
The $Z^C u^C$ component represents random effects with spatial correlation structure. This can be done in a number of ways (e.g. Wakefield, Best and Waller, 2000); we will just describe one of the more common approaches here. Suppose disease incidence data are available over $N$ contiguous regions. The random effect $u^C$ vector is of dimension $N$ and is assumed to follow a Gaussian intrinsic autoregression prior distribution. We can specify the inverse covariance matrix of the Gaussian intrinsic autoregression distribution according to Besag, York and Mollié (1991) and give $u^C$ an improper density proportional to

$$(\sigma_c^2)^{-N/2} \exp \left\{ -\frac{1}{2} \sigma_c^{-2} (U^c_i - U^c_j)^2 \right\},$$

where $i \sim j$ denotes spatially adjacent groups. The conditional distributions of $U^c_i$ given $U^c_j$, $j \neq i$, is a univariate normal distribution with mean equal to the average $U^c_j$ values of neighboring regions, and variance equal to $\sigma_c^2$ divided by the number of spatially contiguous groups of $U^c_i$.

The versatility of (3) can be appreciated by considering the following set of examples. Note that we use truncated linear basis functions for smoothing components to keep the formulations simple (e.g. Brumback, Ruppert and Wand, 1999). In practice these may be replaced by B-splines (Durban and Currie, 2003) or radial basis functions (French, Kammann and Wand, 2001). Knots are denoted by $\kappa_k$ with possible superscripting. Ruppert (2002) discusses choice of knots of univariate smoothings, whereas Nychka and Saltzmann (1998) describe the choice of knots for multivariate smoothing and kriging. In the examples we use $1_d$ to denote a $d \times 1$ vector of ones.

**Example 1: Random intercept**

$$(X \beta + Z u)_{ij} = \beta_0 + U_i + \beta_1 x_{ij}, \quad 1 \leq j \leq n_i, \ 1 \leq i \leq m,$$

$X^R_i = 1_{n_i}, \quad X^G = [x_{ij}], \quad Z^C = Z^C = \emptyset, \quad \Sigma^R = \sigma_u^2.$

**Example 2: Random intercept and slope**

$$(X \beta + Z u)_{ij} = \beta_0 + U_i + (\beta_1 + V_i)x_{ij}, \quad 1 \leq j \leq n_i, \ 1 \leq i \leq m,$$

$$X^R_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} x_{i1}, \quad X^G = Z^G = Z^C = \emptyset,$$

$$\Sigma^R = \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}.$$
Example 3: Crossed random effects model

\((X \beta + Zu)_{i'i'} = \beta_0 + U_i + U'_{i'}, \quad 1 \leq i \leq n, \ 1 \leq i' \leq n',\)

\[X^G = 1_{nn'}, \quad Z^G = [I_n \otimes 1_{n'} \mid 1_n \otimes I_n'], \quad X^R = Z^R = Z^C = \emptyset,\]

\[u^G = [U_1, \ldots, U_n, U'_{1'}, \ldots, U'_{n'}]^T, \quad \text{Cov}(u^G) = \text{blockdiag}(\sigma_u^2 I_n, \sigma_u^2 I_{n'}).\]

An alternative representation of this model is

\[X_R^R = 1_{n' \times 1}, \quad Z^G = [I_n \otimes 1_{n'}], \quad X^G = Z^C = \emptyset,\]

\[\Sigma^R = \sigma_u^2, \quad \text{Cov}(u^G) = \sigma_u^2 I_{n'}.\]

This allows for implementation of hierarchical centering as described in Section 2.2.

Example 4: Nested random effects model

\((X \beta + Zu)_{ijk} = \beta_0 + U_i + V_{j(i)} + \beta_1 x_{ijk}, \quad 1 \leq i \leq m, \ 1 \leq j \leq n, \ 1 \leq k \leq p,\)

\[X^G = [1 x_{ijk}]_{1 \leq i \leq m, \ 1 \leq j \leq n, \ 1 \leq k \leq p}, \quad Z^G = [I_m \otimes 1_{np} \mid I_m \otimes (I_n \otimes 1_p)], \quad X^R = Z^R = Z^C = \emptyset,\]

\[u^G = [U_1, \ldots, U_m, V_{1(1)}, \ldots, V_{n(1)}, \ldots, V_{1(m)}, \ldots, V_{n(m)}]^T, \quad \text{Cov}(u^G) = \text{blockdiag}(\sigma_u^2 I_m, \sigma_v^2 I_{np}).\]

Example 5: Generalized scatterplot smoothing

\((X \beta + Zu)_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^K u_k (x_i - \kappa_k)_+,\)

\[X^G = [1 x_i]_{1 \leq i \leq n}, \quad Z^G = [(x_i - \kappa_k)_+]_{1 \leq i \leq n, \ 1 \leq k \leq K}, \quad X^R = Z^R = Z^C = \emptyset,\]

\[\text{Cov}(u^G) = \sigma_u^2 I_K.\]
Example 6: Generalized additive model

\[
(X \beta + Zu)_i = \beta_0 + \beta_s s_i + \sum_{k=1}^{K^s} u_k^s (s_i - \kappa_k^s)_+ + \beta_t t_i + \sum_{k=1}^{K^t} u_k^t (t_i - \kappa_k^t)_+, \\
X^G = [1 \ s_i \ t_i]_{1 \leq i \leq n}, \quad Z^G = [(s_i - \kappa_k^s)_+ (t_i - \kappa_k^t)_+]_{1 \leq i \leq n}, \quad X^R = Z^R = Z^C = \emptyset, \\
\text{Cov}(u^G) = \text{blockdiag}(\sigma^2_{us} I_{K^s}, \sigma^2_{ut} I_{K^t}).
\]

Example 7: Generalized additive semiparametric mixed model

\[
(X \beta + Zu)_{ij} = \beta_0 + U_i + (\beta_q + V_i) q_{ij} + (\beta_r + W_i) r_{ij} + \beta_1 x_{ij} + \\
+ \beta_s s_{ij} + \sum_{k=1}^{K^s} u_k^s (s_{ij} - \kappa_k^s)_+ + \beta_t t_{ij} + \sum_{k=1}^{K^t} u_k^t (t_{ij} - \kappa_k^t)_+,
\]

\[
X^R = \begin{bmatrix} 1 & q_{i1} & r_{i1} \\ \vdots & \vdots & \vdots \\ 1 & q_{in} & r_{in} \end{bmatrix}, \quad X^C = [s_{ij} \ t_{ij} \ x_{ij}]_{1 \leq j \leq n, 1 \leq i \leq m}, \\
Z^G = [(s_{ij} - \kappa_k^s)_+ (t_{ij} - \kappa_k^t)_+]_{1 \leq j \leq n, 1 \leq k \leq K^s}, \quad Z^C = \emptyset, \\
\Sigma^R = \text{unstructured } 3 \times 3 \text{ covariance matrix}, \quad \text{Cov}(u^G) = \text{blockdiag}(\sigma^2_{us} I_{K^s}, \sigma^2_{ut} I_{K^t}).
\]

Example 8: Generalized bivariate smoothing/low-rank kriging

\[
(X \beta + Zu)_i = \beta_0 + \beta_1^T x_i + \sum_{k=1}^{K} u_k C(\|x_i - \kappa_k\|), \\
X^G = [1 \ x_i^T]_{1 \leq i \leq n}, \quad Z^G = [C(\|x_i - \kappa_k\|)]_{1 \leq i \leq n}, \quad X^R = Z^R = Z^C = \emptyset, \\
\text{Cov}(u^G) = \sigma^2_u I.
\]

Here \(\|v\| \equiv \sqrt{v^T v}\), and \(C(r) = r^2 \log |r|\) corresponding to low-rank thin plate splines with smoothness parameter set to 2 (as defined in Wahba, 1990); \(C(r) = \exp(-|r/\rho|)(1 + |r/\rho|)\) corresponding to Matérn low-rank kriging with range of \(\rho > 0\) and smoothness parameter set to 3/2 (as defined in Stein, 1999; Kammann and Wand, 2003). Several more examples could be added, including some where \(Z^C \neq \emptyset\).
2.1 Covariance Matrix Priors

Over the last decade and a half, prior elicitation for the variance components in Bayesian GLMMs has been an active area of statistical research. Several authors have demonstrated that the use of improper priors for these parameters can lead to improper posteriors, with Gibbs samplers unable to detect such ill-conditioning (Hobert and Casella 1996). As a result, a popular choice is the use of proper but “diffuse” conditionally conjugate priors. In the GLMM setting with normal random effects, this corresponds to an inverse Gamma (IG) distribution for a single variance component, and an inverse Wishart distribution for a variance-covariance matrix. For hierarchical versions of GLMMs, however, recent research has shown that these priors can actually be quite informative, leading to inferences that are sensitive to choice of the hyperparameters for these distributions (Natarajan and McCulloch 1998; Natarajan and Kass 2000; Gelman 2004). Natarajan and Kass (1998) and Gelman (2004) have proposed alternative prior elicitation strategies that improve upon the conditionally conjugate priors. In Section 4, we outline a sensitivity analysis approach that takes these latest proposals into account.

2.2 Hierarchical Centering

Hierarchical centering of parameters has been shown to improve convergence of Markov Chain Monte Carlo schemes (Section 3) for fitting Bayesian mixed models (e.g. Gelfand, Sahu and Carlin, 1995). In the context of this section, hierarchical centering involves reparametrization of \((\beta^R, u^R)\) to \((\beta^R, \gamma)\) where

\[
\gamma \equiv ((Z^R)^T Z^R)^{-1} (Z^R)^T X^R \beta^R + u^R.
\]

The new vector of parameters \(\gamma\) can be further divided into \(m\) sub-vectors \(\gamma_i\) with \(\gamma_i = \beta^R + u_i^R\); so that

\[
\gamma = \begin{bmatrix}
\gamma_1 \\
\vdots \\
\gamma_m
\end{bmatrix}.
\]

Then the general design generalized linear mixed model becomes

\[
X \beta + Zu = Z^R \gamma + X^C \beta^C + \sum_{\ell=1}^{L} Z_{G\ell} u_{G\ell}^C + Z_C u^C.
\]

Note that hierarchical centering is not a well-defined concept for general design or spatial structures since \(u_G^C\) and \(u_C^C\) cannot be centered in a hierarchical way similar to that for \(u^R\). As a result, the general design and spatial structures do not contribute to the model for the mean in a conditionally hierarchical manner.
2.3 Applications

This section describes three public health applications that benefit from general design Bayesian GLMM analysis. The analyses are postponed to Section 4.

2.3.1 Respiratory infection in Indonesian children

Our first example involves longitudinal measurements on 275 Indonesian children. Analyses of these data have appeared previously in the literature (e.g. Diggle, Liang and Zeger, 1995; Lin and Carroll, 2001) so our description of them will be brief. The response variable is binary: the indicator of respiratory infection. The covariate of most interest is the indicator of Vitamin A deficiency. However, the age of the child has been seen to have a non-linear effect in previous analyses.

A plausible model for these data is the Bayesian logistic additive mixed model

\[
\text{logit}\{P(\text{respiratory infection}_{ij} = 1)\} = \beta_0 + U_i + \beta^T x_{ij} + f(\text{age}_{ij}) \tag{5}
\]

where \(1 \leq i \leq 275\) indexes child and \(1 \leq j \leq n_i\) indexes the repeated measures within child. Here \(U_i \overset{\text{ind}}{\sim} N(0, \sigma^2_U)\) is a random child effect, \(x_{ij}\) denote measurements on a vector of 9 covariates: height, indicators for vitamin A deficiency, sex and stunting and visit number, and \(f\) is modelled using penalized splines with spline basis coefficients \(u_k \overset{\text{ind}}{\sim} N(0, \sigma^2_u)\).

2.3.2 Caregiver stress and respiratory health

The Home Allergen and Asthma study is an ongoing longitudinal study that is investigating risk factors for incidence of childhood respiratory problems including asthma, allergy and wheeze (Gold, et al., 1999). The portion of the study data that we will consider consists of 483 families who were followed for two and a half years after the birth of a child. At the start of the study, a number of demographic variables were measured on each family including race, categorized household income, categorized caregiver educational level, and child’s gender. Additionally, one of the hypothesized risk factors for childhood respiratory problems is exposure to a stressful environment (Wright et al., 2004). Each child’s environmental stress level was measured approximately bimonthly by a telephone interview and assessed on a discrete ordinal scale from 0 (no stress) to 16 (very high stress). This assessment was based on the 4-item Perceived Stress Scale (PSS-4) (Cohen, 1988).

Let \(1 \leq i \leq 483\) index family, and \(1 \leq j \leq n_i\) index the repeated measurements within each family. We arrived at the following Bayesian Poisson additive mixed model for stress experience by caregiver \(i\) when the child was \(\text{age}_{ij}\):

\[
\text{stress}_{ij} \sim \text{Poisson} \left[ \exp \left\{ \beta_0 + U_i + \beta^T x_{ij} + f(\text{age}_{ij}) \right\} \right]. \tag{6}
\]
The random intercept, \( U_i \) is a random family effect, and \( x_i \) includes indicators of annual family income and race (see Figure 4 for details). The term \( f(\text{age}_{ij}) \) is a nonparametric term that we model using penalized splines with spline coefficients \( u_k \sim N(0, \sigma_u^2) \). We include the nonparametric term in the model for the effect of stress as a function of child’s age since, outside of anecdotal evidence, we do not know of a biologically motivated parametric model for stress as a function of child’s age. We arrived at the other terms in the model (and removed other demographic terms and interactions from the model) based on discussions with the investigators in the study and exploratory data analyses that we fit via maximum PQL.

### 2.3.3 Standardized cancer incidence and proximity to a pollution source

Elevated cancer rates were observed in a region of Massachusetts, USA, known as Upper Cape Cod during the mid-1980s, and one risk factor of interest is a fuel dump at the Massachusetts Military Reservation (MMR) (Kammann and Wand 2003; French and Wand 2004). For nearly twenty years the Massachusetts Department of Public Health (MDPH) has maintained a cancer registry database which records incident cases for 22 types of cancers, including lung, breast and prostate cancers. In this example we focus on female lung cancer between 1986 and 1994.

We use a semi-parametric Poisson spatial model to investigate the relationship between census tract level female lung cancer SIRs and distance to the MMR. Let \( i = 1, \ldots, 45 \) represent the census tracts in the study, and let \( \text{observed}_i \) and \( \text{expected}_i \) be the observed and expected number of incident cases of female lung cancer in tract \( i \) (i.e. numerator and denominator of the SIR), respectively. After fitting a number of models that included terms for additional demographic factors and water source, we arrived at the following semiparametric Poisson spatial model:

\[
\text{observed}_i \sim \text{Poisson} [\text{expected}_i \exp \{ \beta_0 + U_{ci} + \beta_1 x_i + f(\text{dist}_i) \}],
\]

where \( x_i \) is the percentage of women in tract \( i \) who were over 15 and employed outside the home in 1989, and \( \text{dist}_i \) is the distance from the centroid of census tract \( i \) to the centroid of the MMR. Here, \( \mathbf{u}^c = (U_{1c}^c, \ldots, U_{45c}^c)^T \) is a vector of spatially correlated random effects with Gaussian intrinsic autoregression distribution parameterized by variance component \( \sigma_c^2 \), as defined in (4). To complete the specification of the spatial correlation model, we choose a cutoff distance value \( d \) and treat two census tracts as neighbors if the distance between their centroids is less than or equal to \( d \). We choose \( d=7.5 \) kilometers, which corresponds to the cutoff such that every census tract has at least one neighbor. We model the nonparametric term \( f(\text{dist}_i) \) using penalized splines with coefficients \( u_k \sim N(0, \sigma_u^2) \).
3 FITTING VIA MARKOV CHAIN MONTE CARLO

In the general design GLMM (1) and (2) the posterior distribution of
\[ \nu^T \equiv [\beta^T \ u^T], \]
is
\[ [\nu | y] = \frac{\int \exp\{y^T C \nu - 1^T b(C \nu) - \frac{1}{2}(\log |G| + \nu^T V^{-1} \nu)\} |G| dG}{\int \exp\{y^T C \nu - 1^T b(C \nu) - \frac{1}{2}(\log |G| + \nu^T V^{-1} \nu)\} |G| dG d\nu} \]  
(8)
where \( C \equiv [X \ Z], \ V \equiv \text{blockdiag}(F, G), \) and \( |G| \) is the prior on the variance components in \( G \). These integrals are analytically intractable for most problems. Further, in the applications we consider, the dimensionality precludes the use of numerical integration. A standard remedy is to apply a Markov Chain Monte Carlo (MCMC) algorithm to draw samples from (8). An overview of MCMC is provided by Gilks, Richardson and Spiegelhalter (1996).

MCMC methods break up the model parameters into subsets and then sample from the conditional distribution given the remaining parameters and data; often called “full conditionals”. In the general design GLMM the natural breakdown of the parameters is into \( \nu \) and \( G \); leading to the full conditionals:
\[ [\nu | G, y] \quad \text{and} \quad [G | \nu, y]. \]
The latter full conditional has a standard form when the prior on the variance components is inverse gamma or Wishart, which are “conditionally-conjugate” priors for this model, but not when, say, a folded-Cauchy prior is used (e.g. Gelman, 2004). The first full conditional has the general form
\[ [\nu | G, y] \propto \exp\{y^T C \nu - 1^T b(C \nu) - \frac{1}{2} \nu^T V^{-1} \nu\}, \]
which is a non-standard distribution unless \( y \) is conditionally Gaussian. Clever strategies such as adaptive rejection sampling (Gilks and Wild, 1992) and slice sampling (e.g. Besag and Green, 1993; Neal, 2003) are required to draw samples. The most common versions of these algorithms work with the full conditionals of the components \( \nu \). When \( V \) is diagonal, these full conditionals are of the form
\[ [\nu_k | \nu_{-k}, G, y] \propto \exp\{(C^T y)_k \nu_k - 1^T b(c_k \nu_k + C_{-k} \nu_{-k}) - \frac{1}{2} \nu_k^2 / (V)_{kk}\}, \]  
(9)
Here \( c_k \) is the \( k \)th column of \( C \), \( C_{-k} \) is \( C \) with the \( k \)th column omitted, \( \nu_k \) is the \( k \)th entry of \( \nu \); \( \nu_{-k} \) and \( \nu_{-k} \) is \( \nu \) with the \( k \)th entry omitted. It is easily shown that (9) is log-concave, which permits use of adaptive rejection sampling and simplifies slice sampling. These algorithms can also be used to sample from the full conditionals for the variance components when necessary.
A companion paper to this one (Zhao, Staudenmayer, Coull and Wand, 2004) provides a detailed account of MCMC for general design GLMM and compares several strategies via simulation. One of the conclusions drawn from that paper is that the WinBUGS package (Spiegelhalter, Thomas and Best, 2000) performs excellently amongst various “off-the-shelf” competitors. This is very good news since it saves the user having to write his or her own MCMC code. However, it should be noted that for large models WinBUGS can take quite some time to obtain a fit. Also, the analysis must be performed on a particular platform (Windows). But assuming that computation time is not an issue and that Windows is available we can report that, in 2004, fitting of general design GLMMs via WinBUGS has a large chance of success. For our analyses, we had access to several personal computers and ran multiple chains in parallel to assess convergence and prior sensitivity. This reduced the elapsed time it took to compute each one of the analyses by an order of magnitude.

### 4 DATA ANALYSES

#### 4.1 Input Values and Prior Distributions

We used WinBUGS to fit the models described in Section 2.3. However, several input values and prior distributions needed to be specified, so we preface the analyses with the particular choices that were made.

Based on the recommendations of Gelfand et al. (1995) we used hierarchical centering of random effects. All continuous covariates were standardized to have zero mean and unit standard deviation. A strategy such as this is necessary for the method to be scale invariant given fixed choices for the hyperparameters. In the interest of making the fitted functions smooth and requiring fewer knots we used radial cubic basis functions. This corresponds to  

\[
\phi(x) = \beta_0 + \beta_1 x + Z_x u \]

where

\[
Z_x = \left[ \left| x - \kappa_k \right| \right] \left[ \left| \kappa_k - \kappa_{k'} \right| \right] \left[ \left| \kappa_{k'} \right| \right]^{-1/2} \quad \text{and} \quad u \sim N(0, \sigma^2_u I) \quad (10)
\]

(French, Kammann and Wand, 2001) with \( \kappa_k = \left( \frac{k+1}{K+2} \right) \) th quantile of the unique predictor values. In general, \( K \) can be chosen using rules such as

\[
K = \min \left( \frac{1}{4} \text{(number of unique predictor values, 35)} \right)
\]

or those given in Ruppert (2002). However, often considerably smaller \( K \) can be used through experimentation with the benefit of faster MCMC fitting. This approach was taken in our analyses.

We considered several common variance component priors. These were inverse gamma with equal scale and shape:

\[
[\sigma^2] \propto (\sigma^2)^{-a} e^{-a/\sigma^2}
\]
denoted by $\text{IG}(a, a)$, for $a = 0.001, 0.01, 0.1$ (Spiegelhalter et al. 2003) and the folded-$t$ class of priors for $\sigma$ (Gelman, 2004)

$$[\sigma] \propto \left( 1 + \frac{1}{\nu} \left( \frac{\sigma}{s} \right)^2 \right)^{-(\nu+1)/2},$$

where $s$ and $\nu$ are fixed scale and degrees of freedom hyperparameters, respectively. We investigated the sensitivity of the model fit to the choice of the hyperparameters. Results showed that fits based on the IG priors were stable for $a \geq 0.01$, but those obtained assuming $a = 0.001$ behaved erratically. Out of the remaining choices, the folded-Cauchy prior, a member of the folded-$t$ class of priors, performed well.

As a result of these empirical comparisons, in our examples we take the approach of fitting models under multiple priors distributions for the variance components, and assessing the sensitivity of the results to these assumptions. Due to its popularity, we fit general design GLMMs using independent $\text{IG}(0.01, 0.01)$ priors for each variance components. Results suggest this prior performs well for the examples considered in this paper. We also re-fit the models using independent folded-Cauchy prior distributions for each variance component. For a variance component square-root $\sigma$ and fixed scale parameter $s$, the folded-Cauchy distribution has probability distribution

$$[\sigma] \propto (\sigma^2 + s^2)^{-1}.$$  

Following Gelman (2004), we take $s = 25$ in our examples, and check the sensitivity of results to this choice by also fitting the models for $s = 12$. This prior can be implemented in WinBUGS using the flexible feature that allows a user to code an arbitrary prior distribution for the model parameters (see Appendix). We also ran the models using a Uniform(0,100) prior on $\sigma$. A theoretical comparison of such priors in the general design GLMM setting is a topic worthy of future research.

Table 1 summarizes the input values and prior distributions that were used.

### 4.2 Respiratory Infection in Indonesian Children

Using the prior distributions and input values given in Table 1, WinBUGS produced the output for the $\beta$ coefficients summarized in Figure 1. It is seen that, for this model, the chains mix quite well with little significant autocorrelation and Gelman-Rubin $\sqrt{\hat{R}}$ values (Gelman and Rubin, 1992) all less than 1.01. Vitamin A deficiency is seen to have a borderline positive effect on respiratory infection, which is in keeping with previous analyses. Similar comments apply to sex and some of the visit numbers.
Hierarchical centering used for random intercepts.
Continuous covariates standardized to have zero mean and unit standard deviation.
Radial cubic basis functions for smooth functions.

<table>
<thead>
<tr>
<th>coeff.</th>
<th>trace</th>
<th>lag 1</th>
<th>acf</th>
<th>GR</th>
<th>density</th>
<th>summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>vit. A def.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: 0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−0.402, 1.05)</td>
</tr>
<tr>
<td>sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: 0.519</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0131, 1.02)</td>
</tr>
<tr>
<td>height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: −0.0305</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−0.5931, 1.0238)</td>
</tr>
<tr>
<td>stunted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: 0.458</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−0.203, 1.102)</td>
</tr>
<tr>
<td>visit 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: −1.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−1.34, 0.399)</td>
</tr>
<tr>
<td>visit 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: 0.817</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−1.362, 1.102)</td>
</tr>
<tr>
<td>visit 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: −1.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−2.31, 0.519)</td>
</tr>
<tr>
<td>visit 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: 0.437</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−0.101, 1.039)</td>
</tr>
<tr>
<td>visit 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>posterior mean: −0.571</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>95% credible interval:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(−0.824, 0.064)</td>
</tr>
</tbody>
</table>

Table 1: Input values and prior distributions used in WinBUGS for analyses.

Figure 1: Summary of WinBUGS output for parametric components of (5). The columns are: name of variable, trace plot of sample of corresponding coefficient, plot of sample against 1-lagged sample, sample autocorrelation function, Gelman-Rubin $\sqrt{\hat{\kappa}}$ diagnostic, kernel estimates posterior density and basic numerical summaries.
The estimated effect of age is summarized in the top left panel of Figure 2 and is seen to be significant and non-linear. The remaining panels show good mixing of the chains corresponding to the estimated age effect at quartiles of the age data. Gelman-Rubin $\sqrt{R}$ plots (not shown here) support convergence of these chains.

To assess the sensitivity of our conclusions to the choice of variance component priors, we also ran the Gibbs samplers assuming the folded-Cauchy and Uniform priors for the random effects standard deviations (See Section 2.1). Figure 3 shows the posterior estimates and 95% credible intervals for the regression coefficients of interest using the default independent IG priors, independent folded-Cauchy priors with $s = 25$, independent folded-Cauchy priors with $s = 12$, and independent $U(0, 100)$ priors. This figure shows that results are not sensitive to this choice, with the changes in the posterior means never more than 2% of that obtained from the IG specification and the credible intervals never more than 6.5% wider than their IG counterparts.

### 4.3 Caregiver Stress and Respiratory Health

For this example, we also used the priors and input values given in Table 1 and provide WinBUGS code in the Appendix. For the spline, we used twelve knots...
Figure 3: Results of sensitivity analysis for variance component priors for model 5.

Dots are posterior means. Lines are 95% credible intervals.

that were spaced evenly on the percentiles of age. We found that the fit did not change noticeably if we used more knots, and we chose a small number of knots for computational efficiency.

Figure 4 shows the Bayes estimates and credible intervals for the $\beta$ coefficients as well as an assessment of the convergence of the chains. The coefficients can be interpreted as category specific offsets from the population mean. The chains had a moderate autocorrelation and the Gelman-Rubin $\sqrt{\hat{R}}$ values were all less than 1.04. Figure 5 contains the estimated age effect and trace plots for the effect of age at the quartiles of the data. Again, the Gelman-Rubin $\sqrt{\hat{R}}$ values were less than 1.04 and support convergence. The figures are based on the chain that used the independent inverse gamma priors for the variance components. Fits that used independent Cauchy ($s=25$) priors for the square root of the variance components changed neither the posterior means nor the widths of the credible intervals for the parameters of interest by more than 4.7%. The posterior mean and confidence set for $f(\text{age})$ was also relatively insensitive to the prior on the variance components in this example.

Two aspects of the fit that were interesting to the investigators in the study included the inverse dose response relation between income and stress and that race was significantly related to environmental stress even after accounting for the effect of income. The nonparametric estimate of stress as a function of the
child’s age was also interesting and suggests that relatively stressful times include the first few months, when the child is approximately a year old, and beyond age two.

**Figure 4:** Summary of WinBUGS output for parametric components of (6). The columns are: name of variable, trace plot of sample of corresponding coefficient, plot of 1-lagged sample, sample autocorrelation function, Gelman-Rubin $\sqrt{R}$ diagnostic, kernel estimates posterior density and basic numerical summaries. The coefficients can be interpreted as time invariant offsets to the time varying population mean.

<table>
<thead>
<tr>
<th>coeff.</th>
<th>trace</th>
<th>lag 1</th>
<th>acf</th>
<th>GR</th>
<th>density</th>
<th>summary</th>
</tr>
</thead>
</table>
| white or Asian | ![trace plot](image) | ![lag 1 plot](image) | ![acf plot](image) | ![GR plot](image) | ![density plot](image) | posterior mean: 0.0593
95% credible interval: $(0.0134, 0.0950)$ |
| black or Hispanic | ![trace plot](image) | ![lag 1 plot](image) | ![acf plot](image) | ![GR plot](image) | ![density plot](image) | posterior mean: 0.0738
95% credible interval: $(0.0432, 0.1060)$ |
| $50K+$ | ![trace plot](image) | ![lag 1 plot](image) | ![acf plot](image) | ![GR plot](image) | ![density plot](image) | posterior mean: 0.31
95% credible interval: $(0.276, 0.346)$ |
| $15K-$50K | ![trace plot](image) | ![lag 1 plot](image) | ![acf plot](image) | ![GR plot](image) | ![density plot](image) | posterior mean: 0.0022
95% credible interval: $(0.0015, 0.0029)$ |
| <$15K | ![trace plot](image) | ![lag 1 plot](image) | ![acf plot](image) | ![GR plot](image) | ![density plot](image) | posterior mean: 0.083
95% credible interval: $(0.0475, 0.206)$ |

### 4.4 Standardized Cancer Incidence and Proximity to a Pollution Source

As in the previous examples, we started with the prior distributions and inputs in Table 1. In this case though, the chain required a longer burn in. We found that a burn in of length 15,000 was sufficient to produce acceptable convergence. Figure 6 (bottom panel) contains the resulting convergence diagnostics and inferences for the parameters in the model. The middle panel of Figure 6 contains an estimate of the contribution of distance to the MMR to the standardized incidence and trace plots of the function estimate at the quartiles of distance. The
Gelman-Rubin $\sqrt{R}$ values for the estimates at these quartiles were less than 1.04 and support convergence. Finally, the top panel of Figure 6 maps the estimated SIRs based on the model fit, demonstrating the smoothing achieved by the spatial model. The figures are based on fits that used independent IG(0.01,0.01) priors for the variance components. Fits that used independent Cauchy ($s = 25$) priors for the square root of the variance components decreased the length of the credible interval for the effect of percent working by 6.7% and lowered the posterior mean by 3.1%. The posterior mean and confidence set for $f(\text{dist}_i)$ also changed very little.

The fitted model suggests a nominally positive relation between the percent of women who were working outside the home in 1989 and standardized lung cancer incidence rates at the census tract level. Further, the estimated curve $f(\text{dist}_i)$ suggests an increased standardized incidence rate for census tracts that are closer than about 10 km to the MMR after controlling for other factors, and the map suggests that areas immediately east of the MMR exhibit the highest SIRs. None of the estimated effects of the model covariates are strongly significant. Regardless of statistical significance, however, we emphasize that this type of “cancer cluster” study should be viewed as exploratory since the study design is ecological (e.g. Kelsey et al., 1996, Chapter 10). Additionally, re-analyses of similar studies have demonstrated that unmeasured confounders could radi-
5 DISCUSSION

As illustrated by the analyses in the previous section, general design Bayesian GLMM are a very useful structure. In this article we have demonstrated that WinBUGS provides good “off-the-shelf” MCMC fitting of these models. Some of the reviewers have pointed out the possibility of designing MCMC algorithms that take advantage of the special structure of Bayesian GLMMs that is summarized in Section 2. We have done some exploration in this direction (Zhao et al., 2004) but would welcome such research from MCMC specialists. In the meantime, use of WinBUGS is our recommended fitting method.

APPENDIX: WINBUGS CODE

In this Appendix we list the WinBUGS code used for the data analyses of Sections 4. Note that the spline basis functions and hyperparameters are inputs.

The following code was used for fitting (5) to the data on respiratory infection of Indonesian children. Here inverse gamma priors are used on all variance components.

```r
model
{
  for (i in 1:num.obs)
  {
    X[i,1] <- age[i]
    X[i,2] <- vitAdefic[i]
    X[i,3] <- sex[i]
    X[i,4] <- height[i]
    X[i,5] <- stunted[i]
    X[i,6] <- visit2[i]
    X[i,7] <- visit3[i]
    X[i,8] <- visit4[i]
    X[i,9] <- visit5[i]

    logit(mu[i]) <- gamma[subject[i]] + inprod(beta[],X[i,])
                  + inprod(u.spline[],Z.spline[i,])
    resp[i] ~ dbern(mu[i])
  }

  for (i.subj in 1:num.subj)
  {
    gamma[i.subj] <- beta0 + u.subj[i.subj]
    u.subj[i.subj] ~ dnorm(0,tau.u.subj)
  }

  for (k in 1:num.knots)
  {
```
Figure 6: Summary of WinBUGS output for the fit of (7). The top panel contains a spatial plot of the smoothed SIRs (posterior means of the \( U_i \)'s). The middle panel shows the estimated \( f(\text{dist}) \) and a corresponding pointwise 95% credible set along with trace plots of the samples of the function at the quartiles of distance. The bottom panel displays summaries of other parameters of interest and convergence diagnostics. Additionally, the Gelman-Rubin \( \sqrt{\hat{R}} \) diagnostics were less than 1.03 for all the \( U_i \)'s and for the \( f(\text{dist}) \) at the quartiles of distance. The MMR is the area in the center of the map that is excluded from the analysis.

<table>
<thead>
<tr>
<th>coeff.</th>
<th>trace</th>
<th>lag 1</th>
<th>acf</th>
<th>GR</th>
<th>density</th>
<th>summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>% working (1999)</td>
<td>![trace plot]</td>
<td>![lag 1 plot]</td>
<td>![acf plot]</td>
<td>![GR plot]</td>
<td>![density plot]</td>
<td>![summary plot]</td>
</tr>
<tr>
<td>( \log(a) )</td>
<td>![trace plot]</td>
<td>![lag 1 plot]</td>
<td>![acf plot]</td>
<td>![GR plot]</td>
<td>![density plot]</td>
<td>![summary plot]</td>
</tr>
</tbody>
</table>
The following code was used for fitting (6) to the data on caregiver stress and respiratory health. This code illustrates the use of folded-Cauchy priors on variance components. As noted in the WinBUGS user manual (Spiegelhalter et al., 2000), a single zero Poisson observation with mean $\phi$ contributes a term $\exp(\phi)$ to the likelihood for $\sigma$, which is then combined with a flat prior over the positive real line to produce the folded-Cauchy distribution.

model
{
  for (i in 1:num.obs)
  {
    X[i,1] <- age[i]
    X[i,2] <- income1[i]
    X[i,3] <- income2[i]
    X[i,4] <- race[i]
    log(mu[i]) <- gamma[house[i]] + inprod(beta[],X[i,])
                  + inprod(u.spline[],Z.spline[i,])
    y[i] ˜ dpois(mu[i])
  }
  for (i.house in 1:num.house)
  {
    gamma[i.house] <- beta0+u.subj[i.house]
    u.subj[i.house] ˜ dnorm(0,tau.u.subj)
  }
  for (k in 1:num.knots)
  {
    u.spline[k] ˜ dnorm(0,tau.u.spline)
  }
  beta0 ˜ dnorm(0,tau.beta)
  for (j in 1:num.pred)
  {
    beta[j] ˜ dnorm(0,tau.beta)
  }
  tau.u.spline <- pow(sigma.u.spline,-2)
  zero.u.spline <- 0
  sigma.u.spline ˜ dunif(0,1000)
  phi.u.spline <- log((pow(sigma.u.spline,2)+pow(phi.scale.u.spline,2)))
  zero.u.spline ˜ dpois(phi.u.spline)
  tau.u.subj <- pow(sigma.u.subj,-2)
  zero.u.subj <- 0
  sigma.u.subj ˜ dunif(0,1000)
phi.u.subj <- log((pow(sigma.u.subj,2)+pow(phi.scale.u.subj,2)))
zero.u.subj ~ dpois(phi.u.subj)

Below is the code that we used to fit the spatial model (7) to the Cape Cod female lung cancer data. Please note that the variance components have inverse gamma priors, and adj, weights, and num are inputs to car.normal, the normal conditional autoregressive function in WinBUGS.

model
{
  for (i in 1:num.regions)
  {
    X[i,1] <- working[i,1]
    X[i,2] <- distance[i,1]
    theta[i] <- beta0+u.spatial[i]+inprod(beta[],X[i,])+inprod(u.spline[],Z.spline[i,])
    log(mu[i]) <- log(E[i])+theta[i]
    O[i] ~ dpois(mu[i])
    SIRhat[i] <- 100*mu[i]/E[i]
  }
  u.spatial[1:num.regions]~car.normal(adj[],weights[],num[],tau.u.spatial)
  for (k in 1:num.knots)
  {
    u.spline[k] ~ dnorm(0.0,tau.u.spline)
  }
  for (j in 1:num.pred)
  {
    beta[j] ~ dnorm(0.0,tau.beta)
  }
  beta0 ~ dnorm(0.0,tau.beta)
  tau.u.spatial~dgamma(A.u.spatial,B.u.spatial)
  tau.u.spline~dgamma(A.u.spline,B.u.spline)
}

ACKNOWLEDGMENTS

We are grateful for comments from Ciprian Crainiceanu, Jim Hodges, Scott Sisson, the editor, associate editor and two referees. This research was partially supported by U.S. National Institute of Environmental Health Sciences grant R01-ES10844-01, U.S. National Science Foundation grant NSF-DMS 0306227 and U.S. National Institutes of Health grant ES012044.


