Bayesian prediction of spatial count data using generalised linear mixed models

Ole F. Christensen* **and Rasmus Waagepetersen**

Department of Mathematical Sciences, Aalborg University,

Fredrik Bajers Vej 7G, DK-9220 Aalborg.

**email:* `olefc@math.auc.dk`

November 21, 2001

Summary. Spatial weed count data are modelled and predicted using a generalised linear mixed model combined with a Bayesian approach and Markov chain Monte Carlo. Informative priors for a data set with sparse sampling are elicited using a previously collected data set with extensive sampling. Furthermore, we demonstrate that so-called Langevin-Hastings updates are useful for efficient simulation of the posterior distributions, and we discuss computational issues concerning prediction.

Key words: Bayesian inference; generalised linear mixed model; geostatistics; informative prior; Langevin-Hastings update; Markov Chain Monte Carlo; prediction; weed intensity.

1. Introduction

Site specific farming (also known as precision farming/agriculture) is aiming at targeting inputs of fertiliser, pesticide, and herbicide according to locally determined requirements. In connection with herbicide application on a field, it is important to map the weed intensity so that the dose of herbicide applied at any location can be adjusted to the amount of weed present at the location. In a Danish project on site specific farming (Olesen, 1997) one objective was to investigate whether observations of soil properties could be used for prediction of weed intensity. In practice the farmer or his advisor should then establish a relation between soil properties and weed occurrence from extensive observations collected one year and use this for prediction of the weed intensity in subsequent years where only a limited number of weed count observations would be collected. Many soil properties are fairly constant over time so that observations of soil samples obtained the first year can also be used in subsequent years.

The statistical tasks in relation to the above-mentioned project are 1) to establish a model for the weed counts given the soil properties, and 2) to predict weed counts for one year with sparse sampling using the information obtained from a previous year with extensive sampling. Traditional geostatistical methods for prediction of spatial variables (see Cressie, 1993) are based on an assumption of normality which is not valid for count data. We instead use the approach in Diggle, Tawn and Moyeed (1998) where spatial data are analysed using generalised linear mixed models (GLMMs) and Bayesian methods. As in Diggle, Tawn and Moyeed (1998) we use Markov chain Monte Carlo (MCMC) for posterior computation but consider a more efficient MCMC algorithm based on so-called Langevin-Hastings updates.

The Bayesian approach is well suited for our prediction problem since informative priors can be constructed using information from previous years. We also discuss the choice of vague priors and propose alternatives to the uniform priors considered in Diggle, Tawn and Moyeed (1998). The weed intensity predictions and prediction intervals are obtained from quantiles of the predictive distribution. It turns out that low weed intensities can be predicted accurately. This is important for site specific herbicide application where it is of particular interest to know where spraying is not required.

The paper is organised as follows. The weed data are described in Section 1.1, and Section 2 discusses a model for the weed data and surveys spatial GLMMs. The algorithms for posterior simulation and prediction are described in Section 3. Finally, Section 4 discusses prior elicitation for the weed application, and presents posterior results and weed prediction at unobserved locations.

1.1 *Description of data*

In the project described in Olesen (1997) counts of weed plants on a field were recorded in 1993, 1994 and 1995; the data are partly presented and analysed in Walter, Heisel and Christensen (1997). One objective of the project was to investigate whether weed occurrence could be predicted from observations of soil texture and soil chemical properties. Christensen, Møller and Waagepetersen (2000) considered the counts of the species *Viola arvensis* for 1994 and demonstrated that the intensity of this species is related to the percentage of organic matter in the soil, and that there is a North-South decreasing trend in the data. The percentage of organic matter is known to be fairly constant from one year to another, so the information gained from the 1994 analysis should be helpful for prediction of counts of the species *Viola arvensis* in year 1995.

The left hand plot in Figure 1 displays the observed locations, and the percentages of organic matter in a soil sample at these locations. The contiguous locations with high values correspond to a peatbog. The weed counts from 1995 are displayed in the right hand plot in Figure 1, where observations with a circle are referred to as the reduced weed data set and are used in Section 4.4 for prediction of the remaining observations. The counts are observed within $0.25\,m^2$ circular frames which, except for some missing sites in the first row, form a $20m \times 20m$ rectangular grid.

[Figure 1 about here.]

2. Modelling

Generalised linear mixed models (GLMMs) (Breslow and Clayton, 1993) are extensions of generalised linear models (McCullagh and Nelder, 1989) that allow additional sources of variability due to unobservable random effects. We consider a GLMM for the weed data where the random effects are modelled by a spatially correlated Gaussian field.

2.1 *Model for weed data*

Let $I \subset \mathbb{R}^2$ denote the field where the weed data are observed. If one considers the unobserved positions of weed plants as a realisation of a spatial point process then a key summary of weed occurrence is the weed intensity $\lambda(\cdot)$. For a location $x \in I$ and some small region with area A around x, $\lambda(x)A$ is approximately the expected number of weed plants in the region. For each $x \in I$ we model $\lambda(x)$ as

$$\lambda(x) = \exp(S(x) + d(x)^{\mathrm{T}}\beta), \tag{1}$$

where $d(x)$ is the covariate vector associated with x, $\beta \in \mathbb{R}^p$ is a vector of regression parameters, and $S(x)$ is a random effect which serves to model sources of variation not included in $d(x)^{\mathrm{T}}\beta$. The random effects $\mathcal{S} = \{S(x) : x \in I\}$ are assumed to form a correlated zero-mean Gaussian random field, so that it is possible to model correlation in the weed intensity due to e.g. the reproduction mechanism of the weed or unobserved spatially structured covariates.

Let $Y(x)$ denote the weed count (observed or unobserved) in a frame centered at a location x. At many locations in Figure 1 the counts equal zero, so assuming a normal distribution is certainly inappropriate, even for transformed data. Following the considerations above we model $Y(x)$, $x \in I$, conditionally on $\mathcal{S}$, as independent Poisson distributed random variables with conditional mean $M(x) = 0.25\lambda(x)$. The resulting Poisson-log normal model is an example of a spatial GLMM with Poisson error distribution and a log link function, see Section 2.2.

The Poisson-log normal model can also be seen as an approximation to the distribution of the counts under a so-called log Gaussian Cox point process (Møller, Syversveen and Waagepetersen, 1998; Brix and Møller, 2001) with log normal random intensity surface $\{\lambda(x) : x \in I\}$ given by (1). Under a log Gaussian Cox process and given $\lambda(\cdot)$, the count

$Y(x)$ is Poisson distributed with mean $\int_{B_{0.25}} \lambda(x-z)dz$ where B_a denotes a circle with center 0 and area a.

2.2 *Spatial generalised linear mixed models*

Our model is a special case of a spatial GLMM, which is the framework proposed in Diggle et al. (1998) for modelling of non-Gaussian spatial data. For a spatial GLMM, the random variables $Y(x)$, $x \in I$, are mutually independent given $\mathcal{S}$, and for each $x \in I$, the error distribution $[Y(x)|\mathcal{S}]$ has a density $f(\cdot; M(x))$ which only depends on the conditional mean $M(x) = \mathrm{E}[Y(x)|S(x)]$. Moreover, $M(x)$ is related to the linear predictor $S(x) + d(x)^{\mathrm{T}}\beta$ by a strictly increasing link function g so that

$$M(x) = m(x)g^{-1}\big(S(x) + d(x)^{\mathrm{T}}\beta\big), \tag{2}$$

where $m(x)$ is a known function determined by the design of the experiment. Spatial GLMMs is thus a flexible class of models for e.g. spatially correlated count, binary, and positive data.

Often the Gaussian field $\mathcal{S}$ is assumed to be isotropic and stationary, i.e., the covariance function

$$C(u) = \mathrm{E}\left[S(x)S(x')\right]$$

depends only on the distance $u = ||x - x'||$ between locations $x, x' \in I$. The covariance function is modelled using a positive semi-definite function as in traditional geostatistics, see e.g. Cressie (1993). We consider parametric models

$$C(u) = \sigma^2\rho(u/\alpha), \quad u \geq 0, \tag{3}$$

where ρ is a known correlation function and the parameter $(\sigma, \alpha) \in]0;\infty[^2$ is unknown; α is a correlation scale parameter and σ^2 is the variance.

An alternative would be to model $\mathcal{S}$ by a conditional autoregression (Besag and Kooperberg, 1995). This is advantageous for computational reasons, but on the other hand imposes the restriction that one can only predict or model weed counts located on a predetermined grid.

As in Diggle et al. (1998) we use a Bayesian approach to inference where priors are imposed on the unknown regression parameters and the covariance parameters of the Gaussian random field.

3. Posterior simulation and prediction

In this section we discuss Bayesian inference and prediction for spatial GLMMs. We focus on the Poisson-log normal model, but the techniques are readily applicable to other spatial GLMMs. For the $n = 70$ observations in the reduced weed count data set (see Figure 1) we let $x_1, \ldots, x_n \in I$ denote the centers of the observation frames. We further let $S_i = S(x_i)$, $i = 1, \ldots, n$, $S = (S_1, \ldots, S_n)^{\mathrm{T}}$, and denote the observed realisation of $Y = (Y(x_1), \ldots, Y(x_n))^{\mathrm{T}}$ by $y = (y_1, \ldots, y_n)^{\mathrm{T}}$. For locations $x_{n+1}, \ldots, x_{n+q}$, $q \geq 0$, of interest for prediction, we let $S^* = (S(x_{n+1}), \ldots, S(x_{n+q}))^{\mathrm{T}}$.

Note that S^* and Y are conditionally independent given $(S, \beta, \sigma, \alpha)$. Our Bayesian inference is therefore separated into first posterior simulation of $(S, \beta, \sigma, \alpha)$ given Y and secondly prediction. In Section 3.1 we discuss Metropolis-Hastings algorithms for posterior simulation, and in Section 3.2 we describe our approach to prediction.

3.1 *Posterior simulation for spatial GLMMs*

Briefly, a Metropolis-Hastings algorithm iteratively generates an ergodic Markov chain which yields the posterior sample. In each step a proposal is generated for an update of the current state of the chain. The update is then accepted or rejected according to a certain acceptance probability. The most commonly used algorithm is Gaussian random walk Metropolis, where the proposal distribution is Gaussian with mean equal to the present state. For details on Markov chain Monte Carlo we refer to Robert and Casella (1999).

Diggle et al. (1998) use a so-called fixed scan hybrid algorithm where the covariance parameters, the regression parameters, and each of the random effects $S_1, \ldots, S_n$ are updated in turn in each scan. The update of a random effect S_i is computationally demanding, since it involves the calculation of the conditional variance given the $n-1$ other random effects.

In Section 3.1.2 we instead consider a fixed scan hybrid algorithm where transformed random effects are updated simultaneously using so-called truncated Langevin-Hastings updates. The truncated Langevin-Hastings update utilises the gradient of the log posterior, and can as exemplified in Section 4.2 lead to much reduced Monte Carlo errors compared to the standard alternative of a random walk Metropolis update. It is furthermore easy to implement the Langevin-Hastings update for any spatial GLMM with a canonical link function.

Gamerman (1997) considers random effect updates based on the iterative weighted least squares algorithm for GLMMs. These updates utilise both first and second order derivatives of the log posterior and would imply considerable additional computational complexity for spatial GLMMs.

3.1.1 Reparametrisation Let $\sigma^2\kappa(\alpha)$ denote the covariance matrix of S and let $Q(\alpha)$ be a square-root so that $\kappa(\alpha) = Q(\alpha)(Q(\alpha))^{\mathrm{T}}$ where the square-root for instance may be obtained by Cholesky factorisation. We can then assume that $S = \sigma Q(\alpha)\Gamma$, where Γ follows a n-dimensional standard multivariate Gaussian distribution. Posterior simulations of S can be obtained by transforming MCMC samples of the conditional distribution of Γ given $Y = y$. It is demonstrated in Møller et al. (1998) that this reparametrisation of the random effects can lead to better mixing properties of the MCMC algorithm. In Section 4 β is a priori Gaussian with mean μ_b and covariance matrix Σ_b, so that a reparametrisation $\beta = \mu_b + \Sigma_b^{1/2}\epsilon$, $\epsilon \sim N_p(0, I)$, can be applied to β also. In the weed count application we found that the use of reparametrisation for β gave much better mixing Markov chains.

3.1.2 The hybrid Metropolis-Hastings algorithm In our MCMC algorithm, γ, ϵ, $\log\sigma$, and $\log\alpha$ are updated in turn in each scan using either truncated Langevin-Hastings or random walk Metropolis updates as discussed below.

As in Section 4, suppose that the priors for α, β, σ are independent with prior densities π_a, π_b, π_c, respectively, where π_b is the normal density described in Section 3.1.1. The log posterior density of $(\Gamma, \epsilon, \sigma, \alpha)$ given $Y = y$ is

$$\log f(\gamma, \epsilon, \sigma, \alpha|y) = \text{const}(y) + \log \pi_a(\alpha) - \frac{1}{2}\|\epsilon\|^2 + \log \pi_c(\sigma) - \frac{1}{2}\|\gamma\|^2 + \sum_{i=1}^{n} y_i \log(\mu_i) - \mu_i,$$

with $\mu_i = m(x_i)\exp(s_i + d(x_i)^{\mathrm{T}}\beta)$, where $(s_1, \ldots, s_n)^{\mathrm{T}} = \sigma Q(\alpha)\gamma$, and $\beta = \mu_b + \Sigma_b^{1/2}\epsilon$.

The truncated Langevin-Hastings proposal distribution for γ is $N(\gamma + (h/2)\nabla(\gamma)^{\text{trunc}}, hI)$, where h is a user specified "proposal variance",

$$\nabla(\gamma)^{\text{trunc}} = -\gamma + \sigma Q(\alpha)^{\mathrm{T}} \{y_i - \mu_i \wedge H\}_{i=1}^{n}$$

is obtained by truncating the gradient (with respect to γ) of the log target density, and $0 < H < \infty$ is a truncation constant. A similar simple form is available for any GLMM with g equal to the canonical link function (see McCullagh and Nelder, 1989). A theoretical study on MCMC algorithms with truncated Langevin-Hastings updates for conditional simulation in GLMMs for fixed model parameters is given in Christensen, Møller and Waagepetersen (2001).

For ϵ we similarly use a truncated Langevin-Hastings update where the truncated gradient for ϵ takes the form $-\epsilon + \Sigma_b^{1/2} D^{\mathrm{T}} \{y_i - \mu_i \wedge H\}_{i=1}^{n}$.

The derivative of $\log f(y|\gamma; \beta, \sigma, \alpha)$ with respect to α is complicated, since it involves differentiating the matrix $Q(\alpha)$. In addition Langevin-Hastings updates are most useful for updating high-dimensional quantities (Roberts and Rosenthal, 1998). Therefore, for the one-dimensional parameters $\log\sigma$ and $\log\alpha$ we use standard Gaussian random walk Metropolis updates.

A C-implementation of the MCMC algorithm described above can be downloaded from `http://www.math.auc.dk/~olefc/Programs/mcmcglmm/mcmcglmm.html`.

3.2 *Bayesian prediction*

Here we consider prediction of (functionals of) $(M(x_{n+1}), \ldots, M(x_{n+q}))^{\mathrm{T}}$, where $M(x_i)$, $i = n+1, \ldots, n+q$, are given by (2). The predictive distribution of $M^* = (M(x_{n+1}), \ldots, M(x_{n+q}))^{\mathrm{T}}$ is given by

$$[M^* \mid y] = \mathrm{E}\left[g^{-1}(S^* + D^*\beta) \mid S, \beta, \alpha, \sigma) \mid y\right], \tag{4}$$

where $D^* = (d(x_{n+1}) \cdots d(x_{n+q}))^{\mathrm{T}}$, $S^* = (S(x_{n+1}), \ldots, S(x_{n+q}))^{\mathrm{T}}$, and g^{-1} is applied coordinatewise in (4).

The observed data Y are conditionally independent of S^* given $(S, \beta, \sigma, \alpha)$. An algorithm to obtain a sample $(M^{*kj})_{k=1,\ldots,K, j=1,\ldots,J}$ from the predictive distribution $[M^* \mid y]$ is thus

1. Sample $(S^k, \beta^k, \sigma^k, \alpha^k)_{k=1,\ldots,K}$ from the posterior distribution $[S, \beta, \alpha, \sigma \mid y]$ using the MCMC-algorithm in Section 3.

2. For each $k = 1, \ldots, K$, let $M^{*kj} = g^{-1}(S^{*kj} + D^*\beta^k)$, $j = 1, \ldots, J$, where S^{*kj}, $j = 1, \ldots, J$ are sampled independently from the multivariate Gaussian distribution $[S^* \mid (S^k, \beta^k, \sigma^k, \alpha^k)]$.

Consider for the moment just one location, say x_{n+1}. Common practise is to use $\mathrm{E}[M(x_{n+1}) \mid y]$ as predictor for $M(x_{n+1})$ and $\mathrm{E}[M(x_{n+1}) \mid y] \pm 1.96\sqrt{\mathrm{Var}[M(x_{n+1}) \mid y]}$ as an approximate prediction interval. However, when $g(\mu) = \log(\mu)$ and a vague prior is used for σ, then the mean and the variance of $[M(x_{n+1}) \mid S]$ may be infinite (De Oliveira, Kedem and Short, 1997) whereby also $\mathrm{E}[M(x_{n+1}) \mid y]$ becomes infinite. Further, when the predictive distribution is skew, the use of $\mathrm{E}[M(x_{n+1}) \mid y] \pm 1.96\sqrt{\mathrm{Var}[M(x_{n+1}) \mid y]}$ as a prediction interval is not appropriate. Here we use the median $\mathrm{Med}[M(x_{n+1}) \mid y]$

as predictor for $M(x_{n+1})$ and $[\zeta_{2.5\%}(x_{n+1}); \zeta_{97.5\%}(x_{n+1})]$ as the prediction interval, where $\zeta_\nu(x_{n+1})$ is the ν quantile of $[M(x_{n+1}) \mid y]$.

Also prediction of the observations $Y^* = (Y(x_{n+1}), \dots, Y(x_{n+q}))$ may be of interest (see Section 4.4). The predictive distribution for $Y(x_{n+1})$ is given by the probabilities

$$p(z|y) = P(Y(x_{n+1}) = z \mid y) = \mathrm{E}\left(\mathrm{E}[f(z; M(x_{n+1})) \mid S, \beta, \alpha, \sigma] \mid y\right), \; z = 0, 1, \dots$$

where $f(\cdot; m)$ is the Poisson probability function parametrised by the mean m. As an approximate $1 - \psi$ predictive set for $Y(x_{n+1})$ we use $\{z : \psi/2 \le \sum_{l=0}^{z-1} p(l|y) + p(z|y)/2 \le 1 - \psi/2\}$, $0 < \psi < 1$.

For some prediction problems the second step in the above-mentioned algorithm is unnecessary. Consider for instance prediction of the probability that $M(x_{n+1})$ exceeds a given threshold c. The predictor is

$$P(M(x_{n+1}) > c \mid y) = \mathrm{E}[P(S(x_{n+1}) > g(c) - d^{\mathrm{T}}(x_{n+1})\beta) \mid S, \beta, \alpha, \sigma) \mid y],$$

where $P(S(x_{n+1}) > g(c) - d^{\mathrm{T}}(x_{n+1})\beta) \mid S, \beta, \alpha, \sigma)$ can be computed without Monte Carlo integration.

4. The weed data application

For the weed data we use the organic matter variable and the second spatial coordinate as explanatory variables. The percentages of organic matter are transformed using the logit transform, and the explanatory variables are further standardised by first subtracting the mean and secondly scaling to have maximal absolute value equal to one (this standardisation improves mixing of the MCMC algorithm). Thereby two explanatory variables $d_1(x)$ and $d_2(x)$ are obtained for the locations x given in Figure 1. We further include an intercept β_0 in the regression parameter $\beta = (\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$ so that $d(x) = (1, d_1(x), d_2(x))^{\mathrm{T}}$ is the covariate vector associated to x.

The weed data are modelled using the Poisson-log normal model specified in Section 2.1 with the exponential correlation function $\rho(u) = \exp(-u)$ for the Gaussian random field. We do not believe that much information concerning the type of correlation function is available from our data set, as very close observations do not occur, and the Gaussian random field is not directly observed. For the sake of parsimony we therefore do not include uncertainty concerning the choice of correlation model in our inference.

In Section 4.1 we discuss elicitation of priors for the model parameters, in Section 4.2 we discuss computational issues, Section 4.3 contains posterior distributions of the parameters, and Section 4.4 concerns prediction of the weed intensity.

4.1 *Priors for the weed data*

For the reduced weed dataset in 1995, informative priors are obtained by approximating the posterior distributions from the 1994 data analysis by normal or log-normal distributions. The analysis for the 1994 data was performed with a flat prior $\pi_b^{94}(\beta) \propto 1$ for β, and an improper inverse gamma prior $\pi_c^{94}(\sigma) = 1/\sigma \exp(-10^{-6}/\sigma)$ - this reflects the absence of prior knowledge in 1994. However, for the correlation scale parameter α in (3), the use of an improper prior will typically result in an improper posterior, see p. 224 in Stein (1999) for the case of spatial Gaussian models - a similar argument can be given for spatial GLMMs. Therefore, we used a proper exponential prior $\pi_a^{94}(\alpha) = \exp(-\alpha)$ for α. That the 1994 priors yield a proper posterior is verified by Proposition 1 in Christensen et al. (2000). Condition 2 regarding α in this proposition is seen to hold by combining formulas (14) and (17) in Berger, De Oliveira and Sanso (2001), where the function $\nu(\alpha)$ equals $1/\alpha$ for the exponential correlation model. The resulting independent priors for 1995 are

$$\pi_a^{95}(\alpha) \sim LN(3.00, 0.89); \;\; \pi_c^{95}(\sigma) \sim LN(0.0049, 0.082); \tag{5}$$
$$\pi_b^{95}(\beta) \sim N_3\left(\begin{pmatrix}0.51\\2.04\\2.06\end{pmatrix}, \begin{pmatrix}0.1536 & -0.0603 & 0.0557\\-0.0603 & 0.1267 & 0.0164\\0.0557 & 0.0164 & 0.6386\end{pmatrix}\right),$$

where $LN(\cdot,\cdot)$ denotes a log-normal distribution parametrised in the usual way.

If we disregard for a moment the approximation (5) of the posteriors from 1994, then our approach corresponds to a joint modelling of the weed data for 1994 and 1995 with separate fields of random effects for each year but common values of the regression and covariance parameters. The posteriors presented in Section 4.3 corresponds to posteriors in the joint model given data from both 1994 and 1995 and with the priors equal to the ones used in 1994. Alternatively, one might consider using a larger prior variance for β_0 in $\pi_a^{95}(\beta)$ in order to accommodate possible large year to year changes in the overall weed intensity level.

4.2 *Computations*

For the posterior simulations we use 100,000 scans of the hybrid algorithm, where $Q(\alpha)$ is computed using Cholesky factorisation, the truncation constant in the gradients for γ and ϵ is $H = 50$, and we obtain a Monte Carlo sample of size 10000 by subsampling every 10th scan. The computational cost of the Cholesky factorisation required for each scan is moderate in our application since $Q(\alpha)$ is only a 70×70 matrix.

Theoretical results in Roberts, Gelman and Gilks (1997) and Roberts and Rosenthal (1998) suggest that one should tune the proposal variances to obtain acceptance rates around 0.23 for random walk updates and 0.57 for Langevin-Hastings updates. The overall acceptance rates for the updates of γ, ϵ, $\log\sigma$ and $\log\alpha$ were 0.63, 0.61, 0.30, and 0.39, respectively.

Using the initial monotone sequence estimate in Geyer (1992), we have computed Monte Carlo standard errors for estimates of the posterior means of the model parameters given in Section 4.3, and $S_i + d(x_i)^{\mathrm{T}}\beta$, $i = 1,\ldots,n$. The Monte Carlo standard errors for β_0, β_1, β_2, $\log(\sigma)$, $\log(\alpha)$, are 0.005, 0.007, 0.010 0.004 and 0.020, respectively, and the Monte Carlo errors for $S_i + d(x_i)^{\mathrm{T}}\beta$, $i = 1,\ldots,n$, range between 0.005 and 0.019.

As an alternative to the hybrid Langevin-Hastings/random walk (LH/RW) algorithm, we consider another hybrid algorithm (RW) where random walk Metropolis updates are used for both γ, ϵ, $\log\sigma$, and $\log\alpha$ (with acceptance rates close to 0.23). For the model parameters, the reduction in Monte Carlo standard error obtained by using LH/RW instead of RW is smallest for β_1 (a factor 0.87) and largest for β_2 (a factor 0.45). For $S_i + d(x_i)^{\mathrm{T}}\beta$, $i = 1, \ldots, n$, the reduction factor is between 0.18 and 0.42. The Monte Carlo standard errors are proportional to $1/\sqrt{K}$, where K is the number of subsampled scans. This means that up to $(1/0.18)^2 \approx 31$ times more scans are required for the RW algorithm to obtain the same precision as for the LH/RW algorithm. Notice that the use of Langevin-Hastings updates for γ and ϵ instead of random walk updates also leads to reduced Monte Carlo errors for the covariance parameters $\log\sigma$ and $\log\alpha$. Since the computing times of the two algorithms are almost the same, it is clear that the LH/RW-algorithm performs better than the RW-algorithm.

4.3 *Posterior distributions*

[Figure 2 about here.]

Posterior histograms for the parameters are shown in Figure 2 together with curves indicating the prior distribution. The posterior for β_2 associated with organic matter is similar to the marginal prior whereas the posteriors for β_0 and β_1 compared to the priors are shifted to the left and right, respectively. Posterior means and variances of β_0, β_1, β_2, $\log\sigma$, and $\log\alpha$ are $0.13, 2.46, 1.88, -0.18, 3.92$, and $0.077, 0.101, 0.295, 0.049, 0.699$, respectively.

4.4 *Prediction of weed intensity*

The object of primary interest in connection with weed mapping is the weed intensity $\lambda(x_i)$. In Figure 3, predictive medians $\mathrm{Med}[\lambda(x_i) \mid y]$ and the widths of the prediction intervals are presented for $x_1 \ldots, x_{n+q}$ on the $20 \times 20m$ grid where $x_{n+1}, \ldots, x_{n+q}$ are the locations in Figure 1 without a circle. Note that the prediction intervals are wide where the predictions are large. This is to be expected since the log-link is used.

[Figure 3 about here.]

In order to study the validity of the predictive distribution we compare the weed counts $Y(x_i)$, $i = n+1, \dots, n+q$ with their predictive distributions.

[Figure 4 about here.]

Figure 4 shows the predictive intervals for $M(x_i)$ and the predictive distribution for $Y(x_i)$ at 3 selected locations. The prediction uncertainty for $Y(x_i)$ is large whenever $\mathrm{Med}[M(x_i) \mid y]$ is large and this agrees with the fact that the magnitude of the prediction error $Y(x_i) - \mathrm{Med}[M(x_i) \mid y]$ is positively correlated with $\mathrm{Med}[M(x_i) \mid y]$ (see Figure 4). Among the locations $x_{n+1}, \dots, x_{n+q}$, 4% of the weed counts fall outside the approximate 90% predictive sets for $Y(x_i)$, and 1.5% outside the approximate 95% predictive sets. We have further observed that the 8 observations outside the approximate 90% predictive sets do not form a systematic pattern within the field (figure omitted). The predictive sets thus appear to give a reliable although perhaps a bit conservative measure of the prediction uncertainty.

For site specific herbicide application it is important to know with high certainty where spraying is not required. If a high weed intensity is predicted then the precise value of the intensity is less important. So the fact that the prediction uncertainty is an increasing function of the predicted value does not appear to be a serious problem in practice.

Acknowledgements

This research was supported by the European Union's network "Statistical and Computational Methods for the Analysis of Spatial Data. ERB-FMRX-CT96-0096", and by the Danish Informatics Network in the Agricultural Sciences, funded by a grant from the Danish Research Councils. We wish to thank Svend Elsnab Olesen, Svend Erik Simmelsgaard, and Anne Mette Walter from the Danish Institute of Agricultural Sciences for providing the weed count data, Preben Klarskov Hansen, Danish Institute of Agricultural Sciences, for helpful discussions on herbicide application, and our colleague Jesper Møller for helpful comments on a previous version of the manuscript. The constructive comments from two referees and an associate editor are gratefully acknowledged.

References

Berger, J. O., De Oliveira, V. and Sanso, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statisticial Association* (to appear).

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.

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

Brix, A. and Møller, J. (2001). Space-time multitype log Gaussian Cox processes with a view to modelling weed data. *Scandinavian Journal of Statistics* **28**, 471–488.

Christensen, O. F., Møller, J. and Waagepetersen, R. (2000). Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo. Research Report R-00-2009, Department of Mathematical Sciences, Aalborg University.

Christensen, O. F., Møller, J. and Waagepetersen, R. (2001). Geometric ergodicity of Metropolis Hastings algorithms for conditional simulation in generalised linear mixed models. *Methodology and Computing in Applied Probability* (to appear).

Cressie, N. A. C. (1993). *Statistics for spatial data.* Revised edition. Wiley, New York.

De Oliveira, V., Kedem, B. and Short, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statisticial Association* **92**, 1422–1433.

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

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* **7**, 57–68.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.* **7**, 473–511.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models.* Chapman and Hall, London, second edition.

Møller, J., Syversveen, A. R. and Waagepetersen, R. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.

Olesen, S. E. (1997). *Positionsbestemt dyrkning (forskningsprojekt vedr. gradueret plantedyrkning 1992-96).* Number 17 in SP Rapporter. Danmarks JordbrugsForskning. in Danish.

Robert, C. P. and Casella, G. (1999). *Monte Carlo statistical methods.* Springer-Verlag, New York.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability* **7**, 110–120.

Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statisticial Society, Series* B **60**, 255–268.

Stein, M. L. (1999). *Interpolation of spatial data, some theory for kriging.* Springer Verlag, New York.

Walter, A. M., Heisel, T. and Christensen, S. (1997). Shortcuts in weed mapping. In Stafford, J. V., editor, *Precision agriculture 1997*, pages 777–784. BIOS Scientific Publishers Ltd.

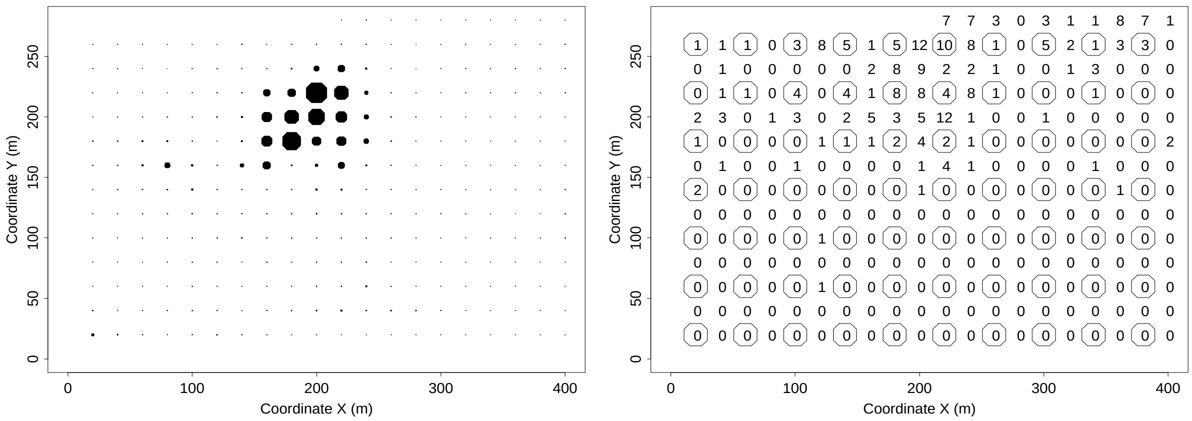

Figure 1. Left: The observed locations, where the size of the dot correspond to the percentage of organic matter (values between 1.6 and 38.8). Right: Counts of *Viola arvensis* in 1995, where observations marked with a circle are used for predicting the remaining observations.

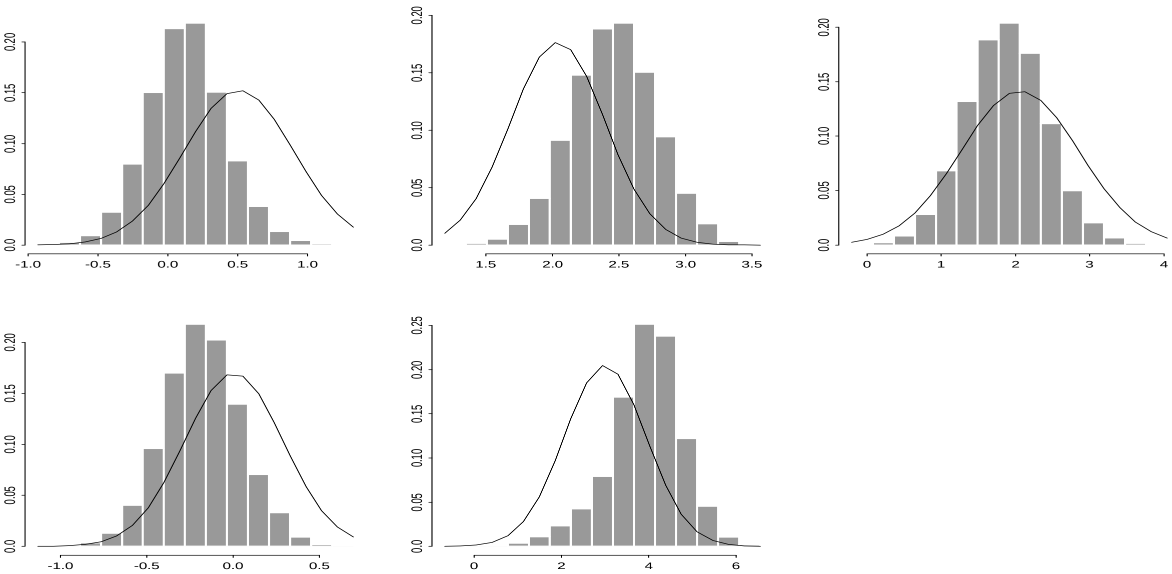

Figure 2. Left to right: marginal posterior distributions for the parameters β_0, β_1, β_2, $\log\sigma$ and $\log\alpha$. The curves indicate marginal prior distributions.

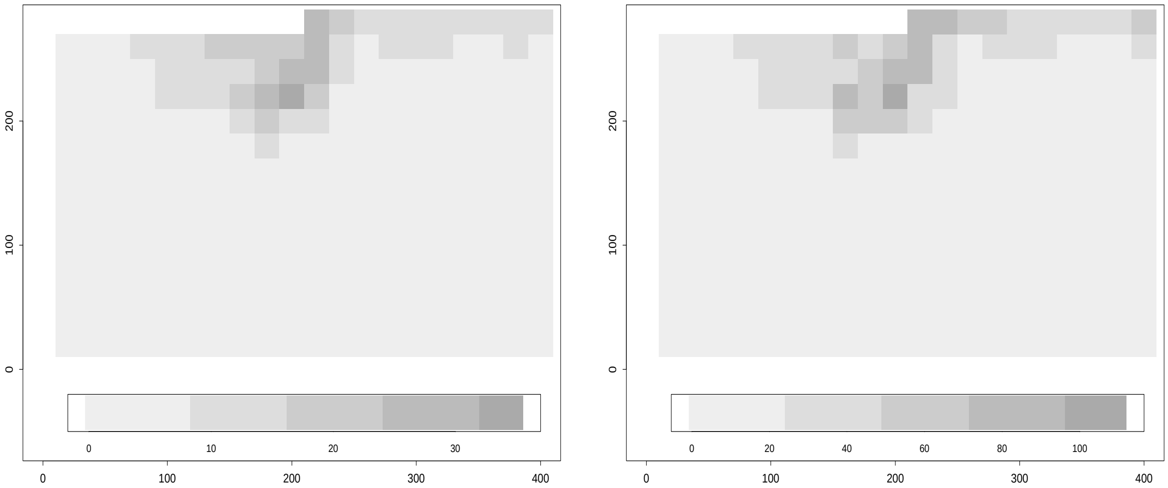

Figure 3. Left: Predicted values Med$[\lambda(x_i) \mid y]$, $i = 1, \ldots, n+q$. Right: Width of prediction intervals for $\lambda(x_i)$, $i = 1, \ldots, n+q$.

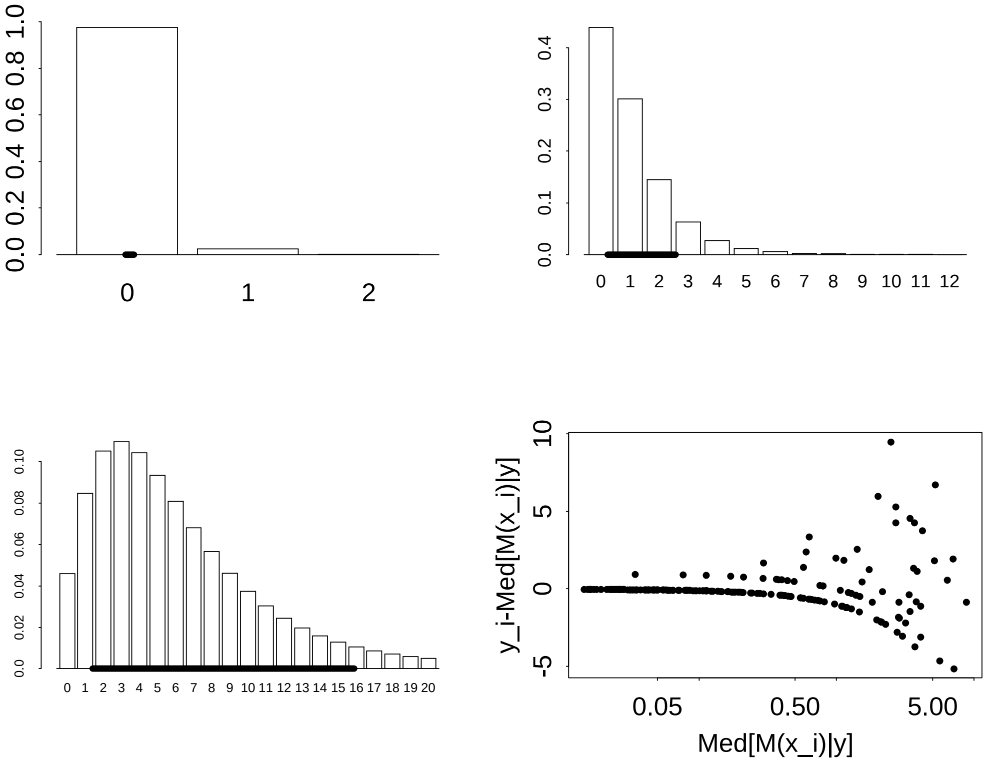

Figure 4. Credibility intervals for $M(x_i)$ (thick line) and the predictive distribution for $Y(x_i)$ at the locations $(320, 20)$, $(160, 160)$ and $(240, 280)$. Lower right plot: Prediction errors $Y(x_i) - \mathrm{Med}[M(x_i) \mid y]$ versus $\mathrm{Med}[M(x_i) \mid y]$, $i = n+1, \ldots, n+q$ (on log scale).