

Bayesian Prediction in Clipped GLG Random Field Using Slice Sampling

Majid Jafari Khaledi*

*Department of Statistics, Tarbiat Modares University,
Tehran, Iran
jafari-m@modares.ac.ir*

Hamidreza Zareifard

*Department of Statistics, University of Jahrom,
Jahrom, Iran
zareifard@jahrom.ac.ir*

Firoozeh Rivaz

*Department of Statistics, Shahid Beheshti University,
Tehran, Iran
f_rivaz@sbu.ac.ir*

Received 29 September 2013

Accepted 20 January 2014

By assuming that an underlying Gaussian-Log Gaussian (GLG) random field clipped to yield binary spatial data, we propose a new model which provides flexibility in capturing the effects of heavy tail in latent variables. For our analysis, we adopt a Bayesian framework and develop a Markov chain Monte Carlo (MCMC) algorithm to carry out the posterior computations. Specifically, we introduce auxiliary variables and employ the slice sampling method to simulate from the full conditional distribution of components which does not define a standard probability distribution. Then, the predictive distribution at unsampled sites is approximated based on acquired samples. Finally, we illustrate our methodology considering simulation and real data sets.

Keywords: Binary spatial data; Bayesian latent model; Heavy tail; Slice sampling.

2000 Mathematics Subject Classification: 62H11, 62F15

1. Introduction

A popular tool for analyzing binary spatial data involves the introduction of latent variables. Based on clipping a Gaussian random field at a fixed threshold, [3] proposed Bayesian prediction of binary

*Corresponding Author

data. Considering a similar framework, [6] also provided composite likelihood approach. [2] demonstrated how the non-identifiable spatial variance parameter can be used to create data augmentation MCMC algorithms in Bayesian probit regression model.

An advantage of the latent variable representation is that the dependency structure can be described in terms of correlation of the latent continuous variables. This simple structure also facilitates generalizations to more complicated data structures. [7] and [10] extended this approach for analyzing ordered categorical spatial data and multivariate mixed discrete and continuous responses, respectively. Despite its mathematical convenience and nice statistical properties, a standard but possible restrictive assumption in modeling of spatial binary data is that the latent random variables are taken to be normal. Based on a process with heavy tailed finite dimensional distributions, [9] and [4] introduced the Gaussian-Log Gaussian (GLG) process to reduce unrealistic normality assumption in spatial models.

The objective of this paper is to propose a modelling strategy for the analysis of spatially correlated binary responses by incorporating a GLG random field as probability model for the latent variables. A Bayesian inference approach for prediction of response variable at new locations is developed using MCMC methods. More specifically, adopting the Gibbs sampling algorithm, we must draw samples from the joint posterior distribution of model parameters and latent variables. With regard to problems remain with convergence and mixing properties of Metropolis-Hastings algorithm, we will introduce auxiliary variables and employ the slice sampling method ([8]) to simulate from full conditional distribution of latent random variables. Actually, this is a technique of generating from arbitrary variables by introducing an auxiliary variable and sampling from two or more uniform distributions. Finally, we apply this method in order to analyze a data set related to soil pollution samples in an area located in North of Iran.

The rest of the article is organized as follows. In Section 2 we state the model and discuss its main features. We describe our model fitting and posterior predictive inference in Section 3. The methodology is illustrated with two examples in Section 4. Finally, we conclude in Section 5.

2. The Model

Let $\mathbf{Z} = (Z(s_1), \dots, Z(s_n))$ be a single realization from a binary random field $Z(\cdot) = \{Z(s); s \in D \subseteq \mathbb{R}^d\}$, $d \geq 1$, at n different locations s_1, \dots, s_n . We assume that the binary random field, $Z(s)$, is created by clipping a GLG random field, $W(s)$, at the threshold ε , defined over the region of interest, D , meaning

$$Z(s) = I_{\{W(s) > \varepsilon\}}, \quad s \in D \quad (2.1)$$

where I denotes the indicator function and

$$W(s) = \mathbf{f}'(s)\boldsymbol{\beta} + \frac{\varepsilon(s)}{\sqrt{\lambda(s)}}, \quad (2.2)$$

with $\varepsilon(s)$ and $\eta(s) = \log \lambda(s)$ are independent Gaussian random fields, $\boldsymbol{\beta} \in \mathbb{R}^k$ is unknown coefficient vector and $\mathbf{f}'(s) = (f_1(s), \dots, f_k(s))$ denotes a vector of k known functions of the spatial coordinates. In this setting, if $W = (W(s_1), \dots, W(s_n))$ and $\eta = (\eta(s_1), \dots, \eta(s_n)) =$

$(\ln \lambda(s_1), \dots, \ln \lambda(s_n))$, then

$$\begin{aligned} W|\eta &\sim N_n(X\beta, \sigma^2 \Lambda^{-\frac{1}{2}} C_{\theta_1} \Lambda^{-\frac{1}{2}}), \\ \eta &\sim N_n(-\frac{\nu}{2} \mathbf{1}, \nu C_{\theta_2}), \end{aligned} \quad (2.3)$$

where $X = (\mathbf{f}(s_1), \dots, \mathbf{f}(s_n))'_{n \times k}$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ with $\lambda_i = \lambda(s_i)$, the scale parameter σ^2 defined in \mathfrak{R}^+ , and C_{θ_1} is the $n \times n$ correlation matrix with $C_{\theta_1}(|s_i - s_j|)$ as its (i, j) th element; $C_{\theta_1}(d)$ is also a valid correlation function of distance d , parameterized by a vector θ_1 . Similarly, C_{θ_2} is defined. Finally, $\mathbf{1}$ is a vector of 1's and $\nu \in \mathfrak{R}^+$ is a scalar parameter. We can easily see that the log-Gaussian assumption for the random field $\lambda(\cdot)$ implies a lognormal distribution for $\lambda_i = \lambda(s_i)$ with $E(\lambda_i) = 1$ and $\text{Var}(\lambda_i) = e^\nu - 1$.

Based on (2.3), the likelihood function of the observed data is given by

$$\begin{aligned} f(z|\beta, \theta_1, \theta_2, \sigma^2, \nu) &= \int_{A(z_1)} \cdots \int_{A(z_n)} \int_{\mathfrak{R}_+^n} f_N^n(w|X\beta, \sigma^2 \Lambda^{-\frac{1}{2}} C_{\theta_1} \Lambda^{-\frac{1}{2}}) dP_\eta dw, \\ A(z_i) &= \begin{cases} (-\infty, \varepsilon], & \text{if } z_i=0, \\ (\varepsilon, \infty), & \text{if } z_i=1, \end{cases} \end{aligned} \quad (2.4)$$

where $f_N^n(w|\cdot)$ denotes the probability density function of a n -variate normal distribution. With regard to the likelihood function, the parameters are not identifiable. To avoid this problem, we fixed $\sigma^2 = 1$ and $\varepsilon = 0$ as proposed by [3] and [7].

The third simplification restricts the spatial correlation function to involve only one parameter, thus restricting θ_1 and θ_2 to a scalar. In fact, [3] points out that after clipping the continuous random field, the binary data contains no information about the smoothness parameter, even if we could observe the complete binary realization rather than a finite number of locations. Thus, we use an isotropic exponential correlation function given as

$$C_\gamma(d) = \exp(-\frac{d}{\phi}) = \gamma^d, \quad \phi > 0, \gamma = \exp(-\phi) \in (0, 1),$$

where γ is the range parameter and controls how fast the correlation decays with distance. Under this assumption our unknown model parameters are $\beta, \gamma_1, \gamma_2$ and ν , being now (likelihood) identifiable.

3. Bayesian Analysis

In order to complete the Bayesian model specification, we assume that the parameters are, priori, mutually independent and use proper priors for all of them. Hence, the prior have a density function of the form

$$\pi(\beta, \gamma_1, \gamma_2, \nu) = \pi(\beta)\pi(\gamma_1)\pi(\gamma_2)\pi(\nu),$$

where

$$\begin{aligned} \pi(\beta) &= f_N^k(\beta|0, c_1 I_k), \\ \pi(\nu) &= f_{IG}(\nu|c_2, c_3), \\ \pi(\gamma_1) &= \pi(\gamma_2) = U(0, 1). \end{aligned} \quad (3.1)$$

$f_{IG}(\cdot|s, r)$ denotes the probability density functions of inverse gamma distributions with shape s and rate r . By adopting very large values for scale hyperparameter c_1 and very small values for hyperparameters c_2 and c_3 , vague prior distributions are obtained for β and ν . To implement MCMC simulation, we explore the complete posterior conditional distributions as:

$$\begin{aligned}
\pi(\beta|z, w, \eta, \gamma_1) &= N_k(\Sigma_\beta X' \Lambda^{\frac{1}{2}} C_{\gamma_1}^{-1} \Lambda^{\frac{1}{2}} w, \Sigma_\beta), \\
\pi(W|z, \eta, \beta, \gamma_1) &= N_n(X\beta, \Lambda^{-\frac{1}{2}} C_{\gamma_1} \Lambda^{-\frac{1}{2}}) \prod_{i=1}^n I_{\{w_i \in A(z_i)\}}, \\
\pi(\eta|z, w, \beta, \gamma_1, \gamma_2, \nu) &\propto \prod_{i=1}^n \lambda_i^{\frac{1}{2}} \exp\left(-\frac{1}{2}(w - X\beta)' \Lambda^{\frac{1}{2}} C_{\gamma_1}^{-1} \Lambda^{\frac{1}{2}} (w - X\beta)\right), \\
&\quad \times \exp\left(-\frac{1}{2\nu}(\eta + \frac{\nu}{2}\mathbf{1})' C_{\gamma_2}^{-1} (\eta + \frac{\nu}{2}\mathbf{1})\right), \\
\pi(\nu|z, \eta, \gamma_2) &\propto f(\eta|\nu, \gamma_2) \pi(\nu), \\
\pi(\gamma_2|z, \eta, \nu) &\propto f(\eta|\nu, \gamma_2) \pi(\gamma_2), \\
\pi(\gamma_1|z, w, \eta, \beta) &\propto f(w|\eta, \beta, \gamma_1) \pi(\gamma_1),
\end{aligned} \tag{3.2}$$

where $\Sigma_\beta = (\frac{1}{c_1} I_k + X' \Lambda^{\frac{1}{2}} C_{\gamma_1}^{-1} \Lambda^{\frac{1}{2}} X)^{-1}$. The full conditional posterior of β is known and easy to sample from. Although the full conditional of W defines a standard probability distribution, sampling of this distribution is simply impracticable. In fact, two methods can be used for sampling of this full conditional. If n is small, we propose the rejection sampling. In this method, we generate proposals of multivariate normal distribution which are accepted, once they are inside the support region otherwise get rejected. However, the rejection sampling may be inefficient when n is big. In this case, the Gibbs sampler is preferable in which each coordinate, W_i say, is generated conditional on all other coordinates of W .

In sum, the full conditional of η , ν , γ_1 and γ_2 do not define a standard probability distribution. To draw samples from the full conditional posterior η , [9] partitioned the elements of η in blocks, each of which corresponds to a cluster of observations that are relatively close together. Indeed, they wanted to confine most of the dependence between the η_i 's to the same cluster. For each cluster, they also used a Metropolis-Hastings step. But, their method has some drawbacks. First, the Metropolis-Hastings algorithm is hard to become automate since it involves tuning tailored to each application. Second, increasing the number of clusters, increases the convergence time of the Markov chain for which sampling of the joint posterior distribution has been designed. Third, the inferences can be affected by the considered clusters.

Recently, auxiliary variable methods based on slice sampler is found to provide an attractive strategy are receiving utmost attention by those who used MCMC algorithms to simulate from complex nonnormalized multivariate densities ([8]). In this paper, for sampling from the full conditional of η , we implement slice sampling algorithm based on three auxiliary variables ([1]). For this purpose, if $U_1|w, \eta, \beta, \gamma_1$, $U_2|\eta, \nu, \gamma_2$ and $U_3|\eta$ have the uniform distribution on the intervals $[0, g_1(w|\eta, \beta, \gamma_1)]$, $[0, g_2(\eta|\nu, \gamma_2)]$ and $[0, \prod_{i=1}^n \lambda_i^{\frac{1}{2}}]$, respectively, where $g_1(w|\eta, \beta, \gamma_1) = \exp(-\frac{1}{2}(w - X\beta)' \Lambda^{\frac{1}{2}} C_{\gamma_1}^{-1} \Lambda^{\frac{1}{2}} (w - X\beta))$ and $g_2(\eta|\nu, \gamma_2) = \exp(-\frac{1}{2\nu}(\eta + \frac{\nu}{2}\mathbf{1})' C_{\gamma_2}^{-1} (\eta + \frac{\nu}{2}\mathbf{1}))$, then

$$\pi(U_1, U_2, U_3, \eta|z, w, \beta, \gamma_1, \gamma_2, \nu) \propto I_{\{U_1 < g_1(w|\eta, \beta, \gamma_1)\}} I_{\{U_2 < g_2(\eta|\nu, \gamma_2)\}} I_{\{U_3 < \prod_{i=1}^n \lambda_i^{\frac{1}{2}}\}}.$$

Thus,

$$\pi(\eta|z, w, \beta, \gamma_1, \gamma_2, v, U_1, U_2, U_3) \propto I_{\{U_1 < g_1(w|\eta, \beta, \gamma_1)\}} I_{\{U_2 < g_2(\eta|v, \gamma_2)\}} I_{\{U_3 < \prod_{i=1}^n \lambda_i^{\frac{1}{2}}\}}. \quad (3.3)$$

Also for sampling from the full conditionals of v , γ_1 and γ_2 , we implement slice sampling algorithm. For this purpose, if $U_4|\eta, v, \gamma_2$, $U_5|\eta, v, \gamma_2$ and $U_6|w, \eta, \beta, \gamma_1$ have the uniform distribution on the intervals $[0, f(\eta|v, \gamma_2)]$, $[0, f(\eta|v, \gamma_2)]$ and $[0, f(w|\eta, \beta, \gamma_1)]$ respectively, then

$$\begin{aligned} \pi(v|z, \eta, \gamma_2, U_4) &\propto I_{\{U_4 < f(\eta|v, \gamma_2)\}} \pi(v), \\ \pi(\gamma_2|z, \eta, v, U_5) &\propto I_{\{U_5 < f(\eta|v, \gamma_2)\}} \pi(\gamma_2), \\ \pi(\gamma_1|z, w, \eta, \beta, U_6) &\propto I_{\{U_6 < f(w|\eta, \beta, \gamma_1)\}} \pi(\gamma_1). \end{aligned}$$

Based on these assumptions, we can summarize the main steps in iteration $(t + 1)$ of the slice sampling algorithm as:

- (1) Draw $\beta^{(t+1)}$ from $\pi(\beta|z, w^{(t)}, \eta^{(t)}, \gamma_1^{(t)})$.
- (2) Draw $w^{(t+1)}$ from $\pi(w|z, \beta^{(t+1)}, \eta^{(t)}, \gamma_1^{(t)})$.
- (3) Draw $e_i^{(t+1)}, i = 1, 2, 3$, of $\exp(1)$, and let $a_t = \log g_1(w^{(t+1)}|\eta^{(t)}, \beta^{(t+1)}, \gamma_1^{(t)}) - e_1^{(t+1)}$, $b_t = \log g_2(\eta^{(t)}|v^{(t)}, \gamma_2^{(t)}) - e_2^{(t+1)}$, $c_t = \frac{1}{2} \sum_{i=1}^n \log \lambda_i^{(t)} - e_3^{(t+1)}$
- (4) Draw $\eta^{(t+1)}$ from a uniform distribution on $\{a_t < \log g_1(w^{(t+1)}|\eta, \beta^{(t+1)}, \gamma_1^{(t)})\} \cap \{b_t < \log g_2(\eta|v^{(t)}, \gamma_2^{(t)})\} \cap \{c_t < \frac{1}{2} \sum_{i=1}^n \log \lambda_i\}$.
- (5) Draw $e_4^{(t+1)}$ of $\exp(1)$, and let $d_t = \log f(\eta^{(t+1)}|v^{(t)}, \gamma_2^{(t)}) - e_4^{(t+1)}$.
- (6) Draw $v^{(t+1)}$ from truncated gamma distribution on $\{v, d_t < \log f(\eta^{(t+1)}|v, \gamma_2^{(t)})\}$.
- (7) Draw $e_5^{(t+1)}$ of $\exp(1)$, and let $e_t = \log f(\eta^{(t+1)}|v^{(t+1)}, \gamma_2^{(t)}) - e_5^{(t+1)}$.
- (8) Draw $\gamma_2^{(t+1)}$ from a uniform distribution on $\{\gamma_2, e_t < \log f(\eta^{(t+1)}|v^{(t+1)}, \gamma_2)\} \cap (0, 1)$.
- (9) Draw $e_6^{(t+1)}$ of $\exp(1)$, and let $f_t = \log f(w^{(t+1)}|\eta^{(t+1)}, \beta^{(t+1)}, \gamma_1^{(t)}) - e_6^{(t+1)}$.
- (10) Draw $\gamma_1^{(t+1)}$ from a uniform distribution on $\{\gamma_1, f_t < \log f(w^{(t+1)}|\eta^{(t+1)}, \beta^{(t+1)}, \gamma_1)\} \cap (0, 1)$.

Iterate above steps till we get the appropriate number of MCMC samples. We now offer a method to generate sample of the uniform distribution on

$$\{a_t < \log g_1(w^{(t+1)}|\eta, \beta^{(t+1)}, \gamma_1^{(t)})\} \cap \{b_t < \log g_2(\eta|v^{(t)}, \gamma_2^{(t)})\} \cap \{c_t < \frac{1}{2} \sum_{i=1}^n \log \lambda_i\}.$$

At first, we have

$$\begin{aligned} a_t < \log g_1(w^{(t+1)}|\eta, \beta^{(t+1)}, \gamma_1^{(t)}) &\Leftrightarrow \varepsilon^{(t+1)'} \Lambda^{\frac{1}{2}} C_{\gamma_1^{(t)}}^{-1} \Lambda^{\frac{1}{2}} \varepsilon^{(t+1)} < a_t^* \\ b_t < \log g_2(\eta|v^{(t)}, \gamma_2^{(t)}) &\Leftrightarrow (\eta + \frac{v^{(t)}}{2} \mathbf{1})' C_{\gamma_2^{(t)}}^{-1} (\eta + \frac{v^{(t)}}{2} \mathbf{1}) < b_t^* \end{aligned} \quad (3.4)$$

where $\varepsilon = w - X\beta$, $a_t^* = \varepsilon^{(t+1)'} \Lambda^{(t)\frac{1}{2}} C_{\gamma_1^{(t)}}^{-1} \Lambda^{(t)\frac{1}{2}} \varepsilon^{(t+1)} + 2e_1^{(t+1)}$ and $b_t^* = (\eta^{(t)} + \frac{v^{(t)}}{2} \mathbf{1})' C_{\gamma_2^{(t)}}^{-1} (\eta^{(t)} + \frac{v^{(t)}}{2} \mathbf{1}) + 2v^{(t)} e_2^{(t+1)}$. Now, we define

$$I_i(\eta_{-i}^{(t)}) = \{\eta^* \in \mathfrak{R}; (\eta + \frac{v^{(t)}}{2} \mathbf{1})' C_{\gamma_2^{(t)}}^{-1} (\eta + \frac{v^{(t)}}{2} \mathbf{1}) < b_t^* \text{ if } \eta = (\eta_1^{(t)}, \dots, \eta_{i-1}^{(t)}, \eta^*, \eta_{i+1}^{(t)}, \dots, \eta_n^{(t)})\}.$$

$$J_i(\eta_{-i}^{(t)}) = \{\eta^* \in \mathfrak{R}^+; \varepsilon^{(t+1)'} \Lambda^{\frac{1}{2}} C_{\gamma_1^{(t)}}^{-1} \Lambda^{\frac{1}{2}} \varepsilon^{(t+1)} < a_t^* \text{ if}$$

$$\Lambda = \text{diag}(\exp(\eta_1^{(t)}), \dots, \exp(\eta_{i-1}^{(t)}), \exp(\eta^*), \exp(\eta_{i+1}^{(t)}), \dots, \exp(\eta_n^{(t)}))\}.$$

Then $I_i(\eta_{-i}^{(t)})$ and $J_i(\eta_{-i}^{(t)})$ contain all possible values of the i th coordinates in order for η to remain in the n dimensional oval while the other $n - 1$ coordinates are fixed. Clearly, $I_i(\eta_{-i}^{(t)})$ and $J_i(\eta_{-i}^{(t)})$ are a non-empty interval because $\eta_i^{(t)} \in I_i(\eta_{-i}^{(t)}) \cap J_i(\eta_{-i}^{(t)})$. Let $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ and (l_1, \dots, l_n) are eigenvectors and eigenvalues of C_{γ_2} , respectively, where $\mathbf{e}_j = (e_{1j}, \dots, e_{nj})'$. It follows, after some algebra, that if

$$a_1 = \sum_{j=1}^n \frac{1}{l_j} e_{ij}^2, \quad a_2 = 2 \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n \frac{e_{ij}}{l_j} e_{kj} v_k^{(t)}, \quad a_3 = v_i^{(t)} a_2 + (v_i^{(t)})^2 a_1 + 2v^{(t)} e_2^{(t+1)},$$

$$f_1 = \frac{-a_2 - \sqrt{a_2^2 + 4a_1 a_3}}{2a_1}, \quad f_2 = \frac{-a_2 + \sqrt{a_2^2 + 4a_1 a_3}}{2a_1},$$

where $v = \eta + \frac{v}{2}$ then

$$I_i(\eta_{-i}^{(t)}) = \{\eta^* \in \mathfrak{R}; \quad f_1 < \eta^* + \frac{v^{(t)}}{2} < f_2\}. \quad (3.5)$$

Similarly, if $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ and (l_1, \dots, l_n) are eigenvectors and eigenvalues of C_{γ_1} , respectively, and

$$b_1 = \sum_{j=1}^n \frac{1}{l_j} e_{ij}^2, \quad b_2 = 2 \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq i}}^n \frac{e_{ij}}{l_j} e_{kj} \varepsilon_k^{(t+1)} \sqrt{\lambda_k^{(t)}}, \quad b_3 = \varepsilon_i^{(t+1)} \sqrt{\lambda_i^{(t)}} b_2 + \varepsilon_i^{(t+1)^2} \lambda_i^{(t)} b_1 + 2e_1^{(t+1)},$$

then we have

$$J_i(\eta_{-i}^{(t)}) = \begin{cases} \{\eta^* \in \mathfrak{R}; \quad 2\log(\frac{\max\{g_1, 0\}}{\varepsilon_i^{(t+1)}}) < \eta^* < 2\log(\frac{g_2}{\varepsilon_i^{(t+1)}})\}, & \text{if } \varepsilon_i^{(t+1)} > 0, \\ \{\eta^* \in \mathfrak{R}; \quad 2\log(\frac{\min\{g_2, 0\}}{\varepsilon_i^{(t+1)}}) < \eta^* < 2\log(\frac{g_1}{\varepsilon_i^{(t+1)}})\}, & \text{if } \varepsilon_i^{(t+1)} < 0, \end{cases} \quad (3.6)$$

where g_1 and g_2 are defined similarly with f_1 and f_2 but based on b_1 , b_2 and b_3 . Thus, under (3.5) and (3.6), i th element $\eta^{(t+1)}$ can be generated uniformly on the interval $I_i(\eta_{-i}^{(t)}) \cap J_i(\eta_{-i}^{(t)}) \cap (\eta_i^{(t)} - 2e_3^{(t+1)}, \infty)$.

Finally, it must be noted that

$$v \in \{v; \quad d_t < \log f(\eta^{(t+1)} | v, \gamma_2^{(t)})\} \Leftrightarrow 0 < v < 2 \frac{-n + \sqrt{n^2 + \mathbf{1}' C_{\gamma_2^{(t)}}^{-1} \mathbf{1} \eta^{(t+1)'} C_{\gamma_2^{(t)}}^{-1} \eta^{(t+1)}}}{\mathbf{1}' C_{\gamma_2^{(t)}}^{-1} \mathbf{1}}.$$

Thus, the full conditional of v is a truncated gamma distribution on mentioned region in the above.

3.1. Prediction

In many applications, prediction of response values at new locations is an important goal. In this subsection, to predict the vector $Z_0 = (Z_{0_1}, \dots, Z_{0_p})'$ at unsampled locations s_{0_1}, \dots, s_{0_p} , the Bayesian posterior predictive distribution, $f(z_0|z)$, are made. In a similar way with [3], we restrict attention to additive loss functions

$$L(Z_0, \tilde{Z}_0) = \frac{1}{p} \sum_{j=1}^p I_{\{Z_{0_j} \neq \tilde{Z}_{0_j}\}},$$

where $\tilde{Z}_0 = (\tilde{Z}_{0_1}, \dots, \tilde{Z}_{0_p})$, is an arbitrary predictor of Z_0 . The optimal Bayes predictor of Z_{0_i} , is $I_{\{\pi(Z_{0_i}=1|z)>0.5\}} = I_{\{\pi(w_{0_i}>0|z)>0.5\}}$. These conditional probabilities can be computed using a sample from

$$\pi(w_0|z) = \int_0^1 \int_0^1 \int_{\mathbb{R}_+} \int_{\mathbb{R}^k} \int_{\mathbb{R}_+^p} \int_{\mathbb{R}_+^n} f(w_0|w, \eta, \eta_0, \beta, \gamma_1) \pi(\eta_0|\eta, v, \gamma_2) \pi(w, \eta, \beta, v, \gamma_1, \gamma_2|z) dw d\eta d\eta_0 d\beta dv d\gamma_1 d\gamma_2,$$

where

$$\pi(\eta_0|\eta, v, \gamma_2) = N_p(-\frac{v}{2}\mathbf{1}_p + C_{\gamma_2}^{po} C_{\gamma_2}^{-1}(\eta + \frac{v}{2}\mathbf{1}_n), v(C_{\gamma_2}^{pp} - C_{\gamma_2}^{po} C_{\gamma_2}^{-1} C_{\gamma_2}^{po'})),$$

$$f(w_0|w, \eta, \eta_0, \beta, \gamma_1) = N_p(X_0\beta + A(w - X\beta), \Lambda_p^{-\frac{1}{2}} C_{\gamma_1}^{pp} \Lambda_p^{-\frac{1}{2}} - A\Lambda^{-\frac{1}{2}} C_{\gamma_1}^{po'} \Lambda_p^{-\frac{1}{2}}),$$

$A = \Lambda_p^{-\frac{1}{2}} C_{\gamma_1}^{po} C_{\gamma_1}^{-1} \Lambda_p^{\frac{1}{2}}$, $X_p = (\mathbf{f}(s_{0_1}), \dots, \mathbf{f}(s_{0_p}))'$, $\Lambda_p = \text{diag}(\lambda_{0_1}, \dots, \lambda_{0_p})$, $C_{\gamma_1}^{pp} = [C_{\gamma_1}(|s_{0_i} - s_{0_j}|)]_{p \times p}$ and $C_{\gamma_1}^{po} = [C_{\gamma_1}(|s_{0_i} - s_j|)]_{p \times n}$. Thus, for each posterior draw $(w, \eta, \beta, v, \gamma_1, \gamma_2)$, we generate a drawing from $\pi(\eta_0|\eta, v, \gamma_2)$ and finally using sampling from $f(w_0|w, \eta, \eta_0, \beta, \gamma_1)$, we can obtain a realization from $\pi(w_0|z)$. Repeating aforementioned steps as many times as required, thereby we generate a sample from $\pi(w_0|z)$ as $\{w_0^{(j)}; j = 1, \dots, J\}$. Then, the Bayesian estimates of the conditional probabilities are given by

$$\hat{p}(z_{0_i} = 1|z) = \frac{1}{J} \sum_{j=1}^J I_{\{w_0^{(j)} > 0\}} \quad i = 1, \dots, p. \quad (3.7)$$

4. Numerical Examples

In this section, we apply our model to simulated and applied data sets. We also compare the results with those obtained from the Gaussian latent variable model.

Example 1: Here, we carry out simulation to compare the prediction error that produced from using the proposed model (CGLG) to that of the clipped Gaussian model (CG). We generated 50 data sets on a regular 15×15 lattice with five units between nearest neighbors, resulted in 225 observations per data set. The exponential correlation function was used to produce the spatial dependence structure of data. The data sets were simulated with the following presumed parameters: $\beta = 10$, $\gamma_1 = \gamma_2 = 0.5$, and $v \in \{1, 5, 10\}$ corresponding to that among the observations there exists a region with larger observational variance relative to the rest with small, medium and large probabilities. To validate the predictive ability of two models, half of the simulated data set, 112

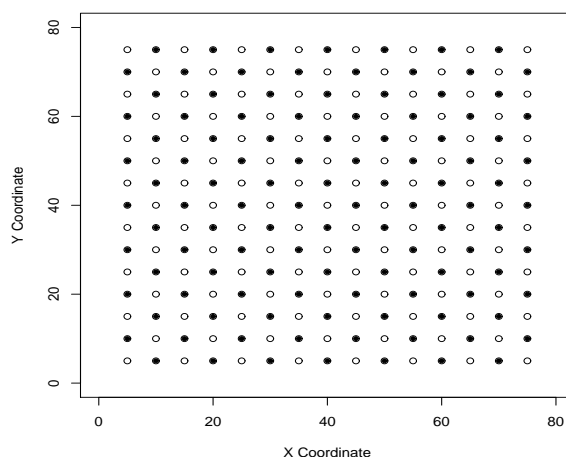


Fig. 1. Locations for simulation study, \circ represent locations used in prediction and \bullet represent locations used for comparison of predictions.

Table 1. Summary of MPS for the 50 simulated data sets under two considered models.

Parameter ν	CGLG			CG		
	Min	Mean	Max	Min	Mean	Max
1	0.929	0.948	0.982	0.938	0.956	0.973
5	0.938	0.951	0.964	0.875	0.893	0.911
10	0.929	0.942	0.973	0.804	0.819	0.848

observations, is withheld and their values are predicted based on the remaining 113 sampled locations. Figure 1 displays the observation locations and the locations set aside for comparisons of predictions.

The Bayesian analysis was specified with proper diffuse priors which centered at the truth. Here, the MCMC chain was run for 200,000 iterations (with a burn-in period of 50,000).

Now we compare the predictive performance of two considered models. For this, we predict the response variable in hold-out data sets and obtain the mean of prediction score (MPS),

$$MPS = \frac{1}{112} \sum_{j=1}^{112} I_{\{Z(s_j) = \hat{Z}(s_j)\}}.$$

Table 1 compares the acquired results under two models. We can see that the CGLG model evidently outperforms the CG one for medium and large values of ν .

Example 2: This example consists of an illustrative application of the proposed methodology. The data set includes binary indicators for a heavy metal (lead or Pb) found in the soil which indicates whether the level of this heavy metal is over the legal pollution threshold or not in 60 locations of a region in north of Iran (see Figure 2). Since this example does not have any explanatory variables, we assume constant mean. To compare the predictive ability of model, we considered cross-validation predictive distributions known as conditional predictive ordinates (CPO). The CPOs which are often used as predictive model checking tools, measure the influence of individual observations ([5]). In cross-validation, a validation of observation Y_i is held to the side and

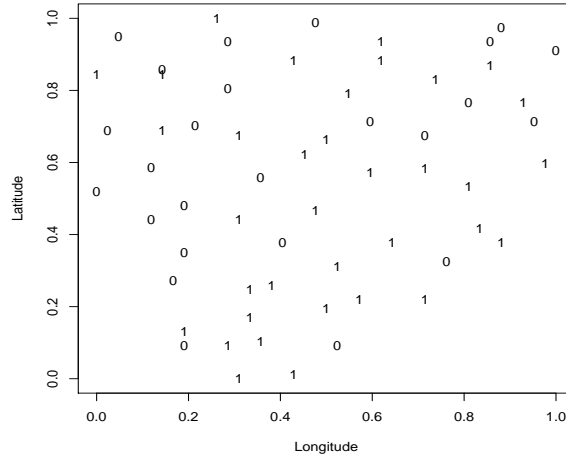


Fig. 2. Sampled locations and values of heavy metal in the soil data.

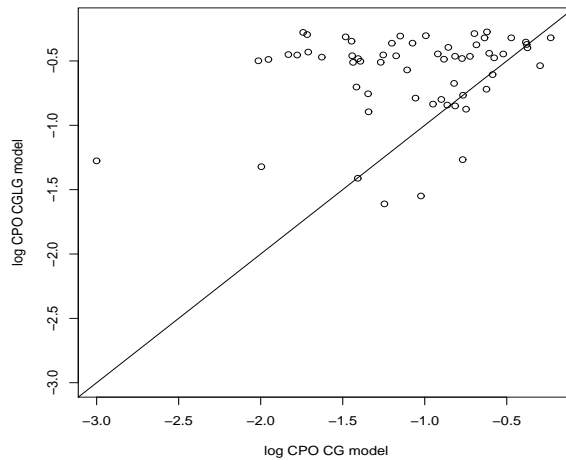


Fig. 3. Result of fitting two models to the heavy metal soil example.

the remaining data denoted Y_{-i} is used as training data to fit the model. By computing the cross-validated conditional predictive distribution for each data, these values can then be summarized and used to compare models. A summary statistic of the CPOs is $B = \sum_{i=1}^{117} \log(CPO_i)$. The larger is the value of B , the better is the fit of the model. Results are obtained based on every fifth draws from an MCMC chain of length 200,000 with a burn-in of 50,000. This proved more than enough for convergence, and much shorter runs led to virtually identical results. Our experiments not reported here show the robustness of the posterior results to the prior changes. We computed the B values for two models CGLG and CG as -32.236 and -65.329, respectively, which indicates our model has a better predictive performance. Furthermore, in Figure 3, the plot shows the $\log CPO$ values for the CGLG versus CG models. In sum, the CPOs are larger for the proposed model. Inference results for

Table 2. Posterior means and 95% credible intervals for model parameters.

Parameter	Posterior mean	95% credible interval
β	0.38	(0.21,0.57)
ν	4.23	(3.72,5.98)
γ_1	0.23	(0.11,0.30)
γ_2	0.12	(0.06,0.17)

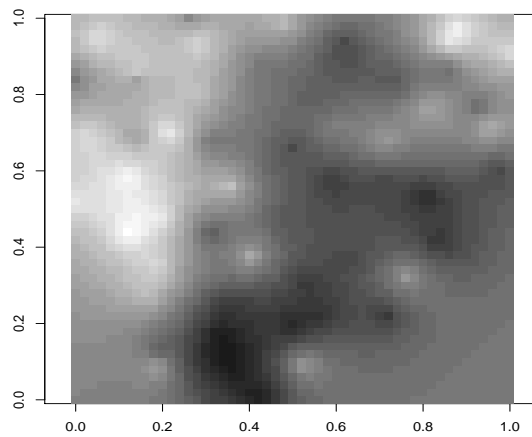


Fig. 4. Prediction map of the latent spatial process based on the posterior median. Regions with dark color have high pollution values.

parameter of distribution of underlying latent random field are summarized in Table 2. Finally, the prediction map corresponding to the predictive median under the CGLG case, is shown in Figure 4. From this figure, the regions with high pollution could be observed.

5. Conclusion

We provide a fully Bayesian approach to analyze the clipped GLG model. We have developed a computationally feasible algorithm based on the slice sampling to simulate from the posterior distribution. Results from numerical examples demonstrate that our proposed model improve the predictive performance in compare to spatial probit model.

The special case considered in this paper includes univariate binary data. Although multivariate binary measurements are very common in many research areas. The extension of our approach to these cases is an interesting area to investigate in further research.

References

- [1] D.K. Agarwal and A.E. Gelfand, Slice sampling for simulation based fitting of spatial data models, *Stat. Comput.* **15** (2005) 61–69.
- [2] C. Berrett and C.A. Calder, Data augmentation strategies for the Bayesian spatial probit regression model, *Comput. Stat. Data An.* **56** (2012) 478–490.

- [3] V. De Oliveira, Bayesian prediction of clipped Gaussian random fields, *Comput. Stat. Data An.* **34** (2000) 299–314.
- [4] T. Fonseca, and M. Steel, Non-Gaussian spatiotemporal modelling through scale mixing, *Biometrika* **98** (2011) 761–774.
- [5] A.E. Gelfand, D.K. Dey and H. Chang, Model determination using predictive distributions with implementation via sampling-based methods, *Bayesian. Stat.* **4** (1992) 147–167.
- [6] P.H. Heagerty and S.R. Lele, A composite likelihood approach to binary spatial data, *J. Amer. Statist. Assoc.* **93** (1999) 1099–1111.
- [7] M.D. Higgs and J.A. Hoeting, A clipped latent variable model for spatially correlated ordered categorical data, *Comput. Stat. Data An.* **54** (2010) 1999–2011.
- [8] R. Neal, Slice sampling, *Ann. Stat.* **31** (2003) 705–767.
- [9] M. Palacios and M. Steel, Non-Gaussian Bayesian geostatistical modeling, *J. Amer. Statist. Assoc.* **101** (2006) 604–618.
- [10] E.M. Schliep and J.A. Hoeting, Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data, *J. Agric. Biol. Environ. Stat.* **93** (2013) 1099–1111.