Biometrics 00, 1–23 DOI: 000
000-000 2014

Efficient estimation and prediction for the Bayesian binary spatial model with flexible link functions

Vivekananda Roy[1] Evangelos Evangelou[2] Zhengyuan Zhu[1]

[1] Department of Statistics, Iowa State University, USA

[2] Department of Mathematical Sciences, University of Bath, UK

Summary: Spatial generalized linear mixed models (SGLMMs) are popular models for spatial data with a non-Gaussian response. Binomial SGLMMs with logit or probit link functions are often used to model spatially dependent binomial random variables. It is known that for independent Binomial data, the robit regression model provides a more robust (against extreme observations) alternative to the more popular logistic and probit models. In this article, we introduce a Bayesian spatial robit model for spatially dependent binomial data. Since constructing a meaningful prior on the link function parameter as well as the spatial correlation parameters in SGLMMs is difficult, we propose an empirical Bayes (EB) approach for the estimation of these parameters as well as for the prediction of the random effects. The EB methodology is implemented by efficient importance sampling methods based on Markov chain Monte Carlo (MCMC) algorithms. Our simulation study shows that the robit model is robust against model misspecification, and our EB method results in estimates with less bias than full Bayesian (FB) analysis. The methodology is applied to a Celastrus Orbiculatus data, and a Rhizoctonia root data. For the former, the robit model is shown to do better for predicting the spatial distribution of an invasive species which is known to contain outlying observations. For the latter, our approach is doing as well as the classical models for predicting the disease severity for a root disease, as the probit link is shown to be appropriate.

Though this paper is written for Binomial SGLMMs for brevity, the EB methodology is more general and can be applied to other types of SGLMMs. In the accompanying R package geoBayes, implementations for other SGLMMs such as Poisson and Gamma SGLMMs are provided.

Key words: Generalized linear mixed models; geostatistics; importance sampling; Markov chain Monte Carlo; robust model; robit model; spatial statistics; spatial prediction.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Spatial nonlinear or non-Gaussian data observed in a continuous region can be analyzed by spatial generalized linear mixed models (SGLMM) introduced by Diggle et al. (1998) and further studied in Diggle et al. (2003); Christensen and Waagepetersen (2002); Zhang (2002); Christensen (2004), and Evangelou et al. (2011). A SGLMM is a generalized linear mixed model (GLMM) where the random effects form a spatial process. Conditional on the random effects, the observations are assumed to have a common sampling distribution with varying site-specific parameters, such as Poisson for count data or binomial for binary data. Spatial count/binary data arise in agriculture, ecology, engineering, epidemiology, and many other fields of study. Typical examples of spatial binary data include incidence counts of diseases and infections (see for example the incidence of campylobacter infections in north Lancashire and south Cumbria considered in the seminal paper of Diggle et al., 1998) and the weed data of Christensen and Waagepetersen (2002) is an example of spatial count data.

Classical analysis of spatial data using SGLMMs often assumes a fixed known link function. On the other hand, there is evidence that fixed link functions may not always be appropriate. Christensen (2004) provides evidence that the log-link, as used by Diggle et al. (1998), may not be a good choice for analyzing a data set of radionuclide concentrations on Rongelap Island and uses a parametric (Box-Cox) family of link functions. Wang et al. (2010) discuss non-symmetric link functions for binary data provided by the inverse CDF of the generalized extreme value (GEV) distribution with varying shape parameter.

In the case for binary data, the majority of the literature uses only two link functions, namely the logit and probit links which are obtained by assuming that the inverse link function is the CDF of the logistic and standard normal distributions respectively. In the case of independent binary data, it is well known that the popular logistic and probit models are not robust against outlying observations (Pregibon, 1982). A robust alternative to these

models is the robit model (Liu, 2004). The robit model is defined by replacing the link function in these models with the inverse CDF of a t-distribution (Albert and Chib, 1993). Unlike the logistic and probit models, the robit link has a degrees of freedom (df) parameter, which provides more flexibility in the response curve and hence for inference. In fact, both the logistic and probit models are well approximated by a robit model with appropriate values of this df parameter. Specifically, a robit link with df value about seven provides an excellent approximation to the logit link, and the probit link is well-approximated by the robit link with large value of df. Roy (2014) who shows that when analyzing independent binary data, for the robit model to provide robust inference, it is necessary to estimate the df parameter from data rather than assuming a fixed df. If the df parameter is chosen appropriately, the robit model replicates the logistic or probit models if the data follows one of those models, but provides a robust alternative when extreme observations are present.

In this paper we develop methods to include the robit link function family to the SGLMMs in a Bayesian framework, which is more robust to extreme observations than the standard SGLMMs with fixed link functions. The challenging part of the new model is how to estimate the df parameter of the robit link function in SGLMMs. Though conjugate normal-inverse gamma priors can be assigned to the regression coefficients and the partial sill parameter, assigning a prior on the df parameter and subsequent posterior simulation is problematic. For example, using a flat prior on df skews the results towards the probit link (Doss, 2012, p. 20). Similarly construction of meaningful priors for some parameters of the correlation function of the underlying Gaussian random fields, such as the range parameter and the nugget parameter, is also difficult, and the choice of prior may influence the inference (Christensen, 2004, p. 716). Use of improper priors on correlation parameters typically results in an improper posterior distribution (Christensen and Waagepetersen (2002, p. 283), Berger et al. (2001); Ren and Sun (2014)). This is why, in practice, these parameters are often dealt

with in an ad hoc way. For example, in the geoRglm package in R, a discrete prior for the range parameter is used to simplify the calculation.

Here we consider a "half-way" Bayesian approach (Christensen, 2004, p. 716) which requires specification of prior distributions on some of the parameters. Instead of a full Bayesian (FB) analysis, we avoid having to specify a prior on df and some parameters of the spatial correlation where little information is available by developing an empirical Bayes (EB) approach. In our approach the prior knowledge on some of the parameters is incorporated through conjugate prior distributions, and the other parameters (denoted altogether by a vector ξ) are estimated by maximizing the marginal likelihood function. That is, we select that value of ξ which maximizes the marginal likelihood, $m_\xi = m_\xi(\mathbf{y})$, of the data $\mathbf{y}$. Note that, for the models discussed here, the marginal likelihood function m_ξ is not available in closed form, and estimating $m_\xi, \xi \in \mathcal{G}$, in general, is very difficult, where we are interested in a family of models indexed by $\xi \in \mathcal{G}$ for some set $\mathcal{G}$. Since the value of ξ that maximizes m_ξ, is the same as the value of ξ that maximizes am_ξ for $\xi \in \mathcal{G}$ where a is a constant, we calculate and subsequently compare the values of $B_{\xi,\xi_1} \equiv m_\xi/m_{\xi_1}$, where ξ_1 is a suitably chosen fixed value. (Note that, in this case $a = 1/m_{\xi_1}$.) The reason for estimating B_{ξ,ξ_1} instead of m_ξ directly is that we need to calculate and compare B_{ξ,ξ_1} for a wide range of values of ξ, and it is much easier to calculate B_{ξ,ξ_1} than m_ξ for all $\xi \in \mathcal{G}$. Doss (2010) considers a method based on importance sampling for selecting prior hyperparameters by estimating a large family of Bayes factors (see also Roy, 2014). Following Doss (2010) and Roy (2014) we estimate the df and the correlation parameters of our spatial robit model by estimating and subsequently maximizing B_{ξ,ξ_1}.

Estimation of the parameters in SGLMMs can be done using the Monte Carlo EM gradient method (Zhang, 2002) or the Monte Carlo maximum likelihood method (Geyer and Thompson, 1992; Geyer, 1994b). The advantage of the proposed methodology over the former is

that we obtain the whole marginal likelihood, not just the MLE (Christensen, 2004, p 708) and the latter can be unstable as the choice of the (importance) sampling distribution has a strong influence on the Monte Carlo maximum likelihood estimator (Christensen, 2004, p 705). Here we use efficient importance sampling methods based on multiple Markov chains for estimating the Bayes factors and subsequently the model parameters.

Another advantage of the proposed method is that we do not need to sample from the nonstandard conditional distributions of the link function and correlation parameters, which do not have closed form and is required in the FB analysis. Further, an MCMC algorithm with updates on such parameters usually does not perform well in terms of mixing and convergence (see e.g. Christensen, 2004, p. 716) and would be slow when the number of sampling locations is large as an inversion of a large covariance matrix will be required at every MCMC iteration. In the simulation studies presented in Section 4, we observe that EB analysis results in estimates with less bias than a FB analysis when each has the same number of MCMC iterations. In our EB approach, by plugging in the point estimate $\hat{\xi}$ of ξ, the posterior distribution of other parameters like the regression parameters do not reflect the uncertainty in ξ. On the other hand, as opposed to the FB analysis (which falls under so-called "Bayesian model averaging") the EB methodology leads to inference that is more parsimonious and interpretable (see Robert, 2007, ch 7).

We compare the proposed spatial robit model with the popular logistic and probit models in terms of prediction. In our simulation examples, we see that the robit model gives better predictions when the true value of the df parameter is low. When the df is large, the three models give similar predictions. The robit model consistently provides good predictions even when the data are simulated from a different model. This shows that the robit model is robust against model misspecification. To the best of our knowledge, we are not aware of

any significant simulation study comparing the robit model with logistic and probit models even in the case of analyzing independent binomial data.

We discuss two applications of the robit model to spatial data. The first example consists of the presence/absence of the invasive vine *Celastrus Orbiculatus*. This plant damages native plants in the United States by girdling so prediction of its spatial distribution is essential for estimating its impact. Wang et al. (2010) compared different binary spatial models, including the aforementioned GEV link model, to these data and found that all models were influenced by outlying observations. A robust model is needed in this case to accurately model the spatial distribution of the plant. In our analysis in Section 5 we find that a robit link function with small df is able to capture the spatial distribution of the plant better than the logit and probit models.

For our second example we consider the *Rhizoctonia root rot* data set collected in the Cunningham farm in the state of Washington (Zhang, 2002). One of the main obstacles to directly seeded (that is, seeded without plowing) wheat and barley is the root disease Rhizoctonia root rot caused by the fungi *Rhizoctonia Solani* and *Rhizoctonia Oryzae*. These fungi attach to the roots and hinder plants from absorbing adequate amount of water and nutrients from soil. A map of severity of the root rot is invaluable in site-specific farming (also known as precision agriculture) that is aimed at targeting inputs of fertilizer, pesticides and fungicides according to local requirements. Zhang (2002) uses the logit link for analyzing this data set and does not consider other alternatives. In our analysis in Section 5 we demonstrate that the probit (instead of logit) link is more appropriate in this example.

The rest of the article is organized as follows. Section 2 introduces our spatial robit model. In Section 3 we present a method based on importance sampling for efficiently estimating B_{ξ,ξ_1} and thus effectively selecting the link function parameter as well as the correlation parameters. In that section we also discuss the computation of the Bayesian predictive

density function. To illustrate the usefulness of spatial robit model and the proposed EB methodology, we present results from simulation studies in Section 4. We apply our model and method to two examples in Section 5. Some remarks and discussion appear in Section 6.

2. A robust spatial generalized linear mixed model for binomial data

Suppose $\mathcal{D}$ is a spatial domain of interest. Let $\{Z(s), s \in \mathcal{D}\}$ be an isotropic Gaussian random field with mean function $E(Z(s)) = \sum_{j=1}^{p} f_j(s)\beta_j$, where $\beta = (\beta_1, \ldots, \beta_p)' \in \mathcal{R}^p$ are the unknown regression parameters, $f(s) = (f_1(s), \ldots, f_p(s))$ are the known location dependent covariates, and the covariance function $\text{cov}(Z(s), Z(s')) = \sigma^2\rho_\theta(s, s') + \tau^2 I_{\{s=s'\}}$. Here $\rho_\theta(s, s') = \rho_\theta(\|s-s'\|) = \text{corr}(Z(s), Z(s'))$, where $\|s-s'\|$ denotes the Euclidean distance between s and s', θ is a vector of parameters which controls the range of correlation and the smoothness/roughness of the random field, τ^2 is called the nugget effect. The nugget effect can be interpreted as micro-scale variation, measurement error, or a combination of both. The two particular families of correlation function $\rho_\theta(u)$, where $u = \|s - s'\|$, that we use in this paper are the *Matérn family* and the *spherical family*. The Matérn family of correlation functions has the form

$$\rho(u; \phi, \kappa) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^\kappa K_\kappa(u/\phi),$$

where $K_\kappa(\cdot)$ denotes the modified Bessel function of order κ. This two-parameter family is very flexible in that $\kappa > 0$ controls the smoothness of the underlying process, while the parameter $\phi > 0$ measures the scale (in units of distance) on which the correlation decays. The spherical family has a single range parameter $\phi > 0$ and is of the form

$$\rho(u; \phi) = \begin{cases} 1 - 1.5(u/\phi) + 0.5(u/\phi)^3 & 0 \leqslant u \leqslant \phi \\ 0 & u > \phi \end{cases}.$$

We assume that, conditional on the latent process $\{z(s), s \in \mathcal{D}\}$, and for any $s_1, \ldots, s_n \in \mathcal{D}$, the corresponding measurement random variables $Y(s_1), \ldots, Y(s_n)$ are independent, that is, $Y(s_i)|z(s_i) \overset{ind}{\sim} \text{Binomial}(\ell_i, p_i)$ with $p_i = h_\nu^{-1}(z(s_i))$ where $h_\nu(\cdot)$ is the link function and ℓ_i is

a known constant. In particular, we introduce the robit link function, that is, $h_\nu^{-1}(\cdot) = F_\nu(\cdot)$, where F_ν is the CDF of the standard Student's t distribution with degrees of freedom ν.

Suppose that the data $\mathbf{y} = (y_1, \ldots, y_n)$ consist of a single realization of the process $\{Y(s), s \in \mathcal{D}\}$ at known sampling locations $s_1, \ldots, s_n \in \mathcal{D}$. The likelihood function is not available in closed form in this case, but only as a high dimensional integral, that is,

$$L(\beta, \sigma^2, \tau^2, \kappa, \nu|\mathbf{y}) = \int_{\mathcal{R}^n} \left[\prod_{i=1}^{n} p(y_i|z_i, \nu)\right] f(\mathbf{z}|\beta, \sigma^2, \tau^2, \kappa) d\mathbf{z}, \tag{1}$$

where $\mathbf{z} = (z(s_1), \ldots, z(s_n))$, $z_i \equiv z(s_i)$, $p(y_i|z_i, \nu)$ is the binomial probability mass function, that is,

$$p(y_i|z_i, \nu) = \binom{\ell_i}{y_i} (p_i)^{y_i} (1 - p_i)^{\ell_i - y_i},$$

with $p_i = F_\nu(z_i)$, and $f(\mathbf{z}|\beta, \sigma^2, \tau^2, \kappa)$ is the multivariate Gaussian density for $\mathbf{z}$, that is,

$$f(\mathbf{z}|\beta, \sigma^2, \tau^2, \kappa) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2}(\mathbf{z} - F\beta)^T \Sigma^{-1} (\mathbf{z} - F\beta)\},$$

where F is the known $n \times p$ matrix defined by $F_{ij} = f_j(s_i)$, Σ is the covariance matrix involving the parameters σ^2, τ^2, and κ.

We use the following Gaussian prior for β (conditional on σ^2) and scaled inverse chi-square prior for σ^2

$$\beta|\sigma^2 \sim N_p(m_b, \sigma^2 V_b), \text{ and } \sigma^2 \sim \chi^2_{ScI}(n_\sigma, a_\sigma), \tag{2}$$

where the hyperparameters $m_b, V_b, a_\sigma, n_\sigma$ are assumed known. (We say $W \sim \chi^2_{ScI}(n_\sigma, a_\sigma)$ with parameters $a_\sigma > 0, n_\sigma > 0$ if the probability density function (pdf) of W is $f(w) \propto w^{-(n_\sigma/2+1)} \exp(-n_\sigma a_\sigma/(2w))$.) We denote $\psi \equiv (\beta, \sigma^2)$. The other parameters are either assumed known or estimated by maximizing the marginal likelihoods. We use ξ to denote the parameters estimated by maximizing marginal likelihoods. Here we take $\xi \equiv (\nu, \phi, \tau)$ or $\xi \equiv (\nu, \phi)$ if τ is assumed known. For fixed ξ the posterior density of ψ is

$$\pi_\xi(\psi|\mathbf{y}) = \frac{L_\xi(\psi|\mathbf{y})\pi(\psi)}{m_\xi(\mathbf{y})}, \tag{3}$$

where the likelihood function $L_\xi(\psi|\mathbf{y}) \equiv L(\psi, \xi|\mathbf{y})$ is defined in (1), $\pi(\psi)$ is the prior given

in (2), and the normalizing constant is

$$m_\xi(\mathbf{y}) = \int_{\mathcal{R}^p \times \mathcal{R}_+} L_\xi(\psi|\mathbf{y})\pi(\psi)d\psi.$$

The EB estimator for ξ is that value of ξ which maximizes the marginal likelihood of the data $m_\xi \equiv m_\xi(\mathbf{y})$.

Since the likelihood function $L_\xi(\psi|y)$ in (1) is not available in closed form, we work with the augmented joint density

$$f(\mathbf{y}, \mathbf{z}|\psi, \xi) = \Big[\prod_{i=1}^{n} p(y_i|z_i, \nu)\Big] f(\mathbf{z}|\beta, \sigma^2, \kappa). \tag{4}$$

and the corresponding so-called *complete* posterior density

$$\pi_\xi(\psi, \mathbf{z}|\mathbf{y}) = \frac{f(\mathbf{y}, \mathbf{z}|\psi, \xi)\pi(\psi)}{m_\xi(\mathbf{y})}. \tag{5}$$

Since $\int_{\mathcal{R}^n} f(\mathbf{y}, \mathbf{z}|\psi, \xi)d\mathbf{z} = L_\xi(\psi|\mathbf{y})$, we have $\int_{\mathcal{R}^n} \pi_\xi(\psi, \mathbf{z}|\mathbf{y})d\mathbf{z} = \pi_\xi(\psi|\mathbf{y})$, that is, integrating the complete posterior density $\pi_\xi(\psi, \mathbf{z}|\mathbf{y})$ we get the target posterior density $\pi_\xi(\psi|\mathbf{y})$. So if we can generate a Markov chain $\{\psi^{(i)}, \mathbf{z}^{(i)}\}_{i\geqslant 1}$ with stationary density $\pi_\xi(\psi, \mathbf{z}|\mathbf{y})$, then the marginal chain $\{\psi^{(i)}\}_{i\geqslant 1}$ has the stationary density $\pi_\xi(\psi|\mathbf{y})$ defined in (3). This is the standard technique of data augmentation and here $\mathbf{z}$ is playing the role of "latent" variables (or "missing data") (Tanner and Wong, 1987). Note that the normalizing constant $m_\xi(\mathbf{y})$ is the same for both the posterior densities (3) and (5). Also note that the priors given in (2) are conjugate priors for (β, σ^2) for the joint density $f(\mathbf{y}, \mathbf{z}|\psi, \xi)$. In fact $\beta|\sigma^2, \mathbf{z}, \mathbf{y}, \xi \sim N(\tilde{\beta}, \sigma^2\tilde{V}_b)$, and $\sigma^2|\mathbf{z}, \mathbf{y}, \xi \sim \chi^2_{ScI}(n + n_\sigma, b_\sigma)$, for some parameters $\tilde{\beta}, \tilde{V}_b$ and b_σ (see e.g. Diggle et al., 2003).

3. Computational Methods

The computations are performed in two stages. In the first step, we estimate the Bayes factors $B_{\xi,\xi_1} = m_\xi/m_{\xi_1}$ (where ξ_1 is a fixed value) using an efficient importance sampling method based on multiple Markov chains. These Bayes factors are then used to obtain the estimate $\hat{\xi}$ of the parameters ξ. This procedure is described in Section 3.1. In the second

stage, described in Section 3.2, we fix $\xi = \hat{\xi}$ and estimate the remaining parameters ψ as well as the underlying spatial random field $\{Z(s)\}$ at observed and unobserved locations.

3.1 *Computation of the Bayes factors via efficient importance sampling methods*

Let $\{\psi^{(i)}, \mathbf{z}^{(i)}\}_{i\geqslant 1}$ be a Markov chain that is appropriately irreducible and has invariant density $\pi_{\xi_1}(\psi, \mathbf{z}|\mathbf{y})$. We use a Metropolis-Hastings within Gibbs algorithm for constructing such a Markov chain. The conditional distribution $\pi_{\xi}(\psi|\mathbf{z}, \mathbf{y})$ is standard, therefore, ψ can be easily updated in the Gibbs sampling algorithm. For sampling from $\pi_{\xi}(\mathbf{z}|\psi, \mathbf{y})$ we use a Metropolis-Hastings algorithm as in Diggle et al. (1998) or Zhang (2002). Since by the ergodic theorem

$$\frac{1}{N}\sum_{i=1}^{N}\frac{f(\mathbf{y}, \mathbf{z}^{(i)}|\psi^{(i)}, \xi)}{f(\mathbf{y}, \mathbf{z}^{(i)}|\psi^{(i)}, \xi_1)} \xrightarrow{\text{a.s.}} \int_{\mathcal{R}^n}\int_{\mathcal{R}^p\times\mathcal{R}_+}\frac{f(\mathbf{y}, \mathbf{z}|\psi, \xi)}{f(\mathbf{y}, \mathbf{z}|\psi, \xi_1)}\pi_{\xi_1}(\psi, \mathbf{z}|\mathbf{y})d\psi d\mathbf{z} = \frac{m_{\xi}(\mathbf{y})}{m_{\xi_1}(\mathbf{y})}, \quad (6)$$

as $N \to \infty$, where $f(\mathbf{y}, \mathbf{z}|\psi, \xi)$ is the joint density given in (4), the entire family, $\{B_{\xi,\xi_1} : \xi \in \mathcal{G}\}$ can be consistently estimated by (6) using samples from only one posterior distribution namely $\pi_{\xi_1}(\psi, \mathbf{z}|\mathbf{y})$. As mentioned in Christensen (2004) and Doss (2010) the estimate (6) is often unstable and only a few terms dominate the estimator. A natural approach for dealing with the instability of the estimator in (6) is to choose a grid of k points $\xi_1, \xi_2, \ldots, \xi_k \in \mathcal{G}$ and replace $f(\mathbf{y}, \mathbf{z}|\psi, \xi_1)$ in the denominator of (6) by a linear combination $\sum_{i=1}^{k} a_i f(\mathbf{y}, \mathbf{z}|\psi, \xi_i)$ for some constants a_i's. In particular, if $\{\psi_j^{(l)}, \mathbf{z}^{(j;l)}\}_{l=1}^{N_j}$ is a Markov chain with stationary density $\pi_{\xi_j}(\psi, \mathbf{z}|\mathbf{y})$ for $j = 1 \ldots, k$, then we have

$$\sum_{j=1}^{k}\sum_{l=1}^{N_j}\frac{f(\mathbf{y}, \mathbf{z}^{(j;l)}|\psi_j^{(l)}, \xi)}{\sum_{i=1}^{k} N_i f(\mathbf{y}, \mathbf{z}^{(j;l)}|\psi_j^{(l)}, \xi_i)/r_i} \xrightarrow{\text{a.s.}} B_{\xi,\xi_1},$$

as $N \equiv \sum_{j=1}^{k} N_j \to \infty$, $N_i/N \to a_i$ where $r_i = m_{\xi_i}/m_{\xi_1}, i = 2, \ldots, k$, $r_1 = 1$.

In practice $\mathbf{r} = (r_1, r_2, \ldots, r_k)$ is of course unknown, however, it can be estimated consistently using the "reverse logistic regression" method as proposed in Geyer (1994a). Let $\hat{\mathbf{r}} = (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_k)$ with $\hat{r}_1 = 1$ be the estimate of $\mathbf{r}$. Then B_{ξ,ξ_1} can be consistently estimated

by

$$\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}}) = \sum_{j=1}^{k} \sum_{l=1}^{N_j} \frac{f(\mathbf{y}, \mathbf{z}^{(j;l)}|\psi_j^{(l)}, \xi)}{\sum_{i=1}^{k} N_i f(\mathbf{y}, \mathbf{z}^{(j;l)}|\psi_j^{(l)}, \xi_i)/\hat{r}_i}. \tag{7}$$

Doss (2010, p.548) gives some guidelines for choosing good values of k and the skeleton points $\xi_1, \xi_2, \ldots, \xi_k$ (see also Buta and Doss, 2011). For our model, this method is illustrated in the supplementary material.

A two-stage procedure for estimating B_{ξ,ξ_1} is used. In stage I, we draw a large sample from $\pi_{\xi_j}(\psi, \mathbf{z}|\mathbf{y})$, for each $j = 1, 2, \ldots, k$ which does not take much time since the MCMC algorithms are quite fast. Note that this step can also be done in parallel. We calculate $\hat{\mathbf{r}}$ using this sample. Since the entire vectors $\mathbf{z}^{(i)}$'s need to be stored, we store only a thinned version of the $\mathbf{z}^{(i)}$'s which reduces the dependence between consecutive elements of the stored values. Independently of stage I, in stage II we get *new* samples $\{\psi_j^{(l)}, \mathbf{z}^{(j;l)}\}_{l=1}^{N_j}$ from the posterior densities $\pi_{\xi_j}(\mathbf{z}, \psi|\mathbf{y}),\ j = 1, 2, \ldots, k$ and use them to estimate B_{ξ,ξ_1} using $\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}})$ defined in (7). As Doss (2010) explains, the reason for using the two-stage procedure is that in stage II we need to calculate B_{ξ,ξ_1} for a large number of values of ξ and for each ξ the amount of computation required to calculate B_{ξ,ξ_1} is linear in N which rules out large N. On the other hand, it is desirable to use a large sample in stage I to estimate $\mathbf{r}$ accurately. The sample sizes of both stages can be determined based the standard errors of $\hat{\mathbf{r}}$ and $\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}})$ (See Supplementary Materials). Finally, a quasi-Newton optimization procedure is used to maximize $\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}})$ and estimate ξ.

3.2 *Prediction*

Let $\hat{\xi}$ be the EB estimate of ξ obtained in Section 3.1. Estimation of the remaining parameters and the spatial random field is done by standard Bayesian methods. To this end, let $\{\psi^{(i)}, \mathbf{z}^{(i)}\}_{i=1}^{M}$ be a MCMC sample with stationary density $\pi_{\hat{\xi}}(\psi, \mathbf{z}|\mathbf{y})$, drawn using the algorithm mentioned at the beginning of Section 3.1. Then, an estimate of ψ is obtained by $\sum_{i=1}^{M} \psi^{(i)}/M$. Let us now consider prediction about Z_0, the values of $Z(s)$ at some

locations of interest, $(s_{01}, s_{02}, \ldots, s_{0k})$ (typically a fine grid of locations covering the observed region) using the posterior predictive distribution. Note that Z_0 and $Y(\cdot)$ are conditionally independent given $(Z(\cdot), \psi, \xi)$ and

$$f(\mathbf{z}_0|\mathbf{y}) = \int_{\mathcal{R}^p \times \mathcal{R}_+} \int_{\mathcal{R}^n} f(\mathbf{z}_0|\mathbf{z}, \psi, \xi) \pi_\xi(\psi, \mathbf{z}|\mathbf{y}) d\mathbf{z} d\psi, \tag{8}$$

where $\mathbf{z}_0 = (z(s_{01}), z(s_{02}), \ldots, z(s_{0k}))$. The conditional density $f(\mathbf{z}_0|\mathbf{z}, \psi, \xi)$ is a multivariate normal density. The mean vector and the dispersion matrix of this normal density are the usual simple kriging mean and covariance matrix (Diggle et al., 2003).

Suppose, we want to estimate $E(t(\mathbf{z}_0)|\mathbf{y})$ for some function t. If $t^*(\mathbf{z}, \psi, \hat{\xi}) \equiv E(t(\mathbf{z}_0)|\mathbf{z}, \psi, \hat{\xi})$ is available in closed form, then we estimate $E(t(\mathbf{z}_0)|\mathbf{y})$ by $\sum_{i=1}^M t^*(\mathbf{z}^{(i)}, \psi^{(i)}, \hat{\xi})/M$. Otherwise, we simulate $\mathbf{z}_0^{(i)}$ from $f(\mathbf{z}_0|\mathbf{z}^{(i)}, \psi^{(i)}, \hat{\xi})$ for $i = 1, \ldots, M$ and calculate the following approximate minimum mean squared error predictor

$$E(t(\mathbf{z}_0)|\mathbf{y}) \approx \frac{1}{M} \sum_{i=1}^{M} t(\mathbf{z}_0^{(i)}). \tag{9}$$

The samples $\{\mathbf{z}_0^{(i)}\}_{i=1}^M$ are also used to compute prediction quantiles corresponding to the predictive density (8) and the predictive distribution of $t(\mathbf{z}_0)$.

Below we summarize the steps involved in the estimation of the parameters and the prediction of the random field.

• Finding EB estimate of ξ.

Stage 1 Generate MCMC samples $\{\psi_j^{(l)}, \mathbf{z}^{(j;l)}\}_{l=1}^{\mathcal{N}_j}$ from $\pi_{\xi_j}(\psi, \mathbf{z}|\mathbf{y})$, for each $j = 1 \ldots, k$, and use these samples to estimate $\mathbf{r}$ by the reverse logistic regression method.

Stage 2 Independently of Stage 1, again generate MCMC samples $\{\psi_j^{(l)}, \mathbf{z}^{(j;l)}\}_{l=1}^{N_j}$ from $\pi_{\xi_j}(\psi, \mathbf{z}|\mathbf{y})$, for each $j = 1 \ldots, k$, and estimate the Bayes factors B_{ξ,ξ_1} by (7) based on these $N = \sum_{j=1}^k N_j$ observations and $\hat{\mathbf{r}}$ computed in Stage 1. Find $\hat{\xi}$ by maximizing $\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}})$.

• Estimating ψ and predicting Z_0.

Once the EB estimate $\hat{\xi}$ of ξ is formed as described above, generate new MCMC samples $\{\psi^{(i)}, \mathbf{z}^{(i)}\}_{i=1}^{M}$ from $\pi_{\hat{\xi}}(\psi, \mathbf{z}|\mathbf{y})$. Estimate ψ by $\sum_{i=1}^{M} \psi^{(i)}/M$, and the minimum mean squared error predictor of $t(\mathbf{z}_0)$ is given by (9).

---

4. Simulations

We perform a simulation study to assess the performance of the EB method and the robit link function. Our simulation study considers several choices for the range ϕ and link function as shown in Table 1. The domain for the simulations is fixed to $\mathcal{D} = [0, 1]^2$ and prediction of the Gaussian random field $\mathbf{z}_0$ is considered at an 11×11 square grid covering $\mathcal{D}$. A realization of the data $\mathbf{y}$ consists of observations from the binomial spatial model at $n = 100$ randomly chosen sites with number of trials $\ell_i = 250$ for all $i = 1, \ldots, 100$. The mean of the random field is set to 1.7 for the left half of the domain and to -1.7 for the right half while its covariance was chosen from the Matérn family with fixed nugget $\tau^2 = 0.2$, partial sill $\sigma^2 = 1$, smoothness $\kappa = 0.5$ and range ϕ. Out of these parameters τ^2 and κ are considered known to simplify the comparison, while σ^2 and ϕ are to be estimated along with the mean parameters and the link function parameter.

4.1 *Estimation performance*

We carry out simulations in order to illustrate the performance of our proposed EB estimation method for $\xi = (\nu, \phi)$. We vary the range parameter and the binomial link function in different simulations and study the performance of our method for estimating these parameters. Our simulations consist of 100 realizations from the binomial spatial generalized linear mixed models for each of the 10 different settings mentioned in Table 1. For the prior on (β, σ^2), we used (2) with mean m_b a vector of zeros and standardized variance $V_b = 100I$ with I the 2×2 identity matrix, and $a_\sigma = 1 = n_\sigma$.

At the first stage of Section 3.1, we estimate $\mathbf{r}$ at a small number of skeleton points. The

skeleton set is chosen to be a square grid that varies depending on the 'true' parameter values. To compute $\hat{\mathbf{r}}$ accurately, we observe that for large values of ν we need less number points in the skeleton set than when ν is small. We use fifteen points in the skeleton set when the true link is logit or probit, and twenty-one points when $\nu = 0.5, 1, 4$. The Bayes factors at the skeleton set are computed using MCMC samples of size 2000 from the posterior densities at each of the skeleton point. These samples are taken after discarding an initial burn-in of 1000 samples and keeping every 5th draw of subsequent random samples. As mentioned in Section 3.1, we use a Metropolis-Hastings within Gibbs algorithm with stationary density $\pi_\xi(\psi, \mathbf{z}|\mathbf{y})$. The full conditional density $\pi_\xi(\psi|\mathbf{z}, \mathbf{y})$ is Gaussian-Scaled inverse chi-square. The conditional distribution of $\mathbf{z}$ given $(\psi, \mathbf{y})$ is not a standard distribution and we used a Metropolis-Hastings algorithm given in Zhang (2002) (see also Diggle et al., 1998) for sampling from this conditional distribution.

For the second stage of Section 3.1, we use *fresh* random samples of size 500 using the same MCMC algorithm to compute the Bayes factors B_{ξ,ξ_1} at other points using (7). The estimate $\hat{\xi} \equiv (\hat{\nu}, \hat{\phi})$ is taken to be the value of ξ where B_{ξ,ξ_1} attains its maximum. We use a quasi-Newton algorithm to maximize the Bayes factors B_{ξ,ξ_1}. Figure 1 shows the contour plots of $\hat{B}_{\xi,\xi_1}$ for four different settings (corresponding to one simulated data set). The EB estimates for ϕ and ν given in Table 1 are the means of 100 estimates corresponding to 100 simulated data sets. Table 1 also shows the root mean squared error (rmse) estimates based on these 100 simulations. The EB method produces good estimates for the two parameters.

For comparison, the parameter ϕ is also estimated by FB under the models with fixed logit ($\hat{\phi}_{\text{lg}}$) or probit ($\hat{\phi}_{\text{pr}}$) link functions and a prior on ϕ is specified. The following priors were considered: uniform in $(0, 1.5)$, exponential with mean the true ϕ, half-normal with mean the true ϕ, and inverse-gamma with shape and scale parameters equal to 1. Of these, the half-normal prior is the most informative for ϕ with coefficient of variation $\sqrt{\pi/2 - 1} \approx 0.76$

and gave the best estimates for that parameter. However, in terms of all other parameters and for prediction they all gave similar answers. The MCMC algorithm for the FB approach uses the Metropolis-Hastings within Gibbs algorithm as before together with a log normal proposal density for the full conditional density of ϕ. The variance of the log normal density is set to achieve 20-30% acceptance rate.

For comparison, Table 1 shows the estimates obtained using the half-normal prior. (The results for all parameters and models are presented in the supplementary material.) Evidently, the EB exhibits less bias in all cases for all priors but the variability is higher compared to the half-normal prior, however we find that the variability of FB depends strongly on the prior for ϕ, e.g. the inverse-gamma prior gives a higher RMSE. The same also holds for the estimates of β and σ^2 (not shown here).

[Figure 1 about here.]

4.2 *Comparison of link functions using prediction performance*

The prediction of random effects is of paramount importance in spatial analysis. In this section we compare the proposed robit model and methodology against the alternative logit and probit models with a FB approach. To the best of our knowledge, we do not know of any simulation study other than Gelman and Hill (2007, ch. 6) that compares the robit link with the logit and probit links even in the case of uncorrelated binomial data.

To assess the performance in terms of prediction for each model we use the logarithmic scoring, defined as follows. Suppose we are interested in the future value of a random variable X from a density $f(x;\theta)$ where θ is unknown. Given data $Y = y$ from a density $p(y;\theta)$, we construct two predictive densities: $f_1(x|y)$ and $f_2(x|y)$, derived by making different assumptions about the joint distribution $f(x, y;\theta)$. Then, given a realized value $X = x_1$, we assign a score s_j to the predictive density f_j equal to

$$s_j = \log f_j(x_1|y) \text{ for } j = 1, 2.$$

If $s_1 > s_2$, then $f_1(x|y)$ is a better predictor of X. When i.i.d. realizations $x_1, \ldots, x_m$ are available, then the average scores are compared.

Here, the data consist of the simulated value $\mathbf{y}$ and the parameter consists of the true values of the model parameters together with the realized random field while prediction is sought at a 11×11 square grid covering the whole domain. The models fitted in Section 4.1 were also compared here.

The Monte Carlo samples corresponding to each prediction location are used to construct the predictive density using kernel density estimation. These predictive densities are then evaluated at 1,000 random samples of the random field from the true conditional distribution at each prediction location given the value of the random field at the observed locations and the true values of the parameters. The average logarithmic score over the 1,000 samples is computed at each location. Table 1 shows the average score over the 121 prediction locations using each model under the different parameter settings. Higher score implies better prediction of random fields. Although we assumed that the range parameter ϕ is unknown, the estimates for the df parameter for the proposed model and the corresponding prediction scores are very similar to the case where ϕ is known (not shown here).

The prediction scores for the three models in Table 1 show that the model with robit link gives better predictions when the true df is low. As the df increases, the three models give similar predictions, however the model with robit link is consistently giving good scores even when it is not the true model. This shows that the robit model is robust against model misspecification.

[Table 1 about here.]

5. Applications

5.1 *Analysis of the Celastrus Orbiculatus data*

The *Celastrus Orbiculatus* (oriental bittersweet) is a perennial, deciduous vine which is native to East Asia but considered an invasive species in North America as it can cause major damage to native plants by girdling. Monitoring the spatial distribution of the plant is necessary to accurately assess its impact to its habitat (Latimer et al., 2009; Mosher et al., 2009).

In this example we examine data on the presence/absence of *C.Orbiculatus* collected from 603 locations in Connecticut, USA, along with some environmental predictors. Of these, 200 randomly chosen samples were deleted from the data set while the remaining $n = 403$ data were used to fit the models described below and predict the probability of presence at the deleted locations. These data were also analyzed by Wang et al. (2010) who also discussed the use of flexible link functions for fitting these data. Indeed the authors detected a number of outlying observations which were not explained by the models that they used, including the GEV model. We advocate that the robit link is more suitable for this example.

The presence is coded as a binary variable and modeled by a binomial SGLMM. As in Wang et al. (2010), the spatial random field was taken to have mean a linear combination of the environmental variables in the data set plus an intercept term, a total of 16 regression coefficients, and variance of the Matérn form with $\kappa = 0.5$ and no nugget effect. For the regression parameters we assign independent normal priors with mean $m_b = 0$ and standardized variance $v_b = 1000$. For the sill parameter we assign a scaled inverse chi-square prior with degrees of freedom $n_\sigma = 4$ and scale $a_\sigma = 0.5$, as in Wang et al. (2010). In our analysis we fit the robit, logit and probit models. We use the proposed EB method for the former and fit the latter two models using both EB and FB.

The skeleton set for the EB was set to

$$(\phi, \nu) \in \{2, 4, 6\} \times \{1, 3, 5\},$$

with reference $\xi_1 \equiv (4, 1)$. In the case of the logit and probit models we only used the part of the skeleton set associated with ϕ. For each point in the skeleton set, we took a sample of size $N = 2000$ from the posterior distributions of the unknown variables after a burn-in of 1000 and thinning 5. This sample was used to estimate the Bayes factors at the skeleton points by reverse logistic regression. A subsequent sample of size 500 was used to estimate ϕ and, if needed, ν. A new MCMC sample of the same size N, burn-in and thinning was sampled from the posterior distributions of the remaining variables with ϕ and ν fixed at their earlier estimates. This new sample was used to estimate the other parameters and for prediction.

For the FB method ϕ was assigned a half-normal prior with mean 2. An MCMC sample of the same size N, burn-in and thinning as above was taken from the posterior distributions of all the variables which was used for parameter estimation and prediction.

A scoring measure was used to assess the predictive performance of each model (see Celeux et al., 2006; Wang et al., 2010). This is defined as

$$S = \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} \log p(\mathbf{y}_0 | \mathbf{z}_0^{(i,j)}, \nu), \tag{10}$$

where $\mathbf{y}_0$ corresponds to the observed presence/absence at the 200 deleted locations and $\mathbf{z}_0^{(i,j)}$ is the jth independent replication of the spatial field $\mathbf{z}_0$ at the same locations using the parameters from the ith step of the MCMC. A higher score corresponds to a better model. For this analysis we used $M = 1000$.

Table 2 shows the results from fitting each model. It is shown that the optimal df parameter for the robit model is small to account for the outlying observations noted in Wang et al. (2010). It is also shown that the robit model outperforms the logit and probit models fitted either by EB or by FB.

[Table 2 about here.]

5.2 *Analysis of the Rhizoctonia root rot data*

The Rhizoctonia root rot is a disease that affects the roots of plants by hindering their process of absorbing water and nutrients from soil. This example was first considered in Zhang (2002); 15 plants were selected from each of $n = 100$ randomly chosen locations in a farm and the number of crown roots and infected crown roots were counted. For the purpose of treating the disease and site-specific farming, we wish to predict the random field over the whole area. As in Zhang (2002), we assume that the number of infected crown roots $Y(s_i)$ at site s_i has a Binomial(ℓ_i, p_i) distribution, where ℓ_i is the total number of crown roots at site s_i and p_i is the varying binomial success probability. As in Zhang (2002) we also assume a constant mean and a spherical correlation function with unknown nugget effect for the underlying Gaussian random field.

A binomial spatial model with robit link is fitted. We use the same prior on (β, σ^2) as in Section 4.1. For the estimation of $\xi \equiv (\phi, \nu, \omega)$, where $\omega = \tau^2/\sigma^2$, we compute $\hat{\mathbf{r}}$ at the following 36 skeleton points,

$$(\phi, \nu, \omega) \in \{100, 140, 180\} \times \{3, 5, 8, 15\} \times \{0.5, 1, 2\}.$$

These points were chosen to minimize the average asymptotic standard error for B_{ξ,ξ_1} for a range of values for ξ as discussed in the supplementary material. The parameter ω is known as the relative nugget parameter.

As in Section 4, we use the reverse logistic regression method using an MCMC sample of size 2,000, which is taken collecting every 5th sample after discarding initial 1,000 samples. A fresh MCMC sample of size 500 is subsequently drawn for the purpose of computing B_{ξ,ξ_1} for all other values of (ϕ, ν, ω) with $\xi_1 \equiv (\phi_1, \nu_1, \omega_1) = (140, 3, 1)$. The EB estimators are obtained by maximizing B_{ξ,ξ_1} using a quasi-Newton algorithm. In Figure 2 we show the profile of the Bayes factors with respect to each of the parameters, that is, one of the parameters is

kept fixed and the maximum of the Bayes factors over the other two parameters is plotted against the values of the first parameter. The maximum is obtained at $\hat{\phi} = 147$, $\hat{\nu} = 30$, and $\hat{\omega} = 0.93$. Figure 2 suggests that probit link can be used for analyzing this data set.

Treating these estimates as fixed, we subsequently estimate the other parameters of the model by running an MCMC. The estimates obtained are $\hat{\beta} = -1.05$ and $\hat{\sigma}^2 = 0.12$ and the 95% prediction intervals of β and σ^2 are given by $(-1.20, -0.90)$ and $(0.09, 0.17)$ respectively.

[Figure 2 about here.]

The robit model was compared against the probit and logit models fitted by EB. The same skeleton points were used as with the robit model except for the df. For comparison we removed 25 observations randomly and used the remaining 75 observations to fit the model and predict at the 25 deleted locations. Table 3 shows the estimates from each model as well as the score, as defined in (10) with $M = 1000$. We note that the three models score similarly in terms of prediction which is not surprising since for high df the predicted probabilities do not differ substantially. Also, the parameter estimates for the robit and probit models are very similar.

[Table 3 about here.]

6. Discussion

The usefulness of the robit link function for spatially correlated binomial data is demonstrated in this article. Since the construction of meaningful priors for the degrees of freedom parameter as well as some correlation parameters is difficult, an empirical Bayes methodology is developed for making inference on these parameters by maximizing the marginal likelihood function. The techniques presented here is useful for other types of geostatistical models, like Poisson SGLMM with a parametric family of link functions (e.g. Box-Cox family) for count data. The proposed methodology can also be extended to spatial ordinal data. Since

the importance sampling based methods proposed in the article requires evaluation of the probability density function of the underlying Gaussian random fields, (which requires $O(n^3)$ calculation for a data set with n observations), extending our methodology to large data sets is challenging. As a possible avenue for future work, it would be interesting to see if some approximate likelihood methods, for example the covariance tapering method (Kaufman et al., 2008), can be implemented together with the importance sampling procedure proposed here.

7. Supplementary Materials

Further details on the selection of the skeleton points, simulations, data analysis, referenced in Sections 3.1, 4.1, 5 and sample R code are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgments

The authors would like to thank Prof. Dipak K. Dey and Prof. John A. Silander for providing the Celastrus Orbiculatus data set. They also thank the editor, an associate editor and two reviewers for helpful comments and valuable suggestions which led to several improvements in the manuscript.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88** 669–679.

Berger, J. O., De Oliveira, V. and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96** 1361–1374.

Buta, E. and Doss, H. (2011). Computational approaches for empirical Bayes methods and Bayesian sensitivity analysis. *The Annals of Statistics*, **39** 2658–2685.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, **1** 651–673.

Christensen, O. F. (2004). Monte Carlo maximum likelihood in model based geostatistics. *Journal of Computational and Graphical Statistics*, **13** 702–718.

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

Diggle, P. J., Ribeiro, P. J. and Christensen, O. F. (2003). An introduction to model-based geostatistics. In *Spatial statistics and computational methods. Lecture notes in statistics*. Springer, 43–86.

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

Doss, H. (2010). Estimation of large families of Bayes factors from Markov chain output. *Statistica Sinica*, **20** 537–560.

Doss, H. (2012). Hyperparameter and model selection for nonparametric Bayes problems via Radon-Nikodym derivatives. *Statistica Sinica*, **22** 1–26.

Evangelou, E., Zhu, Z. and Smith, R. L. (2011). Estimation and prediction for spatial generalized linear mixed models using high order Laplace approximation. *Journal of Statistical Planning and Inference*, **141** 3564–3577.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, Cambridge University Press.

Geyer, C. J. (1994a). Estimating normalizing constants and reweighting mixtures in Markov chain Monte Carlo. Tech. Rep. 568, School of Statistics, University of Minnesota.

Geyer, C. J. (1994b). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society,* Series B, **56** 261–274.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood

for dependent data. *Journal of the Royal Statistical Society,* Series B, **54** 657–699.

Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, **103** 1545–1555.

Latimer, A. M., Banerjee, S., Sang Jr, H., Mosher, E. S. and Silander Jr, J. A. (2009). Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. *Ecology Letters*, **12** 144–154.

Liu, C. (2004). Robit regression: A simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Casual Inference from Incomplete-Data Perspectives* (A. Gelman and X. L. Meng, eds.). Wiley, London, 227–238.

Mosher, E. S., Silander Jr, J. A. and Latimer, A. M. (2009). The role of land-use history in major invasions by woody plant species in the northeastern North American landscape. *Biological Invasions*, **11** 2317–2328.

Pregibon, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, **38** 485–498.

Ren, C. and Sun, D. (2014). Objective Bayesian analysis for autoregressive models with nugget effects. *Journal of Multivariate Analysis*, **124** 260–280.

Robert, C. P. (2007). *The Bayesian Choice*. Springer.

Roy, V. (2014). Efficient estimation of the link function parameter in a robust Bayesian binary regression model. *Computational Statistics and Data Analysis*, **73** 87–102.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation(with discussion). *Journal of the American Statistical Association*, **82** 528–550.

Wang, X., Dey, D. K. and Banerjee, S. (2010). Non-Gaussian hierarchical generalized linear geostatistical model selection. In *Frontiers of Statistical Decision Making and*

Bayesian Analysis (M. H. Chen, D. K. Dey, P. Müller, D. Sun and K. Ye, eds.). Springer, 484–496.

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

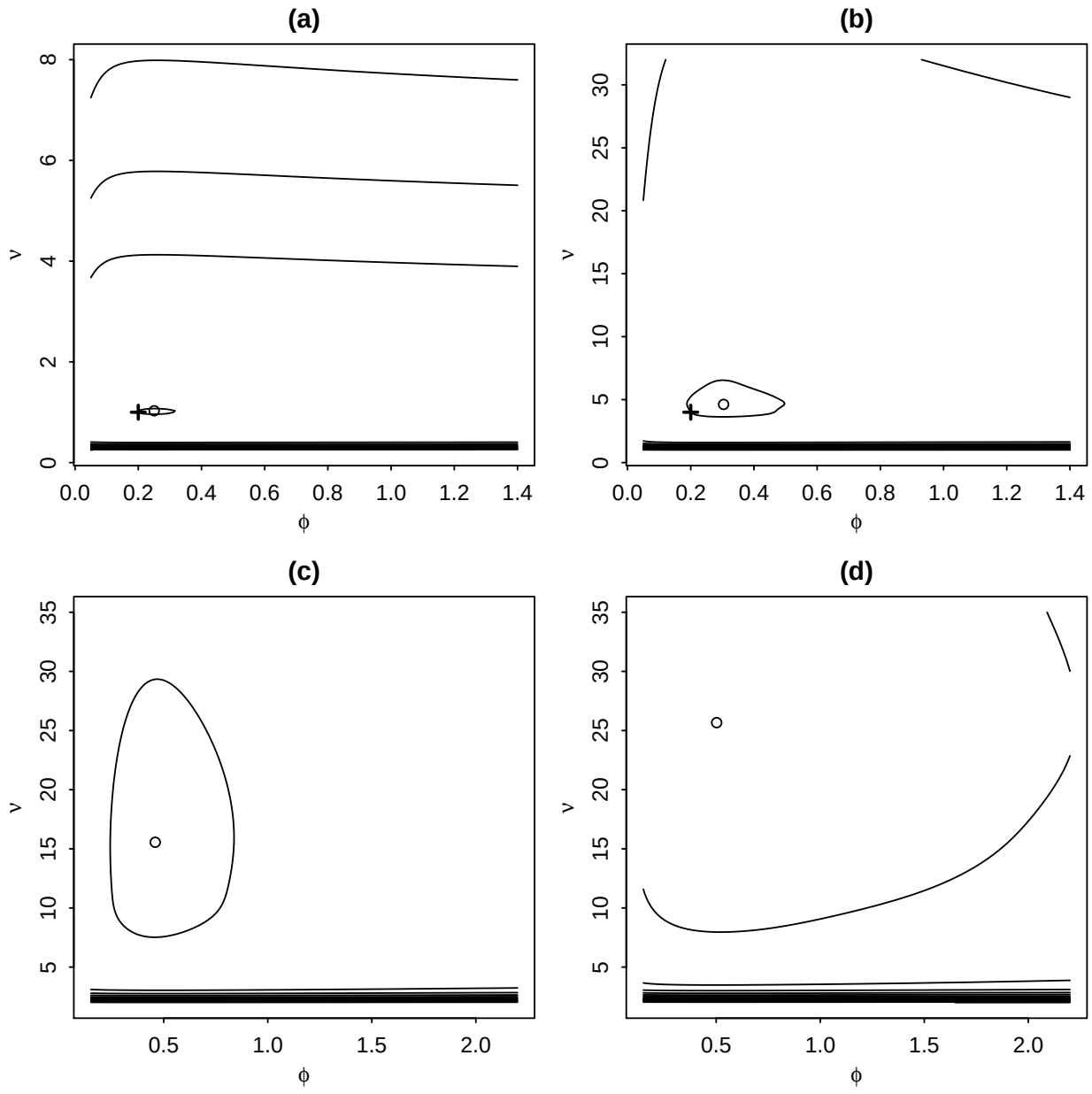


Figure 1. Contour plots for estimates of $B_{(\nu,\phi),(\nu_1,\phi_1)}$ for a simulated data from the following models: (a) Robit(1), $\phi = 0.2$; (b) Robit(4), $\phi = 0.2$; (c) Logit, $\phi = 0.6$; (d) Probit, $\phi = 0.6$. The EB estimate is shown by a $\circ$ and the true value by a $+$.

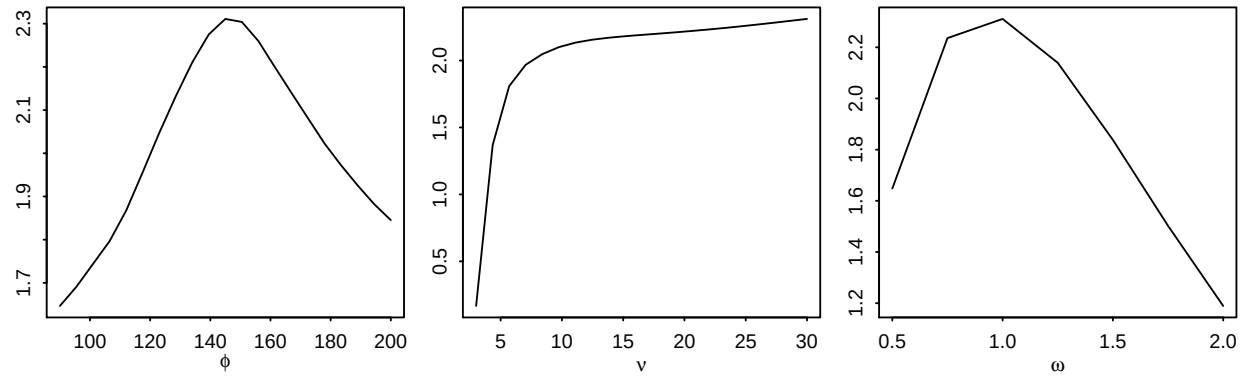


Figure 2. Profile of $B_{(\nu,\phi,\omega),(\nu_1,\phi_1,\omega_1)}$ for the Rhizoctonia data with respect to each parameter. The plot suggests a probit link, ϕ around 147 and τ^2/σ^2 around 0.93.

Table 1

Simulation results showing the true model, parameter estimates and prediction scores for each fitted model. Subscript eb denotes the empirical Bayes robit model, lg denotes the full Bayesian logit model with half-normal prior on ϕ, pr denotes the full Bayesian probit model with the same prior.

True model

	$\phi = 0.2$					$\phi = 0.6$				
ν	0.5	1.0	4.0	logit	probit	0.5	1.0	4.0	logit	probit
Estimates										
$\hat{\nu}_{\text{eb}}$	0.53	1.08	4.61	10.30	23.72	0.52	1.09	4.42	9.84	24.23
rmse($\hat{\nu}_{\text{eb}}$)	0.09	0.24	1.98			0.10	0.27	1.47		
$\hat{\phi}_{\text{eb}}$	0.22	0.20	0.20	0.20	0.21	0.58	0.62	0.55	0.60	0.57
rmse($\hat{\phi}_{\text{eb}}$)	0.15	0.16	0.13	0.15	0.16	0.53	0.53	0.48	0.55	0.46
$\hat{\phi}_{\text{lg}}$	0.31	0.25	0.22	0.22	0.22	0.89	0.74	0.61	0.59	0.56
rmse($\hat{\phi}_{\text{lg}}$)	0.13	0.08	0.06	0.06	0.06	0.37	0.24	0.19	0.18	0.18
$\hat{\phi}_{\text{pr}}$	0.32	0.26	0.23	0.23	0.22	0.92	0.77	0.64	0.62	0.60
rmse($\hat{\phi}_{\text{pr}}$)	0.14	0.09	0.07	0.06	0.06	0.40	0.27	0.20	0.19	0.19
Prediction scores										
Score_{eb}	-1.31	-1.16	-1.15	-1.23	-1.17	-1.86	-1.30	-0.96	-0.97	-0.97
SD Score_{eb}	0.32	0.10	0.11	0.16	0.10	1.07	0.45	0.09	0.11	0.08
Score_{lg}	-68.64	-13.82	-1.21	-1.15	-1.34	-99.92	-12.76	-1.01	-0.93	-1.17
SD Score_{lg}	78.00	17.53	0.18	0.10	0.15	113.62	14.73	0.14	0.07	0.16
Score_{pr}	-95.78	-25.89	-1.74	-1.30	-1.15	-128.95	-25.13	-1.71	-1.12	-0.94
SD Score_{pr}	106.80	30.38	0.74	0.23	0.10	144.86	30.96	0.93	0.24	0.07

Table 2

Estimates and prediction score for the C.Orbiculatus data. EB denotes the empirical Bayes procedure and FB the full Bayesian procedure. The EB Robit Compl denotes the model using all 603 observations.

	EB Robit Compl	EB Robit	EB Logit	FB Logit	EB Probit	FB Probit
$\hat{\nu}$	2.12	1.40				
$\hat{\phi}$	3.20	6.15	3.40	2.78	2.86	2.80
Score		-149	-202	-208	-307	-262

Table 3
Comparison of the robit, logit, and probit models for the rhizoctonia example.

	Robit	Logit	Probit
$\hat{\nu}$	30		
$\hat{\phi}$	142	143	142
$\hat{\omega}$	0.40	0.46	0.38
$\hat{\beta}$	-1.03	-1.11	-1.00
$\hat{\sigma}^2$	0.17	0.21	0.16
score	-1690	-1695	-1693

Web-based Supplementary Materials for "Efficient estimation and prediction for the Bayesian binary spatial model with flexible link functions"

by

Vivekananda Roy, Evangelos Evangelou, Zhengyuan Zhu

Web Appendix A

Further details on the simulations

The data are simulated from the following model, where the sampled locations $s_i = (s_{i,1}, s_{i,2})$, $i = 1, \ldots, 100$ are shown in Web Figure 1.

$$\begin{aligned} Y(s_i)|z(s_i) &\sim \text{Binomial}(250, p(s_i)), \\ h_\nu(p(s_i)) &= z(s_i), \\ Z(s_i) &\sim \text{GRF}(\mu(s_i), \tau^2 + \sigma^2\rho(u;\phi)), \\ \mu(s_i) &= \beta_0 + \beta_1 \times \mathbf{1}(s_{i,1} > 0.5). \end{aligned}$$

In the above $\text{GRF}(\mu(s), c(u))$ denotes the distribution of the Gaussian random field with mean at the spatial location s $\mu(s)$ and covariance function $c(u)$. We set $\beta_0 = 1.70$, $\beta_1 = -3.40$, $\tau^2 = 0.20$, $\sigma^2 = 1.00$, $\rho(u;\phi)$ to be the Matérn correlation function with $\kappa = 0.5$ and ϕ and ν varying according to Web Table 1.

The following models were considered for fitting the data

EB Robit link fitted by EB with skeleton points as in Web Table 1;

MCLg U Logit link with Uniform$(0, 1.5)$ prior for ϕ;

MCLg Ex Logit link with Exponential(ϕ) prior for ϕ;

MCLg HN Logit link with Half-Normal(mean $= \phi$) prior for ϕ;

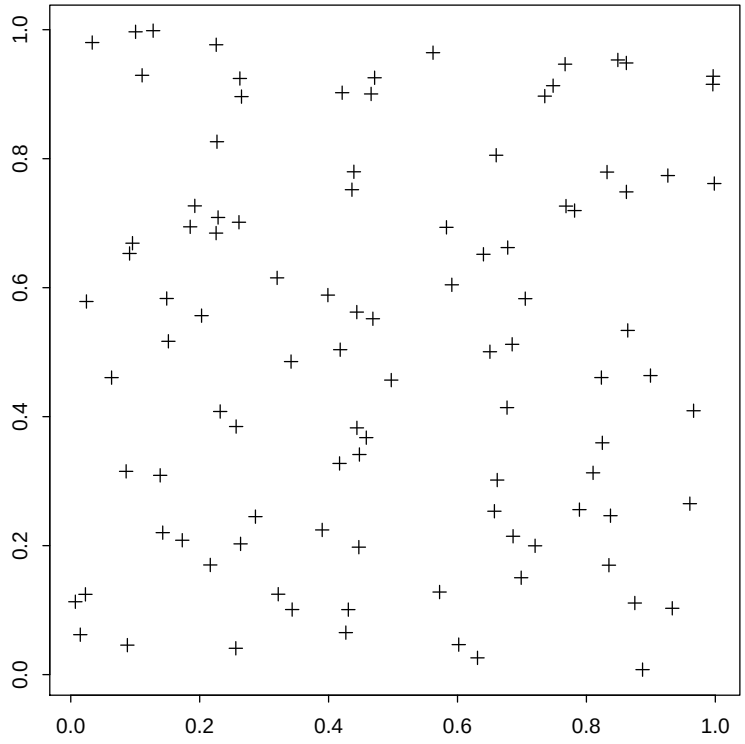
MCLg IG Logit link with Inverse-Gamma(shape $= 1$, scale $= 1$) prior for ϕ;

MCPr U Probit link with Uniform$(0, 1.5)$ prior for ϕ;

MCPr Ex Probit link with Exponential(ϕ) prior for ϕ;

MCPr HN Probit link with Half-Normal(mean $= \phi$) prior for ϕ;

MCPr IG Probit link with Inverse-Gamma(shape $= 1$, scale $= 1$) prior for ϕ.



Web Figure 1 – Sampling locations for the simulations.

ϕ true	0.2 or 0.6	
ϕ skeleton	$\phi \cdot (0.5, 1, 2)$	
ν true	0.5, 1.0, 4.0	logit or probit
ν skeleton	$\nu \cdot (0.5, 0.6, 0.8, 1.0, 1.2, 1.6, 2.0)$	$(4, 6, 8, 12, 15)$

Web Table 1 – Skeleton points for the EB method for the simulations. In the table, the ϕ and ν factors correspond to their true values.

In addition, the following priors were assumed for all models

$$\beta_j|\sigma^2 \overset{\text{ind}}{\sim} N(0, 100\sigma^2), \text{ for } j = 0, 1,$$
$$\sigma^2 \sim \chi^2_{ScI}(1, 1).$$

Prediction is considered at a 11×11 square grid covering the sampling domain and a prediction score is computed for each method. The total computing time for each combination of parameters including all 100 repetitions was between 7 to 9 hours on a computer with Intel(R) Core(TM) i5-2500 CPU, 3.30GHz processor and 4Gb RAM. Web Tables 2–6 show the results derived from 100 simulations corresponding to a new simulated z and y each time.

	$\phi = 0.2$					$\phi = 0.6$				
ν	0.5	1.0	4.0	logit	probit	0.5	1.0	4.0	logit	probit
EB	0.22	0.20	0.20	0.20	0.21	0.58	0.62	0.55	0.60	0.57
	0.15	0.16	0.13	0.15	0.16	0.53	0.53	0.48	0.55	0.46
MCLg U	0.74	0.58	0.46	0.44	0.43	0.95	0.83	0.71	0.69	0.66
	0.58	0.43	0.32	0.30	0.28	0.39	0.29	0.22	0.21	0.20
MCLg Ex	0.36	0.27	0.23	0.23	0.22	1.01	0.78	0.61	0.58	0.56
	0.20	0.12	0.08	0.07	0.07	0.56	0.35	0.24	0.23	0.24
MCLg HN	0.31	0.25	0.22	0.22	0.22	0.89	0.74	0.61	0.59	0.56
	0.13	0.08	0.06	0.06	0.06	0.37	0.24	0.19	0.18	0.18
MCLg IG	1.44	1.03	0.74	0.65	0.71	2.74	1.75	1.04	1.17	1.35
	2.18	1.45	0.84	0.59	1.07	5.64	3.22	0.80	1.95	4.13
MCPr U	0.77	0.61	0.49	0.46	0.44	0.97	0.86	0.75	0.73	0.70
	0.61	0.47	0.35	0.32	0.30	0.40	0.32	0.24	0.23	0.22
MCPr Ex	0.38	0.29	0.24	0.23	0.23	1.05	0.83	0.65	0.63	0.60
	0.22	0.13	0.08	0.08	0.08	0.61	0.39	0.27	0.25	0.24
MCPr HN	0.32	0.26	0.23	0.23	0.22	0.92	0.77	0.64	0.62	0.60
	0.14	0.09	0.07	0.06	0.06	0.40	0.27	0.20	0.19	0.19
MCPr IG	1.74	1.35	0.76	0.68	0.76	3.09	2.00	1.27	1.13	1.14
	2.97	4.04	0.80	0.62	1.31	6.86	3.73	1.99	1.15	1.66

Web Table 2 – Estimates (first row) and RMSE (second row) of the spatial range parameter ϕ under each model.

	$\phi = 0.2$					$\phi = 0.6$				
ν	0.5	1.0	4.0	logit	probit	0.5	1.0	4.0	logit	probit
EB	0.93	0.93	0.96	0.94	1.18	0.92	0.89	0.90	0.92	1.09
	0.36	0.42	0.35	0.35	0.41	0.56	0.46	0.42	0.41	0.37
MCLg U	0.19	0.33	0.94	1.31	2.31	0.14	0.26	0.73	1.01	1.74
	0.82	0.68	0.24	0.44	1.41	0.86	0.75	0.32	0.22	0.85
MCLg Ex	0.14	0.25	0.73	1.03	1.82	0.14	0.24	0.67	0.92	1.59
	0.86	0.75	0.32	0.22	0.91	0.86	0.76	0.37	0.24	0.72
MCLg HN	0.14	0.25	0.73	1.03	1.82	0.14	0.24	0.68	0.94	1.62
	0.87	0.76	0.31	0.21	0.90	0.86	0.76	0.36	0.23	0.74
MCLg IG	0.20	0.38	1.09	1.52	2.70	0.16	0.28	0.79	1.10	1.89
	0.80	0.63	0.27	0.61	1.80	0.84	0.73	0.27	0.26	1.00
MCPr U	0.17	0.28	0.64	0.83	1.32	0.13	0.21	0.50	0.65	1.02
	0.83	0.73	0.40	0.27	0.44	0.87	0.79	0.51	0.38	0.23
MCPr Ex	0.13	0.21	0.49	0.65	1.03	0.13	0.20	0.47	0.60	0.94
	0.87	0.79	0.52	0.38	0.22	0.88	0.80	0.55	0.43	0.24
MCPr HN	0.12	0.21	0.49	0.65	1.03	0.12	0.20	0.47	0.61	0.96
	0.88	0.80	0.52	0.38	0.21	0.88	0.80	0.54	0.42	0.22
MCPr IG	0.18	0.31	0.74	0.96	1.54	0.14	0.23	0.55	0.71	1.11
	0.82	0.69	0.32	0.21	0.62	0.86	0.77	0.47	0.34	0.27

Web Table 3 – Estimates (first row) and RMSE (second row) of the spatial sill parameter σ^2 under each model. The true parameter value is $\sigma^2 = 1$.

	$\phi = 0.2$					$\phi = 0.6$				
ν	0.5	1.0	4.0	logit	probit	0.5	1.0	4.0	logit	probit
EB	1.61	1.62	1.67	1.64	1.80	1.74	1.70	1.75	1.73	1.88
	0.54	0.51	0.48	0.50	0.57	0.88	0.83	0.80	0.82	0.92
MCLg U	0.64	0.90	1.46	1.65	2.07	0.65	0.91	1.52	1.71	2.16
	1.07	0.83	0.48	0.51	0.78	1.07	0.85	0.64	0.75	1.12
MCLg Ex	0.64	0.90	1.47	1.67	2.09	0.65	0.91	1.52	1.71	2.16
	1.07	0.83	0.46	0.48	0.78	1.07	0.85	0.64	0.75	1.13
MCLg HN	0.64	0.90	1.47	1.67	2.09	0.65	0.91	1.52	1.71	2.16
	1.07	0.83	0.46	0.48	0.78	1.07	0.85	0.64	0.75	1.12
MCLg IG	0.64	0.90	1.46	1.65	2.06	0.65	0.91	1.52	1.71	2.16
	1.07	0.84	0.48	0.51	0.79	1.07	0.85	0.64	0.76	1.13
MCPr U	0.60	0.83	1.25	1.38	1.67	0.61	0.84	1.29	1.42	1.72
	1.11	0.90	0.56	0.51	0.51	1.11	0.91	0.64	0.64	0.75
MCPr Ex	0.61	0.83	1.26	1.40	1.68	0.61	0.84	1.30	1.43	1.73
	1.10	0.89	0.54	0.48	0.49	1.11	0.91	0.64	0.63	0.75
MCPr HN	0.61	0.83	1.26	1.40	1.68	0.61	0.84	1.30	1.42	1.73
	1.10	0.89	0.54	0.48	0.49	1.11	0.91	0.63	0.64	0.74
MCPr IG	0.60	0.82	1.25	1.39	1.66	0.62	0.84	1.29	1.42	1.72
	1.11	0.90	0.56	0.51	0.52	1.10	0.91	0.64	0.64	0.74

Web Table 4 – Estimates (first row) and RMSE (second row) of the parameter β_0 under each model. The true parameter value is $\beta_0 = 1.70$.

	$\phi = 0.2$					$\phi = 0.6$				
ν	0.5	1.0	4.0	logit	probit	0.5	1.0	4.0	logit	probit
EB	-3.23	-3.23	-3.33	-3.26	-3.58	-3.37	-3.28	-3.35	-3.34	-3.60
	0.69	0.63	0.55	0.53	0.63	0.79	0.72	0.55	0.54	0.60
MCLg U	-1.28	-1.79	-2.92	-3.31	-4.14	-1.28	-1.78	-2.93	-3.33	-4.15
	2.13	1.63	0.63	0.51	1.03	2.13	1.64	0.59	0.42	0.97
MCLg Ex	-1.28	-1.79	-2.93	-3.32	-4.15	-1.28	-1.78	-2.93	-3.33	-4.15
	2.13	1.62	0.62	0.50	1.03	2.13	1.64	0.59	0.43	0.98
MCLg HN	-1.28	-1.79	-2.93	-3.33	-4.16	-1.28	-1.78	-2.93	-3.33	-4.15
	2.13	1.62	0.62	0.50	1.03	2.13	1.64	0.59	0.43	0.97
MCLg IG	-1.28	-1.79	-2.92	-3.32	-4.14	-1.28	-1.78	-2.93	-3.33	-4.15
	2.13	1.62	0.63	0.51	1.03	2.13	1.64	0.58	0.42	0.97
MCPr U	-1.21	-1.65	-2.50	-2.77	-3.33	-1.20	-1.64	-2.51	-2.78	-3.32
	2.20	1.77	0.96	0.74	0.52	2.20	1.78	0.94	0.70	0.43
MCPr Ex	-1.21	-1.65	-2.51	-2.78	-3.34	-1.20	-1.64	-2.50	-2.78	-3.33
	2.20	1.76	0.94	0.73	0.51	2.20	1.78	0.94	0.70	0.44
MCPr HN	-1.21	-1.65	-2.51	-2.78	-3.34	-1.20	-1.64	-2.51	-2.78	-3.32
	2.20	1.76	0.94	0.72	0.50	2.20	1.78	0.94	0.70	0.43
MCPr IG	-1.21	-1.65	-2.51	-2.78	-3.33	-1.20	-1.64	-2.51	-2.78	-3.32
	2.20	1.76	0.95	0.74	0.52	2.20	1.77	0.94	0.70	0.43

Web Table 5 – Estimates (first row) and RMSE (second row) of the parameter β_1 under each model. The true parameter value is $\beta_1 = -3.40$.

	$\phi = 0.2$					$\phi = 0.6$				
ν	0.5	1.0	4.0	logit	probit	0.5	1.0	4.0	logit	probit
EB	-1.31	-1.16	-1.15	-1.23	-1.17	-1.86	-1.30	-0.96	-0.97	-0.97
	0.32	0.10	0.11	0.16	0.10	1.07	0.45	0.09	0.11	0.08
MCLg U	-68.83	-13.78	-1.22	-1.15	-1.34	-94.53	-13.18	-1.00	-0.93	-1.17
	79.45	15.26	0.18	0.10	0.15	110.78	17.07	0.14	0.07	0.14
MCLg Ex	-68.68	-13.68	-1.22	-1.15	-1.34	-92.88	-12.80	-1.01	-0.93	-1.17
	79.17	16.83	0.19	0.10	0.15	109.26	15.19	0.14	0.07	0.16
MCLg HN	-68.64	-13.82	-1.21	-1.15	-1.34	-99.92	-12.76	-1.01	-0.93	-1.17
	78.00	17.53	0.18	0.10	0.15	113.62	14.73	0.14	0.07	0.16
MCLg IG	-68.31	-13.85	-1.22	-1.15	-1.34	-95.06	-13.73	-1.01	-0.93	-1.17
	79.75	16.15	0.18	0.10	0.14	110.31	18.10	0.13	0.07	0.15
MCPr U	-90.85	-24.23	-1.72	-1.30	-1.15	-131.96	-25.10	-1.71	-1.12	-0.94
	97.05	28.09	0.74	0.22	0.10	151.12	31.99	0.96	0.25	0.07
MCPr Ex	-91.98	-24.51	-1.75	-1.30	-1.15	-132.00	-24.02	-1.74	-1.13	-0.94
	102.95	29.69	0.90	0.23	0.10	145.12	31.22	0.95	0.28	0.07
MCPr HN	-95.78	-25.89	-1.74	-1.30	-1.15	-128.95	-25.13	-1.71	-1.12	-0.94
	106.80	30.38	0.74	0.23	0.10	144.86	30.96	0.93	0.24	0.07
MCPr IG	-92.80	-24.04	-1.74	-1.29	-1.15	-131.38	-26.52	-1.74	-1.12	-0.94
	105.82	27.10	0.80	0.22	0.09	147.83	34.49	0.97	0.28	0.07

Web Table 6 – Prediction score (first row) and standard deviation of the score (second row) under each model.

Web Appendix B

Detailed analysis of the Celastrus Orbiculatus data

The data consist of the presence/absence of the *C.Orbiculatus* species at 603 locations. Web Figure 2 shows the sampling locations along with the presence/absence of the species. The following environmental variables were also used as predictors:

HabitatClass The current state of the habitat: a factor variable of four levels;

LULCChange Land use and land cover (LULC) change: a factor of 5 levels;

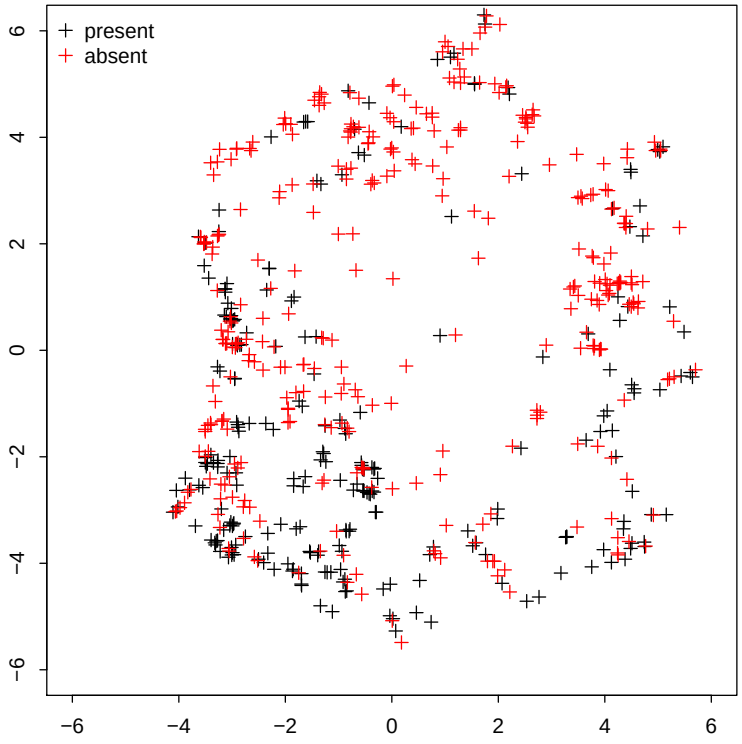
cat1970 What the land use category was in 1970: a factor of 6 levels;

CanopyClosure An ordinal score for the proportion of sky blocked by a canopy of leaves;

HeavilyManagedPts A binary indicator where 1 means heavy land management, e.g. a paved area or a lawn;

LogEdgeDistance The distance (in feet) of the plot to a vegetation edge, such as a forest in the logarithmic scale.

The first three variables are treated as factors and the last three as numerical. An intercept was also included, totaling $m = 16$ regression coefficients.



Web Figure 2 – Sampled locations for the *C.Orbiculatus* example.

Let y_i denote the presence/absence of the *C.Orbiculatus* species at location s_i and $Z(s)$ be the spatial process denoting the intensity of occurrence. We also denote by $f_j(s)$ the value of the jth

regression variable at location s. The following hierarchical model is used.

$$\begin{aligned} Y_i|z_i &\sim \text{Bin}(1, p_i), \\ h_\nu(p_i) &= z_i, \\ Z(s) &\sim \text{GRF}\left(\sum_{j=1}^{m} \beta_j f_j(s), \sigma^2 \rho(u;\phi)\right), \\ \beta_j|\sigma^2 &\overset{\text{ind}}{\sim} N(0, 1000\sigma^2),\ j = 1, \ldots, m, \\ \sigma^2 &\sim \chi^2_{ScI}(4, 0.5). \end{aligned} \tag{1}$$

In the above $\text{GRF}(\mu(s), c(u))$ denotes the distribution of the Gaussian random field with mean at the spatial location s $\mu(s)$ and covariance function $c(u)$. We fix $\rho(u;\phi)$ to the exponential correlation function, i.e. Matérn with $\kappa = 0.5$.

We consider the case where h_ν is the robit(ν), logit and probit link function. The model is fitted by both an empirical Bayes (EB) and a full Bayesian (FB) method. The unknown parameters consist of β_j, σ^2, ϕ and, in the case of the robit link, ν. For the EB method, the skeleton set for (ϕ, ν) was set to

$$(\phi, \nu) \in \{1, 3, 5\} \times \{2, 4, 6\},$$

for the robit model, while for the logit and probit models only the part which corresponds to ϕ was used.

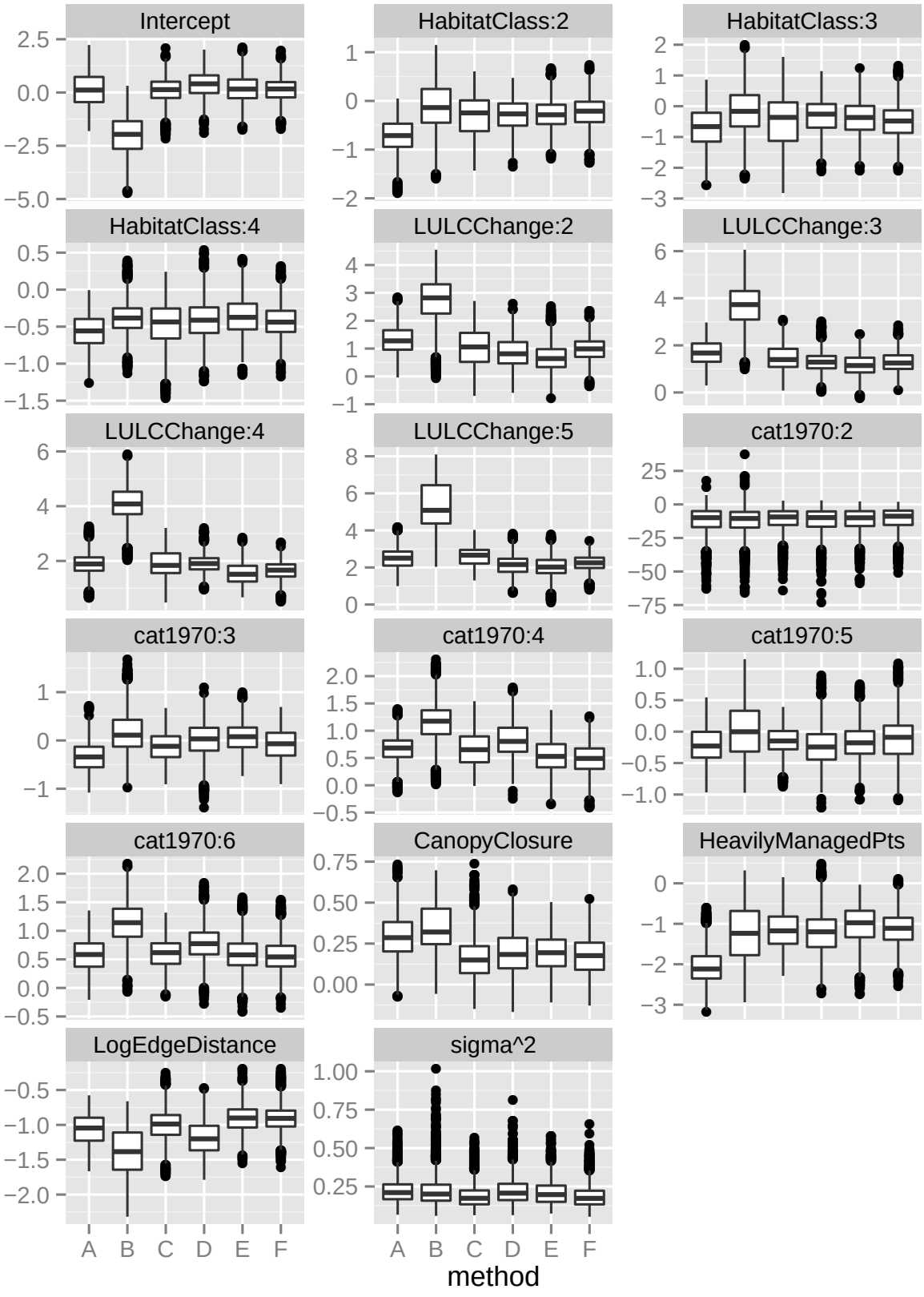
From the whole data set, 200 observations were randomly selected and the absence/presence was assumed unobserved. Each model was fitted using the remaining 403 observations and prediction is considered at the deleted 200 locations. A prediction score is then calculated as discussed in the main paper. Web Table 7 shows the parameter estimates obtained by each model as well as the prediction score. For comparison, we show the estimates obtained when all 603 observations are used in the case of the robit model.

Web Figure 3 shows the distribution of the posterior samples using each method for the parameters β and σ^2. The results from using EB or FB to fit the same model overlap substantially.

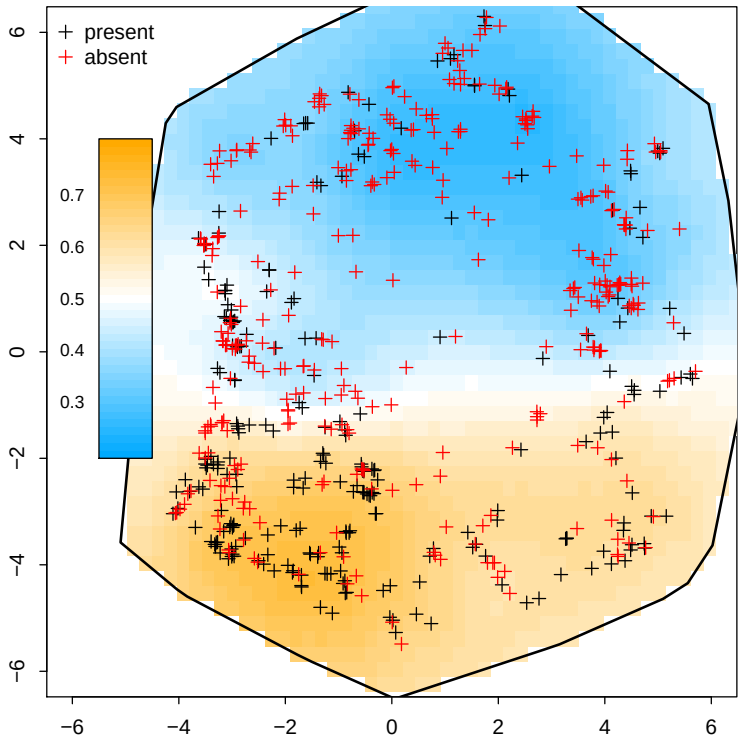
Using the fitted robit model with complete data, we consider prediction at a fine square grid covering the sampling domain. The predicted probability of presence is depicted in Web Figure 4. The figure shows that it is more likely to find the species at the southern part of the domain. It is evident that at some parts where the predicted probability is extreme, some observations appear as outliers.

	Rb EB Compl	Rb EB	Lg EB	Lg FB	Pr EB	Pr FB
Intercept	0.11	-2.02	0.10	0.38	0.18	0.13
HabitatClass:2	-0.72	-0.13	-0.31	-0.28	-0.26	-0.23
HabitatClass:3	-0.71	-0.17	-0.47	-0.31	-0.38	-0.51
HabitatClass:4	-0.57	-0.38	-0.47	-0.42	-0.36	-0.43
LULCChange:2	1.32	2.68	1.04	0.85	0.67	0.98
LULCChange:3	1.68	3.64	1.49	1.30	1.16	1.29
LULCChange:4	1.89	4.10	1.92	1.90	1.54	1.65
LULCChange:5	2.53	5.29	2.61	2.14	2.05	2.23
cat1970:2	-12.00	-12.32	-11.06	-11.96	-11.60	-10.92
cat1970:3	-0.34	0.17	-0.13	0.01	0.07	-0.08
cat1970:4	0.65	1.15	0.67	0.84	0.53	0.48
cat1970:5	-0.21	0.01	-0.14	-0.23	-0.17	-0.12
cat1970:6	0.58	1.13	0.61	0.79	0.60	0.56
CanopyClosure	0.29	0.34	0.15	0.19	0.19	0.18
HeavilyManagedPts	-2.04	-1.25	-1.14	-1.20	-1.04	-1.13
LogEdgeDistance	-1.07	-1.39	-1.00	-1.18	-0.91	-0.91
σ^2	0.23	0.23	0.19	0.22	0.21	0.19
ϕ	3.20	6.15	3.40	2.78	2.86	2.80
ν	2.12	1.40				
Score		-148.54	-202.47	-208.35	-306.81	-261.53

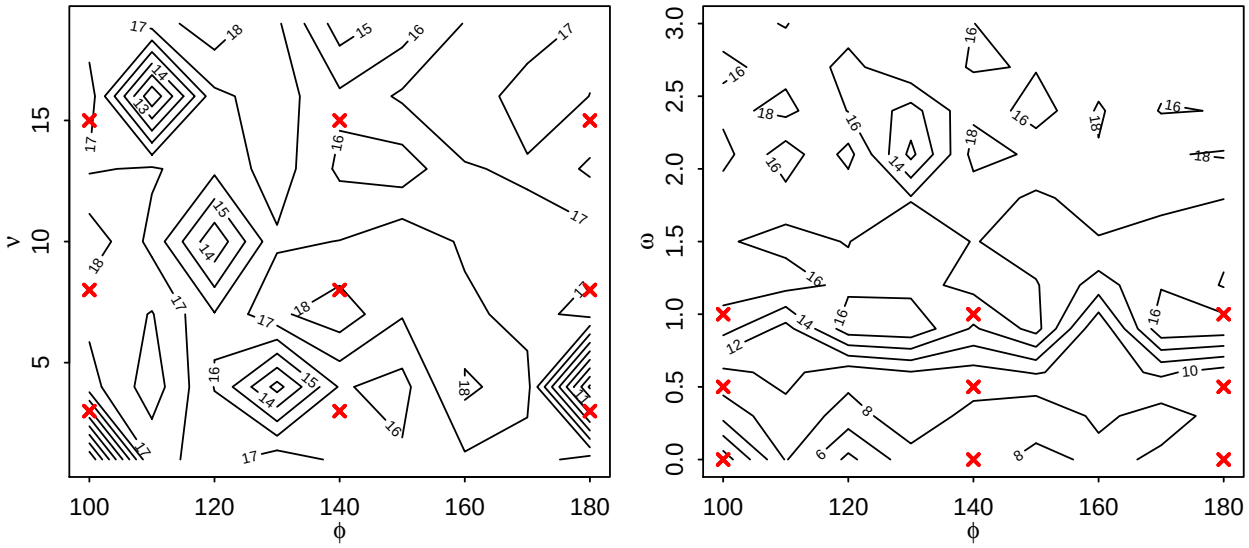
Web Table 7 – Parameter estimates for the *C.Orbiculatus* example from different models. EB denotes a model fitted by empirical Bayes and FB a model fitted by full Bayes. Rb, Lg, Pr are the robit, logit and probit link functions respectively and Compl means the complete data set was used.



Web Figure 3 – Posterior samples for β and σ^2 under different methods. Methods A–F correspond to the columns of Table 7.



Web Figure 4 – Predicted probability of occurrence for the *C.Orbiculatus* species.



Web Figure 5 – Standard errors (in log scale) of the Bayes factor estimates.

Web Appendix C

Illustration of the choice for the skeleton points for the Rhizoctonia example

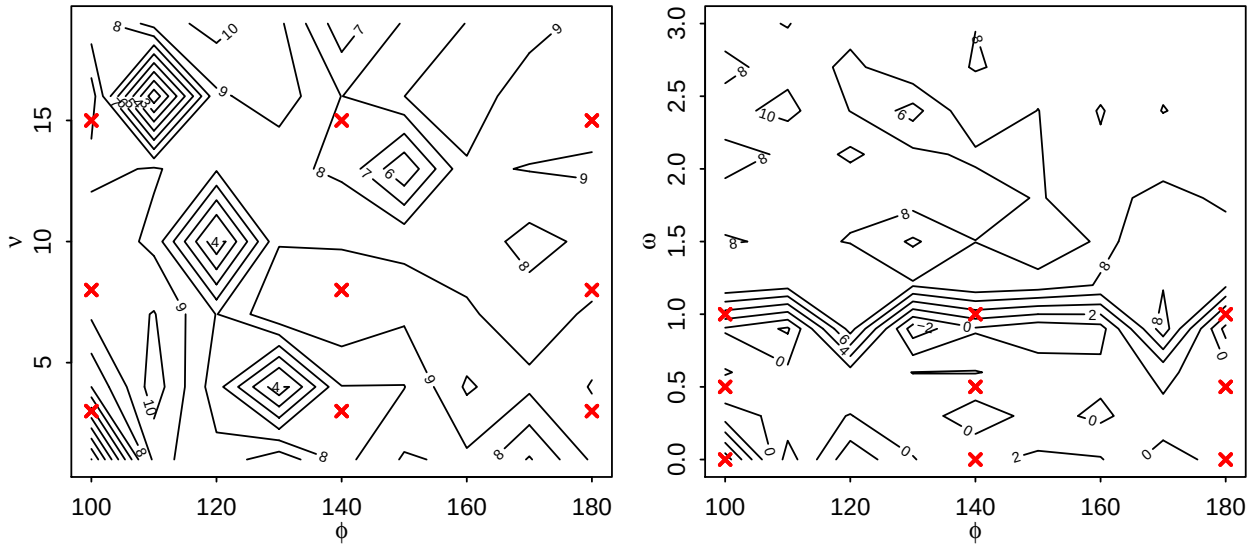
The Rhizoctonia root rot is a disease that affects the roots of plants by hindering their process of absorbing water and nutrients from soil. This example was first considered in Zhang (2002); 15 plants were selected from each of $n = 100$ randomly chosen locations in a farm and the number of crown roots and infected crown roots were counted. For the purpose of treating the disease and site-specific farming, we wish to predict the random field over the whole area. As in Zhang (2002), we assume that the number of infected crown roots $Y(s_i)$ at site s_i has a Binomial (ℓ_i, p_i) distribution, where ℓ_i is the total number of crown roots at site s_i and p_i is the varying binomial success parameter. Following Zhang (2002) we also assume a constant mean and a spherical correlation function with unknown nugget effect for the underlying Gaussian random field.

A binomial spatial model with robit link is fitted. We use the same prior on (β, σ^2) as is discussed in the main text. For the estimation of $\xi \equiv (\phi, \nu, \omega)$ by our proposed EB method, where $\omega = \tau^2/\sigma^2$, we start with the following 27 skeleton points,

$$(\phi, \nu, \omega) \in \{100, 140, 180\} \times \{3, 8, 15\} \times \{0, 0.5, 1\}. \tag{2}$$

The parameter ω is known as the relative nugget parameter. We use the reverse logistic regression method using an MCMC sample of size 800, which is taken collecting every 10th sample after discarding initial 1,000 samples. A fresh MCMC sample of size 200 is subsequently drawn for the purpose of computing B_{ξ,ξ_1} for all other values of (ϕ, ν, ω) with $\xi_1 \equiv (\phi_1, \nu_1, \omega_1) = (100, 3, 0)$. We use the asymptotic standard errors using batchmeans derived in Roy et al. (2015) to calculate the standard errors of our Bayes factor estimator $\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}})$. In Web Figure 5 we plot the profile of the standard errors, that is, the maximum of the standard errors over the free parameter is plotted against the fixed values of the other two parameters. From Web Figure 5 we see that the variance is large when ν is around 5, also when ω is large. The largest standard error is around 2.7×10^8.

Roy et al. (2015) also provides batchmeans estimates of the variance covariance matrix of $\hat{\mathbf{r}}$ based on the first stage samples only. We notice that the maximum variance estimates of the $\hat{r}_i$'s is around 3.7×10^6 and the maximum is attained at the skeleton point (140, 3, 1). Next, we change



Web Figure 6 – Standard errors (in log scale) of the Bayes factor estimates.

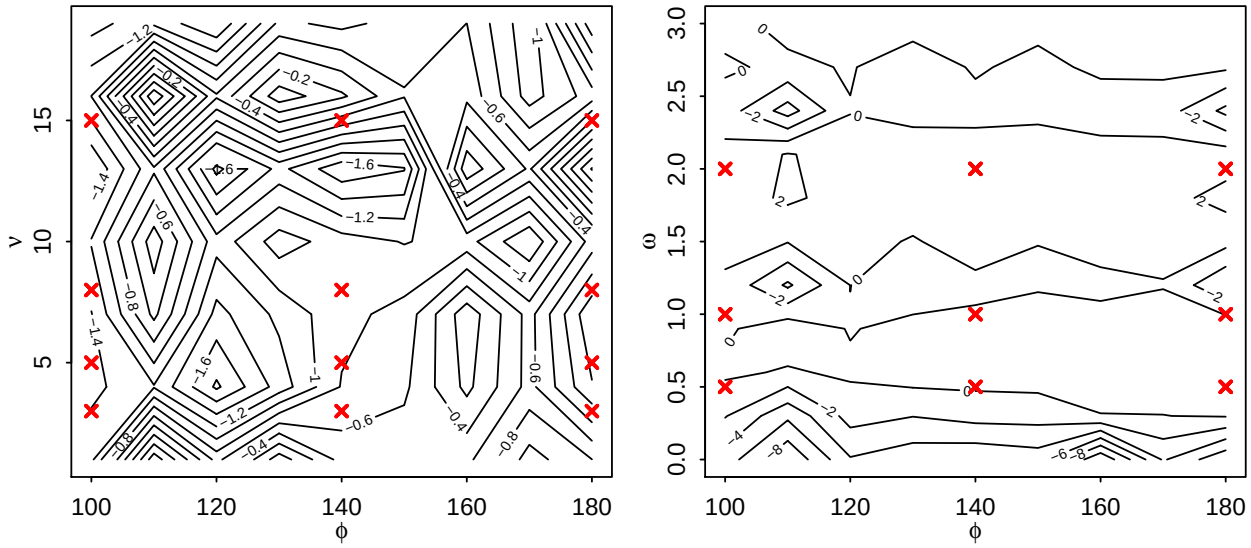
the reference point from $\xi_1 \equiv (\phi_1, \nu_1, \omega_1) = (100, 3, 0)$ to $\xi_1 = (140, 3, 1)$. In this case, the maximum variance estimates of the $\hat{r}_i$'s drastically reduces to 0.45 and the plot the standard errors of $\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}})$ is given in Web Figure 6. Finally, we change the skeleton points to

$$(\phi, \nu, \omega) \in \{100, 140, 180\} \times \{3, 5, 8, 15\} \times \{0.5, 1, 2\}, \tag{3}$$

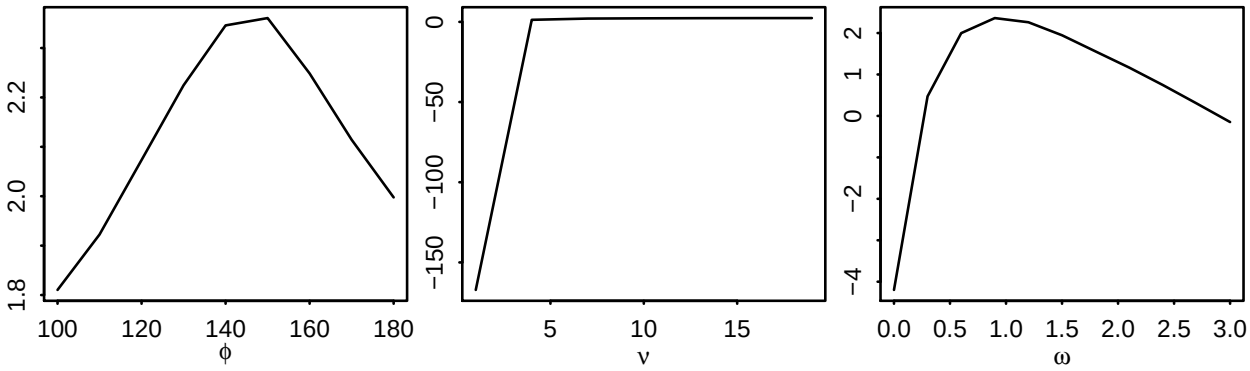
that is, we include $\nu = 5$ in the skeleton set and replace $\omega = 0$ with $\omega = 2$. Web Figure 7 shows the plot of the standard errors of $\hat{B}_{\xi,\xi_1}(\hat{\mathbf{r}})$. We used the same sample size as before, that is, 800 MCMC samples in the first stage and 200 samples in the second stage, but with the set of skeleton points given in (3), the maximum standard error has reduced from about 2.7×10^8 to 2.27. In Web Figure 8 we show the profile of the Bayes factors with respect to each of the parameters, that is, one of the parameters is kept fixed and the maximum of the Bayes factors over the other two parameters is plotted against the values of the first parameter. The maximum is obtained at $\hat{\phi} = 146.22$, $\hat{\nu} = 30.00$, and $\hat{\omega} = 0.95$. Web Figure 8 suggests that probit link can be used for analyzing this data set. Although, the skeleton points given in (2) resulted in large standard errors for the Bayes factor estimates, the EB estimate of the parameters ξ was similar with $(\hat{\phi}, \hat{\nu}, \hat{\omega}) = (146.32, 30.00, 0.97)$. The corresponding profile plot of the Bayes factors estimates is given in Web Figure 9.

Treating these estimates $(\hat{\phi}, \hat{\nu}, \hat{\omega})$ as fixed, we subsequently estimate the other parameters of the model by running a new MCMC chain of length 1,000 collecting every 10th sample after an initial burnin of 1,000 sample. The estimates obtained are $\hat{\beta} = -1.05$ and $\hat{\sigma}^2 = 0.13$ and the 95% prediction intervals of β and σ^2 are given by $(-1.207, -0.895)$ and $(0.092, 0.172)$ respectively. The prediction of the random field and a plot of estimated prediction uncertainty are shown in Web Figure 10. The standard errors estimates for the posterior means of β and σ^2 are obtained using the method of overlapping batch means and equal 0.0003 and 0.0001 respectively.

In comparison with the results of Zhang (2002), who used the logit link function, our prediction has the same pattern and the same range of prediction values. (Note that Web Figure 7 in Zhang (2002) is without the mean.) We also get a similar estimate for the range but our estimate for the partial sill is 0.13 compared to Zhang's 0.18 and the nugget to sill ratio is estimated by our method to be about 1 compared to around 2 in Zhang (2002).



Web Figure 7 – Standard errors (in log scale) of the Bayes factor estimates.



Web Figure 8 – Profile of $B_{(\nu,\phi,\omega),(\nu_1,\phi_1,\omega_1)}$ for the Rhizoctonia data with respect to each parameter corresponding to the skeleton set (3). The plot suggests a probit link, ϕ around 146 and τ^2/σ^2 around 0.95.

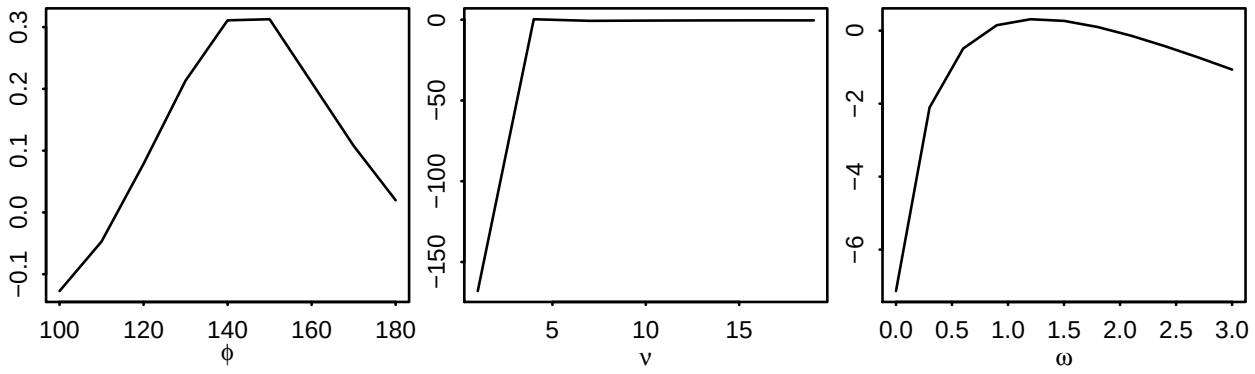
Web Appendix D

R code for analysis of the Rhizoctonia data

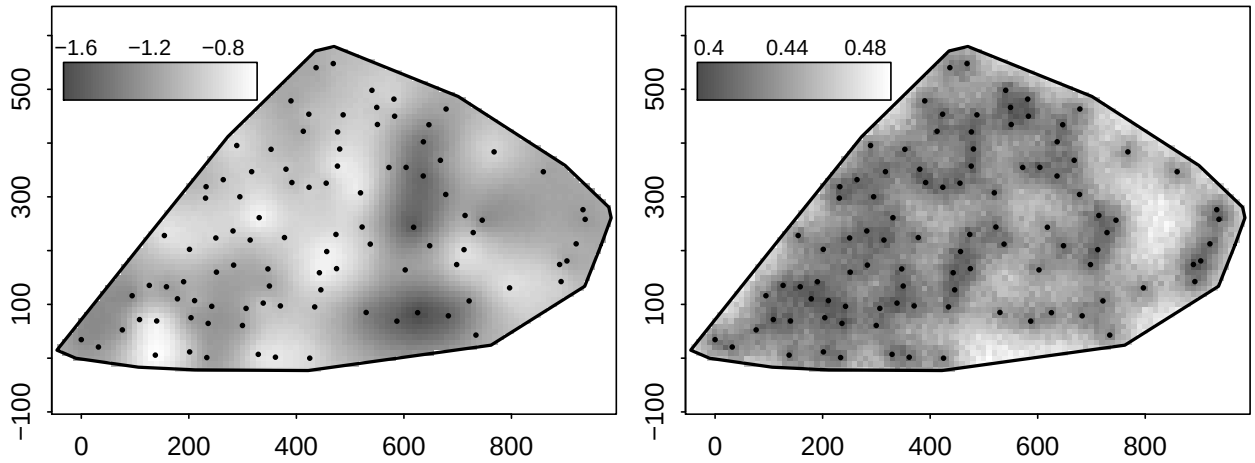
Below we provide the R code for an illustrative analysis of the Rhizoctonia data. More examples can be found in demo(package = "geoBayes").

```
## Load the relevant R package
library(geoBayes)


## Sample size for the MCMC algorithm
Nout <- 1000
Nthin <- 10
Nbi <- 1000
```



Web Figure 9 – Profile of $B_{(\nu,\phi,\omega),(\nu_1,\phi_1,\omega_1)}$ for the Rhizoctonia data with respect to each parameter corresponding to the skeleton set (2).



Web Figure 10 – Estimated means (left plot) and uncertainty defined as the length of 95% prediction interval divided by 4 (right panel) for the predictive distribution of the Gaussian random field for the Rhizoctonia data. The dots in the plots indicate observed locations.

```
## Parameters for the model
kappa <- 0
ssqdf <- 1
ssqsc <- 1
betm0 <- 0
betQ0 <- .01


## Skeleton set
philist <- c(100, 140, 180)
dftlist <- c(3, 5, 8, 15)
nsqlist <- c(.5, 1, 2)
parlist <- expand.grid(phi=philist, dft=dftlist, nsq=nsqlist)
```

```
## Draw MCMC samples corresponding to each parameter combination in the skeleton set
runsa <- list()
for (i in 1:nrow(parlist)) {
  runsa[[i]] <- mcsglmm(Infected ~ 1, "binomial", data = rhizoctonia,
                        Total, , ~ Xcoord + Ycoord,
                        Nout = Nout, Nthin = Nthin, Nbi = Nbi, betm0 = betm0,
                        betQ0 = betQ0, ssqdf = ssqdf, ssqsc = ssqsc,
                        corrfcn = "sph", kappa = kappa, linkp = parlist$dft[i],
                        phisc = 0, omgsc = 0, phistart = parlist$phi[i],
                        omgstart = parlist$nsq[i])
}


## Estimate the Bayes factors at the skeleton points by reverse logistic method
##  and compute the importance weights for each sample (required in the second stage).
bfrl <- bf1skel(runsa, reference = 14)


## Compute contour and plot profile Bayes factors
phivec <- seq(90, 200, len=21)
dftvec <- seq(3, 30, len=21)
nsqvec <- seq(0.5, 2, .25)
bfcontour <- bf2new(bfrl, dftvec, phivec, nsqvec)
plotbf2(bfcontour, c("linkp", "phi", "omg"))


## Estimate nu, phi, omega
paroptim <- list(linkp = c(3, 15, 30), phi = c(90, 150, 200), omg = c(0, .9, 2))
pareb <- bf2optim(bfrl, paroptim)


## Sample the remaining parameters conditioned on the EB estimates
samEB <- mcsglmm(Infected ~ 1, "binomial", data = rhizoctonia,
                 Total, , ~ Xcoord + Ycoord,
                 Nout = Nout, Nthin = Nthin, Nbi = Nbi, betm0 = betm0,
                 betQ0 = betQ0, ssqdf = ssqdf, ssqsc = ssqsc,
                 corrfcn = "sph", kappa = kappa, linkp = pareb$par["linkp"],
                 phisc = 0, omgsc = 0, phistart = pareb$par["phi"],
                 omgstart = pareb$par["omg"])
```

References

Roy, V., Tan, A., and Flegal, J. (2015). Estimating standard errors for importance sampling estimators with multiple markov chains. Technical report, Iowa State University. Preprint available at `http://lib.dr.iastate.edu/stat_las_preprints/34/`.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Bio-*

metrics, 58:129–136.