Statistica Neerlandica (2016) Vol. 70, nr. 1, pp. 47–76
doi:10.1111/stan.12078

Beta spatial linear mixed model with variable dispersion using Monte Carlo maximum likelihood

Oscar O. Melo*

Department of Statistics, Faculty of Sciences, Universidad Nacional de Colombia, Bogotá, Colombia

Carlos E. Melo†

Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

Jorge Mateu‡

Department of Mathematics, Universitat Jaume I, Castellón, Spain

We propose a beta spatial linear mixed model with variable dispersion using Monte Carlo maximum likelihood. The proposed method is useful for those situations where the response variable is a rate or a proportion. An approach to the spatial generalized linear mixed models using the Box–Cox transformation in the precision model is presented. Thus, the parameter optimization process is developed for both the spatial mean model and the spatial variable dispersion model. All the parameters are estimated using Markov chain Monte Carlo maximum likelihood. Statistical inference over the parameters is performed using approximations obtained from the asymptotic normality of the maximum likelihood estimator. Diagnosis and prediction of a new observation are also developed. The method is illustrated with the analysis of one simulated case and two studies: clay and magnesium contents. In the clay study, 147 soil profile observations were taken from the research area of the Tropenbos Cameroon Programme, with explanatory variables: elevation in metres above sea level, agro-ecological zone, reference soil group and land cover type. In the magnesium content, the soil samples were taken from 0- to 20-cm-depth layer at each of the 178 locations, and the response variable is related to the spatial locations, altitude and sub-region.

Keywords and Phrases: clay and magnesium contents, Markov chain Monte Carlo, spatial generalized linear mixed models, spatial mean and spatial precision models, spatial prediction.

*oomelom@unal.edu.co
†cmelo@udistrital.edu.co
‡mateu@mat.uji.es

1 Introduction

The distribution of mineral particles such as clay or magnesium is commonly recorded as basic properties of the soil. The particle size distribution affects many properties of the soil, including its structure, water relations, chemistry, organic carbon dynamics and mechanical properties (Lark and Bishop, 2007). In these studies, we need to predict a fraction of mineral particle in soil, given observations at sample sites and some spatial covariates. Thus, we have a spatial response variable given by a proportion or fraction that is limited to a value in the interval $(0, 1)$ or (l_1, l_2) with $l_1, l_2 > 0$, respectively. This response variable cannot be treated as a realization of unbounded random variable such as a Gaussian process with Gaussian response. For this fact, there are no theoretical reasons to use traditional universal kriging to map particle size distribution when the response variable follows a beta distribution.

The characterization of the spatial variability of soil attributes is essential to achieve a better understanding of complex relations between soil properties and environmental factors (Goovaerts, 1998) and to determine appropriate management practices for soil resources use (Bouma *et al.*, 1999). It also has practical implications in sampling design for ecological, environmental and agricultural studies (Stein and Ettema, 2003). In addition, demands for more accurate information on the spatial distribution of soils have increased with the inclusion of the spatial dependence and scale in ecological models and environmental management systems (Godwin and Miller, 2003).

In this paper, we report two case studies in which we predicted clay and magnesium contents of the soil at test sites. As a first case, we study the data set that contains 147 soil profile observations from the research area of the Tropenbos Cameroon Programme, representative of the humid forest region of south-western Cameroon and adjacent areas of Equatorial Guinea and Gabon (Yemefack, Rossiter and Njomgang, 2005). Specifically, the studied measurements were formed by samples of fixed soil layer at 0–10 cm. The data set comes from two sources: first, 45 representative soil profiles were described and sampled by genetic horizon and, second, 102 plots from various land use/land cover types were sampled at the three fixed depths (see details in Pauwels *et al.*, 1992). For both data sets, samples were located purposively and subjectively to represent soil and land use types. Laboratory analysis was performed by standard local methods (Pauwels *et al.*, 1992). In this study, we want to investigate the relationship between the percentage of clay content and the explanatory variables: elevation in metres above sea level (ELEV), agro-ecological zone (ZONE) and reference soil group (World Reference Base for Soil Resources, WRB). For each observation, the east and north coordinates (UTM zone 32N, WGS84 datum, in metres) were recorded.

In the second case, we study the soil samples that were taken from a 0- to 20-cm-depth layer at each of the 178 locations. The data set has information about the magnesium content, the spatial locations, altitude (ALT) and sub-region (SR) code of each sample, which is associated with three periods of fertilization in different areas. The sampling design forms an incomplete regular lattice at a spacing of approximately 50 m. The data were collected by researchers from Pesquisa Agropecuária (PESAGRO)

and Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)-Solos, Rio de Janeiro, Brazil (Capeche *et al*., 1997; Diggle and Ribeiro, 2007). The main objective of the study was adequate land use planning that allows rational and sustainable management, avoiding the erosion process. Therefore, we should predict magnesium content throughout the study region.

The aforementioned problems can be solved using the beta regression model proposed by Cepeda (2001, pp. 63–64 and 70), Ferrari and Cribari-Neto (2004) and Simas, Barreto-Souza and Rocha (2010), but this method does not consider the spatial correlation among observations. On the other hand, we can employ spatial generalized linear mixed models proposed by Diggle, Tawn and Moyeed (1998), Zhang (2002), Christensen, Diggle and Ribeiro (2001) and Christensen (2004), but these models do not consider beta responses in their fits. Therefore, we propose a methodology that links these two procedures with the goal to give a solution to our problem. Our main innovation here is to incorporate beta spatial dependence directly into the maximum likelihood (ML) estimate (MLE) function. To achieve this, we propose a beta spatial linear mixed model (BSLMM) with variable dispersion using Monte Carlo ML.

We study the effect of each explanatory variable on the response beta variable and develop an approach to the spatial generalized linear mixed models using the Box–Cox transformation. Moreover, the parameter optimization process is developed for both the spatial mean model (SMM) and the spatial variable dispersion model (SVDM). These parameters are estimated using Markov chain Monte Carlo (MCMC) ML, which is a feasible and useful technique and a powerful tool for model selection in BSLMM, exploring other link functions and types of correlation functions. Thus, statistical inference is performed using approximations, which are obtained from the asymptotic normality of the ML estimator, adapting traditional statistics to fit the proposed model. Furthermore, diagnosis and prediction of a new observation are developed.

Section 2 develops the methodological proposal: BSLMM, spectral representation, MCMC ML algorithm for BSLMM, goodness-of-fit measures and spatial prediction of a new subject. Section 3 presents a simulated experiment for BSLMM. Section 4 develops the two applications that illustrate the proposed methodology. Then, the paper ends with some conclusions.

2 A beta spatial linear mixed model

Assume we are interested in any beta response variable that can be associated with a geo-referenced information in each site (such as latitude and longitude) together with a set of binary, categorical and continuous explanatory variables. Let $M = \{M(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ denote a Gaussian random field with $\boldsymbol{s}$ a location in a d-dimensional Euclidean space and with mean function $\mathbb{E}[M(\boldsymbol{s})] = \beta_0 + \boldsymbol{x}^t(\boldsymbol{s})\boldsymbol{\beta}$ and covariance $\mathbb{C}\text{ov}(M(\boldsymbol{s}), M(\boldsymbol{s}')) = \sigma^2\rho(\boldsymbol{s}, \boldsymbol{s}'; \varsigma) + \tau^2 1_{\{\boldsymbol{s}=\boldsymbol{s}'\}}$, where $\beta_0 \in \mathbb{R}$ is a parameter related to the unknown intercept, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown regression parameters, $\boldsymbol{x}(\boldsymbol{s})$'s

are known spatial location-dependent covariates, $\sigma^2 > 0$ is a dispersion parameter, $\rho(s, s', \varsigma)$ is a correlation function in $\mathbb{R}^2$, ς is a correlation parameter and $\tau^2 \geq 0$ is called the nugget effect.

Conditionally on M, the stochastic spatial process, $\{Y(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$, consists of mutually independent random variables. Then, we say that a random variable $\{Y(\boldsymbol{s}) \mid M\}$ follows a beta distribution with parameters $0 < \mu(s) < 1$ and $\phi(\boldsymbol{s}) > 0$, denoted by $\text{Beta}(\mu(\boldsymbol{s})\phi(\boldsymbol{s}), \phi(\boldsymbol{s})(1-\mu(\boldsymbol{s})))$, if the distribution of $\{Y(\boldsymbol{s}) \mid M\}$ admits the following density function with respect to the Lebesgue measure (Ferrari and Cribari-Neto, 2004)

$$f(y(\boldsymbol{s}) \mid \mu(\boldsymbol{s}), \phi(\boldsymbol{s})) = \frac{\Gamma(\phi(\boldsymbol{s}))}{\Gamma[\mu(\boldsymbol{s})\phi(\boldsymbol{s})]\Gamma[(1-\mu(\boldsymbol{s}))\phi(\boldsymbol{s})]} y(\boldsymbol{s})^{[\mu(\boldsymbol{s})\phi(\boldsymbol{s})-1]}[1-y(\boldsymbol{s})]^{[(1-\mu(\boldsymbol{s}))\phi(\boldsymbol{s})-1]} \times I_{(0,1)}(y(\boldsymbol{s})) : \tag{1}$$

where $I(\cdot)$ denotes the indicator function and $\Gamma(\cdot)$ denotes the gamma function. The mean and variance of $\{Y(\boldsymbol{s}) \mid M\}$ are, respectively, $\mathbb{E}(Y(\boldsymbol{s}) \mid M) = \mu(\boldsymbol{s})$ and $\mathbb{V}\text{ar}(Y(\boldsymbol{s}) \mid M) = \mu(\boldsymbol{s})[1-\mu(\boldsymbol{s})]/[1+\phi(\boldsymbol{s})]$, where $\mu(\boldsymbol{s})$ is the mean of the response variable and $\phi(\boldsymbol{s})$ can be interpreted as a precision parameter in the sense that, for a fixed $\mu(\boldsymbol{s})$, a large value of $\phi(\boldsymbol{s})$ leads to a smaller variance of $(Y(\boldsymbol{s}) \mid M)$.

The beta spatial model given in Equation 1 can be appropriate to explain the clay and magnesium contents as a function of the explanatory variables. Note that if the response variable takes values in (l_1, l_2) (with $l_1 < l_2$ known), one can model $(y(\boldsymbol{s}) - l_1)/(l_2 - l_1)$ (Smithson and Verkuilen, 2006). Also, if the response variable takes values close to zero or one, we can employ beta distribution using

$$y^*(\boldsymbol{s}) = (y(\boldsymbol{s}) - \min\{y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n)\}) / (\max\{y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n)\} - \min\{y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n)\})$$

and then using the transformation $(y^*(\boldsymbol{s}) \cdot (n-1) + 0.5)/n$ for the values 0 and 1. However, *ad hoc* re-scalings might provide a nice working solution for small proportions of 0s and 1s, but sensitivity towards parameter estimation can be considerable with higher proportions (Galvis, Badyophadyay and Lachos, 2013).

Consider n distinct locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ and suppose we observe a realization $\boldsymbol{y}_s = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n))^t$ of $\boldsymbol{Y}_s = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^t$, where the conditional density function of $Y(\boldsymbol{s}_i)$ given $M(\boldsymbol{s}_i)$ is given by Equation 1 with mean $\mu(\boldsymbol{s}_i)$ and unknown dispersion $\phi(\boldsymbol{s}_i), i = 1, \ldots, n$.

The proposed interpolation is built for a non-Gaussian random field model by specifically considering categorical, continuous and indicator explanatory variables in the trend model. The data-generating mechanism conditional on the signal of the model follows a classical generalized linear model as described by McCullagh and Nelder (1989). Explicitly, it is assumed that the response variable $Y(\boldsymbol{s}_i), i = 1, 2, \ldots, n$, has a beta distribution, where the SMM and the SVDM are determined as

$$
\begin{aligned}
g_1(\mu(\boldsymbol{s}_i)) = M_1(\boldsymbol{s}_i) &= \sum_{j=0}^{p_1} v_j(\boldsymbol{s}_i)\zeta_j + z_1(\boldsymbol{s}_i) = v^t(\boldsymbol{s}_i)\boldsymbol{\zeta} + z_1(\boldsymbol{s}_i) \\
g_2(\phi(\boldsymbol{s}_i)) = M_2(\boldsymbol{s}_i) &= \sum_{j=0}^{p_2} u_j(\boldsymbol{s}_i)\varphi_j + z_2(\boldsymbol{s}_i) = u^t(\boldsymbol{s}_i)\boldsymbol{\varphi} + z_2(\boldsymbol{s}_i)
\end{aligned}
\tag{2}
$$

where $v_0(\boldsymbol{s}_i) = 1, u_0(\boldsymbol{s}_i) = 1, \boldsymbol{\zeta}^t = (\zeta_0, \zeta_1, \ldots, \zeta_{p_1})$ and $\boldsymbol{\varphi}^t = (\varphi_0, \varphi_1, \ldots, \varphi_{p_2})$ are vectors of unknown trend parameters, with $\boldsymbol{\zeta} \in \mathbb{R}^{p_1+1}$ and $\boldsymbol{\varphi} \in \mathbb{R}^{p_2+1}, p_1 + p_2 < n$. The $z_1(\boldsymbol{s}_i)$ and $z_2(\boldsymbol{s}_i)$ are assumed to have random field structures, whose covariance functions are given by $\mathbb{C}\mathrm{ov}(z_1(\boldsymbol{s}_i), z_1(\boldsymbol{s}_{i'})) = \sigma_1^2\rho_1(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \varsigma_1) + \tau_1^2 1_{\{\boldsymbol{s}_i = \boldsymbol{s}_{i'}\}}$ and $\mathbb{C}\mathrm{ov}(z_2(\boldsymbol{s}_i), z_2(\boldsymbol{s}_{i'})) = \sigma_2^2\rho_2(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \varsigma_2) + \tau_2^2 1_{\{\boldsymbol{s}_i = \boldsymbol{s}_{i'}\}}$, respectively. The $\sigma_j^2 > 0$'s are the dispersion parameters $j = 1, 2$, the $\rho_j(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \varsigma_j)$'s are the correlation functions each one in $\mathbb{R}^2$, the ς_j's are the correlation parameters and the $\tau_j^2 \geq 0$'s are the nugget effects. Additionally, $\boldsymbol{v}^t(\boldsymbol{s}_i) = (v_0(\boldsymbol{s}_i), v_1(\boldsymbol{s}_i), \ldots, v_{p_1}(\boldsymbol{s}_i))$ and $\boldsymbol{u}^t(\boldsymbol{s}_i) = (u_0(\boldsymbol{s}_i), u_1(\boldsymbol{s}_i), \ldots, u_{p_2}(\boldsymbol{s}_i))$ are vectors of observations of the $p_1 + 1$ and $p_2 + 1$ known covariates, respectively. The models (2) are called the SMM and the SVDM of $Y(\boldsymbol{s}_i)$, respectively.

The link functions $g_1 : (0, 1) \to \mathbb{R}$ and $g_2 : (0, \infty) \to \mathbb{R}$ are strictly monotonic and twice differentiable. A rich discussion of the link functions for $g_1(\mu(\boldsymbol{s}_i))$ is presented by Atkinson (1985), McCullagh and Nelder (1989) and Ferrari and Cribari-Neto (2004) and for $g_2(\phi(\boldsymbol{s}_i))$ by Simas *et al.* (2010).

In matrix form, the BSLMMs in Equation 2 can be expressed as

$$
g_1(\boldsymbol{\mu}_s) = \boldsymbol{M}_1 = \boldsymbol{V}_s\boldsymbol{\zeta} + \boldsymbol{E}_1\boldsymbol{z}_1 \qquad\qquad g_2(\boldsymbol{\phi}_s) = \boldsymbol{M}_2 = \boldsymbol{U}_s\boldsymbol{\varphi} + \boldsymbol{E}_2\boldsymbol{z}_2 \tag{3}
$$

where $\boldsymbol{M}_1 = (M_1(\boldsymbol{s}_1), \ldots, M_1(\boldsymbol{s}_n))^t$, $\boldsymbol{M}_2 = (M_2(\boldsymbol{s}_1), \ldots, M_2(\boldsymbol{s}_n))^t$, $\boldsymbol{V}_s = (\mathbf{1}, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_{p_1})$, $\boldsymbol{U}_s = (\mathbf{1}, \boldsymbol{U}_1, \ldots, \boldsymbol{U}_{p_2})$, $\boldsymbol{\mu}_s = \mathbb{E}(\boldsymbol{y}_s \mid \boldsymbol{V}_s, \boldsymbol{z}_1) = (\mu(\boldsymbol{s}_1), \ldots, \mu(\boldsymbol{s}_n))^t$, $\boldsymbol{\phi}_s = (\phi(\boldsymbol{s}_1), \ldots, \phi(\boldsymbol{s}_n))^t$, $\mathbf{1}$ is a vector of ones of size $n \times 1$ and $\boldsymbol{V}_i\,(i = 1, \ldots, p_1)$ and $\boldsymbol{U}_j\,(j = 1, \ldots, p_2)$ are vectors of spatial explanatory variables associated with the SMM and SVDM, respectively. Moreover, both $\boldsymbol{V}_s$ and $\boldsymbol{U}_s$ may involve continuous, categorical and binary variables, or even a mixture of them. Furthermore, $\boldsymbol{z}_1 = (z_1(\boldsymbol{s}_1), \ldots, z_1(\boldsymbol{s}_n))^t$, $\boldsymbol{z}_2 = (z_2(\boldsymbol{s}_1), \ldots, z_2(\boldsymbol{s}_n))^t$ and $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are non-random design matrices that are compatible with the random effects $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, respectively.

For SVDM, all link functions for $g_2(\phi(\boldsymbol{s}_i))$ given by Simas *et al.* (2010) can be considered in a general Box–Cox class of link functions, which has the following form

$$
g_{\nu_2}\left(\phi\left(\boldsymbol{s}_i\right)\right) = \begin{cases} \left(\phi^{\nu_2}\left(\boldsymbol{s}_i\right)\right)/\nu_2 & \text{if} \quad \nu_2 > 0 \\ \log\left(\phi\left(\boldsymbol{s}_i\right)\right) & \text{if} \quad \nu_2 = 0 \end{cases}
$$

As the link function $g_{\nu_2}(\phi(\boldsymbol{s}_i))$ relates $\boldsymbol{\phi}_s$ and $\boldsymbol{M}_2$, we have

$$
\boldsymbol{\phi}_s = m_2 g_{\nu_2}^{-1}\left(\boldsymbol{M}_2\right) \tag{4}
$$

where m_2 is a deterministic function. If $\nu_2 > 0$, then $g_{\nu_2}(\mathbb{R}) = (-1/\nu_2, \infty)$, and it is necessary to define $g_{\nu_2}^{-1}(\phi(\boldsymbol{s}_i)) = 0$ when $\phi(\boldsymbol{s}_i) \notin g_{\nu_2}(\mathbb{R})$. Therefore, we write $f(y(\boldsymbol{s}_i) \mid \mu(\boldsymbol{s}_i), \phi(\boldsymbol{s}_i)) = 1_{\{y(\boldsymbol{s}_i)=0\}}$ when $\phi(\boldsymbol{s}_i) = 0$.

The models in Equation 3 are not a simple reformat because they accommodate more complex data structure beyond spatial data. For example, with properly defined $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ and random effects $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ in both models, respectively, they encompass non-normal clustered data and crossed factor data (Breslow and Clayton, 1993). When $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ are defined as matrices indicating membership of spatial regions (e.g. counties or census tracts), Equation 3 can model areal data as well.

The basis of many procedures, including the one outlined earlier, is a multivariate normal (MN) distribution function. In this case, consider the spatial processes $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, and assume that they have MN distributions, that is, $\boldsymbol{z}_1 \sim MN(\boldsymbol{0}, \boldsymbol{\Sigma}_{z_1})$ and $\boldsymbol{z}_2 \sim MN(\boldsymbol{0}, \boldsymbol{\Sigma}_{z_2})$, respectively. The key to efficient spatial modelling in large-sized problems is parametrization of these MNs in an efficient manner. Rewrite $\boldsymbol{z}_j$, $j = 1, 2$, in terms of a spectral decomposition (see details in Appendix A),

$$\boldsymbol{z}_j = \boldsymbol{\Psi}_j \boldsymbol{\delta}_j, \qquad j = 1, 2 \tag{5}$$

where $\boldsymbol{\Psi}_j = (\boldsymbol{\psi}_{j1}, \ldots, \boldsymbol{\psi}_{jn})$ is a matrix of size $n \times n$ with $\boldsymbol{\psi}_{ji} = (\psi_{ji}(\boldsymbol{s}_1), \ldots, \psi_{ji}(\boldsymbol{s}_n))^t$ ($j = 1, 2$ and $i = 1, \ldots, n$) and $\boldsymbol{\delta}_j$ is a $n \times 1$ vector of spectral coefficients (projections of the process $\boldsymbol{z}_j$ onto the basis functions), with distribution $\boldsymbol{\delta}_j \sim MN(\boldsymbol{0}, \boldsymbol{\Sigma}_{\delta_j})$. Note that if the basis functions are orthogonal, then $\boldsymbol{\delta}_j = \boldsymbol{\Psi}_j^t \boldsymbol{z}_j$ and $\boldsymbol{\Sigma}_{\delta_j} = \boldsymbol{\Psi}_j^t \boldsymbol{\Sigma}_{z_j} \boldsymbol{\Psi}_j$ because $\boldsymbol{\Psi}_j^t \boldsymbol{\Psi}_j = \boldsymbol{\Psi}_j \boldsymbol{\Psi}_j^t = \boldsymbol{I}$.

There are several advantages in writing the spatial process $\boldsymbol{z}_j$ in terms of the spectral process $\boldsymbol{\delta}_j$. First, the spectral operator often acts as a decorrelator so that $\boldsymbol{\Sigma}_{\delta_j}$ is relatively sparse, as compared with $\boldsymbol{\Sigma}_{z_j}$. Other benefits are its efficient computation and that, in some cases to achieve dimension reduction, the basis formulation can be truncated (Royle and Wikle, 2005).

Finally, substituting Equation 5 into Equation 3 using Box–Cox transformation, we have the following BSLMM:

$$\begin{aligned} \boldsymbol{M}_1 = g_1(\boldsymbol{\mu}_s) &= \mathbf{V}_s \boldsymbol{\zeta} + \boldsymbol{E}_1 \boldsymbol{\Psi}_1 \boldsymbol{\delta}_1 \quad & \boldsymbol{M}_2 = g_{\nu_2}(\boldsymbol{\phi}_s) &= \mathbf{U}_s \boldsymbol{\varphi} + \boldsymbol{E}_2 \boldsymbol{\Psi}_2 \boldsymbol{\delta}_2 \\ &= \mathbf{V}_s \boldsymbol{\zeta} + \boldsymbol{E}_1^* \boldsymbol{\delta}_1 & &= \mathbf{U}_s \boldsymbol{\varphi} + \boldsymbol{E}_2^* \boldsymbol{\delta}_2 \end{aligned} \tag{6}$$

where $\boldsymbol{E}_j^* = \boldsymbol{E}_j \boldsymbol{\Psi}_j, j = 1, 2$.

2.1 A Monte Carlo maximum likelihood algorithm for beta spatial linear mixed model

The application of likelihood-based methods to non-Gaussian spatial generalized linear mixed models based on a beta response is hampered by computational difficulties, which arise because of the high dimensionality of the unobserved random vectors, $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$. This subsection considers MCMC ML (Geyer and Thompson, 1992; Geyer, 1994; Højbjerre, 2003; Christensen, 2004) for BSLMM.

As previously discussed, consider n distinct locations $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ and suppose that we observe a realization $\boldsymbol{y}_s = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_2))^t$ of $\boldsymbol{Y}_s = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))^t$. So, $\boldsymbol{M}_1$ follows a multi-normal distribution with mean vector $\boldsymbol{V}_s \boldsymbol{\zeta}$ and covariance matrix $\sigma_1^2 R(\vartheta_1) + \tau_1^2 \boldsymbol{I}_n$, where $R(\vartheta_1)$ is the correlation matrix with entries $R_{ij}(\vartheta_1) = \rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \vartheta_1)$. Similarly, $\boldsymbol{M}_2 \sim MN(\boldsymbol{U}_s \boldsymbol{\varphi}; \sigma_2^2 R(\vartheta_2) + \tau_2^2 \boldsymbol{I}_n)$.

By independence of $Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)$ given $\boldsymbol{M}_1, \boldsymbol{M}_2$ and Equation 4, the conditional density function of $\boldsymbol{Y}_s$ given $\boldsymbol{M}_1 = \boldsymbol{\eta}_1$ and $\boldsymbol{M}_2 = \boldsymbol{\eta}_2$ is given by

$$f\left(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \nu_2\right) = \prod_{i=1}^{n} f\left\{y\left(\boldsymbol{s}_i\right) \mid g_1^{-1}\left[\eta_1\left(\boldsymbol{s}_i\right)\right], m_{i2} g_{\nu_2}^{-1}\left[\eta_2\left(\boldsymbol{s}_i\right)\right]\right\}$$

where $m_{i2} = m_2(\boldsymbol{s}_i)$ and ν_2 determines the link function for SVDM.

Now, from a classical perspective, the likelihood function based on the observed random variables $\boldsymbol{y}_s$ is obtained by marginalizing with respect to the unobserved random variables $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$, leading to the mixed-model likelihood. Then, the likelihood function for a BSLMM is not expressible in closed form, but only as a high-dimensional double integral

$$\begin{aligned} L\left(b, \nu_2\right) &= f\left(\boldsymbol{y}_s \mid b, \nu_2\right) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f\left(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \nu_2\right) f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b\right) d\boldsymbol{\eta}_1 d\boldsymbol{\eta}_2 \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \prod_{i=1}^{n} f\left(y\left(\boldsymbol{s}_i\right) \mid g_1^{-1}\left[\eta_1\left(\boldsymbol{s}_i\right)\right], m_{i2} g_{\nu_2}^{-1}\left[\eta_2\left(\boldsymbol{s}_i\right)\right]\right) f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b\right) \\ &\quad d\eta_1\left(\boldsymbol{s}_1\right) \cdots d\eta_1\left(\boldsymbol{s}_n\right) d\eta_2\left(\boldsymbol{s}_1\right) \cdots d\eta_2\left(\boldsymbol{s}_n\right) \end{aligned} \tag{7}$$

where $b = (\boldsymbol{\zeta}, \boldsymbol{\varphi}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, $f(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b)$ denotes the joint distribution of $(\boldsymbol{M}_1, \boldsymbol{M}_2)$ given the observed covariates $\mathbf{V}_s$ and $\mathbf{U}_s$, with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ as the covariance parameter vectors associated with $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$, respectively. The preceding double integral defines also the normalizing constant in the conditional density function of $(\boldsymbol{M}_1, \boldsymbol{M}_2)$ given $\boldsymbol{Y}_s = \boldsymbol{y}_s$,

$$f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \boldsymbol{y}_s, b, \nu_2\right) \propto \prod_{i=1}^{n} f\left(y\left(\boldsymbol{s}_i\right) \mid g_1^{-1}\left[\eta_1\left(\boldsymbol{s}_i\right)\right], m_{i2} g_{\nu_2}^{-1}\left[\eta_2\left(\boldsymbol{s}_i\right)\right]\right) f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b\right) \tag{8}$$

MCMC provides a method to simulate from Equation 8 and to approximate Equation 7.

The two integrals have a high dimension, and consequently, they are intractable to find the MLEs by direct maximization. Then, we consider the case where ν_2 is fixed, and therefore, it is removed. The likelihood function (Equation 7) can thus be written as

$$\begin{aligned} L(b) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f\left(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right) f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b\right) d\boldsymbol{\eta}_1 d\boldsymbol{\eta}_2 \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{f\left(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right) f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b\right)}{\tilde{f}\left(\boldsymbol{y}_s, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right)} \tilde{f}\left(\boldsymbol{y}_s, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right) d\boldsymbol{\eta}_1 d\boldsymbol{\eta}_2 \\ &\propto \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{f\left(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right) f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b\right)}{f\left(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right) \tilde{f}\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right)} \tilde{f}\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \boldsymbol{y}_s\right) d\boldsymbol{\eta}_1 d\boldsymbol{\eta}_2 \\ &= \widetilde{\mathbb{E}}\left[\frac{f\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \mathbf{V}_s, \mathbf{U}_s; b\right)}{\tilde{f}\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2\right)} \middle| \boldsymbol{y}_s\right] \end{aligned} \tag{9}$$

where $\tilde{f}(\boldsymbol{y}_s, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2) = f(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)\tilde{f}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ and $\tilde{f}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ is some density function with support on $\mathbb{R}^n \times \mathbb{R}^n$, the conditional density function $\tilde{f}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \mid \boldsymbol{y}_s) \propto f(\boldsymbol{y}_s \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)\tilde{f}(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$, and $\widetilde{\mathbb{E}}(\cdot \mid \boldsymbol{y}_s)$ denotes expectation with respect to $\tilde{f}(\cdot \mid \boldsymbol{y}_s)$ and depends on an initial guess of b. MLEs can be calculated by maximizing the Monte Carlo approximation to Equation 9,

$$L_r(b) = \frac{1}{r}\sum_{k=1}^{r} \frac{f\left(\boldsymbol{\eta}_1(k), \boldsymbol{\eta}_2(k) \mid \mathbf{V}_s, \mathbf{U}_s; b\right)}{\tilde{f}\left(\boldsymbol{\eta}_1(k), \boldsymbol{\eta}_2(k)\right)} \tag{10}$$

where the $\boldsymbol{\eta}_1(k)$ and $\boldsymbol{\eta}_2(k)$ $(k = 1, \ldots, r)$ are sampled by MCMC from the distribution function $\tilde{f}(\cdot \mid \boldsymbol{y}_s)$. As noted from Equation 9, one should choose $\tilde{f}(\cdot)$ close to $f(\cdot \mid \hat{b})$, where $\hat{b}$ is the MLE of b, because otherwise one or very few of the terms $f(\boldsymbol{\eta}_1(k), \boldsymbol{\eta}_2(k) \mid \mathbf{V}_s, \mathbf{U}_s; b)/\tilde{f}(\boldsymbol{\eta}_1(k), \boldsymbol{\eta}_2(k))$, $k = 1, \ldots, r$, may dominate the others in $L_r(b)$, which makes the approximation less useful.

Now, a numerical procedure for maximizing the Monte Carlo approximation (Equation 10) is presented. Let $\boldsymbol{\theta}_j = (\sigma_j^2, \Delta_j)$ for $j = 1, 2$ and then $b = (\boldsymbol{\zeta}, \boldsymbol{\varphi}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Delta})$, where $\boldsymbol{\Delta} = (\Delta_1, \Delta_2)$, and each Δ_j denotes the correlation function with values of ϑ_j and τ_j^2, $j = 1, 2$. The maximization of L_r with respect to $\boldsymbol{\zeta}$, $\boldsymbol{\varphi}$, σ_1^2 and σ_2^2 given $\boldsymbol{\Delta}$ is fairly straightforward, because the first-order and second-order derivatives of the normal densities $f(\boldsymbol{\eta}_1(k), \boldsymbol{\eta}_2(k) \mid \mathbf{V}_s, \mathbf{U}_s; b)$, $k = 1, \ldots, r$, with respect to these parameters are simple, which makes an iterative Newton–Raphson procedure feasible and computationally fast. For such an iterative procedure, suitable initial values are

$$\begin{pmatrix} \boldsymbol{\zeta}(k) \\ \boldsymbol{\varphi}(k) \end{pmatrix} = \left[\begin{pmatrix} \mathbf{V}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_s \end{pmatrix}^t \boldsymbol{C}_{M_1,M_2}^{-1} \begin{pmatrix} \mathbf{V}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_s \end{pmatrix}\right]^{-1} \begin{pmatrix} \mathbf{V}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_s \end{pmatrix}^t \boldsymbol{C}_{M_1,M_2}^{-1} \begin{pmatrix} \boldsymbol{\eta}_1(k) \\ \boldsymbol{\eta}_2(k) \end{pmatrix}$$

$$\sigma_1^2(k) = \frac{1}{n}\left[\boldsymbol{\eta}_1(k) - \mathbf{V}_s\boldsymbol{\zeta}(k)\right]^t \left[\boldsymbol{R}\left(\vartheta_1\right) + \tau_{R_1}^2\mathbf{I}_n\right]^{-1} \left[\boldsymbol{\eta}_1(k) - \mathbf{V}_s\boldsymbol{\zeta}(k)\right]$$

$$\sigma_2^2(k) = \frac{1}{n}\left[\boldsymbol{\eta}_2(k) - \mathbf{U}_s\boldsymbol{\varphi}(k)\right]^t \left[\boldsymbol{R}\left(\vartheta_2\right) + \tau_{R_2}^2\mathbf{I}_n\right]^{-1} \left[\boldsymbol{\eta}_2(k) - \mathbf{U}_s\boldsymbol{\varphi}(k)\right]$$

$k = 1, \ldots, r$, corresponding to the MLEs for the normal densities $f(\boldsymbol{\eta}_1(k), \boldsymbol{\eta}_2(k) \mid \mathbf{V}_s, \mathbf{U}_s; b)$ and where

$$\boldsymbol{C}_{M_1,M_2} = \begin{pmatrix} \sigma_1^2\left(\boldsymbol{R}\left(\vartheta_1\right) + \tau_{R_1}^2\mathbf{I}_n\right) & \mathbb{C}\mathrm{ov}\left(\boldsymbol{M}_1, \boldsymbol{M}_2\right) \\ \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{M}_1, \boldsymbol{M}_2\right) & \sigma_2^2\left(\boldsymbol{R}\left(\vartheta_2\right) + \tau_{R_2}^2\mathbf{I}_n\right) \end{pmatrix}$$

where $\sigma_j^2(\boldsymbol{R}(\vartheta_j) + \tau_{R_j}^2\mathbf{I}_n)$ is the covariance matrix of $\boldsymbol{\eta}_j$ $(j = 1, 2)$ and $\tau_{R_j}^2 = \tau_j^2/\sigma_j^2$ is a relative nugget instead of τ_j^2 for $j = 1, 2$. Additionally, $\mathbb{C}\mathrm{ov}(\boldsymbol{M}_1, \boldsymbol{M}_2)$ is the cross-covariance between $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$.

The values of $\boldsymbol{\zeta}$, $\boldsymbol{\varphi}$, σ_1^2 and σ_2^2 that maximize $L_r(b)$ for a fixed value of $\boldsymbol{\Delta}$, $\hat{\boldsymbol{\zeta}}(\boldsymbol{\Delta})$, $\hat{\boldsymbol{\varphi}}(\boldsymbol{\Delta})$, $\hat{\sigma}_1^2(\boldsymbol{\Delta})$ and $\hat{\sigma}_2^2(\boldsymbol{\Delta})$ are plugged into L_r, and we obtain $\tilde{L}_r(\boldsymbol{\Delta}) = L_r(\hat{\boldsymbol{\zeta}}(\boldsymbol{\Delta}), \hat{\boldsymbol{\varphi}}(\boldsymbol{\Delta}), \hat{\sigma}_1^2(\boldsymbol{\Delta}), \hat{\sigma}_2^2(\boldsymbol{\Delta}))$. This function is maximized with respect to ϑ_1, ϑ_2, $\tau_{R_1}^2$ and $\tau_{R_2}^2$ for a given correlation function using numerical optimization. The parameters

ϑ_1, ϑ_2, $\tau^2_{R_1}$ and $\tau^2_{R_2}$ enter $\tilde{L}_r$ via the matrix $\boldsymbol{C}_{M_1,M_2}$, and because the inversion of this matrix is computationally demanding, the maximization would be relatively slow. The maximization can also be sensitive to starting values in this process because the approximation L_r can be multimodal. The result should be investigated carefully by considering a variety of starting values.

In practice, for $k = 1, \ldots, r$, the whole n-dimensional vectors $\boldsymbol{\eta}_1(j)$ and $\boldsymbol{\eta}_2(j)$ must be stored, because there exist no lower-dimensional sufficient statistics for the parameters ϑ_1, ϑ_2, $\tau^2_{R_1}$ and $\tau^2_{R_2}$. This has the consequence that only a relatively small sample can be stored because of limited storage capacity.Therefore, it is advantageous to make an intensive thinning of the simulated sample such that $\boldsymbol{\eta}_1(1), \ldots, \boldsymbol{\eta}_1(r)$ and $\boldsymbol{\eta}_2(1), \ldots, \boldsymbol{\eta}_2(r)$ are approximately uncorrelated for SMM and SVDM, respectively, storing for instance only every 20th simulation.

On the other hand, when interest is on investigating which link functions are suitable in the SVDM, it is useful to integrate with respect to $\boldsymbol{\phi}_s = (\phi(\boldsymbol{s}_1), \ldots, \phi(\boldsymbol{s}_n))^t$, where $\phi(\boldsymbol{s}_i) = m_{i2} g^{-1}_{\nu_2}(\eta_2(\boldsymbol{s}_i))$, $i = 1, \ldots, n$. The determinant of the Jacobian for this transformation is

$$J_{\nu_2}\left(\boldsymbol{\phi}_s\right) = \prod_{i=1}^{n} \frac{g'_{\nu_2}\left(\phi\left(\boldsymbol{s}_i\right)/m_{i2}\right)}{m_{i2}}$$

Assuming that the link function satisfies $g_{\nu_2}(\mathbb{R}) = \mathbb{R}$ for every ν_2 of interest and defining

$$f_{\nu_2}\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s \mid \boldsymbol{V}_s, \boldsymbol{U}_s, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, b\right) = J_{\nu_2}\left(\boldsymbol{\phi}_s\right) f\left(g_1^{-1}\left(\boldsymbol{\eta}_1\right), g_{\nu_2}^{-1}\left(\boldsymbol{\phi}_s\right) \mid \boldsymbol{V}_s, \boldsymbol{U}_s, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, b\right)$$

we obtain

$$\begin{aligned} L\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s, \nu_2\right) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f\left(\boldsymbol{y}_s \mid \boldsymbol{\mu}_s, \boldsymbol{\phi}_s\right) f_{\nu_2}\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s \mid \boldsymbol{V}_s, \boldsymbol{U}_s, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, b\right) d\boldsymbol{\mu}_s d\boldsymbol{\phi}_s \\ &\propto \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \frac{f\left(\boldsymbol{y}_s \mid \boldsymbol{\mu}_s, \boldsymbol{\phi}_s\right) f_{\nu_2}\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s \mid \boldsymbol{V}_s, \boldsymbol{U}_s, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, b\right)}{f\left(\boldsymbol{y}_s \mid \boldsymbol{\mu}_s, \boldsymbol{\phi}_s\right) \tilde{f}_{(\nu_2)_0}\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s\right)} \tilde{f}_{(\nu_2)_0}\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s \mid \boldsymbol{y}_s\right) d\boldsymbol{\mu}_s d\boldsymbol{\phi}_s \\ &= \widetilde{\mathbb{E}}\left[\frac{f_{\nu_2}\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s \mid \boldsymbol{V}_s, \boldsymbol{U}_s, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, b\right)}{\tilde{f}_{(\nu_2)_0}\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s\right)} \middle| \boldsymbol{y}_s\right] \end{aligned} \tag{11}$$

where $\tilde{f}_{(\nu_2)_0}(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s) = J_{(\nu_2)_0}(\boldsymbol{\phi}_s)\tilde{f}(g_1^{-1}(\boldsymbol{\eta}_1), g_{\nu_2}^{-1}(\boldsymbol{\phi}_s))$, the conditional density function $\tilde{f}_{(\nu_2)_0}(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s \mid \boldsymbol{y}_s) \propto f(\boldsymbol{y}_s \mid \boldsymbol{\mu}_s, \boldsymbol{\phi}_s)\tilde{f}_{(\nu_2)_0}(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s)$, and $\widetilde{\mathbb{E}}(\cdot \mid \boldsymbol{y}_s)$ denotes expectation with respect to $\tilde{f}_{(\nu_2)_0}(\cdot \mid \boldsymbol{y}_s)$. The MLE can be calculated by maximizing the Monte Carlo approximation in Equation 11 as

$$L_r\left(\boldsymbol{\mu}_s, \boldsymbol{\phi}_s, \nu_2\right) = \frac{1}{r}\sum_{k=1}^{r} \frac{f_{\nu_2}\left(\boldsymbol{\mu}_s(k), \boldsymbol{\phi}_s(k) \mid \boldsymbol{V}_s, \boldsymbol{U}_s, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, b\right)}{\tilde{f}_{(\nu_2)_0}\left(\boldsymbol{\mu}_s(k), \boldsymbol{\phi}_s(k)\right)}$$

where the $\boldsymbol{\mu}_s(k)$ and $\boldsymbol{\phi}_s(k)$ $(k = 1, \ldots, r)$ are sampled by MCMC from the distribution function $\tilde{f}(\cdot \mid \boldsymbol{y}_s)$. If $g_{\nu_2}(\mathbb{R}) \neq \mathbb{R}$, then Equation 11 may still hold.

Having estimated the trend and the spatial correlation parameters, b, we are ready to discuss the goodness-of-fit spatial measures.

2.2 Goodness-of-fit measures

After fitting a BSLMM, it is important to carry out a diagnostic analysis to verify the goodness of fit of the estimated model. A global measure of explained variation is obtained by computing the pseudo-R_s^2 defined as

$$R_s^2 = \frac{l\left(\widetilde{b}\right) - l\left(\widehat{b}\right)}{l\left(\widetilde{b}\right)}, \qquad 0 \leq R_s^2 \leq 1$$

where $l(\widetilde{b})$ is the log-likelihood function for the saturated model evaluated at $\widetilde{b}$ and $l(\widehat{b})$ denotes the maximum value of the log-likelihood function for the model of interest using MCMC. Note that $l(\widetilde{b})$ will be larger than any other likelihood function for these observations, assuming the same distribution function and link function, because it provides the most complete description of the data.

The discrepancy of the fitted model can be determined through how well the fitted model is significantly different from the saturated model, which contains as many parameters as observations are in the model. For this, let

$$D\left(\boldsymbol{y}_s, b\right) = 2\left[l\left(\widehat{b}\right) - l\left(\widetilde{b}\right)\right]$$

An approximation to this quantity can be given by

$$D\left(\boldsymbol{y}_s, b\right) = \sum_{i=1}^{n} \left(r\left(\boldsymbol{s}_i\right)\right)^2$$

which is known as the *pseudo-deviance* and

$$r_i = r\left(\boldsymbol{s}_i\right) = \text{sign}\left[y\left(\boldsymbol{s}_i\right) - \hat{y}\left(\boldsymbol{s}_i\right)\right]\left\{2\left[l\left(y\left(\boldsymbol{s}_i\right), \widehat{b}\right) - l\left(y\left(\boldsymbol{s}_i\right), \widetilde{b}\right)\right]\right\}^{1/2}$$

where $l(y(\boldsymbol{s}_i), \widetilde{b})$ is the maximum log-likelihood for the saturated model associated with the ith observation and $l(y(\boldsymbol{s}_i), \widehat{b})$ is the maximum value of the log-likelihood function for the model of interest associated with the ith observation. $r(\boldsymbol{s}_i)$ is the ith deviance residual because an observation with a large absolute value of $r(\boldsymbol{s}_i)$ can be viewed as discrepant. As expected, the log-likelihood associated with the saturated model must be greater than a model with $p_1 < n$ parameters.

Having fitted the model, we are ready to discuss spatial techniques to predict the value of a random field at a given location from nearby observations.

2.3 Spatial prediction of a new subject

Specifically, we used the kriging method, which is about interpolating the vector of values $\boldsymbol{y}_0 = (y(\boldsymbol{s}_{n+1}), \ldots, y(\boldsymbol{s}_{n+n'}))^t$ of a random field vector $\boldsymbol{Y}_0$ at n' prespecified locations from observations $y(\boldsymbol{s}_i)$, $i = 1, \ldots, n$.

We focus on interpolation of random effects over a continuous spatial area for beta observations. Now, we can consider prediction for the SMM of $\boldsymbol{\eta}_1^0 = (\eta_1(\boldsymbol{s}_{n+1}), \ldots, \eta_1(\boldsymbol{s}_{n+n'}))^t$ and the SVDM of $\boldsymbol{\eta}_2^0 = (\eta_2(\boldsymbol{s}_{n+1}), \ldots, \eta_2(\boldsymbol{s}_{n+n'}))^t$. In both models, let $f(\boldsymbol{\eta}_j^0, \boldsymbol{\eta}_j)$ be the joint density function of $\boldsymbol{\eta}_j$ and a vector $\boldsymbol{\eta}_j^0, j = 1, 2$. If we confine our interest to pseudo-unbiased linear predictors of the form

$$\tilde{\boldsymbol{\eta}}_j = \boldsymbol{h}_j + \boldsymbol{Q}_j \boldsymbol{\eta}_j, \qquad j = 1, 2 \tag{12}$$

for some conformable vector $\boldsymbol{h}_j$ and matrix $\boldsymbol{Q}_j$ (McCulloch, Searle and Neuhaus, 2008) and where $\tilde{\boldsymbol{\eta}}_j$ is the predictor of $\boldsymbol{\eta}_j^0$, minimizing the mean squared error of the prediction, we find the pseudo best linear unbiased predictor, which is given by (see details in Appendix B)

$$\tilde{\boldsymbol{\eta}}_j = \mathbb{E}\left(\boldsymbol{\eta}_j^0\right) + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)\left[\mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_j\right)\right]^{-1}\left[\boldsymbol{\eta}_j - \mathbb{E}\left(\boldsymbol{\eta}_j\right)\right], \qquad j = 1, 2$$

The covariance matrix for the prediction has the following general form (see details in Appendix B):

$$\mathbb{V}\mathrm{ar}\left(\tilde{\boldsymbol{\eta}}_j \mid \boldsymbol{y}_s\right) \approx \boldsymbol{\Sigma}_j + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)\left[\sigma_j^2\left(\boldsymbol{R}(\vartheta_j) + \tau_{R_j}^2 \boldsymbol{I}_n\right)\right]^{-1}\left[\widetilde{\mathbb{V}\mathrm{ar}}_r(\boldsymbol{\eta}_j \mid \boldsymbol{y}_s)\right]\left[\sigma_j^2\left(\boldsymbol{R}\left(\vartheta_j\right) + \tau_{R_j}^2 \boldsymbol{I}_n\right)\right]^{-1}\mathbb{C}\mathrm{ov}\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)$$

for $j = 1, 2$.

3 A simulated experiment

We present a simulation study to discuss the choice of initial values and determination of the Monte Carlo sample size. We simulated 100 points in a $(0, 1) \times (0, 1)$ regular spatial grid using two random explanatory variables: the first is a continuous variable sampled from a normal distribution with mean 1 and a standard deviation of 4, and the second is a categorical variable sampled from a multinomial distribution assuming three values with probabilities 0.45, 0.25 and 0.3. Because there are three levels in the categorical variable, only two dummy variables (V_2 and V_3) are considered to avoid problems of singularity. For each of the explanatory variables, 100 independent observations were generated. Conditional on $z_1(\boldsymbol{s}_i)$ and $z_2(\boldsymbol{s}_i)$ stationary Gaussian processes, the response variable Y was simulated from a beta distribution with $p(\boldsymbol{s}_i) = \mu(\boldsymbol{s}_i)\phi(\boldsymbol{s}_i)$ and $q(\boldsymbol{s}_i) = \phi(\boldsymbol{s}_i)(1 - \mu(\boldsymbol{s}_i))$, where the SMM and the SVDM are given by

$$\mu\left(\boldsymbol{s}_i\right) = \frac{1}{1 + \exp\left\{-\left[1.3 + 0.01 v_1(\boldsymbol{s}_i) + 0.4 v_2\left(\boldsymbol{s}_i\right) - 0.7 v_3\left(\boldsymbol{s}_i\right) + 1.6 w_1\left(\boldsymbol{s}_i\right) - 1.7 w_2\left(\boldsymbol{s}_i\right) + z_1(\boldsymbol{s}_i)\right]\right\}} \tag{13}$$

$$\phi\left(\boldsymbol{s}_i\right) = \exp\left[2.1 - 0.23 u_1\left(\boldsymbol{s}_i\right) - 0.8 u_2\left(\boldsymbol{s}_i\right) + 0.5 u_3\left(\boldsymbol{s}_i\right) + 2.3 w_1\left(\boldsymbol{s}_i\right) - 1.1 w_2\left(\boldsymbol{s}_i\right) + z_2\left(\boldsymbol{s}_i\right)\right] \tag{14}$$

$i = 1, \ldots, 100$, where $w_1(\boldsymbol{s}_i)$ and $w_2(\boldsymbol{s}_i)$ are the ith spatial coordinates, $v_1(\boldsymbol{s}_i) = u_1(\boldsymbol{s}_i) \sim N(1, 4)$ and $v_2(\boldsymbol{s}_i) = u_2(\boldsymbol{s}_i)$, $v_3(\boldsymbol{s}_i) = u_3(\boldsymbol{s}_i)$ are two dummy variables associated with the categorical variable and $z_1(\boldsymbol{s}_i)$ and $z_2(\boldsymbol{s}_i)$ follow stationary Gaussian processes with isotropic spherical variograms given by $\gamma_1(h_1) = 1+4\left(1.5\left(\frac{h_1}{0.4}\right) - \frac{1}{2}\left(\frac{h_1}{0.4}\right)^3\right)$ and $\gamma_2(h_2) = 5+10\left(1.5\left(\frac{h_2}{0.3}\right) - \frac{1}{2}\left(\frac{h_2}{0.3}\right)^3\right)$, respectively, for $h_1 > 0$, $h_2 > 0$ and $\gamma_1(0) = \gamma_2(0) = 0$.

We first fitted the data by a BSLMM using the logit link function with independent and identically distributed (i.i.d.) random effects in the SMM and a log link function with i.i.d. random effects in the SVDM. For the ML estimation, we used the approach in section 2 with $\tilde{f}(\cdot)$ equal to $f(\cdot \mid b_0)$, where $b_0 = (\boldsymbol{\zeta}_0, \boldsymbol{\varphi}_0, \boldsymbol{\theta}_{01}, \boldsymbol{\theta}_{02})^t$ with $\boldsymbol{\zeta}_0 = (0.6, -0.04, 0.07, -1, 3, -2)^t$, $\boldsymbol{\varphi}_0 = (0, -0.08, -1.5, -0.3, 4, 0.2)^t$, $\boldsymbol{\theta}_{01} = (10, 0.5, 3)^t$ and $\boldsymbol{\theta}_{02} = (10, 0.3, 5)^t$. After an initial run of the MCMC algorithm to determine appropriate values of the parameters b_0, we present the results using a spherical correlation function for both the SMM and SVDM. The Monte Carlo sample size was 1000, and the first 10 000 iterations were discarded as a burn-in period; a total of 1 000 000 simulations were run, taking samples every 1000 iterations. We considered the following models:

- H_0 : No spatial correlation either in the mean model or in the variable dispersion model, but with constant dispersion
- H_1 : No spatial correlation either in the mean model or in the variable dispersion model, but with variable dispersion
- H_2 : No spatial correlation in the mean model but with spatial spherical correlation in the variable dispersion model
- H_3 : Spatial spherical correlation in the mean model but with no spatial correlation in the variable dispersion model
- H_4 : Spherical correlation function in the mean model with $\vartheta_1 > 0$ and $\tau_1^2 = 0$ and spherical correlation function in the variable dispersion model with $\vartheta_2 > 0$ and $\tau_2^2 = 0$
- H_5 : Spherical correlation function in the mean model with $\vartheta_1 > 0$ and $\tau_1^2 \geq 0$ and spherical correlation function in the variable dispersion model with $\vartheta_2 > 0$ and $\tau_2^2 \geq 0$
- H_6 : Spherical correlation function in the mean model with $\vartheta_1 > 0$ and $\tau_1^2 \geq 0$ and Gaussian correlation function in the variable dispersion model with $\vartheta_2 > 0$ and $\tau_2^2 \geq 0$
- H_7 : Gaussian correlation function in the mean model with $\vartheta_1 > 0$ and $\tau_1^2 \geq 0$ and exponential correlation function in the dispersion model with $\vartheta_2 > 0$ and $\tau_2^2 \geq 0$
- H_8 : Gaussian correlation function in the mean model with $\vartheta_1 > 0$ and $\tau_1^2 \geq 0$ and spherical correlation function in the variable dispersion model with $\vartheta_2 > 0$ and $\tau_2^2 \geq 0$

The results obtained after running the MCMC algorithm for the BSLMM method are presented in Table 1. If we compare the difference of the log $\hat{L}$'s between H_1 and

Table 1. Spatial beta maximum likelihood estimates using the models H_0, H_1, H_2, H_3, H_4, H_5, H_6, H_7 and H_8

Spatial mean model									
Model	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{\xi}_3$	$\hat{\xi}_4$	$\hat{\xi}_5$	$\hat{\sigma}_1^2$	$\hat{\vartheta}_1$	$\hat{\tau}_1^2$
H_0	−0.832	0.053	0.211	−0.101	1.226	−0.086	*	*	*
H_1	−0.851	0.056	0.157	−0.209	1.471	0.047	*	*	*
H_2	−0.125	0.060	0.074	−0.302	1.563	−0.142	*	*	*
H_3	−0.028	0.014	0.183	0.152	1.381	−0.955	1.792	0.415	0.183
H_4	1.822	−0.016	0.450	−0.215	1.541	−2.134	5.117	0.406	0.000
H_5	1.652	0.005	0.377	−0.578	1.513	−2.083	3.688	0.429	0.960
	(0.002)	(0.000)	(0.001)	(0.002)	(0.002)	(0.002)	(0.005)	(0.000)	(0.002)
H_6	1.659	0.005	0.376	−0.580	1.504	−2.089	3.695	0.429	0.966
H_7	1.776	0.006	0.299	−0.705	1.423	−2.159	2.929	0.218	1.773
H_8	1.705	0.003	0.402	−0.591	1.448	−2.323	3.419	0.218	1.900

Spatial variable dispersion model												
Model	$\hat{\varphi}_0$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\varphi}_4$	$\hat{\varphi}_5$	$\hat{\sigma}_2^2$	$\hat{\vartheta}_2$	$\hat{\tau}_2^2$	$\log \hat{L}$	AIC	BIC
H_0	−0.792	*	*	*	*	*	*	*	*	191.45	−368.90	−373.69
H_1	−1.391	0.021	−0.403	−0.102	1.059	0.846	*	*	*	199.23	−374.46	−389.25
H_2	−0.620	0.000	−0.350	0.204	1.034	0.206	0.385	0.696	0.344	216.24	−402.48	−423.27
H_3	−0.235	−0.018	−0.656	−4.161	0.693	0.551	*	*	*	225.56	−421.12	−441.91
H_4	2.034	−0.052	−0.632	−0.283	1.850	−1.336	7.558	0.193	0.000	291.68	−551.36	−574.15
H_5	2.784	−0.157	−0.765	0.199	1.891	−1.384	6.798	0.329	5.404	323.12	−610.24	−637.03
	(0.008)	(0.001)	(0.005)	(0.006)	(0.007)	(0.008)	(0.035)	(0.000)	(0.023)			
H_6	2.831	−0.154	−0.791	0.188	1.882	−1.437	5.199	0.174	6.847	322.71	−609.42	−636.21
H_7	2.731	−0.128	−0.756	0.280	1.853	−1.499	10.017	0.099	1.241	317.30	−598.6	−625.39
H_8	2.414	−0.071	−0.911	−0.097	2.116	−0.546	2.705	1.174	6.682	306.66	−577.32	−604.11

AIC, Akaike information criterion; BIC, Bayesian information criterion.
The asterisk corresponds to a model where the parameter is meaningless.

H_0, we find that the dispersion is not constant ($0.5\chi^2_{(5)} = 5.54 < 7.78$). Comparing the difference of the $\log \hat{L}$'s in H_5 and H_1 relative to a $0.5[0.5(\chi^2_{(7)} + \chi^2_{(6)})]$-distribution (whose 95% quantile is 6.665, see details in Stram and Lee, 1994), we note a clear spatial correlation in both models: mean and variable dispersion. Also, when we compare the difference of the $\log \hat{L}$'s in H_3 and H_2 with respect to H_1, we find an evident spatial correlation in SVDM and SMM, respectively ($0.5[0.5(\chi^2_{(4)} + \chi^2_{(3)})] = 4.326$). However, if we compare H_5 with respect to H_3 and H_2, we note that the spatial correlation in the SMM and the SVDM is necessary for BSLMM. In this simulation, we note a presence of nugget effects, which is seen by comparing the $\log \hat{L}$ differences of H_5 and H_4 relative to a $0.5[0.5(\chi^2_{(2)} + \chi^2_{(1)})]$-distribution (whose 95% quantile is 2.458). Note that the estimates $\hat{\vartheta}_1$ and $\hat{\vartheta}_2$ for H_4 are smaller than the estimates $\hat{\vartheta}_1$ and $\hat{\vartheta}_2$ for H_5, which is consistent with the models H_4 having no nugget effects and the models H_5 having nugget effects.

To compare non-nested models, we use Akaike information criterion (AIC) and Bayesian information criterion (BIC) for choosing the best model. The results in Table 1 show that the selection of the correlation function is not important, although the spherical correlation functions (H_5) in both the SMM and SVDM are slightly better than other competitive models (the lowest AIC and BIC). In this way, this result is concordant with the simulation development. Several runs of the MCMC algorithm for the BSLMM approach were performed with different starting values (b_0) and dif-

Table 2. Log-likelihood values using different link functions in spatial mean model and spatial variable dispersion model with spherical correlation functions in both models

	Spatial mean model			
Spatial variable dispersion model (ν_2)	loglog	logit	cloglog	probit
0.0 (log)	300.71	323.12	270.00	288.17
0.5 (square root)	254.30	270.78	252.87	237.83
0.8	249.27	245.54	235.80	243.58
1.0 (identity)	245.54	241.91	233.03	240.22

ferent correlation functions. The results did not differ much, and the pattern was the same as in Table 1. We did not encounter any problems with multimodality of the likelihood function apart from when using the spherical correlation function. The multimodality for the spherical correlation function is a well-known problem (Mardia and Watkins, 1989) caused by this function only being once differentiable at $h_j = \vartheta_j$ (Christensen, 2004).

One important part of the strategy was to use a Gibbs sampling scheme where the components were updated using local-type Metropolis–Hastings updates. The orthogonalization of the components improves the performance of the Gibbs sampling scheme, and the standardization of the individual components makes it easy to scale proposal densities such that the Metropolis–Hastings updates become very efficient.

Model comparison is here used in a somewhat informal way because some of the models that were compared are not nested. The analysis also assumes that the log-likelihood ratio can be reasonably approximated by a chi-squared distribution, but it is often hard to prove this rigorously. Making inferences about spatial correlation based on the log-likelihood ratio will tend to be conservative (in terms of the acceptance of a null model with no spatial dependence) because, in a nested presentation of our model H_1 and any others, one or more parameters go to a bound, or is undefined, in the null case. This fact for spatial data will be studied in more detail in a future research following works such as Stram and Lee (1994), Zhang and Lin (2003) and Crainiceanu *et al.* (2005).

The results of the log-likelihood values for different link functions for the SMM and the SVDM using a spherical correlation function in both models are shown in Table 2. We choose the model with logit link for the SMM and log link for the SVDM to fit the BSLMM because it shows the highest log-likelihood (323.12). Also, note that in Table 1, this model showed the lowest AIC (−610.24) and BIC (−637.03). All these agreed with our simulation case.

Table 1 shows the estimated parameters and standard deviations of the selected model (H_5). It is possible to see that all the estimates agree with the respective true values of the parameters, and all of them have small standard deviations. Figures 1 and 2 for the spatial mean and spatial variable dispersion parameters, respectively, show the chain samples and their distribution function, which confirms these results, noting that all parameters are significant. When judging performance of MCMC, we need to bear in mind that a particular set of realized values of random effects was used in MCMC to fit the BSLMM in both models (SMM and SVDM). Considering this, we consider that the MCMC estimates are satisfactory.

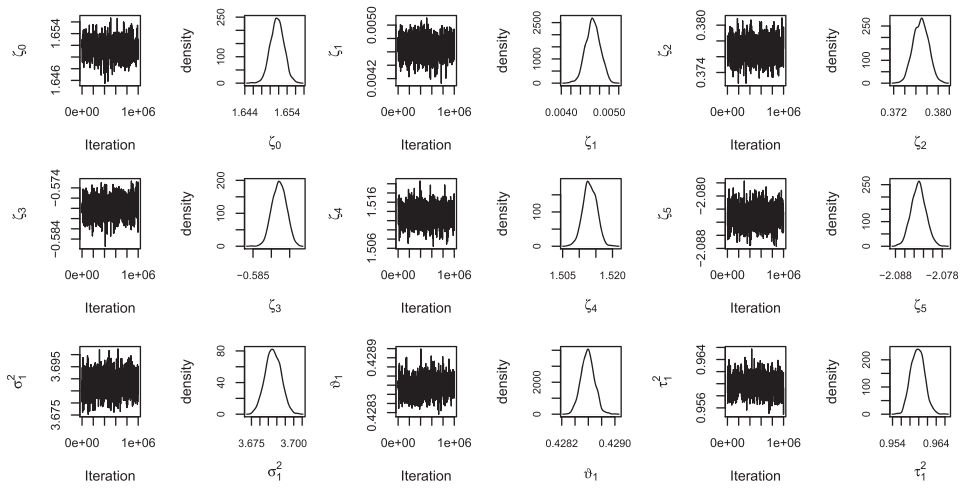

Fig. 1. Behaviour of the chain sample for parameters of the spatial mean model (Equation 13).

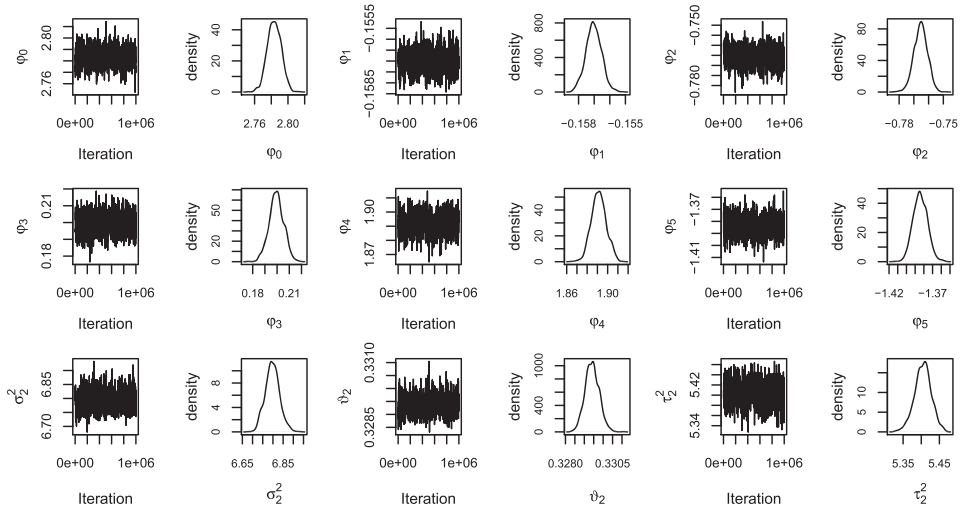

Fig. 2. Behaviour of the chain sample for parameters of the spatial variable dispersion model (Equation 14).

4 Applications

We now consider the two applications presented earlier in the paper: the first is the clay content on samples of fixed soil layer at 0–10 cm at each of 147 soil profile observations, and the second is magnesium content on soil samples taken from 0- to 20-cm-depth layer at each of 178 locations.

4.1 Clay content

The data set contains 147 soil profile observations from the research area of the Tropenbos Cameroon Programme (Yemefack et al., 2005). Specifically, the studied

measurements were made of samples of fixed soil layer at 0–10 cm. We investigate the relationship between the clay content (weight per cent of the mineral fine earth and measured in millimoles$_c$ per cubic decimetre) and the explanatory variables elevation in metres above sea level (ELEV), four agro-ecological zone (ZONE) and two reference soil groups (WRB), which are Acrisols and Ferralsols. For each observation, the east and north coordinates (UTM zone 32N, WGS84 datum, in metres) were recorded.

Let $y(\boldsymbol{s}_i)$ denote the clay content at location $\boldsymbol{s}_i$. In our model, we assume that the $y(\boldsymbol{s}_i)$'s are conditionally independent beta variables given unobserved spatial stochastic processes $z_1(\boldsymbol{s}_i)$ and $z_2(\boldsymbol{s}_i)$ for the SMM and the SVDM, respectively. Additionally, we assume that the SMM at $\boldsymbol{s}_i$ depends on the explanatory variables elevation and agro-ecological zone observed at location $\boldsymbol{s}_i$ on $z_1(\boldsymbol{s}_i)$ and the SVDM at $\boldsymbol{s}_i$ depends on the explanatory variables agro-ecological zone and reference soil group observed at location $\boldsymbol{s}_i$ on $z_2(\boldsymbol{s}_i)$.

Figure 3 shows the relationship between the potential covariates and the clay content. There may be a positive relationship between clay content and the elevation. The boxplots in Figure 3c suggest that the means of the distributions of clay content are different in the distinct agro-ecological zones; the same happened with the boxplots of two reference soil groups (Figure 3d). We did not include the relationship between the locations (E–W and N–S) and clay content because it was less clear.

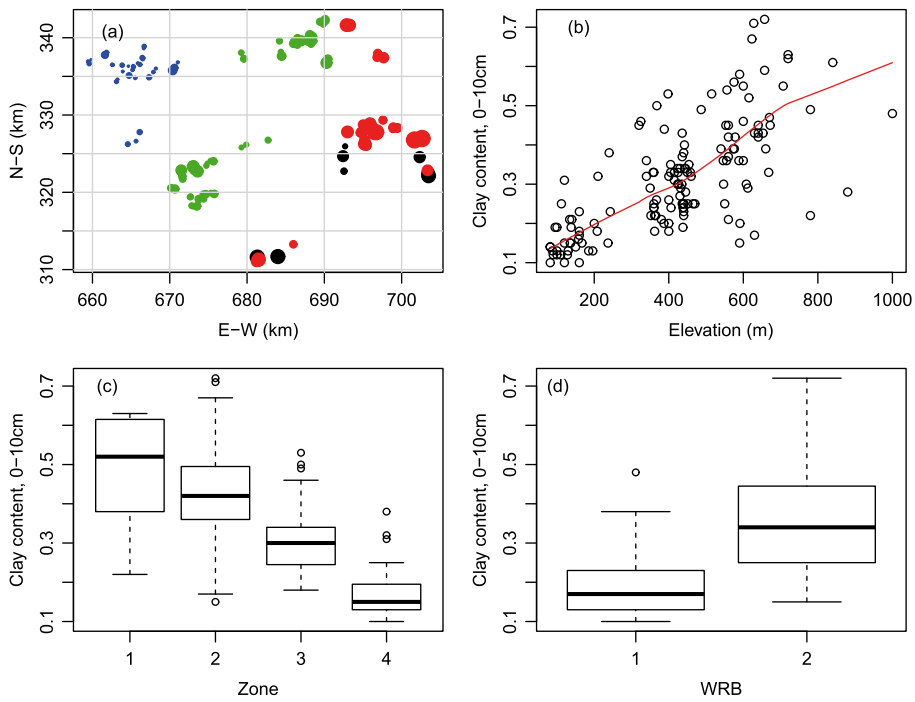

Fig. 3. (a) Locations of the clay content, (b) scatterplot of clay content against elevation where the line is the locally weighted scatterplot smoothing curve, (c) boxplots of clay content in each of the four zones and (d) boxplots of clay content in each of the two reference soil groups.

We did not find any evidence of anisotropy in SMM and SVDM, and thus, we omit the corresponding details. Therefore, we restrict our study to isotropic correlation functions, where the correlation function $\rho(h_j; \vartheta_j)$ depends only on the Euclidean distance $h_j = \|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\|_j$ between locations, $j = 1, 2$, $i, i' = 1, \ldots, n$. The roles of $z_1(\boldsymbol{s}_i)$ and $z_2(\boldsymbol{s}_i)$ in the models are to capture residual spatial variations in the SMM and SVDM, respectively, after adjusting for the explanatory variables.

Similar to the simulation case, we first fitted the data by a BSLMM using the logit link function with i.i.d. random effects in the SMM and the log link function with i.i.d. random effects in the SVDM. For the ML estimation, we used $b_0 = (\boldsymbol{\zeta}_0, \boldsymbol{\varphi}_0, \boldsymbol{\theta}_{01}, \boldsymbol{\theta}_{02})^t$ with $\boldsymbol{\zeta}_0 = (-0.9, 0.002, -0.4, -0.6, -0.9)^t$, $\boldsymbol{\varphi}_0 = (6, -1.7, -0.8, -2.2, -1.8)^t$, $\boldsymbol{\theta}_{01} = (0.3, 0.81, 0.1)^t$ and $\boldsymbol{\theta}_{02} = (0.2, 0.6, 0.1)^t$ to approach $\tilde{f}(\cdot)$ to $f(\cdot \mid b_0)$. An initial run of the MCMC algorithm to determine appropriate values of the parameter b_0 was made. The Monte Carlo sample size was 1000, and the first 10 000 iterations were discarded as a burn-in period; a total of 1 000 000 simulations were run using a spherical correlation function for both the SMM and SVDM, taking samples every 1000 iterations. We analyse the same models (H_1–H_8) as that in the simulation case.

In Table 3, we show the results obtained after running the MCMC algorithm for the BSLMM method. If we compare the difference of the $\log \hat{L}$'s between H_1 and H_0, we find that the dispersion is not constant ($0.5\chi^2_{(4)} = 4.74 < 9.57$). Comparing the difference of the $\log \hat{L}$'s in H_5 (or equivalently H_8) and H_1 (or equivalently H_3) relative to a $0.5[0.5(\chi^2_{(4)} + \chi^2_{(3)})]$-distribution (whose 95% quantile is 4.326), we note that there is a slight significant spatial correlation. Therefore, we chose the spatial model H_5 (or H_8) because there is a slight spatial correlation in the variable dispersion model, although it is not present in the mean model. The nugget effects in both models are not significant according to all proposed models. Table 3 also shows that the choice of the correlation

Table 3. Spatial beta maximum likelihood estimates for the clay content using the models H_0, H_1, H_2, H_3, H_4, H_5, H_6, H_7 and H_8

Spatial mean model											
Model	$\hat{\zeta}_0$	$\hat{\zeta}_1$	$\hat{\zeta}_2$	$\hat{\zeta}_3$	$\hat{\zeta}_4$	$\hat{\sigma}^2_1$	$\hat{\vartheta}_1$	$\hat{\tau}^2_1$			
H_0	−1.012	0.002	−0.376	−0.535	−0.767	*	*	*			
H_1, H_3	−1.878	0.002	0.541	0.343	0.045	0.000	0.000	0.000			
H_2, H_4	−1.850	0.002	0.619	0.359	0.017	*	*	*			
H_5, H_8	−2.085	0.002	0.630	0.457	0.212	0.000	0.000	0.000			
H_6	−1.965	0.002	0.569	0.380	0.110	0.000	0.000	0.000			
H_7	−2.140	0.002	0.655	0.490	0.266	0.000	0.000	0.000			
Spatial variable dispersion model											
Model	$\hat{\varphi}_0$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\varphi}_4$	$\hat{\sigma}^2_2$	$\hat{\vartheta}_2$	$\hat{\tau}^2_2$	$\log \hat{L}$	AIC	BIC
H_0	24.90	*	*	*	*	*	*	*	147.70	−283.40	−285.42
H_1, H_3	3.100	1.123	2.000	0.793	−1.620	*	*	*	156.27	−292.54	−302.56
H_2, H_4	3.141	1.226	2.164	0.759	−1.743	0.048	1.334	0.000	158.86	−293.72	−307.74
H_5, H_8	2.932	1.170	2.434	1.265	−1.655	0.147	5.415	0.000	161.41	−298.82	−312.84
H_6	2.952	1.082	2.271	1.188	−1.569	0.097	1.209	0.000	160.72	−297.44	−311.46
H_7	2.958	1.173	2.402	1.292	−1.659	0.132	2.412	0.000	160.62	−297.24	−311.26

AIC, Akaike information criterion; BIC, Bayesian information criterion.
The asterisk corresponds to a model where the parameter is meaningless.

Table 4. Log-likelihood values for the clay content using different link functions in the spatial mean model and the spatial variable dispersion model with spherical correlation functions in both models

	Spatial mean model			
Spatial variable dispersion model (ν_2)	loglog	logit	cloglog	probit
0.0 (log)	161.36	161.41	160.87	158.23
0.5 (square root)	157.13	159.23	158.60	159.39
0.7	157.77	156.81	158.49	159.17
1.0 (identity)	155.18	157.10	158.49	157.43

function is not relevant, although the spherical correlation function model (H_5 or H_8) for the variable dispersion part is slightly better than other competitive models. Additionally, the results did not differ much, and the pattern was the same as in Table 3 when several runs of the MCMC algorithm for BSLMM approach were performed with different starting values (b_0) and correlation functions.

In Table 4, the log-likelihood values for different link functions for SMM and SVDM using spherical correlation function in both parts of model are shown. We chose the model with logit link for the SMM and log link for the SVDM to fit the BSLMM because it showed the highest log-likelihood (161.41), and according to Table 3, it has the lowest AIC (−298.82) and BIC (−312.84), too.

Then, the fitted BSLMM with spatial mixed models are given by

$$\begin{aligned}\operatorname{logit}\left(\hat{\mu}\left(\boldsymbol{s}_i\right)\right) &= \log\left(\frac{\hat{\mu}\left(\boldsymbol{s}_i\right)}{1-\hat{\mu}\left(\boldsymbol{s}_i\right)}\right) = -2.085 + 0.002 \times \mathrm{ELEV}\left(\boldsymbol{s}_i\right) + 0.630 \times \mathrm{ZONE}_2\left(\boldsymbol{s}_i\right) \\ &\quad + 0.457 \times \mathrm{ZONE}_3\left(\boldsymbol{s}_i\right) + 0.212 \times \mathrm{ZONE}_4\left(\boldsymbol{s}_i\right) \\ \log\left(\hat{\phi}\left(\boldsymbol{s}_i\right)\right) &= 2.932 + 1.170 \times \mathrm{ZONE}_2\left(\boldsymbol{s}_i\right) + 2.434 \times \mathrm{ZONE}_3\left(\boldsymbol{s}_i\right) + 1.265 \times \mathrm{ZONE}_4\left(\boldsymbol{s}_i\right) \\ &\quad - 1.655 \times \mathrm{WRB}_2\left(\boldsymbol{s}_i\right) + \hat{z}_2\left(\boldsymbol{s}_i\right)\end{aligned}$$

where $\hat{y}(\boldsymbol{s}_i)$ is the clay content at the ith location ($i = 1, \ldots, 147$), $\mathrm{ELEV}(\boldsymbol{s}_i)$ is the elevation above sea level at the ith location, $\mathrm{ZONE}_j(\boldsymbol{s}_i)$ is the jth agro-ecological zone at the ith location ($j = 2, 3, 4$), $\mathrm{WRB}_2(\boldsymbol{s}_i)$ is the Ferralsol soil group at the ith location and $\hat{z}_2(\boldsymbol{s}_i)$ is a stationary Gaussian process with fitted isotropic spherical variogram $\hat{\gamma}_2(h_2) = 0.147\left(\frac{3}{2}\left(\frac{h_2}{5.415}\right) - \frac{1}{2}\left(\frac{h_2}{5.415}\right)^3\right)$ for $h_2 > 0$ and $\hat{\gamma}_2(0) = 0$.

Table 5 shows the estimated parameter, standard deviation and the 95% confidence intervals for each of the parameters in BSLMM (H_5 or H_8). In the SMM, the variable elevation has a significant positive effect over clay content, and the agro-ecological zone 1 reduces the clay content. On the other hand, in the SVDM, the Ferralsol soil is not significant and differs from Acrisol, while the agro-ecological zones have positive significant effects. Additionally, the spatial correlation parameters without the nugget effect are relevant to fit the spatial variable dispersion (precision) model. In particular, for the SVDM used in BSLMM, the 95% interval for σ_2^2 is $(0.0001, 0.2471)$ on the log scale, and the interval for ϑ_2 of $(0.0025, 9.6836)$ km indicates that this residual spatial variation operates on a relatively medium scale. Besides, using the logit link function in

Table 5. Estimates and 95% confidence intervals for the parameters involved in beta spatial linear mixed model to fit the clay content

			Quantile	
Parameter	Estimation	Standard deviation	2.5%	97.5%
		Spatial mean model		
ζ_0	−2.085	0.646	−2.9783	−0.4990
ζ_1	0.002	0.000	0.0001	0.0028
ζ_2	0.630	0.349	−0.2751	1.0547
ζ_3	0.457	0.430	−0.6372	1.0343
ζ_4	0.212	0.569	−1.1810	0.9723
		Spatial variable dispersion model		
φ_0	2.932	0.818	1.5422	4.6789
φ_1	1.170	0.628	−0.9638	1.5706
φ_2	2.434	0.633	0.0367	2.5415
φ_3	1.265	0.896	−1.3104	2.3072
φ_4	−1.655	1.102	−3.1615	1.0566
σ_2^2	0.147	0.069	0.0001	0.2471
ϑ_2	5.415	2.557	0.0025	9.6836

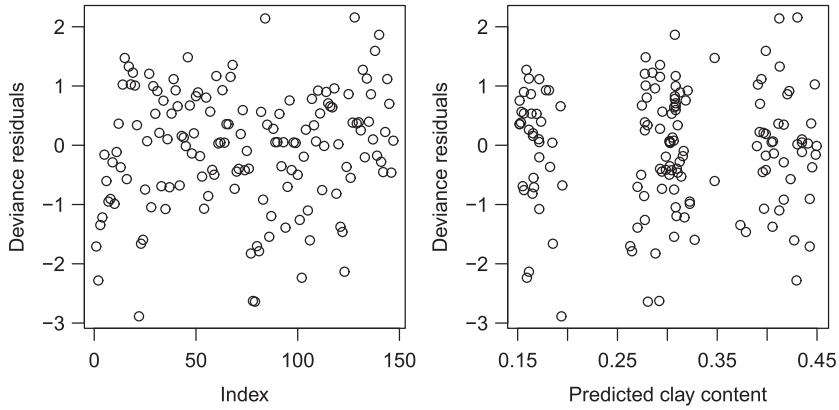

Fig. 4. Deviance residuals against: number of fixed soil layer (*left panel*) and predicted values (*right panel*) using beta spatial linear mixed model for the clay content.

the SMM and log link in the SVDM to fit BSLMM, we found the pseudo-$R_s^2 = 0.51$, which suggests a slight goodness of fit of the proposed model.

In addition, Figure 4 (right panel) plots deviance residuals (r_i) against predicted clay content values, and it shows the desired absence of any obvious relationship, indicating an adequate first-order fit. The left panel of Figure 4 shows absence of any relationship between deviance residuals (r_i's) and the index of fixed soil layer.

Once the model was fitted, the idea in many studies is to show a prediction map of the clay content; however, an analysis of this kind requires the explanatory variables to be recorded at both data and prediction locations. In this study, we do not have this information, and so, we do not perform it here. In the next application, the predicted maps of the response variable and their corresponding standard deviations for the prediction error are obtained.

4.2 Magnesium content

This data set considered soil samples collected with a Dutch-type drill on an incomplete regular lattice at a spacing of approximately 50 m, with geographic coordinates: north and east 900 m apart in both directions. The soil samples were taken from a 0- to 20-cm-depth layer at each of 178 locations (Figure 6). Magnesium content was measured in millimoles$_c$ per cubic decimetre. The study region was divided into three sub-regions, which experienced different soil management regimes. The first, in the upper left corner (Figure 6), is typically flooded during each rainy season and is no longer used as an experimental area because of its varying altitude. The second, corresponding to the lower half of the study region, and the third, in the upper right corner, have received fertilizers in the past: the second is typically occupied by rice fields, while the third is frequently used as an experimental area (Capeche *et al*., 1997; Diggle and Ribeiro, 2007).

So, this data set had information about the magnesium content ($y(\boldsymbol{s}_i)$), the spatial coordinates (w_1, w_2), altitude (ALT) and sub-region (SR) code of each sample, which is associated with three periods of fertilization in different areas. The data are taken from Capeche *et al.* (1997), and the main objective of the study was adequate land use planning that allows rational and sustainable management, avoiding the erosion process. This is important in order to allocate subsidies in the experimental fields to perform searches that will be extrapolated to similar soils and climatic zones. In this study, we assume that the $y(\boldsymbol{s}_i)$'s are conditionally independent beta variables given the unobserved spatial stochastic processes $z_1(\boldsymbol{s}_i)$ and $z_2(\boldsymbol{s}_i)$ for the SMM and the SVDM, respectively.

Similar to the clay content application, we did not find any evidence of anisotropy in both the SMM and SVDM. Furthermore, we first fitted the data by a BSLMM using the cloglog link function with i.i.d. random effects in the SMM and the transformation 0.3 as link function with i.i.d. random effects in the SVDM. For the ML estimation, we used $b_0 = (\boldsymbol{\zeta}_0, \boldsymbol{\varphi}_0, \boldsymbol{\theta}_{01}, \boldsymbol{\theta}_{02})^t$ with $\boldsymbol{\zeta}_0 = (-5.9, -0.0001, 0.0008, 0.2, 0.5, 0.4)^t$, $\boldsymbol{\varphi}_0 = (29.5, -6.6, -3.5, -1.2)^t$, $\boldsymbol{\theta}_{01} = (4, 0.6, 1)^t$ and $\boldsymbol{\theta}_{02} = (5, 0.5, 1)^t$ to approach $\tilde{f}(\cdot)$ to $f(\cdot \mid b_0)$. The Monte Carlo sample size was developed similarly to the simulation case and the clay content application.

In Table 6, we show the results obtained after running the MCMC algorithm for the BSLMM method. If we compare the difference of the log $\hat{L}$'s between H_1 and H_0, we find that the dispersion is not constant ($0.5\chi^2_{(3)} < 8.57$). Comparing the difference of the log $\hat{L}$'s in H_8 and H_1 relative to a $0.5[0.5(\chi^2_{(5)} + \chi^2_{(4)})]$-distribution (whose 95% quantile is 5.140), we can say that there exists a clear spatial correlation in both the SMM and the SVDM. However, when we compare the difference of the log $\hat{L}$'s in H_2 (or H_5) with respect to H_1 (or H_3), we do not find an evident spatial correlation in the SVDM ($0.5[0.5(\chi^2_{(2)}+\chi^2_{(1)})] > 1.74$). If we compare H_8 with respect to H_1 (or H_3) and H_2 (or H_5), we note that it is necessary to consider the spatial correlation in the SMM and the SVDM to fit the BSLMM. Table 6 also shows that the selection of the correlation function is relevant; the Gaussian correlation function for the SMM and spherical

Table 6. Spatial beta maximum likelihood estimates for the magnesium content using the models H_0, H_1, H_2, H_3, H_4, H_5, H_6, H_7 and H_8

				Spatial mean model						
Model	$\hat{\xi}_0$	$\hat{\xi}_1$	$\hat{\xi}_2$	$\hat{\xi}_3$	$\hat{\xi}_4$	$\hat{\xi}_5$	$\hat{\sigma}_1^2$	$\hat{\vartheta}_1$	$\hat{\tau}_1^2$	
H_0	2.099	−0.0001	0.0002	0.028	0.567	0.510	*	*	*	
H_1, H_3	−5.318	−0.0001	0.0007	0.172	0.401	0.375	*	*	*	
H_2, H_5	−5.929	−0.0084	0.0732	0.195	0.399	0.381	*	*	*	
H_4	−5.994	−0.0928	0.760	0.189	0.399	0.370	0.003	10.00	0.00	
H_6	−5.934	−0.0854	0.733	0.195	0.400	0.382	0.000	0.00	0.00	
H_7	−6.917	−0.0834	0.969	0.125	0.471	0.387	0.325	373.00	0.00	
H_8	−7.154	−0.0235	0.964	0.137	0.294	0.215	0.468	497.00	0.00	
				Spatial variable dispersion model						
$\hat{\varphi}_0$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\sigma}_2^2$	$\hat{\vartheta}_2$	$\hat{\tau}_2^2$	$\log \hat{L}$	AIC	BIC	
H_0	60.159	*	*	*	*	*	*	259.13	−504.26	−507.90
H_1, H_3	29.400	0.000	−0.004	−1.169	*	*	*	267.70	−517.40	−525.04
H_2, H_5	30.463	0.043	−1.347	−0.472	0.00	0.00	0.34	269.44	−516.88	−528.52
H_4	101.101	0.705	−12.890	−4.764	1.43	25.60	0.00	328.06	−628.12	−645.76
H_6	105.348	0.453	−13.477	−4.710	1.41	36.50	0.28	328.45	−630.90	−646.54
H_7	88.450	1.844	−12.564	−3.821	3.02	23.80	0.00	337.56	−647.12	−664.76
H_8	88.263	2.571	−13.167	−3.926	2.87	74.80	0.00	338.71	−649.42	−667.06

AIC, Akaike information criterion; BIC, Bayesian information criterion.
The asterisk corresponds to a model where the parameter is meaningless.

Table 7. Log-likelihood values for the magnesium content using different link functions in the spatial mean model and the spatial variable dispersion model with Gaussian and spherical correlation functions, respectively

	Spatial mean model			
Spatial variable dispersion model (ν_2)	loglog	logit	cloglog	probit
0.0 (log)	327.98	337.43	337.16	336.48
0.3	328.55	338.42	338.71	338.24
0.5 (square root)	328.39	338.37	338.60	338.15
1.0 (identity)	337.69	337.41	327.91	337.16

correlation function for the SVDM (H_8) are slightly better than the model H_7. In H_8, the nugget effects in both models are not important. Similar to the aforementioned studies, the results did not differ much, and the pattern was the same as in Table 6 when several runs of the MCMC algorithm for BSLMM approach were performed with different starting values (b_0) and correlation functions.

The results of the log-likelihood values for different link functions for the SMM and the SVDM using Gaussian and spherical correlation functions in the SMM and the SVDM, respectively, are shown in Table 7. We chose the model with cloglog link for the SMM and the transformation 0.3 as link for the SVDM to fit the BSLMM because it showed the highest log-likelihood (338.71). This BSLMM is slightly higher than the others. Note that, also in Table 6, this model showed the lowest AIC (−649.42) and BIC (−667.06).

Therefore, the fitted BSLMM with spatial mixed models are given by

$$\operatorname{cloglog}\left(\hat{\mu}\left(\boldsymbol{s}_i\right)\right) = \log\{-\log\left[1-\hat{\mu}\left(\boldsymbol{s}_i\right)\right]\} = -7.154 - 0.023 \times w_1\left(\boldsymbol{s}_i\right) + 0.964 \times w_2\left(\boldsymbol{s}_i\right) + 0.137 \times \mathrm{ALT}\left(\boldsymbol{s}_i\right) + 0.294 \times \mathrm{SR}_2\left(\boldsymbol{s}_i\right) + 0.215 \times \mathrm{SR}_3\left(\boldsymbol{s}_i\right) + \hat{z}_1\left(\boldsymbol{s}_i\right)$$

$$\left(\hat{\phi}\left(\boldsymbol{s}_i\right)\right)^{0.3} = 88.263 + 2.571 \times w_1\left(\boldsymbol{s}_i\right) - 13.167 \times w_2\left(\boldsymbol{s}_i\right) - 3.926 \times \mathrm{ALT}\left(\boldsymbol{s}_i\right) + \hat{z}_2\left(\boldsymbol{s}_i\right)$$

where $\hat{y}(\boldsymbol{s}_i)$ is the magnesium content at the ith location ($i = 1, \ldots, 178$), $w_1(\boldsymbol{s}_i)$ and $w_2(\boldsymbol{s}_i)$ are the ith spatial coordinates, $\mathrm{ALT}(\boldsymbol{s}_i)$ is the altitude above sea level at the ith location, $\mathrm{SR}_j(\boldsymbol{s}_i)$ is the jth sub-region at the ith location ($j = 2, 3$), $\hat{z}_1(\boldsymbol{s}_i)$ is a stationary Gaussian process with fitted isotropic Gaussian variogram $\hat{\gamma}_2(h_2) = 0.468(1 - \exp(-3h_1^2/497^2))$ for $h_1 > 0$ and $\hat{\gamma}_1(0) = 0$ and $\hat{z}_2(\boldsymbol{s}_i)$ is a stationary Gaussian process with fitted isotropic spherical variogram $\hat{\gamma}_2(h_2) = 2.87\left(\frac{3}{2}\left(\frac{h_2}{74.80}\right) - \frac{1}{2}\left(\frac{h_2}{74.80}\right)^3\right)$ for $h_2 > 0$ and $\hat{\gamma}_2(0) = 0$.

Table 8 shows the estimated parameters, standard deviations and the 95% confidence intervals for each of the parameters in BSLMM (H_8). In the SMM, the variables north–south (w_2), altitude and sub-region have a significant positive effect over

Table 8. Estimates and 95% confidence intervals for the parameters involved in BSLMM to fit magnesium content

			Quantile	
Parameter	Estimation	Standard deviation	2.5%	97.5%
		Spatial mean model		
ζ_0	−7.154	1.758	−8.841	−2.110
ζ_1	−0.023	0.196	−0.234	0.483
ζ_2	0.964	0.317	0.105	1.311
ζ_3	0.137	0.056	0.047	0.255
ζ_4	0.294	0.145	0.012	0.526
ζ_5	0.215	0.130	0.026	0.508
σ_2^2	0.047	0.024	0.001	0.084
ϑ_2	0.497	0.162	0.111	0.949
		Spatial variable dispersion model		
φ_0	88.263	18.662	21.939	94.241
φ_1	2.571	1.492	−2.929	2.961
φ_2	−13.167	2.476	−16.650	−6.793
φ_3	−3.926	0.760	−4.741	−2.228
σ_2^2	0.287	0.281	0.001	0.821
ϑ_2	0.075	0.238	0.051	1.000

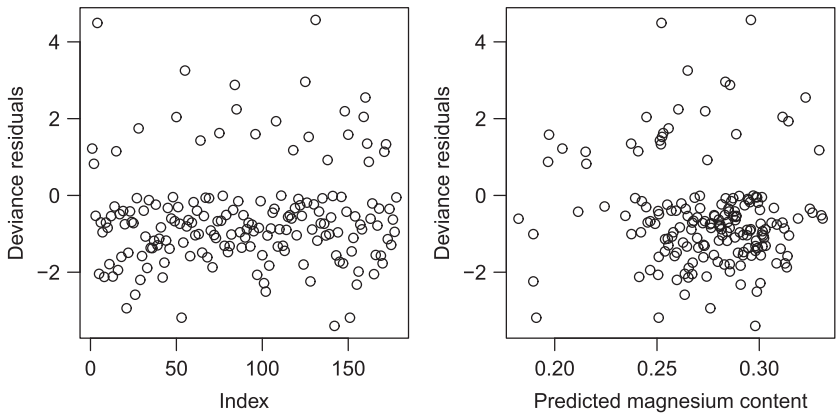

Fig. 5. Deviance residuals against: number of fixed soil layer (*left panel*) and predicted values (*right panel*) using beta spatial linear mixed model for the magnesium content.

magnesium content, while the coordinate east–west (w_1) does not have a significant effect. On the other hand, in the SVDM, the variables north–south and altitude have a significant negative effect over dispersion or precision response, and similarly to the mean model, the coordinate east–west does not have a significant effect. Additionally, the spatial correlation parameters without the nugget effect are relevant to fit the spatial variation in both models (SMM and SVDM). In particular, for the SMM used in BSLMM, the 95% interval for σ_1^2 is $(0.001, 0.084)$ on the cloglog link, and the interval for ϑ_1 of $(0.111, 0.949)$ km. For the SVDM used in BSLMM, the 95% interval for σ_2^2 is $(0.001, 0.821)$ on the transformation 0.3, and the interval for ϑ_2 of $(0.051, 1.000)$ km. In both cases, these confidence intervals indicate that this residual spatial variation operates on a relatively medium scale. Besides, using the cloglog link function in the SMM and the transformation 0.3 as link function in the SVDM to fit BSLMM, we found the pseudo-$R_s^2 = 0.42$, which suggests a slight goodness of fit for the proposed model.

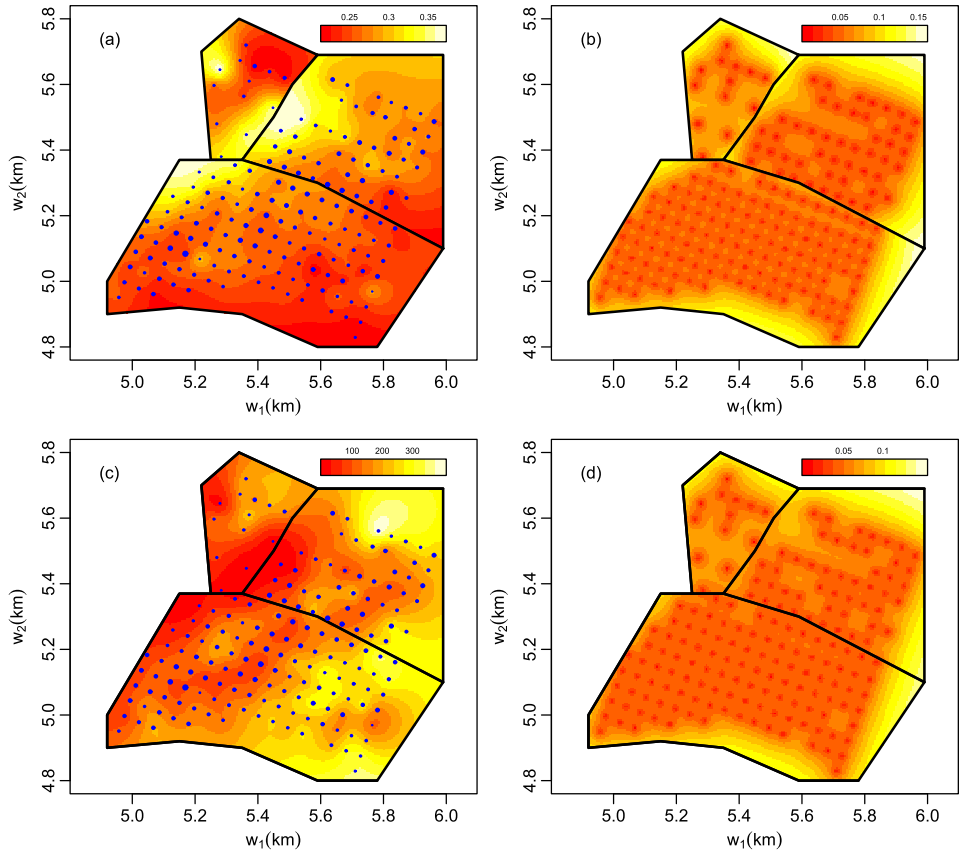

Fig. 6. (a) Point estimates using the spatial mean model (SMM), overlaid with the magnesium contents observed in field studies and each dot proportional to the corresponding measured magnesium content and lines delimiting sub-regions with different soil management practices. (b) Prediction standard errors for the SMM. (c) Point estimates using the spatial variable dispersion model (SVDM), overlaid with the magnesium contents observed in field studies. (d) Prediction standard errors for the SVDM.

Furthermore, Figure 5 (right panel) plots deviance residuals (r_i) against predicted magnesium content values and shows the desired absence of any obvious relationship, indicating an adequate first-order fit. The left panel of Figure 5 shows absence of any relationship between deviance residuals (r_i's) and the index of the fixed soil layer.

Once the model was fitted, the predicted maps of the magnesium content for the SMM and the SVDM and their corresponding standard deviations for the prediction errors were obtained using BSLMM. The results are presented in Figure 6. Note that at locations that are sufficiently close to one or more points, the predictive distribution shows a reduced standard error around the observed points in the sample in both models (Figure 6b,d). In Figure 6a (mean model), we can see that areas within the pale orange and yellow colour ranges are those in which there is a high magnesium content because the $\hat{\phi}(\boldsymbol{s}_i)$'s are high, while areas in the red–orange colour range are those in which there is a low magnesium content, not exceeding the 25% threshold. On the other hand, Figure 6b (dispersion map), we can see the precision over points observed in the mean map. In this panel, areas within the pale orange and yellow colour ranges are those in which there is a high precision of the magnesium content prediction; additionally, areas in the red–orange colour range are those in which there is a low precision of the magnesium content prediction, not exceeding the 100 threshold. Note that the prediction maps show discontinuities at the boundaries between the sub-regions as a consequence of treating the sub-region as a three-level factor.

5 Discussion and conclusions

In general, when there is paucity of field data and inherent uncertainty in the model inputs, decision-makers need to incorporate uncertainties into the model structure. Our proposed BSLMM with variable dispersion using MCMC is useful for situations where the spatial response variable is a rate or a proportion. MCMC proves to be a useful tool for model selection in complicated spatial models; however, it is well known that MCMC algorithms for spatial models (based on random fields) can be very slow to mix. This certainly seems to be the case for our model, where a thinning rate of 1 in 1000 is applied to the MCMC output. On the fitted model, inference can be used with the aim of performing better predictions at unobservable points. For the proposed BSLMM, we built two models (SMM and SVDM) to fit the parameters involved in the beta distribution. In both models, we used explanatory variables to model the trend, and we also considered the spatial correlation. Additionally, we performed inference and showed diagnostics over the parameters of the proposed BSLMM.

Many models require integration over values of latent variables (by evaluating each likelihood and then deriving with respect to each parameter holding all other parameters constant). This can be prohibitively expensive when the number of dimensions is large. When we use the ML estimate of the variance using MCMC, we obtain biased estimators. It can be very difficult to code problems in ML compared with simulating (and using MCMC) on it. It is sensitive to distributional assumptions (e.g. normality),

and assumed distributions of the statistics may not reflect the true distribution of statistics. Here, an alternative can be to incorporate spatial dependence directly into a penalized likelihood function and achieve greater efficiency in the resulting penalized MLEs. Unlike the computation of the penalized likelihood function for a spatial linear model, it can involve operations of a covariance matrix of the same size as the number of observations. Thus, the computational cost can be prohibitively high as the sample size becomes large. This can be studied in a future research using the same way followed in this paper.

In the clay content application, we found that the spatial correlation in the SMM was not significant; however, the spatial correlation without nugget effect in the SVDM was significant. This means that the precision is not constant over the space, and it is necessary to improve the fit of BSLMM. In the magnesium content application, we found that in both models, the spatial correlation without nugget effect was significant. The precision was high at locations with high magnesium content, and it was low at locations with slightly low magnesium content. The maps for mean and dispersion were possible to build because we have information on explanatory variables at locations where the response variable was not observed. This fact shows that we can use BSLMM to perform inference about the effect of the spatial explanatory variables over a spatial response variable and to find areas of high or low presence with its respective precision. The magnesium content maps in general presented different variability standards for the distribution in the non-responding area as well as in the area with different soil management regimens.

Understanding soil variation, and predicting it for agricultural management, is important; while the clay content of the soil is bounded between 0 and 1 as a proportion by mass of the mineral fraction, the researches are usually simultaneously interested in the proportion of other constituents of the same composition (e.g. sand) for management purposes, and so a model that allows us to examine all fractions is needed. So, it is often necessary to predict the distribution of mineral particles in soil between size fractions, given observations at sample sites. Because the contents in each fraction necessarily sum to 100%, these values constitute a composition, which we may assume is drawn from a random compositional variate (Lark and Bishop, 2007).

We can use the BSLMM working hierarchically. For example, if we want to predict three components A, B and C, which sum to 100%, we predict the level A, which is a percentage in the spatial point $\boldsymbol{s}_i$ ($\pi_1(\boldsymbol{s}_i)$), and then, between B and C, we have $(1 - \pi_1(\boldsymbol{s}_i))$. The next step is to build a BSLMM model that is in the interval (0,1) for the B component, so we predict at each spatial location s_i with a probability $\pi_2(\boldsymbol{s}_i)$, and for C component, we immediately assign the probability $1 - \pi_2(\boldsymbol{s}_i)$. Finally, the probability will be $\pi_1(\boldsymbol{s}_i)$ for A component, $(1 - \pi_1(\boldsymbol{s}_i))\pi_2(\boldsymbol{s}_i)$ for B component and $(1 - \pi_1(\boldsymbol{s}_i))(1 - \pi_2(\boldsymbol{s}_i))$ for C component in each location. This procedure makes $\pi_1(\boldsymbol{s}_i) + (1 - \pi_1(\boldsymbol{s}_i))\pi_2(\boldsymbol{s}_i) + (1 - \pi_1(\boldsymbol{s}_i))(1 - \pi_2(\boldsymbol{s}_i)) = 1$. However, this procedure is not studied in this paper, but it can be analysed in a future research.

In the simulation case and the two case studies, the different results presented in this paper show that the selection of the correlation function is not important. In general,

although the spherical correlation functions in both the SMM and SVDM are slightly better than those in other competitive models, several runs of the MCMC algorithm for the BSLMM approach were performed with different starting values (b_0) and different correlation functions; so the results did not differ much, and the pattern was the same. We did not encounter any problems with multimodality of the likelihood function apart from when using the spherical correlation function. The multimodality for the spherical correlation function is a well-known problem caused by this function only being once differentiable at $h_j = \vartheta_j$ (Mardia and Watkins, 1989; Christensen, 2004), but this problem was solved here using different starting values (b_0).

When the sample size increases, the estimates of the parameters become more like the simulated case. However, as the model has two parts (SMM and SVDM), these results are not easy to generalize. For this reason, a more detailed simulation was not performed because there are many possibilities in both models. So, we only present a case to show the potential advantages of the proposed method. Our proposed BSLMM should be used in cases where there exists significant spatial relationship between individuals in at least one of the SMM or the SVDM. In other cases, we may use the classic beta regression model with variable dispersion. Finally, in order to evaluate the predictor and the prediction error variances obtained under BSLMM, a cross-validation could be studied in future works.

On the other hand, the Bayesian approach should provide a natural way of incorporating parameter uncertainty in predictive inference. Meaningful priors on structural parameters such as the type of correlation and choice of link function are laborious to construct, and further, an MCMC algorithm with updates of such parameters may not perform well in BSLMM; however, it should be investigated in future studies. Informative priors are in practice difficult to elicit, and there seems to be no consensus on how to construct non-informative reference priors for these models (Diggle *et al.*, 1998; Christensen, 2004).

Acknowledgements

This work was partially funded and supported by grant MTM2010-14961 from the Spanish Ministry of Science and Education, Carolina Foundation, Applied Statistics in Experimental Research, Industry and Biotechnology (Universidad Nacional de Colombia) and Core Spatial Data Research (Faculty of Engineering, Universidad Distrital Francisco José de Caldas). We thank the AE and two reviewers for providing helpful comments on an earlier version of the manuscript.

Appendix A: Spectral representation

In this section, we show the spectral representation for the z_1 and z_2 processes. According to Royle and Wikle (2005), the z_j's ($j = 1, 2$) processes can be expanded in terms of Fourier basis functions (i.e. sines and cosines). Furthermore, assuming that the spatial processes are defined on a regular lattice, for locations s_i, $i = 1, \ldots, n$, and spatial frequencies $w_q = q/n$ for $q = 0, \ldots, n/2$ (n even), we have

$$z_j(\boldsymbol{s}_i) = \sum_{q=0}^{n/2} \delta_j^{(1)}(q)\cos\left(2\pi \boldsymbol{s}_i w_q\right) + \sum_{q=1}^{n/2-1} \delta_j^{(2)}(q)\sin\left(2\pi \boldsymbol{s}_i w_q\right)$$

$$= \left(\boldsymbol{\psi}_i^{(1)}\right)^t \boldsymbol{\delta}_j^{(1)} + \left(\boldsymbol{\psi}_i^{(2)}\right)^t \boldsymbol{\delta}_j^{(2)} = \boldsymbol{\psi}_i^t \boldsymbol{\delta}_j, \qquad j = 1, 2$$

where $\boldsymbol{\psi}_i^{(1)} = (\psi_i^{(1)}(w_0), \ldots, \psi_i^{(1)}(w_{n/2}))^t$, $\boldsymbol{\psi}_i^{(2)} = (\psi_i^{(2)}(w_1), \ldots, \psi_i^{(2)}(w_{n/2-1}))^t$, $\psi_i^{(1)}(w_q) = \cos(2\pi \boldsymbol{s}_i w_q)$, $\psi_i^{(2)}(w_q) = \sin(2\pi \boldsymbol{s}_i w_q)$, $\boldsymbol{\psi}_i = ((\boldsymbol{\psi}_i^{(1)})^t, (\boldsymbol{\psi}_i^{(2)})^t)^t$, $\boldsymbol{\delta}_j^{(1)} = (\delta_j^{(1)}(0), \ldots, \delta_j^{(1)}(n/2))^t$, $\boldsymbol{\delta}_j^{(2)} = (\delta_j^{(2)}(1), \ldots, \delta_j^{(2)}(n/2-1))^t$ and $\boldsymbol{\delta}_j = ((\boldsymbol{\delta}_j^{(1)})^t, (\boldsymbol{\delta}_j^{(2)})^t)^t$.

It is well known that for second-order stationary random processes, the $\delta_{(q)}$'s coefficients are nearly uncorrelated, and their variances at a given frequency are approximately equal to one-half the power spectral density function at that frequency, except for frequencies w_0 and $w_{n/2}$ at which the variance is equal to the associated power spectral density function (Shumway and Stoffer, 2000). Thus, for $j = 1, 2$, assuming $\boldsymbol{z}_j$ is a second-order stationary process with covariance matrix $\boldsymbol{\Sigma}_{z_j} = \sigma_1^2 \boldsymbol{R}_j$, where $\boldsymbol{R}_j$ is the correlation matrix, then $\mathbb{C}\mathrm{ov}(\boldsymbol{\delta}_j) \approx \sigma_j^2 \boldsymbol{C}_j$, where $\boldsymbol{C}_j$ is a diagonal matrix with the diagonal elements given by $[f_j(w_0), \frac{1}{2}f_j(w_1), \ldots, \frac{1}{2}f_j(w_{n/2}), \frac{1}{2}f_j(w_1), \ldots, \frac{1}{2}f_j(w_{n/2-1})]$, where $f_j(w_q)$ is the spectral density function at frequency w_q corresponding to the correlation function used to construct $\boldsymbol{R}_j$.

In this case, we parameterize the correlation matrix $\boldsymbol{R}_j = \boldsymbol{R}(\boldsymbol{\theta}_j)$, $j = 1, 2$, in terms of a spatial dependence vector parameters $\boldsymbol{\theta}_j$. Thus, the diagonal matrix $\boldsymbol{D}(\boldsymbol{\theta}_j)$ is also a function of $\boldsymbol{\theta}_j$. In the analysis presented here, a Matérn covariance matrix with the associated spectral density function at frequency w is given by (Royle and Wikle, 2005)

$$f_j(w) = \frac{2^{\kappa_j - 1}\sigma_j \Gamma\left(\kappa_j + d_w/2\right)\vartheta_j^{2\kappa_j}}{\pi^{d_w/2}\left(\vartheta_j^2 + w^2\right)^{\kappa_j + d_w/2}}, \quad \sigma_j > 0, \vartheta_j > 0, \kappa_j > 0, \quad j = 1, 2 \qquad \text{(A.1)}$$

where d_w is the dimensionality of the spatial process (Stein, 1999, p. 49), κ_j is related to the degree of smoothness of the jth spatial process, ϑ_j is related to the jth correlation range and σ_j is proportional to the jth variance of the process (Stein, 1999, p. 48). Thus, if one chooses $\boldsymbol{\Psi}_j$ to be the Fourier basis functions, then Equation A.1 suggests the form of $\boldsymbol{\Sigma}_{\delta_j}(\boldsymbol{\theta}_j)$ (diagonal matrix, with the diagonal element corresponding to frequency w given by Equation A.1).

Appendix B: Interpolation of random effects using unbiased linear predictors

The minimization of the mean squared error of the prediction can be performed through

$$\mathbb{E}\left[\left(\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j^0\right)^t \mathsf{A}_j \left(\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j^0\right)\right] = \int\int \left(\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j^0\right)^t \mathsf{A}_j \left(\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j^0\right) f\left(\boldsymbol{\eta}_j^0, \boldsymbol{\eta}_j\right) d\boldsymbol{\eta}_j^0 d\boldsymbol{\eta}_j, \qquad j = 1, 2 \qquad \text{(B.1)}$$

where $f(\boldsymbol{\eta}_j^0, \boldsymbol{\eta}_j)$ is the joint density function of $\boldsymbol{\eta}_j^0$ and $\boldsymbol{\eta}_j$ and A_j is a positive definite symmetric matrix.

Using Equation 12, the left-hand side of Equation B.1 can be written as

$$\begin{aligned} \boldsymbol{q}_j =& \mathbb{E}\left[\left(\boldsymbol{h}_j + \boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)^t \mathsf{A}_j \left(\boldsymbol{h}_j + \boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)\right] \\ =& \boldsymbol{h}_j^t\mathsf{A}_j\boldsymbol{h}_j + 2\boldsymbol{h}_j^t\mathsf{A}_j\mathbb{E}\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right) + \mathbb{E}\left[\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)^t \mathsf{A}_j \left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)\right], \quad j = 1, 2 \end{aligned} \tag{B.2}$$

Differentiating partially Equation B.2 with respect to $\boldsymbol{h}_j$ and equating to $\mathbf{0}$, we have

$$\begin{aligned} \frac{\partial \boldsymbol{q}_j}{\partial \boldsymbol{h}_j} =& 2\mathsf{A}_j\left[\boldsymbol{h}_j + \mathbb{E}\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)\right] = \mathbf{0} \\ \boldsymbol{h}_j =& -\mathbb{E}\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right), \qquad j = 1, 2 \end{aligned} \tag{B.3}$$

Substituting Equation B.3 into Equation B.2 gives

$$\begin{aligned} \boldsymbol{q}_j =& -\left[\mathbb{E}\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_l^0\right)\right]^t \mathsf{A}_j\mathbb{E}\left[\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)\right] + \mathbb{E}\left[\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)^t \mathsf{A}_j \left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)\right] \\ =& \operatorname{tr}\left\{\mathsf{A}_j\mathbb{V}\mathrm{ar}\left(\boldsymbol{Q}_j\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^0\right)\right\} \\ =& \operatorname{tr}\left\{\mathsf{A}_j\left[\boldsymbol{Q}_j\mathbb{V}\mathrm{ar}(\boldsymbol{\eta}_j)\boldsymbol{Q}_j^t + \mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_j^0\right) - \boldsymbol{Q}_j\mathbb{C}\mathrm{ov}\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right) - \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)\boldsymbol{Q}_j^t\right]\right\}, \quad j = 1, 2 \end{aligned} \tag{B.4}$$

where $\mathbb{V}\mathrm{ar}(\boldsymbol{\eta}_j) = \sigma_j^2(\boldsymbol{R}(\vartheta_j) + \tau_{R_j}^2\boldsymbol{I}_n)$, $\mathbb{V}\mathrm{ar}(\boldsymbol{\eta}_j^0)$ is the covariance matrix of $\boldsymbol{\eta}_j^0$ and $\mathbb{C}\mathrm{ov}(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0)$ is the cross-covariance matrix between $\boldsymbol{\eta}_j$ and $\boldsymbol{\eta}_j^0$.

We now wish to minimize Equation B.4 with respect to $\boldsymbol{Q}_j$. To do this, ignore A_j and $\mathbb{V}\mathrm{ar}(\boldsymbol{\eta}_j^0)$ because they do not involve $\boldsymbol{Q}_j$. Then, differentiating partly Equation B.4 with respect to $\boldsymbol{Q}_j$ and equating to $\mathbf{0}$, we have

$$\begin{aligned} \frac{\partial \boldsymbol{q}_j}{\partial \boldsymbol{Q}_j} =& 2\boldsymbol{Q}_j\mathbb{V}\mathrm{ar}(\boldsymbol{\eta}_j) - 2\mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right) = \mathbf{0} \\ \boldsymbol{Q}_j =& \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)\left[\mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_j\right)\right]^{-1}, \qquad j = 1, 2 \end{aligned} \tag{B.5}$$

Thus, by substituting Equations B.3 and B.5 in Equation 12, the pseudo best linear unbiased predictor is given by

$$\tilde{\boldsymbol{\eta}}_j = \mathbb{E}(\boldsymbol{\eta}_j^0) + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)\left[\mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_j\right)\right]^{-1}\left[\boldsymbol{\eta}_j - \mathbb{E}\left(\boldsymbol{\eta}_j\right)\right], \qquad j = 1, 2 \tag{B.6}$$

Specifically, the prediction for the SMM and SVDM is given, respectively, by

$$\begin{aligned} \tilde{\boldsymbol{\eta}}_1 =& \boldsymbol{V}_s^0\boldsymbol{\zeta} + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_1^0\right)\left[\sigma_1^2\left(\boldsymbol{R}\left(\vartheta_1\right) + \tau_{R_1}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\boldsymbol{\eta}_1 - \boldsymbol{V}_s\boldsymbol{\zeta}\right] \\ \tilde{\boldsymbol{\eta}}_2 =& \boldsymbol{U}_s^0\boldsymbol{\varphi} + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_2, \boldsymbol{\eta}_2^0\right)\left[\sigma_2^2\left(\boldsymbol{R}\left(\vartheta_2\right) + \tau_{R_2}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\boldsymbol{\eta}_2 - \boldsymbol{U}_s\boldsymbol{\varphi}\right] \end{aligned} \tag{B.7}$$

where $\boldsymbol{V}_s^0$ and $\boldsymbol{U}_s^0$ are matrices of explanatory variables for the n' new spatial subjects in the SMM and SVDM, respectively.

Taking expectation in Equation B.7, we obtain

$$\begin{aligned}
\mathbb{E}\left(\tilde{\boldsymbol{\eta}}_1 \mid \boldsymbol{y}_s\right) &= \boldsymbol{V}_s^0\boldsymbol{\zeta} + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_1^0\right)\left[\sigma_1^2\left(\boldsymbol{R}\left(\vartheta_1\right) + \tau_{R_1}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\mathbb{E}\left(\boldsymbol{\eta}_1 \mid \boldsymbol{y}_s\right) - \boldsymbol{V}_s\boldsymbol{\zeta}\right] \\
&\approx \boldsymbol{V}_s^0\boldsymbol{\zeta} + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_1, \boldsymbol{\eta}_1^0\right)\left[\sigma_1^2\left(\boldsymbol{R}\left(\vartheta_1\right) + \tau_{R_1}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\tilde{\mathbb{E}}_r\left(\boldsymbol{\eta}_1 \mid \boldsymbol{y}_s\right) - \boldsymbol{V}_s\boldsymbol{\zeta}\right] \\
\mathbb{E}\left(\tilde{\boldsymbol{\eta}}_2 \mid \boldsymbol{y}_s\right) &= \boldsymbol{U}_s^0\boldsymbol{\varphi} + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_2, \boldsymbol{\eta}_2^0\right)\left[\sigma_2^2\left(\boldsymbol{R}\left(\vartheta_2\right) + \tau_{R_2}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\mathbb{E}\left(\boldsymbol{\eta}_2 \mid \boldsymbol{y}_s\right) - \boldsymbol{U}_s\boldsymbol{\varphi}\right] \\
&\approx \boldsymbol{U}_s^0\boldsymbol{\varphi} + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_2, \boldsymbol{\eta}_2^0\right)\left[\sigma_2^2\left(\boldsymbol{R}\left(\vartheta_2\right) + \tau_{R_2}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\tilde{\mathbb{E}}_r\left(\boldsymbol{\eta}_2 \mid \boldsymbol{y}_s\right) - \boldsymbol{U}_s\boldsymbol{\varphi}\right]
\end{aligned}$$

where $\tilde{\mathbb{E}}_r$ is the empirical mean vector based on the sample $\boldsymbol{\eta}_j(1), \ldots, \boldsymbol{\eta}_j(r), j = 1, 2$.

The covariance matrix for the prediction in Equation B.6 is given by

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\tilde{\boldsymbol{\eta}}_j \mid \boldsymbol{y}_s\right) &= \boldsymbol{\Sigma}_j + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)\left[\sigma_j^2\left(\boldsymbol{R}\left(\vartheta_j\right) + \tau_{R_j}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\mathbb{V}\mathrm{ar}\left(\boldsymbol{\eta}_j \mid \boldsymbol{y}_s\right)\right] \\
&\quad \times \left[\sigma_j^2\left(\boldsymbol{R}\left(\vartheta_j\right) + \tau_{R_j}^2\boldsymbol{I}_n\right)\right]^{-1}\mathbb{C}\mathrm{ov}\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right) \\
&\approx \boldsymbol{\Sigma}_j + \mathbb{C}\mathrm{ov}^t\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)\left[\sigma_j^2\left(\boldsymbol{R}\left(\vartheta_j\right) + \tau_{R_j}^2\boldsymbol{I}_n\right)\right]^{-1}\left[\widetilde{\mathbb{V}\mathrm{ar}}_r\left(\boldsymbol{\eta}_j \mid \boldsymbol{y}_s\right)\right] \\
&\quad \times \left[\sigma_j^2\left(\boldsymbol{R}\left(\vartheta_j\right) + \tau_{R_j}^2\boldsymbol{I}_n\right)\right]^{-1}\mathbb{C}\mathrm{ov}\left(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0\right)
\end{aligned}$$

where $\boldsymbol{\Sigma}_j = \mathbb{V}\mathrm{ar}(\boldsymbol{\eta}_j^0) - \mathbb{C}\mathrm{ov}^t(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0)[\sigma_j^2(\boldsymbol{R}(\vartheta_j) + \tau_{R_j}^2\boldsymbol{I}_n)]^{-1}\mathbb{C}\mathrm{ov}(\boldsymbol{\eta}_j, \boldsymbol{\eta}_j^0)$ and $\tilde{\mathbb{V}\mathrm{ar}}_r$ is the covariance matrix based on the sample $\boldsymbol{\eta}_j(1), \ldots, \boldsymbol{\eta}_j(r), j = 1, 2$.

References

Atkinson, A. (1985), *Plots, transformations and regression: an introduction to graphical methods of diagnostic regression analysis*, Clarendon Press, Oxford.

Bouma, J., J. Stoorvogel, B. J. Van Alphen and H. W. G. Booltink (1999), Pedology, precision agriculture, and the changing paradigm of agricultural research, *Soil Science Society of America Journal* **63**, 1763–1768.

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

Capeche, C. L., J. R. Macedo, H. R. H. Manzatto and E. F. Silva (1997), Caracterização pedológica da fazenda angra-pesagro/rio, Informação, globalização, uso do solo, Rio de Janeiro, Estação experimental de Campos.

Cepeda, E. C. (2001), Variability modelling in generalized linear models, Unpublished Ph.D. thesis, Mathematics Institute, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

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

Christensen, O. F., P. J. Diggle and P. J. Ribeiro (2001), Analysing positive-valued spatial data: the transformed Gaussian model, in: Monestiez, P., D. Allard and R. Froidevaux (eds), geoENV III – geostatistics for environmental applications, Boston, Kluwer Academic Publishers, 287–298.

Crainiceanu, C., D. Ruppert, G. Claeskens and M. P. Wand (2005), Exact likelihood ratio tests for penalized splines, *Biometrika* **92**, 91–103.

Diggle, P., A. Tawn and R. Moyeed (1998), Model based geostatistics, *Applied Statistics* **49**, 299–350.

Diggle, P. J. and P. J. Ribeiro (2007), *Model-based geostatistics*, Springer Series in Statistics, New York.

Ferrari, S. and F. Cribari-Neto (2004), Beta regression for modelling rates and proportions, *Journal of Applied Statistics* **31**, 799–815.

Galvis, D. M., D. Badyophadyay and V. H. Lachos, (2013), Augmented mixed beta regression models for periodontal proportion data, *6*, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Brazil.

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

Geyer, C. J. and E. A. Thompson (1992), Constrained Monte Carlo maximum likelihood for dependent data (with discussion), *Journal of the Royal Statistical Society, Series B* **54**, 657–699.

Godwin, R. J. and P. C. H. Miller (2003), A review of the technologies for mapping within-field variability, *Biosystems Engineering* **84**(4), 393–407.

Goovaerts, P. (1998), Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties, *Biology and Fertility of Soils* **27**(4), 315–334.

Højbjerre, M. (2003), Profile likelihood in directed graphical models from BUGS output, *Statistics and Computing* **13**, 57–66.

Lark, R. M. and T. F. A. Bishop (2007), Cokriging particle size fractions of the soil, *European Journal of Soil Science* **58**, 763–774.

Mardia, K. V. and A. J. Watkins (1989), On multimodality of the likelihood in the spatial linear model, *Biometrika* **76**, 289–296.

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

McCulloch, C. E., S. R. Searle and J. M. Neuhaus (2008), *Generalized, linear, and mixed models* second, John Wiley & Sons, New Jersey.

Pauwels, J. M., E. Van-Ranst, M. Verloo and A. Mvondo-Ze (1992), *Manuel de laboratoire de pédologie. méthodes d'analyses de sols et de plantes, equipement, gestion de stocks de verrerie et de produits chimiques* et Centre Universitaire de Dschang, A. (ed.), Publications Agricoles, Bruxelles.

Royle, J. A. and C. K. Wikle (2005), Efficient statistical mapping of avian count data, *Environmental and Ecological Statistics* **12**(2), 225–243.

Shumway, R. H. and D. S. Stoffer (2000), *Time series analysis and its applications*, Springer-Verlag, New York.

Simas, A. B., W. Barreto-Souza and A. V. Rocha (2010), Improved estimators for a general class of beta regression models, *Computational Statistics & Data Analysis* **54**(2), 348–366.

Smithson, M. and J. Verkuilen (2006), A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables, *Psychological Methods* **11**(1), 54–71.

Stein, A. and C. Ettema (2003), An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons, *Agriculture, Ecosystems & Environment* **94**(1), 31–47.

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

Stram, D. O. and J. W. Lee (1994), Variance components testing in the longitudinal mixed effects setting, *Biometrics* **50**(4), 1171–1177.

Yemefack, M., D. G. Rossiter and R. Njomgang (2005), Multi-scale characterization of soil variability within an agricultural landscape mosaic system in southern Cameroon, *Geoderma* **125**(1–2), 117–143.

Zhang, D. and X. Lin (2003), Hypothesis testing in semiparametric additive mixed models, *Biostatistics* **4**, 57–74.

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

Received: June 2013. Revised: September 2014.