Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics and Data Analysis 52 (2008) 3749–3764
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Bayesian spatial prediction of the site index in the study of the Missouri Ozark Forest Ecosystem Project

Xiaoqian Sun[a,*], Zhuoqiong He[b], John Kabrick[c]

[a] *Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA*
[b] *Department of Statistics, University of Missouri, Columbia, MO 65211, USA*
[c] *USDA Forest Service, North Central Research Station, Columbia, MO 65211, USA*

Received 27 June 2007; received in revised form 26 December 2007; accepted 27 December 2007
Available online 9 January 2008

Abstract

This paper presents a Bayesian spatial method for analysing the site index data from the Missouri Ozark Forest Ecosystem Project (MOFEP). Based on ecological background and availability, we select three variables, the aspect class, the soil depth and the land type association as covariates for analysis. To allow great flexibility of the smoothness of the random field, we choose the Matérn family as the correlation function. We adopt the reference prior as an appropriate prior because there is no previous knowledge of the parameters in the model. An efficient algorithm based on the generalized Ratio-of-Uniforms method is developed for the posterior simulation. One advantage of the algorithm is that it generates independent samples from the required posterior distribution, which is much more efficient for both statistical inference of the parameters and prediction of the site indexes at unsampled locations. Our results show that the aspect class and the soil depth are both significant while the land type association is less significant. The model validation is briefly discussed. In addition, our simulation method allows easy realization for computing quantities from the posterior predictive distributions.

1. Introduction

The Missouri Ozark Forest Ecosystem Project (MOFEP) is an ongoing, centuries-long experiment that is designed to monitor and assess the short- and long-term effects of common management practices on Ozark ecosystems. See Brookshire and Shifley (1997), Shifley and Brookshire (2000) and Shifley and Kabrick (2002). The MOFEP will provide a comprehensive evaluation of the impacts of operational management practices on a wide array of ecosystem attributes. The main purpose of this paper is to predict the site index at unmeasured locations.

Site index is a measure of forest productivity based upon the height at a specified base age of dominant or codominant trees. Height is used because it is correlated to site quality and is not very sensitive to differences in stand density. In Missouri, a base age of 50 years is used. Most measured trees are not at the base age, so curves are used to convert heights of tree of any age to the base age height. A compendium of the published site index curves for eastern half of the United States was presented by Carmean et al. (1989).

* Corresponding author. Tel.: +1 864 6565238; fax: +1 864 6565230.
E-mail address: xsun@clemson.edu (X. Sun).

0167-9473/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.12.019

From 1993 and 1996, site index was determined on suitable trees at 648 permanent plots on the nine MOFEP sites. Trees considered suitable were canopy codominants having good form with no indication that they had been suppressed and showing the best growth potential. One to five candidate trees selected for site index determination were sampled outside of the 0.5 acre permanent vegetation plots but within 330 feet of vegetation plots. Candidate trees were also limited to four species — black oak, scarlet oak, white oak and shortleaf pine because these four species are most abundant commercial species in the region and reliable site index curves have been developed locally for them.

For each site index tree, the distance and azimuth from the geo-referenced vegetation plot centre were recorded and later used to determine the location of each site index tree. Trees were assigned a ranking of their perceived quality for indicating site index. Tree heights were measured with clinometer to the nearest feet. A single increment core was extracted at breast height and taken to the lab for age determination. Site index was determined using species, height, age at diameter beast height (d.b.h.), and published site index equations for species in the Missouri Ozarks. See, for example, McQuilkin (1974) and Nash (1978).

Since the site index can only be available on some locations, the prediction of the site index at unmeasured locations is of ecological interest in practice. Recently, Bayesian approach to analysis of spatial data has become of interest, especially when the main goal is prediction. See Handcock and Stein (1993), De Oliveira et al. (1997) and Ecker and Gelfand (1997), and to mention just a few. The main advantage of the Bayesian approach is that parameter uncertainty is fully accounted for when performing prediction and inference, even in small samples.

In this paper, we will propose Bayesian spatial model to achieve this goal. We consider black oak site index on two of the nine MOFEP sites. However, our proposed method can easily be applied to other species, and other sites at the MOFEP or elsewhere. Section 2 will deal with how to set up an appropriate Bayesian spatial model. In Section 3, an efficient algorithm based on the generalized Ratio-of-Uniforms method is developed for the posterior simulation. One advantage of the algorithm is that it generates independent samples from the required posterior distribution, which is much more efficient for both statistical inference of the parameters and prediction of the site indexes at unsampled locations. Model validation will be briefly studied in Section 4. In Section 5, we will discuss spatial prediction of the site index and concluding remarks will be given in Section 6. The proofs of two main theorems are shown in the Appendix.

2. The Gaussian model

2.1. Data structure and the likelihood

There are 173 sampled *black oaks* of high quality as judged by technicians, irregularly located in sites one and two, see Fig. 1. The location of each site index tree was used to identify geo-referenced environmental characteristics such as soil type, aspect, land type association, at the location of each site index tree, which will be partly used as covariates in our model.

We choose the Gaussian process to model our spatial data. Let $\{Z(\mathbf{s}), \mathbf{s} \in \mathcal{D}\}$, $\mathcal{D} \subseteq \mathbb{R}^2$, be the random field of interest. The data consist of n observations $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \ldots, Z(\mathbf{s}_n))'$ where $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ are known distinct sampling locations in $\mathcal{D}$. We assume that $\mathbf{Z}(\cdot)$ is a Gaussian random field with $\mathbb{E}\{Z(\mathbf{s})\} = \beta_0 + \beta_1 X_1(\mathbf{s}) + \cdots + \beta_p X_p(\mathbf{s})$ and $\mathrm{cov}\{Z(\mathbf{s}), Z(\mathbf{u})\} = \sigma^2 K_{\boldsymbol{\theta}}(\|\mathbf{s} - \mathbf{u}\|)$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)' \in \mathbb{R}^{p+1}$ are unknown regression parameters, $X_1(\mathbf{s}), \ldots, X_p(\mathbf{s})$ are known location-dependent covariates, $\sigma^2 = \mathrm{var}\{Z(\mathbf{s})\}$, and $K_{\boldsymbol{\theta}}(\|\mathbf{s} - \mathbf{u}\|) = \mathrm{corr}\{Z(\mathbf{s}), Z(\mathbf{u})\}$ is an isotropic correlation function with $\|\cdot\|$ denoting Euclidean distance, and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^c$ controlling the range of correlation and the smoothness of the random field. Thus the likelihood function of the model parameters $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$, based on the observed data $\mathbf{z} = (z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n))'$, is given by

$$L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}; \mathbf{z}) = (2\pi\sigma^2)^{-n/2} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right\} \tag{1}$$

where $\mathbf{X} = (x_{ij})$ is the known $n \times (p+1)$ matrix with its first column as $\mathbf{1} = (1, \ldots, 1)'$, assumed to have full rank, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (K_{\boldsymbol{\theta}}(\|\mathbf{s}_i - \mathbf{s}_j\|))$, assumed to be positive definite for any $\boldsymbol{\theta} \in \Theta$.

For our data set, $\mathcal{D}$ represents the area of sites one and two, and $Z(\mathbf{s})$ denotes the site index of black oak at location $\mathbf{s}$. Of 173 sampled black oaks, we choose $n = 113$ trees for modelling based on three considerations. First, the distances among 173 points located in sites one and two range from 3.32 m to 3928.96 m. In the spatial setting, if we

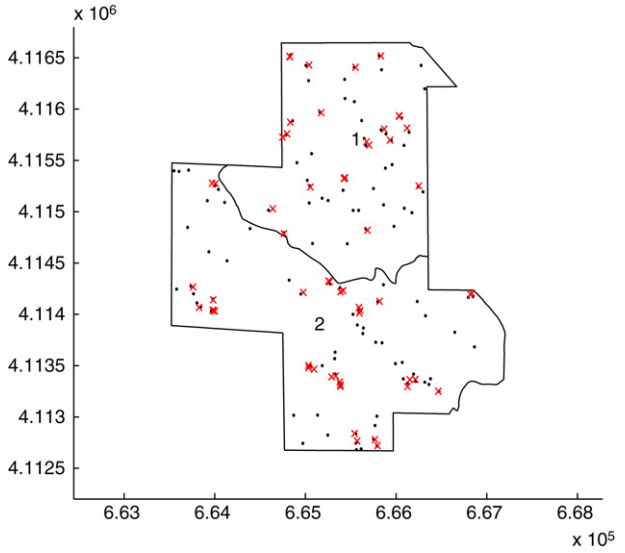

Fig. 1. 173 sampled black oaks on sites one and two of MOFEP study in UTM coordinates: Locations for modelling marked as "·" and those for model validation marked as "×".

Table 1
Summary of covariates

Name	Symbol	Type	Categories	Values and description
Aspect class	X_1 (**s**)	Categorical	2	1 — protected 0 — exposed
Soil depth	X_2 (**s**)	Categorical	2	1 — deep to very deep soil 0 — shallow to moderate deep soil
LTA	X_3 (**s**)	Categorical	2	1 — OZ9b 0 — OZ9e

add a location that is very close to an existing location, the data from the new location will not add much information to the inference about the spatial model. Second, if the model contains two very close locations, then the associated correlation matrix $\mathbf{\Sigma}_{\boldsymbol{\theta}}$ will be nearly singular, which will result in numerical difficulties for parametric inference with either the frequentist method or the Bayesian method. Third, if the model has contained enough points, then prediction at a new spatial location will improve very slowly with increasing the sample size, as mentioned by Banerjee et al. (2004). For 113 locations chosen, the minimum distance among them is 50 m. We will reserve some of the remaining 60 trees for empirical validation of the resulting predictions.

Covariates were chosen based on availability and ecological background. We selected three variables, the aspect class, the soil depth and and land type association (LTA), denoted as $X_1(\mathbf{s})$, $X_2(\mathbf{s})$ and $X_3(\mathbf{s})$ respectively. Aspect classes are described as 'protected' if the slope aspect azimuth is within 316-135 degrees, and 'exposed' if the slope aspect azimuth is within 136-315 degrees. There are two land type associations, the Current River Oak-Pine Woodland/Forest Hills (coded as OZ9b) and the Current River Oak Forest Breaks (coded as OZ9e). See Shifley and Kabrick (2002) and Nigh and Schroeder (2002) for detail. Soil depth is created by the soil type and categorized into two classes, deep to very deep soil and shallow to moderate deep soil. See the detail relationship between soil types and soil characteristics in Kabrick et al. (2000). Table 1 summarizes the covariates in the model.

2.2. *Correlation function*

The previous subsection did not specify a correlation function for the spatial model. Actually, there are many correlation functions, such as spherical, power exponential, rational quadratic, that are commonly used in spatial statistics, see Cressie (1993), Chiles and Delfiner (1999), Stein (1999) and Banerjee et al. (2004), *etc*. Amongst the

various families of correlation function which have been proposed, the Matérn family is particularly attractive. Its algebraic form is given by

$$K_{\theta,\nu}(d) = \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{d}{\theta}\right)^{\nu}\mathcal{K}_{\nu}\left(\frac{d}{\theta}\right), \quad \theta > 0, \nu > 0, \tag{2}$$

where θ is the spatial range parameter, which measures how quickly the correlations of the random field decay with distance, ν is the smoothness parameter, which measures the differentiability of the random field, and $\mathcal{K}_{\nu}(\cdot)$ is the modified Bessel function of the second kind and of order ν; see Abramowitz and Stegun (1965) for details on the behavior of this special function. What makes this family particularly attractive is that the corresponding process $Z(\cdot)$ is mean-square $[\nu] - 1$ times differentiable where $[\nu]$ represents the largest integer less than or equal to ν. So the Matérn family does allow for considerable flexibility in the smoothness of the random field while keeping the number of parameters manageable. Furthermore, note that the exponential family is the special case with $\nu = 1$ and the Gaussian family is the case when $\nu \to \infty$.

The Matérn family was strongly recommended by Stein (1999) because there is the parameter ν to control the smoothness of the random field. We choose this family as the correlation function for our model.

2.3. The prior

The parameters in our model are the regression coefficient $\boldsymbol{\beta}$, the variance σ^2, the range parameter θ and the smoothness parameter ν. The smoothness of a random field plays a critical role in spatial data analysis. However, there is often no basis for knowing a priori of the degree of some physical process in a random field. Thus, we will not assign a prior for the smoothness parameter ν and we assume for the moment that it is known. We will discuss how to estimate the smoothness parameter ν before we make Bayesian inference on the parameters $\boldsymbol{\beta}$, σ^2 and θ.

Selection of the prior is based upon previous knowledge of the model parameters. Often, there is little information available on the model parameters which prompts the use of noninformative priors. In this paper, we consider the following reference prior for $(\boldsymbol{\beta}, \sigma^2, \theta)$ developed by Berger et al. (2001),

$$\pi(\boldsymbol{\beta}, \sigma^2, \theta) \propto \frac{\pi(\theta)}{\sigma^2}, \qquad \boldsymbol{\beta} \in R^{p+1}, \sigma^2 > 0, \theta > 0, \tag{3}$$

where

$$\pi(\theta) \propto \left\{\mathrm{tr}[\mathbf{W}_{\theta}^2] - \frac{1}{n-p}(\mathrm{tr}[\mathbf{W}_{\theta}])^2\right\}^{1/2}, \tag{4}$$

with

$$\mathbf{W}_{\theta} = \frac{\partial \boldsymbol{\Sigma}_{\theta}}{\partial \theta}\boldsymbol{\Sigma}_{\theta}^{-1}\mathbf{P}_{\theta}^{\boldsymbol{\Sigma}} \quad \text{and} \quad \mathbf{P}_{\theta}^{\boldsymbol{\Sigma}} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\theta}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\theta}^{-1}. \tag{5}$$

Here $(\partial/\partial\theta)\boldsymbol{\Sigma}_{\theta}$ denotes the matrix obtained by differentiating $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ with respect to θ element by element and $\mathbf{I}$ is an identity matrix of order n.

The reference prior was originally proposed by Bernardo (1979) and it is expected to have a minimal effect on the posterior inference about the unknown parameters and thus can produce "objective" Bayesian inference. Berger and Bernardo (1989, 1992) further developed crucial mathematical extensions through Fisher information matrix. Interested readers may find a nice review article written by Bernardo (2005) and the references therein.

Note that the reference prior (3) is a non-informative and improper prior and the posterior propriety under this prior has fully been studied by Berger et al. (2001). They prove that for any design matrix $\mathbf{X}$ with full rank, the posterior is always proper.

Remark 1. The prior $\pi(\boldsymbol{\beta}, \sigma^2, \theta)$ in (3) with $\pi(\theta) = 1$ was proposed by Kitanidis (1986) and Handcock and Stein (1993). However, as shown by Berger et al. (2001), it will result in an improper posterior.

3. Statistical inference for parameters

3.1. Estimating the smoothness parameter

Before making Bayesian inference on $\boldsymbol{\beta}$, σ^2 and θ, the smoothness parameter ν must be estimated.

Because we are now considering the smoothness parameter ν, it is useful to explicitly recognize that the reference prior was defined with ν considered given, so we now write $\pi(\boldsymbol{\beta}, \sigma^2, \theta|\nu)$ instead of $\pi(\boldsymbol{\beta}, \sigma^2, \theta)$, and $\pi(\theta|\nu)$ instead of $\pi(\theta)$ in this subsection. Although the reference prior (3) is improper, Berger et al. (2001) prove that the marginal prior $\pi(\theta|\nu)$ is proper. This makes it possible to apply the idea in Berger et al. (1998) to estimate the smoothness parameter ν. The procedure is as follows:

For each ν, the reference prior used is

$$\pi(\boldsymbol{\beta}, \sigma^2, \theta|\nu) = \frac{C(\nu)\pi(\theta|\nu)}{\sigma^2}, \tag{6}$$

where

$$C(\nu) = \frac{1}{\int_0^\infty \pi(\theta|\nu)\mathrm{d}\theta}.$$

Note that computation of $C(\nu)$ must be done numerically. We compute $C(\nu)$ by the function *quad* with MATLAB software. Using this prior (6), we compute the marginal integrated likelihood for each ν as

$$\begin{aligned} m(\mathbf{z}|\nu) &= \int L(\boldsymbol{\beta}, \sigma^2, \theta, \nu; \mathbf{z}) \frac{C(\nu)\pi(\theta|\nu)}{\sigma^2} \mathrm{d}\boldsymbol{\beta}\mathrm{d}\sigma^2\mathrm{d}\theta \\ &= \int_0^\infty L^I(\theta, \nu; \mathbf{z}) C(\nu)\pi(\theta|\nu)\mathrm{d}\theta \\ &= \int_0^\infty |\boldsymbol{\Sigma}_{\theta,\nu}|^{-1/2} |\mathbf{X}'\boldsymbol{\Sigma}_{\theta,\nu}^{-1}\mathbf{X}|^{-1/2} (S^2_{\theta,\nu})^{-(n-p)/2} C(\nu)\pi(\theta|\nu)\mathrm{d}\theta \end{aligned} \tag{7}$$

where

$$\begin{aligned} L^I(\theta, \nu; \mathbf{z}) &= \int_{R^{p+1}\times(0,+\infty)} L(\boldsymbol{\beta}, \sigma^2, \theta, \nu; \mathbf{z}) \frac{1}{\sigma^2} \mathrm{d}\boldsymbol{\beta}\mathrm{d}\sigma^2 \\ &\propto |\boldsymbol{\Sigma}_{\theta,\nu}|^{-1/2} |\mathbf{X}'\boldsymbol{\Sigma}_{\theta,\nu}^{-1}\mathbf{X}|^{-1/2} (S^2_{\theta,\nu})^{-(n-p)/2}, \end{aligned} \tag{8}$$

and

$$S^2_{\theta,\nu} = (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\theta,\nu})'\boldsymbol{\Sigma}_{\theta,\nu}^{-1}(\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\theta,\nu}), \tag{9}$$

$$\hat{\boldsymbol{\beta}}_{\theta,\nu} = (\mathbf{X}'\boldsymbol{\Sigma}_{\theta,\nu}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\theta,\nu}^{-1}\mathbf{z}. \tag{10}$$

Note that $S^2_{\theta,\nu}$, is the generalized residual sum of squares and $\hat{\boldsymbol{\beta}}_{\theta,\nu}$, is the generalized least squares estimator of $\boldsymbol{\beta}$ given θ and ν.

Based on the idea of Berger et al. (1998), ν can be estimated by maximizing the marginal density $m(\mathbf{z}|\nu)$, that is,

$$\hat{\nu} = \arg\max_{\nu} m(\mathbf{z}|\nu).$$

Fig. 2 shows the marginal integrated likelihood $m(\mathbf{z}|\nu)$ for our dataset. It follows that for our model, the most likely value of ν is 0.13. Thus we will assume $\nu = 0.13$ for the posterior simulation in the next subsection.

3.2. Posterior simulation

Standard Bayesian theory tells us that the posterior distribution is determined by

$$p(\boldsymbol{\beta}, \sigma^2, \theta|\mathbf{z}) \propto L(\boldsymbol{\beta}, \sigma^2, \theta; \mathbf{z})\pi(\boldsymbol{\beta}, \sigma^2, \theta) \tag{11}$$

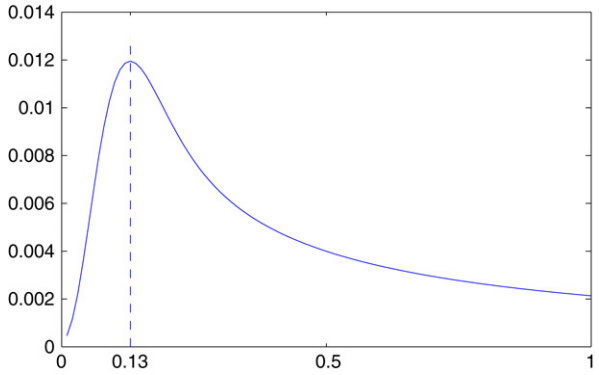

Fig. 2. The marginal density $m(\mathbf{z}|\nu)$ in terms of ν.

where $L(\boldsymbol{\beta}, \sigma^2, \theta; \mathbf{z})$ is given by (1) and $\pi(\boldsymbol{\beta}, \sigma^2, \theta)$ by (3), assuming that the smoothness parameter ν is 0.13. Berger et al. (2001) also proved that the posterior distribution $p(\boldsymbol{\beta}, \sigma^2, \theta|\mathbf{z})$ under the reference prior is proper, but did not provide an efficient algorithm for the posterior simulation. Here we will propose an efficient algorithm for the posterior simulation based on the generalized Ratio-of-Uniforms method developed by Wakefield et al. (1991). The following theorem plays an important role in our posterior simulation:

Theorem 1. *The joint posterior distribution of* $\boldsymbol{\beta}$, σ^2, θ *has the following decomposition:*

$$p(\boldsymbol{\beta}, \sigma^2, \theta|\mathbf{z}) = p(\boldsymbol{\beta}|\sigma^2, \theta; \mathbf{z})p(\sigma^2|\theta; \mathbf{z})p(\theta|\mathbf{z}), \tag{12}$$

where

$$(\boldsymbol{\beta}|\sigma^2, \theta; \mathbf{z}) \sim N_p(\hat{\boldsymbol{\beta}}_\theta, \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}) \tag{13}$$

$$(\sigma^2|\theta; \mathbf{z}) \sim IG((n-p)/2+1, \mathbf{z}'\{\boldsymbol{\Sigma}_\theta^{-1} - \boldsymbol{\Sigma}_\theta^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\}\mathbf{z}/2), \tag{14}$$

$$[\theta|\mathbf{z}] \propto L^I(\theta; \mathbf{z})\pi(\theta), \tag{15}$$

with $L^I(\theta; \mathbf{z}) = L^I(\theta, \nu; \mathbf{z})$ *given by (8) and* $\hat{\boldsymbol{\beta}}_\theta = \hat{\boldsymbol{\beta}}_{\theta,\nu}$, *by (10), assuming the smoothness parameter* ν *being known.* $IG(\alpha_1, \alpha_2)$ *represents the inverse gamma distribution with shape parameter* α_1 *and scale parameter* α_2.

Sampling from (13) and (14) is straightforward. We will apply the generalized Ratio-of-Uniforms method proposed by Wakefield et al. (1991) to sample θ from (15) because it is not of standard form.

In order to apply the generalized Ratio-of-Uniforms method to sample θ from (15), it is important to study the analytical properties of the function $L^I(\theta; \mathbf{z})\pi(\theta)$. Fig. 3 shows the plot of the integrated likelihood function $L^I(\theta; \mathbf{z})\pi(\theta)$ with $\nu = 0.13$. The following theorem is essential to generate samples from $p(\theta|\mathbf{z})$ by the generalized Ratio-of-Uniforms method.

Theorem 2. *For* $\nu \neq 1$, *there exists a positive number* r, *such that both* $[L^I(\theta; \mathbf{z})\pi(\theta)]^{1/(r+1)}$ *and* $\theta[L^I(\theta; \mathbf{z})\pi(\theta)]^{r/(r+1)}$ *are bounded on* $(0, +\infty)$.

By choosing an appropriate number r, the algorithm of the posterior simulation works as follows (notice that $\inf_\theta\{\theta[L^I(\theta; \mathbf{z})\pi(\theta)]^{r/(r+1)}\} = 0$).

Algorithm for the posterior simulation from $p(\boldsymbol{\beta}, \sigma^2, \theta|\mathbf{z})$:

Step 1*: Compute*

$$a(r) = \sup_{\theta>0}\{[L^I(\theta; \mathbf{z})\pi(\theta)]^{1/(r+1)}\},$$

$$b(r) = \sup_{\theta>0}\{\theta[L^I(\theta; \mathbf{z})\pi(\theta)]^{r/(r+1)}\}.$$

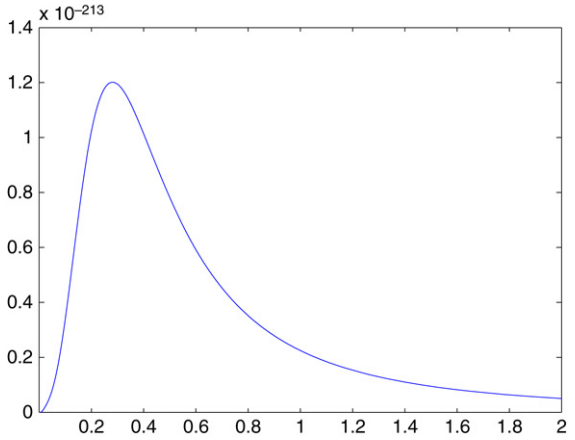

Fig. 3. The plot of $L^I(\theta;\mathbf{z})\pi(\theta)$ with $\nu = 0.13$.

Step 2*: Simulate*

$$U \sim \textit{Uniform}\,[0, a(r)],$$
$$V \sim \textit{Uniform}\,[0, b(r)],$$

and compute $\rho = V/U^r$;
Step 3*: If* $U \le [L^I(\rho;\mathbf{z})\pi(\rho)]^{1/(r+1)}$*, we accept* ρ *as a sample* θ *from* $L^I(\theta;\mathbf{z})\pi(\theta)$*, otherwise go back to Step* 2;
Step 4*: For each* θ*, simulate*

$$\sigma^2 \sim IG((n-p)/2+1, \mathbf{z}'\{\boldsymbol{\Sigma}_\theta^{-1} - \boldsymbol{\Sigma}_\theta^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\}\mathbf{z}/2);$$

Step 5*: For each* (θ, σ^2)*, simulate*

$$\boldsymbol{\beta} \sim N_{p+1}(\hat{\boldsymbol{\beta}}_\theta, \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1});$$

Step 6*: Go back to Step* 2.

As discussed above, an appropriate estimate of ν is 0.13 for our dataset, and thus we may take any $r > 0.5747$ from Theorem 2. For simplicity, we choose $r = 1$, which means that the basic Ratio-of-Uniforms method proposed by Kinderman and Monahan (1977) can be applied. In this case, $a(1) = 0.1349$ and $b(1) = 0.0584$. The acceptance area with the Ratio-of-Uniforms method is shown in Fig. 4 and the theoretical acceptance rate for the simulation is around 76%. In fact, we obtained 10 000 samples by sampling 13 101 pairs of (u, v), which means the actual acceptance rate is 76.33% in our simulation. The simulation took about 52 min on a 2.20 GHz AMD Athlon XP 3200+ PC. In addition, notice that our simulation produces independent samples for $(\boldsymbol{\beta}, \sigma^2, \theta)$, which is advantageous over other traditional MCMC algorithms when making inference of parameters or prediction.

The acceptance rate of the generalized Ratio-of-Uniforms method depends on the value r chosen. It is unclear how to get the best choice of r. Our experience shows that a good choice is the integer that is closest to $1/(2|\nu - 1|)$ and greater than or equal to 1. If the smoothness parameter ν is exactly equal to one, it is still unclear whether the generalized Ratio-of-Uniforms method can be applied theoretically.

Fig. 5 gives the histogram of θ based on these 10 000 independent samples. We can see that the marginal posterior density of the range parameter θ is positively skewed and heavy-tailed, which commonly appears in the area of spatial statistics. This property usually results in difficulties in spatial simulation because it is often difficult to get samples in the tail. See Banerjee et al. (2004) and Møller (2003) and the references therein. The histograms of σ^2 and $\beta_0, \beta_1, \beta_2, \beta_3$ are shown in Fig. 6 and Fig. 7, respectively. Table 2 shows some posterior quantities from the simulation.

From Table 2, we know that both the aspect class and the soil depth are significant, but the land type association is less significant in the model. Therefore, we consider the new spatial model with two covariates, the aspect class and the soil depth. Similar simulation shows that the aspect class and the soil depth are still significant in the new model and thus the new model with the aspect class and the soil depth only is appropriate for further analysis such as

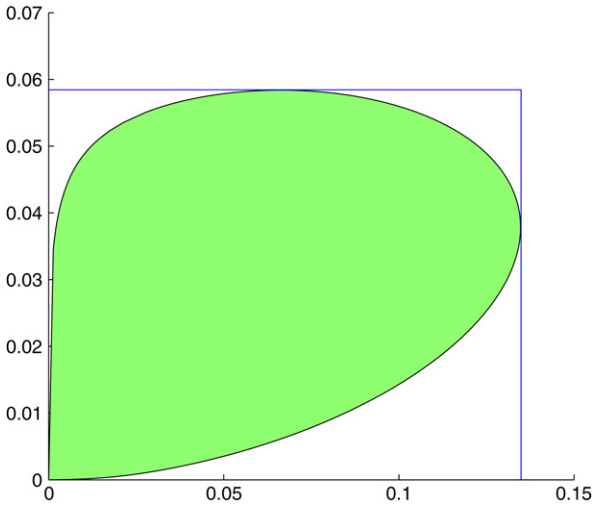

Fig. 4. The acceptance area with the Ratio-of-Uniforms method.

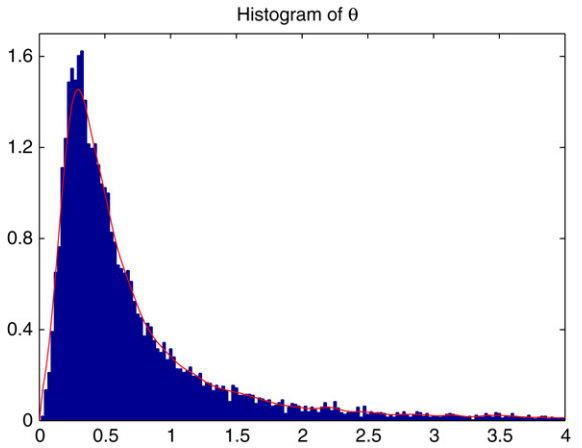

Fig. 5. Histogram for θ in the model.

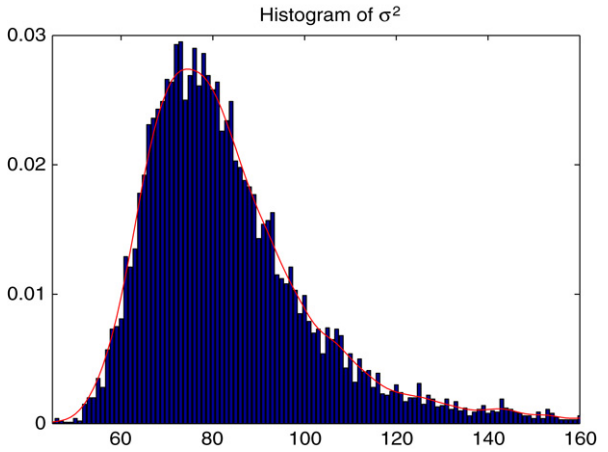

Fig. 6. Histogram for σ^2 in the model.

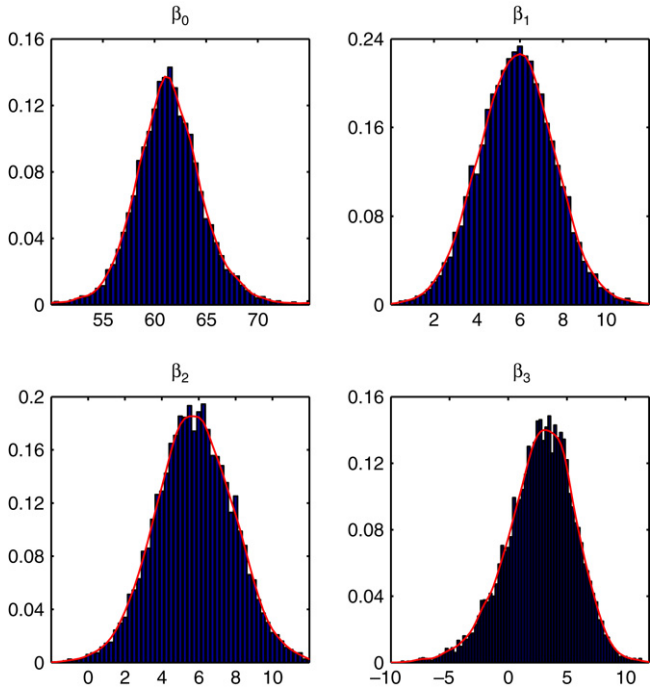

Fig. 7. Histograms for $\beta_0, \beta_1, \beta_2, \beta_3$ in the model.

Table 2
Posterior quantities for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$, σ^2 and θ

Parameter	Mode	Median	Mean	Standard deviation	95% Bayesian credible interval
β_0	61.32	61.34	61.40	3.48	[54.91, 68.36]
β_1	5.86	5.86	5.84	1.74	[2.44, 9.27]
β_2	5.49	5.48	5.47	2.11	[1.40, 9.59]
β_3	2.94	2.93	2.73	2.92	[−3.65, 7.93]
σ^2	78.06	79.99	84.88	22.70	[58.34, 135.54]
θ	0.28	0.50	1.20	10.78	[0.12, 5.37]

prediction. For brevity, we just present the histograms of the parameters β_1, β_2, θ and σ^2 in Fig. 8 based on 10 000 independent posterior samples. Notice that for the new model, the estimated smoothness parameter in Section 3.1 becomes 0.12.

Fig. 8 also tells us that the approximate mode of the marginal posterior density of the range parameter θ is around 0.30, which is smaller than the posterior median 0.49. As discussed in Sun (2006), we may choose the posterior mode as an appropriate point estimate of the range parameter θ for our model. This implies that the *effective range*, the distance at which the correlation drops to only 0.05, is about 460 m, which is a reasonable distance in the MOFEP study.

In the following sections, we work with the model with the aspect class and the soil depth only because the land type association is not significant.

4. Model validation

Assessing model adequacy is very important and fundamental in Bayesian data analysis, since the analysis can be misleading when the model is not adequate. The literature on Bayesian model adequacy is very extensive; for example, see Box (1980), Geisser (1993), Gelfand et al. (1992), Dey et al. (1997), and many others. When the observations are independent, the cross-validation approach, in which the predictive distribution is used in various ways to assess model adequacy, are popularly used. The main idea of this cross-validation approach is to validate conditional predictive

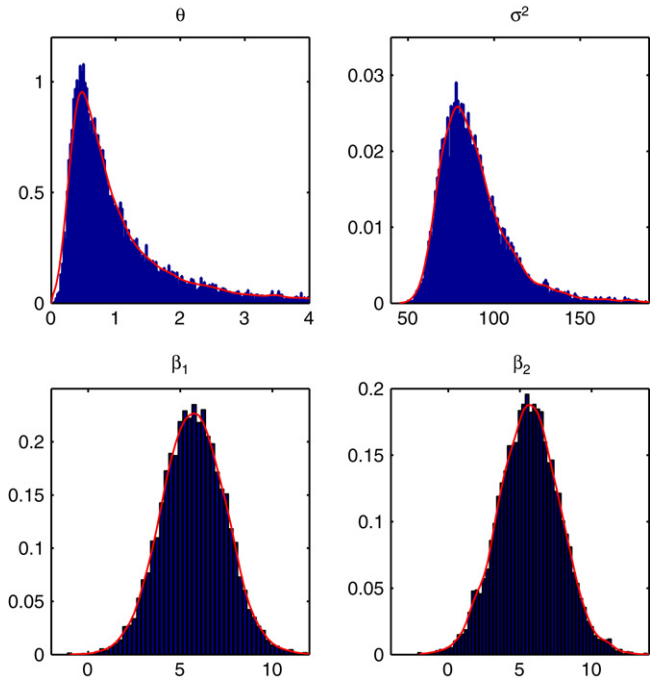

Fig. 8. Histograms for θ, σ^2, β_1, β_2 in the new model with two covariates, the aspect class and the soil depth based on 10 000 independent posterior samples.

distribution arising from single observation deletion against observed responses. In the area of spatial statistics, the observations measured at different locations are correlated. If the sample size is small, this approach is still applicable, for example, see De Oliveira et al. (1997). However, this approach is not appropriate for a large spatial dataset because of very expensive computation. Now we will use some of 60 sampled trees that are not contained in the model for assessing model adequacy.

From the 60 sampled trees that are not used in building the model, we selected 29 trees, such that the distance between each of 29 trees and any of 113 trees in the model is at least 20 m. We then obtain the predictive distribution of the site index at each of 29 locations as follows:

$$p(z_0|\mathbf{z}) = \int p(z_0|\mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta)\, p(\boldsymbol{\beta}, \sigma^2, \theta|\mathbf{z}) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\sigma^2 \mathrm{d}\theta \tag{16}$$

where $p(\boldsymbol{\beta}, \sigma^2, \theta|\mathbf{z})$ is the posterior distribution described by (11) and $p(z_0|\mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta)$ is determined by

$$p(\mathbf{z}, z_0|\boldsymbol{\beta}, \sigma^2, \theta) \sim N_{n+1}\left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{x}'(\mathbf{s}_0)\boldsymbol{\beta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{\Sigma}_\theta & \mathbf{k}_\theta \\ \mathbf{k}'_\theta & 1 \end{pmatrix}\right). \tag{17}$$

Here $\mathbf{x}(\mathbf{s}_0)$ is the covariate vector evaluated at $\mathbf{s}_0$ and $\sigma^2\mathbf{k}_\theta$ is $n \times 1$ vector of covariances of z_0 with $(z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n))$. In fact, we have

$$p(z_0|\mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta) \sim N(\mu_0, \sigma_0^2) \tag{18}$$

where

$$\mu_0 = \mathbf{x}'(\mathbf{s}_0)\boldsymbol{\beta} + \mathbf{k}'_\theta \boldsymbol{\Sigma}_\theta^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}), \quad \sigma_0^2 = \sigma^2(1 - \mathbf{k}'_\theta \boldsymbol{\Sigma}_\theta^{-1}\mathbf{k}_\theta). \tag{19}$$

Based on the samples generated from the posterior distribution $p(\boldsymbol{\beta}, \sigma^2, \theta|\mathbf{z})$ in the previous section, we can easily obtain the samples from the predictive distribution and then calculate 95% Bayesian credible interval for each z_0 at location s_0. The detailed algorithm will be described in the next section. Of 29 locations for the purpose of model validation, four observed measurements are outside their corresponding 95% Bayesian credible intervals. If

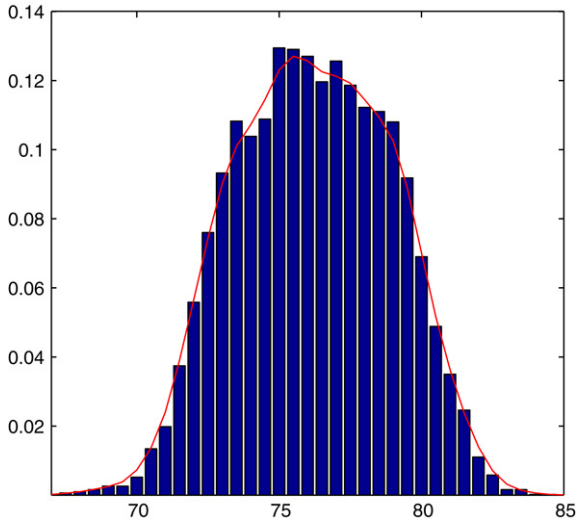

Fig. 9. Histogram of z_0 sampling from the predictive distribution at (665 614.875, 4115 457.5), covariate information: aspect class — exposed and soil depth — deep.

we consider 90% Bayesian credible prediction intervals, then only two measurements are outside their corresponding prediction intervals. In addition, no overprediction or underprediction tendency is noted. So the model in Section 2 with two covariates, the aspect class and the soil depth, seems to perform adequately. We will also justify our model by discussing the variability of the predictive distribution in the next section.

5. Spatial prediction of the site index

Modelling point-referenced data is not only useful for identifying significant covariates but for producing smooth maps of the outcome by predicting it at unsampled locations. Spatial prediction is usually refereed as kriging and popularly used in spatial area. In the classical framework a lot of effort is devoted to the determination of the optimal estimates to use in the predictive equation. However, as pointed out by Le and Zidek (1992), classical kriging methodology fails to incorporate parameter uncertainty when performing prediction and inference. This deficiency leads to unwarranted confidence in the interpolated values and, essentially, to seemingly valid decisions or regulatory actions which are, in fact, unjustified. Bayesian approaches to spatial interpolation avoid this deficiency by considering uncertainty about the parameters in the model.

For each unmeasured site $\mathbf{s}_0$, we obtain its covariate information from the GIS database provided by the Missouri Department of Conservation. Given samples from the posterior distribution, simulation of realizations from (16) is straightforward. In fact, by (16) and (18), we just need to add

Step $5^\star$*: for given* $(\boldsymbol{\beta}, \sigma^2, \theta)$, *simulate* z_0 *from* $N(\mu_0, \sigma_0^2)$ *with* μ_0, σ_0^2 *given by* (19);

before *Step* 6 of the algorithm in Section 3.2 and thus we get independent samples of z_0 from the posterior predictive distribution $p(z_0|\mathbf{z})$ in (16).

Fig. 9 shows the histogram of the z_0 at (665614.875, 4115457.5), which is located at the centre of Site one. Fig. 10 shows the histograms of the z_0 at other four locations, which are near the border of Site one. Each histogram is based on a sample of size 10 000. In fact, we found that the sample size 300 is sufficient enough to get the approximate predictive density of z_0. Note that each posterior predictive distribution is unimodal and approximately symmetric, which implies that its mean, median and mode are almost the same. In addition, the variation of each predictive distribution is relatively small compared to the posterior distribution of the variance σ^2. For example, the standard deviation of the predictive distribution at (665564.875, 4115337.5) is about 2.76 for our 10 000 samples. This also gives us much confidence that our proposed model is appropriate.

In order to get a prediction map of the site index, we make a grid of 10 m by 10 m on Site one providing 38 744 unmeasured locations. Figs. 11 and 12 present the covariate information for the aspect class and the soil depth in Site one, respectively. Based on the prediction procedure described in this section, Fig. 13 shows the prediction map for

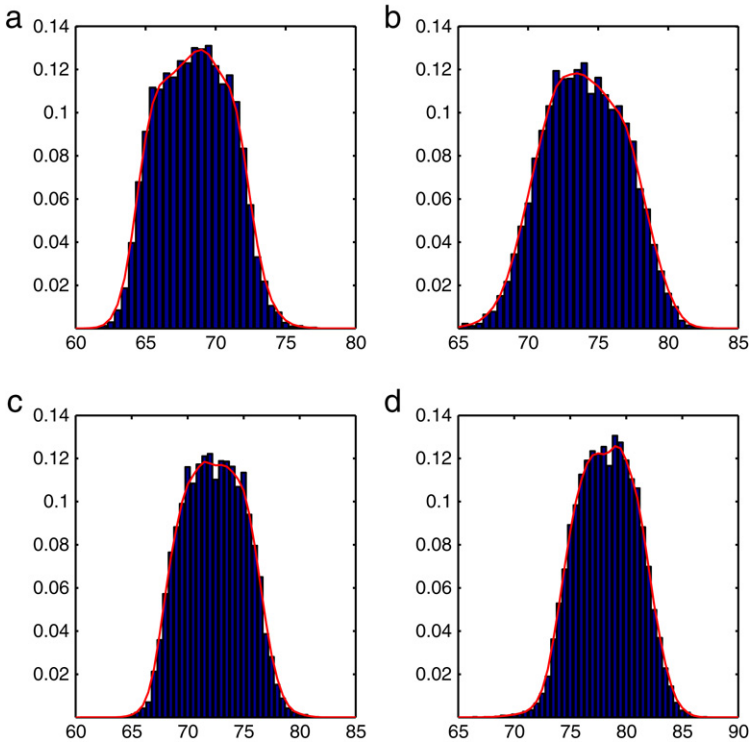

Fig. 10. Histograms of the predictive distribution at four different locations: (a) Location (666 194.875, 4116 397.5), covariate information: aspect class—exposed and soil depth—deep to very deep; (b) location (666 124.875, 4114 517.5), covariate information: aspect class—exposed and soil depth—deep to very deep; (c) location (664 474.875, 4115 407.5), covariate information: aspect class—protected and soil depth—shadow to deep; (d) location (665 564.875, 4115 337.5), covariate information: aspect class—exposed and soil depth—deep to very deep.

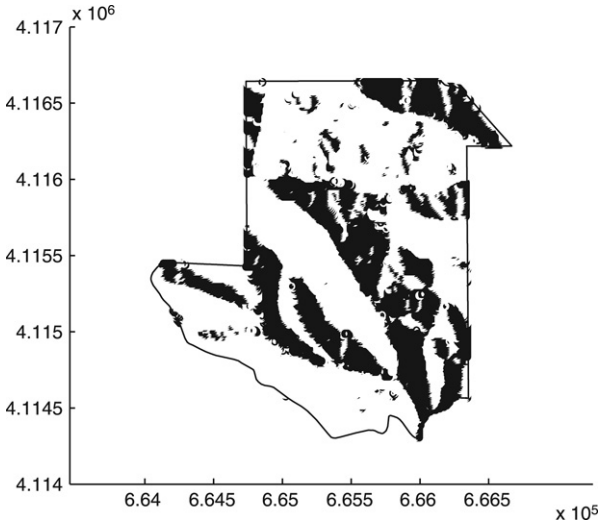

Fig. 11. The map of the aspect class in Site one: The white represents the area with protected and the black represents the area with exposed.

Site one. The whole simulation of prediction using the above method took about 75 h on a 3.20 GHz Pentium IV PC. For convenience, we also provide the map of the variability of the prediction in Fig. 14.

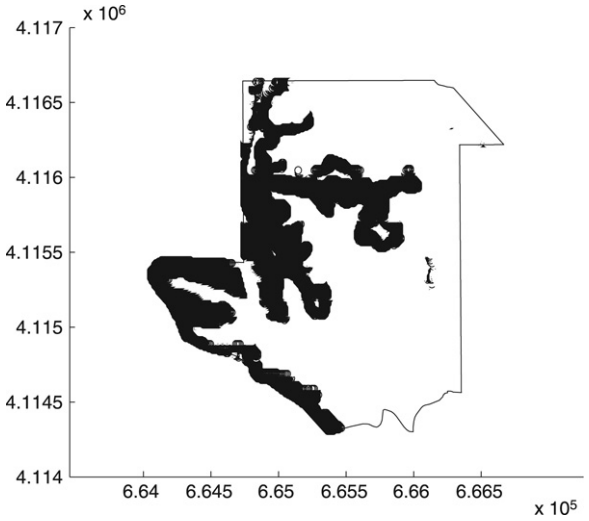

Fig. 12. The map of the soil depth in Site one: The white represents the soil depth varying from deep to very deep and the black represents the soil depth varying from shallow to deep.

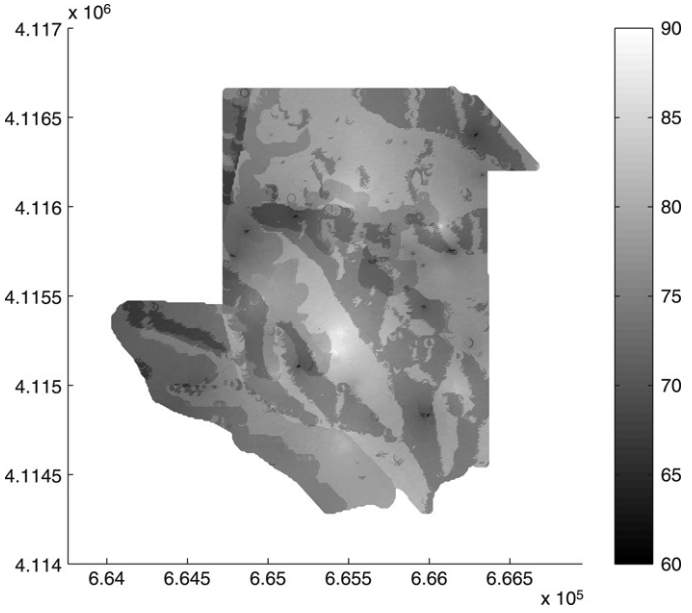

Fig. 13. The prediction map of the site index in Site one.

6. Concluding remarks and discussions

In this paper we discuss how to model the site index dataset provided by the Missouri Department of Conservation and then predict the site index at unsampled locations. As an example, we chose black oaks in sites one and two for analysis. Based on ecological background and availability, we selected three variables, the aspect class, the land type association and the soil depth as covariates. To allow great flexibility of the smoothness of the random field, we adopted the Matérn family as the correlation function. We also chose the reference prior as an appropriate prior because there is no previous knowledge of the parameters in the model and thus an appropriate Bayesian spatial model is established. An efficient algorithm based on the generalized Ratio-of-Uniforms method is developed for the posterior simulation. One advantage of the algorithm is that it generates independent samples from the required posterior distribution, which is much more efficient for both statistical inference of the parameters and prediction of

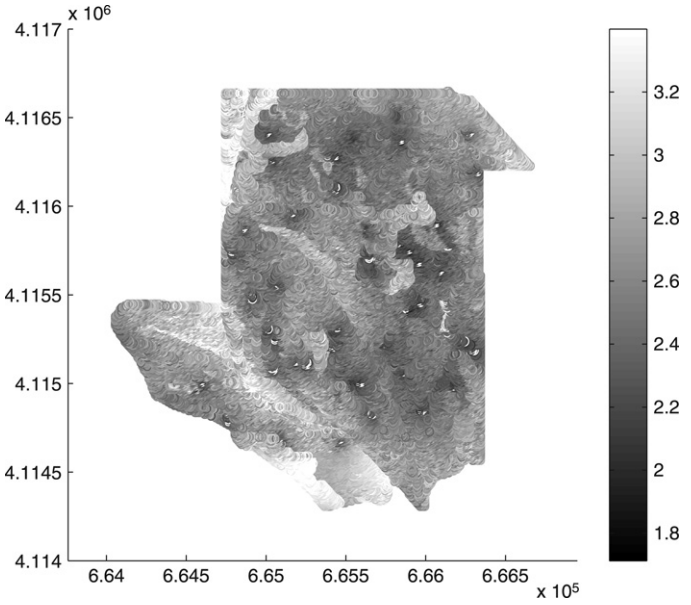

Fig. 14. The standard deviation map of the site index prediction at Site one.

the site indexes at unsampled locations. Our results show that the aspect class and the soil depth are both significant while the land type association is less significant.

One possible extension is to include the nugget effect in the model, but a new efficient algorithm is needed because there are too many unsampled locations to be predicted. We may also consider develop a multivariate Bayesian spatial model by considering four species together to enhance the prediction.

Acknowledgments

The authors wish to thank two anonymous referees for their constructive comments and suggestions. They are also grateful to the Missouri Department of Conservation for providing useful datasets for analysis.

Appendix

Proof of Theorem 1. From (1) and (3), we can easily obtain (13). Since

$$(\mathbf{z}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{z}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{z} + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}})'\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_{\theta}) - \hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}'\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}},$$

it follows

$$\begin{aligned} p(\sigma^2|\boldsymbol{\theta};\mathbf{z}) &= \int L(\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta};\mathbf{z})\pi(\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta})\mathrm{d}\boldsymbol{\beta} \\ &\propto (\sigma^2)^{-n/2-a}\int \exp\{-(\mathbf{z}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z}-\mathbf{X}\boldsymbol{\beta})/(2\sigma^2)\}\mathrm{d}\boldsymbol{\beta} \\ &\propto (\sigma^2)^{-(n-p)/2-a}\exp[-\mathbf{z}'\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\}\mathbf{z}/(2\sigma^2)] \end{aligned}$$

and hence (14). (15) can easily be seen from (8), which completes the proof. □

Proof of Theorem 2. For the Matérn family and the reference prior $\pi(\boldsymbol{\beta},\sigma^2,\theta)$, Lemma 1 in Berger et al. (2001) showed that $L^I(\theta;\mathbf{z}) \to c_0$ as $\theta \to 0^+$ where c_0 is a positive number, $L^I(\theta;\mathbf{z}) = O(1)$ as $\theta \to +\infty$ and $\pi(\theta) \to 0^+$ as $\theta \to 0$. Consider three cases in the following:

(i) $0 < \nu < 1$. Corollary 1 in Berger et al. (2001) stated that

$$\pi(\theta) \propto \frac{1}{\theta^{1+2(1-\nu)}}$$

and thus we may take any $r > 1/[2(1-\nu)]$.

(ii) ν *is greater than* 1 *and is non-integer*. Similarly, we have

$$\pi(\theta) \propto \frac{1}{\theta^{1+2(\nu-1)}}$$

and thus we may take any $r > 1/[2(\nu-1)]$.

(iii) ν *is greater than* 1 *but is integer*. We have

$$\pi(\theta) \propto \frac{1+2(\nu-1)|\log(\theta)|}{\theta^{1+2(\nu-1)}}$$

and thus we may still take $r > 1/[2(\nu-1)]$. Thus the theorem follows.

References

Abramowitz, M., Stegun, I., 1965. Handbook of Mathematical Functions, New York.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2004. Hierarchical Moddeling and Analysis for Spatial Data. Chapman & Hall/CRC.

Berger, J.O., Bernardo, J.M., 1989. Estimating a product of means: Bayesian analysis with reference priors. Journal of the American Statistical Association 84, 200–207.

Berger, J.O., Bernardo, J.M., 1992. On the development of reference priors (Disc: p49–60). In: Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting.

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

Berger, J.O., Pericchi, L.R., Varshavsky, J.A., 1998. Bayes factors and marginal distributions in invariant situations. Sankhya, Series A, Indian Journal of Statistics 60, 307–321.

Bernardo, J.M., 1979. Reference posterior distributions for Bayesian inference (C/R p128–147). Journal of the Royal Statistical Society, Series B: Methodological 41, 113–128.

Bernardo, J.M., 2005. Reference analysis. In: Dey, D.K., Rao, C.R. (Eds.), Handbook of Statistics, vol. 25. Elsevier, pp. 17–90.

Box, G.E.P., 1980. Sampling and Bayes' inference in scientific modelling and robustness (C/R: p404–430). Journal of the Royal Statistical Society, Series A, General 143, 383–430.

Brookshire, B., Shifley, S., 1997. Proceedings of the Missouri Ozark Forest Ecosystem Project Symposium: An Experimental Approach to Landscape Research (eds.). General Technical Report NC-193. The United States Department of Agriculture, Forest Service, North Central Experiment Station. St. Paul, MN.

Carmean, W.H., Hahn, J.T., Jacobs, R.D., 1989. Site index curves for forest tree species in the eastern United States. General Technical Report NC-128. USDA Forest Service, North Central Forest Experiment Station.

Chiles, J.-P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons, New York.

Cressie, N.A.C., 1993. Statistics for Spatial Data (Revised edition). Wiley-Interscience.

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

Dey, D.K., Chen, M.-H., Chang, H., 1997. Bayesian approach for nonlinear random effects models. Biometrics 53, 1239–1252.

Ecker, M.D., Gelfand, A.E., 1997. Bayesian variogram modeling for an isotropic spatial process. Journal of Agricultural, Biological, and Environmental Statistics 2, 347–369.

Geisser, S., 1993. Predictive Inference. An Introduction. Chapman & Hall Ltd.

Gelfand, A.E., Dey, D.K., Chang, H., 1992. Model determination using predictive distributions, with implementation via sampling-based methods (Disc: p160–167). In: Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting.

Handcock, M.S., Stein, M.L., 1993. A Bayesian analysis of kriging. Technometrics 35, 403–410.

Kabrick, J., Meinert, D., Nigh, T., Gorlinsky, B.J., 2000. Physical environment of the Missouri Ozark forest ecosystem project sites. In: Missouri Ozark Forest Ecosystem Project Site History, Soils, Landforms, Woodys and Herbaceous Vegetation, Down Wood, and Inventory Methods for Landscape Experiment, NC-208, 41–70.

Kinderman, A., Monahan, J.F., 1977. Computer generation of random variables using the ratio of uniforms deviates. ACM Transactions on Mathematical Software 3, 257–260.

Kitanidis, P.K., 1986. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resources Research 22, 499–507.

Le, N.D., Zidek, J.V., 1992. Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging. Journal of Multivariate Analysis 43, 351–374.

McQuilkin, R.A., 1974. Site index prediction tables for black, scarlet and white oaks in south-eastern Missouri. Research Paper NC-108. U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station. St. Paul, MN.

Møller, J. (Ed.), 2003. Spatial Statistics and Computational Methods. In: Lecture Notes in Statistics, vol. 173. Springer-Verlag, New York, Papers from the TMR and MaPhySto Summer School held at Aalborg University, Aalborg, August 19–22, 2001.

Nash, A.J., 1978. A method for classifying shortleaf pine sites in Missouri. In: Reseach Bulletin 824. Missouri Agricultural Experiment Station, Columbia, MO.

Nigh, T.A., Schroeder, W.A., 2002. Atlas of Missouri Ecoregions. Missouri Department of Conservation, Jefferson City, Missouri.

Shifley, S., Brookshire, B., 2000. Missouri Ozark Forest Ecosystem Project: site history, soils, landforms, woody and herbaceous vegetation, down wood, and inventory methods for the landscape experiment (Eds.). General Technical Report NC-208. The United States Department of Agriculture, Forest Service, North Central Research Station. St. Paul, MN.

Shifley, S., Kabrick, J., 2002. Proceedings of the Missouri Ozark Forest Ecosystem Project Symposium: post-treatment results of the landscape experiment (Eds.). General Technical Report NC-227. U.S. Dept. of Agriculture, Forest Service, North Central Research Station. St. Paul, MN.

Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer-Verlag Inc., New York.

Sun, X., 2006. Bayesian spatial data analysis with application to the Missouri Ozark Forest Ecosystem Project. Ph.D. Thesis. University of Missouri, Columbia.

Wakefield, J.C., Gelfand, A.E., Smith, A.F.M., 1991. Efficient generation of random variates via the ratio-of-uniforms method. Statistics and Computing 1, 129–133.