Autologistic Regression Model for the Distribution of Vegetation

Fangliang HE, Julie ZHOU, and Hongtu ZHU

Modeling the contagious distribution of vegetation and species in ecology and biogeography has been a challenging issue. Previous studies have demonstrated that the autologistic regression model is a useful approach for describing the distribution because spatial correlation can readily be accounted for in the model. So far studies have been mainly restrained to the first-order autologistic model. However, the first-order correlation model may sometimes be insufficient as long-range dispersal/migration can play a significant role in species distribution. In this study, we used the second-order autologistic regression model to model the distributions of the subarctic evergreen woodland and the boreal evergreen forest in British Columbia, Canada, in terms of climate covariates. We investigated and compared three estimation methods for the second-order model—the maximum pseudo-likelihood method, the Monte Carlo likelihood method, and the Markov chain Monte Carlo stochastic approximation. Detailed procedures for these methods were developed and their performances were evaluated through simulations. The study demonstrates the importance for including the second-order correlation in the autologistic model for modeling vegetation distribution at the large geographical scale; each of the two vegetations studied was strongly corrected not only in the south-north direction but also in the northwest-southeast direction. The study further concluded that the assessment of climate change should be performed on the basis of individual vegetation or species because different vegetation or species likely respond differently to different sets of climate variables.

Key Words: Biogeography; Climate change; Markov chain Monte Carlo; Pseudo-likelihood; Presence/absence data; Second-order neighbors; Spatial correlation; Stochastic approximation.

Fangliang He is Research Scientist, Pacific Forestry Center, Canadian Forest Service, Victoria, BC, Canada V8Z 1M5 (E-mail: fhe@pfc.forestry.ca). Julie Zhou is Associate Professor, Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3P4 (E-mail: jzhou@math.uvic.ca). Hongtu Zhu is a Postdoctoral Associate, Department of Epidemiology and Public Health, Yale School of Medicine, New Haven, CT 06511 (E-mail: htzhu@masal.med.yale.edu).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 8, Number 2, Pages 205–222
DOI: 10.1198/1085711031508

1. INTRODUCTION

Recording the presence/absence of a species or a vegetation type in landscapes is perhaps the simplest and easiest survey method in biogeography. With the aid of remote sensing and aerial photogrammetric technologies, information on species occurrence is increasingly documented by this format (Little 1971; Arnold 1993, 1995; Mitchell-Jones et al. 1999). In practice, presence/absence data are typically presented in the form of an atlas map which is divided into a grid system with equal pixel size. Species occurrence and other covariates in each pixel may or may not be fully observed.

Interests in modeling such atlas data have largely focused on two issues. One is to smooth or restore an observed atlas in order to estimate the distribution range for a species (Heikkinen and Högmander 1994; Augustin, Mugglestone, and Buckland 1996; Hoeting, Leecaster, and Bowden 2000). This restoration procedure is necessary in many occasions as the occurrence of a species in some pixels is sometimes less certain than others due to discrepancy in survey intensity in some area, or worse, occurrence information in some pixels may not be available at all. Another issue is to model the spatial distribution rather than to estimate the range. This latter emphasis is particularly pertinent if the goal of a study is to interpret species distribution in terms of covariates, for example, climatic variables (Box, Crumpacker, and Hardin 1993; Wu and Huffer 1997). On this account, the questions of interest then are: Why does the species occur there? Which covariates determine its distribution? and How would its distribution shift if covariates change? It is these questions that motivate this study, in which we investigate the response of vegetation distribution to climate change in British Columbia (BC), Canada.

Several methods have been developed for modeling spatially correlated presence/absence data (Besag 1972, 1974; Albert and McShane 1995; Diggle and Tawn 1998). Among them, the autologistic regression is perhaps the most widely used one, thanks to the statistical theory of the model developed by Besag (1972, 1974). Indeed, previous studies have demonstrated the usefulness of the autologistic regression in modeling binary data with observed covariates (Augustin et al. 1996; Gumpertz, Graham, and Ristaino 1997; Wu and Huffer 1997; Huffer and Wu 1998; Hoeting et al. 2000). In these studies, spatial correlation, a key component in the autologistic regression model, is commonly modeled by considering the effect of neighboring cells. This structure captures the nature of species dispersal or the influence of underlying environmental factors that make species occurrence in neighboring pixels tend more similar than pixels far apart. Several structures can be proposed to reflect this spatial correlation (Figure 1). Although any of them may be reasonable, previous studies usually consider no more than the first-order correlation (e.g., Huffer and Wu 1998) or occasionally consider higher order correlation but depend on the pseudo-likelihood function for parameter estimation (e.g., Gumpertz et al. 1997). This study employs autologistic regression to model the distribution of vegetation in BC by considering the second-order correlation structure. We viewed this as a necessary structure in describing the spatial distribution of vegetation, because the landscapes in BC are typically featured by mountain ranges in the northwest-southeast direction, and dispersal/migration processes in the distribution of vegetation/species may impose a long-range effect. First-order correlation models

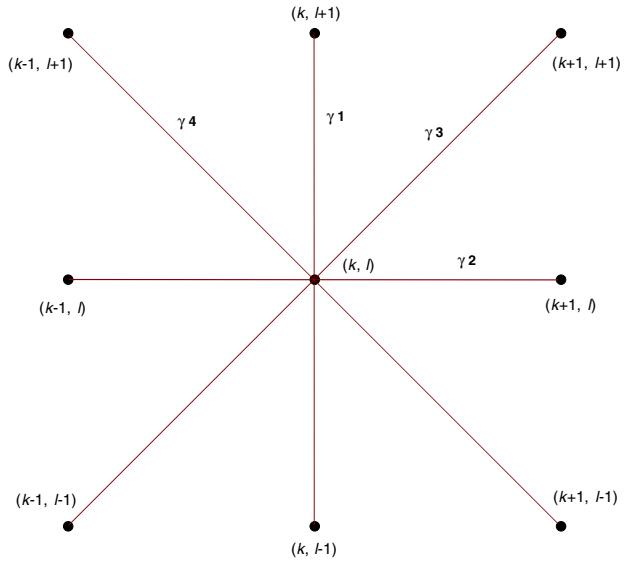

Figure 1. Structure of the first-order and second-order neighborhoods. The vertical (γ_1) and horizontal (γ_2) are the first-order neighborhoods, whereas the second-order neighborhoods further include the diagonal directions (γ_3: southwest-northeast direction; γ_4: northwest-southeast direction).

are not so effective for these geographic structures. In this study three methods were used to estimate the parameters of the second-order autologistic model, including the maximum pseudo-likelihood method (Besag 1977), the Monte Carlo likelihood method (Geyer and Thompson 1992; Geyer 1994; Huffer and Wu 1998) and the Markov chain Monte Carlo stochastic approximation method (Gu and Kong 1998; Gu and Zhu 2001). Simulation studies were conducted to compare the performance of the three methods. The model was then applied to the vegetation data of BC.

2. AUTOLOGISTIC REGRESSION MODEL

Suppose data are recorded at M locations (sites) forming a subset $\mathcal{D}$ of a rectangular lattice. Each site in $\mathcal{D}$ is described by coordinates (k, l) specifying the row and column of the lattice at which it is located. At each site (k, l), we observe a binary response $y_{k,l}$ and a $p \times 1$ vector of covariates $\mathbf{x}_{k,l}$, where $y_{k,l}$ equals 1 if the site is occupied by vegetation, otherwise 0. Taken altogether, the M binary responses $\mathbf{Y} = (y_{k,l}, (k, l) \in \mathcal{D})$ constitute a map of the distribution of that vegetation type. The second-order autologistic regression model specifies the conditional probability $\Pr_{k,l}(\theta)$ that $y_{k,l} = 1$ given all the other values $y_{m,n}$ $((m, n) \neq (k, l))$ as follows

$$\Pr_{k,l}(\theta) = P(y_{k,l} = 1 \mid \text{all other values}) = \frac{\exp(f_{k,l}(\theta))}{1 + \exp(f_{k,l}(\theta))}, \tag{2.1}$$

where $f_{k,l}(\theta) = \beta_0 + \mathbf{x}_{k,l}^T \beta_1 + \gamma_1 y_{k,l}^{(1)} + \gamma_2 y_{k,l}^{(2)} + \gamma_3 y_{k,l}^{(3)} + \gamma_4 y_{k,l}^{(4)}, \theta = (\beta_0, \beta_1^T, \gamma_1, \gamma_2, \gamma_3, \gamma_4)^T$

$\in \Theta$ (a parameter space for θ), $y_{k,l}^{(1)}$ is the number of occupied sites in $\{(k,l-1),(k,l+1)\}$, $y_{k,l}^{(2)}$ is the number of occupied sites in $\{(k-1,l),(k+1,l)\}$, $y_{k,l}^{(3)}$ is the number of occupied sites in $\{(k-1,l-1),(k+1,l+1)\}$, and $y_{k,l}^{(4)}$ is the number of occupied sites in $\{(k-1,l+1),(k+1,l-1)\}$. Thus, $\gamma_1,\gamma_2,\gamma_3$ and γ_4 are the parameters for describing various spatial correlation structures (Figure 1).

Let P_θ be the probability measure of a random map $\mathbf{Y}=(y_{k,l},(k,l)\in\mathcal{D})$ generated from model (2.1). It can be shown that this measure P_θ is a member of an exponential family defined as follows

$$P_\theta(\mathbf{Y}) = C(\theta)^{-1}\exp\{\theta^T T(\mathbf{Y})\}, \tag{2.2}$$

where $T(\mathbf{Y}) = \sum_{(k,l)\in\mathcal{D}} y_{k,l}\tilde{\mathbf{x}}_{k,l}$, $\tilde{\mathbf{x}}_{k,l} = (1,\mathbf{x}_{k,l}^T, y_{k,l}^{(1)}/2, y_{k,l}^{(2)}/2, y_{k,l}^{(3)}/2, y_{k,l}^{(4)}/2)^T$ and $C(\theta)$ is a normalizing factor. The normalizing constant $C(\theta)$ (called the partition function) is obtained by summing over all possible configurations $\mathbf{Y}$, namely

$$C(\theta) = \sum_{\text{all } \mathbf{Y}} \exp\{\theta^T T(\mathbf{Y})\}.$$

Let $\mathbf{Y}_0$ be the observed map, the log-likelihood function for θ is

$$\ell(\theta) = \ell(\theta;\mathbf{Y}_0) = \ln P_\theta(\mathbf{Y}_0) = \theta^T T(\mathbf{Y}_0) - \ln C(\theta). \tag{2.3}$$

The first and second derivatives of $\ell(\theta)$ are, respectively,

$$\bigtriangledown\, \ell(\theta) = T(\mathbf{Y}_0) - \bigtriangledown \ln C(\theta) \quad \text{and} \quad \bigtriangledown^2\, \ell(\theta) = -\bigtriangledown^2 \ln C(\theta), \tag{2.4}$$

where $\bigtriangledown$ and $\bigtriangledown^2$ are the first and second derivative operators with respect to θ. Using the identities $E_\theta[\bigtriangledown \ell(\theta;\mathbf{Y})] = \mathbf{0}$ and $E_\theta[\bigtriangledown^2 \ell(\theta;\mathbf{Y})] = -E_\theta[\bigtriangledown \ell(\theta;\mathbf{Y})^{\otimes 2}]$, where E_θ denotes expectation with respect to the density in (2.2), we can show that

$$\begin{aligned} \bigtriangledown \ln C(\theta) &= E_\theta[T(\mathbf{Y})], \\ \bigtriangledown^2 \ln C(\theta) &= E_\theta\{T(\mathbf{Y})\}^{\otimes 2} - \{E_\theta[T(\mathbf{Y})]\}^{\otimes 2} = \text{var}_\theta[T(\mathbf{Y})], \end{aligned} \tag{2.5}$$

where for a vector $\mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$.

3. ESTIMATION METHODS

This section describes three estimation methods—the maximum pseudo-likelihood (MPL) estimate, the maximum likelihood estimate (MLE), obtained via the Monte Carlo likelihood (MCL) method, and the maximum likelihood estimate obtained via the Markov chain Monte Carlo stochastic approximation method (MCMC-SA), and develop detailed algorithms for model (2.1). We have developed C-code for computing the above three estimates with various constraints on $(\gamma_1,\gamma_2,\gamma_3,\gamma_4)^T$. The codes are available upon request from the authors.

3.1 MAXIMUM PSEUDO-LIKELIHOOD ESTIMATION

The MPL estimate, proposed by Besag (1977), maximizes the sum of the log pseudo-likelihood function

$$\begin{aligned} \ell_{ps} &= \sum_{(k,l)\in\mathcal{D}} \ln[P(Y_{k,l} = y_{k,l} \mid \text{all other values})] \\ &= \sum_{(k,l)\in\mathcal{D}} \{y_{k,l} f_{k,l}(\theta) - \ln[1 + \exp(f_{k,l}(\theta))]\}. \end{aligned} \tag{3.1}$$

Since the objective function in (3.1) is convex in θ, the standard Newton–Raphson algorithm can be used. The approximate variance for the MPL estimate $\hat{\theta}_{\text{MPL}}$ is taken to be the diagonal elements of

$$\left[\sum_{(k,l)\in D} \text{Pr}_{k,l}(\hat{\theta}_{\text{MPL}})(1 - \text{Pr}_{k,l}(\hat{\theta}_{\text{MPL}}))\tilde{\mathbf{x}}_{k,l}\tilde{\mathbf{x}}_{k,l}^T\right]^{-1}. \tag{3.2}$$

However, it should be noted that this estimate does not give a valid variance estimate. The result should be interpreted cautiously. In addition to being easy to compute, it has been proved that the MPL estimate is consistent and asymptotically normally distributed (Comets 1992). Although the MPL estimate is very efficient in the case of weak spatial correlation, it is not so in the case of strong spatial correlation (Diggle et al. 1994).

3.2 MONTE CARLO LIKELIHOOD METHOD

Geyer and Thompson (1992) considered the log-likelihood ratio against a "reference" point $\psi \in \Theta$,

$$\ell(\theta) - \ell(\psi) = (\theta - \psi)^T T(\mathbf{Y}_0) - \ln\frac{C(\theta)}{C(\psi)}. \tag{3.3}$$

From (2.2), we have

$$\frac{C(\theta)}{C(\psi)} = \int \exp\{(\theta - \psi)^T T(\mathbf{Y})\} P_\psi(\mathbf{Y}) d\mathbf{Y}. \tag{3.4}$$

Using Monte Carlo, we can approximate the integral in (3.4) based on a random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ from $P_\psi(\mathbf{Y})$, that is,

$$\frac{C(\theta)}{C(\psi)} \approx \frac{1}{n}\sum_{i=1}^{n} \exp\{(\theta - \psi)^T T(\mathbf{Y}_i)\}. \tag{3.5}$$

Substituting (3.5) into (3.3), we obtain an approximation to the log-likelihood ratio in (3.3)

$$\ell_n(\theta;\psi) = (\theta - \psi)^T T(\mathbf{Y}_0) - \ln\left[\frac{1}{n}\sum_{i=1}^{n} \exp\{(\theta - \psi)^T T(\mathbf{Y}_i)\}\right].$$

Directly maximizing $\ell_n(\theta;\psi)$ yields the MCL estimate $\hat{\theta}_n$.

Define

$$\omega_{n,\theta,\psi}(\mathbf{Y}) = \frac{\exp\{(\theta-\psi)^T T(\mathbf{Y})\}}{\sum_{i=1}^n \exp\{(\theta-\psi)^T T(\mathbf{Y}_i)\}}$$

and for any function g, $\mathbf{E}_{n,\theta,\psi} g(\mathbf{Y}) = \sum_{i=1}^n g(\mathbf{Y}_i)\omega_{n,\theta,\psi}(\mathbf{Y}_i)$. Using these notations, we get

$$\bigtriangledown \ell_n(\theta;\psi) \approx T(\mathbf{Y}_0) - \mathbf{E}_{n,\theta,\psi}[T(\mathbf{Y})], \tag{3.6}$$

and

$$-\bigtriangledown^2 \ell_n(\theta;\psi) \approx \text{var}_{n,\theta,\psi} T(\mathbf{Y}) = \mathbf{E}_{n,\theta,\psi}[T(\mathbf{Y})]^{\otimes 2} - [\mathbf{E}_{n,\theta,\psi} T(\mathbf{Y})]^{\otimes 2}.$$

Based on these results, the Newton–Raphson algorithm can be applied to compute $\hat{\theta}_n$. The details of the algorithm are:

M1: Select an initial value θ^0 for θ and a fixed point $\psi \in \Theta$.

M2: Generate a random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ from $P_\psi(\mathbf{Y})$ using a Gibbs sampler below. Each site (k,l) is selected in the lexicographical order.

1. Calculate $\text{Pr}_{k,l}(\psi)$ using (2.1).
2. Generate a random number u uniformly distributed on (0, 1]. If $u \le \text{Pr}_{k,l}(\psi)$, set $y_{k,l} = 1$; otherwise, set $y_{k,l} = 0$.

After updating all sites in $\mathcal{D}$, we treat the observations on $\mathcal{D}$ as one map.

Note: An initial map is needed to carry out the above procedure, and we usually take the observed map $\mathbf{Y}_0$. Also a number of burn-in maps are required to reach an equilibrium state before selecting a random sample (Geyer and Thompson 1992). In our computation in Sections 4 and 5, 200 burn-in maps were used and a spacing of 2, that is, every other map generated after that was taken to obtain the random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. Experiments with different number of burn-ins and spacing did not yield different results.

M3: For $i \ge 1$, update estimate θ^i using the Newton–Raphson algorithm

$$\theta^i = \theta^{i-1} - \rho[\bigtriangledown^2 \ell_n(\theta^{i-1};\psi)]^{-1} \bigtriangledown \ell_n(\theta^{i-1};\psi), \tag{3.7}$$

where $0 < \rho = 1/2^{k_0} \le 1$ for some $k_0 \ge 0$ such that $\ell_n(\theta^i;\psi) \ge \ell_n(\theta^{i-1};\psi)$. Repeat this step until θ^i converges, and at the final step we set $\hat{\theta}_n = \theta^i$.

Under some mild conditions, Geyer (1994) showed that $\hat{\theta}_n \to \hat{\theta}$ (MLE) in probability as $n \to \infty$. The success of the MCL method depends on the choice of ψ. If ψ is far from the exact MLE, the above method may fail to have a solution. Huffer and Wu (1998) proposed to use the MPL estimate for the initial value θ^0 and reference point ψ.

3.3 MARKOV CHAIN MONTE CARLO STOCHASTIC APPROXIMATION

Gu and Zhu (2001) recently proposed a two-stage stochastic approximation for computing maximum likelihood estimates for a class of spatial models. Their approach combines

the methods of adaptive search direction and off-line average. Assume that at the ith iteration θ^i is the current estimate of $\hat{\theta}$, $\mathbf{h}_i$ is the current estimate of $E_{\hat{\theta}}[T(\mathbf{Y})]$ and Γ_i is the current estimate of $E_{\hat{\theta}}\{T(\mathbf{Y})\}^{\otimes 2}$. In the algorithm m is a preselected integer. In practice, we suggest choosing m to be about $20M$. The two basic steps for the stochastic approximation for model (2.1) are:

Step 1. At the ith iteration, set $\mathbf{Y}_1^{(i)} = \mathbf{Y}_m^{(i-1)}$. We generate $m-1$ maps $\{\mathbf{Y}_j^{(i)}, j = 2, \ldots, m\}$ as follows. We start with map $\mathbf{Y}_1^{(i)}$ and update one randomly selected site to get the next map $\mathbf{Y}_2^{(i)}$. Therefore the possible difference between any two consecutive maps $\mathbf{Y}_j^{(i)}$ and $\mathbf{Y}_{j+1}^{(i)}$ is the value at the randomly selected site. Here is the detail to update one site.

1. Randomly select a site $(k, l) \in \mathcal{D}$ with probability $1/|\mathcal{D}|$.
2. Calculate $\Pr_{k,l}(\theta^{i-1})$ using (2.1).
3. Generate a random number u uniformly distributed on (0, 1]. If $u \le \Pr_{k,l}(\theta^{i-1})$, set $y_{k,l} = 1$; otherwise, set $y_{k,l} = 0$.

Step 2. Update θ^{i-1} to θ^i, $\mathbf{h}_{i-1}$ to $\mathbf{h}_i$ and Γ_{i-1} to Γ_i by

$$\begin{cases} \mathbf{h}_i = \mathbf{h}_{i-1} + \gamma_i(\sum_{j=1}^m T(\mathbf{Y}_j^{(i)})/m - \mathbf{h}_{i-1}), \\ \Gamma_i = \Gamma_{i-1} + \gamma_i(\sum_{j=1}^m T(\mathbf{Y}_j^{(i)})^{\otimes 2}/m - \Gamma_{i-1}), \\ \theta^i = \theta^{i-1} + \gamma_i[\Gamma_i - \mathbf{h}_i^{\otimes 2}]^{-1}[T(\mathbf{Y}_0) - \sum_{j=1}^m T(\mathbf{Y}_j^{(i)})/m]. \end{cases} \tag{3.8}$$

Note that the last formula in (3.8) is just the stochastic approximation version of the Newton–Raphson update of (3.7). The main procedure is implemented as follows.

Stage I. Choose an initial point θ^0, an initial matrix Γ_0, an initial vector $\mathbf{h}_0$ and an initial spatial configuration $\mathbf{Y}_m^{(0)}$, set $i = 1$ and iterate Steps 1 and 2 with $i = 1, \ldots, K_1$. The gain constants are defined by

$$\gamma_i = \gamma_{1i} = b_1/(i^{a_1} + b_1 - 1), \quad i = 1, \ldots, K_1,$$

where $K_1 \ge K_0$ is determined by

$$K_1 = \inf\left\{K \ge K_0, \quad \left\| \sum_{i=K-K_0+1}^{K} \text{Sign}(\theta^i - \theta^{i-1})/K_0 \right\| \le \eta_1 \right\}, \tag{3.9}$$

in which $\|\cdot\|$ denotes the commonly used $\mathcal{L}^2$ norm. The function $\text{Sign}(\theta)$ is a vector of 1, 0 or -1 according to whether the component of θ is positive, zero or negative respectively. The integers b_1 and K_0, the real numbers $a_1 \in (0, 1)$ and η_1 are preassigned constants. Gu and Zhu (2001) suggested to choose a_1 to be close to zero, b_1 to be relatively large, and η_1 to be a relatively small value, say $a_1 = 0.3$, $b_1 = 5$ and $\eta_1 = 0.1$.

Stage II. Take the final values of θ, $\mathbf{h}$, Γ and $\mathbf{Y}$ of Stage I as its initial values. We iterate Steps 1 and 2 with $i = 1, \ldots, K_2$. The gain constants are defined by

$$\gamma_i = \gamma_{2i} = b_2/(i^{a_2} + b_2 - 1), \quad i = 1, \ldots, K_2,$$

where the integer b_2 and the real number $a_2 \in (1/2, 1)$ are preassigned. We choose a_2 close to 1/2, and a small integer for b_2, say $a_2 = 0.8$ and $b_2 = 2$. At the same time, an averaging procedure is used with $\tilde{\theta}^0 = \theta^0$, $\tilde{\mathbf{h}}_0 = \mathbf{h}_0$ and $\tilde{\Gamma}_0 = \Gamma_0$,

$$\tilde{\theta}^i = \tilde{\theta}^{i-1} + (\theta^i - \tilde{\theta}^{i-1})/i, \ \tilde{\mathbf{h}}_i = \tilde{\mathbf{h}}_{i-1} + (\mathbf{h}_i - \tilde{\mathbf{h}}_{i-1})/i, \ \tilde{\Gamma}_i = \tilde{\Gamma}_{i-1} + (\Gamma_i - \tilde{\Gamma}_{i-1})/i.$$

Taking account of a possibly large Monte Carlo error, Gu and Zhu (2001) defined the following criteria function to determine K_2,

$$\hat{\Delta}_i = (T(\mathbf{Y}_0) - \tilde{\mathbf{h}}_i)^T \left[\tilde{\Gamma}_i - \tilde{\mathbf{h}}_i^{\otimes 2}\right]^{-1} (T(\mathbf{Y}_0) - \tilde{\mathbf{h}}_i) + \text{tr}\left\{\left[\tilde{\Gamma}_i - \tilde{\mathbf{h}}_i^{\otimes 2}\right]^{-1} \hat{\Sigma}_i\right\}/i,$$

where $\hat{\Sigma}_i$ denotes the sample covariance of $\{\sum_{j=1}^m T(\mathbf{Y}_j^{(k)})/m, k = 1, \ldots, i\}$ and K_2 is defined as

$$K_2 = \inf\left\{i, \quad \hat{\Delta}_i \le \eta_2\right\},$$

where η_2 (usually taken to be around 0.002) is a preassigned small number. After the K_2th iteration, we use the off-line average $(\tilde{\theta}^{K_2}, \tilde{\mathbf{h}}_{K_2}, \tilde{\Gamma}_{K_2})$ as our final estimate of $(\hat{\theta}, E_{\hat{\theta}}[T(\mathbf{Y})], E_{\hat{\theta}}\{T(\mathbf{Y})\}^{\otimes 2})$.

Note: In practice, to save computer time, we can run an MCMC-SA algorithm with a relatively large η_1, for example, $\eta = 0.05$, and then use the obtained (rough) estimate as the starting value to obtain the MCL estimate. This approach is very useful for a large dataset and large K_2. Moreover, the Step 2 in the MCMC-SA algorithm can be regarded as a stochastic version of the Newton–Raphson update given in (3.7).

Gu and Zhu (2001) showed that, as $K_2 \to \infty$,

$$(\tilde{\theta}^{K_2}, \tilde{\mathbf{h}}_{K_2}, \tilde{\Gamma}_{K_2}) \longrightarrow (\hat{\theta}, E_{\hat{\theta}}[T(\mathbf{Y})], E_{\hat{\theta}}\{T(\mathbf{Y})\}^{\otimes 2}) \quad \text{in probability.}$$

3.4 GOODNESS-OF-FIT STATISTICS

After estimating the parameters in the autologistic model, we can obtain an estimate of the conditional probability $\text{Pr}_{k,l}(\hat{\theta})$ in (2.1) at each site $(k, l) \in D$. Based on these probabilities, we considered three goodness-of-fit statistics:

1. sum of the absolute errors: $\text{SAE} = \sum_{(k,l)\in D} |y_{k,l} - \text{Pr}_{k,l}(\hat{\theta})|/M$,
2. sum of the squares of errors: $\text{SSE} = \sum_{(k,l)\in D} (y_{k,l} - \text{Pr}_{k,l}(\hat{\theta}))^2/M$,
3. sum of correct predictions: $\text{SCP} = \sum_{(k,l)\in D} \delta(y_{k,l}, \hat{y}_{k,l})$,

where $\delta(\cdot, \cdot)$ is the Kronecker function and

$$\hat{y}_{k,l} = \begin{cases} 1, & \text{Pr}_{k,l} \ge 0.5, \\ 0, & \text{Pr}_{k,l} < 0.5. \end{cases}$$

Adequate fits have small SAE and SSE and large SCP. The AIC for MCMC-SA was calculated following Huang and Ogata (2001).

4. SIMULATION STUDY

To compare the MCMC-SA algorithm with MPL and MCL, we considered the following autologistic regression model on a 40×40 lattice

$$f_{k,l}(\theta) = \beta_0 + \beta_1 x_{k,l} + \tilde{\gamma}_1(y_{k,l}^{(1)} + y_{k,l}^{(2)}) + \tilde{\gamma}_2(y_{k,l}^{(3)} + y_{k,l}^{(4)}),$$

where $x_{k,l} = 2.5 \times \sin(0.1 \times (k+l))$. Thus, we have four parameters to estimate. In the simulation study, we set $\beta_0 = 1$, $\beta_1 = -1$ and

$$(\tilde{\gamma}_1, \tilde{\gamma}_2) \in \{(0,0), (0.5, 0.5), (0.5, -0.5), (-0.5, -0.5), (-0.5, 0.5)\}.$$

Parameter $\beta_1 = -1$ represents a strong covariate effect, and $(\tilde{\gamma}_1, \tilde{\gamma}_2)$ ranges in four different directions. The corner sites only have three neighbors, and the sites along the edges have five neighbors. To simulate the process, we used the Gibbs sampler. The initial state of the process was taken at random such that $y_{k,l}$ at each site (k,l) was taken independently to be 1 or 0 with 1/2 probability. We repeated the Gibbs sampler 2001 times (2001 Monte Carlo steps) to ensure that the equilibrium states were achieved.

For each parameter vector $\theta = (\beta_0, \beta_1, \tilde{\gamma}_1, \tilde{\gamma}_2)^T$, we generated $N = 500$ datasets. We applied the proposed three methods to obtain estimates of θ for each pseudo-observed dataset. For the MPL method and the MCMC-SA algorithm, the initial value of θ was set at $(0,0,0,0)$. The MPL estimate was used as the starting value and the reference point for the MCL method. In fact, we took the first step θ^1 of the Newton–Raphson iteration as a new value for θ to generate another random sample from the Gibbs sampler. This is different from Huffer and Wu (1998) who used θ^1 only when the sample produced using the MPL did not lead to adequate estimates.

For the MCL method, the random samples were generated from a Gibbs sampler described in Section 3.2. We used a burn-in of 200 iterations and a spacing of 2. A total of 10,200 (40×40) random samples were generated, and only 5,000 random samples were used to obtain the MCL estimates. In the MCMC-SA algorithm, we set $(a_1, a_2, b_1, b_2, m, K_0, \eta_1, \eta_2)$ as $(0.30, 0.60, 4.0, 1.0, 20{,}000, 200, 0.144, 0.001)$, $\mathbf{h}_0 = \mathbf{0}$ and $\Gamma_0 = \mathbf{I}$ the identity matrix.

To compare the performance of the proposed approaches, we examined only univariate aspects of the sampling distribution of the four estimates based on the 500 parameter estimates. These included the estimated sample mean (ESM), the estimated sample standard deviation (ESD), the mean of the standard error estimates (MSD), the skewness coefficient (SK) $g_1 = m_3/m_2^{3/2}$ and the excess kurtosis coefficient (KT) $g_2 = m_4/m_2^2 - 3$, where m_2, m_3 and m_4 are the estimated centered sample moments. The ratio of ESD over MSD can be used to check accuracy of the standard error estimates, whereas the coefficients of skewness and excess kurtosis are used to test the hypotheses that the obtained estimates come

Table 1. Summary Statistics for 500 Samples Simulated from the Scenario with True Parameters: $(\beta^T, \gamma) = (1, -1, 0.5, 0.5)$. ESM: the estimated sample mean, ESD: the estimated sample standard deviation, MSD: the mean of the standard error estimates, SK: the skewness coefficient, KT: the kurtosis coefficient, 95%, 90%, 70%, 50%, 10% are the true confidence levels.

	ESM	*ESD*	*MSD*	*SK*	*KT*	*95%*	*90%*	*70%*	*50%*	*10%*
					MPL					
β_0	1.180	0.784	0.733	0.349	−0.194	0.954	0.874	0.668	0.470	0.084
β_1	−1.047	0.189	0.192	−0.875	1.324	0.970	0.942	0.734	0.508	0.104
$\tilde{\gamma}_1$	0.509	0.283	0.199	0.228	−0.078	0.842	0.750	0.524	0.380	0.058
$\tilde{\gamma}_2$	0.468	0.240	0.182	−0.204	−0.227	0.848	0.784	0.582	0.372	0.082
					MCL					
β_0	1.174	0.725	0.722	0.426	−0.057	0.970	0.912	0.686	0.504	0.086
β_1	−1.041	0.184	0.180	−0.954	1.652	0.958	0.938	0.736	0.492	0.110
$\tilde{\gamma}_1$	0.495	0.267	0.266	0.198	−0.155	0.960	0.904	0.678	0.494	0.100
$\tilde{\gamma}_2$	0.483	0.223	0.227	−0.135	−0.163	0.958	0.906	0.708	0.486	0.124
					MCMC-SA					
β_0	1.173	0.725	0.720	0.436	−0.044	0.962	0.912	0.686	0.502	0.088
β_1	−1.041	0.184	0.179	−0.962	1.684	0.958	0.934	0.732	0.498	0.106
$\tilde{\gamma}_1$	0.495	0.268	0.266	0.193	−0.154	0.956	0.890	0.676	0.492	0.096
$\tilde{\gamma}_2$	0.482	0.223	0.227	−0.136	−0.178	0.960	0.900	0.708	0.484	0.114

from a normal distribution. Under the normal assumption and when N is large enough, g_1 and g_2 are approximately normally distributed with mean zero and standard error $(6/N)^{1/2}$ and $(24/N)^{1/2}$, respectively. In our simulation study, $N = 500$ and the standard deviations of g_1 and g_2 are, respectively, 0.11 and 0.22. Formula (3.2) was used to obtain an estimate of the covariance matrix for the MPL estimate. The inverse of the observed information matrix was used as an estimate of the covariance matrix for the MLEs via the MCL and MCMC-SA methods.

Like g_1 and g_2, $r_i = (\hat{\theta}_i - \theta_i)/\hat{\sigma}_i$ is also approximately $N(0, 1)$, where $\hat{\theta}_i$ is the ith element of vector $\hat{\theta}$ and $\hat{\sigma}_i$ is the standard error for $\hat{\theta}_i$. Therefore we can form the $(1 - \alpha)$ confidence interval for θ_i using $\hat{\theta}_i \pm z_{\alpha/2}\hat{\sigma}_i$, where $z_{\alpha/2}$ is the normal quartile corresponding to the $(1 - \alpha)$ confidence level. Based on the obtained parameter estimates and standard error estimates, for each set of parameter values, we can construct confidence intervals and calculate the actual coverage probabilities of the 95%, 90%, 70%, 50%, and 10% confidence intervals. The exact Binomial two-sided test is used to test the hypothesis that the coverage probabilities are equal to the nominal levels. Under the level of significance $\alpha = 0.05$, the critical values for the various confidence levels are: $95\%(0.930, 0.969)$, $90\%(0.872, 0.926)$, $70\%(0.659, 0.740)$, $50\%(0.456, 0.544)$ and $10\%(0.074, 0.128)$.

The results for one of the five simulation scenarios are illustrated in Table 1, similar results were obtained for other simulations. From the simulations, we seem to be able to make the following conclusions. (i) The MCL and MCMC-SA algorithms are superior to that of MPL in terms of the accuracy of the standard error estimates and of the estimated sample standard deviations. The estimates from MCL and MCMC-SA are often too close to be distinguishable. (ii) In comparison with the MCL and MPL methods, the MCMC-SA

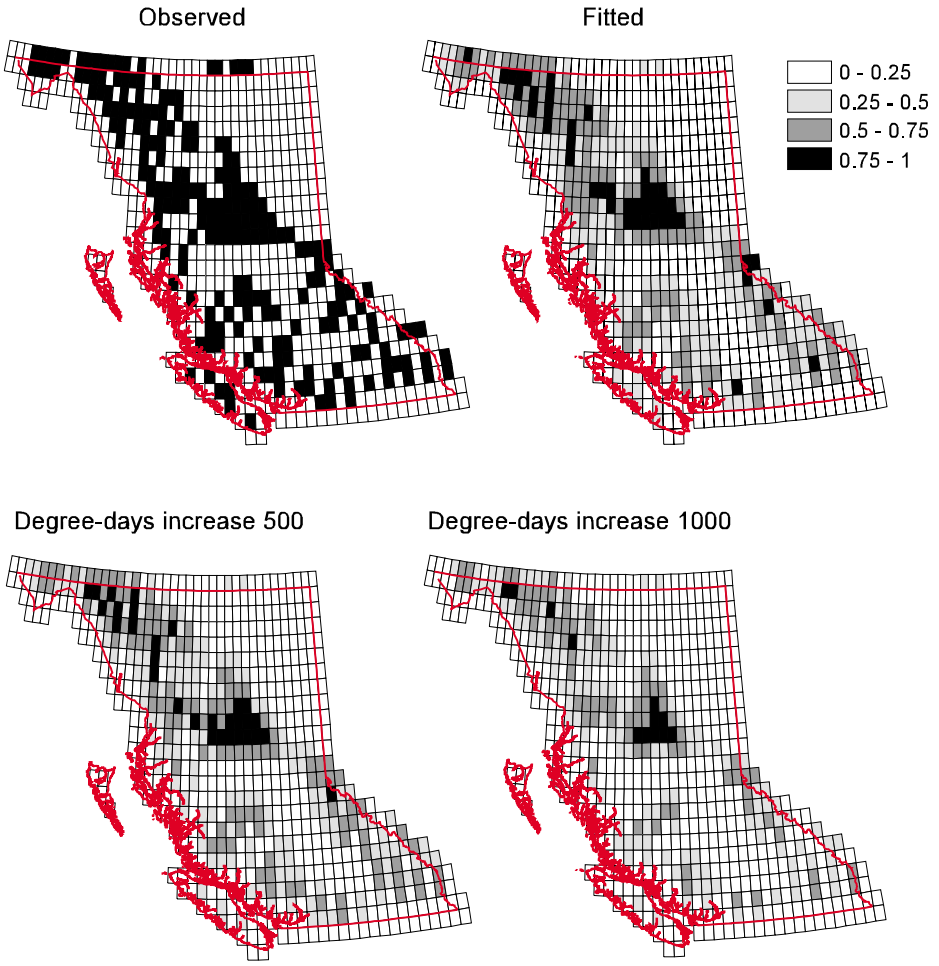

Figure 2. Distribution of the subarctic evergreen woodland in BC. The top maps are the observed and fitted distributions and the bottom maps are the effect of the increase in degree-days (X_2) on the distribution of the woodland, all produced using the reduced model in Table 2.

algorithm demands much more computer time, although the latter algorithm possibly has more chance to find the true MLE's in a certain sense. According to our experience, adjusting η_2 is important in improving the estimate. We suggest choosing $\eta_2 \in (0.001, 0.01]$ which works very well in most cases in this study. (iii) Our simulations show that the MCL method works well if the MPL estimates are used as a starting point, confirming the conclusion of Huffer and Wu (1998).

5. APPLICATIONS

Understanding the relationship between the geographical distribution of vegetation and climatic variables is essential in studying the potential effects of climate change on

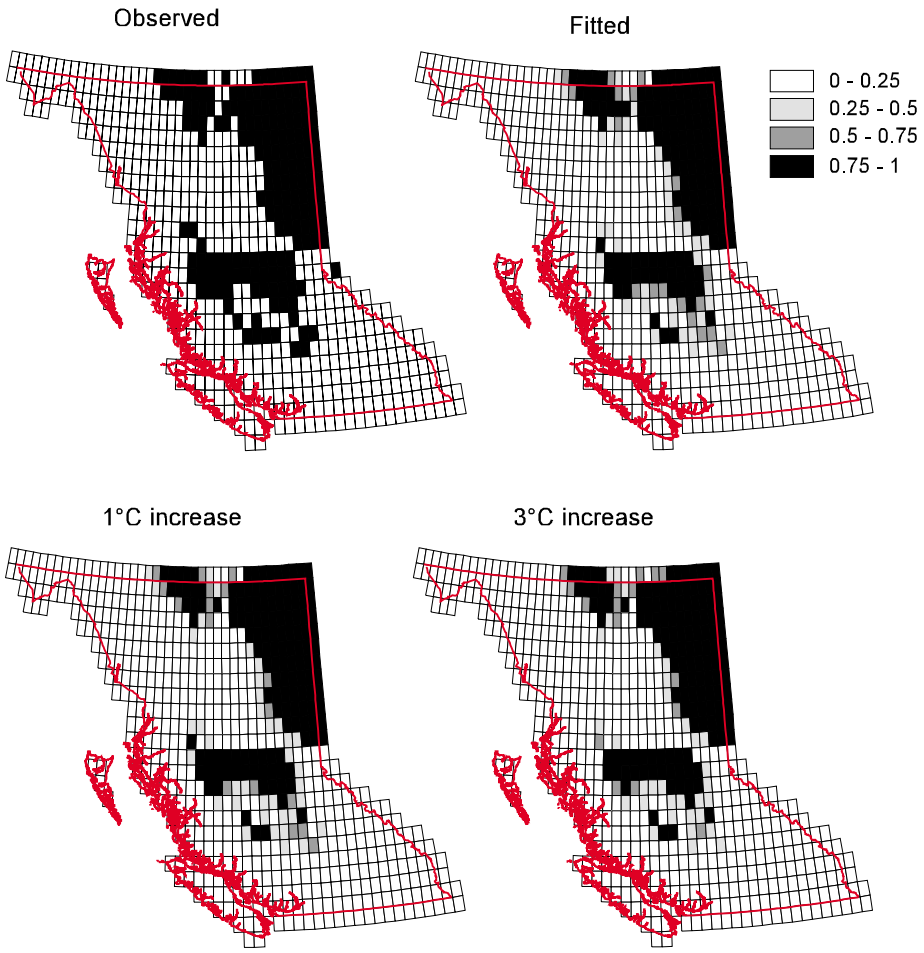

Figure 3. Distribution of the boreal evergreen forest in BC. The top maps are the observed and fitted distributions and the bottom maps are the effect of the increase in temperatue (X_1) on the distribution of the boreal forest, all produced using the reduced model in Table 3.

vegetation. The data we analyze are from an earlier study that addressed the vegetation distribution of all of Canada from ecophysiological perspectives (Lenihan and Neilson 1993). In this study we concentrate on modeling the distributions of two types of vegetation in British Columbia in terms of five climatic variables. Our particular goal at this point is to develop suitable estimation methods for the second-order autologistic model. The two vegetation types are the subarctic evergreen woodland and the boreal evergreen forests in the form of an atlas map with pixel resolution 0.5° latitude $\times 0.5^\circ$ longitude (Figures 2 and 3). There are total of 707 grid cells. The subarctic evergreen woodland is the broad ecotone between tundra (at higher elevation) and closed forest (at lower elevation), that is, between the tree line and the continuous forest line. The boreal evergreen forest is dominated by black spruce (*Picea mariana* (Mill.) B.S.P.), white spruce (*Picea glauca* (Moench) Voss), jack pine (*Pinus banksiana Lamb.*) and balsam fir (*Abies balsamea* (L.) Mill.), and occupies

the central mountainous region and the northeastern upland.

In each $0.5^\circ \times 0.5^\circ$ pixel, there are records for each of the following five climatic variables: absolute minimum temperature (°C) for the coldest month (X_1), annual degree-days with base temperature = 0°C (X_2), total actual evapotranspiration (mm) for summer months (X_3), annual soil moisture deficit (mm) (X_4) and annual snowpack (mm) (X_5). These variables were commonly derived from two fundamental climatic factors: temperature and precipitation on a monthly basis for a 30-year period [see Lenihan and Neilson (1993) for further details]. These variables are expected to be among those that determine the distribution of vegetation at the geographical scale and they are also the variables likely to respond to global warming.

The three estimation methods detailed in Section 3 were applied to produce estimates for the second-order autologistic model of the two vegetation types. The initial values for the MPL and MCMC-SA methods were set to be $\theta^0 = \mathbf{0}$ whereas the initial values and the reference point for the MCL method were the MPL estimates. For the MCL method, we used a burn-in of 200 iterations and a spacing of 2. A total of 10,200 random samples were generated, but only 5,000 random samples were used. In the MCMC-SA algorithm, we set $(a_1, a_2, b_1, b_2, m, K_0, \eta_1, \eta_2) = (0.4, 0.8, 2.0, 1.0, 10000, 100, 0.1, 0.001)$, $\mathbf{h}_0$ and $\Gamma_0 = I$ the identity matrix. However, for the boreal evergreen forest, using the MPL estimate as the reference point for the MCL estimate did not yield reasonable results. Thus, the MLEs obtained from the MCMC-SA algorithm were instead used as the reference point for the boreal forest. The estimated results are presented in Tables 2 and 3. For comparison, the results for the standard logistic regression assuming no spatial correlation and for the first-order logistic regression are given in the tables. Furthermore, for the application purpose, reduced models estimated via the MCMC-SA are also presented in the tables for the subarctic woodland and the boreal forest, respectively. It is worth mentioning that the nearly identical reduced models were also produced using the MCL.

The various goodness-of-fit statistics and the AIC in Tables 2 and 3 show that the second-order regression model improves the fitting of the standard and the first-order logistic models to the data and is consistent with the results of Huang and Ogata (2001). For both the subarctic evergreen woodland and the boreal forest, only γ_1 and γ_4 are significant according to the outputs of the full models of MCL and MCMC-SA and the corresponding reduced models (Tables 2, 3). This spatial structure is rather reasonable considering the geographical variation of BC. The significance of γ_1 describes the south-north tendency in the distribution of vegetation, echoing the climate gradient from south to north, whereas the significance of γ_4 reflects the distribution of landscapes and vegetation in BC which typically lay in the northwest-southeast direction. In contrast, the variation in west-east and northeast-southwest directions is not important to the vegetation. In addition to the spatial coefficients, degree-days (X_2) seems to be most important in limiting the distribution of the subarctic woodland, suggesting that the accumulation of physiological heat is critical for the vegetation. The other significant variable is the depth of snowpack (X_5) which is known to control the upper limit (i.e., the tree line) of the subartic woodland. For the boreal forest, the temperature for the coldest month (X_1) and the degree-days (X_2) significantly restrain

the distribution of the forest, suggesting that the heat sum (X_2) and the lethal temperature (X_1) together determine the boreal forest.

From resource management and policy-making perspectives, there is enormous interest in projecting the redistribution of vegetation at various global warming scenarios. It is predicted that an enhanced greenhouse effect (e.g., the doubling of atmospheric concentration of CO_2) would increase the global mean temperature from 1.5 to 4.5°C in the future 30 to 50 years, while the increase in temperature in Northern Hemisphere might be even as high as 5-9°C (Lamb 1977; Houghton, Jenkins, and Ephraums 1990). One advantage of the autologistic regression model is that the climate change effect can be easily quantified

Table 2. Estimation Results for the Subarctic Evergreen Woodland in BC. EST: estimate, SD: standard deviation. The reduced model was estimated using the MCMC-SA algorithm.

					Logistic model (MLE)					
	β_0	X_1	X_2	X_3	X_4	X_5				
EST	1.5510	0.0011	−0.0012	−0.0027	−0.0032	−0.0011				
SD	1.1045	0.0214	0.0005	0.0069	0.0028	0.0003				
	SAE	SSE	SCP	AIC	Iteration					
	0.382	0.193	469	805.365	5					
					First-order auto-logistic model (MCMC-SA)					
	β_0	X_1	X_2	X_3	X_4	X_5	γ_1	γ_2		
EST	−0.8741	0.0106	−0.0007	0.0021	−0.0004	−0.0004	1.1987	0.7797		
SD	0.8395	0.0145	0.0003	0.0048	0.0019	0.0002	0.1832	0.1722		
	SAE	SSE	SCP	AIC	Iteration					
	0.313	0.164	535	736.44	2548					
					Second-order auto-logistic model					
	β_0	X_1	X_2	X_3	X_4	X_5	γ_1	γ_2	γ_3	γ_4
					MPL					
EST	−1.0611	0.0184	−0.0011	0.0094	0.0014	−0.0004	0.7339	0.5316	0.3168	0.5659
SD	1.3640	0.0246	0.0005	0.0077	0.0030	0.0004	0.1403	0.1558	0.1571	0.1507
	SAE	SAE	SCP	Iteration						
	0.311	0.157	537	5						
					MCL					
EST	−1.1001	0.0124	−0.0006	0.0019	−0.0002	−0.00036	1.0121	0.3213	0.2931	0.5896
SD	0.8743	0.0137	0.0004	0.0052	0.0019	0.0002	0.2502	0.2436	0.2098	0.2285
	SAE	SSE	SCP	Iteration						
	0.311	0.157	537	5						
					MCMC-SA					
EST	−1.2147	0.0109	−0.0007	0.0029	0.0001	−0.0003	0.9677	0.3254	0.2851	0.6070
SD	0.7959	0.0134	0.0003	0.0044	0.0017	0.0002	0.2055	0.2347	0.2033	0.2021
	SAE	SSE	SCP	AIC	Iteration					
	0.312	0.159	533	729.77	3087					
					Reduced model					
EST	–	–	−0.0012	–	–	−0.0008	0.8234	–	–	0.5216
SD	–	–	0.0001	–	–	0.0002	0.1559	–	–	0.1511
	SAE	SSE	SCP	Iteration						
	0.324	0.163	529	2688						

through odds ratio of the conditional probabilities. For example, if the minimum temperature X_1 increases by one degree while other variables remain constant, the odds ratio of the conditional probabilities of vegetation presence is supposed to increase by a factor e^{β_1}. For the boreal evergreen forest, $\beta_1 = -0.1404$ according to the reduced model (Table 3) that equates the odds ratio of 0.8690 $(= e^{-0.1404})$ for increasing 1°C in temperature, that is, the odds of forest presence in each 0.5° latitude $\times 0.5^\circ$ longitude cell will decrease approximately by 13.10% with a 95% confidence interval $(7.34\%, 18.86\%)$. This result suggests that the boreal evergreen forest in BC will be reduced by global warming. Similarly, if degree-days (X_2) increases 350 (approximately equivalent to 1°C increase in daily

Table 3. Estimation Results for the Boreal Evergreen Forest in BC. EST: estimate, SD: standard deviation. The reduced model was estimated using the MCMC-SA algorithm.

					Logistic model (MLE)					
	β_0	X_1	X_2	X_3	X_4	X_5				
EST	−23.3924	−0.3654	0.0031	0.0306	0.0040	−0.0095				
SD	2.4621	0.0482	0.0008	0.0118	0.0043	0.0019				
	SAE	SSE	SCP	AIC	Iteration					
	0.162	0.078	631	383.49	9					
					First-order auto-logistic model (MCMC-SA)					
	β_0	X_1	X_2	X_3	X_4	X_5	γ_1	γ_2		
EST	−11.0660	−0.1180	0.0014	0.0075	0.0004	−0.0020	2.3902	0.8973		
SD	1.8504	0.0373	0.0006	0.0083	0.0027	0.0012	0.3430	0.2463		
	SAE	SSE	SCP	AIC	Iteration					
	0.095	0.048	657	280.55	4646					
					Second-order auto-logistic model					
	β_0	X_1	X_2	X_3	X_4	X_5	γ_1	γ_2	γ_3	γ_4
					MPL					
EST	−13.1997	−0.1263	0.0012	0.0204	0.0029	−0.0008	2.2362	0.0305	0.5001	1.1820
SD	3.4094	0.0705	0.0012	0.0018	0.0060	0.0020	0.3529	0.3545	0.3370	0.3504
	SAE	SSE	SCP	Iteration						
	0.087	0.046	662	8						
					MCL					
EST	−10.9824	−0.1099	0.0013	0.0099	0.0008	−0.0022	2.2699	0.3567	−0.2709	1.0795
SD	1.7925	0.0346	0.0006	0.0082	0.0027	0.0011	0.3876	0.5546	0.4065	0.4145
	SAE	SSE	SCP	Iteration						
	0.092	0.046	661	3						
					MCMC-SA					
EST	−10.9264	−0.1096	0.0013	0.0095	0.0008	−0.0019	2.2634	0.3612	−0.2796	1.0671
SD	1.8074	0.0363	0.0006	0.0084	0.0027	0.0012	0.3868	0.5541	0.4094	0.4198
	SAE	SSE	SCP	AIC	Iteration					
	0.092	0.046	661	276.96	4546					
					Reduced model					
EST	−12.1182	−0.1404	0.0017	–	–	–	2.3385	–	–	1.3289
SD	1.8532	0.0335	0.0003	–	–	–	0.3302	–	–	0.2573
	SAE	SSE	SCP	Iteration						
	0.092	0.046	656	4460						

temperature), the odds ratio for the subarctic woodland is 0.6610 ($= e^{-0.4140}$), that is, the presence of the woodland in each cell is estimated to be reduced by 33.90% with a 95% confidence interval $(29.4\%, 38.60\%)$. The impacts of climate change on the distributions of the two vegetations assessed using the reduced models are shown in Figures 2 and 3.

6. DISCUSSION

The logistic regression model has been used to model spatially correlated binary data in two different ways. One is the so-called marginal model in which spatial correlation is explicitly defined by fitting theoretical variogram models (Albert and McShane 1995; Gotway and Stroup 1997; Gumpertz, Wu, and Pye 1999). Another is the autologistic model studied in this article that models incidence probability by conditioning on the incidence in neighboring cells. In this study we focused on comparing three estimation methods for the second-order autologistic regression model in the application to the distribution of vegetation.

The MCL and MCMC-SA methods give nearly identical results in all aspects of estimation including ESM, ESD, and goodness-of-fit statistics both in the simulations and applications, while the MPL tends to produce somewhat different results. Based on the various goodness-of-fit statistics, the fit to the two vegetation types by the MPL estimates seems no worse than those obtained from the MCL and MCMC-SA methods. This result suggests that the MPL may be sufficient for most applications if statistical inference is not the major concern. However, proper precaution is necessary when using these goodness-of-fit statistics since it is not yet known what is the best way to measure the goodness-of-fit for autologistic models.

In all simulation studies, we used the MPL estimates as initial values and reference points for the MCL method and the algorithm worked well as suggested by Huffer and Wu (1998). However, for the boreal forest the MCL method using the MPL estimates as a starting point failed to produce reasonable results. The problem is likely caused by the large differences in β_0 and γ_3 between the MPL and the MCL methods (Table 3). This finding suggests that the MCL algorithm may not converge when the MPL estimates as initial values are not close enough to the resulting MCL estimates. This is not unusual with real data involving large spatial correlations or many covariates. For such data it frequently happens that the MPL estimates gives a bad reference point which leads the MCL algorithm to produces either no parameter estimates at all (the program crashes) or very bad ones. As suggested by a reviewer, when this happens various ad hoc approaches may be used to find a better reference point (e.g., Geyer and Thompson 1992). Huffer and Wu (1998) used the first step of the Newton–Raphson output as a new reference point. A better approach sometimes used is to obtain the reference point by a crude Robbins–Monro style stochastic approximation scheme. Stage I of the MCMC-SA algorithm is probably an excellent way to obtain these reference points, as suggested by the reviewer.

In summary, this study aims to evaluate the usefulness of the autologistic regression model for modeling the impacts of climate change on the distribution of vegetation and

species, and to compare the three algorithms. The results suggest that for species distribution data any of the three methods is probably adequate in estimating the autologistic model although the Monte Carlo likelihood method is favored because of its superior performance in the simulations. In our experience, the Markov chain Monte Carlo stochastic approximation is the most robust one. It worked rather well even for some very "nasty" problems (e.g., highly correlated data) on which both the pseudo-likelihood method and the Monte Carlo likelihood method failed. The disadvantage of the stochastic approximation method is its high demand on computation time. Considering the superiority and disadvantage of MCMC-SA, an ideal approach may be to combine the MCMC-SA and MCL algorithm as follows: compute the parameter estimates via MCL, but use the first stage of the MCMC-SA algorithm to compute the reference point required by MCL.

The applications of the autologistic regression model to the vegetation in BC have led to an important conclusion for studying the impact of climate change: different vegetation or species likely respond differently to different sets of climate variables. The effective assessment of climate change should be performed on the basis of individual vegetation or species.

ACKNOWLEDGMENTS

The authors are grateful to Julian Besag and three anonymous reviewers for their constructive and helpful comments, which greatly improved this article. Jim Lenihan and Ron Neilson kindly provided us with the vegetation and climate data. This research is supported by BC Advanced Systems Institute, Canadian Forest Service, Pacific Institute of Mathematical Sciences, and Natural Science and Engineering Research Council of Canada.

[Received April 2001. Revised April 2002.]

REFERENCES

Albert, P. S., and McShane, L. M. (1995), "A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data," *Biometrics*, 51, 627–638.

Arnold, H. R. (1993), *Atlas of Mammals in Britain*, London: Her Majesty's Stationery Office.

——— (1995), *Atlas of Amphibians and Reptiles in Britain*, London: Her Majesty's Stationery Office.

Augustin, N. H., Mugglestone, M. A., and Buckland, S. T. (1996), "An Autologistic Model for the Spatial Distribution of Wildlife," *Journal of Applied Ecology*, 33, 339–347.

Besag, J. (1972), "Nearest-Neighbour Systems and the Auto-logistic Model for Binary Data," *Journal of the Royal Statistical Society*, Series B, 34, 75–83.

——— (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Series B, 36, 192–236.

——— (1977), "Efficiency of Pseudo Likelihood Estimators for Simple Gaussian Fields," *Biometrika*, 64, 616–618.

Box, E. O., Crumpacker, D. W., and Hardin, E. D. (1993), "A Climatic Model for Location of Plant Species in Florida, USA," *Journal of Biogeography*, 20, 629–644.

Comets, F. (1992), "On Consistency of a Class of Estimators for Exponential Families of Markov Random Fields on the Lattice," *The Annals of Statistics*, 20, 455–468.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D., and Tanemura, M. (1994), "On Parameter Estimation for Pairwise Interaction Point Processes," *International Statistical Review*, 62, 99–117.

Diggle, P. J., and Tawn, J, A. (1998), "Model-Based Geostatistics" (with discussion), *Applied Statistics*, 47, 299–350.

Geyer, C. J. (1994), "On the Convergence of Monte-Carlo Maximum Likelihood Calculations," *Journal of the Royal Statistics Society*, Series B, 56, 261–274.

Geyer, C. J., and Thompson, E. A. (1992), "Constrained Monte Carlo Maximum Likelihood for Dependent Data" (with discussion), *Journal of the Royal Statistics Society*, Series B, 54, 657–699.

Gotway, C., and Stroup, W. (1997), "A Generalized Linear Model Approach to Spatial Data Analysis and Prediction," *Journal of Agricultural, Biological and Environmental Statistics*, 2, 157–178.

Gu, M. G., and Kong, F. H. (1998), "A Stochastic Approximation Algorithm with Markov Chain Monte Carlo Method for Incomplete Data Estimation Problems," *Proceedings of National Academic Science of USA*, 95, 7270–7274.

Gu, M. G., and Zhu, H. T. (2001), "Maximum Likelihood Estimation for Spatial Models by Markov Chain Monte Carlo Stochastic Approximation," *Journal of the Royal Statistics Society*, Series B, 63, 339–355.

Gumpertz, M. L., Graham, J. M., and Ristaino, J. B. (1997), "Autologistic Model of Spatial Pattern of Phytophthora Epidemic in Bell Pepper: Effects of Soil Variables on Disease Presence," *Journal of Agricultural, Biological and Environmental Statistics*, 2, 131–156.

Gumpertz, M. L., Wu, C. T., and Pye, J. M. (1999), "Logistic Regression for Southern Pine Beetle Outbreaks with Spatial and Temporal Autocorrelation," *Forest Science*, 46, 95–107.

Houghton, J. T., Jenkins, G. J., and Ephraums, J. J. (1990), *Climate Change, the IPCC Scientific Assessment*, Cambridge: Cambridge University Press.

Heikkinen, J., and Högmander, H. (1994)."Fully Bayesian Approach to Image Restoration With an Application in Biogeography," *Applied Statistics*, 43, 569–582.

Hoeting, J. A., Leecaster, M., and Bowden, D. (2000), "An Improved Model for Spatially Correlated Binary Responses," *Journal of Agricultural, Biological, and Environmental Statistics*, 5, 102–114.

Huang, F., and Ogata, Y. (2001), "Comparison of Two Methods for Calculating the Partition Functions of Various Spatial Statistical Models," *The Australian & New Zealand Journal of Statistics*, 43, 47–65.

Huffer, F. W., and Wu, H. L. (1998), "Markov Chain Monte Carlo for Auto-logistic Regression Models with Application to the Distribution of Plant Species," *Biometrics*, 54, 509–524.

Lamb, H. H. (1977), *Climate: Past, Present, and Future. Volume 2. Climatic History and the Future*, London: Methuen.

Lenihan, J. M., and Neilson, R. P. (1991),"A Rule-Based Vegetation Formation Model for Canada," *Journal of Biogeography*, 20, 615–628.

Little, E. J. Jr. (1971), *Atlas of United States Trees, Volume 1–5*, Washington, DC: U.S. Government Printing Office.

Mitchell-Jones, A. J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P. J. H., Spitzenberger, F., Stubbe, M., Thissen, J.B.M., Vohralk, V., and Zima, J. (1999), *The Atlas of European Mammals*, London: Poyser.

Wu, H., and Huffer, F. W. (1997), "Modelling the Distribution of Plant Species Using the Autologistic Regression Model," *Environmental and Ecological Statistics*, 4, 49–64.