Auto-multicategorical regression model for the distribution of vegetation

Hongtu Zhu[*], Fangliang He[†] and Julie Zhou[‡]

Modeling the contagious distribution of species has long been a challenging issue in ecology. Autologistic regression modeling has been a primary approach used to describe the spatially correlated distribution of single species on landscapes. Here we introduce a generalized auto multicategorical regression method to model the simultaneous distribution of multiple species. The auto multicategorical regression model includes the autologistic model as a special case. We develop a stochastic approximation algorithm for calculating the maximum likelihood estimates of parameters in the auto-multicategorical model. Based on the pseudo-likelihood likelihood function, we propose three model selection criteria for the selection of a 'good' model. A simulation study is carried out to examine the performance of the maximum likelihood estimation method and the three model selection criteria. An application of the model is provided through the analysis of the distribution of vegetation types in terms of climate variables in British Columbia, Canada.

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35.
Keywords and phrases: Auto-multicategorical regression, Maximum likelihood, Model selection, Pseudo-likelihood, Stochastic approximation, Vegetation.

1. INTRODUCTION

The spatial distribution of species is typically documented as the presence or absence of a species in a gridded map of equal cell size. Such data are essential for studying biodiversity patterns in ecology [10], understanding environmental determination of species distribution [9, 18, 14], predicting distribution in areas where observations are not available [15, 23], and planning biological conservation [25]. The majority of these studies have thus far focused on the distribution of a single species and its dependence on environmental covariates [3, 11, 31, 16, 14]. However, species rarely grow alone on landscapes; they live together, interacting and competing with each other for light, space, and nutrients. As a result, different species may dominate different local sites, forming a distribution as shown in Figure 1 (a).

An important feature in the distribution of species is that the distribution is spatially correlated. Statistical models for analyzing the correlated distribution can be divided roughly into two categories: (i) marginal logistic models that specify the marginal mean of a dependent variable as a function of covariates; and (ii) conditional auto-logistic models that define the full conditional distribution of the observation in each focal cell as a function of covariates and observations in other (neighboring) cells [30]. The inclusion of observations from neighboring cells as explanatory terms captures both the nature of species dispersal as well as the influence of underlying environmental factors which may lead species occurrence in neighboring pixels to be more similar than that in distant pixels. While several structures have been proposed to describe this spatial correlation, the first-order and second-order neighboring correlations [32, 13, 18, 16, 14] are most commonly used (Figure 2).

This study develops an auto-multicategorical regression model which is then used to analyze the distribution of several species of vegetation in British Columbia (BC) by considering the second-order correlation structure. The second-order correlation is viewed as a necessary structure in describing the spatial distribution of vegetation because the physiography of BC is characterized by mountain ranges in the northwest-southeast direction; the dispersal/migration processes in the distribution of vegetation may therefore impose a long-range effect. The secondary goals of this study are to (i) develop the maximum likelihood method for estimating parameters in the auto-multicategorical model; and (ii) develop and evaluate model selection criteria for choosing a 'good' model from a group of candidate models.

The remainder of the paper is organized as follows. In Section 2, the auto-multicategorical regression model is introduced. Section 3 describes a Markov chain Monte Carlo (MCMC) stochastic approximation (SA) algorithm for computing the maximum likelihood (ML) estimate. Three model selection criteria are proposed based on the pseudo-likelihood function. In Section 4, a simulation study is conducted to investigate the empirical performance of the ML estimation method and the three model selection criteria. In Section 5, the auto-multicategorical model is applied in an analysis of the distribution of four types of vegetation in BC. The study concludes with a discussion in Section 6.
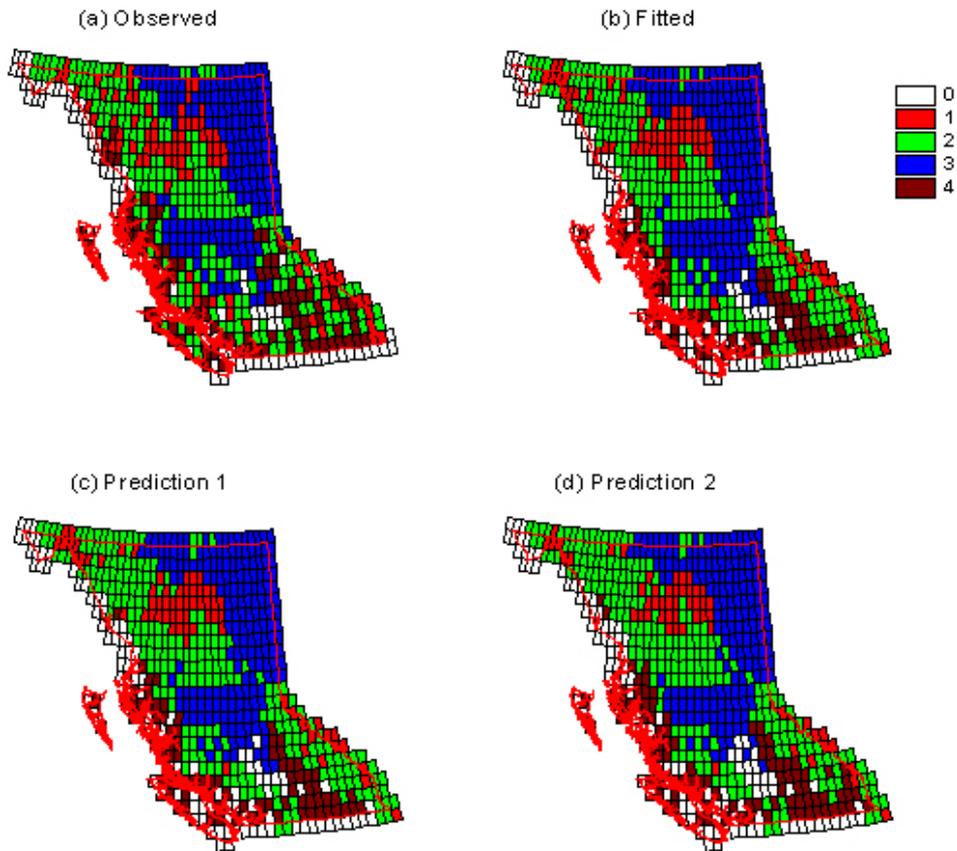
Figure 1. *Vegetation Distribution in British Columbia, Canada: 0 = Background, 1 = Low Arctic Shrub Tundra, 2 = Subarctic Evergreen Woodland, 3 = Boreal Evergreen Forest, and 4 = Temperate Evergreen Forest. (a) The Observed Distributions of Four Vegetation Types; (b) the Predicted Distributions of the Four Vegetations Based on MD(8); (c) the Predicted Distributions of the Four Vegetations with $1°C$ Increase in Temperature (X_1) Based on MD(8); and (d) the Predicted Distributions of the Four Vegetations with $3°C$ Increase in Temperature Based on MD(8). In Panels (b), (c), and (d), we Present Only the Species with the Largest Conditional Probability in Each Cell of BC Based on MD(8).*
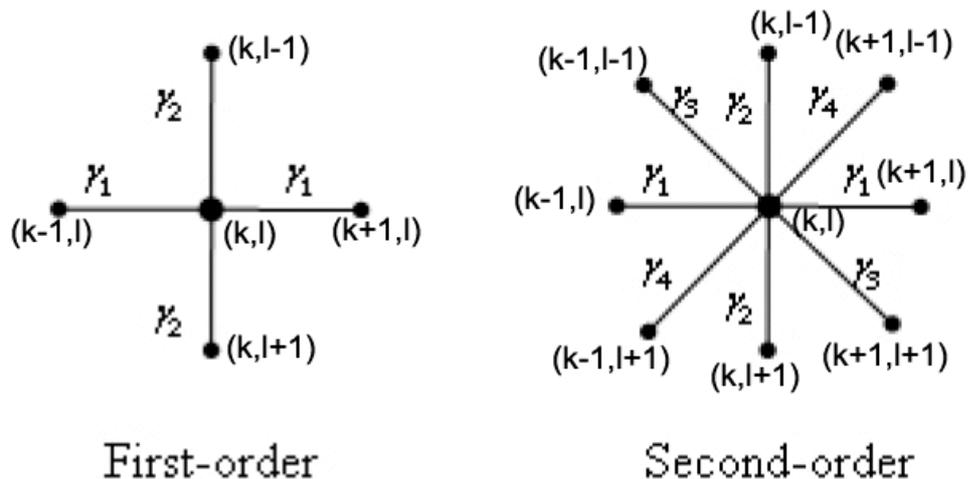

Figure 2. *Structure of the First-Order and Second-Order Neighborhoods. The Vertical (γ_1) and Horizontal (γ_2) are the First-Order Neighborhoods, whereas the Second-Order Neighborhoods Further Include the Diagonal Directions $(\gamma_3$: Southwest-Northeast Direction; γ_4: Northwest-Southeast Direction).*

2. THE AUTO-MULTICATEGORICAL MODEL

We consider a dataset $\{(X_{k,l}, Y_{k,l}) : (k,l) \in D\}$ recorded at n_D locations on a subset D of a finite lattice, where the coordinate (k,l) specifies the row and column of the lattice D at which it is located. At the site (k,l), $X_{k,l}$ is a $q \times 1$ covariate vector and response variable $Y_{k,l}$ has C possible values, say $0, 1, \ldots, C-1$. In this paper, $Y_{k,l} = 0$ denotes the absence of all species and $Y_{k,l} = j$ ($\neq 0$) denotes the presence of the jth species. Neither the ordering nor the difference between species numbers is meaningful.

A common model to analyze the data $\{(X_{k,l}, Y_{k,l}) : (k,l) \in D\}$ is a multicategorical response model [24], which assumes that the $Y_{k,l}$s are independent and satisfy

(1)
$$\Pr(Y_{k,l} = i | X_{k,l}) = \exp\{X_{k,l}^T \beta(i)\} \Big/ \sum_{j=0}^{C-1} \exp\{X_{k,l}^T \beta(j)\}$$

for $i = 0, \ldots, C-1$, where the superscript T denotes transpose, $\beta(0)$ is set to $\mathbf{0}$ to avoid redundancy, and $\beta(i)$ denotes the coefficients vector for the odds of $Y_{k,l} = i$ relative to $Y_{k,l} = 0$ for $i = 1, \ldots, C-1$. A key assumption of this model is independence among all $Y_{k,l}$. However, the assumption of independence is often unrealistic in ecology, as the presence/absence of a species in one site may have a potential effect on the presence/absence of the same species in the neighborhood. Therefore, an accurate model should incorporate the spatial dependence among all $Y_{k,l}$s.

A class of random fields that captures spatial correlation with wide applications in ecology is the Markov random field (MRF) on a finite lattice [5, 7]. Because D is usually an irregular lattice in most applications, we consider the joint distribution of the internal site responses $\mathbf{Y}^I = \{Y_{k,l} : (k,l) \in D^o\}$ conditional upon fixed boundary values $\mathbf{Y}^B = \{Y_{k,l} : (k,l) \in \partial D\}$, where ∂D and D^o, respectively, denote the set of all sites forming the boundary of D and the set of all internal sites of D. Explicitly, the probability function of the MRF of $\mathbf{Y}^I = \{Y_{k,l}, (k,l) \in D^o\}$ in a configuration space Ω given $\mathbf{Y}^B$ and covariates $\mathbf{X} = \{X_{k,l}, (k,l) \in D\}$ can be written in a Gibbsian form as follows:

(2)
$$p(\mathbf{Y}^I | \theta, \mathbf{Y}^B, \mathbf{X}) = \exp\{Q(\mathbf{Y}, \mathbf{X}; \theta)\} / Z_n(\theta),$$

where $\mathbf{Y} = (\mathbf{Y}^I, \mathbf{Y}^B)$, n denotes the number of sites in D^o, θ denotes all unknown parameters in $Q(\mathbf{Y}, \mathbf{X}; \theta)$, $Q(\mathbf{Y}, \mathbf{X}; \theta)$ is a potential (or energy) function, and $Z_n(\theta) = \sum_{\text{all } \mathbf{Y}^I \in \Omega} \exp\{Q(\mathbf{Y}, \mathbf{X}; \theta)\}$ is the normalizing factor, called a *partition function*. The potential function $Q(\mathbf{Y}, \mathbf{X}; \theta)$ usually includes contributions from each site of D and the second- and higher order interactions from multiple distinct neighboring sites.

For computational simplicity, we focus on a first- or second-order auto-multicategorical regression model. In this model the *potential/energy function* $Q(\mathbf{Y}, \mathbf{X}; \theta)$ depends only on contributions from cliques in which every pair of distinct sites are neighbors containing no more than two sites. For example, in the left panel of Figure 2, the site (k,l) has four first-order neighbors $(k, l-1)$, $(k-1, l)$, $(k, l+1)$, and $(k+1, l)$. Specifically, we assume that $Q(\mathbf{Y}, \mathbf{X}; \theta)$ is given by

(3)
$$Q(\mathbf{Y}, \mathbf{X}; \theta) = \sum_{(k,l) \in D^o} X_{k,l}^T \beta(y_{k,l})$$
$$+ \sum_{(k,l) \sim (k',l')} \phi_{k,l;k',l'}(y_{k,l}, y_{k',l'}),$$

where $\phi_{k,l;k',l'}(y_{k,l}, y_{k',l'})$ characterizes the relationship between $y_{k,l}$ and $y_{k',l'}$ and $(k,l) \sim (k',l')$ denotes that (k,l) and (k',l') are either first-order or second-order neighbors and that either (k,l) or (k',l') is in D^o (Figure 2). It should be noted that although we may include more complicated interactions between the sites in D, such models with complicated interactions can be computationally prohibitive. For simplicity, we consider only a special case of the *second-order* model, in which $\phi_{k,l;k',l'}(y_{k,l}, y_{k',l'})$ is defined as follows:

$$\phi_{k,l;k',l'}(y_{k,l}, y_{k',l'}) = \delta(y_{k,l}, y_{k',l'}) \times$$
$$\big[\gamma_1(y_{k,l})\mathbf{1}((k',l') = (k+1, l) \text{ or } (k-1, l)) +$$
$$\gamma_2(y_{k,l})\mathbf{1}((k',l') = (k, l+1) \text{ or } (k, l-1)) +$$
$$\gamma_3(y_{k,l})\mathbf{1}((k',l') = (k+1, l+1) \text{ or } (k-1, l-1)) +$$
$$\gamma_4(y_{k,l})\mathbf{1}((k',l') = (k+1, l-1) \text{ or } (k-1, l+1))\big],$$

where $\mathbf{1}(\cdot)$ is an indicator function and $\delta(y_{k,l}, y_{k',l'})$ equals one whenever $y_{k,l} = y_{k',l'}$ and zero otherwise. We set $\gamma_i(0) = 0$ for $i = 1, 2, 3$ and 4 to avoid redundancy. Parameters $\gamma_i(j)$ represent spatial correlations along certain directions. Specifically, $\gamma_1(\cdot)$, $\gamma_2(\cdot)$, $\gamma_3(\cdot)$, and $\gamma_4(\cdot)$ denote spatial correlations along the north-south, west-east, southwest-northeast, and northwest-southeast directions, respectively (Figure 2). Moreover, because different species have different distribution patterns (Figure 1), we allow $\gamma_i(j)$ to vary with j. Furthermore, if we set $\gamma_3(y_{k,l}) = \gamma_4(y_{k,l}) = 0$, then the second-order auto-multicategorical regression model reduces to the *first-order* auto-multicategorical regression model. If $X_{k,l} = 0$ and $\gamma = \gamma(1) = \cdots = \gamma(C-1)$, then the potential function of model (3) reduces to $\gamma \sum_{(k,l) \sim (k',l')} \delta(y_{k,l}, y_{k',l'})$, which is the well-known *Potts model* [27].

For the general first- and second-order auto-categorical regression models, the potential function $Q(\mathbf{Y}, \mathbf{X}; \theta)$ can be written as follows:

(4)
$$Q(\mathbf{Y}, \mathbf{X}; \theta) = S(\mathbf{Y})^T \theta.$$

For instance, in the second-order auto-categorical regression model, we have

$$\theta = (\beta(1)^T, \gamma_1(1), \ldots, \gamma_4(1), \ldots, \beta(C-1)^T,$$
$$\gamma_1(C-1), \ldots, \gamma_4(C-1))^T$$
$$S(\mathbf{Y}) = \left(\sum_{(k,l) \in D^o} \delta(y_{k,l}, 1) X_{k,l}^T, F_1(1), \ldots, F_4(1), \right.$$
$$\ldots, \sum_{(k,l) \in D^o} \delta(y_{k,l}, C-1) X_{k,l}^T, F_1(C-1),$$
$$\left. \ldots, F_4(C-1) \right)^T,$$

where

$$F_1(i) = \sum_{(k,l) \in D^o \text{ or } (k+1,l) \in D^o} \delta(y_{k+1,l}, i),$$
$$F_2(i) = \sum_{(k,l) \in D^o \text{ or } (k,l+1) \in D^o} \delta(y_{k,l+1}, i),$$
$$F_3(i) = \sum_{(k,l) \in D^o \text{ or } (k+1,l+1) \in D^o} \delta(y_{k+1,l+1}, i),$$
$$F_4(i) = \sum_{(k,l) \in D^o \text{ or } (k+1,l-1) \in D^o} \delta(y_{k+1,l-1}, i).$$

Thus, θ includes both $\beta(j)$ for the odds of $Y_{k,l} = j$ relative to $Y_{k,l} = 0$ and $\gamma_i(j)$ for describing the spatial correlations among multiple species.

It follows from (2) and (4) that the conditional probability distribution of $Y_{k,l} \in D^o$ given its neighborhood is

(5)

$$\Pr(Y_{k,l} = i | \text{all other sites}) = \Pr_{k,l}(i|\theta)$$
$$= \frac{\exp\{f_{k,l}(i|\theta, N(y_{k,l}))\}}{\sum_{j=0}^{C-1} \exp\{f_{k,l}(j|\theta, N(y_{k,l}))\}},$$

where $f_{k,l}(y_{k,l}|\theta, N(y_{k,l})) = \theta^T G_{k,l}(y_{k,l}; N(y_{k,l}))$ and $G_{k,l}(y_{k,l}; N(y_{k,l}))$ is a function of $y_{k,l}$ and the set of its neighbors, denoted by $N(y_{k,l})$. For instance, for the second-order model, $f_{k,l}(y_{k,l} = i|\theta, N(y_{k,l}))$ is given by

(6)

$$f_{k,l}(y_{k,l} = i|\theta, N(y_{k,l})) = \theta^T G_{k,l}(y_{k,l}; N(y_{k,l}))$$
$$= X_{k,l}^T \beta(i) + \sum_{m=1}^{4} \gamma_m(i) y_{k,l}^{*(m)}(i)$$

for $i = 0, \ldots, C-1$, where $y_{k,l}^{*(1)}(i)$ is the number of two sites $\{(k, l-1), (k, l+1)\}$ coded i, $y_{k,l}^{*(2)}(i)$ is the number of two sites $\{(k-1, l), (k+1, l)\}$ coded i, $y_{k,l}^{*(3)}(i)$ is the number of two sites $\{(k-1, l-1), (k+1, l+1)\}$ colored i, and $y_{k,l}^{*(4)}(i)$ is the number of two sites $\{(k-1, l+1), (k+1, l-1)\}$ colored i. Specifically, for the second-order model, $G_{k,l}(y_{k,l}; N(y_{k,l}))$ is a $(q+4)(C-1)$ vector such that $G_{k,l}(y_{k,l} = 0; N(y_{k,l})) = \mathbf{0}$

and $G_{k,l}(y_{k,l} = i; N(y_{k,l}))$ is given by

$$(\underbrace{0, \ldots, 0}_{(i-1)(q+4)}, X_{k,l}^T, y_{k,l}^{*(1)}(i), y_{k,l}^{*(2)}(i), y_{k,l}^{*(3)}(i),$$
$$y_{k,l}^{*(4)}(i), \underbrace{0, \ldots, 0}_{(C-i-1)(q+4)})^T$$

for $i = 1, \ldots, C-1$. Furthermore, for the first-order model, because $\gamma_3(i) = \gamma_4(i)$ are set to 0, $G_{k,l}(y_{k,l} = i; N(y_{k,l}))$ is given by

$$(\underbrace{0, \ldots, 0}_{(i-1)(q+2)}, X_{k,l}^T, y_{k,l}^{*(1)}(i), y_{k,l}^{*(2)}(i), \underbrace{0, \ldots, 0}_{(C-i-1)(q+2)})^T$$

for $i = 1, \ldots, C-1$.

3. ESTIMATION METHODS AND MODEL SELECTION

3.1 Maximum likelihood estimate

The maximum pseudo-likelihood (MPL) and ML methods have been used widely to estimate θ in spatial models such as the autologistic regression model [5, 18, 12, 14]. Each estimation method has characteristic advantages and disadvantages. The MPL method is computationally simple and easy to implement in statistical packages such as SAS and SPlus. The ML method, by comparison, is computationally intensive due to the intractable partition function. However, when the spatial correlation is relatively high, the MPL estimate is inefficient in comparison to the ML estimate [26, 12].

We propose to use the MPL estimate as the starting point and then use a stochastic approximation algorithm to compute the ML estimate, denoted by $\hat{\theta}$. For model (2), the MPL estimate $\hat{\theta}_P$ maximizes the log pseudo-likelihood function $\ell_{P,n}(\theta) = \ell_{P,n}(\theta, \mathbf{Y})$, which is given by

$$\ell_{P,n}(\theta) = \sum_{(k,l) \in D^o} \left\{ f_{k,l}(y_{k,l}|\theta) - \log\left[\sum_{i=0}^{C-1} \exp\{f_{k,l}(i|\theta)\} \right] \right\}$$
$$= \sum_{(k,l) \in D^o} \left\{ \theta^T G_{k,l}(y_{k,l}; N(y_{k,l})) \right.$$
$$\left. - \log\left[\sum_{i=0}^{C-1} \exp\{\theta^T G_{k,l}(y_{k,l} = i; N(y_{k,l}))\} \right] \right\}.$$

Letting $U_{k,l}(\theta) = G_{k,l}(y_{k,l}; N(y_{k,l})) - E_{k,l}\{G_{k,l}(y_{k,l}; N(y_{k,l}))\}$, we get

(7)
$$\nabla \ell_{P,n}(\theta) = \sum_{(k,l) \in D^o} U_{k,l}(\theta) \quad \text{and}$$
$$-\nabla^2 \ell_{P,n}(\theta) = \sum_{(k,l) \in D^o} \text{Var}_{k,l}\{U_{k,l}(\theta)\},$$

where $E_{k,l}$ denotes the expectation with respect to (5). Because $-\bigtriangledown^2 \ell_{P,n}(\theta)$ is always positive semi-definite, the log pseudo-likelihood function is concave. Furthermore, $\hat{\theta}_P$ is unique when $\ell_{P,n}$ is locally strictly concave in a neighbor of θ_0. We use the Newton-Raphson algorithm to estimate $\hat{\theta}_P$ directly.

We note several important facts about the log-likelihood function for observed-data $\mathbf{y}_{0,n}$ given by

$$(8) \qquad \ell_n(\theta) = \ell(\theta; \mathbf{y}_{0,n}) = \theta^T S(\mathbf{y}_{0,n}) - \log Z_n(\theta).$$

Because the partition function $Z_n(\theta)$ involves summing over all possible configuration $\mathbf{Y}^I$, the observed-data log-likelihood function is computationally intractable. Furthermore, the first and second derivatives of $\ell_n(\theta)$ can be written by

$$(9) \qquad \bigtriangledown \ell_n(\theta) = S(\mathbf{y}_{0,n}) - \bigtriangledown \log Z_n(\theta) \quad \text{and}$$
$$\bigtriangledown^2 \ell_n(\theta) = -\bigtriangledown^2 \log Z_n(\theta),$$

where $\bigtriangledown$ and $\bigtriangledown^2$ denote the first and second derivatives with respect to θ, respectively. 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 the expectation with respect to the density in (2), we can show that

$$(10) \qquad \bigtriangledown \log Z_n(\theta) = E_\theta\{S(\mathbf{Y})\}, \quad \text{and}$$
$$\bigtriangledown^2 \log Z_n(\theta) = E_\theta\{S(\mathbf{Y})^{\otimes 2}\} - [E_\theta\{S(\mathbf{Y})\}]^{\otimes 2}$$
$$= \text{Var}_\theta\{S(\mathbf{Y})\},$$

where for vector $\mathbf{a}$, $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. Thus, if $\text{Var}_\theta\{S(\mathbf{Y})\}$ is always positive semi-definite, then the log-likelihood function $\ell_n(\theta)$ is globally concave.

Calculating the ML estimate $\hat{\theta} = \text{argmax}_\theta \ell_n(\theta)$ requires computation of $Z_n(\theta)$. Because the number of terms in $Z_n(\theta)$ is enormous even for today's computers, it is not feasible to maximize $\ell_n(\theta)$ directly, even in the modern computing environment. Instead, we approximate the first and second derivatives of $\ell_n(\theta)$ and then use the Newton-Raphson algorithm to calculate the ML estimate in spatial models. Such estimation methods for calculating $\hat{\theta}$ include the Monte Carlo likelihood method [18] and the Markov chain Monte Carlo stochastic approximation (MCMC-SA) algorithm [12].

We adapt the MCMC-SA algorithm of [12] to calculate $\hat{\theta}$ for the auto-multicategorical regression model. The MCMC-SA algorithm is valid for calculating the ML estimate for a large class of spatial models which includes model (2) as a special case. The MCMC-SA algorithm combines MCMC methods, such as the Gibbs sampler, with SA methods, such as adaptive search direction. In particular, the MCMC methods can be used to simulate samples from model (2) approximately, while the SA methods can be used to handle noise in approximating $\nabla \ell_n(\theta)$ and $\nabla^2 \ell_n(\theta)$. The final estimates of $\hat{\theta}$ and $-\nabla^2 \ell_n(\hat{\theta})$ are used as the ML estimate and its covariance matrix, respectively. For model (2), we include

detailed steps of the MCMC-SA algorithm in the Appendix. Although the MCMC-SA algorithm usually converges to a local optimum as pointed out by a reviewer, it does converge to a global maximum of model (2) when $\text{Var}_\theta\{S(\mathbf{Y})\}$ is positive semi-definite for all θ, because $\ell_n(\theta)$ is globally concave.

3.2 Model selection procedures based on the pseudo-likelihood function

We now discuss methods for selecting the "best" approximate model from a group of candidate models $MD(m) = \{p(\mathbf{Y}|\theta_{(m)}) : \theta_{(m)} \in \Theta_m\}$ with $m = 1, \ldots, M_0$ based on both the pseudo-likelihood functions. Each candidate model $MD(m)$ is determined by unknown parameters $\theta_{(m)}$, the dimension of $\theta_{(m)}$ denoted by $p(m)$, and the associated neighborhood system.

We develop three model selection criteria based on the pseudo-likelihood function. Because direct computation of the likelihood function $\ell_n(\theta)$ is extremely difficult, we propose to use the pseudo-likelihood function, which is computationally easy, to construct a model selection criterion. The model selection criteria for the pseudo-likelihood function are pseudo AIC (PAIC), pseudo CAIC (PCAIC), and pseudo BIC (PBIC). These are defined as follows:

$$(11) \qquad \text{PCAIC(m)} = -2\ell_{P,n}(\hat{\theta}_{P,(m)})$$
$$+ 2\text{tr}\{\hat{I}_{P,n}(\hat{\theta}_{P,(m)})\hat{V}(\theta_0)^{(3)}\}$$
$$+ p(m) \log n,$$

$$(12) \qquad \text{PAIC(m)} = -2\ell_{P,n}(\hat{\theta}_{P,(m)})$$
$$+ 2\text{tr}\{\hat{I}_{P,n}(\hat{\theta}_{P,(m)})\hat{V}(\theta_0)^{(3)}\}, \quad \text{and}$$

$$(13) \qquad \text{PBIC(m)} = -2\ell_{P,n}(\hat{\theta}_{P,(m)}) + p(m) \log n.$$

In practice, we select the best model with minimum PAIC(m) (or PCAIC(m), or PBIC(m)) over $m = 1, \ldots, M_0$. A simulation study in Section 4.2 provides an empirical performance of {PAIC, PCAIC, PBIC}. Following Sin and White (1996), we may derive the consistency of {PAIC, PCAIC, PBIC} under some conditions, which are beyond the scope of this paper. It has been well known that for independent data structure, information criteria, such as AIC, with the constant penalty tends to overfit the data. Thus, we expect that PAIC(m) with the penalty $2\text{tr}\{\hat{I}_{P,n}(\hat{\theta}_{P,(m)})\hat{V}(\theta_0)^{(3)}\}$ may overfit the data in this spatial data structure.

4. A SIMULATION STUDY

4.1 Performance of the ML estimates

We considered a 40×40 lattice and the observed categorical response had three possible values: 0, 1, and 2 at each site (k, l). We considered a first-order auto-multicategorical

Table 1. Summary Statistics and their Sample Standard Deviations (in Parenthesis) for 500 Samples Based on the Three Different True Values of θ

$\theta = (1, -1, 0.2, -1, 1, 0.2)^T$ AVEN: 947					
θ	True	EST	MSD	0.95	0.90	0.30
$\beta_1(1)$	1	1.022 (0.126)	0.132 (0.010)	0.978 (0.007)	0.936 (0.011)	0.298 (0.021)
$\beta_1(2)$	-1	-1.005 (0.129)	0.131 (0.010)	0.956 (0.009)	0.900 (0.013)	0.318 (0.020)
$\gamma(1)$	0.2	0.189 (0.061)	0.064 (0.004)	0.960 (0.009)	0.918 (0.012)	0.264 (0.020)
$\beta_2(1)$	-1	-1.027 (0.131)	0.130 (0.010)	0.954 (0.009)	0.892 (0.014)	0.300 (0.021)
$\beta_2(2)$	1	1.032 (0.134)	0.131 (0.010)	0.952 (0.014)	0.896 (0.021)	0.270 (0.020)
$\gamma(2)$	0.2	0.187 (0.069)	0.064 (0.005)	0.936 (0.011)	0.884 (0.014)	0.274 (0.020)
$\theta = (1, -1, 0.5, -1, 1, 0.5)^T$ AVEN $= 1003$					
θ	True	EST	MSD	0.95	0.90	0.30
$\beta_1(1)$	1	1.029 (0.182)	0.179 (0.017)	0.946 (0.010)	0.884 (0.014)	0.316 (0.021)
$\beta_1(2)$	-1	-1.017 (0.178)	0.178 (0.017)	0.946 (0.010)	0.896 (0.014)	0.286 (0.020)
$\gamma(1)$	0.5	0.493 (0.073)	0.071 (0.005)	0.942 (0.011)	0.896 (0.014)	0.334 (0.021)
$\beta_2(1)$	-1	-1.024 (0.173)	0.174 (0.017)	0.954 (0.009)	0.916 (0.012)	0.310 (0.021)
$\beta_2(2)$	1	1.032 (0.181)	0.175 (0.017)	0.938 (0.011)	0.900 (0.013)	0.284 (0.020)
$\gamma(2)$	0.5	0.496 (0.072)	0.070 (0.005)	0.944 (0.010)	0.884 (0.014)	0.308 (0.021)
$\theta = (1, -1, 1, -1, 1, 1)^T$ AVEN $= 1239$					
θ	True	EST	MSD	0.95	0.90	0.30
$\beta_1(1)$	1	1.066 (0.353)	0.337 (0.054)	0.954 (0.009)	0.904 (0.013)	0.286 (0.020)
$\beta_1(2)$	-1	-1.081 (0.354)	0.338 (0.055)	0.942 (0.011)	0.894 (0.014)	0.270 (0.020)
$\gamma(1)$	1	1.010 (0.109)	0.102 (0.010)	0.950 (0.010)	0.884 (0.014)	0.264 (0.020)
$\beta_2(1)$	-1	-1.059 (0.339)	0.327 (0.051)	0.958 (0.009)	0.900 (0.013)	0.312 (0.021)
$\beta_2(2)$	1	1.046 (0.332)	0.329 (0.052)	0.964 (0.008)	0.916 (0.012)	0.274 (0.020)
$\gamma(2)$	1	1.012 (0.105)	0.101 (0.009)	0.932 (0.011)	0.894 (0.014)	0.294 (0.020)

model with $f_{k,l}(i|\theta, N(y_{k,l}))$ given by

$$(14) \quad f_{k,l}(i|\theta, N(y_{k,l})) = \beta_1(i)x_{k,l}^{(1)} + \beta_2(i)x_{k,l}^{(2)}$$
$$+ \gamma(i)\{y_{k,l}^{*(1)}(i) + y_{k,l}^{*(2)}(i)\},$$

for $i = 1, 2$, where $x_{k,l}^{(1)} = 2.0 \times \cos(0.2 \times (k + 1))$ and $x_{k,l}^{(2)} = 2.0 \times \sin(0.2 \times (k + l))$. In our simulation, $\beta_1 = (\beta_1(1), \beta_1(2))^T = (1, -1)^T$, $\beta_2 = (\beta_2(1), \beta_2(2))^T = (-1, 1)^T$, and $(\gamma(1), \gamma(2))$ varied as $(0.2, 0.2)$, $(0.5, 0.5)$, and $(1.0, 1.0)$. The total number of the unknown parameters was six.

To simulate the process, the Gibbs sampler was used. The initial state of the process was determined at random such that $Y_{k,l}$ at each site (k, l) was taken independently to be $2/1/0$ with $1/3$ probability. Then, the Gibbs sampler was repeated 2000×40^2 times (2000 Monte Carlo steps) to ensure that equilibrium states are achieved. For each parameter vector $\theta = (\beta_1(1), \beta_1(2), \gamma(1), \beta_2(1), \beta_2(2), \gamma(2))^T$, 500 datasets were generated. We set $(a_1, b_1, K_0, \eta_1; a_2, b_2, N, \eta_2)$ as $(0.6, 2.0, 100, 0.1; 0.8, 1.0, 10000, 0.003)$ in the MCMC-SA algorithm.

Based on 500 parameter estimates, we calculated the following quantities: the estimated sample mean (EST) and its sample standard deviation, the mean of the estimated standard error estimates (MSD) and its sample standard deviation, and the mean number of iterations for each estimate (AVEN). All relative efficiencies (the ratio of the mean of the standard deviation estimates to the root mean-square error) are close to 1.0 (Table 1). To evaluate whether $r_i = (\hat{\theta}_i - \theta_i)/\hat{\sigma}_i$ is distributed approximately $N(0, 1)$, we constructed the $(1-\alpha)$ confidence interval of θ as $\hat{\theta}_i \pm z_{\alpha/2}\hat{\sigma}_i$, where $\hat{\theta}_i$ was an estimate of θ, $\hat{\sigma}_i$ was the standard error estimate, and $z_{\alpha/2}$ was the normal quartile corresponding to the $(1 - \alpha)$ confidence level. Based on the parameter estimates and standard error estimates, we constructed confidence intervals and calculated the actual coverage probabilities of the $95\%, 90\%$, and 30% confidence intervals and their corresponding sample standard deviation for each set of true parameter values [18]. To test whether the coverage probabilities were equal to the nominal value, the exact Binomial two-sided test was used. Under the significance level $\alpha = 0.05$, the critical values for the various confidence levels were: $95\%(0.930, 0.969)$, $90\%(0.872, 0.926)$, and $30\%(0.260, 0.341)$. The ML estimates obtained via the MCMC-SA algorithm and the estimated standard error estimates were accurate (Table 1).

4.2 Comparison of model selection criteria

In this section, we compare the empirical performance of the six model selection criteria based on the pseudo-likelihood and likelihood functions. Following the simulation study in Section 4.1, we considered model (2) with $f_{k,l}(y_{k,l} = i|\theta, N(y_{k,l}))$ given in (14). Moreover, we chose the following $M_0 = 4$ models as our candidate models with

Table 2. Relative Performance of Six Model Selection Criteria in 500 Replications of the Monte Carlo Experiment

	Candidate models				$\theta_0 = (1, -1, 0 - 1, 1, 0)^T$	
Criterion	1*	2	3	4	Pr(overfit)	Pr(underfit)
PAIC	396	54	41	9	0.208	0.0
PCAIC	499	1	0	0	0.002	0.0
PBIC	498	1	0	0	0.002	0.0
	Candidate models				$\theta_0 = (1, -1, 0.5, -1, 1, 0.5)^T$	
Criterion	1	2*	3	4	Pr(overfit)	Pr(underfit)
PAIC	0	383	76	41	0.234	0.0
PCAIC	0	499	1	0	0.002	0.0
PBIC	0	487	13	0	0.026	0.0
	Candidate models				$\theta_0 = (1, -1, 1, -1, 1, 1)^T$	
Criterion	1	2*	3	4	Pr(overfit)	Pr(underfit)
PAIC	0	365	85	50	0.270	0.0
PCAIC	0	500	0	0	0.000	0.0
PBIC	0	491	9	0	0.018	0.0

their $f_{k,l}(y_{k,l} = i|\theta, N(y_{k,l}))$ given by:

MD(1): $\quad f_{k,l}(i|\theta, N(y_{k,l})) = \beta_1(i)x_{k,l}^{(1)} + \beta_2(i)x_{k,l}^{(2)};$

MD(2): $\quad f_{k,l}(i|\theta, N(y_{k,l})) = \beta_1(i)x_{k,l}^{(1)} + \beta_2(i)x_{k,l}^{(2)}$
$$+ \gamma(i)[y_{k,l}^{*(1)}(i) + y_{k,l}^{*(2)}(i)];$$

MD(3): $\quad f_{k,l}(i|\theta, N(y_{k,l})) = \beta_1(i)x_{k,l}^{(1)} + \beta_2(i)x_{k,l}^{(2)}$
$$+ \gamma_1(i)y_{k,l}^{*(1)}(i) + \gamma_2(i)y_{k,l}^{*(2)}(i);$$

MD(4): $\quad f_{k,l}(i|\theta, N(y_{k,l})) = \beta_1(i)x_{k,l}^{(1)} + \beta_2(i)x_{k,l}^{(2)}$
$$+ \sum_{m=1}^{4} \gamma_m(i)y_{k,l}^{*(m)}(i).$$

We chose the following two scenarios. In the first scenario, we set $\gamma(1) = \gamma(2) = 0$, and thus MD(1) was the true model. In the second scenario, we set either $\gamma(1) = \gamma(2) = 0.5$ or $\gamma(1) = \gamma(2) = 1.0$, and thus $MD(2)$ was the true model. We again used 500 simulated datasets to compare the finite performance of the three model selection criteria under those two scenarios.

The PCAIC selection criterion performed best with the smallest probability of overfit (Table 2). Moreover, the PBIC criterion worked very well with smaller probability of overfit in all three cases, whereas PAIC had relatively high probability of overfit. The probabilities of underfit for all three selection criteria were zero in this simulation study.

5. MODELING THE DISTRIBUTION OF VEGETATION IN BRITISH COLUMBIA

In this section we use the auto-multicategorical model to analyze the distribution of vegetation in British Columbia. The data we used are a subset of a large dataset from a previous study that addresses the vegetation distribution of all of Canada from an ecophysiological perspective [22]. Our goals were to model the relationship between the distribution of vegetation and the climate, and to assess the

potential impact of climate change on vegetation distribution.

Four major types of vegetation in BC and a general background condition are described below:

V0—background-consisting of human habitation, glaciation, and other land cover;

V1—low arctic shrub tundra-dominated by dwarf shrubs and graminoids but absent of upright trees;

V2—subarctic evergreen woodland—the broad ecotone between tundra (at higher elevation) and closed forest (at lower elevation), i.e., between the treeline and the continuous forest line;

V3—boreal evergreen forest-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 occupying the central mountainous region and the northeastern upland;

V4—temperate evergreen forest-characterized by the dominance of coniferous species such as Sitka spruce (*Picea sitchensis* (Bong.) Car.), western hemlock (*Tsuga heterophylla* (Raf.) Sarg.), western red cedar (*Thuja plicata Donn.*) and Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), and restricted to the coastal and intermountain region of southern BC.

The entirety of BC is divided into a lattice of 707 cells with cell size $= 0.5° \times 0.5°$ (Figure 1). The number of observations for each type of vegetation is summarized in Table 3.

In each cell, there are records for 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

Table 3. Number of Observations for Background Landbase and Four Vegetation Types in British Columbia

	V0	V1	V2	V3	V4	Total
NUM	127	89	211	188	92	707
%	17.9	12.6	29.8	26.7	13.0	100

Table 4. ML Estimates of MD(8) for the Vegetation Data

	$\beta(1)$						$\gamma_1(1)$
EST	−3.771	−0.201	−0.004	0.016	−0.012	−0.004	0.132
SD	3.735	0.065	0.001	0.015	0.009	0.0008	0.205
	$\beta(2)$						$\gamma_1(2)$
EST	−3.898	−0.159	−0.0007	0.005	−0.008	−0.003	0.314
SD	2.425	0.043	0.0008	0.010	0.004	0.0006	0.125
	$\beta(3)$						$\gamma_1(3)$
EST	−12.834	−0.262	0.0009	0.008	−0.006	−0.0035	1.356
SD	3.151	0.059	0.001	0.012	0.004	0.001	0.158
	$\beta(4)$						$\gamma_1(4)$
EST	−0.846	−0.029	−0.001	0.010	−0.001	−0.002	0.883
SD	2.260	0.040	0.001	0.008	0.003	0.0005	0.16

months (X_3), annual soil moisture deficit (mm) (X_4), and annual snowpack (mm) (X_5). These climate variables are calculated from two fundamental climatic factors: monthly temperature and monthly precipitation over a 30 year period (Lenihan and Neilson, 1993). These five variables are expected to have significant effects on the distribution of vegetation as global warming continues to change climactic conditions.

We considered 13 different auto-multicategorical models, each having different correlation structures distinguished by different orders and constraints. See Figure 2 for the notation of spatial structures. To avoid clutter, we denote the first- and second-order structures by FST and SEC, respectively. MD(1) denotes the multicategorical response model, while the candidate models, MD(2) to MD(7), are given as follows:

> MD(2): FST with $\gamma_1 = \gamma_2$;
> MD(3): FST with γ_1, γ_2;
> MD(4): SEC with $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4$;
> MD(5): SEC with $\gamma_1 = \gamma_2, \gamma_3 = \gamma_4$;
> MD(6): SEC with $\gamma_1, \gamma_2, \gamma_3 = \gamma_4$;
> MD(7): SEC with $\gamma_1, \gamma_2, \gamma_3, \gamma_4$.

In the above models, we assume the implicit constraint that $\gamma_j(i)$ is independent of species i, thus omitting (i) in the $\gamma_j(i)$. For the candidate models MD(8) to MD(13), we relaxed the above constraint and considered some models in which the $\gamma_j(i)$s depend on species i. Those models are given as follows:

> MD(8): FST with $\gamma_1(i) = \gamma_2(i)$;
> MD(9): FST with $\gamma_1(i), \gamma_2(i)$;
> MD(10): SEC with $\gamma_1(i) = \gamma_2(i) = \gamma_3(i) = \gamma_4(i)$;
> MD(11): SEC with $\gamma_1(i) = \gamma_2(i), \gamma_3(i) = \gamma_4(i)$;
> MD(12): SEC with $\gamma_1(i), \gamma_2(i), \gamma_3(i) = \gamma_4(i)$;
> MD(13): SEC with $\gamma_1(i), \gamma_2(i), \gamma_3(i), \gamma_4(i)$.

We applied the PAIC, PCAIC and PBIC selection criteria to select the best model among the above 13 candidate models. Judged by the model selection criteria, our results (not presented) showed that neither the multicategorical regression model nor the first-order model was favored. This indicates the need to incorporate the second-order and higher-order correlation structures. As expected (Section 4), the three selection criteria {PAIC, PCAIC, PBIC} did not agree with each other. For example, PCAIC favored MD(8), PAIC favored MD(12), while PBIC favored MD(10). Based on our simulation study (Table 2), we chose MD(8) as the best model. We used the two-stage MCMC-SA algorithm and set $(a_1, b_1, K_0, \eta_1, N, a_2, b_2, \eta_2) = (0.6, 2, 100, 0.1, 10000, 0.8, 1, 0.002)$ to obtain the ML estimates. We calculated the ML estimates of θ in MD(8) (Table 4).

The significance of the ML estimates of $\{\gamma_1(2), \gamma_1(3), \gamma_1(4)\}$ reflected the importance of the spatial correlation structure. In addition to the spatial coefficients, the coldest month (X_1) and the annual snowpack (X_5) were important covariates determining the distribution of the vegetation in BC. We also calculated the conditional probability of each vegetation type in each cell in BC and presented the vegetation type with the largest conditional probability in each cell of BC (Figure 1 (b)).

It is interesting to predict the redistribution of vegetation under various scenarios of global warming. It has been predicted that an enhanced greenhouse effect (e.g., doubling atmospheric CO_2 concentration) would increase the global mean temperature at a magnitude of 1.5 to 4.5 °C in next 50 years, with even a higher increase in temperature in the Northern Hemisphere [21, 17]. One advantage of the auto-multicategorical model is that the effect of climate change can be quantified through the odds ratio of the conditional probabilities. For example, if the minimum temperature X_1 increases by one degree while other variables remain constant, then the odds ratio of the conditional probabilities of the j-th vegetation presence is predicted to increase by a

factor $e^{\beta_2(j)}$. For the four vegetation types V1–V4, if temperature increases $1\,^\circ$C, then the odds ratios are multiplied by $0.818(= e^{-0.201})$, $0.853(= e^{-0.159})$, $0.770(= e^{-0.262})$, and $0.977(= e^{-0.029})$, respectively. The odds of V1, V2, V3, and V4 presence with 95% confidence interval in each 0.5° latitude and 0.5° longitude cell would decrease approximately by 18.2% (4.78%, 29.74%), 14.7% (4.97%, 23.33%), 23.0% (12.02%, 32.70%), and 2.3% (-5.23%, 11.3%), respectively. This result suggests that low arctic shrub tundra, subarctic evergreen woodlands, and boreal evergreen forests in BC would all be substantially reduced by global warming. More formally, we used the Gibbs sampler to simulate from the fitted model MD(8) incorporating the change of temperature. We then estimated the conditional probability of each vegetation type in each cell. The predicted redistributions of the vegetation types with the largest conditional probability in each cell of BC due to the impact of climate change are shown in Figure 1 (c) and (d).

6. DISCUSSION

Although the autologistic regression model has been widely used to model spatial distribution of species [4, 3, 31, 18, 16, 14], the application of auto-multicategorical regression to modeling the distribution of multiple species has not yet been reported. A key limitation of the use of the autologistic regression model is that it does not distinguish between all other species and the true empty cells, treating them together as 'absence'. As such, it can only model one species at a time. Because the environmental conditions in those 'absent' cells are not homogeneous, it is an inefficient model both statistically and biologically. Different species clearly have different preferences for varied environmental conditions, and the auto-logistic regression model cannot model these different preferences simultaneously. In such cases, the auto-multicategorical regression model must be used. It is clear, then, that the auto-multicategorical regression model provides a useful approach to analyzing multiple species simultaneously, such as in the current example involving the distribution of vegetation in BC. In addition, the autocorrelation structure in model (3) describes spatially contagious processes in species distribution, which extends the Potts model. Furthermore, by using the auto-multicategorical regression model, we can directly estimate the probability of the presence of each species in a given focal cell and predict the impact of temperature changes on their distributions.

In this study we have developed the ML estimate for the auto-multicategorical model and proposed three model selection criteria based on the pseudo-likelihood function. Our simulation and application results have shown that our proposed methods work well for the auto-multicategorical regression model. Although the ML estimate is more efficient than the MPL estimate in some cases [12, 26], the disadvantage of the ML method is that it is little more computationally costly than the MPL method.

A toolbox for implementing statistical procedure for the auto-multicategorical regression has been made available at http://www.bios.unc.edu/~hzhu/autocat/autocat.tar.

APPENDIX

The MCMC-SA algorithm for the auto-multicategorical regression model consists of two basic steps. At the t-th iteration, θ^t is the current estimate of $\hat{\theta}$, $\mathbf{h}_t$ is the current estimate of $E_{\hat{\theta}}\{S(\mathbf{Y})\}$, and Γ_t is the current estimate of $E_{\hat{\theta}}\{S(\mathbf{Y})\}^{\otimes 2}$.

Step 1. At the t-th iteration, set $\mathbf{Y}_1(t) = \mathbf{Y}_N(t-1)$. Start with map $\mathbf{Y}_1(t)$ and update one randomly selected site to get the next map $\mathbf{Y}_2(t)$ as follows.
i) Randomly select a site $(k, l) \in D^o$ with probability $1/n$.
ii) Calculate $\Pr_{k,l}(j|\theta^{t-1})$ using (5) for $j = 0, 1, \ldots, C - 1$.
iii) Generate a random number u uniformly distributed on $(0, 1]$ and set $y_{k,l} = j$ if $\sum_{i=0}^{j} \Pr_{k,l}(i|\theta^{t-1}) \le u < \sum_{i=0}^{j+1} \Pr_{k,l}(i|\theta^{t-1})$. Continue this procedure to generate $N - 1$ maps $\{\mathbf{Y}_s(t), s = 2, \ldots, N\}$. Therefore, the possible difference between any two consecutive maps $\mathbf{Y}_s(t)$ and $\mathbf{Y}_{s+1}(t)$ is at most a value at a randomly selected site in i).

Step 2. Update θ^{t-1} to θ^t, $\mathbf{h}_{t-1}$ to $\mathbf{h}_t$, and Γ_{t-1} to Γ_t by

$$
\begin{cases}
\mathbf{h}_t = \mathbf{h}_{t-1} + \gamma_t\{\sum_{s=1}^{N} S(\mathbf{Y}_s(t))/N - \mathbf{h}_{t-1}\}, \\
\Gamma_t = \Gamma_{t-1} + \gamma_t\{\sum_{s=1}^{N} S(\mathbf{Y}_s(t))^{\otimes 2}/N - \Gamma_{t-1}\}, \\
\theta^t = \theta^{t-1} + \gamma_t(\Gamma_t - \mathbf{h}_t^{\otimes 2})^{-1}\{S(\mathbf{y}_{0,n}) \\
\qquad - \sum_{s=1}^{N} S(\mathbf{Y}_s(t))/N\}.
\end{cases}
$$

Note N is a pre-selected integer. In practice, we suggest to choose N to be about $5n \sim 20n$.

The main procedure of the MCMC stochastic approximation algorithm consists of two stages. At the first step of Stage I, we choose an initial point θ^0, an initial matrix Γ_0, an initial vector $\mathbf{h}_0$, and an initial spatial configuration $\mathbf{Y}^{(0)}$. Then, we set $t = 1$ and iterate Steps 1 and 2 with $t = 1, \ldots, K_1$. The gain constants in Stage 1 are defined by

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

where $K_1 \ge K_0$ is determined by

$$
K_1 = \inf\left\{ K \ge K_0, \ \left\| \sum_{t=K-K_0+1}^{K} \mathrm{Sign}(\theta^t - \theta^{t-1})/K_0 \right\| \le \eta_1 \right\}.
$$

In addition, K_0 is a prefix integer and function $\mathrm{Sign}(\theta)$ is a vector of 1, 0 or -1 according to whether the component of θ is positive, zero or negative, respectively. The real number

a_1 is usually set to be close to 0.5 and b_1 is set to be relatively large (e.g., $a_1 = 0.3$ and $b_1 = 5$) to obtain large gain constants to ensure that θ^t can move quickly into a neighborhood of the ML estimate. Also, we choose a relatively small value of η_1 and K_0 (e.g., $\eta_1 = 0.1$ and $K_0 = 100$).

Stage II starts with the final values of θ, $\mathbf{h}$, Γ and $\mathbf{Y}$ of Stage I as initial values and iterates the basic steps 1 and 2 with $t = 1, \ldots, K_2$ for the stochastic approximation algorithm as described above. The gain constants are defined by

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

where integer b_2 and real number $a_2 \in (1/2, 1)$ are preassigned. Usually, to stabilize the MCMC-SA algorithm, a_2 is set to be close to 1 and a small integer is chosen 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 $\tilde{\Gamma}_0 = \Gamma_0$,

$$\tilde{\theta}^t = \tilde{\theta}^{t-1} + (\theta^t - \tilde{\theta}^{t-1})/t,$$
$$\tilde{\mathbf{h}}_t = \tilde{\mathbf{h}}_{t-1} + (\mathbf{h}_t - \tilde{\mathbf{h}}_{t-1})/t, \quad \text{and}$$
$$\tilde{\Gamma}_t = \tilde{\Gamma}_{t-1} + (\Gamma_t - \tilde{\Gamma}_{t-1})/t.$$

The following criterion function is used to determine K_2, given by

$$\hat{\Delta}_t = \{S(\mathbf{y}_{0,n}) - \tilde{\mathbf{h}}_t\}^T (\tilde{\Gamma}_t - \tilde{\mathbf{h}}_t^{\otimes 2})^{-1} \{S(\mathbf{y}_{0,n}) - \tilde{\mathbf{h}}_t\}$$
$$+ \operatorname{tr}\left\{ (\tilde{\Gamma}_t - \tilde{\mathbf{h}}_t^{\otimes 2})^{-1} \hat{\Sigma}_t \right\} / t,$$

where $\hat{\Sigma}_t$ denotes the sample covariance of $\{\sum_{s=1}^N S \times (\mathbf{Y}_s(k))/N, k = 1, \ldots, t\}$ and K_2 is defined as $K_2 = \inf\{t : \hat{\Delta}_t \leq \eta_2\}$, in which η_2 (usually taken to be around 0.002) is a preassigned small number. Finally, under some conditions (Zhu, Gu, and Peterson, 2007; Andrieu, Moulines, and Priouret, 2005), the off-line average $(\tilde{\theta}^{K_2}, \tilde{\mathbf{h}}_{K_2}, \tilde{\Gamma}_{K_2})$ is used as our final estimate of $(\hat{\theta}, E_{\hat{\theta}}\{S(\mathbf{Y})\}, E_{\hat{\theta}}\{S(\mathbf{Y})\}^{\otimes 2})$, because

$$(\tilde{\theta}^{K_2}, \tilde{\mathbf{h}}_{K_2}, \tilde{\Gamma}_{K_2}) \to (\hat{\theta}, E_{\hat{\theta}}\{S(\mathbf{Y})\}, E_{\hat{\theta}}\{S(\mathbf{Y})\}^{\otimes 2})$$

almost surely, as $K_2 \to \infty$.

REFERENCES

[1] AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19** 716–723. MR0423716

[2] ANDRIEU, C., MOULINES, E., and PRIOURET, P. (2005). Stability of stochastic approximation under verifiable conditions. *SIAM Journal on Control and Optimization* **44** 283–312. MR2177157

[3] 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.

[4] BESAG, J. E. (1972). Nearest-neighbour systems and the autologistic model for binary data. *Journal of the Royal Statistics Society, Series B* **34** 75–83. MR0323276

[5] BESAG, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistics Society, Series B* **36** 192–236. MR0373208

[6] BESAG, J. E. (1977). Efficiency of pseudo likelihood estimators for simple Gaussian fields. *Biometrika* **64** 616–618. MR0494640

[7] BESAG, J. E. (1986). On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B* **48** 259–302. MR0876840

[8] CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley and Sons, New York. MR1239641

[9] CURRIES, D. J. (1991). Energy and large scale patterns of animal and plant species richness. *American Naturalist* **137** 27–49.

[10] GASTON, K. J. and BLACKBURN, T. M. (2000). *Patterns and Process in Macroecology*. Blackwell Science, Oxford.

[11] 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. MR1812255

[12] 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. MR1841419

[13] 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. MR1812254

[14] HE, F., ZHOU, J., and ZHU H. T. (2003). Autologistic regression model for the distribution of vegetation. *Journal of Agricultural, Biological, and Environmental Statistics* **8** 205–222.

[15] HEIKKINEN, J. and HÖGMANDER, H. (1994). Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics* **43** 569–582.

[16] 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. MR1817027

[17] HOUGHTON, J. T., JENKINS, G. J., and EPHRAUMS, J. J. (1990). *Climate Change, the IPCC Scientific Assessment*. Cambridge University Press, Cambridge.

[18] 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.

[19] IRIZARRY, R. (2001). Information and posterior probability criteria for model selection in local likelihood estimation. *Journal of the American Statistical Association* **96** 303–315. MR1952740

[20] KAISER, M. S., and CRESSIE, N. (2000). The construction of multivariate distributions from Markov random fields. *Journal of Multivariate Analysis* **73** 199–220. MR1763324

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

[22] LENIHAN, J. M. and NEILSON, R. P. (1993). A rule-based vegetation formation model for Canada. *Journal of Biogeography* **20** 615–628.

[23] MACKENZIE, D. I., NICHOLS, J. D., LACHMAN, G. B., DROEGE, S., ROYLE, J. A., and LANGTIMM, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83** 2248–2255.

[24] McCULLAGH, P., and NELDER, J. A. (1989). *Generalized Linear Models* (2nd edition). Chapman and Hall, London. MR0727836

[25] MYERS, N., MITTERMEIER, R. A., MITTERMEIER, C. G., DA FONSECA, G. A. B., and KENT, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* **403** 853–858.

[26] PETTITT, A. N., FRIEL, N., and REEVES, R. (2003). Efficient calculation of the normalization constant of the auto-logistic model on the lattice. *Journal of Royal Statistical Society, Series B* **65** 235–247.

[27] POTTS, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophic Society* **48** 106–109. MR0047571

[28] SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6** 461–464. MR0468014

[29] SIN, C. Y., and WHITE, H. (1996). Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics* **71** 207–225. MR1381082

[30] WALLER, L. A., and GOTWAY, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. Wiley & Sons, New Jersey. MR2075123

[31] 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.

[32] ZHANG, H. P. (1993). Image restoration: flexible neighborhood systems and iterated conditional expectations. *Statistica Sinica* **3** 117–139.

[33] ZHU, H. T., GU, M. G., and PETERSON, B. G. (2007). Maximum likelihood from spatial random effects models via the stochastic approximation expectation maximization algorithm. *Statistics and Computing* **17** 163–177.

Hongtu Zhu
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7420, USA
E-mail address: hzhu@bios.unc.edu

Fangliang He
Department of Renewable Resources
University of Alberta
Edmonton, AB, T6G 2H1, Canada
E-mail address: fhe@ualberta.ca

Julie Zhou
Department of Mathematics and Statistics
University of Victoria
Victoria, BC, V8W 3P4, Canada
E-mail address: jzhou@math.uvic.ca