Received: 10 January 2012, Revised: 1 November 2012, Accepted: 2 November 2012, Published online in Wiley Online Library: 7 January 2013

(wileyonlinelibrary.com) DOI: 10.1002/env.2189

Spatial multinomial regression models for nominal categorical data: a study of land cover in Northern Wisconsin, USA

Chongyang Jin[a], Jun Zhu[b]*, Michelle M. Steen-Adams[c], Stephan R. Sain[d] and Ronald E. Gangnon[e]

We develop statistical tools for regression analysis of nominal categorical data on a spatial lattice that are becoming increasingly abundant because of the advances of geographic information systems in environmental science. In a generalized linear mixed model framework, we model the response variable by a multinomial distribution. There are two additive components in the linear predictor: a linear regression on covariates and a spatial random effect such that the spatial dependence in the random effect is induced by a multivariate conditional autoregressive model. Bayesian hierarchical modeling is used for statistical inference, and Markov chain Monte Carlo algorithms are devised to obtain posterior samples. The methodology is applied to analyze a northern Wisconsin land cover data set in a study that assesses the relationship between forest landscape structure and past social conditions, expanding the analytical tools available in landscape ecology and environmental history. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: Bayesian hierarchical model; environmental history; geographic information systems; landscape ecology; Multivariate CAR model; spatial statistics

1. INTRODUCTION

Spatial categorical data on a lattice are becoming increasingly abundant because of the advances of computer technologies in environmental and ecological studies. In a study of northern Wisconsin, an interdisciplinary team of ecologists and environmental historians assess the influence of past social conditions on forest landscape structure and use the insight gained from historical data to inform future forest landscape management. Using a novel spatial multinomial regression for nominal categorical data, we analyze land cover data from this particular study to integrate historic conditions into landscape ecological analysis. In particular, we build on landscape ecology research that demonstrates relationships between land ownership characteristics and forest landscape composition (Theobald *et al.*, 1996; Crow *et al.*, 1999; Brown, 2003). The response variable is land cover defined in terms of dominance by a land cover class within each quarter section in the study area. The covariates consist of several land ownership metrics: ownership category, ownership size (or, property area), and parcel size. See Figures 1 and 2 and Section 5 for more details.

In recent years, ecologists have called for integration of historical data derived from land surveys and other sources (Whitney, 1994) to illuminate landscape processes, thereby guiding management (Turner, 2005); numerous studies demonstrate the utility of such an approach (e.g, Foster and Aber, 2004; Rhemtulla *et al.*, 2007; Foster *et al.*, 2008; Steen-Adams *et al.*, 2011). Environmental history data can meaningfully improve models of landscape change and past structure in several ways. One, land-use history records have provided information about effects of soil alterations (e.g., presence vs. absence of past cultivation) on nutrient cycling and ecological community composition (Dupouey *et al.*, 2002; Fraterrigo *et al.*, 2005). This body of work demonstrates persisting legacy effects of land-use history on biogeochemical processes and landscape structure (Turner, 2005). Two, investigators have tapped land ownership history records to illuminate landscape structure (Foster *et al.*, 1992; Medley *et al.*, 2003). Similarly, research on current ownership condition documents a positive relationship between parcel size and forest patch size (Stanfield *et al.*, 2002). Three, historical forest survey records have enabled ecologists to characterize the temporal and spatial scale of past disturbance regimes (e.g., Schulte and Mladenoff, 2005). Four, historic vegetation data have

* *Correspondence to: Jun Zhu, Department of Statistics and Department of Entomology, University of Wisconsin–Madison, Madison, WI, U.S.A. E-mail: jzhu@stat.wisc.edu*

a *Department of Statistics, University of Wisconsin–Madison, Madison, WI, U.S.A.*

b *Department of Statistics and Department of Entomology, University of Wisconsin–Madison, Madison, WI, U.S.A.*

c *Department of Environmental Studies, University of New England, Biddeford, ME, U.S.A.*

d *Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO, U.S.A.*

e *Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, Madison, WI, U.S.A.*

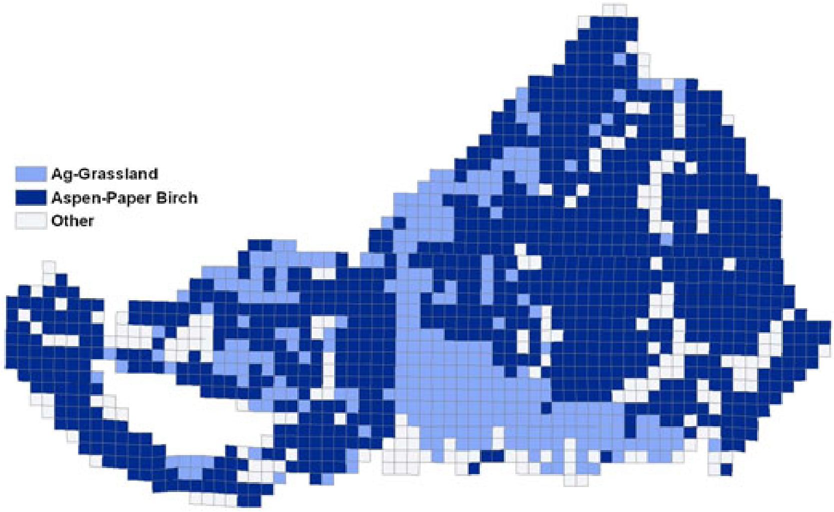

Figure 1. Map of the response variable land cover type

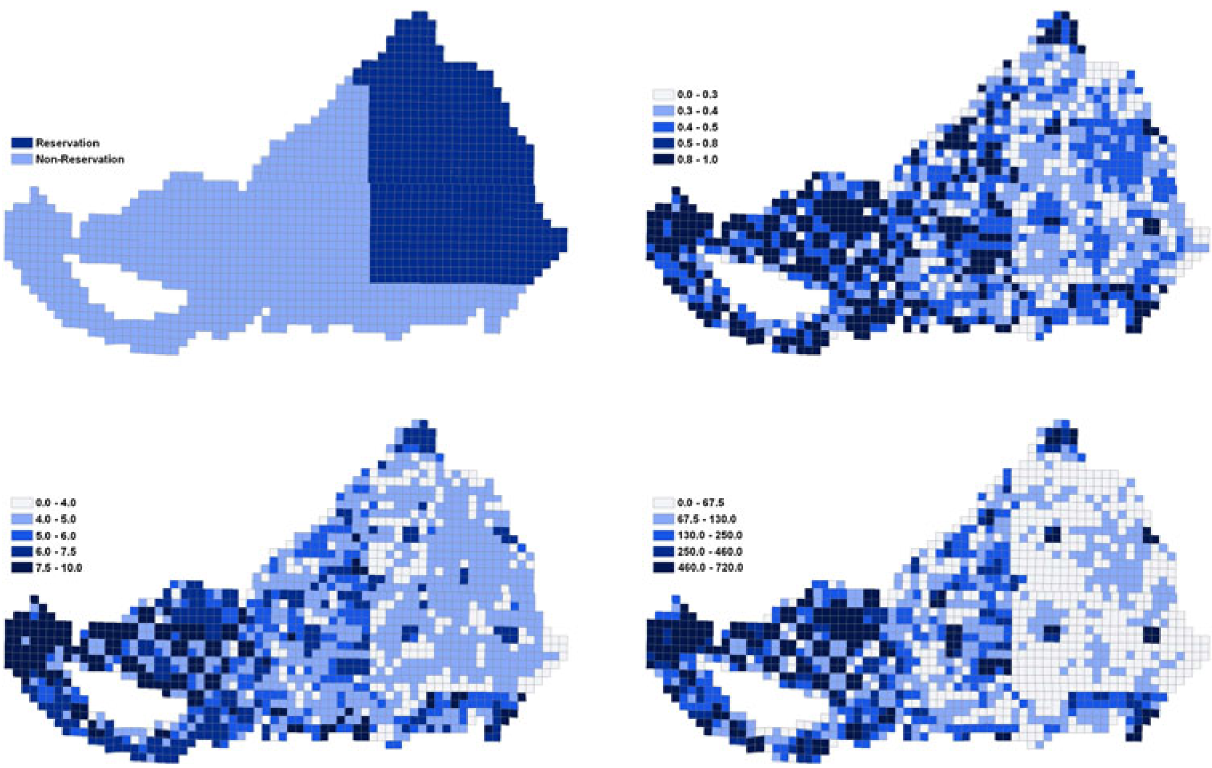

Figure 2. Map of covariates: `Reserv` (top left), `Fragment` (top right), `log TotOwn` (bottom left), and `AvParcel` (bottom right)

characterized baseline ecological conditions and thereby enabled ecologists to estimate the historic ranges of variability of desired parameters (e.g., patch size, ecological community composition, species' geographic ranges, and climatic conditions); such estimates can inform restoration and management (Swetnam *et al.*, 1999). Furthermore, advances in geographic information system technology enable investigators to tap the information contained in historic records (Knowles, 2002; Heasley, 2003). In sum, historic records can reveal important, otherwise unavailable information to improve models of long-term landscape ecological structure and process. Yet, because these records are often best capitalized upon when processed into categorical datasets, ecologists need advanced statistical techniques to model nominal categorical data.

Regression is suitable for addressing the ecological questions in our study, as it provides a flexible, yet principled, approach to linking human dimensions to ecological patterns. Practical statistical methods for analyzing spatially referenced data with multiple categories are limited, despite the increasing abundance of such data. One possible approach is auto-multinomial models where spatial dependence is accounted for by autoregression (Cressie, 1993; Li, 2006). However, the existing models tend to be restricted to either a constant mean or only one covariate at a time. Introduction of autoregression also results in an unknown normalizing constant in the likelihood function, which makes computation for statistical inference challenging (Besag, 1974; Cressie, 1993; Zhu *et al.*, 2005; Zheng and Zhu, 2008). An alternative approach is to formulate the problem in the framework of spatial generalized linear mixed models (GLMM), where the response variable is categorical and is linked to a regression on covariates and a spatial random effect as two additive components in a link function. Diggle *et al.* (1998) proposed such a modeling framework for non-Gaussian response variables, but did not consider multinomial models explicitly. For spatial *ordinal* data, Higgs and Hoeting (2010) extended the traditional probit models to a general class of models that impose a spatial process on the continuous latent variable. Statistical inference was carried out using Bayesian hierarchical modeling and Markov chain Monte Carlo (MCMC) algorithms (see also De Oliveira, 2000). Both Diggle *et al.* (1998) and Higgs and Hoeting (2010) considered geostatistical data as opposed to lattice data. In contrast, we believe that methodologies for the analysis of spatial *nominal* data on a lattice are not well developed. An exception is Paciorek and McLachlan (2009) who did model nominal data on a spatial lattice but did not focus on regression.

Here, we develop a set of statistical tools for the regression analysis of land cover data on a lattice by bringing together several ideas in recent development of spatial statistics. In Sections 2 and 3, we develop a spatial multinomial model in the spatial GLMM framework, such that the response variable follows a multinomial model and the latent spatial random effect follows a lattice model. As we will show, there are several challenging issues to address in both modeling and computation. First, the spatial random effect here is multivariate at each spatial site rather than univariate as in most other spatial GLMMs. Our strategy is to utilize a multivariate lattice model (Mardia, 1988; Sain and Cressie, 2007), which can be viewed as an extension from the popular, univariate conditional autoregressive (CAR) model. Like the CAR model, precision rather than variance matrices are specified in these multivariate lattice models, providing a distinct computational advantage. However, because these lattice models for the latent spatial random effects are embedded in a multinomial model, computation can be still demanding. Thus, we further utilize dimension-reduction techniques to help alleviate the computational burden. In particular, we consider spatial linear latent variable models, such that the latent variables at each spatial location are reduced to a smaller number of latent factors (Wang and Wall, 2003; Zhu *et al.*, 2005).

In Section 4, we reparameterize the models in Sections 2 and 3 and adopt Bayesian hierarchical modeling for statistical inference. For computation, we devise MCMC algorithms to obtain the posterior samples. In Section 5, we return to analyze the northern Wisconsin land cover data using our new methodology. In Section 6, we present the result of the data analysis followed by a discussion of the ecological and management implications. Technical details regarding the link function, a comparison with a multivariate intrinsic CAR model, and the MCMC algorithms are given in the Appendices as web-based supplementary materials.[†]

2. RESPONSE VARIABLE MODEL

2.1. Multinomial model

At site $i = 1, \ldots, I$ on a spatial lattice of size $I \in \mathbb{N}$, we let n_i denote the total number of trials and $y_{ji} \in \{0, \ldots, n_i\}$ denote the number of trials that result in the jth category, where $j = 0, \ldots, J$ and $J \in \mathbb{N}$. Let π_{ji} denote the probability of the jth outcome at site i. Thus, $\sum_{j=0}^{J} y_{ji} = n_i$ and $\sum_{j=0}^{J} \pi_{ji} = 1$. Let $\boldsymbol{y}_i = (y_{0i}, \ldots, y_{Ji})'$ denote a $(J+1)$-dimensional multinomial response variable with n_i trials and $J+1$ possible categories with probabilities $\boldsymbol{\pi}_i = (\pi_{0i}, \ldots, \pi_{Ji})'$. The corresponding probability density function for the ith site is

$$p(\boldsymbol{y}_i | n_i, \boldsymbol{\pi}_i) = \begin{pmatrix} n_i \\ \boldsymbol{y}_i \end{pmatrix} \pi_{0i}^{y_{0i}} \cdots \pi_{Ji}^{y_{Ji}} \tag{1}$$

(McCullagh and Nelder, 1989).

2.2. Link function

Following the convention of multinomial models, we consider a log-ratio link of the probability π_{ji} for the jth category relative to π_{0i} for the baseline category (Aitchison, 1987; Sain *et al.*, 2006),

$$Z_{ji} = \log(\pi_{ji}/\pi_{0i}), \quad j = 1, \ldots, J \tag{2}$$

Thus, with $Z_{0i} \equiv 0$,

$$\pi_{ji} = \exp(Z_{ji}) \left\{ \sum_{j=0}^{J} \exp(Z_{ji}) \right\}^{-1} \tag{3}$$

where $i = 1, \ldots, I$ and $j = 0, \ldots, J$. Let $\boldsymbol{Z}_i = (Z_{1i}, \ldots, Z_{Ji})'$ denote a J-dimensional vector of latent variables at the ith site. The dimension of $\boldsymbol{Z}_i$ is for J categories all in reference to the baseline category. We will describe models for $\{\boldsymbol{Z}_i : i = 1, \ldots, I\}$ in Section 3.

A practical issue to consider is the choice of the baseline category in a multinomial model for nominal categorical data, which we will discuss in detail in Section 6.3 and Appendix A.

[†]Supporting information may be found in the online version of this article.

3. LATENT VARIABLE MODEL

3.1. Full-dimensional model

Following Mardia (1988) and Sain and Cressie (2007), we model the latent variables $\{\boldsymbol{Z}_i : i = 1, \ldots, I\}$ by a multivariate Gaussian process such that the spatial dependence is induced by a multivariate CAR (MCAR) model. Specifically, the Gaussian MCAR model is specified via the conditional probability distributions

$$\mathrm{E}\left(\boldsymbol{Z}_i | \boldsymbol{Z}_{i'} : i' \in \mathcal{N}_i\right) = \boldsymbol{\mu}_i + \sum_{i' \in \mathcal{N}_i} \boldsymbol{\Lambda}_{ii'}(\boldsymbol{Z}_{i'} - \boldsymbol{\mu}_{i'}), \quad \mathrm{Var}\left(\boldsymbol{Z}_i | \boldsymbol{Z}_{i'} : i' \in \mathcal{N}_i\right) = \boldsymbol{\Gamma}_i \tag{4}$$

We let the mean μ_{ji} follow a linear regression $\mu_{ji} = \boldsymbol{x}_i' \boldsymbol{\beta}_j$, where $\boldsymbol{x}_i$ is a K-dimensional vector of covariates at the ith site ($i = 1, \ldots, I$) and $\boldsymbol{\beta}_j$ is a K-dimensional vector of regression coefficients for the jth category ($j = 1, \ldots, J$). Thus, $\boldsymbol{\mu}_i = \left(\boldsymbol{x}_i' \boldsymbol{\beta}\right)'$, where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J]$ is a $K \times J$ matrix of all the regression coefficients.

Furthermore, $\mathcal{N}_i = \{i' : i'$ is a neighbor of $i\}$ denotes the site indices of the neighbors of the ith site according to a neighborhood structure. The $J \times J$ matrix $\boldsymbol{\Lambda}_{ii'}$ consists of autoregressive coefficients between the ith site and its neighbor at the i'th site both within the same category and among different categories and thus, induces spatial dependence across sites and categories. The $J \times J$ conditional variance matrix $\boldsymbol{\Gamma}_i$ allows dependence among categories at a given site.

Let $\boldsymbol{Z} = \left(\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_I'\right)'$ denote an (IJ)-dimensional vector of all the latent variables and let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I]'$ denote an $I \times K$ design matrix. Then, by (4), the joint probability distribution of $\boldsymbol{Z}$ is multivariate Gaussian

$$\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{5}$$

where the mean vector and the variance matrix are

$$\boldsymbol{\mu} = \left(\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_I'\right)' = \mathrm{vec}\{(\boldsymbol{X}\boldsymbol{\beta})'\}, \quad \boldsymbol{\Sigma} = \left\{\mathrm{blk}(-\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}_{ii'})\right\}^{-1}$$

$\mathrm{blk}(\boldsymbol{A}_{ii'})$ denotes a matrix with blocks $\boldsymbol{A}_{ii'}$ row indexed by i and column indexed by i' for $i, i' = 1, \ldots, I$. The following assumptions ensure that the variance matrix $\boldsymbol{\Sigma}$ is symmetric and positive definitive (Mardia, 1988; Sain and Cressie, 2007):

(A.1) $\boldsymbol{\Lambda}_{ii'}\boldsymbol{\Gamma}_{i'} = \boldsymbol{\Gamma}_i \boldsymbol{\Lambda}_{i'i}'$, where $\boldsymbol{\Lambda}_{ii} = -\boldsymbol{I}_J$ and $\boldsymbol{\Lambda}_{ii'} = \boldsymbol{0}$ for $i' \notin \mathcal{N}_i \cup \{i\}$
(A.2) $\mathrm{blk}\left(-\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}_{ii'}\right)$ or $\mathrm{blk}(-\boldsymbol{\Lambda}_{ii'})$ is positive definite.

We will refer to the model specified by (1) and (2) and (4) and (5) as a spatial multinomial model with a full-dimensional latent MCAR model or, simply, the *full spatial multinomial model*.

3.2. Reduced-dimensional model

For more than two categories $J > 1$, we further develop a lower-dimensional model for the latent variables $\{Z_{ji}\}$ as

$$\boldsymbol{Z}_i = \boldsymbol{\mu}_i + \boldsymbol{\Omega}\boldsymbol{Z}_i^* \tag{6}$$

where $\boldsymbol{\mu}_i = \left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)'$, $\boldsymbol{\Omega}$ is a $J \times J^*$ matrix, and $\boldsymbol{Z}_i^*$ is a J^*-dimensional vector of zero-mean latent variables with $1 \leqslant J^* \leqslant J - 1$. There are at least two reasons to consider a lower-dimensional model. One is to reduce the computational cost, as the dimension is reduced from J to $J^* < J$. When the primary focus is on regression and modeling the spatial dependence is of secondary interest, this may be a relatively small price to pay for a gain in computational efficiency. Two, the new latent variable $\boldsymbol{Z}_i^*$ can be interpreted as shared spatial effects due to unobserved environmental factors. When $J^* = 1$, the shared spatial effect is also known as a common spatial factor (Wang and Wall, 2003).

We then impose an MCAR model on $\boldsymbol{Z}_i^*$ via conditional probability distributions

$$\mathrm{E}\left(\boldsymbol{Z}_i^* | \boldsymbol{Z}_{i'}^* : i' \in \mathcal{N}_i\right) = \sum_{i' \in \mathcal{N}_i} \boldsymbol{\Lambda}_{ii'}^* \boldsymbol{Z}_{i'}^*, \quad \mathrm{Var}\left(\boldsymbol{Z}_i^* | \boldsymbol{Z}_{i'}^* : i' \in \mathcal{N}_i\right) = \boldsymbol{\Gamma}_i^* \tag{7}$$

where $\boldsymbol{\Lambda}_{ii'}^*$ and $\boldsymbol{\Gamma}_i^*$ are $J^* \times J^*$ matrices satisfying conditions similar to (A.1) and (A.2) for $\boldsymbol{\Lambda}_{ii'}$ and $\boldsymbol{\Gamma}_i$. We will refer to the model specified by (1) and (2) and (6)–(7) as a spatial multinomial model with a reduced-dimensional latent MCAR model or, simply, the *reduced spatial multinomial model*. In the case $J^* = 1$, we will refer to the model as a *common-factor spatial multinomial model*.

Other alternative models for the latent variables are possible such as multivariate intrinsic CAR models (MICAR) (Jin *et al.*, 2005). We compare MCAR models with the MICAR models and demonstrate that there is a close relationship between the two types of models in Appendix B.

4. STATISTICAL INFERENCE

In this section, we develop statistical inference for both the full and reduced spatial multinomial models.

4.1. Full spatial multinomial model

For a full spatial multinomial model, we let $\mathbf{\Lambda}_{ii'} = \mathbf{\Lambda}$ for $i' \in \mathcal{N}_i$ and $\mathbf{\Gamma}_i = \mathbf{\Gamma}$, where $\mathbf{\Lambda}$ is a symmetric matrix and $\mathbf{\Gamma}$ is a variance matrix, both of dimension $J \times J$. Thus, $\mathbf{\Lambda}_{i'i} = \mathbf{\Lambda}_{ii'}$ and from (A.1), $\mathbf{\Lambda}\mathbf{\Gamma} = \mathbf{\Gamma}\mathbf{\Lambda}$. Further, the precision matrix $\mathbf{\Sigma}^{-1}$ has $\mathbf{\Gamma}^{-1}$ as the ith diagonal block, $-\mathbf{\Gamma}^{-1}\mathbf{\Lambda}$ as the the upper-triangular and lower-triangular blocks for the (i, i')th block such that $i' \in \mathcal{N}_i$, and $\mathbf{0}$ otherwise. It can be shown that $\mathbf{\Sigma} = \text{diag}\{\mathbf{\Gamma}\}(\boldsymbol{I} - \boldsymbol{C} \otimes \mathbf{\Lambda})^{-1}$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{C} = \left[\mathcal{I}(i \sim i')\right]_{i,i'=1}^{I}$ indicates whether site i' is a neighbor of site i according to a neighbor structure. Here $\mathcal{I}(\cdot)$ is an indicator function and $i \sim i'$ denotes that $i' \in \mathcal{N}_i$. Also, let $\boldsymbol{V} = \mathbf{\Gamma}^{-1}$. Thus, the model parameters are $\{\boldsymbol{\beta}_j\}_{j=1}^{J}$, $\boldsymbol{V}$, and $\mathbf{\Lambda}$.

We adopt Bayesian hierarchical modeling for statistical inference and devise MCMC algorithms to obtain the posterior. Independent prior distributions for the model parameters are assumed, with probability density functions $p(\boldsymbol{\beta}_j)$, $p(\boldsymbol{V})$, and $p(\mathbf{\Lambda})$. In particular, we follow Sain and Cressie (2007) and let

$$\boldsymbol{\beta}_j \sim \text{N}\left(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}\right), \quad \boldsymbol{V} \sim \text{W}(\rho, \psi \boldsymbol{I}), \rho > J$$
$$p(\mathbf{\Lambda}) \propto \exp\left\{-\text{vec}(\mathbf{\Lambda})'\text{vec}(\mathbf{\Lambda})/\xi^2\right\} \tag{8}$$

where W refers to a Wishart distribution, σ_β^2, ρ, $\psi \boldsymbol{I}$, and ξ^2 are the hyperparameters corresponding to variance, degrees of freedom, scale matrix, and scale parameter, respectively. The resulting posterior distribution of $\boldsymbol{Z}$, $\{\boldsymbol{\beta}_j\}$, $\boldsymbol{V}$, and $\mathbf{\Lambda}$ has, up to a proportionality constant, the following probability density function:

$$p(\boldsymbol{Z}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}, \mathbf{\Lambda}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{Z})p(\boldsymbol{Z}|\{\boldsymbol{\beta}_j\}, \boldsymbol{V}, \mathbf{\Lambda}) \prod_{j=1}^{J} p(\boldsymbol{\beta}_j)p(\boldsymbol{V})p(\mathbf{\Lambda}) \tag{9}$$

Even though the prior specification (8) is similar to Sain and Cressie (2007), the posterior distribution (9) is different because the data distribution $p(\boldsymbol{y}|\boldsymbol{Z})$ is multinomial, not Poisson as in Sain and Cressie (2007). Thus, new MCMC algorithms need to be devised.

To sample from (9), we use a Gibbs sampler and sample from the full conditional distributions of $\boldsymbol{Z}$, $\{\boldsymbol{\beta}_j\}$, $\boldsymbol{V}$, and $\mathbf{\Lambda}$. It can be shown that the full conditional distribution of $\boldsymbol{\beta}_j$ is Gaussian with mean and variance

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_j} = \mathbf{\Sigma}_{\boldsymbol{\beta}_j}\left\{\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{Z}_{j\cdot} + \boldsymbol{X}'\sum_{k \neq j}\boldsymbol{S}_{jk}(\boldsymbol{Z}_{k\cdot} - \boldsymbol{X}\boldsymbol{\beta}_k)\right\}, \quad \mathbf{\Sigma}_{\boldsymbol{\beta}_j} = \left(\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{X} + \sigma_\beta^{-2}\boldsymbol{I}\right)^{-1}$$

where $\{\boldsymbol{S}_{jk}\}$ is defined in Appendix C.1. Further, the full conditional distribution of $\boldsymbol{V}$ is $\text{W}\left(I + \rho, \bar{\mathbf{\Psi}}\right)$, where $\bar{\mathbf{\Psi}}$ is defined in Appendix C.2. Thus, we may directly sample $\{\boldsymbol{\beta}_j\}$ and $\boldsymbol{V}$ from their full conditional distributions. For the remainder model parameters, the full conditional distributions are not in closed form, and thus, we use Metropolis–Hastings (MH) algorithms to simulate from the full conditional distributions. In particular, the proposal distributions for $\mathbf{\Lambda}$ and $\boldsymbol{Z}$ are set to be uniform and Gaussian, respectively. The parameters of these proposal distributions are tuned to improve convergence and mixing of the Markov chains. For details of the MCMC algorithms, see Appendix C.

4.2. Reduced spatial multinomial model

For a reduced spatial multinomial model, we reparameterize as in the previous subsection but replace $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ of dimension $J \times J$ with $\mathbf{\Lambda}^*$ and $\mathbf{\Gamma}^*$ of a lower dimension $J^* \times J^*$. Analogously define $\boldsymbol{V}^* = \mathbf{\Gamma}^{*-1}$. The model parameters are $\{\boldsymbol{\beta}_j\}_{j=1}^{J}$, $\boldsymbol{V}^*$, $\mathbf{\Lambda}^*$, and $\mathbf{\Omega}$.

Again, independent prior distributions for the model parameters are assumed, with probability density functions $p(\mathbf{\Omega})$, $p(\boldsymbol{\beta}_j)$, $p(\boldsymbol{V}^*)$, and $p(\mathbf{\Lambda}^*)$. In particular, let

$$\text{vec}(\mathbf{\Omega}) \sim \text{N}\left(\mathbf{0}, \sigma_\Omega^2 \boldsymbol{I}\right), \quad \boldsymbol{\beta}_j \sim \text{N}\left(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}\right)$$
$$\boldsymbol{V}^* \sim \text{W}(\rho^*, \psi^* \boldsymbol{I}), \rho^* > J^*, \quad p(\mathbf{\Lambda}^*) \propto \exp\left\{-\text{vec}(\mathbf{\Lambda}^*)'\text{vec}(\mathbf{\Lambda}^*)/\xi^{*2}\right\} \tag{10}$$

where W refers to a Wishart distribution, σ_Ω^2, σ_β^2, ρ^*, $\psi^* \boldsymbol{I}$, and ξ^{*2} are the hyperparameters corresponding to variances, degrees of freedom, scale matrix, and scale parameter, respectively. The resulting posterior distribution of $\boldsymbol{Z}^*$, $\mathbf{\Omega}$, $\{\boldsymbol{\beta}_j\}$, $\boldsymbol{V}^*$, and $\mathbf{\Lambda}^*$ has the following probability density function, up to a proportionality constant:

$$p\left(\boldsymbol{Z}^*, \mathbf{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*, \mathbf{\Lambda}^*|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{Z}^*, \mathbf{\Omega}, \{\boldsymbol{\beta}_j\}\right) p(\boldsymbol{Z}^*|\boldsymbol{V}^*, \mathbf{\Lambda}^*)p(\mathbf{\Omega}) \prod_{j=1}^{J} p(\boldsymbol{\beta}_j)p(\boldsymbol{V}^*)p(\mathbf{\Lambda}^*) \tag{11}$$

Again, we use a Gibbs sampler to sample $\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*$, and $\boldsymbol{\Lambda}^*$ from their respective full conditional distributions. Because only $\boldsymbol{V}^*$ has conjugacy, we choose to use a MH algorithm to sample the other parameters from their corresponding full conditional distributions. In the case $J^* = 1$, the matrix $\boldsymbol{V}^*$ is reduced to a scalar v^*. The prior of v^* is set to be a gamma distribution to achieve conjugacy for v^*. Thus, we may directly sample v^* from its full conditional distribution. Details of the MCMC algorithms are given in Appendix D.

5. NORTHERN WISCONSIN LAND COVER DATA

5.1. Data description

A primary research objective was to assess the relationship between land ownership history characteristics and forest landscape structure in northern Wisconsin and inform future forest landscape management. Data are derived from 1915 plat maps, which indicate ownership conditions, as well as the Wisconsin Land Economic Inventory, which indicates land cover in 1930, at an important turning point in the region's history (for details, see Koch, 2006; Steen-Adams *et al.*, 2011). Geographic information system software packages, ArcView 3.3 and ArcMap 9.2 (Environmental Systems Resource Institute, 2002; 2006), are employed to process, map, and display the data. To control for the influence of land form and soil type, analysis is restricted to parcels that lie within a single Land Type Association, here, the Ashland Lake-Modified Till Plain (LTA 212-Ya03), which is hierarchically nested within the Province 212, Laurentian Mixed Forest, as defined by the USDA Forest Service National Hierarchical Framework of Ecological Units (Cleland *et al.*, 1997). Data are on a square lattice and employ spatial characteristics that conform to the rectilinear township and range grid of the U.S. Public Land Survey System (White, 1983; Hubbard, 2009). The unit of analysis of this study is the quarter section (1/36 township, 160 acres $\approx$65 ha), the minimum acceptable spatial scale for the use of U.S. Public Land Survey records as ecological data (Delcourt and Delcourt, 1996; Schulte and Mladenoff, 2001).

The response variable is the dominant land cover in 1930 of each quarter section. Dominant land cover is defined as the land cover that accounts for the largest proportion of all component classes in the quarter-section unit of analysis. We selected 1930 land cover data because of its significance to the region's environmental history and landscape ecology: the data represent landscape conditions after original forest clearance and at the peak of agricultural land use (White and Mladenoff, 1994). There are three land cover types in the response variable, which correspond with the historical period: aspen-paper birch forest (APB), which can develop after forest clearance, agriculture-grassland (AG), and all others in one category (OTH). See Figures 1 and 2. Thus, $J = 2$. In addition, $n_i \equiv 1$ because each quarter section can take on only one cover type. That is, $y_{ji} = 0$ or 1, such that $y_{ji} = 1$ if the ith site is in the jth category, where $j = 0, \ldots, J = 2$ and $i = 1, \ldots, I = 1429$. The probability density (1) simplifies to $p(\boldsymbol{y}_i|\boldsymbol{\pi}_i) = \prod_{j=0}^{J} \pi_{ji}^{y_{ji}}$.

The covariates are various land ownership characteristics, as indicated by the 1915 plat maps, and are designed to assess important aspects of ownership. Four covariates are of interest: `Reserv` is a binary variable indicating whether a quarter section is on an Indian reservation or not, designed to detect the influence of contrasting social-economic histories; `Fragment` is the proportion of the average size of parcel units that lie within the quarter section; `TotOwn` is the total property area (measured in acres) associated with the owner of the largest parcel unit of the quarter section and designed to assess whether land cover varies with ownership size; and `AvParcel` is the average size (measured in acres) of all parcels associated with a quarter section. See Figure 2. The covariates `Fragment` and `AvParcel` measure different aspects of

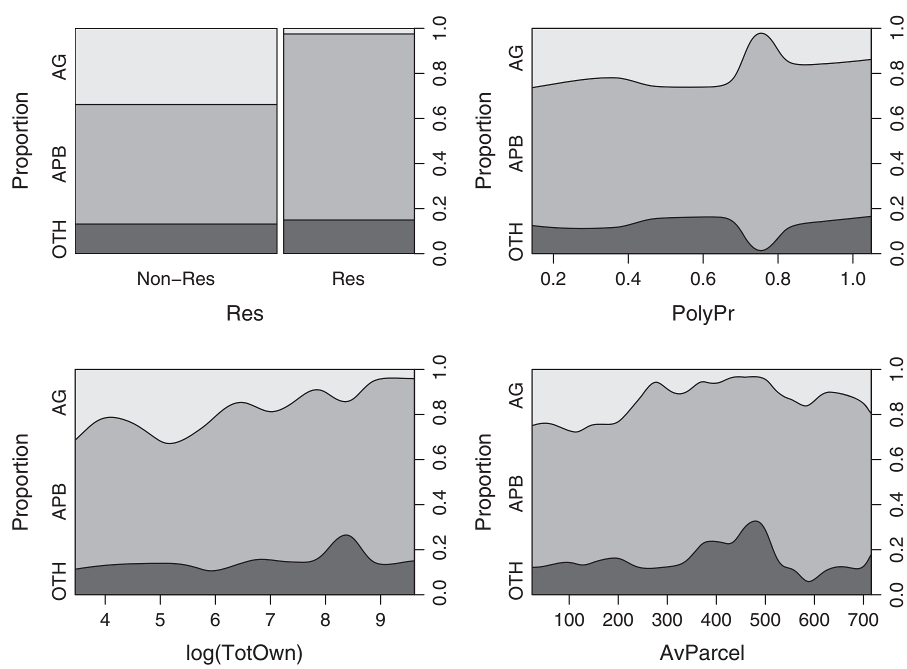

Figure 3. Sample proportions of land cover type versus covariates: `Reserv` (top left), `Fragment` (top right), `log TotOwn` (bottom left), and `AvParcel` (bottom right)

ownership: calculation of `Fragment` is limited to parcel units within the quarter-section unit of analysis and is a measure of fragmentation; calculation of `AvParcel` is based on entire parcel units, which may extend beyond the quarter-section boundary, and is a measure of parcel size. Moreover, the covariate `TotOwn` is highly skewed and thus, transformed to the log scale. Finally, the continuous covariates are standardized to have mean 0 and standard deviation 1. Figure 3 plots the proportions of the three land cover types against the four covariates.

5.2. Model fitting

We apply the methodology developed in Sections 2–4 to analyze this data. Both full and reduced spatial multinomial models are considered. Moreover, in the MCAR models for the latent variables, we consider a first-order neighborhood that consists of the four nearest neighbors on the grid in the north, south, west, and east.

When implementing the MCMC algorithms, we select model parameter values for the prior distributions to be diffuse. We also tune the model parameters in the proposal distributions to facilitate the convergence and mixing of the MCMC. Convergence of the MCMC algorithms is judged by examining the trace plots and Geweke's convergence diagnostics (Geweke, 1992). In total, there are 50,000 iterations in each MCMC chain, and the first 10,000 iterations are set aside as the burn-in period. The medians and the 95% credible intervals are constructed for the inference of model parameters.

We use deviance information criterion (DIC) for model selection (Spiegelhalter *et al.*, 2002). A smaller DIC value indicates a better model in the sense of model fitting and model parsimony.

5.3. Alternative models

We consider the case $\boldsymbol{\Lambda}_{ii'} = \mathbf{0}$ in models (1) and (2) and models (4) and (5), where the latent variables are independent across sites. We then let $\boldsymbol{Z}_i \sim \mathrm{N}_J(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{\boldsymbol{Z}})$, where $\boldsymbol{\mu}_i = (\boldsymbol{x}_i'\boldsymbol{\beta})'$, $i = 1, \ldots, I$ as before and $\boldsymbol{\Sigma}_{\boldsymbol{Z}} = \mathrm{diag}\left\{\sigma_1^2, \ldots, \sigma_J^2\right\}$. We will refer to the resulting model as an independent multinomial model. Under the assumption of independent prior distributions $\boldsymbol{\beta}_j \sim \mathrm{N}\left(0, \sigma_\beta^2 \boldsymbol{I}\right)$ and $\sigma_j^2 \sim \mathrm{IG}(\tilde{\alpha}_2, \tilde{\gamma}_2)$ where IG denotes an inverse gamma distribution, the posterior distribution of $\boldsymbol{Z}$, $\{\boldsymbol{\beta}_j\}$, and $\left\{\sigma_j^2\right\}$ is, up to a proportionality constant,

$$p\left(\boldsymbol{Z}, \{\boldsymbol{\beta}_j\}, \left\{\sigma_j^2\right\} | \boldsymbol{y}\right) \propto p(\boldsymbol{y}|\boldsymbol{Z})\, p\left(\boldsymbol{Z} | \{\boldsymbol{\beta}_j\}, \left\{\sigma_j^2\right\}\right) \left\{\prod_{j=1}^{J} p(\boldsymbol{\beta}_j)\, p\left(\sigma_j^2\right)\right\} \tag{12}$$

We use a Gibbs sampler and sample from the full conditional distributions. We directly sample from the full conditional distributions of $\boldsymbol{\beta}_j$ and σ_j^2 because they have conjugate priors. For $\boldsymbol{Z}$, we use an MH algorithm with a Gaussian proposal to sample from the full conditional distribution. Details for the MCMC algorithms are given in Appendix E.

6. DISCUSSION

6.1. Data analysis results

Tables 1–3 give the medians of the posterior samples of the parameters, along with the 95% credible intervals. The DIC values for the three models fitted are 1997, 1326, and 1497, for the independent, reduced spatial, and full spatial multinomial models, respectively. That is, the best model is the reduced spatial multinomial model as it has the smallest DIC value.

Table 1. Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from the *reduced spatial multinomial model*

Percentile	Intercept	`Reserv`	`Fragment`	`TotOwn`	`AvParcel`
Aspen-paper birch					
2.5%	0.28	0.14	−0.70	0.11	−0.09
50%	0.97	1.36	−0.16	0.63	0.49
97.5%	1.75	2.70	0.33	1.28	1.15
Agriculture grassland					
2.5%	−1.11	−4.53	−0.96	−1.58	−0.30
50%	−0.31	−3.62	−0.52	−1.08	0.21
97.5%	0.38	−2.82	−0.12	−0.66	0.73
	$\Omega_{1,1}$	$\Omega_{2,1}$	v^*	λ^*	
2.5%	−3.82	0.53	0.49	0.249	
50%	−2.67	1.08	0.79	0.251	
97.5%	−1.75	−1.88	1.20	0.252	

The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are `Reserv`, `Fragment`, `log TotOwn`, and `AvParcel`.

Table 2. Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from the *full spatial multinomial model*

Percentile	Intercept	`Reserv`	`Fragment`	`TotOwn`	`AvParcel`	
Aspen-paper birch						
2.5%	0.70	0.30	−0.66	−0.027	−0.087	
50%	1.16	1.09	−0.31	0.31	0.37	
97.5%	1.61	1.90	0.046	0.66	0.77	
Agriculture grassland						
2.5%	−0.37	−3.82	−1.06	−1.26	−0.36	
50%	0.15	−2.79	−0.63	−0.80	0.23	
97.5%	0.63	−1.73	−0.22	−0.33	0.77	
	$V_{1,1}$	$V_{1,2}$	$V_{2,2}$	$\Lambda_{1,1}$	$\Lambda_{1,2}$	$\Lambda_{2,2}$
2.5%	0.42	−0.15	0.49	0.249	−0.001	0.251
50%	0.63	−0.022	0.73	0.251	0.000	0.252
97.5%	0.88	0.087	0.96	0.252	0.000	0.252

The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are `Reserv, Fragment, log TotOwn`, and `AvParcel`.

Table 3. Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from the *independent multinomial model*

Percentile	Intercept	`Reserv`	`Fragment`	`TotOwn`	`AvParcel`
Aspen-paper birch					
2.5%	1.08	0.46	−0.51	−0.082	−0.001
50%	1.31	0.89	−0.22	0.17	0.34
97.5%	1.58	1.34	0.063	0.43	0.67
Agriculture grassland					
2.5%	0.71	−4.41	−0.82	−1.33	−0.30
50%	0.97	−3.75	−0.46	−0.98	0.10
97.5%	1.24	−3.17	−0.13	−0.64	0.55
	σ_1^2	σ_2^2			
2.5%	0.68	0.65			
50%	1.03	0.94			
97.5%	1.63	1.45			

The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are `Reserv, Fragment, log TotOwn`, and `AvParcel`.

On the basis of the results of the best model in Table 1, there appears to be a relationship between land cover, and `Reserv`, `Fragment`, and `TotOwn`. In particular, the regression coefficient for `Reserv` is in the positive range for APB and is in the negative range for agriculture grassland with all others as the baseline. In a quarter section that is on an Indian reservation, the odds of APB increase by 3.91, but the odds of agriculture grassland decrease by 37.19, relative to all others. The regression coefficient for `Fragment` is in the negative range for agriculture grassland with all others as the baseline. That is, as the `Fragment` increases by one sample standard deviation 0.29 (or, by one unit), the odds of agriculture grassland decrease by 1.68 (or, by 5.89). Further, the regression coefficient for `TotOwn` (on the log scale) is in the positive range for APB and is in the negative range for agriculture grassland with all others as the baseline. When the log of `TotOwn` increases by one sample standard deviation 1.76 (or, by one unit), the odds of APB increase by 1.89 (or, by 1.43), whereas the odds of agriculture grassland decrease by 2.95 (or, by 1.85), relative to all others. Finally, the 95% credible intervals for the regression coefficients of `AvParcel` cover zero for both APB and agriculture grassland, indicating no relationship between land cover types and `AvParcel` after accounting for the other covariates. These results are in line with the plots of the data shown in Figure 3. Although multicollinearity does not appear to be an issue here, it can be with such regression analysis in general.

On the basis of the results of the other two models in Tables 2 and 3, similar results are observed with the exception that the 95% credible interval for the regression coefficient of `TotOwn` covers zero for APB.

6.2. Ecological and management implications

Two covariates have a positive association with APB: `Reserv`, a measure of ownership type, and `TotOwn`, a measure of ownership size. This finding illuminates the factors that influence landscape structure in northern Wisconsin, a major topic in landscape ecology. Whereas the influence of biotic and abiotic factors on landscape pattern is well documented (Turner, 2005), social factors are less well understood. The relationship between `Reserv` and APB is supported by related studies of the region's environmental and landscape history (Steen-Adams *et al.*, 2007, 2011): historical and ecological factors (e.g., U.S. Indian policy and legislation, forest succession dynamics on cutover sites, which were harvested later on-reservation) promoted nearly exclusive management for APB on the Indian reservation, in contrast to off-reservation, during the 1930 study date. The positive relationship between `TotOwn` and APB is also plausible. Regional history investigation shows that foresters frequently introduced practices to promote this early successional, pulpwood forest type. For reasons detailed elsewhere (timber contracts, maintenance of stumpage supply to the mills; Steen-Adams *et al.*, 2007, 2011), timber companies tended to maintain large land holdings, which in northern Wisconsin in the early 20th century meant APB.

Three covariates have a negative association with agriculture-grassland: `Reserv`, `TotOwn`, and `Fragment`. These results are consistent with related research of the region's environmental and landscape history (Steen-Adams *et al.*, 2007, 2011). Permanent mixed husbandry agriculture (in contrast to shifting cultivation), which 1930 land surveyors would have classified as agriculture, successfully established off the Indian reservation but generally failed on-reservation, corresponding with the variable `Reserv`. The negative relationship between land ownership size (`TotOwn`) and AG is also plausible: the conditions of the study area region and historical period promoted small-scale farmsteads, railroad land grantees and real estate companies subdivided land into small parcels affordable to the generally poor, immigrant, farm settlers. Likewise, the negative relationship between AG and quarter-section fragmentation (`Fragment`) makes sense for historical land economic reasons: time-consuming, enormous investments of labor and capital were required to develop cutover farmsteads, which promoted agricultural land use on small parcels for practical reasons.

The results of our statistical method to analyze nominal data, specifically historic (1930) land cover pattern, bear several implications for current forest management. One, this approach expanded our understanding of the development of several land covers and, more broadly, the landscape ecology of northern Wisconsin. Specifically, our analysis revealed that several land ownership covariates function as important drivers of landscape pattern (`Reserv`, also for AG cover, `Fragment`, `TotOwn`); this finding highlights the utility of tailoring policies to specific ownership characteristics to achieve management objectives. Our finding corresponds with those of studies in other regions (e.g., Crow *et al.*, 1999; Stanfield *et al.*, 2002; Spies *et al.*, 2007), which demonstrate that ownership variables can influence forest landscape pattern, including land cover diversity and patch size, and consequently, the delivery of ecological services to society.

Two, our new statistical methodology, which models the relationship between covariates and categorical response variable while accounting for spatial dependence, contributed capacity to assess historical factors. Our study builds on the expanding literature that demonstrates the utility of historical and long-term data to ecology (Turner, 2005). Investigators must work with the format of existing historical records, however. For example, the source of this study's outcome variable, the Wisconsin Land Economic Inventory, is most appropriately analyzed as nominal data. Our method expands the analytical tools to assess multiple covariates (or, factors). An important management implication is the persistence of social-ecological relationships: the finding of relationships between of 1915 ownership and 1930 land cover (e.g., ownership covariates `TotOwn`, `Fragment` and 1930 agriculture-grassland) suggests that social conditions may have a persisting influence on the landscape, indicating the importance of a relatively long-term perspective in developing policies.

Three, our finding of relationships between ownership type (`Reserv`) and land cover type suggests the need for greater coordination among ownerships across ecological units (a "multi-ownership perspective"), as other studies have concluded (Spies *et al.*, 2007). In the absence of a coordinated approach, contrasting motivations, policies, and applicable laws among ownerships can stymie achievement of ecosystem-level management goals.

Four, our statistical method substantially enhanced our capacity to interpret ownership-ecological relationships for specific land covers of interest, which thereby helps managers to discern influential factors on which to focus their interventions. For example, our findings can assist goals to promote agriculture grassland, a regionally important open land cover. Our result that this land cover type is negatively related to ownership size (i.e., increased likelihood on small ownerships) can signal managers to develop policies tailored to small owners when the management objective is to promote agriculture grassland.

6.3. Further remarks

We have demonstrated via the northern Wisconsin land cover data example that our methodology is useful for quantifying relationships between a nominal categorical response and covariates. We have used a reduced-dimensional model to help further reduce the computational cost. It appears that the resulting model is not only somewhat faster to implement but is also a better model.

We contend that the spatial multinomial models are worth the extra effort, as demonstrated by a number of results. One, judging from the DIC values, the spatial multinomial models outperform the independent models. Two, the inference for regression coefficients identifies relationship between `TotOwn` and land cover in the reduced spatial multinomial, which the independent multinomial model did not detect. In general, the Bayesian hierarchical model provides a rigorous, yet flexible, framework for statistical inference. For example, it is straightforward to obtain the posterior samples of the regression coefficients for any desirable choice of baseline category, which may not be easy to do using frequentist inference.

Under the log-ratio link function (2), it can be shown that there is a one-to-one correspondence between the model under two different baseline category choices in terms of the mean vector and the variance-covariance matrix of the latent variables. However, the interpretation of dependence structure may not be straightforward. An alternative approach is to consider the log of the probability π_{ji} for the ith site and the jth category,

$$\log(\pi_{ji}) = \tilde{Z}_{ji} - \log(C_i), \quad j = 0, \ldots, J \tag{13}$$

where $\tilde{Z}_{ji}$ is a latent variable (different from Z_{ji}) and $C_i = \sum_{j=0}^{J} \exp\left(\tilde{Z}_{ji}\right)$. That is, $\pi_{ji} = \exp\left(\tilde{Z}_{ji}\right)/C_i$ and is proportional to $\exp\left(\tilde{Z}_{ji}\right)$. An advantage of (13) is that it overcomes the difficulty in interpretation encountered in (2), as the dependence structures of the latent variables are not subject to the choice of baseline category. See Appendix A for more details. A potential disadvantage of (13) is that the dimension of the latent variable increases. Even though the increase is only from dimension J to $J+1$, we have encountered numerical instability when fitting the model to the land cover data. Further investigation will be needed to understand the cause of the computational difficulty and to develop counter measures. We leave this for future research.

Acknowledgements

Funding has been provided for this research from a USDA Cooperative State Research, Education and Extension Service (CSREES) Hatch project. The authors thank professors David J. Mladenoff and Nancy Langston for guidance and support to develop the northern Wisconsin land cover study. We thank Mark D.O. Adams for GIS assistance. A National Science Foundation IGERT Grant 9870703 (Human Dimensions of Social and Aquatic Systems Interactions) contributed to database development. We also thank two anonymous referees, an associate editor, and the editor for constructive comments that helped improve this work and the presentation.

REFERENCES

Aitchison J. 1987. *The Statistical Analysis of Compositional Data*. Chapman and Hall: London.

Besag J. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society, Series B* **36**: 192–236.

Brown DG. 2003. Land use and forest cover on private parcels in the Upper Midwest, USA , 1970 to 1990. *Landscape Ecology* **18**: 777–790.

Cleland DT, Avers PE, McNab WH, Jensen ME, Bailey RG, King T, Russell WE. 1997. National hierarchical framework of ecological units. In *Ecosystem Management Applications for Sustainable Forest and Wildlife Resources*, Boyce MS, Haney A (eds). Yale University Press: New Haven, CT; 181–200.

Cressie N. 1993. *Statistics for Spatial Data, Rev. Ed*. Wiley: New York.

Crow TR, Host GE, Mladenoff DJ. 1999. Ownership and ecosystem as sources of spatial heterogeneity in a forested landscape, Wisconsin, USA. *Landscape Ecology* **14**: 449–463.

Delcourt HR, Delcourt PA. 1996. Presettlement landscape heterogeneity:evaluating grain of resolution using General Land Office Survey data. *Landscape Ecology* **11**: 363–381.

De Oliveira V. 2000. Bayesian prediction of clipped Gaussian random fields. *Computational Statistics and Data Analysis* **34**: 299–314.

Diggle PJ, Tawn JA, Moyeed RA. 1998. Model-based geostatistics (with discussion). *Applied Statistics* **47**: 299–350.

Dupouey JL, Dambrine E, Laffite JD, Moares D. 2002. Irreversible impact of past land use on forest soils and biodiversity. *Ecology* **83**: 2978–2984.

Environmental Systems Resource Institute. 2002. *ArcView GIS 3.3*. ESRI: Redlands, California.

Environmental Systems Resource Institute. 2006. *ArcGIS 9.2*. ESRI: Redlands, California.

Foster DR, Aber JD (eds). 2004. *Forests in time: the environmental consequences of 1,000 years of change in New England*. Yale University Press: New Haven.

Foster DR, Donahue B, Kittredge DB, Motzkin G, Hall B, Turner BL, Chilton E. 2008. New England's forest landscape: ecological legacies and conservation patterns shaped by agrarian history. In *Agrarian Landscapes in Transition*, Redman CL, Foster DR (eds). Oxford University Press: New York.

Foster DR, Zebryk T, Schoonmaker P, Lezberg A. 1992. Post-settlement history of human land-use and vegetation dynamics of a Tsuga Canadensis (Hemlock) woodlot in central New England. *Journal of Ecology* **80**: 773–786.

Fraterrigo JM, Turner MG, Pearson SM, Dixon P. 2005. Effects of past land use on spatial heterogeneity of soil nutrients in southern Appalachian forests. *Ecological Monographs* **75**: 215–230.

Geweke J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, Bernado JM, Berger JO, Dawid AP, Smith AFM (eds). Clarendon Press: Oxford; 169–193.

Heasley L. 2003. Shifting boundaries on a Wisconsin landscape: can GIS help historians tell a complicated story? *Human Ecology* **31**: 183–211.

Higgs MD, Hoeting JA. 2010. Clipped latent-variable model for spatially correlated ordered categorical data. *Computational Statistics and Data Analysis* **54**: 1999–2011.

Hubbard BT. 2009. *American Boundaries: The Nation, the States, the Rectangular Survey*. University of Chicago: Chicago.

Knowles AK (ed.). 2002. *Past Time, Past Place: GIS for History*. ESRI Press: Redlands.

Koch J. 2006. Touching every forty: John Bordner and the Wisconsin land economic inventory. *Wisconsin Magazine of History* **89**: 14–25.

Jin X, Carlin BP, Banerjee S. 2005. Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61**: 950–961.

Li X. 2006. Bayesian Analysis of Cross-classified Spatial Data with Autocorrelation. *PhD thesis*, University of Wisconsin, Madison.

Mardia KV. 1988. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis* **24**: 265–284.

McCullagh P, Nelder JA. 1989. *Generalized Linear Models, 2nd Ed*. Chapman and Hall: Boca Raton.

Medley KE, Pobocik CM, Okey BW. 2003. Historical changes in forest cover and land ownership in a Midwestern U.S. landscape. *Annals of the Association of American Geographers* **93**: 104–120.

Paciorek CJ, McLachlan JS. 2009. Mapping ancient forests: Bayesian inference for spatial-temporal trends in forest composition using the fossil pollen proxy record. *Journal of the American Statistical Association* **104**: 608–621.

Rhemtulla JM, Mladenoff DJ, Clayton MK. 2007. Regional land cover conversion in the U.S. upper Midwest: magnitude of change and limited recovery (1850-1930-1993). *Landscape Ecology* **22**: 57–75.

Sain SR, Cressie N. 2007. A spatial model for multivariate lattice data. *Journal of Econometrics* **140**: 226–259.

Sain SR, Jagtap S, Mearns L, Nychka D. 2006. A multivariate spatial model for soil water profiles. *Journal of Agricultural, Biological, and Environmental Studies* **11**: 462–480.
Schulte LA, Mladenoff DJ. 2001. The original U.S. Public Land Survey records: their use and limitations in reconstruction pre-European settlement vegetation. *Journal of Forestry* **99**: 5–10.
Schulte LA, Mladenoff DJ. 2005. Severe wind and fire regimes in northern forests: historical variability at the regional scale. *Ecology* **86**: 431–445.
Spiegelhalter DJ, Best NG, Carlin BP, Van DL A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B* **64**: 583–639.
Spies TA, Johnson KN, Burnett KM, Ohmann JL, McComb BC, Reeves GH, Bettinger P, Kline JD, Garber-Yonts B. 2007. Cumulative ecological and socioeconomic effects of forest policies in coastal Oregon. *Ecological Applications* **17**: 5–17.
Stanfield BJ, Bliss JC, Spies TA. 2002. Land ownership and landscape structure: a spatial analysis of sixty-six Oregon (USA) Coast Range watersheds. *Landscape Ecology* **17**: 685–697.
Steen-Adams MM, Mladenoff DJ, Langston NE. 2007. White pine in the northern forests: an ecological and management history of white pine on the Bad River Reservation of Wisconsin. *Environmental History* **12**: 614–648.
Steen-Adams MM, Mladenoff DJ, Langston NE, Liu F, Zhu J. 2011. Influence of biophysical factors and differences in Ojibwe reservation versus Euro-American social histories on forest landscape change in northern Wisconsin, USA. *Landscape Ecology* **26**: 1165–1178.
Swetnam TW, Allen CD, Betancourt JL. 1999. Applied historical ecology: using the past to manage for the future. *Ecological Applications* **9**: 1189–1206.
Theobald DM, Gosnell H, Riebsame WE. 1996. Land use and landscape change in the Colorado Mountains II: a case study of the East River Valley. *Mountain Research and Development* **16**: 407–418.
Turner MG. 2005. Landscape ecology: what is the state of the science? *Annual Review of Ecology, Evolution, and Systematics* **36**: 319–344.
Wang F, Wall MM. 2003. Generalized common spatial factor model. *Biostatistics* **4**: 569–582.
White MA, Mladenoff DJ. 1994. Old-growth forest landscape transitions from pre-European settlement to present. *Landscape Ecology* **9**: 191–205.
White CS. 1983. *A History of the Rectangular Survey System*. U.S. Government Printing Office: Washington DC.
Whitney GG. 1994. *From Coastal Forest to Fruited Plain: A History of Environmental Change in Temperate North America*. Cambridge University Press: New York.
Zheng Y, Zhu J. 2008. Markov chain Monte Carlo for spatial-temporal autologistic regression model. *Journal of Computational and Graphical Statistics* **17**: 123–127.
Zhu J, Eickhoff JC, Yan P. 2005. Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* **61**: 674–683.
Zhu J, Huang H-C, Wu J-P. 2005. Modeling spatial-temporal binary data using Markov random fields. *Journal of Agricultural, Biological, and Environmental Statistics* **10**: 212–225.

