A Bayesian Hierarchical Model for Spatial Mapping of Multinomial Proportions

Joint work with
Simon Goring
Jason McLachlan
Chris Paciorek
Jack Williams
Jun Zhu

Abstract

In ecological studies, it is often desirable to characterize the composition of different species across a spatial domain. For a large number of spatial units, applying spatial models can be computationally cumbersome, especially in a Bayesian context. Here, I propose a hierarchical spatial model for estimating species compositions using a Bayesian framework for inference that allows the full conditional distributions to be sampled directly, so that spatial information is handled in a computationally efficient manner. For illustration, I apply this method to public land survey data containing approximately thirty different tree species at over 2,000 spatial locations in the state of Wisconsin.

1. Introduction

This paper concerns spatial categorical data on a lattice commonly encountered in environmental and ecological studies. For example, in an effort to map forest composition and structure during the European settlement era, a research team of ecologists compiled data including forest composition, stem density, basal area, and biomass based on the Public Land Survey System (PLSS) records in the upper Midwestern United States (TO ADD SIMONS MANUSCRIPT INPREP). The spatial unit for data analysis is a 5-arcminute by 5-arcminute cell. Here we focus on the forest composition data which were recorded as the observed numbers of trees for different tree species at the sampling locations within each cell. Of interest is a rigorous yet computationally feasible statistical approach to estimate

the proportions of different tree species on this 5-arcminute raster grid of cells. The data size is large in this study, because both the number of grid cells and the number of tree species can be quite large. For example, in the state of Wisconsin alone, the total number of grid cells is 2,368, and there are 29 tree species under study.

A multinomial model that accounts for spatial dependence is suitable for addressing the ecological questions in our study. One possible approach is a type of Markov random field, known as the auto-multinomial models, where spatial dependence is accounted for by autoregressive terms (Cressie, 1993). However, an unknown normalizing constant in the likelihood function makes the computation of maximum likelihood estimates time consuming (Besag, 1974; Cressie, 1993; Li, 2006). 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 in a link function, in the spirit of Diggle et al. (1998). In particular, for spatial ordinal data, Higgs and Hoeting (2010) developed a probit model with a latent continuous spatial process and for spatial nominal data, Paciorek and McLachlan (2009) developed spatial multinomial models for the purpose of mapping the probability of categories whereas Jin et al. (2013) developed spatial multinomial regression models for associating nominal categorical responses to covariates. In Higgs and Hoeting (2010), Paciorek and McLachlan (2009), and Jin et al. (2013), Bayesian hierarchical modeling was applied for statistical inference.

Here we focus on spatial nominal data as Paciorek and McLachlan (2009) and Jin et al. (2013). Because the Markov chain Monte Carlo (MCMC) algorithms for computing the posterior distributions involved Gibbs sampler with Metropolis-Hastings updates in Paciorek and McLachlan (2009) and Jin et al. (2013), the computation tended to be intensive and thus may not be suitable for analyzing big ecological data. Alternatively, McCulloch and Rossi (1994) proposed a multinomial probit model for categorical data indexed by time. Their model used a latent variable interpretation of the multinomial probit model and developed inference in a Bayesian framework that avoided the computational difficulties of both maximum likelihood and method of simulated moments approaches to estimation. The full conditional distributions are sampled directly without requiring Metropolis-Hastings updates, so that estimation is computationally efficient.

The primary purpose of this paper is to develop a set of statistical tools that are suited for mapping the forest composition and are fast to compute for large sample size. In Section 2, we develop a spatial multinomial model for spatial nominal data on a lattice by utilizing the latent variable interpretation of the multinomial probit model and specifiying an intrinsic conditional autoregression (ICAR) model as the prior distribution for the first-order parameters. In Section 3, we apply Bayesian hierarchical modeling for statistical inference. In particular, we derive the full conditional distributions in closed forms and implement a Gibbs sampler based on these full conditional distributions. In Section 4, we present a numerical example to demonstrate the performance of the method described in Section 3. In Section 5, we return to analyze the forest composition data using our new methodology and present the result of the data analysis. We conclude with a discussion of the proposed method and results in Section 6.

2. Model

Suppose that I have data aggregated to I locations. Each location i consists of n_i objects that can be classfied into P distinct classes. For a given object j at location i, let Y_{ij}^W be the response variable indicating the class label, and without loss of generality, let $\{1, ..., P\}$ denote the set of class labels. I associate Y_{ij}^W with P latent variables $W_{ij1}, ..., W_{ijP}$ such that $Y_{ij}^W = p$ if and only if $W_{ijp} = \max_{p*} W_{ijp*}$; in other words, the maximum of a set of latent variables $\{W_{ijp}\}_{p=1}^P$ determines the class label of object j at location i. Let $\mathbf{W}$= $(W_{ijp}, j = 1, ..., n_i, i = 1, ..., I, p = 1, ..., P)^T$, $\mathbf{Y}$= $(Y_{ij}^W, j = 1, ..., n_i, i = 1, ..., I)^T$, and $\boldsymbol{\alpha}_p = (\alpha_{1p}, ..., \alpha_{Ip})^T$. Also, let $N(\mu, \Sigma)$ denote the normal distribution with mean vector μ and covariance matrix Σ and $IG(a, b)$ denote the inverse gamma distribution with shape parameter a and scale parameter b. Then I parametrize the model as follows:

$$(\mathbf{Y}|\mathbf{W}) \sim \prod_{i=1}^{I} \prod_{j=1}^{n_i} \prod_{p=1}^{P} 1(W_{ijy_{ij}^W} \geq W_{ijp})$$

$$(W_{ijp}|\alpha_{ip}) \sim N(\alpha_{ip}, 1), j = 1, 2, ..., n_i; i = 1, 2, ..., I; p = 1, 2, ..., P$$

$$(\boldsymbol{\alpha}_p|\sigma_p^2) \sim N(0, \sigma_p^2\big(D - C\big)^{-1}), p = 1, 2, ..., P$$

$$\sigma_p^2 \sim IG(\beta_p, \gamma_p), p = 1, 2, ..., P$$

Here, C is the $I \times I$ adjacency matrix defining the neighborhood relation of the locations; that is, $(C)_{ii'} = I\{\text{locations } i \text{ and } i' \text{ are neighbors}\}$. The matrix D is an $I \times I$ diagonal matrix with diagonal entry $(D)_{ii} = \sum_{i'=1}^{I} (C)_{ii'}$, $i = 1, ..., I$. The covariance matrix specification $\sigma_p^2\big(D - C\big)^{-1}$ is called an *intrinsic conditional autoregression (ICAR)*. Formally, the matrix $D - C$ is singular, but for the purposes of a Bayesian analysis, posterior distributions containing $D - C$ are proper. In addition, note that this model is not identifiable since, for any scalar c, $W_{ijp} + c$ produces the same likelihood as W_{ijp}, $p = 1, 2, ..., P$.

Consider the following example with I=2 locations that are neighbors and P=2 class labels. Each object j at location i is associated with two variables W_{ij1} and W_{ij2}, governed by the parameters α_{i1} and α_{i2}, respectively. Suppose that $\alpha_{i1} > \alpha_{i2}$. Then this model implies that any object j is more likely to be labeled 1 than 2 at location i. Also, because the two locations are neighbors, observe that the vector $(\alpha_{1p}, \alpha_{2p})^T$ has precision matrix

$$\sigma_p^{-2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The difference between α_{i1} and α_{i2} explains the *difference* in composition between *labels* 1 and 2 at location i, and the similarity between α_{1p} and α_{2p} explains the *correlation* between composition at *locations* 1 and 2 for label p.

3. Computation

3.1. Full conditional distributions

The latent variable specification allows the full conditional distributions to be sampled directly. Define $\bar{w}_{i.p} = n_i^{-1} \sum_{j=1}^{n_i} W_{ijp}$, $\bar{\mathbf{w}}_p = (\bar{w}_{1.p}, ..., \bar{w}_{I.p})^T$, $A = diag(n_1, ..., n_I)$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, ..., \boldsymbol{\alpha}_P^T)^T$, $\mathbf{v} = (\sigma_1^2, ..., \sigma_P^2)^T$, and $V_p^{-1} = \sigma_p^{-2}\big(D - C\big)$. Also, let $\phi(z)$ denote the standard normal density evaluated at z. Then the full conditional distributions are

$$\pi(w_{ijp}|\mathbf{Y}, \mathbf{W}_{-ijp} = \mathbf{w}_{-ijp}, \boldsymbol{\alpha}, \mathbf{v}) \propto \begin{cases} \phi(w_{ijy_{ij}^W} - \alpha_{iy_{ij}^W}) I(w_{ijy_{ij}^W} \geq \max_{p^* \neq y_{ij}^W} w_{ijp^*}), & \text{if } p = y_{ij}^W \\ \phi(w_{ijp} - \alpha_{ip}) I(w_{ijp} \leq w_{ijy_{ij}^W}), & \text{if } p \neq y_{ij}^W \end{cases}$$

$$(\boldsymbol{\alpha}_p|\mathbf{Y}, \mathbf{W}, \boldsymbol{\alpha}_{-p}, \mathbf{v}) \sim N\left(\left[A + V_p^{-1}\right]^{-1} A\bar{\mathbf{w}}_p, \left[A + V_p^{-1}\right]^{-1}\right)$$
$$(\sigma_p^2|\mathbf{Y}, \mathbf{W}, \boldsymbol{\alpha}, \mathbf{v}_{-\mathbf{p}}) \sim IG\left(\frac{I}{2} + \beta_p, \gamma_p + \frac{1}{2}\boldsymbol{\alpha}_p^T (D - C)\boldsymbol{\alpha}_p\right).$$

In order to speed up computations, I take advantage of the sparsity of C to sample $\boldsymbol{\alpha}_p$ using sparse matrix computations.

3.2. Posterior distribution of $\mathbf{W}$

Let $\mathbf{W}_{-ijp}$ denote the set of all variables in $\mathbf{W}$ except W_{ijp}. Then

$$\begin{aligned} \pi(w_{ijp}|\mathbf{Y}, \mathbf{W}_{-ijp} = \mathbf{w}_{-ijp}, \boldsymbol{\alpha}, \mathbf{v}) &\propto \pi(\mathbf{Y}|\mathbf{W})\pi(\mathbf{W}|\boldsymbol{\alpha}) \\ &\propto \left(\prod_{i=1}^{I}\prod_{j=1}^{n_i}\prod_{p=1}^{P} I(w_{ijy_{ij}^W} \geq w_{ijp})\right)\left(\prod_{i=1}^{I}\prod_{j=1}^{n_i}\prod_{p=1}^{P} \phi(w_{ijp} - \alpha_{ip})\right) \end{aligned}$$

Observe that

$$\pi(w_{ijp}|\mathbf{Y}, \mathbf{W}_{-ijp} = \mathbf{w}_{-ijp}, \boldsymbol{\alpha}, \mathbf{v}) \propto \begin{cases} \phi(w_{ijy_{ij}^W} - \alpha_{iy_{ij}^W}) I(w_{ijy_{ij}^W} \geq \max_{p^* \neq y_{ij}^W} w_{ijp^*}), & \text{if } p = y_{ij}^W \\ \phi(w_{ijp} - \alpha_{ip}) I(w_{ijp} \leq w_{ijy_{ij}^W}), & \text{if } p \neq y_{ij}^W \end{cases}$$

Thus, W_{ijp} is drawn from a truncated normal distribution, and the truncation value is determined by the class label of object j in cell i.

3.3. Posterior distribution of $\boldsymbol{\alpha}_p$

Let $\boldsymbol{\alpha}_{-p}$ denote the vector of α_{iq} for all $q \neq p$. The posterior distribution of $\boldsymbol{\alpha}_p$ is

$$
\begin{aligned}
\pi(\boldsymbol{\alpha}_p|\mathbf{Y},\mathbf{W},\boldsymbol{\alpha}_{-p},\mathbf{v}) &\propto \pi(\mathbf{W}|\boldsymbol{\alpha})\pi(\boldsymbol{\alpha}|\mathbf{v}) \\
&\propto \Big[\prod_{i=1}^{I}\exp\Big\{-\frac{1}{2}\sum_{j=1}^{n_i}(w_{ijp}-\alpha_{ip})^2\Big\}\Big]\exp\Big[-\frac{1}{2}\boldsymbol{\alpha}_p^T\big\{\sigma_p^2(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big] \\
&\propto \Big[\prod_{i=1}^{I}\exp\Big\{-\frac{1}{2}\sum_{j=1}^{n_i}(w_{ijp}^2-2w_{ijp}\alpha_{ip}+\alpha_{ip}^2)\Big\}\Big]\exp\Big[-\frac{1}{2}\boldsymbol{\alpha}_p^T\big\{\sigma_p^2(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big] \\
&\propto \Big[\prod_{i=1}^{I}\exp\Big\{-\frac{1}{2}(-2\alpha_{ip}\sum_{j=1}^{n_i}w_{ijp}+n_i\alpha_{ip}^2)\Big\}\Big]\exp\Big[-\frac{1}{2}\boldsymbol{\alpha}_p^T\big\{\sigma_p^2(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big] \\
&\propto \Big[\prod_{i=1}^{I}\exp\Big\{-\frac{1}{2\frac{1}{n_i}}(-2\bar{w}_{i.p}\alpha_{ip}+\alpha_{ip}^2)\Big\}\Big]\exp\Big[-\frac{1}{2}\boldsymbol{\alpha}_p^T\big\{\sigma_p^2(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big] \\
&\propto \Big[\prod_{i=1}^{I}\exp\Big\{-\frac{1}{2}\frac{(\alpha_{ip}-\bar{w}_{i.p})^2}{\frac{1}{n_i}}\Big\}\Big]\exp\Big[-\frac{1}{2}\boldsymbol{\alpha}_p^T\big\{\sigma_p^2(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big]
\end{aligned}
$$

Next, I combine the two terms:

$$
\begin{aligned}
\pi(\boldsymbol{\alpha}_p|\mathbf{Y},\mathbf{W},\boldsymbol{\alpha}_{-p},\mathbf{v}) &\propto \Big[\exp\Big\{-\frac{1}{2}(\boldsymbol{\alpha}_p-\bar{\mathbf{w}}_p)^T A(\boldsymbol{\alpha}_p-\bar{\mathbf{w}}_p)\Big\}\Big]\exp\Big[-\frac{1}{2}\boldsymbol{\alpha}_p^T\big\{\sigma_p^2(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big] \\
&\propto \exp\Big[\Big(-\frac{1}{2}\Big)\Big\{(\boldsymbol{\alpha}_p-\bar{\mathbf{w}}_p)^T A(\boldsymbol{\alpha}_p-\bar{\mathbf{w}}_p)+\boldsymbol{\alpha}_p^T V_p^{-1}\boldsymbol{\alpha}_p\Big\}\Big] \\
&\propto \exp\Big[\Big(-\frac{1}{2}\Big)\Big\{\boldsymbol{\alpha}_p^T A\boldsymbol{\alpha}_p-2\bar{\mathbf{w}}_p^T A\boldsymbol{\alpha}_p+\bar{\mathbf{w}}_p^T A\bar{\mathbf{w}}_p+\boldsymbol{\alpha}_p^T V_p^{-1}\boldsymbol{\alpha}_p\Big\}\Big] \\
&\propto \exp\Big[\Big(-\frac{1}{2}\Big)\Big\{\boldsymbol{\alpha}_p^T\big(A+V_p^{-1}\big)\boldsymbol{\alpha}_p-2\bar{\mathbf{w}}_p^T A\boldsymbol{\alpha}_p\Big\}\Big] \\
&\propto \exp\Big[\Big(-\frac{1}{2}\Big)\Big\{\big(A+V_p^{-1}\big)^{\frac{1}{2}}\boldsymbol{\alpha}_p-\big(A+V_p^{-1}\big)^{-\frac{1}{2}}A\bar{\mathbf{w}}_p\Big\}^T \\
&\qquad \Big\{\big(A+V_p^{-1}\big)^{\frac{1}{2}}\boldsymbol{\alpha}_p-\big(A+V_p^{-1}\big)^{-\frac{1}{2}}A\bar{\mathbf{w}}_p\Big\}\Big] \\
&\propto \exp\Big[\Big(-\frac{1}{2}\Big)\Big\{\boldsymbol{\alpha}_p-\big(A+V_p^{-1}\big)^{-1}A\bar{\mathbf{w}}_p\Big\}^T\big(A+V_p^{-1}\big) \\
&\qquad \Big\{\boldsymbol{\alpha}_p-\big(A+V_p^{-1}\big)^{-1}A\bar{\mathbf{w}}_p\Big\}\Big]
\end{aligned}
$$

Thus, the posterior distribution $(\boldsymbol{\alpha}_p|\mathbf{Y},\mathbf{W},\boldsymbol{\alpha}_{-p},\mathbf{v})$ is the multivariate normal distribution

$$(\boldsymbol{\alpha}_p|\mathbf{Y},\mathbf{W},\boldsymbol{\alpha}_{-p},\mathbf{v}) \sim N\bigg(\big(A+V_p^{-1}\big)^{-1}A\bar{\mathbf{w}}_p, \big(A+V_p^{-1}\big)^{-1}\bigg).$$

3.4. Posterior distribution of hyperparameter σ_p^2

Let $\mathbf{v_{-p}}$ denote the vector of σ_q^2 for all $q \neq p$. The posterior distribution of σ_p^2 is

$$\begin{aligned}\pi(\sigma_p^2|\mathbf{Y},\mathbf{W},\boldsymbol{\alpha},\mathbf{v_{-p}}) &\propto \pi(\boldsymbol{\alpha}_p|\sigma_p^2)\pi(\sigma_p^2)\\ &\propto \bigg(|\sigma_p^2(D-C)^{-1}|^{-1/2}\exp\Big[(-\frac{1}{2})\boldsymbol{\alpha}_p^T\big\{\sigma_p^2(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big]\bigg)\exp(-\gamma_p\sigma_p^2)(\sigma_p^2)^{-\beta_p-1}\\ &\propto (\sigma_p^2)^{-I/2}\exp\bigg(-\Big[\gamma_p+\frac{1}{2}\boldsymbol{\alpha}_p^T\big\{(D-C)^{-1}\big\}^{-1}\boldsymbol{\alpha}_p\Big]\sigma_p^{-2}\bigg)(\sigma_p^2)^{-\beta_p-1}.\end{aligned}$$

Thus, $(\sigma_p^2|\mathbf{Y},\mathbf{W},\boldsymbol{\alpha},\mathbf{v_{-p}}) \sim IG\bigg(\dfrac{I}{2}+\beta_p, \gamma_p+\dfrac{1}{2}\boldsymbol{\alpha}_p^T(D-C)\boldsymbol{\alpha}_p\bigg)$.

3.5. Inference

The quantity

$$\theta_{ip} = P(W_{ijp} = \max_{p^*} W_{ijp^*}|\alpha_{ip'}, p' = 1,...,P)$$

defines the composition of class p at location i. Consider a Markov chain Monte Carlo (MCMC) chain with length M. Let $\boldsymbol{\alpha}^{(m)}$ denote the set of variates at the mth step in the chain. I sample a new set of variates $\{{W^*_{ijp}}^{(m)}, i = 1,...,I, j = 1,...,n_i, p = 1,...,P\}$ given $\boldsymbol{\alpha}^{(m)}$. For each sample m, I define

$$\hat{\theta}_{ip}^{(m)} = n_i^{-1}\sum_{j=1}^{n_i} 1\big\{{W^*_{ijp}}^{(m)} = \max_{p^*} {W^*_{ijp^*}}^{(m)}\big\}.$$

Then an estimate of the probability of class p at location i is

$$\bar{\theta}_{ip} = M^{-1}\sum_{m=1}^{M}\hat{\theta}_{ip}^{(m)}$$

and an estimate of the variance of $\bar{\theta}_{ip}$ is

$$\hat{\sigma}^2_{\bar{\theta}_{ip}} = (M-1)^{-1} \sum_{m=1}^{M} \left(\hat{\theta}^{(m)}_{ip} - \bar{\theta}_{ip}\right)^2.$$

4. Numerical Example

Here I present a numerical example to investigate the adequacy of inference for estimating the surface of composition estimates as a function of n_i, $i = 1, ..., I$ and σ^2_p, $p = 1, ..., P$. Data was generated from the proposed model with $P = 5$ species in a square grid of $I = 100$ cells. The neighborhood was defined to be the four nearest neighbors of each cell, which are the adjacent cells in the four cardinal directions. Because the ICAR model cannot be used to generate data, data was generated from a CAR model specification, which has the covariance matrix $\sigma^2(I_{I\times I} - \rho C)^{-1}$, where $I_{I\times I}$ is the $I \times I$ identity matrix. The parameter ρ is constrained to be between $\lambda^{-1}_{\min}$ and $\lambda^{-1}_{\max}$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of C, respectively. Here, I set $\rho = \lambda^{-1}_{\max}$ - .01 to induce high correlation. For simplicity, I fixed $n = n_i$ for all i, and $\sigma^2 = \sigma^2_p$ for all p. The values of n and σ^2 were varied to investigate the influence of these two quantities on inference. First, $n = 100$ trees were generated in each cell. Next, a sample of size $n = 50$ trees were randomly subsampled from these 100 trees in each cell to form a new set of data. Then a subsample of $n = 10$ trees were randomly subsampled from these 50 trees in each cell. This subsampling scheme was repeated for data sets with $n = 5$ and $n = 1$ trees per cell.

Three methods were applied. The *independence method* applies the proposed method but treats the data in different cells as independent. The *CAR method* fixes the correlation parameter of the CAR model to be $\rho = \lambda^{-1}_{\max}$ - .0001. The *ICAR method* uses an intrinsic CAR model to perform inference. For each method, I ran the Gibbs sampler for a chain of length 1,000 and used the last 500 variates to perform inference. Also, the prior distribution of σ^2_p was assumed to be the improper prior distribution $IG(-1/2, 0)$.

Figures 1–3 plot the true probabilities, sample proportions of the simulated data, and estimates computed under the independence, CAR, and ICAR methods for species 1 with $\sigma^2 = .01$, 1, and 10, respectively. The true probabilities and the sample proportions become

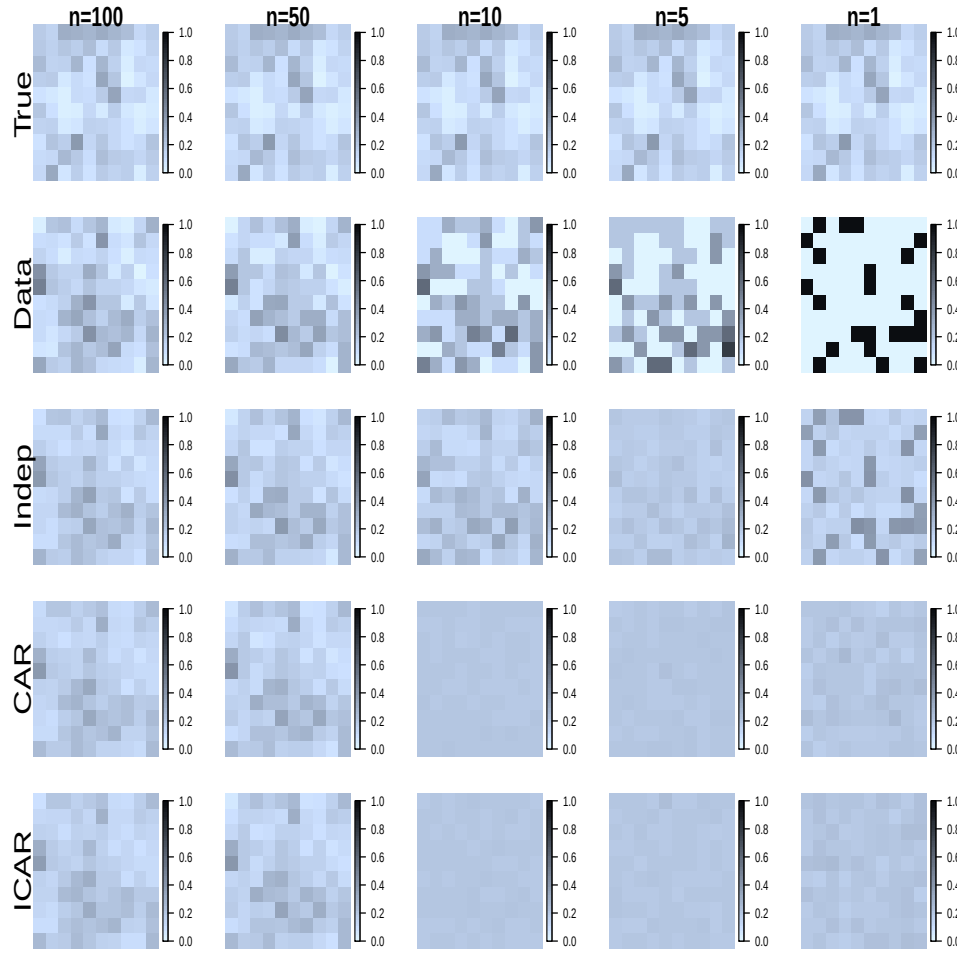


Figure 1: *Composition estimates for the simulated data,* $\sigma^2 = .1$

more heterogeneous as σ^2 increases and for the two largest values of σ^2, their values are almost the same when $n = 100$. As n increases, the composition estimates from all three methods approach the sample proportions. In Figures 2 and 3, the estimates from the models are very similar to the true probabilities.

Also, the smoothness of the probability surface depends on the magnitude of n. The images corresponding to the smaller values of n in each plot tend to be smoother. Also, the composition estimates can be quite different from the true probabilities for very small values of n. Nevertheless, for a reasonably large value of n, the proposed method accurately captures the values of the true probability surface.

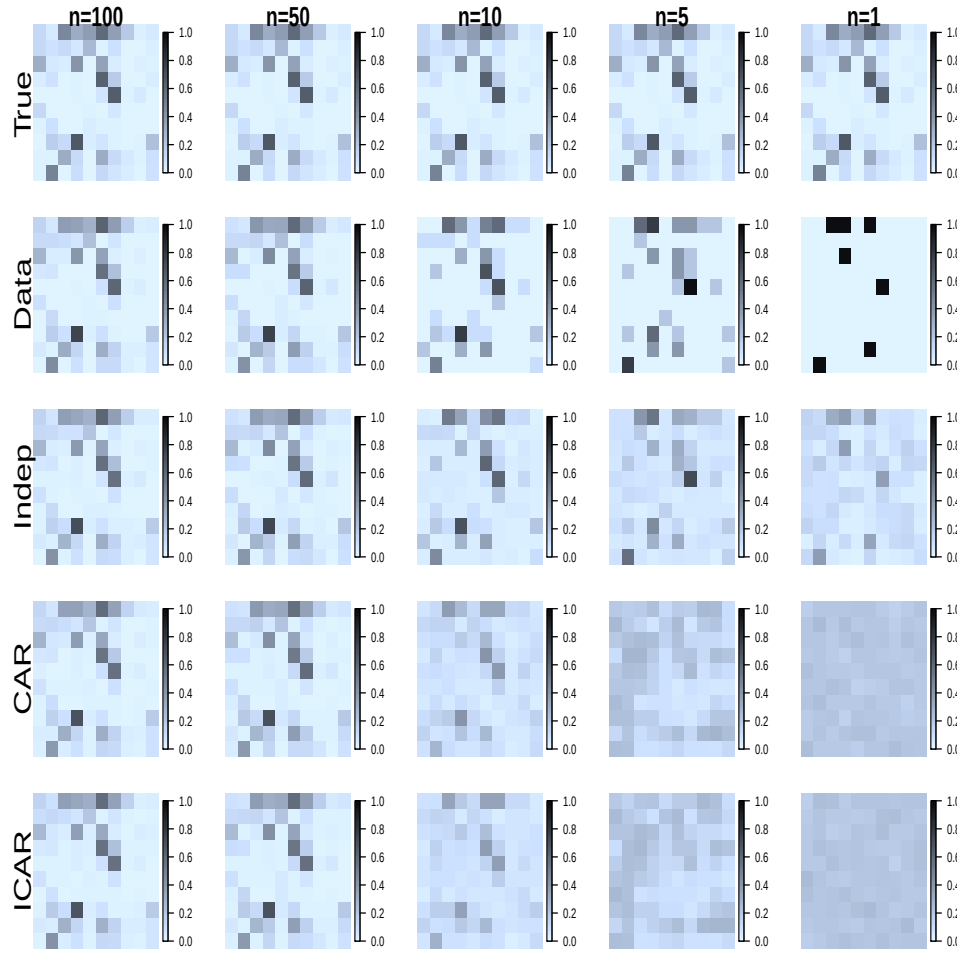


Figure 2: *Composition estimates for the simulated data,* $\sigma^2 = 1$

5. Application

The proposed model was applied to public land survey (PLS) data consisting of 338,546 trees located in I=2,368 cells in the state of Wisconsin. Trees were classified into P=29 taxa. All of the cells had at least 50 trees except for one cell with only 22 trees, and the median and interquartile range of the number of trees per cell were 144 and 14, respectively. A diamond-shaped neighbor relation, with two neighbors in each of the four cardinal directions and four neighbors in the diagonal directions, was used to produce a neighborhood structure for the ICAR model. Figure 4(a) plots the locations of the centroids of the 2,368 cells with the neighborhood structure for one of the cells in bold, and Figure 4(b) gives a plot of the number of distinct species observed in each cell.

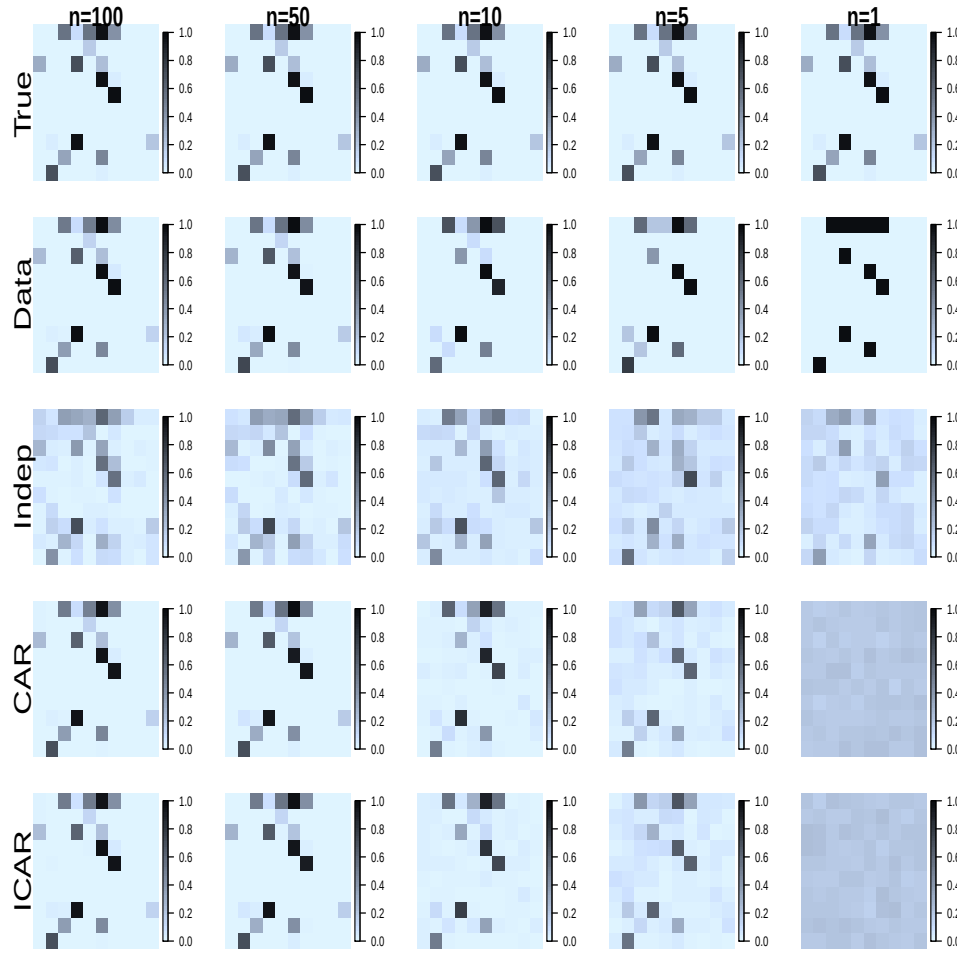
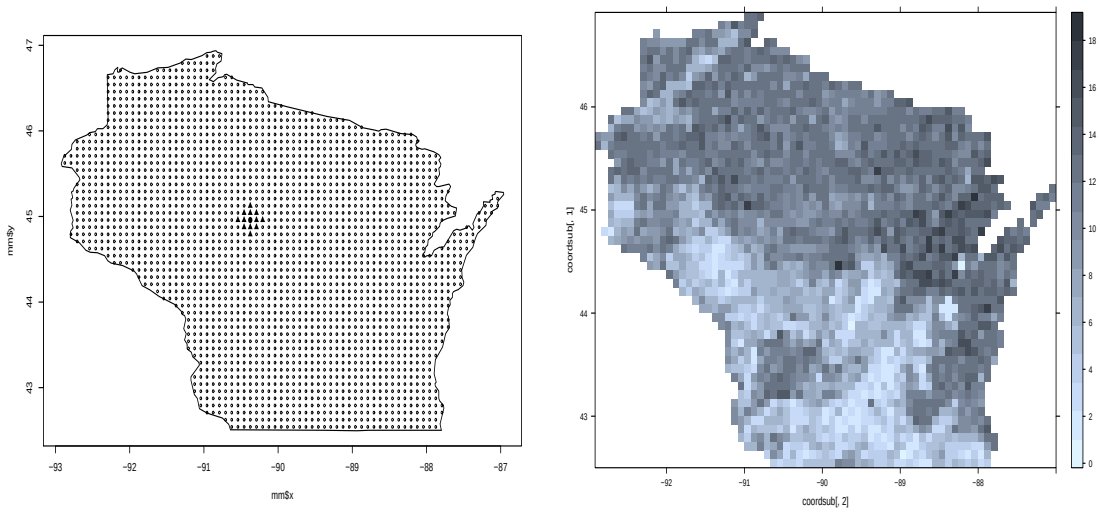


Figure 3: *Composition estimates for the simulated data,* $\sigma^2 = 10$

An MCMC chain of length 1000 was generated according to the sampling scheme developed in Section 3.1. The first 500 variates were discarded for burn-in, and every five of the other 500 variates were subsampled, resulting in 100 of each variate to be used for inference. The hyperparameters β_p and γ_p were set to be 1 for all $p = 1, ..., P$.

Figures 5 and 6 give plots of the sample proportions and composition estimates $\bar{\theta}_{ip}$ for the 29 taxa across the domain, respectively. Generally, the composition estimates are very similar to the sample proportions across the landscape. Specifically, I focus on two taxa, *Quercus* and *Cornus*, for further investigation. *Quercus* represents an abundant taxon with 95,542 stems in the region, while *Cornus* represents an extremely sparse taxon with only one stem in the entire region.

Figure 7 provides a plot of the sample proportions and estimates $\bar{\theta}_{ip}$, as in Figures 5



(a) (b)

Figure 4: *PLS data – (a) the centroids of the 2,368 cells containing in Wisconsin. The darkened circle indicates a cell for which the darkened triangles are neighbors. (b) the number of observed species per cell.*

and 6, but only for *Quercus*. Figure 8 provides a similar plot for *Cornus*. The composition estimates for an abundant taxon like *Quercus* are very similar to the sample proportions in the southwestern region of Wisconsin, where *Quercus* is relatively abundant, but in the northern half of Wisconsin, where *Quercus* is relatively sparse, the model seems to produce composition estimates of *Quercus* much closer to zero than the sample proportions in that region. On the other hand, composition estimates and sample proportions differ for an extremely rare taxon like *Cornus*. The model seems to produce small but nonzero estimates of *Cornus* composition at locations where no *Cornus* stems were observed. In fact, the plot of the composition estimates for *Cornus* and the plot of species richness in Figure 4(b) are very similar. Despite the apparent structure of the composition estimates in Figure 8(b), the 95^{th} percentile is .0089, indicating that most of the composition estimates are very close to zero.

Figures 9 and 10 provide plots of the standard errors of the estimates given in Figures 7 and 8, respectively. For the sample proportion $\hat{p}$ in a cell, I use the following variance estimator for n observed stems in that cell:

$$\hat{\sigma}^2_{\hat{p}} = n^{-1}\hat{p}(1-\hat{p}).$$

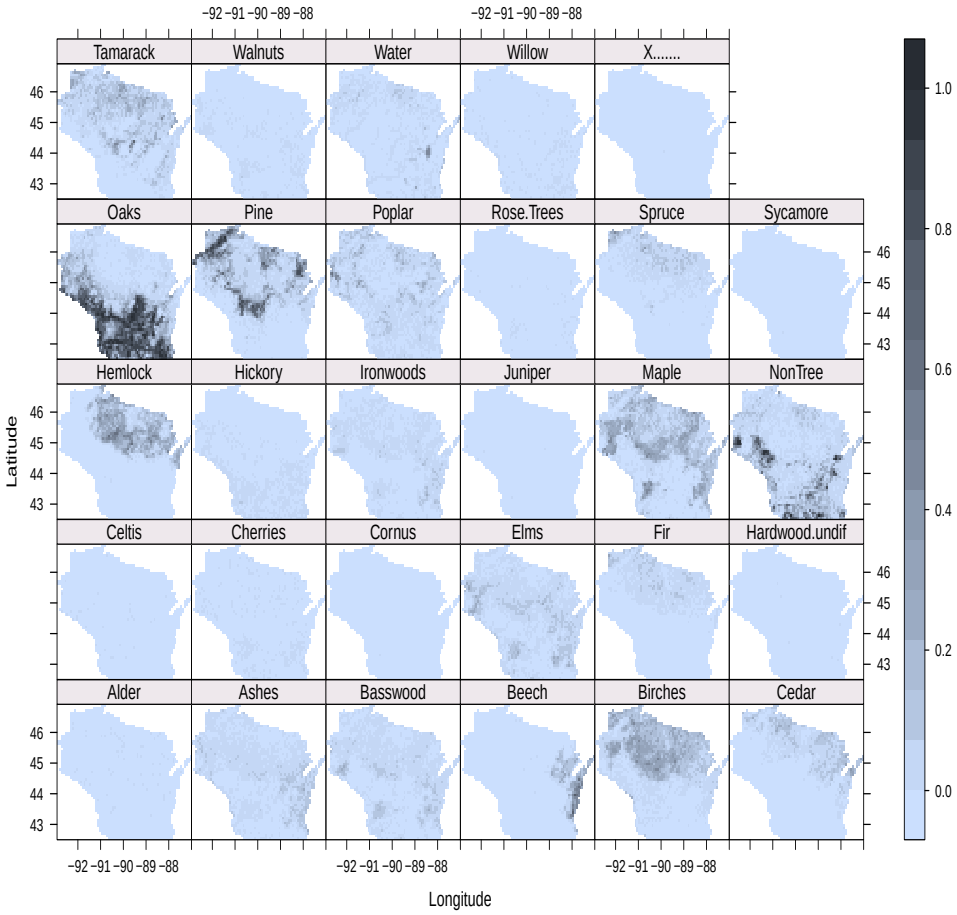


Figure 5: *Plot of the sample proportions of the 29 taxa across the domain. Dark colors indicate higher proportions of the taxon in each cell.*

The standard errors of the composition estimates seem to be very similar to those of the sample proportions for an abundant taxon like *Quercus*. On the other hand, for an extremely sparse taxon like *Cornus*, because the standard errors of the sample proportions are zero for cells in which no *Cornus* stems were observed, the composition estimates have larger standard errors.

5.1. Composition estimates for rare species

For rare species, the proposed method produces small but nonzero estimates of composition. I constructed another numerical example to investigate this issue. I simulated data according to the procedure described in Section 4. In addition to the five classes in Section 4, I added a single object with class label 6 in one cell. In addition to implementing the

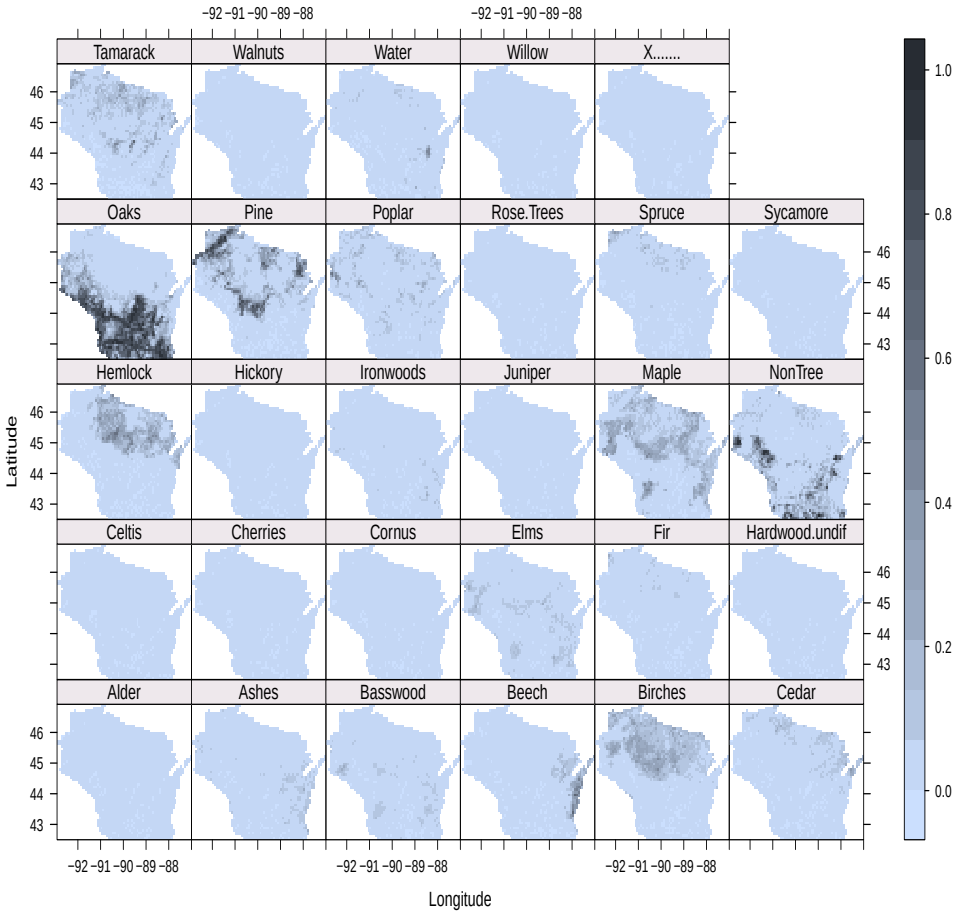
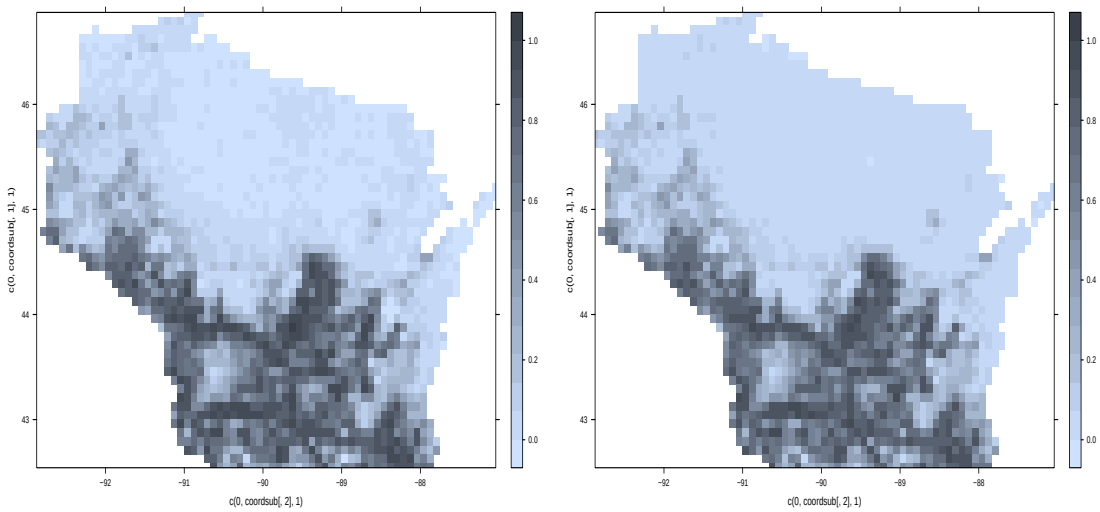


Figure 6: *Plot of $\bar{\theta}_{ip}$ for the 29 taxa across the domain. Dark colors indicate higher proportions of the taxon in each cell.*

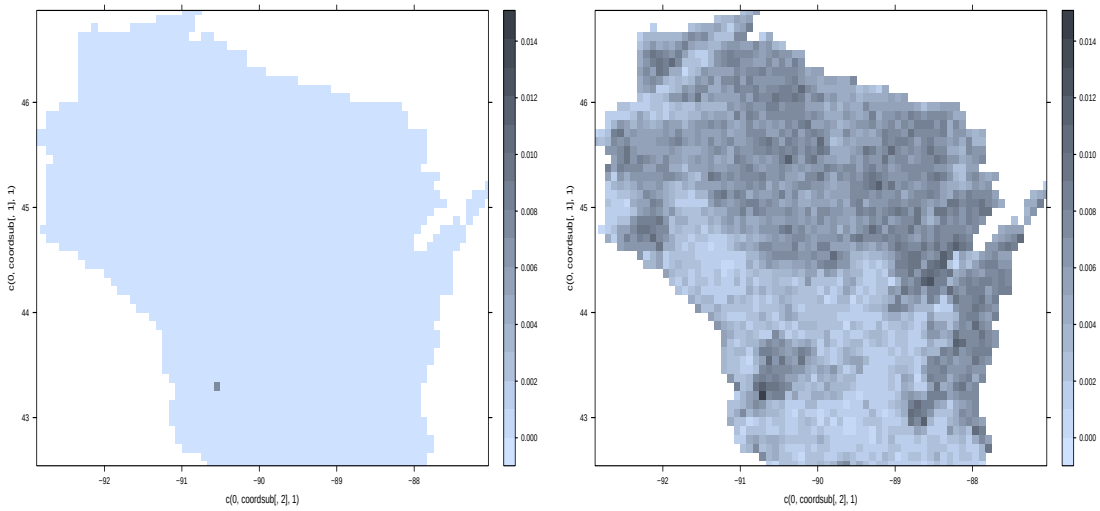
independence method, I modeled the cell counts as a multinomial distribution with three different noninformative priors from the Dirichlet($\boldsymbol{\gamma}$) family of distributions and estimated the composition of species 6 in the 100 cells. Figure 11 and 12 plot the composition estimates in the 100 cells for species 1 and 6, respectively. Each column corresponds to a different prior distribution from the Dirichlet($\boldsymbol{\gamma}$) family. The three parameter specifications were $\boldsymbol{\gamma} = (0, 0, 0, 0, 0, 0)^T$, $\boldsymbol{\gamma} = (1/2, 1/2, 1/2, 1/2, 1/2, 1/2)^T$ (Jeffreys' prior), and $\boldsymbol{\gamma} = (1, 1, 1, 1, 1, 1)^T$ (Uniform prior). The two rows in each plot correspond to different y-axis scales.

For species 1, the composition estimates are very similar to the sample proportions in all cells. Also, the prior distribution of the multinomial-Dirichlet method seems to have little influence on the composition estimates. For species 6, the prior distribution for the



(a) sample proportions (b) $\bar{\theta}_{ip}$

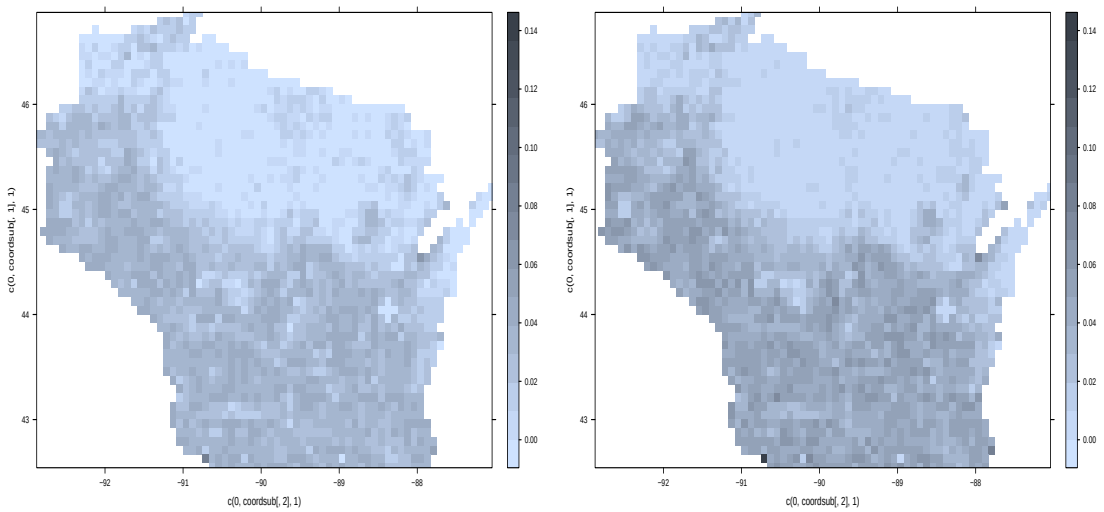
Figure 7: *Plot of the (a) sample proportions and (b) $\bar{\theta}_{ip}$ for Quercus. Dark colors indicate higher proportions of the taxon in each cell.*



(a) sample proportions (b) $\bar{\theta}_{ip}$

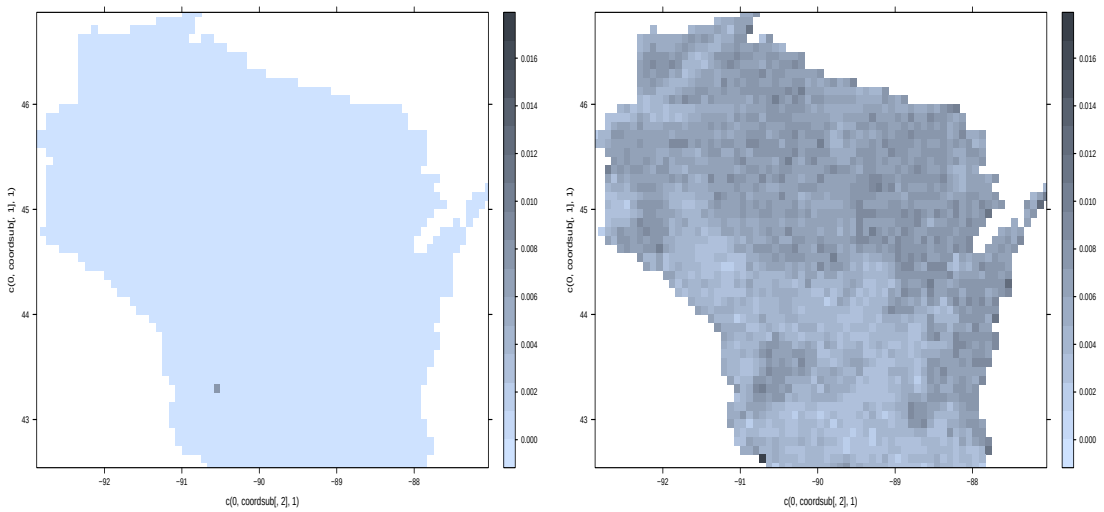
Figure 8: *Plot of the (a) sample proportions and (b) $\bar{\theta}_{ip}$ for Cornus. Dark colors indicate higher proportions of the taxon in each cell.*

multinomial-Dirichlet method has a large influence on the composition estimates, but on the [0,1] scale, the composition estimates for both methods are similar to the sample proportions. This behavior could be due to estimating a parameter on the boundary of the parameter space.



(a) $\hat{\sigma}_{\hat{p}}$ (b) $\hat{\sigma}_{\bar{\theta}_{ip}}$

Figure 9: *Plot of the standard errors of the Joint work with Simon Goring (University of Wisconsin–Madison), Jason McLachlan (University of Notre Dame), Chris Paciorek (University of California–Berkeley), Jack Williams (University of Wisconsin–Madison), and Jun Zhu (University of Wisconsin–Madison)(a) sample proportions and (b) $\bar{\theta}_{ip}$ for Quercus. Dark colors indicate higher values.*



(a) $\hat{\sigma}_{\hat{p}}$ (b) $\hat{\sigma}_{\bar{\theta}_{ip}}$

Figure 10: *Plot of the standard errors of the (a) sample proportions and (b) $\bar{\theta}_{ip}$ for Cornus. Dark colors indicate higher values.*

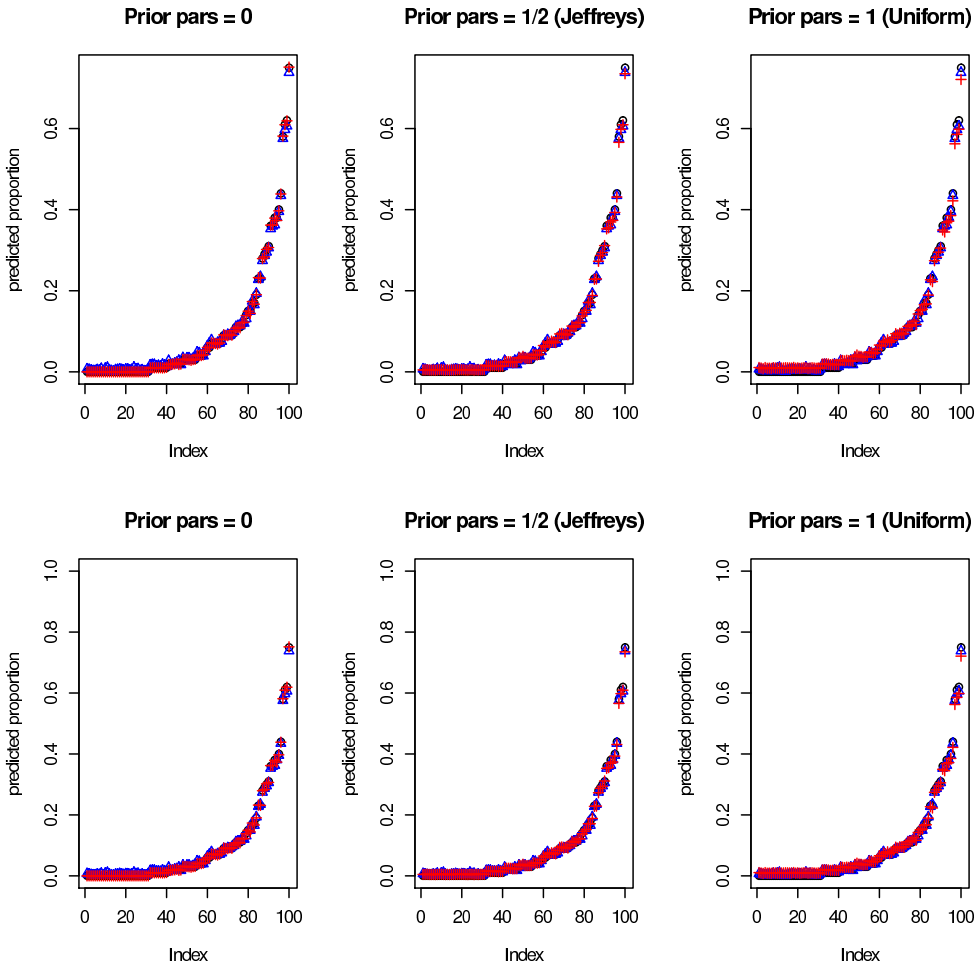


Figure 11: *Sample proportions (black circles), independence method (blue triangles), and multinomial-Dirichlet (red plus signs) composition estimates for species 1 – Top row gives a small scale on the y-axis and the bottom row gives the [0,1] scale.*

6. Conclusion and Discussion

Here I have presented a hierarchical spatial model for composition and a method for inference in a Bayesian framework. Because of the latent variable specification, the full conditional distributions can be sampled directly. Also, because of the sparse neighbor relation, high dimensional multivariate normal distributions can be sampled using efficient sparse matrix computations. The numerical example showed that the true compositions could be estimated accurately when there was a sufficient degree of spatial heterogeneity and large enough number of objects per location. The application of this method to PLS data in Wisconsin showed that the compositions of 29 species of trees were generally similar

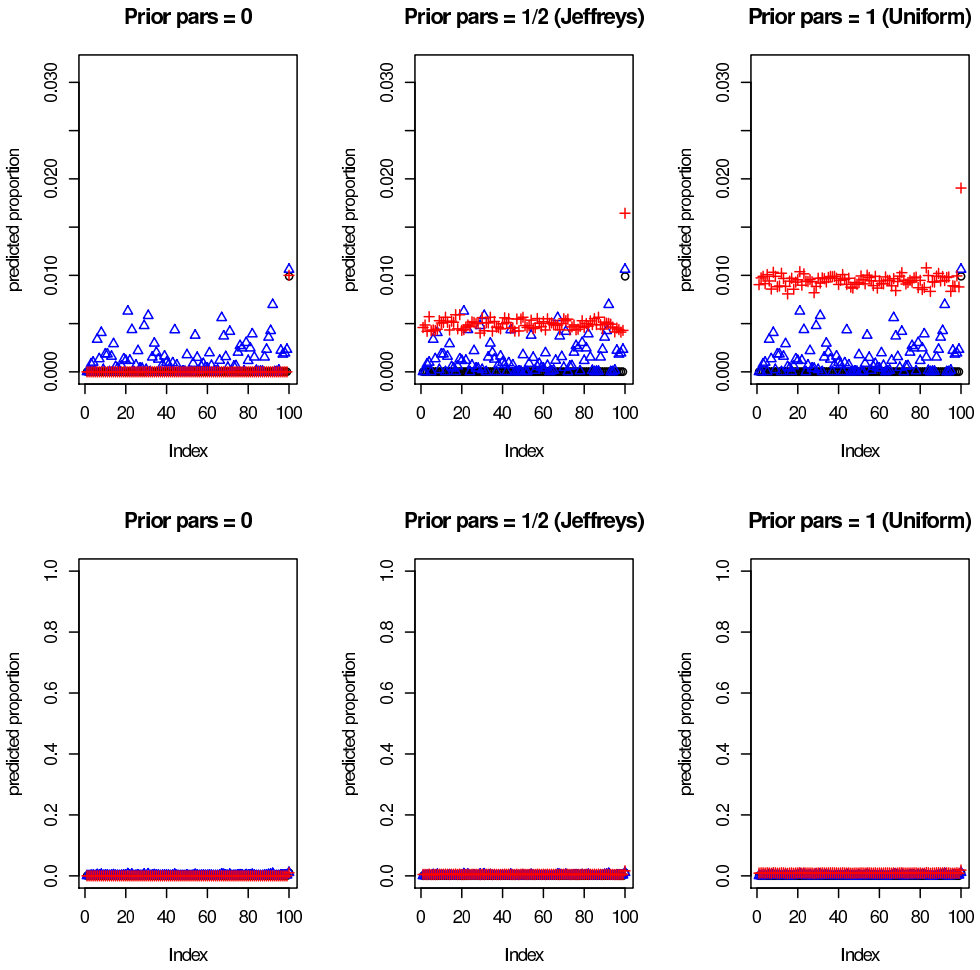


Figure 12: *Sample proportions (black circles), independence method (blue triangles), and multinomial-Dirichlet (red plus signs) composition estimates for species 6 – Top row gives a small scale on the y-axis and the bottom row gives the [0,1] scale.*

to the sample proportions at each location, but for extremely rare species, the composition estimates were small but nonzero at location where no trees were observed.

In both the numerical example and application, the estimated compositions were similar to the sample proportions, and there was little evidence of spatial smoothing when the cell sample sizes were large. It may be useful to include more than one variance parameter, σ_p^2, for each species. For example, *Quercus* composition is large in the southern half of Wisconsin but small in the northern half. If there were two variance parameters to govern the ICAR models in these two halves of the state, the composition estimates may be estimated more smoothly because the variability in each half is much smaller than the variability in the entire state.

Also, one modification would be to investigate other covariance matrix specifications, for example, a Matérn covariance model approximation for lattice data (Lindgren et al., 2011). The covariance between locations is an approximation of the well-known Matérn covariance model used in geostatistics, and it could potentially provide smoother estimates of composition.

Also, it would be useful to investigate how integrated nested Laplace approximation (INLA) could be used to obtain composition estimates more quickly than MCMC (Rue et al., 2009). INLA takes advantage of the latent Gaussian process specification to directly approximate the posterior distribution of the variables of interest instead of using MCMC to sample those variables. The difference in computational time between INLA and MCMC can be quite drastic.

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

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

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-based geostatistics (with discussion)," *Applied Statistics*, 47, 299–350.

Higgs, M. D. and Hoeting, J. A. (2010), "Clipped latent-variable model for spatially correlated ordered categorical data," *Computational Statistics and Data Analysis*, 54, 1999–2011.

Jin, C., Zhu, J., Steen-Adams, M., Sain, S., and Gangnon, R. (2013), "Spatial multinomial regression models for nominal categorical data: a study of land cover in Northern Wisconsin, USA," *Environmetrics*, 24, 98–108.

Li, X. (2006), "Bayesian Analysis of Cross-Classified Spatial Data with Autocorrelation," Ph.D. thesis, University of Wisconsin, Madison.

Lindgren, F., Rue, H., and Lindström, J. (2011), "An explicit link between Gaussian fields

and Gaussian Markov random fields: the stochastic partial differential equation approach," *Journal of the Royal Statistical Society: Series B*, 73, 423–498.

McCulloch, R. and Rossi, P. (1994), "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 64, 207–240.

Paciorek, C. and McLachlan, J. (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.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society: Series B*, 71, 319–392.