Spatial Regression Modeling for Compositional Data With Many Zeros

Thomas J. LEININGER, Alan E. GELFAND, Jenica M. ALLEN, and John A. SILANDER Jr.

Compositional data analysis considers vectors of nonnegative-valued variables subject to a unit-sum constraint. Our interest lies in spatial compositional data, in particular, land use/land cover (LULC) data in the northeastern United States. Here, the observations are vectors providing the proportions of LULC types observed in each 3 km × 3 km grid cell, yielding order 10^4 cells. On the same grid cells, we have an additional compositional dataset supplying forest fragmentation proportions. Potentially useful and available covariates include elevation range, road length, population, median household income, and housing levels.

We propose a spatial regression model that is also able to capture flexible dependence among the components of the observation vectors at each location as well as spatial dependence across the locations of the simplex-restricted measurements. A key issue is the high incidence of observed zero proportions for the LULC dataset, requiring incorporation of local point masses at 0. We build a hierarchical model prescribing a power scaling first stage and using latent variables at the second stage with spatial structure for these variables supplied through a multivariate CAR specification. Analyses for the LULC and forest fragmentation data illustrate the interpretation of the regression coefficients and the benefit of incorporating spatial smoothing.

Key Words: Areal data; Conditionally autoregressive model; Continuous ranked probability score; Hierarchical modeling; Markov chain Monte Carlo.

1. INTRODUCTION

Compositional data analysis is concerned with inference for data in the form of vectors of nonnegative observations that are subject to a constant-sum constraint (without loss of

Thomas J. Leininger is a Ph.D. Candidate (E-mail: *thomas.leininger@stat.duke.edu*) and Alan E. Gelfand (✉) is a Professor (E-mail: *alan@stat.duke.edu*), Department of Statistical Science, Duke University, Box 90251, Durham, NC 27708-0251, USA. Jenica M. Allen is an Assistant Professor in Residence (E-mail: *jenica.allen@uconn.edu*) and John A. Silander Jr. is a Professor (E-mail: *john.silander@uconn.edu*), Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Road, Unit 3043, Storrs, CT 06269, USA.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 18, Number 3, Pages 314–334
DOI: 10.1007/s13253-013-0145-y

generality, a unit-sum constraint). This defines, for a D-dimensional vector $\mathbf{Y}$, a sample space referred to as the simplex $\mathcal{S}^{D-1} = \{\mathbf{Y} : 0 \leq Y_k \leq 1, \ \mathbf{Y}^T\mathbf{1} = 1\}$ (Aitchison 1986). One setting where compositional data arises is in the analysis of proportions of land cover composing geographical regions. Each observation is a vector of proportions, with entries specifying the proportion of that region covered by a specific land cover category. Early work ignored these constraints and applied standard multivariate analysis. However, much effort has been devoted in the last 30 years to understanding the nature of such data and developing appropriate methods for their analysis, as discussed in, e.g., Aitchison and Egozcue (2005).

The contribution of this paper is to present a novel model specification for spatial regression on compositional data. The novelty addresses the issue that, for our observation vectors, components take on the value 0 sufficiently often to require point masses at 0. We view a 0 entry for a component of a vector at a location as arising at random, rather than as an *essential zero*, i.e., an entry that must be 0 for that component at that location. It is also possible that a zero value is *observed* due to rounding or to the true value being below some minimum detection level. Observed zeros arise frequently with nominal categorical data if we seek to avoid collapsing and aggregating of categories. Current methods lack ease of implementation and/or suitable interpretation when having to handle zero values, complicating regression and inference. A model that permits point masses to explain the zero values becomes useful in this context. Moreover, the compositional data vectors are each associated with a distinct areal unit and compositions in neighboring cells are expected to be similar. As a result, we first present the non-spatial case but readily move to the spatial setting through introduction of spatial random effects. In fact, with dynamic spatial compositional data, spatio-temporal random effects could also be introduced. The model is specified hierarchically, where fitting is done within a Bayesian framework.

One setting where compositional data arises is in the analysis of proportions of land cover types comprising geographical units. We illustrate the compositional data regression problem using data consisting of the proportions of a set of land use/land cover (LULC) classes at the scale of 3 km $\times$ 3 km grid cells over a region covering the northeastern United States, resulting in $n = 19{,}210$ cells. Individual proportions are often zero. For instance, many cells in rural areas exhibit a high concentration of forest but no urbanization. In fact, for the *Developed* classification in our dataset, zero values occur in 21 % of the cells. Potential explanatory variables for the regression modeling are drawn from US Census data along with measurements of road length and elevation change. The observations are also collected at four different times, providing the possibility of modeling change in land use/land cover over time. A second setting considered is that of forest fragmentation. Here, working again in the northeastern United States, grid cells describe a forest fragmentation process which yields the classifications: *Patch*, *Edge*, *Perforated*, or *Core* forest as well as *Other*. Potential explanatory variables for regression modeling include the same covariates as above.

2. PREVIOUS WORK

While the Dirichlet distribution might seem a natural choice for modeling data with a unit-sum constraint, the implicit negative pairwise correlations between *all* components can be overly restrictive. Additionally, the fact that the mean of a Dirichlet distribution determines the covariance structure is also restrictive. Compositional data are therefore often analyzed using transformations such as the additive logratio (alr) transformation (Aitchison 1986) or isometric logratio (ilr) transformation (Egozcue et al. 2003), whence standard multivariate methods can then be applied to the transformed data. Model fitting software for both transformations is available, e.g., using CoDaPack or the `compositions` package (van den Boogaart and Tolosana-Delgado 2008) in R (R Core Team 2012).

One shortcoming of these methods is that they do not allow zero values in the components. Handling rounded zeros can be done by assigning a pre-determined small value to the component or through an imputation algorithm (Fry, Fry, and McLaren 2000; Martín-Fernández, Barcelo-Vidal, and Pawlowsky-Glahn 2003). However, adding a small value seems arbitrary and may not be satisfying when there are numerous zero values.

Scealy and Welsh (2011), building on work by Stephens (1982), employ a square-root transformation to convert the compositional data to directional data, where the occurrence of zeros is no longer an issue. They then specify the Kent distribution (Kent 1982), a distribution defined on the hypersphere, to provide regression models. Butler and Glasbey (2009) handle zeros by introducing a latent Gaussian variable $\mathbf{Z}$, which maps the data from the simplex to the unit hyperplane $H_D = \{\mathbf{W} \in \mathcal{R}^D : \mathbf{W}^T\mathbf{1} = 1\}$. They use the transformation $\mathbf{Y} = g(\mathbf{W})$, which minimizes the squared Euclidean distance $(\mathbf{Y} - \mathbf{W})^T(\mathbf{Y} - \mathbf{W})$. The truncation imposed by their method enables a point mass at 0. However, accommodating $D > 3$ seems very challenging. Stewart and Field (2010) model zero values separately using mixture specifications. Conditional models are specified given the set of nonzero components and usually take the form of a multiplicative logistic (possibly skewed) normal distribution.

Salter-Townshend and Haslett (2006) suggest a one-stage approach representing components of $\mathbf{Y}$ in the form $Y_k = \max(0, Z_k)/(\sum_{k'=1}^{D} \max(0, Z_{k'}))$, where Z_k is a latent random variable that is presumed normally distributed. The function $\max(0, Z_k)$ takes the maximum between 0 and Z_k, which ensures that negative values of Y_k are not allowed by mapping negative values of Z_k to a point mass at $Y_k = 0$. The use of a Box–Cox transformation from the data $\mathbf{Y}$ to transformed data $\mathbf{Z}$ using $Z_k = ((Y_k/Y_D)^\gamma - 1)/\gamma$, for $k = 1, \ldots, D-1$, with a (possibly component-specific) scaling parameter γ is considered in Aitchison (1986), Fry, Fry, and McLaren (2000), and Tsagris, Preston, and Wood (2011). This approach requires that one component, taken here to be the last component, Y_D, is always nonzero so it can serve as a baseline. Our proposed method, building on these approaches, uses a simple transformation by employing a latent Gaussian variable, allowing direct incorporation of regression and spatial effects. We create a point mass at 0 similar to the Butler–Glasbey and Stewart–Field models, but through a comparison of the components to a baseline, similar to the alr transformation.

There has been some previous spatial compositional data work. Billheimer et al. (1997) studied the composition of benthic species across sample sites. Using the alr transforma-

tion to $\mathbf{Z} = alr(\mathbf{Y})$, they modeled $\mathbf{Z}$ with regression and random spatial effects, the latter being modeled with a multivariate spatial conditional autoregressive model (Mardia 1988). Tjelmeland and Lund (2003) extended the logistic normal distribution, incorporating Gaussian processes to model spatial structure. Similarly, Haslett et al. (2006) introduce a spatial process using a linear variogram approach, incorporating spatial effects in a two-stage model.

When working with data on, for example, a regular 3 km grid, one customarily adopts areal data spatial models. Conditionally autoregressive (CAR) models, dating to Besag (1974), are a common choice. Multivariate CAR (MCAR) specifications (Mardia 1988) are required for multivariate observations at areal units. Further development is provided in Gelfand and Vounatsou (2003). Coregionalizaton, that is, linear transformation of independent univariate CAR processes, is considered in Gelfand et al. (2004).

The format of the remainder of the paper is as follows. In Section 3 we introduce the compositional datasets used to investigate land use/land cover and forest fragmentation changes in the northeastern United States. In Section 4, we present a compositional model that allows point masses at zero and discuss properties of the model along with model fitting and inference. In Sections 5 and 6 we present analyses with simulated and real data to highlight the usefulness and performance of our proposed model. We offer a brief summary and possibilities for future work in Section 7.

3. DATASETS

With the goal of studying changes in land use and land cover in the northeastern United States, we analyze LULC data collected as part of the National Land Cover Dataset (NLCD) (Fry et al. 2009) and the Coastal Change Analysis Program (CCAP) (National Oceanic Atmospheric Administration 2006) from LANDSAT imagery at 30 meter resolution. Maps generated from these satellite images describe the usage type of each 30 meter pixel and provide a summary of the proportions of each LULC class within each of the 3 km $\times$ 3 km grid cells, altogether a total of 19,210 grid cells. There are 14 available LULC classes, which we collapsed into four broader classes, as shown in Table 1. (Still, many zero values are observed, as we quantify below.) The collapsing decisions were made in consultation with ecologists involved in this project.

Explaining the observed proportions of these resulting classes illuminates the LULC process. The data are available at four time periods: 1992, 1996, 2001, and 2005/2006 (hereafter called 2006). There was little observed change in the components over time; most of cells showed less than 5 % change from 1992 to 2006. Therefore, we do not consider dynamics in LULC and focus on the most recent time period.

The *Forest* classification is of primary interest, with goals of better understanding relationships with other classes and identifying factors that explain forest cover variation across the region. Forest fragmentation effects on invasive plant species is of special interest. Figure 1(a) shows, for each classification in the LULC data, the locations where there is an observed zero proportion. The observed incidence of 0's for each class is: *Developed*,

Table 1. New LULC classifications based on collapsing the original classes.

New Class	Developed	Crops and Grass	Forest	Other
Original Classes	High Intensity Developed Medium Intensity Developed Low Intensity Developed	Urban Grassland Pastures and Grassland Crops	Deciduous Forest Evergreen Forest Mixed Forest Woody Wetland	Scrub and Shrubland Open Water Emergent Wetland Barren Land

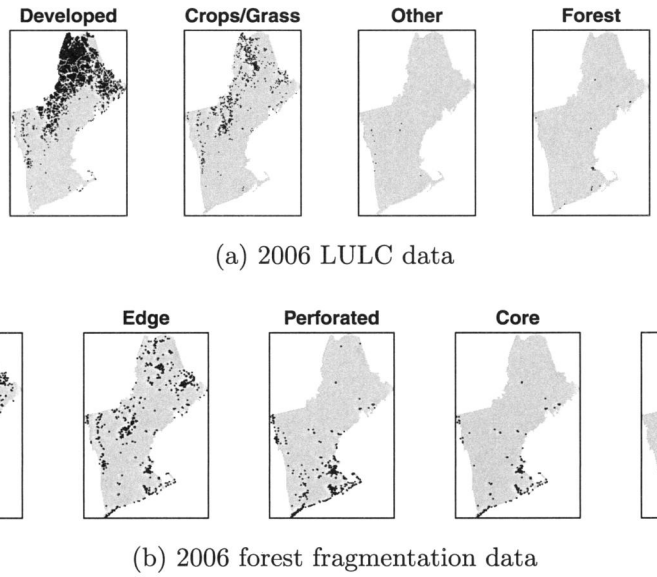


Figure 1. The black dots mark the incidence of a zero value in the (a) 2006 LULC data and (b) 2006 forest fragmentation data.

0.2102; *Crops/Grass*, 0.0433; *Other*, 0.0006; *Forest*, 0.0014. Figure 1(a) emphasizes the need for a model for LULC data that accommodates many 0's.

Using a GIS tool (Parent and Hurd 2010), the forest cover classes were divided into subclasses of core, patch, perforated, and edge forest. These subclasses address the issue of forest fragmentation. Core forest represents regions of forest that are a pre-specified distance from any non-forest regions. Edge forest represents parts of the forest surrounding the core forest. Patch forest represents small groups of forest that are too small to include any core forest. Perforated forest accounts for any forest cover surrounding a non-forest clearing in the midst of a larger forest region containing areas of core forest. Forest fragmentation occurs as the core forest class decreases and the remaining classes increase.

For the forest fragmentation data, the proportions of zeros are: *Patch*, 0.0369; *Edge*, 0.0340; *Perforated*, 0.0216; *Core*, 0.0095; *Other*, 0. Figure 1(b) shows the locations of the observed zero values for each classification. Evidently, incidence of 0's is less of an issue but our model can still be applied; our analysis can be compared with a customary alr or ilr specification.

For both datasets, the covariates considered in the regression model include measurements for each cell regarding change in elevation, road length, median household income, population, and a single-family housing ratio. Change in elevation measurements are drawn from the National Elevation Dataset (U.S. Geological Survey 1999) and are taken to be the difference between the maximum and minimum elevations within the cell. High values will indicate a large range of elevations and possibly imply a hilly location where wooded regions would be more likely. The road lengths variable records the sum of road lengths in each cell in 2008 (U.S. Census Bureau 2008); more roads will usually indicate higher incidence of developed regions.

The remaining covariates are taken from the 2000 US Census (most appropriate for examining compositional data for 2006), where the values for a grid cell were an areally weighted average from the possibly multiple census tracts observed in the cell (Minnesota Population Center 2004). The usual housing indicators in the census are highly correlated with population, so we employ the single-family housing ratio, the ratio of single-family housing to total housing in a region. A higher single-family housing ratio would imply that single-family homes are more common in the cell and possibly indicate a suburban or rural region. Highly developed regions with housing complexes and a dense population have a lower single-family housing ratio.

4. A POWER SCALING MODEL ALLOWING POINT MASSES AT ZERO

We first present the *local* model and its properties before discussing the full spatial specification. Hence, here and in Section 4.1, locations are suppressed. For any cell, we assume the compositional data vector, $\mathbf{Y}$, is generated from a latent multivariate Gaussian random variable, $\mathbf{Z}$, where the transformation from $\mathbf{Z}$ to $\mathbf{Y}$ sets the kth component $Y_k = 0$ when $Z_k \leq 0$ and $Y_k > 0$ when $Z_k > 0$. As in the Introduction, the primary assumption is that there exists one component that is nonzero for all observed Y_D and is used as a *baseline*. Without loss of generality, Y_D denotes the baseline component. With our land cover data, the *Forest* category is almost always nonzero. The 27 zero values in the *Forest* category out of 19,210 cells were set to the value 0.01 for computational stability. Thus, it seems appropriate to also set the nonzero values in the *Forest* component smaller than 0.01 to this value. Then, all observations were rescaled to sum to 1.

The general transformation is given by

$$Y_k = \frac{(\max(0, Z_k))^\gamma}{1 + \sum_{k'=1}^{d} (\max(0, Z_{k'}))^\gamma}, \quad k = 1, 2, \ldots, d, \tag{4.1}$$

with $d \equiv D - 1$, $\gamma > 0$, and $Y_D = (1 + \sum_{k'=1}^{d} (\max(0, Z_{k'}))^\gamma)^{-1}$. The corresponding inverse transformation is

$$\begin{aligned} Z_k &= (Y_k / Y_D)^{1/\gamma} && \text{if } Y_k > 0, \\ Z_k &\leq 0 \text{ (latent)} && \text{if } Y_k = 0, \end{aligned} \quad k = 1, 2, \ldots, d. \tag{4.2}$$

The vector $\mathbf{Z}$ is now modeled as a d-variate normal distribution where regression and spatial effects are easily incorporated, as described below. The latent Z_k where $Y_k = 0$ are

considered as censored data; hence, $\Pr(Y_k = 0) = \Pr(Z_k \leq 0)$. In MCMC model fitting, the Z_k can be sampled from a truncated multivariate normal distribution. It might be natural to allow γ to be unknown, assigning a prior to it and allowing the data to inform about it. However, model fitting is easier if γ is fixed, so we ran the model for several choices of γ and converted the selection of γ to a model choice problem. Perhaps $\gamma = 1$ would provide the simplest interpretation of the model but $\gamma = 0.5, 1, 2$, and 3 are considered below.

In this regard, with areal dependence in our model, internal validation for choosing γ is all that is available, since prediction with CAR models is incoherent (see Banerjee, Carlin, and Gelfand 2004, Chapter 5). We fitted the models for various values of γ and found $\gamma = 2$ to be an appropriate choice, in particular, preferable to $\gamma = 1$ under the criterion described in Section 4.4. For $\gamma = 2$ and 3, the effects of very large values in the non-baseline category are mitigated, making the model more robust to non-baseline components being close to one. This, of course, makes the model more robust to baseline values close to zero, a potential issue noted above. As a result, we prefer $\gamma = 2$ to $\gamma = 1$ and below argue that $\gamma = 3$ and $\gamma = 2$ are not much different, suggesting no reason to explore larger values.

4.1. PROPERTIES

Suppose we specify the mean of the kth component as $\mathbb{E}(Z_k) \equiv \mu_k = \mathbf{x}^T \boldsymbol{\beta}_k$ with a general covariance $\mathbf{V}$ for the vector $\mathbf{Z}$, i.e., $\mathbf{Z} \sim MVN_d(\mathbf{B}^T\mathbf{x}, \mathbf{V})$, where $\mathbf{B} = [\boldsymbol{\beta}_1 \ldots \boldsymbol{\beta}_d]$. This specification provides a simple calculation for the probability of a zero component: $\Pr(Y_k = 0) = \Pr(Z_k \leq 0) = \Phi(-\mathbf{x}^T \boldsymbol{\beta}_k / \sqrt{V_{kk}})$, where Φ denotes the cumulative distribution function of the standard normal distribution.

The regression relationship can be understood in terms of the change in the kth component *relative* to the baseline Dth component. Let $\Omega_\gamma = \mathbb{E}(Y_k / Y_D)$ for a given γ in (4.2). Then, $\Omega_1 = \mathbb{E}(\max(0, Z_k)) = \sqrt{V_{kk}}\varphi(-\mu_k/\sqrt{V_{kk}}) + \mu_k(1 - \Phi(-\mu_k/\sqrt{V_{kk}}))$, where φ denotes the probability density function of the standard normal distribution (see Appendix A.1). Similarly, $\Omega_2 = \mathbb{E}(\max(0, Z_k)^2) = (\mu_k + V_{kk})(1 - \Phi(\mu_k/\sqrt{V_{kk}}))$. In fact, Ω_γ is monotonic in μ_k for any $\gamma > 0$, as proven in Appendix A.2. In turn, this means that if, e.g., $\beta_{kl} > 0$ for covariate x_l, then μ_k will increase in x_l and hence, so will $\mathbb{E}(Y_k / Y_D)$. This does not, however, imply that the effect will yield monotonicity for $\mathbb{E}(Y_k)$ itself.

In fact, primary interest lies in the behavior of the posterior mean of Y_k, i.e., $\mathbb{E}(Y_k|\text{data}, \mathbf{x})$ as a function of the covariates. In order to assess the effect of a particular covariate x_l on the expected response of a particular component, Y_k, we would like to track the movement of the posterior mean over $[0, 1]$ as x_l is varied, holding the other covariates fixed (say, at their mean levels). Unfortunately, $\mathbb{E}(Y_k|\text{data}, \mathbf{x}) = \mathbb{E}(g_k(\mathbf{B}, \mathbf{V}, \mathbf{x})|\text{data}, \mathbf{x})$ involves an untractable function $g_k(\mathbf{B}, \mathbf{V}, \mathbf{x}) = \mathbb{E}(Y_k|\mathbf{B}, \mathbf{V}, \mathbf{x})$. However, with MCMC model fitting, the conditional distribution $[Y_k|\text{data}, \mathbf{x}]$ can be sampled by composition using posterior samples of $\mathbf{B}$ and $\mathbf{V}$ and making draws of $\mathbf{Y}$ from $[\mathbf{Y}|\mathbf{B}, \mathbf{V}, \mathbf{x}]$. These posterior samples will provide a Monte Carlo integration for $\mathbb{E}(Y_k|\text{data}, \mathbf{x})$ over a range of x_l's. The analyses in Section 6 supply such a posterior summary across components and covariates for the LULC and forest fragmentation datasets.

4.2. THE HIERARCHICAL MODEL AND PRIOR SPECIFICATIONS

With cells indexed by $i = 1, 2, \ldots, n$, in the non-spatial model we have conditionally independent $\mathbf{Y}_i$'s generated from latent $\mathbf{Z}_i$'s, with component k having mean $\mathbb{E}(Z_{ik}) = \mathbf{x}_i^T \boldsymbol{\beta}_k$. To introduce spatial dependence, $\mathbb{E}(Z_{ik})$ is revised to $\mathbb{E}(Z_{ik}) = \mathbf{x}_i^T \boldsymbol{\beta}_k + \phi_{ik}$. Here, $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2}, \ldots, \phi_{id})^T$ is a vector of spatial random effects for cell i and $\{\boldsymbol{\phi}_i\}_{i=1}^n$ follow a d-dimensional multivariate CAR model specified explicitly below. The regression coefficients are fitted in the presence of the spatial random effects but the resulting interpretation with regard to regression relationships is as in the non-spatial case. In our first example, *Forest* is used as the baseline variable, which allows us to capture how changes in covariates affect the proportions of the other components relative to the *Forest* category. We can also implement the covariate analysis proposed above for $\mathbb{E}(Y_k|\text{data}, \mathbf{x})$ in the spatial case, setting the spatial effects to 0.

The Bayesian model is fully specified after including the prior distributions for the parameters, as shown below in (4.3). The notation $j \sim i$ means that cell j is a contiguous neighbor of cell i and n_i is the number of neighbors that cell i has. The number of covariates is denoted by p and the number of grid cells is n. The full multi-level model is:

$$\begin{aligned}
\mathbf{Y}_i|\mathbf{Z}_i &= g(\mathbf{Z}_i) \text{ as in (4.1)}, \quad \text{for } i = 1, \ldots, n \\
\mathbf{Z}_i &\overset{\text{ind}}{\sim} \text{Normal}_d\left(\mathbf{B}^T \mathbf{x}_i + \boldsymbol{\phi}_i, \mathbf{V}\right) \\
vec(\mathbf{B}) &\sim \text{Normal}_{d(p+1)}\left(vec\left([\mu \mathbf{1}_{d\times 1} \quad \mathbf{0}_{d\times p}]^T\right), \lambda \mathbf{I}_{d(p+1)}\right) \\
\boldsymbol{\phi}_i|\{\boldsymbol{\phi}_j\}_{j\neq i} &\sim \text{Normal}_d\left(\frac{1}{n_i}\sum_{j\sim i}\boldsymbol{\phi}_j, \frac{1}{n_i}\boldsymbol{\Sigma}\right) \\
\mathbf{V} &\sim \text{Inverse Wishart}_d(\mathbf{M}_V, m_V) \\
\boldsymbol{\Sigma} &\sim \text{Inverse Wishart}_d(\mathbf{M}_\Sigma, m_\Sigma)
\end{aligned} \tag{4.3}$$

We place independent normal priors on each $\beta_{kl} \equiv (\mathbf{B})_{lk} = (\mathbf{B}^T)_{kl}$ with a large variance and $\mathbb{E}(\beta_{kl}) = 0$ for $l > 0$. However, setting $\mathbb{E}(\beta_{k0}) = 0$ is not sensible. More precisely, let $U_k = \max(0, Z_k)^2$ and suppose, a priori, we think that μ_k and V_{kk} are constant across k. Then, all U_k have the same marginal distribution, and therefore all Y_k have the same distribution. Now, let $U = \sum_{k=1}^d U_k$. Then, $\sum_{k=1}^d Y_k = U/(1+U)$ and $Y_D = 1/(1+U)$ so $\mathbb{E}(Y_k) = (1/(D-1))\,\mathbb{E}(U/(1+U)) = (1/(D-1))(1-\mathbb{E}(Y_D))$, i.e., we have obtained the common $\mathbb{E}(Y_k)$ in terms of $\mathbb{E}(Y_D)$. Therefore, $\mathbb{E}(Y_k) = 1/D$ if and only if $\mathbb{E}(Y_D) = 1/D$, i.e., if and only if $\mathbb{E}(U/(1+U)) = (D-1)/D$. Hence, there is a value of $\mu_k = \mu$ such that this holds and this value could be obtained numerically, given D. However, for simplicity, in the sequel, we set $\mathbb{E}(\beta_{k0}) = 1$. Centering at 1 for the intercepts roughly corresponds to centering $\mathbf{Y}$, a priori, in the middle of the simplex. That is, if the prior mean for $\mathbf{Z}$ is $\mathbb{E}(\mathbf{Z}) = \mathbf{1}_{d\times 1}$, a vector of ones, then, from (4.1), $\mathbb{E}(\mathbf{Y})$ is approximately centered at $(1/D)\mathbf{1}_{D\times 1}$. We make the prior for $\mathbf{V}$ vague so we anticipate little sensitivity to this prior centering.

As noted previously, we model the spatial effects $\{\boldsymbol{\phi}_i\}$ using an intrinsic multivariate CAR model MCAR(1, $\boldsymbol{\Sigma}$), as in Gelfand and Vounatsou (2003). This corresponds to the

prior given above where $\mathbb{E}(\boldsymbol{\phi}_i|\{\boldsymbol{\phi}_j\}_{j\neq i})$ is the average of its neighbors. Alternatively, we could use a coregionalized model $\boldsymbol{\phi}_i = \mathbf{A}\mathbf{u}_i$, where $\mathbf{A}$ is a lower diagonal matrix and $\mathbf{u}_i$ is a vector of independent univariate intrinsic CAR $(1, \sigma^2)$ models (Gelfand et al. 2004). Specification to make these priors equivalent can be provided. The usual benefits of coregionalization include univariate updates of each univariate CAR process. However, we do not gain that advantage here because of the nonzero off-diagonals in the covariance matrix $\mathbf{V}$. To identify the intrinsic specification of the MCAR model, we assume that each dimension of the MCAR process sums to zero, i.e., $\sum_{i=1}^{n} \phi_{ik} = 0$ for each $k = 1, \ldots, d$. This is implemented computationally by centering each dimension on the fly, as in Besag et al. (1995).

4.3. Model Fitting

With the prior specification complete, posterior computation is done by Gibbs sampling. The full conditional distributions are presented in Appendix B. The regression coefficients $\mathbf{B}$ are updated simultaneously, using the fact that the prior and full conditional can be written as a matrix Normal distribution. Assuming that $\mathbf{X}$ has been centered and scaled, we set $\lambda = 10$ in the covariance for $\mathbf{B}$, a weak specification since prior predictive simulations show that effects for a scaled covariate will rarely be larger than 1. $\mathbf{V}$ and $\boldsymbol{\Sigma}$ are updated from Inverse Wishart distributions, with scale matrices $\mathbf{M}_V = \mathbf{M}_\Sigma = 2I_d$ and degrees of freedom $m_V = m_\Sigma = d+3$. This specification provides $\mathbb{E}(\mathbf{V}) = \mathbb{E}(\boldsymbol{\Sigma}) = \mathbf{I}_d$, which implies a priori independence among the components.

The MCAR prior for the spatial effects provides local updating in the MCMC, i.e., the vector $\boldsymbol{\phi}_i$, consisting of d random spatial effects, is updated as a unit. Though this is straightforward, the need to update one location at a time still leads to a bottleneck with regard to computational run time. However, due to the local structure of the MCAR model, it is possible to usefully parallelize the updating process by dividing the region into blocks and updating within the blocks in parallel, followed by updating the locations on the borders of the blocks (Chakraborty et al. 2010). The number of blocks must be chosen to balance the speed gains by adding more blocks with the speed costs due to increased overhead of more blocks. For our analysis, with over 19,000 locations, the ability to update the spatial effects in parallel blocks provides significant speed gains; we achieved good efficiency with four blocks. Furthermore, the covariance matrix for each of these local updates is just a function of the number of neighbors, which implies that there are only a few unique combinations to be calculated each MCMC iteration.

We note that the negative, latent $\mathbf{Z}_i$'s can also be updated in parallel fashion due to the conditional independence of the $\mathbf{Z}_i$'s. That is, a latent component of $\mathbf{Z}_i$, corresponding to a zero $\mathbf{Y}_i$ component, needs to be imputed through Gibbs sampling in order to enable sampling of the model parameters and the spatial effects. When a large number of zero components are observed, the need to sample the latent $\mathbf{Z}_i$'s can significantly decrease the efficiency of the computation unless parallelization is introduced. Fortunately, the parallelization scheme for the $\mathbf{Z}_i$'s is even simpler to implement than that for the $\boldsymbol{\phi}_i$'s.

Table 2. The CRPS values when fitting the 2006 LULC data with the specified γ values.

γ	Developed	Crops/Grass	Other	Forest	Average
0.5	0.071	0.071	0.165	0.458	0.191
1	0.065	0.042	0.040	0.137	0.071
2	0.010	0.008	0.013	0.012	0.013
3	0.006	0.007	0.011	0.015	0.010

4.4. CHOOSING γ

Choosing a value for γ is a model selection issue which we addressed by fitting the 2006 LULC data for several values of γ. Given that the predicted values contain point masses at zero proportions, we used the continuous ranked probability score (CRPS) (Unger 1985) to provide comparison between observed values and the sampled posterior predictive values. The CRPS is a proper scoring rule (Gneiting and Raftery 2007) which compares the posterior predictive CDF F to the empirical (degenerate) CDF for an observed value y using

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} \left[F(x) - \mathbf{1}(x \geq y)\right]^2 dx, \tag{4.4}$$

where $\mathbf{1}(\cdot)$ is the indicator function.

The CRPS is especially useful in our context because the comparison of distribution functions seamlessly integrates the point masses at zero and values in the unit interval, which will arise from predictions using our compositional data model. For an observed value $y = 0$, the CRPS will prefer the point mass over any predictions in the unit interval and will prefer small nonzero predictions to large ones. Roughly independent posterior parameter samples provide roughly independent predictive draws $\tilde{y}_{ik}^{(t)}$ associated with each observation y_{ik}, where $\tilde{y}_{ik}^{(t)}$ denotes a predicted value for y_{ik} using the tth set of posterior samples obtained via MCMC. Using these predictive draws and the data, the computational form given in Gneiting and Raftery (2007) computes the average CRPS as

$$\text{CRPS} = \frac{1}{nDT} \sum_{i=1}^{n} \sum_{k=1}^{D} \sum_{t=1}^{T} |y_{ik} - \tilde{y}_{ik}^{(t)}| - \frac{1}{2nDT^2} \sum_{i=1}^{n} \sum_{k=1}^{D} \sum_{t=1}^{T} \sum_{s=1}^{T} |\tilde{y}_{ik}^{(t)} - \tilde{y}_{ik}^{(s)}|. \tag{4.5}$$

We fit the 2006 LULC data for $\gamma = 0.5, 1, 2$, and 3 using the spatial version of our model and compared the CRPS values for each case. The CRPS provides no penalty for model complexity, but since these four models all have the same dimension, the comparison is fair. The results shown in Table 2 reveal that as γ increases, the CRPS values get smaller, indicating better fit. We chose to use $\gamma = 2$ in our subsequent analyses in light of the convenient analytical calculations earlier in this section as well as the fact that it provided a large gain over using $\gamma = 1$ with a relatively much smaller additional gain for using $\gamma = 3$ (of course, the corresponding analysis can be provided for $\gamma = 3$).

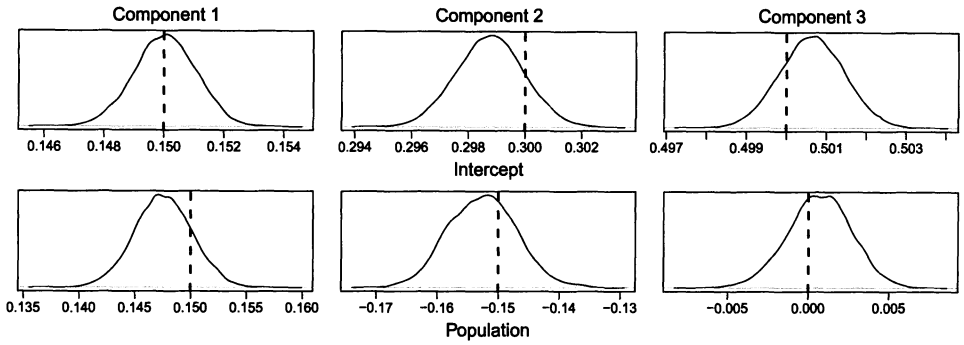


Figure 2. Posterior density plots for regression coefficients of simulated data, with the intercept coefficients on the top row and the population coefficients on the bottom row. The dashed lines indicate the known values.

5. SIMULATION EXAMPLE

Before presenting the data analysis, we offer some simulation results to assess the performance of our model in parameter estimation. To illustrate the performance of our model, we use the same 19,210 locations as in the LULC example in Section 6.1 and use population as a covariate. We chose to have four components in the response ($D = 4$). The parameter values considered are

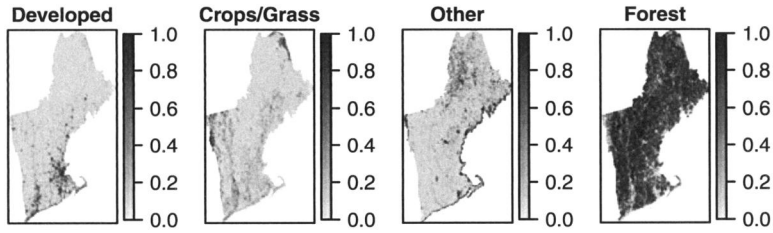
$$\mathbf{B}^T = \begin{bmatrix} 0.15 & 0.15 \\ 0.3 & -0.15 \\ 0.5 & 0 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} 0.01 & 0 & -0.005 \\ 0 & 0.02 & 0.005 \\ -0.005 & 0.005 & 0.01 \end{bmatrix}, \quad \text{and}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.2 & 0.02 & -0.05 \\ 0.02 & 0.1 & 0.05 \\ -0.05 & 0.05 & 0.1 \end{bmatrix}.$$

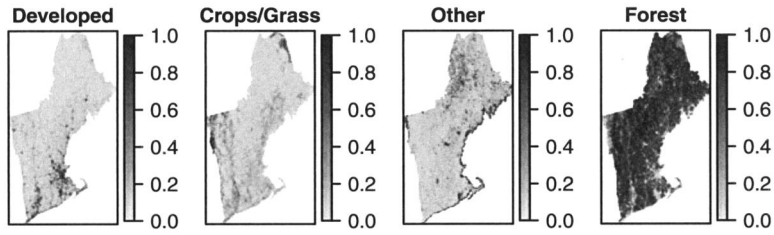
Since the multivariate CAR prior is improper, for simplicity, the $\{\boldsymbol{\phi}_i\}$ are generated using the known value of $\boldsymbol{\Sigma}$ and looping through a few iterations of Gibbs sampling updates from the prior. The proportions of zero values for the resulting data are: Y_1, 0.2820; Y_2, 0.0859; Y_3, 0.0026; Y_4, 0. We fit the spatial version of our model, collecting 10,000 MCMC samples after a burn-in of 5,000 samples. Figure 2 shows that the model was able to recover the regression parameters fairly well.

6. DATA ANALYSES

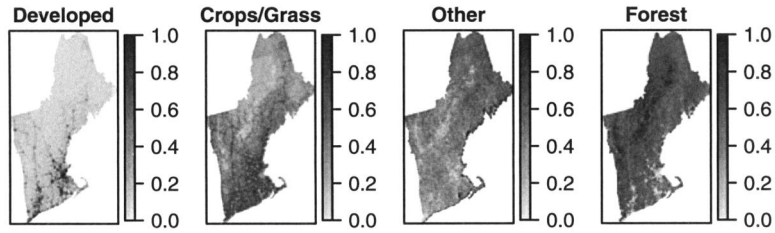
We now consider the two earlier datasets. Again, we use the LULC data in 2006, with high incidence of zeros, as well as the forest fragmentation data, where the low incidence of zeros suggests that we can compare the performance of our model to that of the alr model. In both cases, we use the covariates described in Section 3 to explain changes in the composition of the responses at each location. The models fitted in each analysis ran for 50,000 iterations after a 5,000 iteration burn-in and we retained every fifth sample,



(a) 2006 LULC data



(b) Posterior predictive means from the spatial model



(c) Posterior predictive means from the non-spatial model

Figure 3. Data and posterior predictive means for the 2006 LULC data example.

resulting in 10,000 posterior samples. MCMC convergence was monitored using standard algorithms provided by the `coda` package in R (Plummer et al. 2006).

6.1. LAND USAGE/LAND COVER ANALYSIS

For the 2006 LULC data, we fit spatial and non-spatial (e.g., set each $\boldsymbol{\phi}_i = \mathbf{0}$) versions of our model. The classes used are those given in Table 1 with the *Forest* component being used as the baseline. The region is mostly forest, as shown in Figure 3a, with many zeros in the *Developed* category as previously noted in Figure 1.

Figure 3 shows the fit of the posterior predictive means to the LULC data for both the spatial and non-spatial models. We note that the spatial model yields a very similar picture when compared with the actual data, capturing the local variation in each of the components while introducing a small amount of smoothing. The non-spatial version, however, only seems to perform well for the *Developed* category, failing to capture much of the detail in the remaining components. Though including additional covariates might help to

Table 3. Comparison of CRPS for non-spatial and spatial versions of our model on the 2006 LULC data.

Model	Developed	Crops/Grass	Other	Forest	Average
Non-spatial	0.023	0.042	0.095	0.119	0.070
Spatial	0.010	0.008	0.014	0.020	0.013

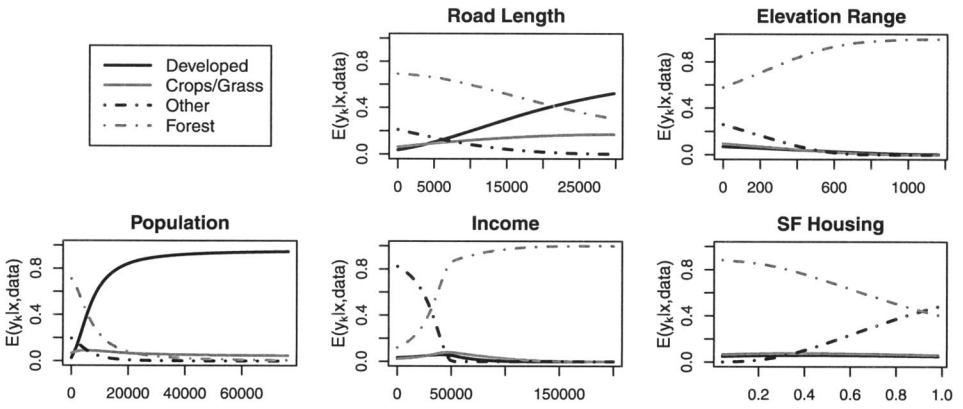


Figure 4. Posterior mean plots for the spatial model in the 2006 LULC data example. The plots show the effect of each covariate on the posterior mean of each component, holding the other covariates at their means and assuming no spatial effects.

ameliorate this issue, the spatial random effects are evidently successful in capturing the local dependence structure.

We then compared the models with and without spatial random effects using continuous ranked probability scores as a rough guide for model comparison, though the issue with model dimension is at hand here. The CRPS for each model is shown in Table 3, indicating a very strong preference for the spatial version of the model. In fact, in terms of CRPS, the spatial model was, on average, more than five times closer to the observed values than the non-spatial version.

Illustration of the effects of the regression coefficients is given in Figure 4, where the change in the posterior mean for each component can be seen as a function of the change in the covariate level. These plots show $\mathbb{E}(Y_k|\text{data}, \mathbf{x})$, the posterior mean for a component Y_k given the data and a covariate value x_l, changing one covariate at a time while holding the other covariates at their mean. We see how changing each covariate affects the posterior mean for each component, taking into account the indirect effects of the regression coefficients, the intercepts for each component, and the correlation structure in $\mathbf{V}$. These plots assume no spatial effect is present; adding a spatial effect will modify the local regression intercepts.

Figure 4 shows that road length, population, and single family housing had strong negative correlations with the amount of forest cover. These findings seem intuitive, reflecting the general effect of people and development on loss of forest. Change in elevation and income had strong positive relationships with forest cover.

Table 4. Posterior means and 95 % credible intervals for the regression coefficients for the spatial model using the 2006 LULC data. The *Forest* category is used as the baseline.

Covariate	Developed	Crops/Grass	Other
Intercept	0.277 (0.274, 0.280)	0.326 (0.325, 0.327)	0.513 (0.512, 0.515)
Road Length	0.114 (0.109, 0.120)	0.047 (0.043, 0.050)	−0.057 (−0.067, −0.047)
Elevation Range	−0.047 (−0.056, −0.038)	−0.059 (−0.065, −0.054)	−0.129 (−0.145, −0.113)
Population	0.435 (0.427, 0.444)	0.087 (0.080, 0.093)	−0.030 (−0.046, −0.013)
Income	−0.098 (−0.111, −0.084)	−0.050 (−0.060, −0.041)	−0.874 (−0.900, −0.844)
SF Housing	0.026 (0.011, 0.041)	0.023 (0.013, 0.033)	0.245 (0.214, 0.275)

The posterior means and 95 % credible intervals for the regression coefficients in the spatial model are given in Table 4. The signs and magnitudes of the coefficient estimates generally agree with the relationships shown in Figure 4. The coefficients with smaller magnitudes in Table 4, such as the income coefficient for the *Developed* component, may not have a clear, monotonic relationship in Figure 4. The larger magnitudes in Table 4, such as the population coefficient for the *Developed* component, show a corresponding significant trend in Figure 4.

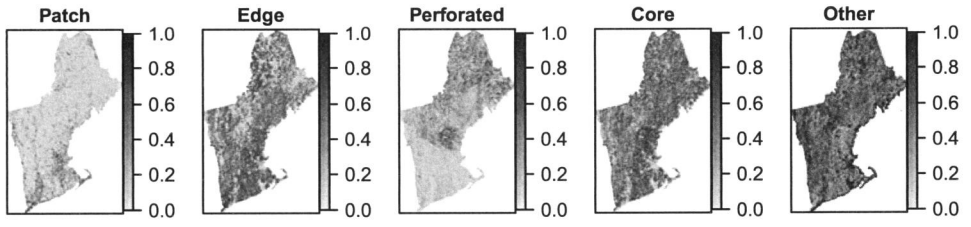
6.2. FOREST FRAGMENTATION ANALYSIS

We now investigate changes in the specific components of forest cover. We use the same covariates and locations as above, but now have a five-component response at each location covering the four forest-type components and one non-forest component in each cell, labeled as the *Other* category. The *Other* category is used as the baseline here, since it has no zero values.

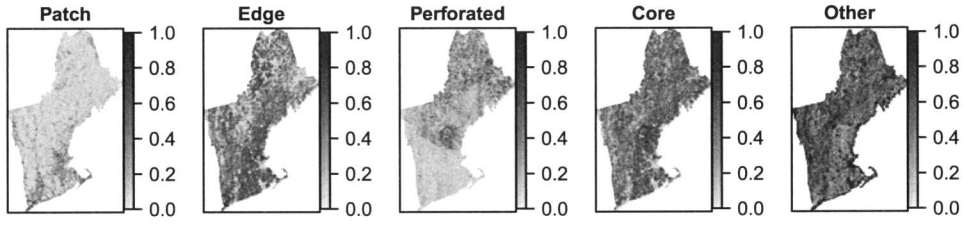
Again, we fit a spatial and non-spatial version of our model, but we now compare the results to the alternative of using the alr transformation on the data. For the alr, we assign a small value, 0.01, where a zero occurs, and rescale the remaining nonzero components as in the multiplicative strategy of Martín-Fernández, Barcelo-Vidal, and Pawlowsky-Glahn (2003). This imputed dataset can now be used to fit the alr model, though we use the original, non-imputed dataset for model comparison via CRPS. Figure 5 shows the data and the posterior predictive means for the 2006 forest fragmentation data for the spatial and non-spatial versions of our model. Again, the spatial version strongly outperforms the non-spatial version in terms of model fit.

We use CRPS to compare the spatial and non-spatial models, also adding the CRPS for the alr transformation with spatial effects. From Table 4, again, the non-spatial version of our model had a much higher CRPS, roughly three times larger than the spatial version of our model. We see, in Table 5, that our model seems to do a bit worse than the alr model for components such as *Core* forest and *Other* where there were few or no zeros. Since there were not many zeros present in the data, perhaps the recommendation would be to continue to use the alr model in such cases, rather than our purpose-built model.

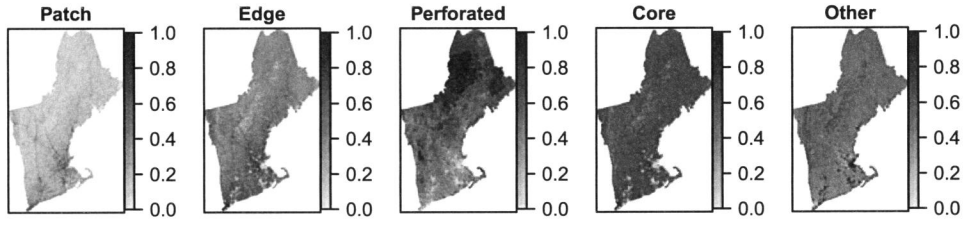
Figure 6 shows the posterior mean plots, as explained before, where the effect of each covariate on the expected mean of the forest components is illustrated. Population and elevation range have negative relationships with *Core* forest, while income shows a strong



(a) 2006 forest fragmentation data



(b) Posterior predictive means from our spatial model



(c) Posterior predictive means from our non-spatial model

Figure 5. Data and posterior predictive means for 2006 forest fragmentation data.

Table 5. Comparison of CRPS for non-spatial and spatial versions of our point mass model (PM) and the alr model using the 2006 forest fragmentation data.

Model	Spatial effects	Patch	Edge	Perforated	Core	Other	Average
PM	Non-spatial	0.010	0.023	0.015	0.056	0.085	0.038
PM	Spatial	0.004	0.007	0.005	0.020	0.022	0.012
alr	Spatial	0.006	0.007	0.004	0.012	0.011	0.008

positive relationship with *Core* forest. The posterior estimates for the regression coefficients (not shown) agree with the general trends seen in Figure 6 for the forest fragmentation data.

6.3. Sensitivity to Grid Cell Resolution

Finally, we compare the results at different resolutions. We present the posterior predictive means using our spatial model with the 2006 LULC data at 3 km and at 6 km resolution. The 6 km resolution dataset contains 4,614 observations, compared to the 19,210

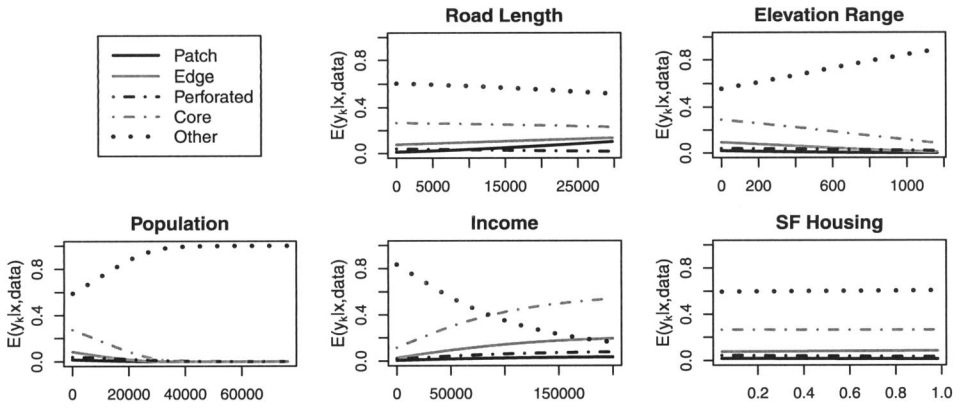


Figure 6. Posterior mean plots for the spatial version of our model on 2006 forest fragmentation data.

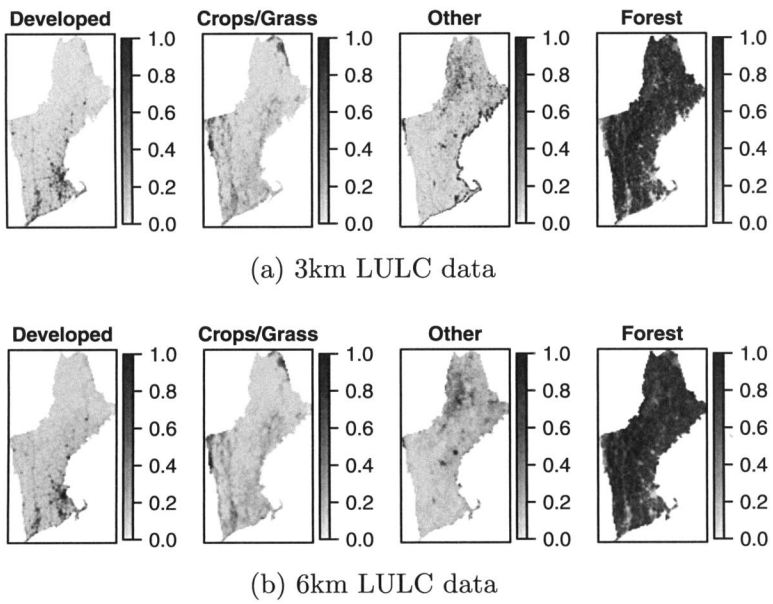


Figure 7. Comparison of posterior predictive means for 3 km and 6 km 2006 LULC data using the spatial version of our model.

observations in the 3 km dataset, and the developed class still has a large incidence of observed zero proportions.

The averaging to achieve the lower resolution implies less local detail in the 6 km dataset. In Figure 7, we see that very similar predicted values are obtained for both the 3 km and 6 km datasets. However, for the latter, the resultant smoothing succeeds in losing many of the high values in the *Other* category along the eastern coast. The fitted regression coefficients tell the same story at both scales (not shown). The variance components, i.e., the entries in $\mathbf{V}$, tend to be a bit larger at the 3 km scale, reflecting the increased detail, hence increased component variability, at the higher resolution (again, not shown). In sum-

mary, the 3 km and 6 km grid cells are artificial partitions of a continuous landscape but we find consistent inference results between the resolutions.

7. SUMMARY AND FUTURE DIRECTIONS

We have provided a model which allows zeros in the components through a latent variable modification of the basic alr transformation. By specifying a nonzero baseline component, our model allows a transformation to a lower dimensional space where we can easily handle regression and the inclusion of spatial random effects. Moreover, the resulting models can be fitted efficiently, providing ready interpretation. Illustration with two datasets shows that, with a substantial incidence of observed 0 components over the spatial region, our model performs very well. However, with very few zeros, an alr model, replacing the zero values with a small value, performs at least as well. The noteworthy limitation of our model is the need to specify a baseline classification, a category that has no 0's over the region; alr models suffer the same limitation. One can envision spatial datasets where such a category is not available across the region.

Extension to include temporal structure is straightforward. For example, the data we analyzed are available at the four different time points, so, in principle, we could introduce temporal dependence into the model, perhaps even to forecast estimates of future land cover patterns. However, as there was little change evident in the data, we have not investigated this. Another path to explore would be joint modeling of LULC and forest fragmentation in order to see if joint analysis benefits our explanatory story.

One issue we have not addressed here is that of potential spatial confounding between the fixed effects in the regression and random effects in the multivariate CAR model. The results presented here exhibited some spatial confounding when comparing the spatial and non-spatial models. The results were similar for both versions, though the regression effects in the spatial model were generally stronger than in the non-spatial model, making the effects in Figure 4 more dramatic than those in the non-spatial model. In one case, for the population effect on the *Other* component in the LULC analysis, the regression coefficient switched signs when removing the spatial effects.

Hence, future work to address spatial confounding might prove beneficial. The papers of Reich, Hodges, and Zadnik (2006) and Hughes and Haran (2013) provide methods for alleviating this problem, though these approaches would be more challenging for multivariate spatial random effects.

ACKNOWLEDGEMENTS

The authors thank Daniel Civco and James Hurd for acquiring and processing the data and Adam Wilson and Corey Merow for useful conversations. This work was supported in part by USDA-NRI 2008-003237.

APPENDIX A: CALCULATIONS USING Ω_γ

Without loss of generality, assume Z_k has mean μ and variance v and let ϕ and Φ denote the probability density and distribution functions of the standard normal distribution.

A.1 Value of Ω_1

We calculate, letting $\gamma = 1$, the value of Ω_1 as

$$\begin{aligned}
\Omega_1 &= \mathbb{E}\max(0, Z_k) \\
&= \int_{-\infty}^{\infty} \max(0, z)\, p_{Z_k}(z)\, dz = \int_0^{\infty} z\, p_{Z_k}(z)\, dz \\
&= \int_0^{\infty} z \frac{1}{\sqrt{2\pi v}} \exp\left\{\frac{-(z-\mu)^2}{2v}\right\} dz \\
&= \int_0^{\infty} (z-\mu) \frac{1}{\sqrt{2\pi v}} \exp\left\{\frac{-(z-\mu)^2}{2v}\right\} dz + \int_0^{\infty} \mu \frac{1}{\sqrt{2\pi v}} \exp\left\{\frac{-(z-\mu)^2}{2v}\right\} dz \\
&= \left(\frac{-\sqrt{v}}{\sqrt{2\pi}} \exp\left\{\frac{-(z-\mu)^2}{2v}\right\}\right)\Bigg|_0^{\infty} + \mu\big(1 - \Phi(-\mu/\sqrt{v})\big) \\
&= \sqrt{v}\phi(-\mu/\sqrt{v}) + \mu\big(1 - \Phi(-\mu/\sqrt{v})\big).
\end{aligned}$$

The proof for Ω_2 is similar.

A.2 Proof of monotonicity of Ω_γ

Start by modifying the notation to write $\Omega_\gamma(\mu, v)$ as an explicit function of μ and v. Without loss of generality, consider increasing μ to $\mu + \delta$ for some $\delta > 0$. We see that

$$\begin{aligned}
\Omega_\gamma(\mu+\delta, v) &= \int_0^{\infty} z \frac{1}{\sqrt{2\pi v}} \exp\{-(z-\mu-\delta)^2/2v\}\, dz \\
&= \int_{-\delta}^{\infty} (y+\delta) \frac{1}{\sqrt{2\pi v}} \exp\{-(y-\mu)^2/2v\}\, dy \\
&= \underbrace{\int_{-\delta}^{0} (y+\delta) \frac{1}{\sqrt{2\pi v}} \exp\{-(y-\mu)^2/2v\}\, dy}_{>0} \\
&\quad + \underbrace{\int_0^{\infty} (y+\delta) \frac{1}{\sqrt{2\pi v}} \exp\{-(y-\mu)^2/2v\}\, dy}_{>\,\Omega_\gamma(\mu, v) \text{ because } (y+\delta) > y} \\
&> \Omega_\gamma(\mu, v).
\end{aligned}$$

Hence Ω_γ is a monotonically increasing function in μ_k for arbitrary γ.

APPENDIX B: FULL CONDITIONALS FOR SPATIAL REGRESSION MODEL

$$[vec(\mathbf{B})|-] \sim \text{Normal}\left(\boldsymbol{\Omega}(\mathbf{V}^{-1} \otimes \mathbf{X}^T\mathbf{X})(\mathbf{X}^T\mathbf{X})^{-1}\sum_i \mathbf{x}_i^T(\mathbf{Z}_i - \boldsymbol{\phi}_i),\ \boldsymbol{\Omega}\right)$$

$$\text{where } \boldsymbol{\Omega} = \left(\lambda^{-1}\mathbf{I}_{d\times p} + \mathbf{V}^{-1} \otimes \mathbf{X}^T\mathbf{X}\right)^{-1}$$

$$[\mathbf{V}|-] \sim \text{Inverse Wishart}\left(m_V + n,\ \mathbf{M}_V + \sum_i (\mathbf{Z}_i - \mathbf{B}^T\mathbf{x}_i - \boldsymbol{\phi}_i)(\mathbf{Z}_i - \mathbf{B}^T\mathbf{x}_i - \boldsymbol{\phi}_i)^T\right)$$

$$[\boldsymbol{\phi}_i|-] \sim \text{Normal}_d\left(\left(\mathbf{V}^{-1} + n_i\boldsymbol{\Sigma}^{-1}\right)^{-1}\left(\mathbf{V}^{-1}(\mathbf{Z}_i - \mathbf{B}^T\mathbf{x}_i) + \boldsymbol{\Sigma}^{-1}\left(\sum_{i\sim j}\boldsymbol{\phi}_j\right)\right),\right.$$

$$\left.\left(\mathbf{V}^{-1} + n_i\boldsymbol{\Sigma}^{-1}\right)^{-1}\right)$$

$$[\boldsymbol{\Sigma}|-] \sim \text{Inverse Wishart}\left(m_\Sigma + n,\ \mathbf{M}_\Sigma + \sum_i\sum_j (\mathbf{D}_W - \mathbf{W})_{ij}\boldsymbol{\phi}_i\boldsymbol{\phi}_j^T\right)$$

where $\mathbf{D}_W$ is a diagonal matrix with elements $(\mathbf{D}_W)_{ii} = n_i$

and $\mathbf{W}$ is the spatial adjacency matrix with $W_{ij} = 1$ if $i \sim j$ and 0 otherwise

Sample code can be found at *http://stat.duke.edu/~tjl13/research.html.*

[Published Online May 2013.]

REFERENCES

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, New York: Chapman and Hall.

Aitchison, J., and Egozcue, J. J. (2005), "Compositional Data Analysis: Where Are We and Where Should We Be Heading?" *Mathematical Geology*, 37, 829–850.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman and Hall/CRC Press.

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

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems," *Statistical Science*, 10, 3–66.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H.-W., and Silkey, M. (1997), "Natural Variability of Benthic Species Composition in the Delaware Bay," *Environmental and Ecological Statistics*, 4, 95–115.

Butler, A., and Glasbey, C. (2009), "Corrigendum: A Latent Gaussian Model for Compositional Data With Zeros," *Journal of the Royal Statistical Society. Series C. Applied Statistics*, 58, 141.

Chakraborty, A., Gelfand, A., Wilson, A. M., Latimer, A. M., and Silander, J. A. (2010), "Modeling Large Scale Species Abundance With Latent Spatial Processes," *The Annals of Applied Statistics*, 4, 1403–1429.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003), "Isometric Logratio Transformations for Compositional Data Analysis," *Mathematical Geology*, 35, 279–300.

Fry, J., Fry, T., and McLaren, K. (2000), "Compositional Data Analysis and Zeros in Micro Data," *Applied Economics*, 32, 953–959.

Fry, J. A., Coan, M. J., Homer, C. G., Meyer, D. K., and Wickham, J. (2009), "Completion of the National Land Cover Database (NLCD) 1992–2001 Land Cover Change Retrofit Product," U.S. Geological Survey Open-File Report 2008–1379, 18 p.

Gelfand, A. E., and Vounatsou, P. (2003), "Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis," *Biostatistics*, 4, 11–25.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. F. (2004), "Nonstationary Multivariate Process Modelling Through Spatially Varying Coregionalization" (with discussion), *Test*, 13, 1–50.

Gneiting, T., and Raftery, A. E. (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J. R. M., Huntley, B., and Mitchell, F. J. G. (2006), "Bayesian Palaeoclimate Reconstruction," *Journal of the Royal Statistical Society. Series A. Statistics in Society*, 169, 395–438.

Hughes, J., and Haran, M. (2013), "Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 75, 139–159.

Kent, J. T. (1982), "The Fisher-Bingham Distribution on the Sphere," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 44, 71–80.

Mardia, K. V. (1988), "Multi-dimensional Multivariate Gaussian Markov Random Fields With Application to Image Processing," *Journal of Multivariate Analysis*, 284, 265–284.

Martín-Fernández, J. A., Barcelo-Vidal, C., and Pawlowsky-Glahn, V. (2003), "Dealing With Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation," *Mathematical Geology*, 35, 253–278.

Minnesota Population Center (2004), "National Historical Geographic Information System: Pre-release Version, 0.1," University of Minnesota, Minneapolis, MN, available at: *http://www.nhgis.org/*.

National Oceanic Atmospheric Administration (2006), "Coastal Change Analysis Program Land Cover," available at: *http://www.csc.noaa.gov/crs/lca/northeast.html*.

Parent, J., and Hurd, J. (2010), "Landscape Fragmentation Tool (LFT v2.0)." Center for Land Use Education and Research, available at: *http://clear.uconn.edu/tools/lft/lft2/index.htm*.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11.

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna. ISBN:3-900051-07-0.

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006), "Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models," *Biometrics*, 62, 1197–1206.

Salter-Townshend, M., and Haslett, J. (2006), "Modelling Zero Inflation of Compositional Data," in *Proceedings of the 21st International Workshop on Statistical Modelling*, pp. 448–456.

Scealy, J. L., and Welsh, A. H. (2011), "Regression for Compositional Data by Using Distributions Defined on the Hypersphere," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 73, 351–375.

Stephens, M. A. (1982), "Use of the von Mises Distribution to Analyse Continuous Proportions," *Biometrika*, 69, 197–203.

Stewart, C., and Field, C. (2010), "Managing the Essential Zeros in Quantitative Fatty Acid Signature Analysis," *Journal of Agricultural, Biological, and Environmental Statistics*, 16, 45–69.

Tjelmeland, H., and Lund, K. V. (2003), "Bayesian Modelling of Spatial Compositional Data," *Journal of Applied Statistics*, 30, 87–100.

Tsagris, M. T., Preston, S., and Wood, A.T. (2011), "A Data-Based Power Transformation for Compositional Data," in *Proceedings of CoDaWork: 4th International Workshop on Compositional Data Analysis*, eds. J. Egozcue, R. Tolosana-Delgado, and M. Ortego.

Unger, D. A. (1985), "A Method to Estimate the Continuous Ranked Probability Score," in *Preprints of the Ninth Conference on Probability and Statistics in Atmospheric Sciences, Virginia Beach, Virginia*, Boston: American Meteorological Society, pp. 206–213.

U.S. Census Bureau (2008), "TIGER/Line Shapefiles [machine-readable data files]," available at: *http://www.census.gov/geo/maps-data/data/tiger.html*.

U.S. Geological Survey (1999), "National Elevation Dataset," available at: *http://nationalmap.gov/viewer.html*.

van den Boogaart, K. G., and Tolosana-Delgado, R. (2008), "Compositions: A Unified R Package to Analyze Compositional Data," *Computers and Geosciences*, 34, 320–338.