Spatial Models for Discrete Compositional Data

Dean Billheimer

Department of Mathematical Sciences

University of Alaska Fairbanks

Peter Guttorp

Department of Statistics

University of Washington

21 November 1995

Abstract

We develop a spatial model for compositional data based on Aitchison's (1986) model of multiplicative error structure for proportions. An unobservable "state" proportion vector (composition) is assumed to exist for each sampled location. The state compositions may be related through time or space, and may also depend on covariates. Discrete observations are accommodated by a conditional multinomial observation model (given the state). Markov chain Monte Carlo (MCMC) is used to provide inference about the species compositions, covariates, and temporal and spatial dependence parameters. Graphical presentation of parameter estimates simplify model building and interpretation. The methods developed include models for independent, discrete compositions, time series of discrete and continuous compositions, and spatially related discrete and continuous compositions.

The methods are illustrated by an application. evaluating the ecological condition of the Delaware Bay estuary via benthic invertebrate composition.

We gratefully acknowledge the financial support of cooperative agreement CR821799 between EPA and the University of Washington. This manuscript has not been subjected to EPA's peer and policy review, and does not necessarily reflect the the views of the agency.

1 Introduction

The relative abundance of different species in a biological communtiy provides an example of compositions derived from discrete data. Such data are widespread in biological studies (e.g. ecological monitoring, fossil palynology), as well as market research (see e.g., Allenby and Lenk, 1994 JASA pp. 1218-1231). Data sampled from different spatial locations often exhibit spatial structure in the compositions. Our approach to such problems is a state-space model where each observation, $\mathbf{y}^j$, $j = 1, 2, ...n$ is a k-vector of counts of different groups. We posit that each observation is a multinomial realization with the proportion parameter vector equal to an unobservable state composition, $\mathbf{z}^j$. We use Aitchison's (1982) Logistic Normal distribution (LN) in conjunction with a Markov random field approach to spatial modeling (Besag, 1974; Mardia, 1988) to incorporate spatial dependence into the state distribution. In addition, methods to evaluate the effect of covariates are incorporated in the model specification. Markov chain Monte Carlo (MCMC) is used in a Bayesian setting for inference about the unobservable states and the state distribution parameters.

The method described above appears to be the first spatially explicit statistical approach for modeling compositions arising from count data. Mardia (1988) describes the Markov random field approach for multivariate normal observation vectors. He provides an example using logistic transformed compositions as the "observation vector" (derived from landsat classification). His emphasis is to illustrate the methodology, and estimate the spatial dependence parameter. There is no attempt to interpret the results in terms of the original compositions.

The LN distribution for compositional data was introduced by Aitchison and Shen (1980), who studied its properties and potential uses. They also compare the LN class with the Dirichlet class of distributions for compositions. Aitchison (1982) presents the LN as an analysis tool for compositional data, and established many of its mathematical and statistical properties. These results include the perturbation operation and its relevance to a central limit theorem for compositions. For a comprehensive account of statistical issues and analysis methods for independent, continuous compositions see Aitchison (1986).

Several researchers have developed methods for spatially related compositions and cat-

egorical data. For categorical spatial data, Upton and Fingleton (1989) and Cerioli (1992) focus on the analysis of spatial contingency tables. Observations occur on a regular spatial grid, and are assumed to be size one realizations from a Multinomial distribution (at each site). The Multinomial parameter, $\pi \in \nabla^{k-1}$, the $(k-1)$-dimensional simplex, is identical for all spatial locations. The general approach is to modify the standard contingency table methodology for Chi-square tests of independence and goodness of fit to account for the spatial dependence between locations. The main emphasis is on evaluating and correcting for the effect of spatial dependence between locations.

Pawlowsky and Burger (1992) use the additive logratio transformation (Aitchison and Shen, 1980) to analyze spatially distributed continuous compositions. They term such data regionalized compositions, since the underlying random functions have a constant sum at each point in the sampling region. The additive logratio transformation is used to transform the composition to unconstrained Euclidean space (multivariate logit scale). The spatial covariance structure of these transformed compositions is then modeled by co-kriging. The authors note that a difficult part of the analysis is that problems have to be formulated in terms of logratios, and interpretation and description of spatial dependencies are also on the same scale.

Finally, Besag, et al., (1995) describe a spatial logistic regression formulation for ninomial data. The response is mortality from prostate cancer for men classified by age group, time period, and birth cohort (date of birth). Spatial aspects of cancer mortality are not modeled directly. Instead, pairwise-difference prior distributions are used to accommodate the temporal ordering of parameters. That is, age, period and cohort effects are assumed to be similar for parameters that are adjacent in time. The prior distributions specify a "spatial smoothing" effect for mortality probability for contiguous age, period and birth cohort groups.

2 Perturbations and the Logistic Normal Distribution

Aitchison (1986) describes statistical analysis methods for compositional data with independent observations. These methods rely on the additive logratio transform (alr$(\cdot)$) to take observations from the $(k-1)$-dimensional simplex (∇^{k-1}) to $(k-1)$-dimensional Euclidean space ($\Re^{k-1}$). Aitchison assumes that the transformed data can be adequately modeled by the $(k-1)$ Multivariate Normal distribution. This transformation and assumption of Multivariate Normality define a distribution on ∇^{k-1}: the Additive Logistic Normal (LN) distribution.

A central idea behind the choice of the alr transform is a perturbation operator that can be used to model errors for compositional data (Aitchison, 1982). This model produces a structure for errors on ∇^{k-1} that is more natural than the usual additive error model used in other areas of statistics. Briefly, an observed proportion vector, $\mathbf{z}$, is modeled as a location vector ($\boldsymbol{\xi}$) "perturbed" by an error ($\boldsymbol{\alpha}$). The i^{th} element of $\mathbf{z}$ is

$$z_i = \frac{\xi_i \, \alpha_i}{\sum_{j=1}^{k} \xi_j \, \alpha_j} \qquad \text{with} \quad \alpha_i > 0 \quad \text{for all} \quad i = 1, 2, ..., k \quad \text{elements.}$$

This section introduces the perturbation operator, and shows how it leads to the LN distribution. We first describe the properties of the perturbation operator, and its interpretation for changes in compositions. The relation to the LN distribution is also described. The section concludes with a summary of other distributions for proportions, and illustrates how the LN can be superior in fitting biological data.

2.1 Intuition of the Perturbation Operator

The perturbation operator can be considered an "addition" operator for proportions. That is, for $\mathbf{z}^1, \, \mathbf{z}^2 \, \in \, \nabla^{k-1}$,

$$\mathbf{z}^1 \circ \mathbf{z}^2 = \text{“}\mathbf{z}^1 + \mathbf{z}^2\text{”} \quad \in \nabla^{k-1} \quad .$$

Aitchison (1986, section 2.8, p. 42) describes a simple motivating example that is related here to introduce this operation.

Suppose at time $t = 0$, we have a unit of perishable substance consisting of components 1, 2, and 3 in the proportions (x_1, x_2, x_3). Further suppose that component 1 degrades as a rate of u_1/unit time, while components 2 and 3 degrade at rates u_2 and u_3/unit time, respectively. At time $t = 1$, the remaining relative amounts of the three components are $(x_1u_1/c, x_2u_2/c, x_3u_3/c)$, where $c = \sum_{i=1}^{3} x_i u_i$. In this example, it is clear that each u_i must be between zero and one. However, it is easy to see that in biological situations where one or more of the components are allowed to grow, u_i might be greater than one. A property of this operator (shown in the next section) is that only the relative size of the components of $\mathbf{u}$ affect the resulting composition. So, $\mathbf{u}$ can be normalized to sum to one without affecting the result. Thus, we may consider $\mathbf{x}$ and $\mathbf{u}$ to be proportion vectors, and their "sum" to also be a composition. This operation is commonly used in statistics in Bayes' rule (Aitchison, 1986, section 2.10, p. 45). A prior distribution for a discrete parameter is "perturbed" by a vector of likelihood functions to form a posterior distribution. In biological applications, the perturbation operation is used by Edwards (1977) to describe relative gene frequencies after selection. In addition, Aebischer, et al. (1993) use this operation to model the active selection of available resources by animals.

2.2 Properties of Perturbations

Here we define the composition and perturbation operations, and describes their mathematical properties. Much of the development follows Aitchison (1986, section 2.8, pp. 42-43). The scalar multiplication and vector space properties are introduced here for the first time. Proofs of these properties are deferred to Appendix I. To ease notation we use lower-case to represent random variables as well as their corresponding values. The distinction should be clear from the context. When needed we follow the usual notation of using upper-case to denote the random variable, and lower-case to denote a realization.

Definition 2.1 *Composition Operator ($\mathcal{C}$) (Aitchison, 1986, p. 31)*
Suppose $\mathbf{a}$ *is a* k*-dimensional vector in positive Euclidean space* ($\Re_+^k$)*. Define* $\mathcal{C}(\mathbf{a})$ *by the following operation:*

$$[\,\mathcal{C}(\mathbf{a})\,]_i = \frac{a_i}{\sum_{j=1}^{k} a_j}$$

where $[\ \mathcal{C}(\mathbf{a})\ ]_i$ *denotes the* i^{th} *element of the* k*-vector* ($i = 1, 2, ..., k$).

Thus, the composition operator normalizes a positive k-vector to sum to one, and $\mathcal{C}(\mathbf{a}) \in \nabla^{k-1}$.

Definition 2.2 *Perturbation Operator (Aitchison, 1986, p. 42)*
Let $\mathbf{u}$ *be a* k*-part composition and* $\mathbf{a}$ *be a* k*-vector with positive elements. Define the perturbation operator as follows:*

$$\mathbf{u} \circ \mathbf{a} = \mathcal{C}(\mathbf{u} \cdot \mathbf{a}) \quad \textit{where } (\cdot) \textit{ denotes element-wise multiplication.}$$

Thus, the composition $\mathbf{u}$ is mapped to a location in ∇^{k-1} by the perturbing vector $\mathbf{a}$.

Aitchison (1986, pp. 43) shows several other simple properties of perturbations.

Property 2.1 *The operation* $\circ\, \mathbf{a}$ *is a one-to-one transformation between* ∇^{k-1} *and* ∇^{k-1}. *The inverse transformation is* $\circ\, \mathbf{a}^{-1}$, *where*

$$\mathbf{a}^{-1} = (\frac{1}{a_1}, \frac{1}{a_2}, ..., \frac{1}{a_k})$$

Property 2.2 *The effect of any perturbing vector* $\mathbf{a}$ *is the same as that for the composition* $\mathcal{C}(\mathbf{a})$.

$$\mathbf{u} \circ \mathbf{a} = \mathbf{u} \circ \mathcal{C}(\mathbf{a})$$

Therefore without loss of generality, we need only consider perturbing vectors in ∇^{k-1}.

We also present properties of the perturbation operator that are implied, or left as exercises in Aitchison (1986, p.47).

Property 2.3 *For* $\mathcal{I}_{k-1} = (\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k})$, *the operation* $\circ\, \mathcal{I}_{k-1}$ *is the identity operator for any* $\mathbf{u} \in \nabla^{k-1}$, *i.e.,* $\mathbf{u} \circ \mathcal{I}_{k-1} = \mathbf{u}$

Property 2.4 *The operation* $\circ$ *is commutative. For* $\mathbf{u}$ *and* $\mathbf{a}$ *in* ∇^{k-1},

$$\mathbf{u} \circ \mathbf{a} = \mathbf{a} \circ \mathbf{u}$$

Property 2.5 *The operation* $\circ$ *is associative. For* $\mathbf{u}$, $\mathbf{a}$, *and* $\mathbf{z}$ *in* ∇^{k-1},

$$(\mathbf{u} \circ \mathbf{a}) \circ \mathbf{z} = \mathbf{u} \circ (\mathbf{a} \circ \mathbf{z})$$

Notice that properties 1 through 5 imply that the perturbation operator defines an abelian group on ∇^{k-1}.

Theorem 2.1 *The perturbation operator defines an abelian group on* ∇^{k-1}.

In general, one may consider the perturbation operator to define an "addition" operator on the $(k-1)$-dimensional simplex. By adding the inverse of a composition, we also obtain a "subtraction" operator. This analogy with simple mathematical operations on $\Re$ leads to the corresponding multiplication and division analogs.

Scalar Multiplication of a Composition

Let $a \in \Re$ be any scalar, and $\mathbf{u} \in \nabla^{k-1}$ be a k-component composition.

Definition 2.3 *Scalar Multiplication*
Define scalar multiplication of a composition $\mathbf{u}$ *by* a *in the following way*

$$\mathbf{u}^a = \mathcal{C}(u_1^a, u_2^a, ..., u_k^a)$$

This defines a "multiplication" operator that is consistent with the perturbation "addition" analogy. To see this, note that for $a \in Z^+$ (a positive integer),

$$\begin{aligned}
\mathbf{u}^a &= \mathcal{C}(u_1^a, u_2^a, ..., u_k^a) \\
&= \mathbf{u} \circ \mathcal{C}(u_1^{a-1}, u_2^{a-1}, ..., u_k^{a-1}) \\
&= \mathbf{u} \circ \mathbf{u} \circ \mathcal{C}(u_1^{a-2}, u_2^{a-2}, ..., u_k^{a-2}) \\
&\vdots \\
&= \mathbf{u} \circ \mathbf{u} \circ ... \circ \mathbf{u} \qquad (a \text{ times})
\end{aligned}$$

Similarly, for $a \in Z^-$ (negative integers), the symmetric result holds for $\mathbf{u}^{-1}$. Taking $a \in Z^-$ corresponds to a division operation. This multiplication definition is consistent for

all real a. To show this we require the perturbation and scalar multiplication to define a vector space in ∇^{k-1}.

In addition, through use of the alr$(\cdot)$ transform ∇^{k-1} is shown to be an inner product space. (We use this to obtain the Cauchy–Schwarz inequality and a norm.)

Definition 2.4 *Inner Product for Compositions*
For $\mathbf{u}, \mathbf{z} \in \nabla^{k-1}$, *let* $\boldsymbol{\theta} = alr(\mathbf{u})$, *and* $\boldsymbol{\phi} = alr(\mathbf{z})$. *Define*

$$\langle \mathbf{u}, \mathbf{z} \rangle = \boldsymbol{\theta}' \mathcal{N}^{-1} \boldsymbol{\phi}$$

to be the inner product of $\mathbf{u}$ *and* $\mathbf{z}$.

Here, $\mathcal{N} = [I_{k-1} | -\mathbf{j}_{k-1}]$, where I_{k-1} is a $(k-1)$-dimensional identity matrix, and $\mathbf{j}_{k-1}$ is a $(k-1)$ column vector of ones. Note that

$$\mathcal{N}^{-1} = I_{k-1} - \frac{1}{/} k \mathbf{j}_{k-1} \mathbf{j}_{k-1}'$$

Theorem 2.2 $\langle \mathbf{u}, \mathbf{z} \rangle$ *is an inner product.*

An immediate consequence of this theorem is that the norm of $\mathbf{u} \in \nabla^{k-1}$ is defined as the non-negative square root of $\langle \mathbf{u}, \mathbf{u} \rangle$.

Definition 2.5 *Norm on* ∇^{k-1}
Define $\|\mathbf{u}\|$ *for* $\mathbf{u} \in \nabla^{k-1}$ *as* $\langle \mathbf{u}, \mathbf{u} \rangle^{1/2}$.

Note that the inner product and norm are invariant to permutations of components of $\mathbf{u}$. Proof of this fact is also included in Appendix 1.

The Cauchy–Schwarz and triangle inequalities follow immediately from the theory of inner product spaces (c.f. Rudin, 1974, Chapter 4, p. 80.). They are stated here for completeness.

Cauchy–Schwarz Inequality

$$|\langle \mathbf{u}, \mathbf{z} \rangle| \leq \|\mathbf{u}\| + \|\mathbf{z}\|$$

Triangle Inequality

$$\|\mathbf{u} \circ \mathbf{z}\| \leq \|\mathbf{u}\| + \|\mathbf{z}\|$$

Further, it follows from the triangle inequality that for $\mathbf{u}, \mathbf{w}$, and $\mathbf{z} \in \nabla^{k-1}$

$$\|\mathbf{u} \circ \mathbf{z}^{-1}\| \leq \|\mathbf{u} \circ \mathbf{w}^{-1}\| + \|\mathbf{w} \circ \mathbf{z}^{-1}\|$$

Thus, the distance between $\mathbf{u}$ and $\mathbf{z}$ is $\|\mathbf{u} \circ \mathbf{z}^{-1}\|$. The analogies with addition and subtraction of the perturbation operator are important for the statistical model development in subsequent chapters.

Finally, ∇^{k-1}, with the perturbation operator and scalar multiplication is a Hilbert space (i.e., a complete inner product space).

Theorem 2.3 *∇^{k-1} is a Hilbert space.*

2.3 Logistic Normal Distribution

The perturbation operation leads to the LN distribution as the limit distribution of a sequence of independent perturbations (section 2.4). The LN distribution was introduced by Aitchison and Shen (1980). Its use in the analysis of compositional data is chronicled by Aitchison (1986). Because of its flexibility in fitting compositional data and its role as a limit distribution, the LN is important for statistical modeling. The density function and the relevant properties of the LN distribution are shown here following the development of Aitchison (1986, Chapter 6, pp. 112–125).

To begin, we first define the additive logistic transformation.

Definition 2.6 *The additive logistic transformation is the one-to-one transformation of $\mathbf{y} \in \Re^{k-1}$ to $\mathbf{z} \in \nabla^{k-1}$ defined by*

$$z_i = \frac{\exp(y_i)}{\sum_{j=1}^{k-1} \exp(y_j) + 1} \quad , (i = 1, ..., k-1)$$

$$\textit{and} \quad z_k = \frac{1}{\sum_{j=1}^{k-1} \exp(y_j) + 1}$$

The inverse of this transformation is the additive logratio transformation (alr).

$$y_i = \log(\frac{z_i}{z_k})$$

The jacobian of the additive logistic transformation is $(\prod_{i=1}^{k} z_i)^{-1}$.

Denote the additive logratio transformation by $\text{alr}(\cdot)$, and its inverse by $\text{ialr}(\cdot)$.

Definition 2.7 *A k-part composition,* $\mathbf{z}$*, is said to have a Logistic Normal distribution* $L^{k-1}(\boldsymbol{\mu}, \Sigma)$ *when* $\mathbf{y} = \log(\frac{\mathbf{z}_{-k}}{z_k})$ *has a* $k-1$ *dimensional Multivariate Normal distribution with mean* $\boldsymbol{\mu}$ *and covariance matrix* Σ*, where* $\mathbf{z_{-k}} = (z_1, z_2, ..., z_{k-1})'$.

We use the notation $L^{k-1}(\boldsymbol{\mu}, \Sigma)$ to denote the Logistic Normal distribution.

Density Function

The density function for $L^{k-1}(\boldsymbol{\mu}, \Sigma)$ is, for $\mathbf{z} \in \nabla^{k-1}$

$$f(\mathbf{z} \mid \boldsymbol{\mu}, \Sigma) = \left(\frac{1}{2\pi}\right)^{\frac{k-1}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \left(\frac{1}{\prod_{i=1}^{k} z_i}\right) \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu})\right]$$

where

$$\boldsymbol{\theta} = \text{alr}(\mathbf{z}) = \log(\frac{\mathbf{z_{-k}}}{z_k})$$

Note that the density is defined only on the interior of the simplex. This is necessary because of the log transform.

Moments

All moments $(E\{\prod_{i=1}^{k} z_i^{a_i}\})$ and log moments $(E\{\prod_{i=1}^{k} \log(z_i)^{a_i}\})$ of all positive orders exist $(a_i > 0, i = 1, 2, ..., k)$. However, the integral expressions are not reducible to any simple form (Aitchison, 1986, section 6.4, p.116).

Expressions for moments of logarithms of ratios, z_i/z_j, can be obtained from multivariate Normal–Lognormal theory (with the convention $\mu_k = 0$, and $\sigma_{ik} = 0, \quad i = 1, ..., k$) .

$$E\{\log(\frac{z_i}{z_j})\} = E\{\log(\frac{z_i/z_k}{z_j/z_k})\} = \mu_i - \mu_j$$

$$\text{cov}\{\log(\frac{z_i}{z_l}), \log(\frac{z_j}{z_m})\} = \sigma_{ij} + \sigma_{lm} - \sigma_{im} - \sigma_{jl}$$

where σ_{ij} is the i, j element of Σ.

Alternatively, one may interpret μ_i as $E\{\log(z_i/z_k)\}$, and σ_{ij} as $\text{cov}\{\log(z_i/z_k), \log(z_j/z_k)\}$ Hence, $\boldsymbol{\mu}$ and Σ are the mean vector and covariance matrix for the multivariate logit $\text{alr}(\mathbf{z})$, with a Multivariate Normal distribution.

Interpretation of $\boldsymbol{\mu}$ as a Composition

The location parameter $\boldsymbol{\mu}$ can be expressed as a composition via the additive logistic transformation. That is,

$$\text{ialr}(\boldsymbol{\mu}) = \boldsymbol{\xi} \ , \ \text{where} \quad \boldsymbol{\xi} \in \nabla^{k-1}$$

As a point on the simplex, this value is directly interpretable as a composition. This is much simpler than interpreting $\boldsymbol{\mu}$, a multivariate vector of logits. However, some of the statistical properties of $\boldsymbol{\mu}$ are lost with the transformation to the simplex. $\boldsymbol{\mu}$ is the mean and mode of the Multivariate Normal logits (i.e., the $\text{alr}(\mathbf{z})$). The $\text{ialr}(\cdot)$ transform does not preserve these properties. However, $\text{ialr}(\cdot)$ transform is monotone in each of the $(k-1)$ components of $\boldsymbol{\mu}$. As a consequence, ordering of values is preserved under this transformation. Hence, $\boldsymbol{\xi} = \text{ialr}(\boldsymbol{\mu})$ can be interpreted as a component-wise multivariate median for the ALN distribution in ∇^{k-1}. As is shown in the following chapters, this interpretation is a useful characterization for point estimates of parameters, and as a "center" for the asymmetric ALN distribution. It remains to show that the transformation is monotone in each of the $(k-1)$ components.

For $i = 1, 2, ..., k-1$, $\boldsymbol{\xi} = \text{ialr}(\boldsymbol{\mu})$, implies

$$\xi_i = \frac{\exp(\mu_i)}{1 + \sum_{j=1}^{k-1} \exp(\mu_j)}$$

Taking logs of both sides we obtain,

$$\log(\xi_i) = \mu_i - \log(1 + \sum_{j=1}^{k-1} \exp(\mu_j))$$

Now, taking partial derivatives with respect to μ_i yields

$$\frac{\partial \log(\xi_i)}{\partial \mu_i} = 1 - \frac{\exp(\mu_i)}{1 + \sum_{j=1}^{k-1} \exp(\mu_j)}$$

Since, $\exp(\mu_i) < 1 + \sum_{j=1}^{k-1} \exp(\mu_j)$, the last term is less than one, and

$$\frac{\partial \log(\xi_i)}{\partial \mu_i} > 0$$

for all $\mu_i \in \Re$. Hence the transformation is monotone for each of the $(k-1)$ components of $\boldsymbol{\mu}$. Note that $\boldsymbol{\xi}_k$ does not present a concern since $\boldsymbol{\xi}_k = 1 - \sum_{i=1}^{k-1} \boldsymbol{\xi}_i$.

2.4 A Central Limit Theorem

Aitchison (1986, section 6.9, p. 125) shows that a sequence of independent errors, through the perturbation operation, leads to a central limit theorem. This result shows that it is at least possible that some processes might be well approximated by the LN distribution.

Theorem 2.4 *If* $\mathbf{x}^n$, $(n = 1, 2, ...)$ *is a sequence of compositions generated by successive, independent perturbations,* $\mathbf{u}^n$

$$\mathbf{x}^n = \mathbf{x}^{n-1} \circ \mathbf{u}^n$$

Then for large n *(under certain regularity conditions),* $\mathbf{x}^n$ *will follow a LN distribution.*

Sketch of proof. Let $\mathbf{y}^n = \text{alr}(\mathbf{x}^n)$ and $\mathbf{v}^n = \text{alr}(\mathbf{u}^n)$. Then

$$\begin{aligned} \mathbf{y}^n &= \mathbf{y}^{n-1} + \mathbf{v}^n \\ &= \mathbf{y}^0 + \mathbf{v}^1 + \mathbf{v}^2 + ... + \mathbf{v}^n \end{aligned}$$

The asymptotic normality of $\mathbf{y}^n$, and consequently the asymptotic additive logistic normality of $\mathbf{x}^n$, follow from the usual version of the multivariate central limit theorem.

2.5 Composition of a Lognormal Basis

Aitchison (1986, section 6.5, p. 117) shows that a Multivariate Lognormal distribution can be used to construct the Logistic Normal distribution.

Property 2.6 *Suppose* $\mathbf{v} \in \Re_+^k$ *has a Multivariate Lognormal distribution, with* $E\{\log(\mathbf{v})\} = \boldsymbol{\xi}$, *and* $Var\{\log(\mathbf{v})\} = \Omega$. *Then, its composition,* $\mathcal{C}(\mathbf{v})$, *is* $L^{k-1}(\mathbf{F}\boldsymbol{\xi}, \mathbf{F}\Omega\mathbf{F}')$, *where* $\mathbf{F} = [I_{k-1} \mid -j_{k-1}]$ *and* I_{k-1} *is the* $(k-1)$*-dimensional identity matrix, and* j_{k-1} *is a* $(k-1)$*-vector of ones.*

This Logistic Normal result is analogous to the Dirichlet result that the composition of independent Gamma random variables with equal scale yields a Dirichlet random vector. However, this result states that any Lognormal basis can give rise to a Logistic Normal composition. The Multivariate Lognormal can have a general covariance structure so long as it is a valid covariance matrix. The added generality is advantageous in modeling data where different components of a composition may have a covariance structure in addition to that induced by the summation constraint.

2.6 Permutation Properties

Aitchison(1986, section 5.5, pp. 93–96) shows that the LN density is invariant to permutations of the components of the composition vector $\mathbf{z}$. His proofs, though not presented here, rely on two key facts to show the density is not affected by permutations. These are

$$\mid \Sigma_p \mid = \mid \Sigma \mid$$

and

$$(\mathbf{y}_p - \boldsymbol{\mu}_p)' \Sigma_p^{-1} (\mathbf{y}_p - \boldsymbol{\mu}_p) = (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})$$

where $\mathbf{y} = \text{alr}(\mathbf{z})$, and the subscript p denotes the observation, parameter vector or matrix obtained from a permutation of the components of $\mathbf{z}$. The equality of the determinants of the variance-covariance matrices, and of the quadratic forms shows that the LN density is not affected by permutations. Consequently, it does not matter which component is chosen as the denominator of the transformation.

There is also a geometric interpretation of permuting the components of $\mathbf{z}$ that suggests intuitively that the density should be invariant. Begin by considering the case where k=1. The 1-dimensional simplex is the open interval (0, 1), and the LN density is a function defined on this interval. A permutation of the two components, (z_1, z_2) to (z_2, z_1) simply flips the endpoints of the interval so that 0 and 1 are interchanged; the resulting density is the mirror image of the original.

Adding another dimension results in $k = 2$, and the simplex is now represented by a ternary diagram. Graphically, the density is a surface sitting on top of the triangle. Permutations of the components of $\mathbf{z} = (z_1, z_2, z_3)$ now result in a rotation and/or an exchange

of two of the axes labels. In either case, the resulting density surface is a rotated, mirror image of the original surface. Although permutations affect the values of the individual parameters, they do not affect the value of the density.

2.7 Symmetric Logratio Transformation

An alternative approach for dealing with the summation constraint of compositional data is the symmetric logratio transform (see, e.g., Aitchison, 1986, section 4.6, p. 79). The transformation takes a k-component composition from ∇^{k-1} to a $(k-1)$ dimensional hyperplane in $\Re^k$. The transformation is defined as follows.

Definition 2.8 *The symmetric logratio transformation (slr) is the transformation of* $\mathbf{z} \in \nabla^{k-1}$ *to* $\mathbf{y} \in \Re^k$ *defined by*

$$y_i = \log\left(\frac{z_i}{(\prod_{i=1}^{k} z_i)^{\frac{1}{k}}}\right) \qquad , (i = 1, ..., k)$$

Note that the resulting vector $\mathbf{y}$ is also constrained by $\sum_{i=1}^{k} y_i = 0$. As a consequence, the transformed compositions lie in a $(k-1)$-dimensional hyperplane in $\Re^k$.

The advantage of the slr transform is that all k components of the composition are treated symmetrically. Further, even though the transformed components are restricted to sum to zero, they may otherwise range through $\Re^k$. The disadvantage is that the summation constraint results in a singularity in the variance-covariance matrix of the transformed components.

Aitchison (1986, section 4.8, p. 82) show that the slr transformation is related to the alr (additive logratio) by the following operation. For $\mathbf{z} \in \nabla^{k-1}$ (and $F = [I_{k-1}| - \mathbf{j}_{k-1}]$),

$$\text{alr}(\mathbf{z}) = \mathbf{F}\,\text{slr}(\mathbf{z})$$

This result holds since the denominator of the slr transformed composition cancels in $\log(z_i/\prod z_j) - \log(z_k/\prod z_j)$.

The result implies that if $\text{slr}(\mathbf{z})$ (of dimension k) has a degenerate Multivariate Normal distribution with mean $\boldsymbol{\mu}$ ($\sum \mu_i = 0$) and singular variance-covariance matrix Γ (rank $(k-1)$), then $\mathbf{z}$ has distribution $L^{k-1}(\mathbf{F}\,\boldsymbol{\mu}, \mathbf{F}\,\Gamma\,\mathbf{F}')$. Here, $\mathbf{F}\,\Gamma\,\mathbf{F}'$ is now full rank, and

the distribution is well defined. Consequently, if the Multivariate Normality assumption is made, there appears to be little advantage to working with the overspecified slr transformation. The exception to this occurs if the variance-covariance structure of the components is of particular interest. Aitchison (1986, section 8.4. p. 194) discusses advantages of the symmetric logratio parameterization for principal components analysis of compositional data.

2.8 Distribution Properties

Below I summarize several other properties of the LN distribution presented by Aitchison (1986, section 6.8, p. 123).

Property 2.7 *Let* $\mathbf{x}$ *be a* k *part composition with distribution* $L^{k-1}(\boldsymbol{\mu}, \Sigma)$*, and* $\mathbf{u}$ *a* k*-vector of positive components distributed independently of* $\mathbf{x}$*. Then the distribution of the vector* $\mathbf{z} = \mathbf{x} \circ \mathbf{u}$ *is given below for three different distribution assumptions on* $\mathbf{u}$*.*

1. For $\mathbf{u}$ distributed Lognormal with parameters $(\boldsymbol{\xi}, \Omega)$,

$$\mathbf{z} \sim L^{k-1}(\boldsymbol{\mu} + \mathbf{F}\boldsymbol{\xi},\ \Sigma + \mathbf{F}\Omega\mathbf{F}')$$

2. For $\mathbf{u}$ distributed $L^{k-1}(\boldsymbol{\theta}, \Theta)$,

$$\mathbf{z} \sim L^{k-1}(\boldsymbol{\mu} + \boldsymbol{\theta},\ \Sigma + \Theta)$$

3. For $\mathbf{u}$ a constant vector

$$\mathbf{z} \sim L^{k-1}(\boldsymbol{\mu} + \log(\frac{\mathbf{u}_{-k}}{u_k})\ ,\ \Sigma)$$

3 Logistic Normal – Multinomial State-Space Model

3.1 LN–Multinomial Model

The LN–Multinomial model is specified through its hierarchical structure. Suppose that at site j, the k-component composition, $\mathbf{z}^j$, given location parameter vector $\boldsymbol{\mu}$ and variance matrix Σ has $L^{k-1}(\boldsymbol{\mu}, \Sigma)$ distribution. Given a realization, $\mathbf{z}^j$, the observation, $\mathbf{y}^j$, has a Multinomial distribution. Combining these we see that the joint distribution for $\mathbf{y}^j$, $\mathbf{z}^j$, has the following density function.

$$f\left(\mathbf{y}^j,\, \mathbf{z}^j \mid \sum_{i=1}^{k} y_i^j = m^j,\, \boldsymbol{\mu}, \Sigma\right) =$$
$$\frac{m^j!}{\prod_{i=1}^{k} y_i^j!} \prod_{i=1}^{k} (z_i^j)^{(y_i^j - 1)} \left(\frac{1}{2\pi}\right)^{\frac{k-1}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathrm{alr}(\mathbf{z}^j) - \boldsymbol{\mu})' \Sigma^{-1} (\mathrm{alr}(\mathbf{z}^j) - \boldsymbol{\mu})\right]$$

where $\mathrm{alr}(\mathbf{z}^j) = \log(\mathbf{z}^j_{-k}/z^j_k)$.

Recall that the invertibility of $\mathrm{alr}(\cdot)$ allows a one-to-one mapping from $\Re^{k-1}$ to ∇^{k-1}. Thus, the location parameter, $\boldsymbol{\mu}$, can be expressed as a composition. Let $\boldsymbol{\xi} = \mathrm{ialr}(\boldsymbol{\mu})$, $(\boldsymbol{\xi} \in \nabla^{k-1})$ be the species composition location parameter. Prior distributions for $\boldsymbol{\mu}$ and Σ complete the model specification. Typically, we use Multivariate Normal prior distribution for $\boldsymbol{\mu}$, and a Wishart prior distribution for Σ^{-1}. Note that the LN-Multinomial model allows super-multinomial variability in the marginal distribution of $\mathbf{y}$.

Covariates

To incorporate the effect of covariates, the location parameter, $\boldsymbol{\mu}$, may depend on explanatory variables (for continuous compositions see Aitchison, 1986; section 7.6, p. 158). For a scalar covariate measured at site j, x_j (say), $\boldsymbol{\mu}^j$ can be replaced in the density expression by $\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(x_j - \bar{x})$. Here, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are vectors in $\Re^{k-1}$, and $\bar{x}$ is the mean of the observed covariate values. This parameterization allows interpretation of $\boldsymbol{\beta}_0$ as the overall location, and $\boldsymbol{\beta}_1$ as the change in location for a unit change in x.

Equivalently, the regression expression $\boldsymbol{\mu}^j = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(x_j - \bar{x})$ can be written as a perturbation of compositions. This is accomplished by taking the additive logistic transformation

of both sides,

$$\text{ialr}(\boldsymbol{\mu}^j) = \text{ialr}(\boldsymbol{\beta}_0) \circ \text{ialr}(\boldsymbol{\beta}_1)^{(x_j - \bar{x})}$$

We may write this equation more conveniently in the following form,

$$\boldsymbol{\xi}^j = \boldsymbol{\xi} \circ \boldsymbol{\gamma}^{u_j}$$

where $\boldsymbol{\xi}^j = \text{ialr}(\boldsymbol{\mu}^j)$, $\boldsymbol{\xi} = \text{ialr}(\boldsymbol{\beta}_0)$, and $\boldsymbol{\gamma} = \text{ialr}(\boldsymbol{\beta}_1)$. The scalar u_j is the centered covariate. In this parameterization, $\boldsymbol{\xi}$ is the overall location on the simplex. Further, the role of the regression composition parameter, $\boldsymbol{\gamma}$, is clear: the location parameter for $\mathbf{z}^j$ is the overall location ($\boldsymbol{\xi}$) perturbed by $\boldsymbol{\gamma}$ (for $u_j = 1$). Thus, $\boldsymbol{\gamma}$ is directly interpretable as a composition. It is the amount by which a location is shifted by a unit change in the covariate, via a perturbation. Finally, deviations in $\boldsymbol{\gamma}$ from the identity composition, $\mathcal{I}_{k-1}$ indicate the direction and magnitude of the change. Through this parameterization and the perturbation operator, regression parameters can be interpreted by their effect on compositions. This is more informative than the alternative interpretation on the log-odds scale that results from the alr($\cdot$) transform.

Figures 1 – 3 illustrate covariate dependence for compositional data. Figure 1 shows simulated compositions from the following model:

$$\mathbf{z}^j = \boldsymbol{\xi} \circ \boldsymbol{\gamma}^{u_j} \circ \boldsymbol{\alpha}^j$$

for $j = 1, 2, ..., 30$. For this simulation, $\boldsymbol{\xi} = (0.7, 0.2, 0.1)$, and $\boldsymbol{\gamma} = (0.350, 0.333, 0.317)$. Values of the covariate $\mathbf{u}$ are the integers from 1 to 30 centered to have mean zero. The errors, $\boldsymbol{\alpha}^j$ are independent, identically distributed $L^2(\boldsymbol{\mu}, \Sigma)$, where

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{bmatrix}$$

The figure shows that large positive values of the covariate are associated with a high proportion of component 1, while large negative values are associated with a smaller proportion of component 1, and relatively higher proportions of components 2 and 3.

Figure 2 shows the regression function plotted with the compositions. Hence, for a value of the covariate, u_j, the resulting composition $\mathbf{z}^j$ has location on the line ($\boldsymbol{\xi} \circ \boldsymbol{\gamma}^{u_j}$),

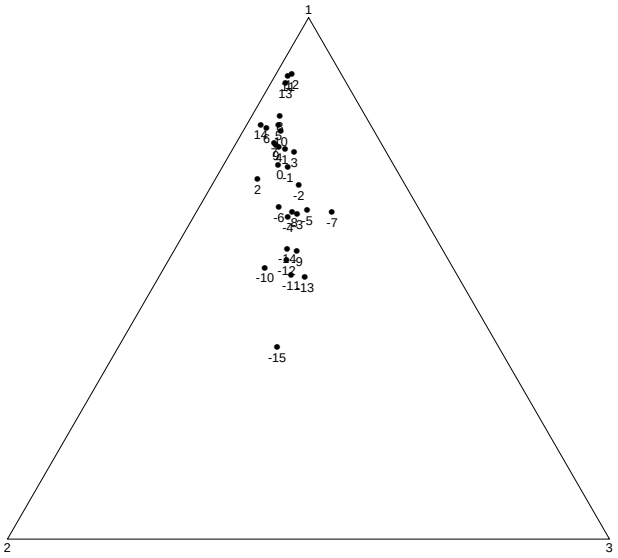

Figure 1: Simulated observations from a regression model. The numbers indicate the value of the centered covariate.

and an error distribution described by $L^2(\boldsymbol{\mu}, \Sigma)$. Note that this function describes linear relationships between the covariate and $\log(z_1^j/z_3^j)$, and the covariate and $\log(z_2^j/z_3^j)$. For this reason, regression diagnostics are more easily interpreted on the logit scale.

Finally, Figure 3 shows the errors from the simulated compositions. The errors are obtained by perturbing $\mathbf{z}^j$ by the inverse of its location ("subtraction"). That is,

$$\boldsymbol{\alpha}^j = \mathbf{z}^j \circ (\boldsymbol{\xi} \circ \gamma^{u_j})^{-1} \quad .$$

The errors can be used to assess the adequacy of the form of the regression function, and to estimate statistical properties (that is, to estimate Σ).

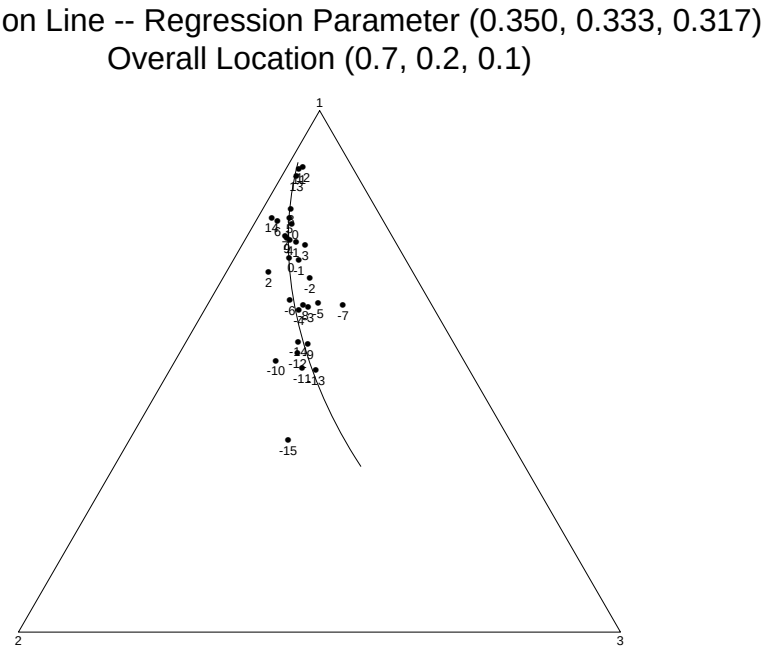

Figure 2: Actual regression line. The regression parameter vector is (0.350, 0.333, 0.317), with overall location (0.7, 0.2, 0.1).

3.2 Spatial Model

The LN–Multinomial state-space model can be extended to account for spatial structure between observations. This is done by specifying a Markov random field (MRF) prior distribution for for the compositions. Here we describe the conditional autoregressive (CAR; Besag, 1974; Mardia, 1988) MRF model. We present the general model specification and outline the MCMC implementation. Details of the specification for a particular application are presented in section 4. Note that the intrinsic autoregressive process (IAR; Künsch, 1987) can also be used to specify the MRF. (see Billheimer, 1995 for details regarding the IAR formulation for spatial compositions.)

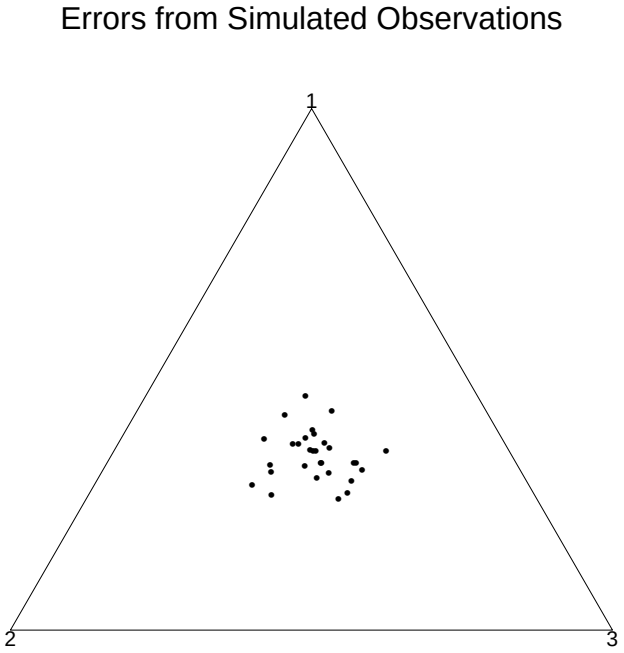

Figure 3: Errors from simulated regression.

3.3 Conditional Autoregressive Model Specification

Mardia (1988) describes the theoretical background for a multivariate normal Markov random field specification. Here we extend Mardia's result to a multi-site, logistic normal setting.

The typical method of specifying a CAR model is via the conditional distribution of the composition at site j, given all of the other sites. We let j index sites on a regular spatial lattice, $j = 1, 2, ..., n$. The location parameter at site j given all other sites is

$$E\{\text{alr}(\mathbf{z}^j) \mid \mathbf{z}^{-j}\} = \boldsymbol{\mu}^j + \sum_{r \in \delta j} \Lambda_{jr} \left(\text{alr}(\mathbf{z}^r) - \boldsymbol{\mu}^r\right)$$

where δj is the set of neighbors of site j, and $\mathbf{z}^{-j}$ denotes the compositions of all sites except site j. The conditional variance matrix for $\text{alr}(\mathbf{z}^j)$ given $\mathbf{z}^{-j}$ is

$$\text{Var}\left(\text{alr}(\mathbf{z}^j) \mid \mathbf{z}^{-j}\right) = \Gamma_j$$

Note that Γ_j and Λ_{jr} are $(k-1)\times(k-1)$ matrices, and Γ_j is positive definite for all j. Assuming $\mathbf{z}^j \mid \mathbf{z}^{-j}$ follows the logistic normal distribution, $\text{alr}(\mathbf{z}^j) \mid \mathbf{z}^{-j}$ is conditionally multivariate normal for all n sites.

Adapting the result of Mardia (1988), the joint distribution for $(\text{alr}(\mathbf{z}^1), \text{alr}(\mathbf{z}^2), ..., \text{alr}(\mathbf{z}^n))$ is $n(k-1)$ multivariate normal with mean vector

$$\boldsymbol{\mu}^{'} = (\boldsymbol{\mu}^{1'}, \boldsymbol{\mu}^{2'}, ..., \boldsymbol{\mu}^{n'})$$

and variance matrix

$$\Sigma = \{\text{Block}(-\Gamma_j^{-1}\Lambda_{jr})\}^{-1}$$

provided $\Lambda_{jr}\Gamma_r^{'} = \Gamma_j\Lambda_{rj}^{'}$ (for symmetry of Σ), and $\text{Block}(-\Lambda_{jr})$ is positive definite (define $\Lambda_{jj} = -I_{k-1}$). The term "Block" refers to a large matrix comprised of sub-matrices, each of dimension $(k-1)\times(k-1)$, where the $(j,r)^{th}$ sub-matrix of the large matrix is $-\Gamma_j^{-1}\Lambda_{jr}$. (Note that in the symmetry condition we correct a typographic error in Mardia, 1988.) It follows that the distribution for $\mathbf{z}^{'} = (\mathbf{z}^{1'}, \mathbf{z}^{2'}, ..., \mathbf{z}^{n'})$ is a multivariate version of the logistic normal distribution with density given by

$$\pi(\mathbf{z}^1, \mathbf{z}^2, ..., \mathbf{z}^n \mid \boldsymbol{\mu}, \Sigma) =$$
$$\left(\frac{1}{2\pi}\right)^{\frac{n(k-1)}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \prod_{j=1}^{n}\left(\frac{1}{\prod_{i=1}^{k} z_i^j}\right) \exp\left[-\frac{1}{2}\sum_{j=1}^{n}\sum_{r=1}^{n}(\boldsymbol{\theta}^j - \boldsymbol{\mu}^j)'\Gamma_j^{-1}\Lambda_{jr}(\boldsymbol{\theta}^r - \boldsymbol{\mu}^r)\right]$$

for $\mathbf{z} \in (\nabla^{k-1})^n$. Here, $\boldsymbol{\theta}^j = \text{alr}(\mathbf{z}^j)$ and $\prod_{j=1}^{n}\prod_{i=1}^{k}(1/z_i^j)$ is the Jacobian of the ialr transformation. The variance-covariance matrix is $\Sigma = \left\{\text{Block}(-\Gamma_j^{-1}\Lambda_{jr}\right\}^{-1}$, where $\Lambda_{jj} = -I_{k-1}$ (from Mardia, 1988). Σ must be symmetric and positive definite. The symmetry condition requires

$$\Lambda_{jr}\Gamma_r^{'} = \Gamma_j\Lambda_{rj}^{'}$$

(Again, this expression corrects a typographic error in Mardia, 1988.) Also, the positive definiteness of Σ holds if the matrix defined by $\text{Block}(-\Lambda_{jr})$ is positive definite. This result follows from the factorization of Σ^{-1} (see Mardia, 1988). The variance-covariance matrix

can be factored to the product of a block diagonal matrix, $\text{Diag}(\Gamma_j^{-1})$, and the matrix of spatial dependence parameters as follows.

$$\Sigma^{-1} = \text{Diag}(\Gamma_j^{-1})\,\text{Block}(-\Lambda_{jr})$$

Since Γ_j is assumed positive definite, the condition for Block $(-\Lambda_{jr})$ ensures that Σ is positive definite.

3.4 Inference

The LN-Multinomial model provides a framework for incorporating observed counts from different groups, covariate information, and spatial structure for making inferences. The questions under study may be about the composition at a specific site (or over a region). Alternatively, the quantity of primary interest may be the effect of a covariate or characteristics of the spatial dependence structure. We use a Bayesian approach for making inferences about these quantities. Advances in MCMC, particularly in the last five years, make Bayesian inference flexible and straight-forward in complicated hierarchical modeling problems. These developments allow the use of non-conjugate prior distributions and likelihoods. In addition, missing observations and prediction of latent (state) variables are easily accommodated. Finally, posterior distributions of the quantities of interest are directly approximated from the Monte Carlo realizations. (for a comprehensive review see Besag, et al., 1995)

Markov Chain Monte Carlo

MCMC was introduced as an inference method for Bayesian statistics by Grenander (1983), and Geman and Geman (1984). It's use in Bayesian inference has expanded tremendously in the last five years (see among others Gelfand and Smith, 1990, and Gelfand, Hills, Racine-Poon, and Smith, 1990, for introductory applications; Besag et al., 1995, for an extensive review of the methodology; and Tierney, 1994, for a rigorous theoretical treatment for continuous state-spaces). The idea of MCMC is to sample from an ergodic Markov chain with limit distribution, $\pi(\mathbf{x})$ say, that is equal to the posterior distribution (the dependence on the data $\mathbf{y}$ is subsumed in the notation). A realization of the Markov

chain, $\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N$ is used to construct an empirical distribution, $\hat{\pi}_N(\mathbf{x})$ that directly approximates $\pi(\mathbf{x})$.

For inference, functionals of $\pi(\mathbf{x})$, $T(\pi)$, are estimated by $T(\hat{\pi}_N)$. If the Markov chain is positive recurrent and aperiodic with stationary distribution $\pi(\mathbf{x})$, the functionals of the empirical distribution converge to functionals of the true distribution (Levental, 1988). The only constraint on the functional, $T(\cdot)$, is that small changes in $\hat{\pi}_N$ should not lead to large changes in $T(\hat{\pi}_N)$. Technically, the requirement is that $T(\cdot)$ must be continuous with respect to the *sup norm* at π for convergence (Pollard, 1984, section 4.2, p. 70).

As an example, consider

$$T(\pi) = \int g(\mathbf{x}) d\pi = E_\pi \, g(\mathbf{x})$$

where $g(\cdot)$ is a function defined on the range of $\mathbf{x}$. Then

$$\frac{1}{N} \sum_{j=1}^{N} g(\mathbf{x}^j) = T(\hat{\pi}_N) \rightarrow E_\pi \, g(\mathbf{x}) \text{ a.s.}$$

Thus, if $g(\mathbf{x}) = \mathbf{x}$, $T(\hat{\pi}_N) \rightarrow E_\pi(\mathbf{x})$. Tierney (1994) provides the details of almost sure convergence and a central limit theorem for functionals of $\pi(\mathbf{x})$. Higdon (1994) provides a less technical description for expectations, quantiles and credible intervals.

Hastings Algorithm for Compositions Hastings algorithms (Hastings, 1970) can be used to construct a general class of transition kernels that maintain the stationary distribution of the Markov chain. Because this method has not been widely applied for updating compositional quantities, we describe its application here. This construction will be described briefly for general random vectors, and then shown in detail for compositional data. For a complete description of Hastings algorithm see Besag et al., (1995).

To begin, suppose that we are given the current state vector $\mathbf{x}$ with n components, and the density of the stationary distribution $\pi(\mathbf{x})$ (equivalently, the posterior distribution). Typically, this density is known only up to a constant of proportionality. In this description, the components of $\mathbf{x}$ are unknown parameters. Our notation, $\pi(\mathbf{x})$ subsumes the dependence on the data, $\mathbf{y}$. That is, we denote $\pi(\mathbf{x} \mid \mathbf{y})$ by $\pi(\mathbf{x})$.

Take $\mathbf{x}_i$ to be a subset of the components of $\mathbf{x}$. Subsets are used in the development so that several components of $\mathbf{x}$ may be updated simultaneously. (For example, one may

consider $(k-1)$ elements of a k-composition to be a subset of $\mathbf{x}$. Updating $(k-1)$ of the elements also updates the k^{th} by the summation constraint.) We generate a proposed new value of the state, $\mathbf{x}^*$, from a transition kernel with density $P_i(\mathbf{x}_i \rightarrow \mathbf{x}_i^* \mid \mathbf{x}_{-i})$. The proposed value has components $\mathbf{x}_{-i}^* = \mathbf{x}_{-i}$, where $\mathbf{x}_{-i}$ denotes all components of $\mathbf{x}$ not in $\mathbf{x}_i$. The proposal $\mathbf{x}^*$ is accepted as the new state with probability

$$\min\left(1, \frac{\pi(\mathbf{x}^*)P_i(\mathbf{x}_i^* \rightarrow \mathbf{x}_i \mid \mathbf{x}_{-i})}{\pi(\mathbf{x})P_i(\mathbf{x}_i \rightarrow \mathbf{x}_i^* \mid \mathbf{x}_{-i})}\right)$$

Otherwise, $\mathbf{x}$ is retained as the state. Because $\pi(\mathbf{x})$ and $\pi(\mathbf{x}^*)$ differ only in the i-th subset of the components, many terms in the posterior densities cancel. Typically, $\pi(\mathbf{x}^*)/\pi(\mathbf{x})$ simplifies to $\pi(\mathbf{x}_i^* \mid ...)/\pi(\mathbf{x}_i \mid ...)$, the ratio of the full conditional densities.

The Hastings algorithm can be implemented for a k-component composition, $\mathbf{z}$, in the following manner ($\mathbf{z}$ is used here to denote the proportion vector at a single site, but this method can be used for any composition vector). A composition, $\mathbf{v}$, is sampled from $L^{k-1}(0_{k-1}, b\mathcal{N})$, where $\mathcal{N}$ is a $(k-1)\times(k-1)$ positive definite variance–covariance matrix. Typically, we specify this matrix as

$$\mathcal{N} = I_{k-1} + \mathbf{j}_{k-1}\mathbf{j}_{k-1}^{'}$$

where I_{k-1} is a $(k-1)$-dimensional identity matrix, and $\mathbf{j}_{k-1}$ is a $(k-1)$ vector of ones. Sampling from this distribution is straightforward since the sample value can be constructed from the composition of k independent, identically distributed Lognormal random variables, with location parameter 0, and scale parameter b. The scalar b is a scaling constant selected to maintain an acceptance rate of 30–80%.

The proposal composition, $\mathbf{z}^*$ is constructed by perturbing the current state $\mathbf{z}$ by the composition $\mathbf{v}$. That is,

$$\mathbf{z}^* = \mathbf{z} \circ \mathbf{v}$$

Note that this proposal distribution is not symmetric in $\mathbf{z}$ and $\mathbf{z}^*$. That is, the density of $\mathbf{z}^*|\mathbf{z}$ is not equal to the density of $\mathbf{z}|\mathbf{z}^*$. To compute the Hastings acceptance probability, we account for the asymmetry (to maintain the correct limit distribution) through the ratio $P(\mathbf{z}^* \rightarrow \mathbf{z} \mid ...)/P(\mathbf{z} \rightarrow \mathbf{z}^* \mid ...)$, where $P(\mathbf{z} \rightarrow \mathbf{z}^* \mid ...)$ is the LN density with location

parameter alr($\mathbf{z}$) and variance-covariance matrix $b\mathcal{N}$. This distributional result follows directly from property 2.7. The proposed composition $\mathbf{z}^*$ is accepted as the new state with probability

$$\min\left(1,\ \frac{\pi(\mathbf{z}^*)P(\mathbf{z}^* \to \mathbf{z} \mid ...)}{\pi(\mathbf{z})P(\mathbf{z} \to \mathbf{z}^* \mid ...)}\right)$$

Otherwise, $\mathbf{z}$ is retained as the state. This approach is used to update all compositionsl quantities.

Gibbs Sampler The Gibbs sampler (Geman and Geman, 1984) is a special case of Hastings algorithm where the proposal $\mathbf{x}_i^*$ is from the full conditional distribution of $\mathbf{x}_i$, $\pi(\mathbf{x}_i \mid \mathbf{x}_{-i})$. For this algorithm, the proposal $\mathbf{x}_i^*$ is always accepted. The Gibbs Sampler is appropriate when realizations can be generated easily from the conditional distribution of $\mathbf{x}_i$. Typically, once $\mathbf{x}_i$ has been updated, another subset of components is selected from $\mathbf{x}_{-i}$, and is updated accordingly. In this manner, all elements of $\mathbf{x}$ are updated to obtain another realization from the Markov chain. Typically, the Gibbs sampler is used for Uniform or (Multivariate) Gaussian distributed quantities. The algorithm is described briefly following Higdon (1994).

The Gibbs sampler is based on the following result. Given $\mathbf{X} = \mathbf{x}$, define the random vector $\mathbf{X}^* = (\mathbf{X}_i^*, \mathbf{X}_{-i}^*)$ such that $\mathbf{X}_{-i}^* = \mathbf{x}_{-i}$, and $\mathbf{X}_{-i}^* \mid \mathbf{x}_{-i}$ has density $\pi(\mathbf{x}_i^* \mid \mathbf{x}_{-i})$. It follows that $\mathbf{X}^*$ has marginal density $\pi(\mathbf{x})$.

$$\begin{aligned} Pr(\mathbf{X}^* = \mathbf{x}^*) &= Pr(\mathbf{X}_i^* = \mathbf{x}_i^*; \mathbf{X}_{-i}^* = \mathbf{x}_{-i}^*) \\ &= Pr(\mathbf{X}_i^* = \mathbf{x}_i^*; \mathbf{X}_{-i} = \mathbf{x}_{-i}^*) \\ &= \pi(\mathbf{x}_{-i}^*)\pi(\mathbf{x}_i^* \mid \mathbf{x}_{-i}^*) \\ &= \pi(\mathbf{x}_i^*) \end{aligned}$$

Thus, replacing $\mathbf{x}_i$ by $\mathbf{x}_i^* \sim \pi(\mathbf{x}_i \mid \mathbf{x}_{-i})$ maintains $\pi(\mathbf{x})$. This defines a Markov chain transition kernel

$$P_i(\mathbf{x} \to \mathbf{x}^*) = \pi(\mathbf{x}_i^* \mid \mathbf{x}_{-i}) I_{(\mathbf{x}_{-i} = \mathbf{x}_{-i}^*)}$$

where π is the stationary distribution for P_i. Cycling through the subsets of elements of $\mathbf{X}$ allows all components to be updated.

3.5 Estimation

3.5.1 Composition Estimates

MCMC is used to construct point estimates and credible regions for inference about site compositions. The approach is to (temporarily) approximate the posterior distribution of each site composition by a LN distribution, and to use the MCMC realizations to estimate the LN parameters. These estimates are the point estimates for location and variance parameters. To construct a size $(1-\alpha)$ credible region, the MCMC realizations are first ordered according to their (approximate) posterior density (assuming the LN parameter estimates). After ordering, the $(1-\alpha)$ fraction of realizations with highest density are retained. The estimated credible region is constructed by a convex hull circumscribing the retained high-density realizations.

To illustrate this process, consider a composition at site j, and suppose $\mathbf{z}^{j1}, \mathbf{z}^{j2}, ..., \mathbf{z}^{jN}$ is a dependent realization from the marginal posterior, $\pi(\mathbf{z}^j)$. Assume (for the moment) that this posterior marginal distribution is approximated by $\mathbf{z}^j \sim L^{k-1}(\boldsymbol{\mu}^j, \Sigma^j)$. The parameters, $\boldsymbol{\mu}^j$ and Σ^j are estimated from the MCMC realizations.

$$\hat{\boldsymbol{\mu}^j} = \frac{1}{N}\sum_{i=1}^{N} \text{alr}(\mathbf{z}^{ji})$$

and

$$\hat{\Sigma^j} = \frac{1}{N}\sum_{i=1}^{N} \left[\text{alr}(\mathbf{z}^{ji}) - \hat{\boldsymbol{\mu}^j}\right] \left[\text{alr}(\mathbf{z}^{ji}) - \hat{\boldsymbol{\mu}^j}\right]'$$

$\text{ialr}(\hat{\boldsymbol{\mu}^j}) = \boldsymbol{\xi}^j$ is the point estimate for the location of $\mathbf{z}^j$. Recall from 2.3 that $\boldsymbol{\xi}^j$ is a "multivariate median" for the species composition at site j.

To construct a size $(1-\alpha)$ credible region, order the N MCMC realizations according to their approximate posterior density with $\hat{\boldsymbol{\mu}^j}$ and $\hat{\Sigma^j}$ as parameters. A $100\,(1-\alpha)\%$ credible region is then constructed by taking the convex hull circumscribing the $(1-\alpha)$ fraction of the $\mathbf{z}^{ji}$ values with highest (approximate) density. If the marginal distribution were truly LN, the hull would be a Monte Carlo estimated $(1-\alpha)$ HPD region (assuming that iso-density contours are convex). Because the analytic form of the marginal distribution is not known, the hull describes a Monte Carlo estimated $(1-\alpha)$ size credible region.

Regression Parameter Estimates

The posterior distribution of $\boldsymbol{\gamma}$ is approximated by its inclusion in the MCMC updating scheme. Point estimates and credible regions are calculated in the same manner as described for species compositions. Values of $\boldsymbol{\gamma}$ near the center of the simplex, $\mathcal{I}_{k-1}$, indicate that the effect of the covariate is not strong. In particular, $\boldsymbol{\gamma} = \mathcal{I}_{k-1}$ indicates that the covariate has no predictive effect for species compositions.

3.6 Prediction

Prediction regions for compositions can be constructed via the MCMC methodology (Besag, et al., 1995). Let $\boldsymbol{\theta}, \mathbf{z}$, and $\mathbf{y}$ denote the state distribution parameters, estimated site compositions and observed counts, respectively, for sampled sites. Let $\mathbf{z}_p$ denote the compositions to be predicted at unmonitored sites. Following Besag, et al. (1995) predictive distributions for the unmonitored sites can be constructed by including $\mathbf{z}_p$ as additional components to be updated via MCMC. This results in an approximate joint posterior distribution of parameters, observed site compositions and predicted site compositions given the observed counts, $\mathbf{y}$. The marginal distributions of $\mathbf{z}_p$ are obtained by ignoring all other components of the posterior distribution, and working directly with the prediction site Monte Carlo realizations. Prediction regions of size $(1-\alpha)$ can be constructed by ordering the realizations by their respective densities from the LN distribution with parameters estimated from the realizations.

Diagnostics

The major difficulty in diagnosing model inadequacy for state-space models is that the state variable is not observed. Either the state or observation model can be mis-specified. However, only the fitted and observed compositions can be examined. Since the fitted composition depends on both the state and the observation models, a source of mis-specification cannot be identified . For Bayesian hierarchical models the standard approach for assessing model inadequacy is through the posterior predictive distribution (Geisser, 1980; Rubin, 1984; Meng, 1994). The idea is to construct a predictive distribution for a future observable quantity. This prediction is conditional on the observed data and the model specification. If the model is adequate, the future observable will be well predicted. The degree of being

"well predicted" can be quantified by calculating the tail area probability corresponding to the observable. This quantity has been termed the posterior predictive p-value (Meng, 1994).

To diagnose model inadequacy for state-space models, we use this idea in combination with other standard diagnostic procedures. These are

- Prediction of "omitted observations".
- Regression lack of fit.
- Scientific interpretation of regression parameters.

Prediction of Omitted Observations The MCMC method for constructing predictive distributions at unmonitored sites can be used to identify and evaluate influential observations. This is done by simply ignoring the observed counts at the site in question, and computing its predictive distribution. The actual counts at that site can then be evaluated with respect to the predictive distribution. Sites can be evaluated sequentially in a "leave one out" diagnosis scheme. This method is particularly valuable for diagnosis in spatial models where the counts at a single site can have considerable influence over predictions at neighboring sites.

Regression Lack of Fit For models including a covariate, we can construct diagnostic plots to evaluate the adequacy of the fitted form of the covariate. To see this, recall that a composition, $\mathbf{z}^j$ can be expressed as

$$\mathbf{z}^j = \boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{u_j}$$

This implies that the effect of the covariate can be evaluated by examining $\mathbf{z}^j \circ (\boldsymbol{\xi}^j)^{-1} = \boldsymbol{\gamma}^{u_j}$. This effect is more easily evaluated on the logit scale by taking the alr transform of $\mathbf{z}^j \circ (\boldsymbol{\xi}^j)^{-1}$. That is,

$$\text{alr}\left(\mathbf{z}^j \circ (\boldsymbol{\xi}^j)^{-1}\right) = u_j \boldsymbol{\beta}$$

where $\boldsymbol{\beta} = \text{alr}(\boldsymbol{\gamma}) \in \Re^{k-1}$. Thus we may examine plots for each of the $(k-1)$ components of $\text{alr}\left(\mathbf{z}^j \circ (\boldsymbol{\xi}^j)^{-1}\right)$ versus the centered predictor to assess whether the linear fit is adequate. Such plots allow evaluation of the functional form of the covariate for inclusion in the statistical model.

Interpretation of Regression Parameters Interpretation of regression parameters in terms of their scientific meaning provide a useful model diagnostic. Parameter estimates failing to confirm a known (or strongly suspected) covariate effect can indicate a model specification problem.

4 CAR Spatial Model for Biological Pollution Monitoring

Section 5 presents modelling results from a biological monitoring study evaluating the natural variability in benthic invertebrate compositions in Delaware Bay. Here we describe the modifications to the general methodology of section 3 used to model the benthic invertebrate data. In addition to spatial dependence, we also demonstrate the inclusion of a covariate effect (salinity). Hence the model specification incorporates regression and spatial components.

To begin, suppose the group composition at site j, $\mathbf{z}^j$, $j = 1, 2, ..., n$, depends on a mean zero (scalar) covariate x_j through the regression relationship described in section 3.1. That is,

$$\mathbf{z}^j = \boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{x_j}$$

where $\boldsymbol{\gamma} \in \nabla^{k-1}$ is the regression parameter composition, and $\boldsymbol{\xi}^j$ (also in ∇^{k-1}) is the composition of the spatial process "adjusted" for the effect of the covariate. We may interpret $\boldsymbol{\xi}^j$ as a hypothetical composition at site j assuming that the covariate is at its mean value. To account for spatial structure, let the prior (state) distribution for $\boldsymbol{\xi}^j$ follow a CAR model (as described in section 3.2). Then,

$$E\left[\text{alr}(\boldsymbol{\xi}^j) \mid \boldsymbol{\xi}^{-j}\right] = \boldsymbol{\mu} + \sum_{r \in \delta j} \Lambda_{jr} \left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]$$

where $\boldsymbol{\eta} = \text{ialr}(\boldsymbol{\mu})$ is the overall location (level) of the spatial process, and Λ_{jr} is a $(k-1) \times (k-1)$ matrix of spatial dependence parameters. (Recall that δj is the set of neighbors of site j.) In addition, the conditional variance for the composition at site j is specified.

$$\text{Var}\left[\text{alr}(\boldsymbol{\xi}^j) \mid \boldsymbol{\xi}^{-j}\right] = \Gamma_j$$

Assuming $\boldsymbol{\xi}^j \mid \boldsymbol{\xi}^{-j}$ is LN, Mardia's (1988) result for Multivariate Normal random vectors leads to the following result for LN random vectors. The random quantity $\boldsymbol{\xi}' = (\boldsymbol{\xi}^{1'}, \boldsymbol{\xi}^{2'}, ..., \boldsymbol{\xi}^{n'})$ follows a multivariate version of the ALN distribution with density given by

$$\pi(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, ..., \boldsymbol{\xi}^n \mid \boldsymbol{\mu}, \Sigma) =$$
$$\left(\frac{1}{2\pi}\right)^{\frac{n(k-1)}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \prod_{j=1}^{n} \left(\frac{1}{\prod_{i=1}^{k} \xi_i^j}\right) \exp\left[-\frac{1}{2}\sum_{j=1}^{n}\sum_{r=1}^{n}(\boldsymbol{\theta}^j - \boldsymbol{\mu})' \Gamma_j^{-1} \Lambda_{jr} (\boldsymbol{\theta}^r - \boldsymbol{\mu})\right]$$

where $\boldsymbol{\theta}^j = \text{alr}(\boldsymbol{\xi}^j)$. The variance-covariance matrix is $\Sigma = \left\{\text{Block}(-\Gamma_j^{-1}\Lambda_{jr}\right\}^{-1}$, where $\Lambda_{jj} = -I_{k-1}$ (from Mardia, 1988). Σ must be symmetric and positive definite.

To implement this model as the prior distribution for the spatial process, several simplifying assumptions are made. First, because sites may have different numbers of neighbors (from 1 to 6 "first-order" neighbors), assume that the conditional variance at site j depends on the number of neighbors as follows:

$$\Gamma_j = \frac{1}{n_j}\Gamma$$

where n_j is the number of neighbors of site j. The site composition (adjusted for the covariate) $\boldsymbol{\xi}^j$ is predicted with greater precision as the number of neighbors increases. The matrix Γ describes the relative variability and covariance relationships between the different groups (given the neighboring sites). This assumption provides a mechanism for allowing increased variability at "edge" sites.

Combining this assumption with the symmetry condition $\Lambda_{jr}\Gamma_r' = \Gamma_j \Lambda_{rj}'$ implies the following relationship:

$$\frac{1}{n_r}\Lambda_{jr}\,\Gamma' \quad = \quad \frac{1}{n_j}\Gamma\,\Lambda_{rj}'$$

We also simplify the covariance structure by assuming the spatial influence of the neighbors of site j is the same for all neighbors. Hence,

$$\Lambda_{j,r} = \begin{cases} \Lambda_j & \text{if} \quad r \in \delta j \\ -I_{k-1} & \text{if} \quad r = j \\ 0_{(k-1)\times(k-1)} & \text{otherwise} \end{cases}$$

As a final simplification, assume that $\Lambda_j = \lambda/n_j \;\; I_{k-1}$. Hence, the spatial dependence is assumed to be the same for all groups (actually the same for all logits $\log(\xi_i^j/\xi_k^j)$). Further, $\log(\xi_i^j/\xi_k^j)$ and $\log(\xi_m^r/\xi_k^r)$ are conditionally independent, given all other logits.

Note that the final assumption, $\Lambda_j = \lambda/n_j\, I_{k-1}$ (when site r is a neighbor of site j), combined with $\Gamma_j = \Gamma/n_j$ implies that the spatial dependence is the same for all neighbor pairs, regardless of direction.

These assumptions result in a simplification of the form for the matrix Block$(-\Lambda_{jr})$. This matrix can be written as follows.

$$\text{Block}(-\Lambda_{jr}) = \begin{bmatrix} I_{k-1} & -\frac{1}{n_1}\lambda \mathrm{I}_{(2\in\delta 1)} & \cdots & -\frac{1}{n_1}\lambda \mathrm{I}_{(n\in\delta 1)} \\ -\frac{1}{n_2}\lambda \mathrm{I}_{(1\in\delta 2)} & I_{k-1} & \cdots & -\frac{1}{n_2}\lambda \mathrm{I}_{(n\in\delta 2)} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_{(k-1)}}\lambda \mathrm{I}_{(1\in\delta n)} & -\frac{1}{n_{(k-1)}}\lambda \mathrm{I}_{(2\in\delta n)} & \cdots & I_{k-1} \end{bmatrix}$$

Here, I_{k-1} denotes the $(k-1)$ identity matrix, and $\mathrm{I}_{(2\in\delta n)}$ denotes the indicator function for site 2 being a neighbor of site n. Thus, for each row (j) of the matrix, the (j,j) cell is 1, and for all neighbors of site j $(r \in \delta j)$, there is a single non-zero element equal to $-\lambda/n_j$. That is, for any row, the non-zero elements are a single 1 and n_j identical elements $-\lambda/n_j$. As a consequence of this simplified form, a sufficient condition for positive definiteness of Block$(-\Lambda_{jr})$ is that $\mid \lambda \mid < 1$ (each row sum is less than one).

Expressions for the observation density (likelihood) and prior distributions complete the model specification. The observed group counts are assumed conditionally multinomial given the unobservable site composition, $\mathbf{z}^j$. Since $\mathbf{z}^j$ is comprised of spatial and regression components, $\boldsymbol{\xi}^j$ and $\boldsymbol{\gamma}^{x_j}$ are parameters in the likelihood expression.

$$p(\mathbf{y}^j \mid \boldsymbol{\xi}^j, \boldsymbol{\gamma}, m_j) \;=\; \frac{m_j!}{\prod_{i=1}^k y_i^j!} \prod_{i=1}^k (\boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{x_j})_i^{y_i^j}$$

where $m_j = \sum_{i=1}^k y_i^j$, and $(\,\cdot\,)_i$ denotes the i^{th} component of $(\boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{x_j})$.

Prior distributions are required for λ, $\boldsymbol{\gamma}$, $Q = \Gamma^{-1}$, and $\boldsymbol{\mu}$, the (alr transformed) overall location (level) of the spatial process. We assume the following prior distributions.

$$\pi(\lambda) = \frac{1}{2}\, \mathrm{I}_{(-1<\lambda<1)}$$

$$\pi(\boldsymbol{\gamma}) = L^{k-1}\,(\,0_{k-1},\, a\,\mathcal{N})$$

$$\pi(Q) = \text{Wishart}(\,[b\,\mathcal{N}]^{-1}\,,\, \rho)$$

$$\pi(\boldsymbol{\mu}) = N_{k-1}\,(\,0_{k-1},\, c\,\mathcal{N})$$

Recall that $\mathcal{N} = I_{k-1} + \mathbf{j}_{k-1}\mathbf{j}_{k-1}'$. Typical choices for a, b, and c are $a = c = 0.5$, and $b = 1$. The hyperparameter ρ must be at least $(k-1)$ to make $\pi(Q)$ a proper distribution.

Combining the prior distributions with the likelihood, we obtain the following expression for the posterior distribution (up to a constant of proportionality).

$$
\begin{aligned}
&\pi(\boldsymbol{\xi}, \boldsymbol{\mu}, \lambda, \boldsymbol{\gamma}, Q \mid \mathbf{y}) \propto \\
&\quad \prod_{j=1}^{n} \left(\prod_{i=1}^{k} (\boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{x_j})_i^{y_i^j} \right) \mid Q \mid^{\frac{n}{2}} \mid \text{Block}(-\Lambda_{jr}) \mid^{\frac{1}{2}} \times \\
&\quad \prod_{j=1}^{n} \left(\prod_{i=1}^{k} \frac{1}{\xi_i^j} \right) \exp\left\{ \frac{1}{2} \sum_{j=1}^{n} \sum_{r=1}^{n} \left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu} \right]' Q \Lambda_{jr} \left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu} \right] \right\} \times \\
&\quad \left(\prod_{i=1}^{k} \frac{1}{\gamma_i} \right) \exp\left[-\frac{1}{2} \text{alr}(\boldsymbol{\gamma})' (a\mathcal{N})^{-1} \text{alr}(\boldsymbol{\gamma}) \right] \exp\left[-\frac{1}{2} \boldsymbol{\mu}' (c\mathcal{N})^{-1} \boldsymbol{\mu} \right] \times \\
&\quad \mid b\mathcal{N} \mid^{\frac{\rho}{2}} \mid Q \mid^{\frac{\rho-k}{2}} \exp\left[-\frac{1}{2} tr(b\mathcal{N} Q) \right] \times \mathrm{I}_{(-1<\lambda<1)}
\end{aligned}
$$

Note that the determinant $\mid \text{Diag}(n_j Q) \mid$ simplifies to $\prod_{j=1}^{n} \mid n_j Q \mid$. The n_j can be factored from each of the n determinants and incorporated in the normalizing constant. Also note the omission of a minus sign in the exponential term for $\boldsymbol{\xi}$. This is implicitly included in Λ_{jr}.

$$
\Lambda_{jr} = \begin{cases} \frac{1}{n_j}\lambda\, I_{k-1} & \text{if} \quad r \in \delta j \\ -I_{k-1} & \text{if} \quad r = j \\ 0_{(k-1)\times(k-1)} & \text{otherwise} \end{cases}
$$

This posterior distribution leads to the following full conditional distributions for $\boldsymbol{\xi}^j$, $\boldsymbol{\mu}$, λ, and Q.

$$
\begin{aligned}
\pi(\boldsymbol{\xi}^j \mid ...) \quad &\propto \quad \prod_{i=1}^{k} \left[(\boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{x_j})_i^{y_i^j} \left(\frac{1}{\xi_i^j} \right) \right] \times \\
&\qquad \exp\left\{ \frac{1}{2} \sum_{j=1}^{n} \sum_{r=1}^{n} \left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu} \right]' Q \Lambda_{jr} \left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu} \right] \right\} \\
&\propto \quad \prod_{i=1}^{k} \left[(\boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{x_j})_i^{y_i^j} \left(\frac{1}{\xi_i^j} \right) \right] \exp\left\{ -\frac{1}{2} \left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\theta}^j \right]' n_j Q \left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\theta}^j \right] \right\}
\end{aligned}
$$

where $\boldsymbol{\theta}^j = \boldsymbol{\mu} + \sum_{r \in \delta j} \Lambda_{jr} \left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu} \right]$ is the conditionally expected value of $\text{alr}(\boldsymbol{\xi}^j)$ given all other sites.

$$\pi(\boldsymbol{\mu} \mid ...) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}'(c\,\mathcal{N})^{-1}\boldsymbol{\mu} - \sum_{j=1}^{n}\sum_{r=1}^{n}\left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu}\right]' Q\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]\right)\right\}$$

$$\pi(\boldsymbol{\gamma} \mid ...) \;\propto\; \prod_{j=1}^{n}\left[\prod_{i=1}^{k}(\boldsymbol{\xi}^j \circ \boldsymbol{\gamma}^{x_j})_i^{y_i^j}\right]\prod_{i=1}^{k}\left(\frac{1}{\gamma_i}\right)\exp\left\{-\frac{1}{2}\,\text{alr}(\boldsymbol{\gamma})'\,(a\,\mathcal{N})^{-1}]\,\text{alr}(\boldsymbol{\gamma})\right\}$$

$$\pi(\lambda \mid ...) \propto$$
$$\mid \text{Block}(-\Lambda_{jr}) \mid^{\frac{1}{2}} \exp\left\{\frac{1}{2}\sum_{j=1}^{n}\sum_{r=1}^{n}\left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu}\right]' Q\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]\right\} \times \text{I}_{(-1<\lambda<1)}$$

$$\pi(Q \mid ...) \;\propto\; \mid Q \mid^{\frac{n}{2}} \mid b\,\mathcal{N} \mid^{\frac{\rho}{2}} \mid Q \mid^{\frac{\rho-k}{2}} \exp\left\{-\frac{1}{2}\,tr(b\,\mathcal{N}\,Q)\right\} \times$$
$$\exp\left\{\frac{1}{2}\sum_{j=1}^{n}\sum_{r=1}^{n}\left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu}\right]' Q\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]\right\}$$

This last expression implies that the full conditional distribution for Q is Wishart with parameter matrix

$$b\,\mathcal{N} - \sum_{j=1}^{n}\sum_{r=1}^{n} tr\left\{\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]\left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu}\right]'\right\}$$

and $n+\rho$ degrees of freedom. This result follows from re-arranging terms in the exponential expression for $\boldsymbol{\xi}$. Because

$$\left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu}\right]' Q\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]$$

is a scalar quantity, it is equal to its own trace. A property of the trace operator is that $tr(AB) = tr(BA)$ as long as both multiplications are defined (see e.g., Hocking, 1985). Applying this operation twice, the following result is obtained.

$$tr\left\{\left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu}\right]' Q\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]\right\}$$
$$= \quad tr\left\{\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r) - \boldsymbol{\mu}\right]\left[\text{alr}(\boldsymbol{\xi}^j) - \boldsymbol{\mu}\right]' Q\right\}$$

It follows that

$$\sum_{j=1}^{n}\sum_{r=1}^{n}\left[\text{alr}(\boldsymbol{\xi}^j)-\boldsymbol{\mu}\right]' Q\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r)-\boldsymbol{\mu}\right]$$
$$= \sum_{j=1}^{n}\sum_{r=1}^{n} tr\left\{\Lambda_{jr}\left[\text{alr}(\boldsymbol{\xi}^r)-\boldsymbol{\mu}\right]\left[\text{alr}(\boldsymbol{\xi}^j)-\boldsymbol{\mu}\right]' Q\right\}$$

leading to the distributional result. These full conditional distributions are used directly in the MCMC algorithm.

MCMC Implementation

The MCMC algorithm is used to obtain a Markov chain realization from the joint posterior distribution. The algorithm updates $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, ..., \boldsymbol{\xi}^n$, $\boldsymbol{\mu}$, λ, $\boldsymbol{\gamma}$, and Q, each conditional on all other parameters (and on the data, $\mathbf{y}$). Hastings algorithm (1970) for compositions, described in section 3.4, is used to update the $\boldsymbol{\xi}^j$'s, and $\boldsymbol{\gamma}$. This method is also used to update $\boldsymbol{\mu}$ via its additive logistic transformed value, $\boldsymbol{\eta} = \text{ialr}(\boldsymbol{\mu})$. That is. $\boldsymbol{\eta}^* \in \nabla^{k-1}$ is proposed via the composition proposal method of section 3.4, and it is transformed to $\boldsymbol{\mu}^*$ to compute its Hastings acceptance probability. λ is updated via a symmetric, Uniform proposal density and Metropolis algorithm acceptance probability (Metropolis, et al., 1953). We also use Hastings algorithm to update the matrix Q. Proposals, Q^*, are generated from the Wishart$(Q^{(c)}, \text{df})$ distribution where $Q^{(c)}$ is the current value of the matrix Q, and df is the proposal degrees of freedom. df is adjusted to maintain a 30–80% acceptance rate.

Initial values for all parameters are required to begin the MCMC realization. The initial site compositions (ignoring the covariate), $\mathbf{z}^{j(1)}$, are computed from the observed group counts at each site j. That is,

$$\mathbf{z}^{j(1)} = \mathcal{C}(\mathbf{y}^j), \quad \text{for } j = 1, 2, , ..., n.$$

For any site with a zero count, the initial composition is adjusted to the interior of the simplex by adding one to each category and then normalizing. Initial values of $\boldsymbol{\xi}^{j(1)}$, $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\gamma}^{(1)}$, and Q are computed from the $\mathbf{z}^{j(1)}$ values. Let $\mathbf{v}^j = \text{alr}(\mathbf{z}^{j(1)})$, and $X_j = (1, x_j)$ (recall that x_j is the centered covariate). Combining the rows X_j into an $(n \times 2)$ matrix, X, the

initial values of $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are obtained from the least squares solutions to the multivariate regression equation (see, Aitchison, 1986, section 7.6, p. 158).

$$\left[\boldsymbol{\mu}^{(1)},\, \mathrm{alr}(\boldsymbol{\gamma}^{(1)})\right]' = (X'X)^{-1}X'\mathbf{v}$$

where $\mathbf{v}_{n\times(k-1)} = (\mathbf{v}^1, \mathbf{v}^2, ..., \mathbf{v}^n)'$. The initial values of $\boldsymbol{\xi}$ are set by removing the effect of the covariate from the initial site composition via the perturbation operator. Thus,

$$\boldsymbol{\xi}^{j^{(1)}} = \mathbf{z}^{j^{(1)}} \circ (\boldsymbol{\gamma}^{(1)})^{-x_j}$$

Finally, an initial value for Q is obtained as follows

$$Q^{(1)} = \frac{1}{n}\sum_{j=1}^{n}\left[\mathrm{alr}(\boldsymbol{\xi}^{j^{(1)}}) - \boldsymbol{\mu}^{(1)}\right]\left[\mathrm{alr}(\boldsymbol{\xi}^{j^{(1)}}) - \boldsymbol{\mu}^{(1)}\right]'$$

λ is initialized at zero.

Several test runs of the MCMC algorithm using the benthic invertebrate data indicate that a burn-in of 2000 cycles is sufficient to achieve convergence to the stationary distribution. These runs also suggest that $\boldsymbol{\eta}$ and the site compositions are "fast mixing". That is, the posterior distribution is adequately explored by the Markov chain with relatively few cycles. Conversely, realizations for λ and $\boldsymbol{\gamma}$ do not mix well, and require a longer MCMC runs for inference. A MCMC run of length 1.25 million was used to explore the posterior distribution. To ease memory and disk storage requirements, every 50^{th} realization from the chain was collected for statistical inference. Subsequent runs evaluating model performance used run lengths of 20,000, collecting every realization. Appendix II summarizes MCMC convergence issues for this application.

5 Natural Variability of Benthic Invertebrates in Delaware Bay

To illustrate the use of these methods, we present an analysis of the natural variability of benthic invertebrate populations in Delaware Bay. The data are from the US EPA's Environmental Monitoring and Assessment Program Estuaries Resource group. An extensive analysis of these data is presented in Billheimer, et al. (1995). This presentation is to illustrate how the general methodology is adapted to a specific problem.

In 1990, 25 locations in Delaware Bay were sampled to evaluate the benthic communities, as well as physical and chemical characteristics at each sample site. The locations of these sites are shown in Figure 4. The stations are identified by their station ID number.

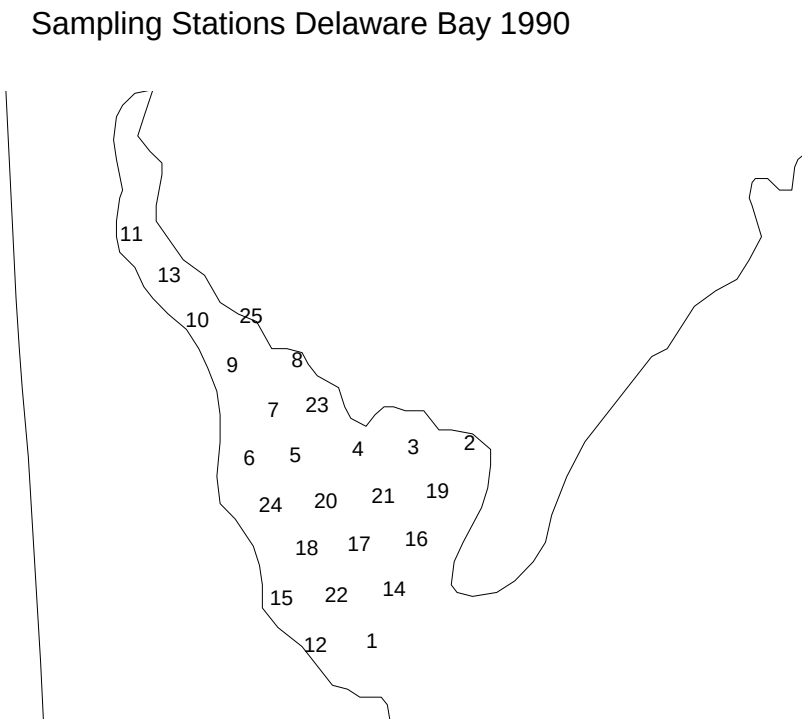

Figure 4: Sampling stations from Delaware Bay. These are sampling locations from the EMAP–Estuaries group, Virginia Province, 1990

As the figure indicates, a triangular lattice was used in locating the sample sites. Overton, et al., (1990) provide details of the sample design. This sampling strategy allows each site to have as many as six equidistant near-neighbors, and is advantageous for assessing the spatial dependence structure.

At each site, three grab samples (subsamples) of the bottom sediment were collected. These samples were later processed to remove and identify benthic organisms, as well as to determine the physical and chemical characteristics of the substrate. In addition, depth, salinity, dissolved oxygen, temperature, pH and other characteristics were measured at the time of sampling. The observed species were classified into one of three groups: disturbance tolerant, disturbance intolerant, and suspension feeders. (see Billheimer, et al., 1995) for details)

Figure 5 graphically displays the relative abundance of the three benthic groups for each of the Delaware Bay sampling stations. In this figure, the stations are identified by their station ID numbers (corresponding to the ID's in Figure 4). From this display it is clear that the benthic compositions range from almost exclusively tolerant organisms (station #23) to nearly all pollution intolerant organisms (station #11). Sites #2 and #8 have the greatest proportions of suspension feeders.

Figure 6 shows the group composition information with spatial information. The plotting symbols are known as "stars". They consist of a central point (located at the sample station coordinates), with arms radiating outward. Each arm corresponds to one of the three group proportions. The length of each arm denotes the magnitude of the corresponding proportion. The endpoints of the arms are joined by line segments. Note that the star diagram allows an arbitrary number of categories to be displayed.

In addition to the stars, salinity contours of 22.5 and 28 parts per thousand (ppt) are identified. This figure indicates that much of the observed spatial structure in the compositions may be explained by salinity.

A preliminary analysis was conducted to investigate the within site sampling variability. The goal was to determine if the three grab samples collected at each site could be considered three Multinomial samples with a single proportion parameter vector. A Chi-square test of homogeneity was performed for the grab samples at each site (see Billheimer, et al.,

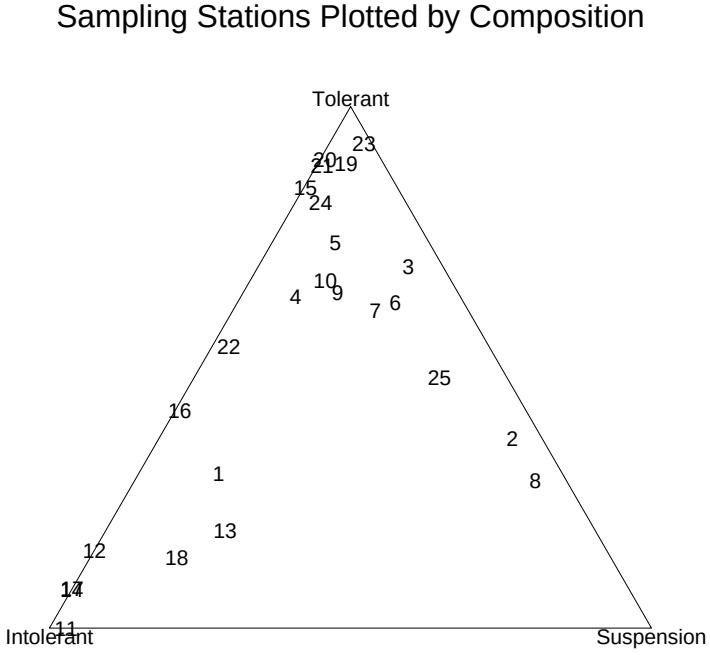

Figure 5: Sampling stations plotted by composition. Each sampling station is plotted by its observed composition.

1995 for details). The preliminary analysis clearly indicated that the observed variability is not explained by a Multinomial sampling model. As is typical of many biological sampling problems, the grab samples exhibit super-Multinomial variability.

Statistical Modeling Results

The statistical model described in section **??** was used to analyze the benthic compositions of Delaware Bay. The model uses a spatial structure defining neighbors of station j as those stations (when present) at the vertices of a hexagon centered at j. Any hexagon with a "missing" vertex (i.e., no station) simply has fewer neighbors. Inference about the site compositions, the spatial dependence parameter (λ), and the salinity regression composition ($\boldsymbol{\gamma}$) result from a MCMC run with a burn-in of 2000 cycles, and a collection phase

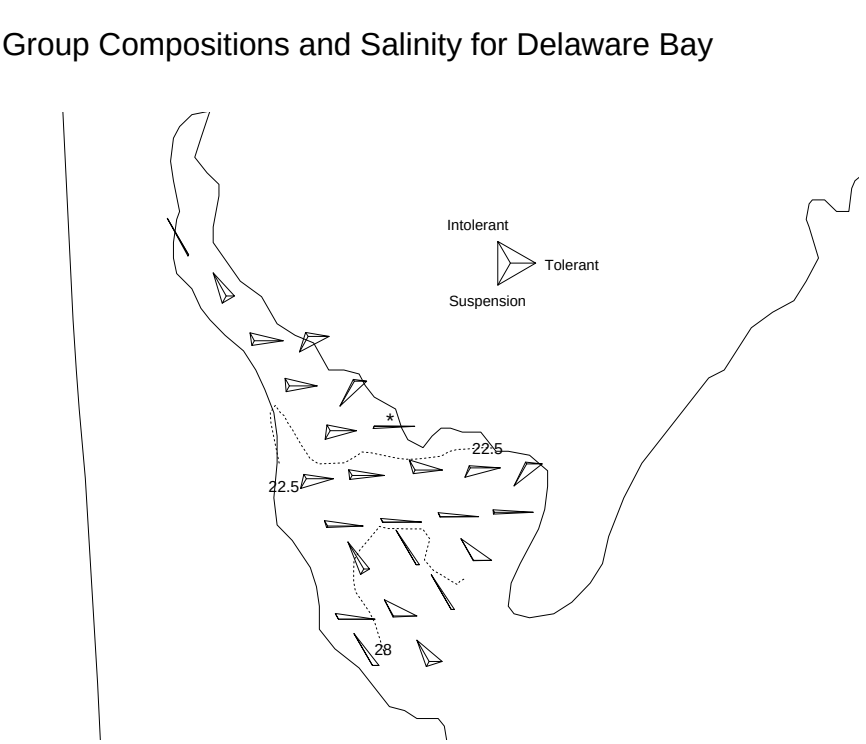

Figure 6: Group compositions from Delaware Bay. These are observed compositions at each sampling station. The contour lines denote low (< 22.5 ppt), medium ($22.5 - 28$ ppt), and high (> 28ppt) salinity.

of 1.25 million cycles subsampling every 50^{th} realization. Credible regions for all compositional quantities were constructed by the method described in section 3.5. Diagnostics evaluating MCMC performance indicate that the 1.25 million cycles may not be adequate to adequately evaluate the posterior distribution. Indeed, γ appears to mix quite slowly, and may require a run length of about 3 million cycles for reliable inference. The slow mixing behavior may be related to partial confounding of the salinity gradient with the spatial structure of the observations. Such confounding makes separation of the salinity and spatial effects difficult. Diagnostics for assessing convergence of the Markov chain and precision of the credible regions are discussed in Appendix II.

Figure 7 shows the 95% credible region for the benthic composition at each sample site.

The observed composition from the sample is also shown for each site. For many of the sites, the large number of observed organisms (and conditional Multinomial observation model) limits the range of plausible compositions. For a number of sites the size of the plotting symbol for the observed composition is roughly the same size as the credible region. However, for sites with fewer total organisms (e.g., #9, #10, #13) the credible region extends to a range of values around the observed composition. The point estimate for the overall benthic composition in Delaware Bay is (0.53, 0.32, 0.15).

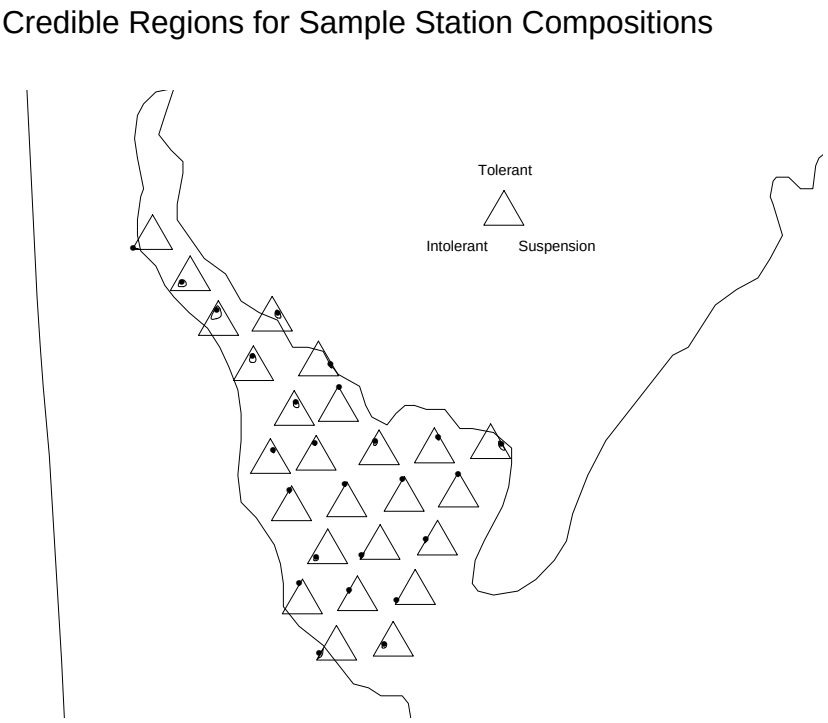

Figure 7: 95% credible regions for compositions for Delaware Bay stations. The observed composition for the site is plotted with the prediction region from the spatial model.

The credible region for the salinity effect is shown in Figure 8. The point estimate for this composition is (0.33, 0.38, 0.29). 50% and 95% credible regions are also shown in the figure. These regions indicate that the association between salinity and benthic composition is not neutral. The point estimate can be interpreted in the following way: an increase

in salinity of 1 ppt has the effect of perturbing a benthic composition by (0.33, 0.38, 0.29) (over the observed range of 15 – 30 ppt salinity). This point estimate indicates that as salinity increases, the proportion of suspension feeders decreases. Suspension feeders are replaced by pollution intolerant organisms. The proportion of tolerant organisms is not much affected by salinity.

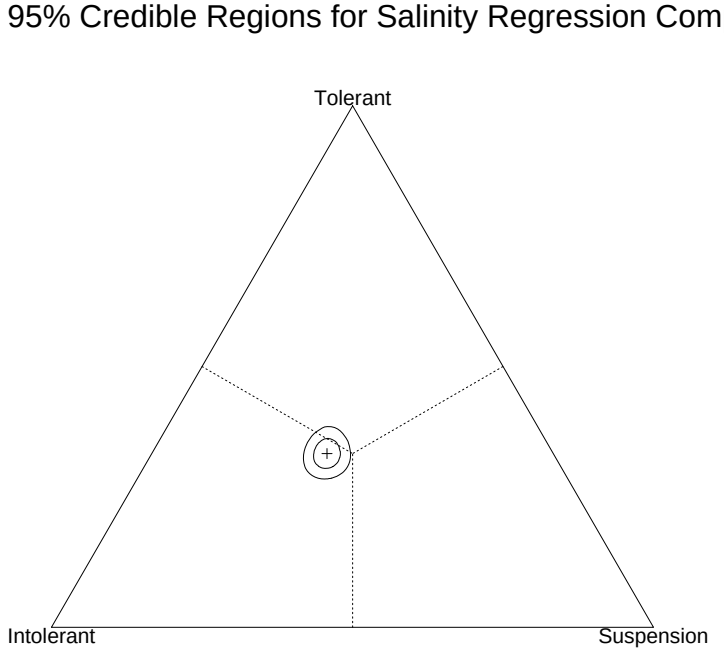

Figure 8: Point estimate and 50% and 95% credible regions for salinity regression parameter. The point estimate is (0.33, 0.38, 0.29).

For comparison, the association between water depth and benthic composition is also explored. The point estimate for this covariate effect is (0.330, 0.336, 0.334). This result indicates little association between water depth and benthic composition. Further, the 95% credible region for this composition covers the identity element, $\mathcal{I}_2 = (1/3,\ 1/3,\ 1/3)$, further suggesting little association between depth and benthic composition.

The realized values of the spatial dependence parameter (λ) are shown in Figure 9. This

figure suggests that there is spatial similarity between neighboring sites (i.e., $\lambda > 0$). The median value for this realization is 0.68, while the observed mean is 0.61. The observed mode is about 0.87. Nearly 96% of the realized values are greater than zero, and 70% are greater than 0.5.

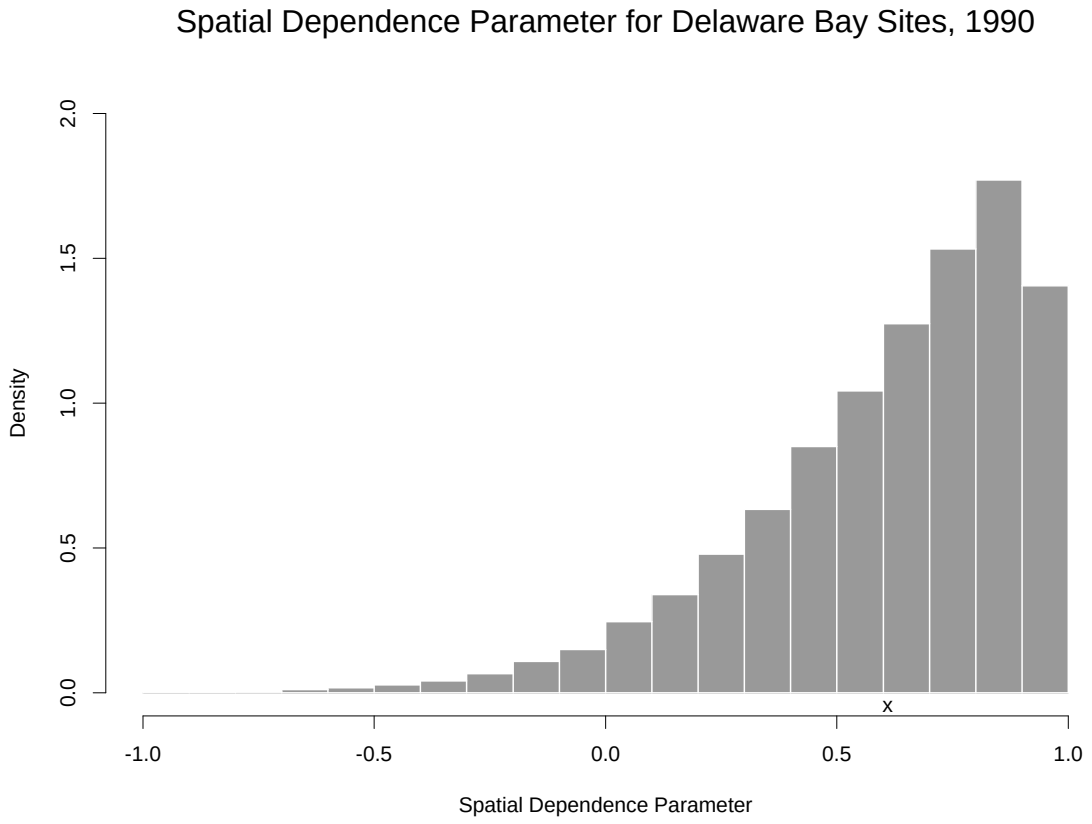

Figure 9: MCMC realizations for the spatial dependence parameter, λ. The median for λ is 0.68, and the mode is about 0.87. The x indicates the mean of the realizations of 0.61.

To better evaluate the evidence of spatial dependence, a Bayes factor is computed using the Savage density ratio (see Kass and Raftery, 1995 for a review). This ratio compares the prior density for λ with the posterior density; both evaluated at $\lambda = 0$ (spatial independence). A large value for the ratio indicates that the posterior density is shifted away from zero, and that the data provide evidence against spatial independence. The posterior density is approximated using a kernel density estimator with the MCMC realizations of λ. Note that these realizations approximate the posterior distribution of λ

integrated over all other parameters. The kernel estimator results in a value of 0.18 for the posterior density at $\lambda = 0$. The prior distribution for λ, Uniform(-1, 1), gives a prior density of 0.5. Hence, the Bayes factor is $0.5/0.18 = 2.8$. This value indicates moderate evidence of positive spatial dependence.

Note that the spatial dependence and effect of salinity are estimated simultaneously. Salinity is a spatially varying covariate that (generally) increases along the gradient from river to ocean across the estuary. The observed spatial dependence is present after accounting for the effect of salinity. Thus, λ denotes spatial dependence above that explained by the salinity gradient. The consequences of the positive spatial dependence are examined further in the Discussion section. The identification of salinity and spatial dependence effects is discussed in Appendix A.

Finally, the compositions "adjusted" for salinity are shown in Figure 10. The adjusted compositions are the 95% prediction regions for benthic compositions that would be expected if all sites had the mean salinity value (22.7 ppt). The figure shows that stations with high salinity have compositions shifted toward the suspension feeding vertex. Conversely, those with low salinity values exhibit shifts toward reduced suspension feeders (and more intolerant and tolerant organisms). This figure provides another method of interpreting the association of changing salinity and benthic invertebrate composition.

5.1 Model Evaluation

Prediction at omitted sites and residual diagnostics for the salinity covariate were used to evaluate the adequacy of the statistical model.

5.1.1 Leave-one-out Prediction.

Leave-one-out predictions regions were constructed for the sampling stations to evaluate the adequacy of the statistical model for spatial prediction. To construct these regions (for site j, say), the MCMC algorithm was used with the benthic counts for site j replaced with zeros. All other site counts were unchanged. The zero counts maintain the neighborhood structure for site j, and allow its composition to be updated as a regular part of the MCMC algorithm. This is the recommended method for accommodating missing observations

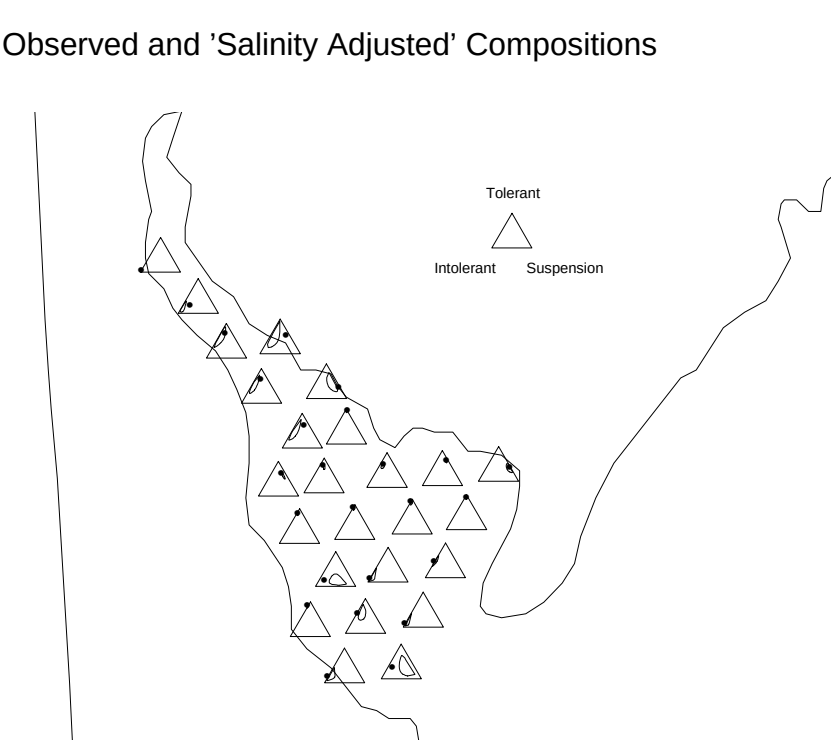

Figure 10: .

Observed and "salinity adjusted" compositions. The adjusted compositions are 95% prediction regions for compositions with salinity set to its mean observed value (23.7 ppt).

for MCMC (Besag et al., 1995). A 95% prediction region for the site composition was constructed via the method in section 3.5. Once this region was defined, a Multinomial random vector with sample size equal to the total number of organisms for the omitted site was generated. A Multinomial vector was constructed for each MCMC realization in the region. Finally, a convex hull circumscribing the composition of the Multinomial vectors was used to construct the 95% prediction region for the omitted benthic composition.

Figures 11–**??** show the 95% prediction regions for sites #5, #17, and #20. These sites were randomly selected from the five sites $\{5, 17, 20, 21, 22\}$ having 6 neighbors (all other sites had 5 or fewer neighbors).

These figures show that the spatial-regression statistical model does well in predicting

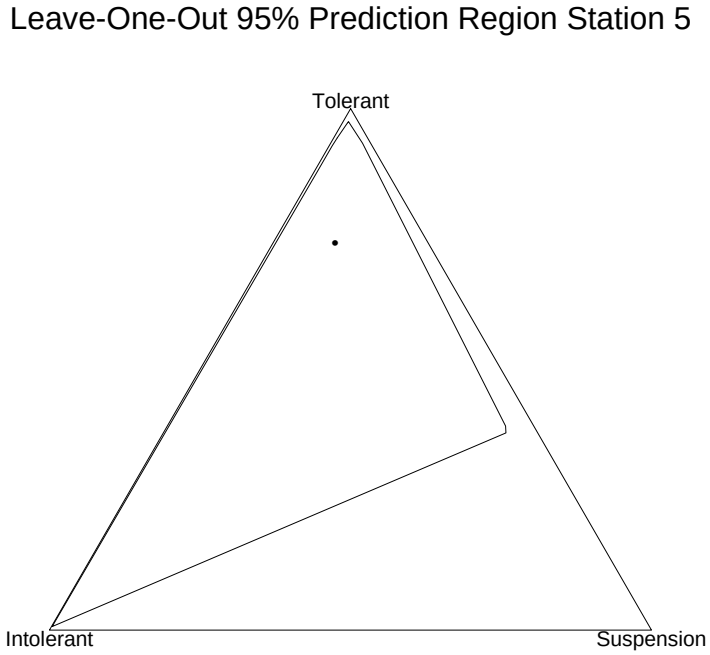

Figure 11: Leave-one-out 95% prediction regions for station 5. The observed composition for sample station 5 is plotted with the prediction region.

compositions at sites with omitted data. All three observed benthic compositions fall in the respective prediction regions.

5.1.2 Regression Residual Diagnostics.

Residuals from the regression on salinity can be computed, and used to examined the adequacy of the form of the salinity covariate in the statistical model. These residuals are not the usual linear regression residuals, but are adapted from the Bayesian model fitting of compositional data. First, the construction of these regression residuals is described. Then, they are used for evaluating the adequacy of the salinity regression model.

The residuals for Bayesian regression for compositional data differ from the usual least squares residuals in two ways: 1) there is no fixed estimate of the regression parameter, but

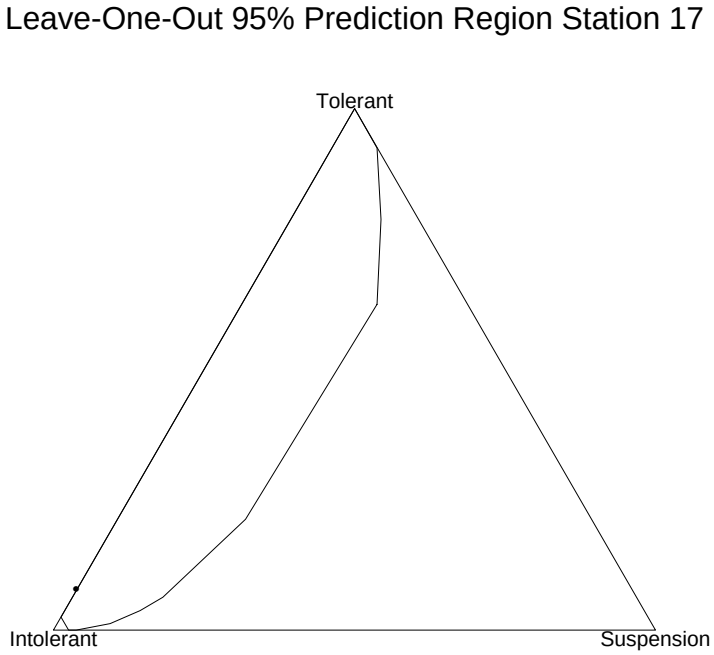

Figure 12: Leave-one-out 95% prediction regions for station 17. The observed composition for sample station 17 is plotted with the prediction region.

a distribution of parameter values, and 2) one cannot simply take the difference between observed and predicted values, because the observed values (the composition of discrete counts) may be on the boundary of the simplex, and the difference is not defined.

The first of these problems is easily fixed by taking the point estimate of the regression composition from the MCMC realizations (the multivariate median). Recall that this point estimate is the mean of the alr$(\cdot)$ transformed compositions in $\Re^2$, expressed as a composition.

The second problem is more difficult. The usual "adjustment" of boundary values, by adding a small constant to all categories and re-normalizing is not adequate because 1) the corresponding "residual" is sensitive to the value of the constant used, and 2) we ignore the discrete observation structure and the information provided by the total number of

organisms observed at each site.

To sidestep this second problem, I use the MCMC point estimate of the site composition for each sample station. Although this is a fitted value, and depends on the regression composition estimate, it is also constrained to be "close" to the actual observation by the Multinomial observation distribution. Further, the point estimate must lie in the interior of the simplex.

The residuals are evaluated on the multivariate logit scale induced by the alr$(\cdot)$ transformation. For regression expressions of the form $\mathbf{z} = \boldsymbol{\xi} \circ \boldsymbol{\gamma}^x$ (described in section **??**), the transformed expression reduces to a linear relationship in two real variables. The residuals can be constructed either via the point estimate perturbed by the inverse of the regression expression $(\boldsymbol{\xi} \circ \boldsymbol{\gamma}^x)$, and then transformed to $\Re^2$, or the alr$(\cdot)$ transformation can be taken first, and the residual is the difference of the resulting vectors.

Figures 13 and 14 show the residuals for the first and second logit components plotted against salinity. These plots do not suggest any substantial lack of fit in the benthic composition regression on salinity. Note that because these are not least squares residuals, they do not have the usual properties of residuals (i.e., they do not sum to zero). Indeed, the residuals for the first logit component sum to 8.3, and the second component sum to 6.0. This may be caused by fitting the regression via compositions, and transforming to the logit scale for diagnosis. However, neither the cause nor the implication of the positive sums is clear.

Finally, note in Figure 14 the large residual associated with a salinity of 17.0 ppt. The value of this residual is 4.9, and is about 2.4 standard deviations above the mean (mean = 0.242, standard deviation = 1.98). This residual corresponds to site #11. From Figure 4, site #11 is located at the mouth of the Delaware river. The composition at this site has a very large proportion of intolerant organisms, and almost no tolerant or suspension feeders.

One possible explanation of the large residual is that site #11 is a "river" site and not a bay site. That is, the relationship between salinity and composition that fits for Delaware Bay may not hold for sites up-river.

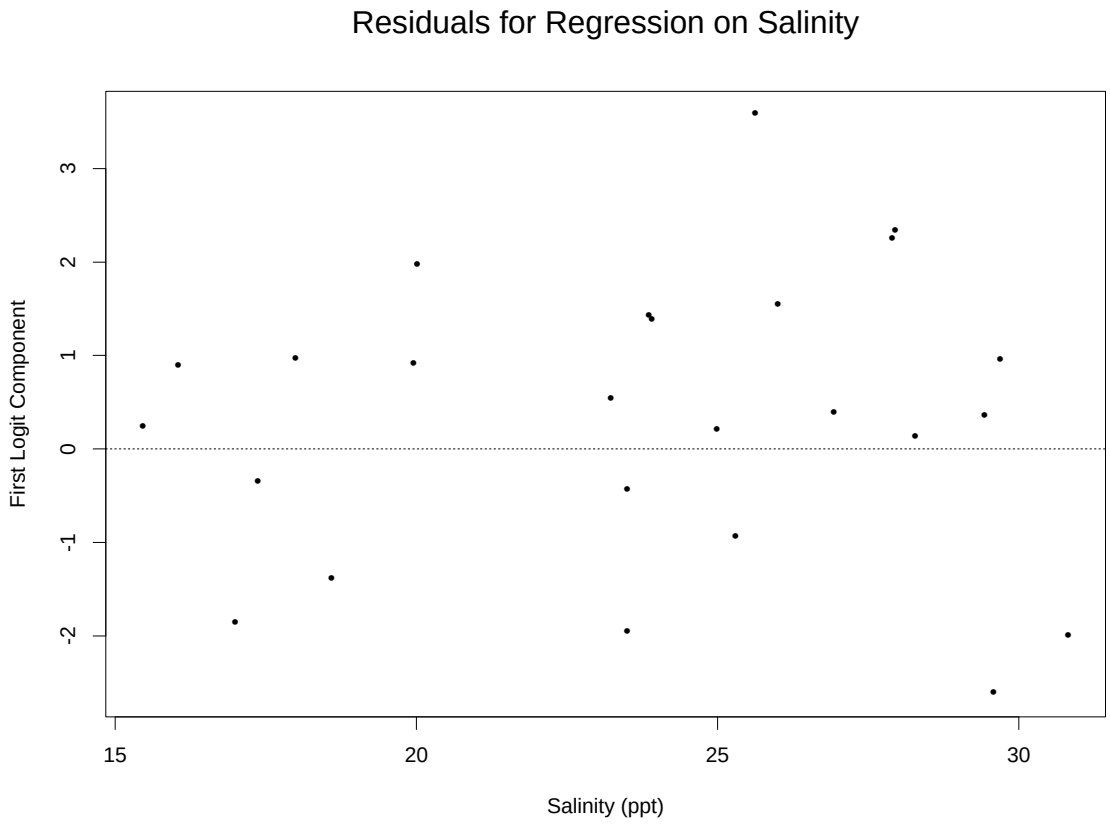

Figure 13: Residuals for first logit component regression on salinity. The residuals do not suggest substantial misspecification in fitting the salinity component.

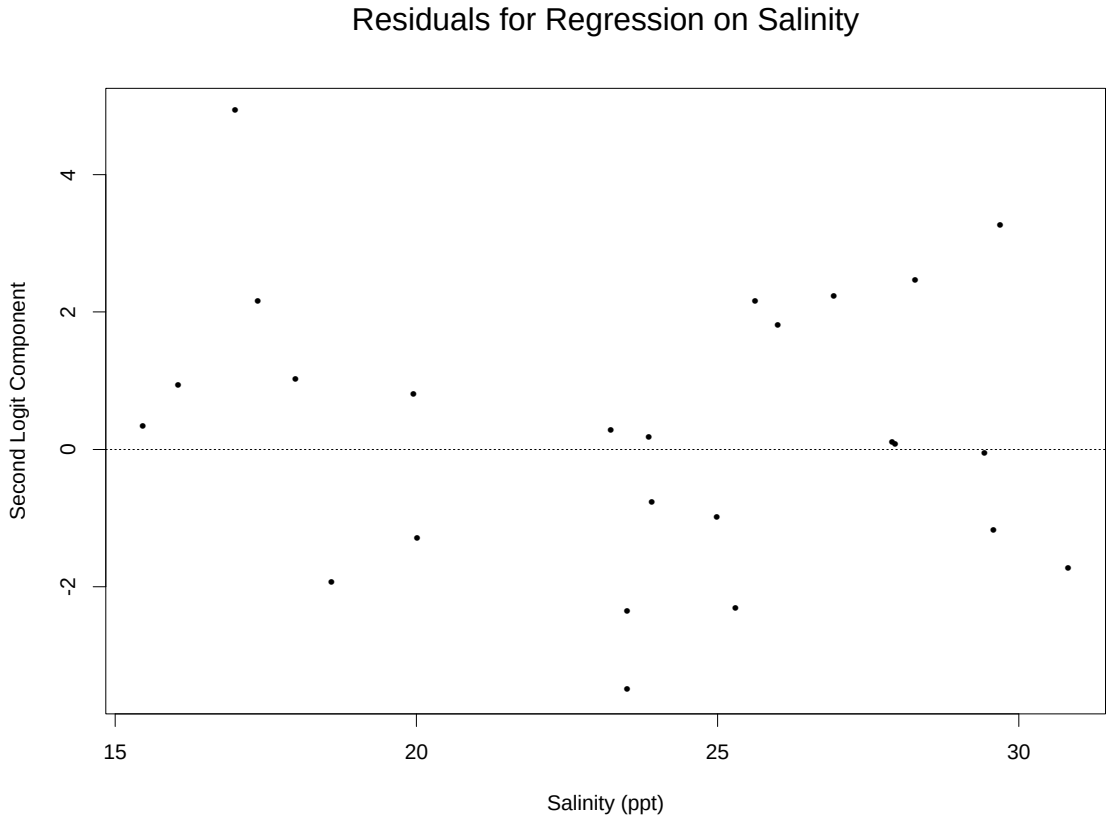

Figure 14: Residuals for second logit component regression on salinity. The residuals do not suggest substantial misspecification in fitting the salinity component.

6 Appendix I

Property 6.1 *For* $\mathcal{I}_{k-1} = (\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k})$, *the operation* $\circ\, \mathcal{I}_{k-1}$ *is the identity operator for any* $\mathbf{u} \in \nabla^{k-1}$, *i.e.,* $\mathbf{u} \circ \mathcal{I}_{k-1} = \mathbf{u}$

Proof.

$$\begin{aligned}
\mathbf{u} \circ \mathcal{I}_{k-1} &= \mathcal{C}\left(u_1\frac{1}{k}, u_2\frac{1}{k}, ..., u_k\frac{1}{k}\right) \\
&= \left(\frac{u_1\frac{1}{k}}{\frac{1}{k}\sum_{i=1}^{k} u_i}, \frac{u_2\frac{1}{k}}{\frac{1}{k}\sum_{i=1}^{k} u_i}, ..., \frac{u_k\frac{1}{k}}{\frac{1}{k}\sum_{i=1}^{k} u_i}\right) \\
&= \mathcal{C}(\mathbf{u}) \\
&= \mathbf{u}
\end{aligned}$$

Property 6.2 *The operation* $\circ$ *is commutative. For* $\mathbf{u}$ *and* $\mathbf{a}$ *in* ∇^{k-1},

$$\mathbf{u} \circ \mathbf{a} = \mathbf{a} \circ \mathbf{u}$$

Proof.

$$\begin{aligned}
\mathbf{u} \circ \mathbf{a} &= \mathcal{C}(u_1a_1, u_2a_2, ..., u_ka_k) \\
&= \mathcal{C}(a_1u_1, a_2u_2, ..., a_ku_k) \\
&= \mathbf{a} \circ \mathbf{u}
\end{aligned}$$

Property 6.3 *The operation* $\circ$ *is associative. For* $\mathbf{u}$, $\mathbf{a}$, *and* $\mathbf{z}$ *in* ∇^{k-1},

$$(\mathbf{u} \circ \mathbf{a}) \circ \mathbf{z} = \mathbf{u} \circ (\mathbf{a} \circ \mathbf{z})$$

Proof.

$$\begin{aligned}
(\mathbf{u} \circ \mathbf{a}) \circ \mathbf{z} &= \mathcal{C}(u_1a_1, u_2a_2, ..., u_ka_k) \circ \mathbf{z} \\
&= \mathcal{C}(u_1a_1z_1, u_2a_2z_2, ..., u_ka_kz_k) \\
&= \mathbf{u} \circ \mathcal{C}(a_1z_1, a_2z_2, ..., a_kz_k) \\
&= \mathbf{u} \circ (\mathbf{a} \circ \mathbf{z})
\end{aligned}$$

Theorem 6.1 ∇^{k-1} *is a vector space with addition defined by the perturbation operator and scalar multiplication defined as* $\mathbf{u}^a = \mathcal{C}(u_1^a, u_2^a, ..., u_k^a,)$ *for the scalar* a.

Proof. To show ∇^{k-1} is a vector space the four following properties must hold.

1. There is an identity scalar multiplier.

   Clearly, $a = 1$ is the identity scalar multiplier.

2. Scalar multiplication is associative,

$$\begin{aligned}
(\mathbf{u}^a)^b &= \left(\mathcal{C}(u_1^a, u_2^a, ..., u_k^a)\right)^b \\
&= \left[\frac{u_1^a}{\sum_{i=1}^k u_i^a}, \frac{u_2^a}{\sum_{i=1}^k u_i^a}, ..., \frac{u_k^a}{\sum_{i=1}^k u_i^a}\right]^b \\
&= \mathcal{C}\left(\left(\frac{u_1^a}{\sum_{i=1}^k u_i^a}\right)^b, \left(\frac{u_2^a}{\sum_{i=1}^k u_i^a}\right)^b, ..., \left(\frac{u_k^a}{\sum_{i=1}^k u_i^a}\right)^b\right) \\
&= \mathcal{C}\left(u_1^{ab}, u_2^{ab}, ..., u_k^{ab}\right) \\
&= \mathbf{u}^{ab}
\end{aligned}$$

3. $(\mathbf{u} \circ \mathbf{z})^a = \mathbf{u}^a \circ \mathbf{z}^a$

$$\begin{aligned}
(\mathbf{u} \circ \mathbf{z})^a &= \left[\mathcal{C}(\mathbf{u} \cdot \mathbf{z})\right]^a \\
&= \left(\frac{u_1 z_1}{\sum_{i=1}^k u_i z_i}, \frac{u_2 z_2}{\sum_{i=1}^k u_i z_i}, ..., \frac{u_k z_k}{\sum_{i=1}^k u_i z_i}\right)^a \\
&= \left(\frac{(u_1 z_1)^a}{\sum_{i=1}^k (u_i z_i)^a}, \frac{(u_2 z_2)^a}{\sum_{i=1}^k (u_i z_i)^a}, ..., \frac{(u_k z_k)^a}{\sum_{i=1}^k (u_i z_i)^a}\right) \\
&= \mathcal{C}\left(\mathbf{u}^a \cdot \mathbf{z}^a\right) \\
&= \mathbf{u}^a \circ \mathbf{z}^a
\end{aligned}$$

4. $\mathbf{u}^{a+b} = \mathbf{u}^a + \mathbf{u}^b$.

$$\begin{aligned}
\mathbf{u}^{a+b} &= \left(\frac{u_1^{a+b}}{\sum_{i=1}^k u_i^{a+b}}, \frac{u_2^{a+b}}{\sum_{i=1}^k u_i^{a+b}}, ..., \frac{u_k^{a+b}}{\sum_{i=1}^k u_i^{a+b}}\right) \\
&= \mathcal{C}\left(\mathbf{u}^a \cdot \mathbf{u}^b\right) \\
&= \mathbf{u}^a \circ \mathbf{u}^b
\end{aligned}$$

Theorem 6.2 *Let* $\mathbf{u}$ *and* $\mathbf{z}$ *be elements of* ∇^{k-1}, *and* $\boldsymbol{\theta} = alr(\mathbf{u})$ *and* $\boldsymbol{\phi} = alr(\mathbf{z})$. *Then* $\langle \mathbf{u}, \mathbf{z} \rangle = \boldsymbol{\theta}' \mathcal{N}^{-1} \boldsymbol{\phi}$ *is an inner product.*

where $\mathcal{N} = I_{k-1} + \mathbf{j}_{k-1} \mathbf{j}_{k-1}'$.

Before proving the theorem, we first show the following proposition.

Proposition

$$\mathcal{N}^{-1} = \left(I_{k-1} \; - \; \frac{1}{k} \mathbf{j}_{k-1} \mathbf{j}_{k-1}^{'} \right)$$

Proof of Proposition. From Rao (1973, p. 33), for $\mathbf{A}$ a nonsingular matrix, and $\mathbf{U}$ and $\mathbf{V}$ two column vectors, then

$$\left(\mathbf{A} \; + \; \mathbf{U}\mathbf{V}^{'} \right)^{-1} = \mathbf{A}^{-1} \; - \; \frac{(\mathbf{A}^{-1}\,\mathbf{U})(\mathbf{V}^{'}\,\mathbf{A}^{-1})}{1 \; + \; \mathbf{V}^{'}\,\mathbf{A}^{-1}\,\mathbf{U}}$$

Applying this result to compute $\mathcal{N}^{-1}$ we obtain the following.

$$\begin{aligned}
\mathcal{N}^{-1} &= \left(I_{k-1} \; + \mathbf{j}_{k-1} \mathbf{j}_{k-1}^{'} \right)^{-1} \\
&= I_{k-1} \; - \; \frac{(I_{k-1}\,\mathbf{j}_{k-1})(\mathbf{j}_{k-1}^{'}\,I_{k-1})}{1 \; + \; \mathbf{j}_{k-1}^{'}\,I_{k-1}\,\mathbf{j}_{k-1}} \\
&= I_{k-1} \; - \; \frac{\mathbf{j}_{k-1}\,\mathbf{j}_{k-1}^{'}}{1 \; + \; (k-1)} \\
&= I_{k-1} \; - \; \frac{1}{k} \mathbf{j}_{k-1} \mathbf{j}_{k-1}^{'}
\end{aligned}$$

Note that $\mathcal{N}^{-1}$ is symmetric.

Proof. To show this operation defines an inner product, the following conditions are required

1. $\langle \mathbf{u}, \mathbf{z} \rangle = \langle \mathbf{z}, \mathbf{u} \rangle$

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{z} \rangle &= \boldsymbol{\theta}^{'}\,\mathcal{N}^{-1}\,\boldsymbol{\phi} \\
&= \left(\boldsymbol{\theta}^{'}\,\mathcal{N}^{-1}\,\boldsymbol{\phi} \right)^{'} \qquad \text{(since the product is a scalar)} \\
&= \boldsymbol{\phi}^{'}\,\mathcal{N}^{-1}\,\boldsymbol{\theta} \\
&= \langle \mathbf{z}, \mathbf{u} \rangle
\end{aligned}$$

2. $\langle \mathbf{u} \circ \mathbf{w}, \mathbf{z} \rangle = \langle \mathbf{u}, \mathbf{z} \rangle + \langle \mathbf{w}, \mathbf{z} \rangle$ for $\mathbf{u}, \mathbf{w}, \mathbf{z} \in \nabla^{k-1}$.

$$\begin{aligned}
\langle \mathbf{u} \circ \mathbf{w}, \mathbf{z} \rangle &= (\boldsymbol{\theta} \; + \; \boldsymbol{\eta})^{'}\,\mathcal{N}^{-1}\,\boldsymbol{\phi} \\
&= \boldsymbol{\theta}^{'}\,\mathcal{N}^{-1}\,\boldsymbol{\phi} \; + \; \boldsymbol{\eta}^{'}\,\mathcal{N}^{-1}\,\boldsymbol{\phi} \\
&= \langle \mathbf{u}, \mathbf{z} \rangle + \langle \mathbf{w}, \mathbf{z} \rangle
\end{aligned}$$

3. $\langle \mathbf{u}^a, \mathbf{z} \rangle = a \langle \mathbf{u}, \mathbf{z} \rangle$ for a scalar a.

   Recall $\boldsymbol{\theta} = \log\left(\frac{\mathbf{u}_{-k}}{u_k}\right)$. So, $\text{alr}(\mathbf{u}^a) = a\,\boldsymbol{\theta}$. The inner product is written as follows:

$$\begin{aligned} \langle \mathbf{u}^a, \mathbf{z} \rangle &= a\,\boldsymbol{\theta}'\,\mathcal{N}^{-1}\,\boldsymbol{\phi} \\ &= a \langle \mathbf{u}, \mathbf{z} \rangle \end{aligned}$$

4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, for all $\mathbf{u} \in \nabla^{k-1}$.

$$\langle \mathbf{u}, \mathbf{u} \rangle = \boldsymbol{\theta}'\,\mathcal{N}^{-1}\,\boldsymbol{\theta}$$

   $\mathcal{N}$ is positive definite, hence $\mathcal{N}^{-1}$ is p.d. This implies

$$\boldsymbol{\theta}'\,\mathcal{N}^{-1}\,\boldsymbol{\theta} \geq 0$$

   for all $\mathbf{u} \in \nabla^{k-1}$.

5. $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ only if $\mathbf{u} = \mathcal{I}_{k-1}$.

   Since $\mathcal{N}^{-1}$ is p.d.

$$\boldsymbol{\theta}'\mathcal{N}^{-1}\boldsymbol{\theta} = 0 \quad \Leftrightarrow \quad \boldsymbol{\theta} = 0_{k-1} \quad \Leftrightarrow \quad \frac{u_i}{u_k} = 1$$

   for all $i = 1, 2, ..., k$. This implies the elements of $\mathbf{u}$ are all equal. Since $\mathbf{u} \in \nabla^{k-1}$,

$$\mathbf{u} = (1/k, 1/k, ..., 1/k) = \mathcal{I}_{k-1}$$

All the conditions are satisfied, and $\langle , \rangle$ defines an inner product on ∇^{k-1}.

Theorem 6.3 *The norm on ∇^{k-1} defined by the inner product is invariant to permutations of the components of* $\mathbf{u}$.

Proof. To show that the norm is invariant to permutations, we will show that it is symmetric in all k components of $\mathbf{u}$. Let $\mathbf{a} = \log(\mathbf{u})$. Then,

$$\boldsymbol{\theta} = \log\left(\frac{\mathbf{u}_{-k}}{u_k}\right) = [I_{k-1} \mid -\mathbf{j}_{k-1}]\,\mathbf{a}$$

We then get the following expression for $\langle \mathbf{u}, \mathbf{u} \rangle$.

$$\begin{aligned}
\langle \mathbf{u}, \mathbf{u} \rangle &= \boldsymbol{\theta}' \, \mathcal{N}^{-1} \, \boldsymbol{\theta} \\
&= \mathbf{a}' \left[\frac{I_{k-1}}{-\mathbf{j}_{k-1}'} \right] \left[I_{k-1} - \frac{1}{k} \mathbf{j}_{k-1} \mathbf{j}_{k-1}' \right] \left[I_{k-1} \mid -\mathbf{j}_{k-1} \right] \mathbf{a} \\
&= \mathbf{a}' \left[\frac{I_{k-1} - \frac{1}{k} \mathbf{j}_{k-1} \mathbf{j}_{k-1}'}{-\mathbf{j}_{k-1}' + \frac{k-1}{k} \mathbf{j}_{k-1}'} \right] \left[I_{k-1} \mid -\mathbf{j}_{k-1} \right] \mathbf{a} \\
&= \mathbf{a}' \left[\begin{array}{c|c} I_{k-1} - \frac{1}{k} \mathbf{j}_{k-1} \mathbf{j}_{k-1}' & -\mathbf{j}_{k-1} + \frac{k-1}{k} \mathbf{j}_{k-1} \\ \hline -\mathbf{j}_{k-1}' + \frac{k-1}{k} \mathbf{j}_{k-1}' & k - 1 - \frac{k-1}{k} (k-1) \end{array} \right] \mathbf{a} \\
&= \mathbf{a}' \left[\begin{array}{c|c} I_{k-1} - \frac{1}{k} \mathbf{j}_{k-1} \mathbf{j}_{k-1}' & -\frac{1}{k} \mathbf{j}_{k-1} \\ \hline -\frac{1}{k} \mathbf{j}_{k-1}' & 1 - \frac{1}{k} \end{array} \right] \mathbf{a} \\
&= \mathbf{a}' \left[I_k - \frac{1}{k} \mathbf{j}_k \mathbf{j}_k' \right] \mathbf{a}
\end{aligned}$$

This shows that $\|\mathbf{u}\|^2$ is symmetric in all components of $\mathbf{a} = \log(\mathbf{u})$. So, the norm defined by the inner product above is invariant to the ordering of the elements of $\mathbf{u}$.

Theorem 6.4 *∇^{k-1} is a Hilbert space (a complete, inner product space).*

Proof. It remains to show completeness of the space. That is, we require that every Cauchy sequence, $\{\mathbf{z}_n\} \in \nabla^{k-1}$, converges in ∇^{k-1}.

Suppose $\{\mathbf{u}_n\} \in \nabla^{k-1}$ is a Cauchy sequence. Then, for every $\epsilon > 0$, there is an integer, N, such that $m, n > N$ imply $\|\mathbf{u}_m \circ \mathbf{u}_n^{-1}\| < \epsilon$.

Let $\boldsymbol{\theta}_n = \text{alr}(\mathbf{u}_n)$. Then $\boldsymbol{\theta}_n \in \Re^{k-1}$ for all n. Note the for the norm defined above

$$\begin{aligned}
\|\mathbf{u}_m \circ \mathbf{u}_n^{-1}\|^2 &= (\boldsymbol{\theta}_n - \boldsymbol{\theta}_m)' \, \mathcal{N}^{-1} \, (\boldsymbol{\theta}_n - \boldsymbol{\theta}_m) \\
&= (\boldsymbol{\theta}_n - \boldsymbol{\theta}_m)' \left[I_{k-1} - \frac{1}{k} \mathbf{j}_{k-1} \mathbf{j}_{k-1}' \right] (\boldsymbol{\theta}_n - \boldsymbol{\theta}_m) \\
&\leq (\boldsymbol{\theta}_n - \boldsymbol{\theta}_m)' \, (\boldsymbol{\theta}_n - \boldsymbol{\theta}_m)
\end{aligned}$$

with equality holding only when $\boldsymbol{\theta}_n - \boldsymbol{\theta}_m$ is equal to the zero vector. Note that this final expression, $(\boldsymbol{\theta}_n - \boldsymbol{\theta}_m)' \, (\boldsymbol{\theta}_n - \boldsymbol{\theta}_m) = \sum_{i=1}^{k-1} (\theta_{ni} - \theta_{mi})^2$ is the square of the usual L^2 norm for vectors in $\Re^{k-1}$. By the completeness of $\Re^{k-1}$ (under L^2 norm), the limit of $\{\boldsymbol{\theta}_n - \boldsymbol{\theta}_m\} \in \Re^{k-1}$. By the inequality above, limit points under the L^2 norm are also limits under the norm defined above. Further, since the $\text{alr}(\cdot)$ transform is bijective all

limit points in $\Re^{k-1}$ can be transformed to points in ∇^{k-1}. Hence, any Cauchy sequence in ∇^{k-1} (as measured by the norm defined on ∇^{k-1}) has a limit in ∇^{k-1}, and ∇^{k-1} is complete. So, ∇^{k-1}, with the perturbation operator and scalar multiplication is a Hilbert space.