Statistical Analysis and Interpretation of Discrete Compositional Data

Dean Billheimer **Peter Guttorp** **William F. Fagan**

NRCSE

Technical Report Series

NRCSE-TRS No. 011

Statistical Analysis and Interpretation of Discrete Compositional Data

Dean Billheimer *
The Boeing Company, Seattle
Peter Guttorp
University of Washington, Seattle
William F. Fagan
Arizona State University, Tempe, AZ

6 March 1998

Abstract

A composition is a vector of proportions describing the contribution of each of k components to the whole. We introduce an algebra for compositions that provides a natural definition for additive statistical models. The algebra eases interpretation of treatment effects, treatment interactions, and covariates. Our developments extend the logistic normal modeling framework of Aitchison (1982, 1986), and further extend Aitchison's approach to incorporate discrete observations present in many applications (i.e., counts of objects in different groups). We demonstrate these methods in two examples. The first is a designed experiment evaluating the effect of omnivory on the recovery of arthropod communities to disturbance. The second evaluates the natural variability and spatial dependence of benthic invertebrate communities in the Delaware Bay.

1 Introduction

Compositional data are vectors of proportions describing the relative contributions of each of k categories to the whole. Mathematically, $\mathbf{z} = (z_1, z_2, ..., z_k)'$, where $z_i > 0$, for all $i = 1, 2, ..., k$ and $\sum_{i=1}^{k} z_i = 1$. The summation constraint and bounded support require special techniques for compositional data. Aitchison (1982, 1986) introduces the logistic normal (LN) distribution as a framework for analysis of compositional data. This

*We are grateful to Melissa Hughes of EPA Delaware and Tony Olsen of EPA Corvallis for making these data available. This research had partial support from the United States Environmental Protection Agency under a cooperative agreement with the University of Washington. This document has not undergone Agency review, and reflects in no way official Agency policy.

techniques assumes multivariate normality of additive log-ratio transformed data. Thus, the inference tools developed for multivariate normal data can be applied to the transformed compositions. Unfortunately, as Aitchison (1986) and others (e.g., Pawlowsky and Burger, 1992) describe, interpretation of parameter estimates on the multivariate log-odds scale is difficult. Specifically, location parameters are $\mu_i = E(\log(z_i/z_k))$ for $i = 1, 2, ..., k-1$, and elements of the covariance matrix, $\sigma_{ij} = \text{cov}(log(z_i/z_k), log(z_j/z_k))$. It is often challenging to interpret these parameters (or their estimates) in terms of the motivating scientific problem.

We present parameterization and analysis tools for improved interpretation of statistical modeling results. Foremost is an algebra for compositions that includes addition, scalar multiplication, and a norm. In turn, these tools provide intuitive definitions for additive error, introduction and interpretation of covariates, and distances between composition vectors. We demonstrate these methods in two applications. In addition to the analysis tools, we extend Aitchison's approach to problems with discrete data, and to spatially dependent data. This is accomplished via a hierarchical model structure, and a discrete observation distribution.

The logistic normal (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 establishes many of its mathematical and statistical properties. These results include the perturbation operation (section 2) and the LN's relevance as a limit distribution for compositions. A key benefit of the LN distribution is its ability to model complicated covariance structure among the k categories. 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 categorical data. For categorical spatial data, Upton and Fingleton (1989) and Cerioli (1992) focus on the analysis of spatial contingency tables. They consider observations on a regular spatial grid, and assume a single multinomial observation at each site. The multinomial parameter vector is assumed 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.

Mardia (1988) introduces a Markov random field approach for multivariate normal observation vectors. He considers an example using logistic transformed compositions as bivariate observations. (The original compositions were derived from Landsat classification). Mardia's emphasis is to illustrate conditional autoregressive (CAR) methods, and to estimate the spatial dependence parameter. It appears that the compositional nature of the data is merely a nuisance, and is quickly remedied via the logistic transformation. There is no attempt to interpret the results in terms of the original compositions.

Pawlowsky and Burger (1992) use the logistic transformation 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 logistic transformation converts the sum-constrained 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 must be formulated in terms of logratios (i.e., the log of the ratio of proportions), and interpretation and description of spatial dependencies are also on the same scale.

Albert and Chib (1993) address the problem of a categorical response regression model using a hierarchical Bayesian model formulation. They formulate a probit regression model for binary outcomes (and ordered and unordered multinomial outcomes) with an underlying normal regression structure on latent continuous data. Data augmentation is combined with Gibbs sampling to approximate the posterior distribution of regression parameters.

Allenby and Lenk (1994) use logistic normal regression models to relate covariates to household purchase decisions. They incorporate random effects and serial correlation to describe household purchase behavior. These authors also use a hierarchical Bayes formulation and Gibbs sampling for inference. They apply their model to scanner panel data for ketchup purchases to explore household preference, brand switching, and dependence on past purchases. We note that Allenby and Lenk interpret their modeling results on the logit scale with respect to a fixed category. Further, parameter estimates are interpreted qualitatively, and with respect to the baseline brand.

The remainder of this paper is organized as follows. The next section reviews the LN distribution and Aitchison's (1982) perturbation operator. It also introduces an algebra for composition vectors and demonstrates its operations on compositions. Section 3 illustrates how the algebra, along with graphical analysis tools can be used to interpret and visualize statistical modeling activities. Section 4 describes the coupling of the conditional multinomial observation model with the logistic normal. This hierarchical approach is used in the applications of sections 5 and 6. Finally, section 7 discusses issues associated with this modeling approach.

2 Logistic Normal Distribution and an Algebra for Compositions

Aitchison (1986) describes statistical analysis methods for compositional data with independent observations. These methods rely on the additive logratio transform ($\text{alr}(\cdot)$) to take observations from the $(k-1)$-dimensional simplex (∇^{k-1}) to $(k-1)$-dimensional Euclidean space ($\Re^{k-1}$). The additive logratio transform of $\mathbf{z} \in \nabla^{k-1}$ to $\Re^{k-1}$ is defined as

$$\text{alr}(\mathbf{z}) = \left[\log\left(\frac{z_1}{z_k}\right), \log\left(\frac{z_2}{z_k}\right), \cdots, \log\left(\frac{z_{k-1}}{z_k}\right)\right]$$

This transformation is a bijection with inverse transformation denoted by alr^{-1}. Aitchison (1986) terms the inverse transformation the additive logistic transform.

Aitchison models the transformed data via the $(k-1)$ multivariate normal distribution. This transformation

and assumption of multivariate normality define a distribution on ∇^{k-1}: the logistic normal (LN) distribution. Aitchison further describes that the rich covariance structure of the multivariate normal distribution transfers to the logistic normal, and allows positive or negative covariances between pairs of the k elements of the composition. The density function is

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

where

$$\boldsymbol{\theta} = \text{alr}(\mathbf{z}) = \log\left(\frac{\mathbf{z_{-k}}}{z_k}\right) = \left[\log\left(\frac{z_1}{z_k}\right), \log\left(\frac{z_2}{z_k}\right), ..., \log\left(\frac{z_{k-1}}{z_k}\right)\right]'$$

for $\mathbf{z} \in \nabla^{k-1}$ and $\mathbf{z}_{-k} = (z_1, z_2, ..., z_{k-1})$ is a vector containing the first $(k-1)$ components of $\mathbf{z}$. We denote the density function by $L^{k-1}(\boldsymbol{\mu}, \Sigma)$, It is clear that the parameters depend on the ordering of the k elements of $\mathbf{z}$. However, Aitchison (1986) shows that the density is invariant with respect to permutations of the components. He also establishes other properties and moments of this distribution.

Associated with 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 ($\boldsymbol{\alpha}$). For $\boldsymbol{\xi}, \boldsymbol{\alpha} \in \nabla^{k-1}$,

$$\mathbf{z} = \boldsymbol{\xi} \oplus \boldsymbol{\alpha} = \left(\frac{\xi_1\alpha_1}{\sum_{i=1}^{k} \xi_i\alpha_i}, \frac{\xi_2\alpha_2}{\sum_{i=1}^{k} \xi_i\alpha_i}, \cdots, \frac{\xi_k\alpha_k}{\sum_{i=1}^{k} \xi_i\alpha_i}\right)$$

and $\mathbf{z} \in \nabla^{k-1}$. The vector, $\boldsymbol{\alpha}$, need not be an element of ∇^{k-1} for the perturbation operator to be defined. It is sufficient that $\alpha_i > 0$ for all $i = 1, 2, ..., k$. Note that the perturbation operator leads to the LN distribution as the limit distribution of a sequence of independent perturbations (Aitchison, 1986; p. 124).

Aitchison (1986) shows a number of properties of the perturbation operator including an inverse perturbation, and an identity element

$$\mathcal{I}_{k-1} = \left(\frac{1}{k}, \frac{1}{k}, \cdots, \frac{1}{k}\right)$$

Finally, he defines a power-transformation for compositions (Aitchison 1986, p. 120).

Algebra for Compositions

Clearly, we may consider the perturbation operator to define an addition operator for compositions. Further, the power transformation allows us to define scalar multiplication of a composition $\mathbf{z}$ by a scalar a by,

$$\mathbf{z}^a = \left(\frac{z_1^a}{\sum_{i=1}^{k} z_i^a}, \frac{z_2^a}{\sum_{i=1}^{k} z_i^a}, \cdots, \frac{z_k^a}{\sum_{i=1}^{k} z_i^a}\right)$$

We show that ∇^{k-1} equipped with the perturbation operator and scalar multiplication constitutes complete inner product space. (See Appendix I for details.) This additional abstraction allows the Cauchy-Schwartz

and triangle inequalities, and the definition of a norm on ∇^{k-1}. In turn, these constructs allow us to interpret operations on compositions in terms of the component proportions. First we show the inner product and norm, and then describe interpretation of parameters.

Definition 2.1 *For* $\mathbf{u}, \mathbf{z} \in \nabla^{k-1}$, *let* $\boldsymbol{\theta} = alr(\mathbf{u})$, *and* $\boldsymbol{\phi} = alr(\mathbf{z})$. *Define by*

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

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

Here, $\mathcal{N} = \left[I_{k-1} + \mathbf{j}_{k-1}\mathbf{j}_{k-1}' \right]$, 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

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

Definition 2.2 *Define the norm for* $\mathbf{u} \in \nabla^{k-1}$, $\|\mathbf{u}\|$, *by* $\langle \mathbf{u}, \mathbf{u} \rangle^{1/2}$.

Note that the inner product and norm are invariant to permutations of components of $\mathbf{u}$ (See Appendix I for details).

Differences Between Compositions

The definition of an (inverse) addition operation and a norm allow us to measure the difference between compositions. For demonstration, consider three compositions in ∇^2, $\mathbf{z}_1 = \mathcal{I}_2 = (1/3,\, 1/3,\, 1/3)$, $\mathbf{z}_2 = (0.80, 0.10, 0.10)$, and $\mathbf{z}_3 = (0.98, 0.01, 0.01)$. For reference, we show these compositions in the ternary diagram of figure 1.

*** figure 1 about here ****

We first note the norms of these compositions are

$$\|\mathbf{z}_1\| = 0, \qquad \|\mathbf{z}_2\| = 1.698, \text{ and} \qquad \|\mathbf{z}_3\| = 3.744$$

Thus, the defined norm measures the distance of a composition from $\mathcal{I}_{k-1}$, the "center" of ∇^{k-1}.

Next, using the inverse of the perturbation operator, we find the difference between pairs $\mathbf{z}_1$ and $\mathbf{z}_2$, and $\mathbf{z}_2$ and $\mathbf{z}_3$. To find the difference between two compositions we perturb the second by the elementwise inverse of the first. That is,

$$\mathbf{z}_2 \ominus \mathbf{z}_1 \;=\; \mathbf{z}_2 \oplus \mathbf{z}_1^{-1} \;=\; \mathbf{z}_2$$

since $\mathbf{z}_1$ is the identity element. Similarly,

$$\mathbf{z}_3 \ominus \mathbf{z}_2 = \left(\frac{[z_3]_1 [z_2]_1^{-1}}{\sum_{i=1}^{3} [z_3]_i [z_2]_i^{-1}}, \frac{[z_3]_2 [z_2]_2^{-1}}{\sum_{i=1}^{3} [z_3]_i [z_2]_i^{-1}}, \frac{[z_3]_3 [z_2]_3^{-1}}{\sum_{i=1}^{3} [z_3]_i [z_2]_i^{-1}} \right)$$

$$= (0.860,\ 0.070,\ 0.070)$$

where $[\mathbf{z}_i]_j$ is the j^{th} element of the composition $\mathbf{z}_i$. Thus, $(0.86, 0.07, 0.07)$ is the composition by which we need to perturb $\mathbf{z}_2$ to obtain $\mathbf{z}_3$. By taking the norm of the difference composition, we measure the distance between $\mathbf{z}_2$ and $\mathbf{z}_3$.

$$\|\mathbf{z}_3 \ominus \mathbf{z}_2\| = \|\,(0.86,\ 0.07,\ 0.07)\,\| = 2.046$$

Note that the distance from $\mathbf{z}_1$ to $\mathbf{z}_2$ is 1.698, while the distance from $\mathbf{z}_2$ to $\mathbf{z}_3$ is larger at 2.046. This demonstrates two points,

1. Interpretation of distances between compositions is difficult without a carefull definition of a norm.
2. Graphical interpretation in the simplex (e.g., ternary diagram) is complicated by the compression of distances near the boundaries of the simplex.

An (invertible) addition operation and norm allow interpretation of differences in compositions. Specifically, if $\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\xi}}_2$ are estimated location parameter vectors for treatments 1 and 2, resectively, we may easily obtain information about the direction and distance between them.

3 Interpretation and Visualization of Parameters

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

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

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

Interpretation of $\boldsymbol{\xi}$ is much simpler than for $\boldsymbol{\mu}$ on the multivariate logit scale. However, some of the statistical properties of $\boldsymbol{\mu}$ are lost with the transformation to the simplex. Specifically, $\boldsymbol{\mu}$ is the mean and mode of the multivariate normal logit (i.e., the alr($\mathbf{z}$)). The $\text{alr}^{-1}(\cdot)$ transform does not preserve these properties. However, $\text{alr}^{-1}(\cdot)$ transform is monotone in each of the $(k-1)$ components of $\boldsymbol{\mu}$ (Billheimer and Guttorp, 1995). As a consequence, ordering of values is preserved under this transformation. Hence, $\boldsymbol{\xi} = \text{alr}^{-1}(\boldsymbol{\mu})$ can be interpreted as a component-wise multivariate median for the LN distribution in ∇^{k-1}. As is shown in the sequel, this interpretation is a useful characterization for point estimates of parameters, and as a "center" for the asymmetric LN distribution.

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 x_j, indexed by $j = 1, 2, ..., n$ observations, $\boldsymbol{\mu}_j$ 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,

$$\text{alr}^{-1}(\boldsymbol{\mu}_j) = \text{alr}^{-1}(\boldsymbol{\beta}_0) \oplus \text{alr}^{-1}(\boldsymbol{\beta}_1)^{(x_j - \bar{x})}$$

We write this more compactly as

$$\boldsymbol{\xi}_j = \boldsymbol{\xi} \oplus \boldsymbol{\gamma}^{u_j}$$

where $\boldsymbol{\xi}_j = \text{alr}^{-1}(\boldsymbol{\mu}_j)$, $\boldsymbol{\xi} = \text{alr}^{-1}(\boldsymbol{\beta}_0)$, and $\boldsymbol{\gamma} = \text{alr}^{-1}(\boldsymbol{\beta}_1)$. 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 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. An illustration of this model is given later in this section. Finally, deviations in $\boldsymbol{\gamma}$ from the identity composition, $\mathcal{I}_{k-1}$, indicate the direction and magnitude of the change. Note that $\boldsymbol{\gamma} = \mathcal{I}_{k-1}$ implies the covariate has no effect on the composition location. Figure 2 shows the curves of $\boldsymbol{\xi}_j = \boldsymbol{\xi} \oplus \boldsymbol{\gamma}^{u_j}$ for different values of $\boldsymbol{\gamma}$.

*** Figure 2 about here. ***

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.

Additive Statistical Models

To illustrate the usefulness of the algebra developed in section 2 for interpreting compositional data, we present three simple, additive statistical models. Suppose $\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n$ are elements of ∇^2, and are independent realiza-

tions from the following statistical model

$$\mathbf{z}_j = \boldsymbol{\xi} \oplus \boldsymbol{\epsilon}_j \quad \text{where } \boldsymbol{\epsilon}_j \sim LN(\mathbf{0}_2, \Sigma)$$

Here, $\mathbf{0}_2$ is a vector of zeros of length 2, and Σ is a two-by-two variance-covariance matrix.

Figure 3 shows 30 realizations from this model with $\boldsymbol{\xi} = (0.7,\ 0.2,\ 0.1)$ and $\Sigma = 0.2\mathcal{N}$. (Recall that $\mathcal{N} = \left[I_{k-1} + \mathbf{j}_{k-1}\mathbf{j}_{k-1}^{'}\right]$.)

*** Figure 3 about here. ***

We can estimate $\boldsymbol{\xi}$ via maximum likelihood (Aitchison, 1986), and compute the residuals by "subtracting" the estimate from each observation.

$$\boldsymbol{\epsilon}_j = \mathbf{z}_j \ominus \hat{\boldsymbol{\xi}}$$

where $\hat{\boldsymbol{\xi}}$ denotes the maximum likelihood estimator. The residuals are shown in figure 4. Note that centering the observations removes the (visual) effect of location ($\boldsymbol{\xi}$) on the shape of the residual distrubution. It is also straightforward to estimate Σ from the residuals by maximum likelihood.

*** Figure 4 about here. ***

As discussed earlier, we may also allow the location to depend on a (centered) covariate,

$$\boldsymbol{\xi}_j = \boldsymbol{\xi} \oplus \boldsymbol{\gamma}^{u_j}$$

Figure 5 shows this "linear" relationship for 21 realizations from the logistic normal model described above, and $\boldsymbol{\gamma} = (0.40,\ 0.35,\ 0.25)$. The covariate takes integer values between -10 and 10, inclusive, and are displayed beside their respective observations. The maximimum likelihood estimated regression line is denoted by the solid line.

*** Figure 5 about here. ***

The effect of the covariate, expressed as a composition, is directly interpretable graphically and by comparison with $\mathcal{I}_2 = (0.333,\ 0.333,\ 0.333)$. We see that large positive values of the covariate are associated with a high

proportion of component 1. Conversely, negative values correspond to smaller proportions of component 1, and relatively higher proportions of components 2 and 3. The effect of the covariate is clearly non-linear as depicted in the ternary diagram. However it is linear on the log-odds scale. For this reason we believe departures from typical regression assumptions are more easily detected on the logistic transformed scale.

Aitchison (1986) suggests using the perturbation operator to define a Markov chain for compositional data. Billheimer (1995) extends this approach to autoregressive and seasonal time series models. We provide an AR(1) model example to illustrate how the composition algebra is useful for defining time series models. Figure 6 shows realizations of AR(1) processes from the model

$$\mathbf{z}_t = (\mathbf{z}_{t-1})^{\phi} \oplus \boldsymbol{\epsilon}_t$$

for ϕ values of 0.2, 0.6, 0.95, and 1 (independent increments), respectively. All realizations begin at the center of the simplex, $\mathcal{I}_2$, and use identical sequences of errors from the error distribution described above. Identical errors allow us to focus on the effect of the autoregressive parameter.

*** Figure 6 about here. ***

For small values of ϕ, values of $\mathbf{z}_t$ are clustered around the initial starting value (also the location parameter for this model). The effect of the AR parameter is to shrink the location for subsequent observations toward the center of the simplex (the unconditional location for the process). Small values of ϕ have a greater shrinking effect. As ϕ increases the paths wander farther from the initial location. Finally, for $\phi = 1$, the increments $\mathbf{z}_t \ominus \mathbf{z}_{t-1}$ are independent realizations from $L_2(\mathbf{0}_2, \Sigma)$. While this type of dependence on ϕ holds for all autoregressive models, its effect is more noticable for time series not defined on the real line ($\Re^1$) and centered at zero.

Graphical Tools for Higher Dimensions

In our approach to analysis of compositional data, we make extensive use of graphical tools to aid visualization of data and results. The ternary diagrams (above) have long been used in soil science for displaying the relative amounts of a three-part composition. Their structure is easily derived by considering the 2-dimensional submanifold in $\Re^3$ defined by positive values constrained to sum to 1.

Extending the graphical structure to the 3-dimensional simplex (a tetrahedron) for a four-part composition is useful in a dynamic graphics environment. Graphics software allowing brushing and rotation (e.g., Xgobi, Swayne, et al., 1991) allow effective visualization in this high dimensional space. Unfortunately, generalization to higher dimensional simplices is limited by our knowledge of how to do such things.

To aid visualization of compositions with more than 3 groups, we have developed a graphical tool we call a webplot (arachnid, not world-wide). The structure of this tool is similar to the star-plot available in Splus (Statistical Sciences, Inc., 1995), and is constructed by representing each element of a k-part composition as a radial distance from the origin. The radii are arrayed, with equal angle spacing, around the 360^o of a circle. A line segment joins contiguous radii. Thus, each k-part composition is represented by a k-sided polygon. Figure 7a shows a 5-part composition (0.40, 0.20, 0.10, 0.05, 0.25).

*** Figure 7 about here. ***

Figure 7b shows realizations from two distributions with different median vectors, but identical variance-covariance structure. Population 1, denoted by the dark lines has median (0.40, 0.20, 0.10, 0.05, 0.25), while the median for population 2 (light lines) is (0.2, 0.2, 0.2, 0.2, 0.2).

Both have variance covariance matrix

$$\Sigma = 0.1\,\mathcal{N} = 0.1 \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

4 State-Space Model for Discrete Compositions

In the two following applications we combine the logistic normal model for continuous compositions with a conditional multinomial observations distribution. Briefly, we posit a latent compositions vector associated with each sampled observation. That is, for observation $\mathbf{y}_j$, a k-vector of counts from site j, given $m_j = \sum_{i=1}^{k} [\mathbf{y}_j]_i$, and $\mathbf{z}_j$ (an unobservable composition vector), the probability mass function for $\mathbf{y}_j$ is

$$p(\mathbf{y}_j \mid \mathbf{z}_j, m_j) \;=\; \frac{m_j!}{\prod_{i=1}^{k} [y_j]_i!} \prod_{i=1}^{k} [\mathbf{z}_j]_i^{[y_j]_i}$$

where $[\cdot]_i$ denotes the i^{th} element of the vector. We suppose $\mathbf{z}_j \sim L_{k-1}(\boldsymbol{\mu}_j, \Sigma_j)$. Covariate and spatial dependence are introduced via the logistic normal location parameter vector.

Markov chain Monte Carlo (MCMC) is used for inference about the unkonwn logistic normal population parameters and the unobservable latent vectors. The MCMC is performed in a Bayesian setting. In the first application (section 5), the observations are obtained from a designed experiment with factorial treatment structure. Each treatment corresponds to a (possibly) different location parameter vector. The second application

(section 6) is from a biological monitoring study. Here we seek to evaluate natural variability in benthic invertebrate populations, and identify important covariates. In addition, we anticipate spatial dependence to be present between sample sites.

Vegetation Disturbance	Predator Manipulation		
	Increased Omnivores	Increased Specialists	Control Level
50% Removal	OV	SV	CV
Control	OC	SC	CC

Table 1: Treatment structure for arthropod community stability experiment. The codes OV, SV, etc., denote the predator–vegetation factor treatment combinations.

5 Stability of Arthropod Food Webs – Analysis of a Designed Experiment

Here, we demonstrate the LN-Multinomial model for analysis of independent observations from a designed experiment. We show that the algebraic and graphical methods presented in sections 2 and 3 allow direct interpretation of the analysis results to address the scientific questions of interest.

A series of experiments were conducted (Fagan, 1996, 1997) to evaluate the factors affecting the stability of arthropod communities in the presence of environmental disturbance. Classical food web models (e.g., Pimm and Castor, 1978) predict that omnivory – defined broadly as feeding on multiple trophic levels (Pimm, 1982; Menge and Sutherland, 1987; Polis, 1994) – destabilizes ecological communities. However, recent empirical evidence (Strong, 1992; Fagan, 1997) suggests that omnivory is a stabilizing factor in reticulate food webs. We use Fagan's (1997) definition of ecological stability as the capacity of a community to recover from an external disturbance (i.e., a "shock" to the system), and measure the community's response by the relative abundance of individuals in different trophic classes. The critical question to be addressed is, "How does the degree of omnivory in an ecological assemblage influence its recovery from an environmental disturbance?"

The experimental protocol is described by Fagan (1996). In summary, five plots are assigned to each of six (6) experimental treatments. The treatment design is a two-way factorial design with predator manipulation (3 levels) and vegetation disturbance (2 levels) as the factors. The levels of the predator manipulation are increased dominance by omnivores (*Pardosa*, wolf spiders), increased dominance by specialist predators (*Nabis* bugs) and no change (control). The vegetative disturbance consists of removing 50% of the existing fireweed (*Epilobuim*) and pearly-everlasting (*Anaphalis*). The treatment structure is summarized in Table 5. The treatments are assigned to plots using a completely random design. All increased predator densities are within the naturally occurring density range.

In control plots, removing vegetation tends to increase the abundance of several herbivorous species (Fagan,

1996). Both specialist and generalist feeding herbivores increase in abundance, but specialists tend to increase more. These herbivores seem attracted to, and may enjoy greater survival in, areas with decreased plant density. Futhermore, decreased plant density may allow greater productivity through increased growth of the remaining plants. When large numbers of omnivorous spiders are present, these effects on community composition are hypothesized to be reduced or eliminated. Because these spiders eat both specialist and generalist feeding herbivores, they limit increases in the abundance of these group, preventing compositional shifts.

The goal of the experiment is to evaluate whether the disturbed treatment (OV) exhibits a species composition similar to the double control treatment (CC), while the disturbed treatments dominated by specialists or featuring the natural predator assemblage (SV, CV) show compositions different from CC. This goal can be evaluated by assessing the predator–vegetative disturbance interactions (i.e., the difference in vegetative disturbance treatments across the different predator treatments).

Assume that for each observation (plot), there is a latent species composition with k components, $\mathbf{z} \in \nabla^{k-1}$. Conditionally on this composition, the observation for plot j from treatment t (say), is Multinomial($m_{tj} = \sum_{i=1}^{k} [y_{tj}]_i, \mathbf{z}^{tj}$). The compositions are modeled as independent from $L^{k-1}(\boldsymbol{\mu}_{tj}, \Sigma)$. Note that explanatory variables at the plot level can be included in the mean structure as

$$\boldsymbol{\mu}_{tj} = \boldsymbol{\mu}_t + \boldsymbol{\beta}(x_j - \bar{x})$$

For the remainder of this section, we assume that $\boldsymbol{\mu}_{tj} = \boldsymbol{\mu}_t,$ for all $j = 1, 2, ..., n_t$ plots with treatment t. This specification gives the likelihood

$$L(\mathbf{y},\, \mathbf{z} \mid \mathbf{m},\, \boldsymbol{\mu}, \Sigma) = \prod_{t=1}^{T} \left[\prod_{j=1}^{n_t} \left(\frac{m^{tj}!}{\prod_{i=1}^{k} [y_{tj}]_i!} \prod_{i=1}^{k} ([z_{tj}]_i)^{([y_{tj}]_i - 1)} \left(\frac{1}{2\pi}\right)^{\frac{k-1}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \right.\right.$$
$$\left.\left. \times \exp\left[-\frac{1}{2} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t)' \Sigma^{-1} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t) \right] \right) \right]$$

where $\boldsymbol{\theta}_{tj} = \text{alr}(\mathbf{z}_{tj})$.

The model formulation is completed by specifying prior distributions for $\boldsymbol{\mu}_t$ and Σ. Let $\boldsymbol{\mu}_t$ have a $(k-1)$-dimensional Multivariate Normal distribution with mean vector $\boldsymbol{\eta}$, and variance-covariance matrix Ω. Further, assume that $\Sigma^{-1} \sim \text{Wishart}(\Psi^{-1}, \rho)$, where Ψ is a $(k-1) \times (k-1)$ positive definite matrix, and ρ denotes the degrees of freedom. Typical choices for the hyperparameters are

$$\boldsymbol{\eta} = 0_{k-1} \quad \Omega = a\mathcal{N} \text{ and } \Psi = c\mathcal{N}$$

and

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

where 0_{k-1} is a $(k-1)$-vector of 0's, I_{k-1} is a $(k-1)$ identity matrix, $\mathbf{j}_{k-1}$ is a $(k-1)$-vector of ones, and a and c are scalars.

For hyperconstants, typical values are $a = 0.5$, $c = 0.1$, and $\rho = k - 1$. The value of a is selected to allow the 95% prior probability contour for $\boldsymbol{\xi}_t = \text{alr}^{-1}(\boldsymbol{\mu}_t)$ to reach at least 0.05 for each component. The value of c is chosen so that the observed variance of simulated compositions approximates that observed in the data. The value of ρ is the smallest allowable (least informative) that still maintains a proper Wishart distribution. I let $\mathcal{N}$ denote the "null" variance-covariance matrix as defined in section 2. The prior distribution for $\boldsymbol{\xi}^t$ is centered at $\mathcal{I}_{k-1}$, and is disperse (but proper) over the simplex.

Combining the likelihood with the prior distributions, the posterior distribution can be written as (up to a constant of proportionality)

$$\begin{aligned}
\pi(\mathbf{z}, \boldsymbol{\xi}, \Sigma \mid \mathbf{y}) \quad \propto \quad & \prod_{t=1}^{T} \left\{ \prod_{j=1}^{n_t} \left(\prod_{i=1}^{k} ([z_{tj}]_i)^{([y_{tj}]_i - 1)} \mid \Sigma \mid^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t)' \Sigma^{-1} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t) \right] \right) \right. \\
& \left. \times \mid \Omega \mid^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\boldsymbol{\mu}_t - \boldsymbol{\eta})_{'} \Omega^{-1} (\boldsymbol{\mu}_t - \boldsymbol{\eta}) \right] \right\} \\
& \times \mid \Psi \mid^{\frac{\rho}{2}} \mid \Sigma \mid^{-\frac{\rho - k}{2}} \exp\left[-\frac{1}{2} tr(\Psi \Sigma^{-1}) \right]
\end{aligned}$$

The full conditionals for $\mathbf{z}_{tj}, \boldsymbol{\mu}_t = \text{alr}(\boldsymbol{\xi}_t)$, and Σ^{-1} follow immediately from the posterior density.

$$\pi(\mathbf{z}_{tj} | ...) \propto \prod_{i=1}^{k} ([z_{tj}]_i)^{[y_{tj}]_i - 1} \exp\left[-\frac{1}{2} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t)' \Sigma^{-1} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t) \right]$$

$$\pi(\text{alr}(\boldsymbol{\xi}_t) | ...) \propto \exp\left[-\frac{1}{2} (\boldsymbol{\mu}_t - \boldsymbol{\eta})' \Omega^{-1} (\boldsymbol{\mu}_t - \boldsymbol{\eta}) - \frac{1}{2} \sum_{j=1}^{n_t} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t)' \Sigma^{-1} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t) \right]$$

$$\begin{aligned}
\pi(\Sigma^{-1} \mid ...) \quad \propto \quad & \mid \Sigma \mid^{-(\rho - k + \sum_{t=1}^{T} n_t)/2} \times \\
& \exp\left\{ -\frac{1}{2} \left[\sum_{t=1}^{T} \sum_{j=1}^{n_t} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t)' \Sigma^{-1} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t) + tr(\Psi \Sigma^{-1}) \right] \right\}
\end{aligned}$$

Notice that this expression for Σ^{-1} specifies a Wishart distribution with parameter matrix

$$\left\{ \frac{1}{\rho + \sum_{t=1}^{T} n_t} \left[\sum_{t=1}^{T} \sum_{j=1}^{n_t} (\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t)(\boldsymbol{\theta}_{tj} - \boldsymbol{\mu}_t)' + \rho \Psi \right] \right\}^{-1}$$

and $\rho + \sum_{t=1}^{T} n_t$ degrees of freedom. This follows from the conjugacy of the Wishart distribution for the precision matrix of the (Additive Logistic) Normal distribution.

With the full conditional distributions specified, implementation of MCMC is now straightforward. Using the conditional distribution for $\mathbf{z}_{tj}$, the $\sum_{t=1}^{T} n_t$ different species compositions (a composition for each observation vector) can be generated in turn, conditionally on the current values of $\boldsymbol{\mu}_t$ and Σ. Then, the values of $\boldsymbol{\mu}_t$

(equivalently, $\boldsymbol{\xi}_t$) and Σ are updated accordingly. Because the LN distribution is not conjugate for the Multinomial observation distribution, we use Hastings' algorithm (Hastings, 1970) for updating the compositions $\mathbf{z}_{tj}$. The Gibbs sampler is used to update $\boldsymbol{\mu}_t$ and Σ since their conditional distributions are easily sampled directly (multivariate normal and inverse Wishart, respectively, see, e.g., Gelfand, et al, 1990).

Examination of MCMC realizations indicates that the algorithm converges to the limiting distribution in 50-100 Monte Carlo iterations. Further, the convergence is not affected by changes in the (hyper)prior distribution scale parameters. Several trial runs on simulated data suggest a Hastings proposal standard deviation for $\mathbf{z}_{tj}$ of 0.1. This value results in proposal acceptance probabilities of 50–60%.

Results – Stability in Arthropod Food Webs

Arthropods (insects and spiders – hereafter referenced as "bugs") were counted on each of the 30 experimental plots 2, 4, and 6 weeks after treatment application. Here we consider only the 6 weeks data. Note that experimentally manipulated species (*Pardosa* and *Nabis*) are not included in the counts. Eleven different species of bugs (partitioned into three trophic categories: predators, generalist herbivores, and specialist herbivores) were observed and included in the analysis. The proportion attributable to each category was constructed as the composition of bug counts for each plot.

The total number of observed bugs ranged from 7 (on plots 1 and 2 of the SC treatment) to 34 (on plot 5 of the SV treatment). Consequently, plots with the most bugs provide almost five times as much information about the treatment location ($34/7 = 4.9$) as do plots with the fewest bugs. That none of the plots yielded a large total number of bugs suggests that the observed composition (counts) for any single plot is subject to substantial variability, and may be (qualitatively) quite different from the actual (unobservable) plot composition. Also note that for four of the experimental plots, no predators were observed.

The statistical model described above was evaluated using MCMC. A systematic updating scheme was used where each plot composition ($\mathbf{z}_{tj}$), treatment location ($\boldsymbol{\xi}_t$), and the common variance-covariance matrix (Σ) were updated in turn. A sequence of 500 Monte Carlo iterations was used for "burn-in", and the subsequent 10,000 Monte Carlo realizations were collected for each of the updated components. Visual inspection of the realized values and convergence diagnostics indicate that the run length is adequate. Point estimates and credible regions were constructed for each treatment location. The observed plot compositions and the location parameter estimates are shown in Figure 8.

*** Figure 8 about here ***

The figure shows that the treatment with increased omnivory (OV) is less affected than similarly disturbed treatments with either background levels of predators (CV), or increased specialist predators (SV). For treatments that shifted from the CC composition (i.e., SC, SV, and CV), the change is toward compositions with increased relative abundance of specialist herbivores.

To better evaluate the magnitude of the treatment differences, approximate 95% credible regions were constructed for the CC and CV treatment locations. The regions and the location point estimates are shown in Figure 9. The regions shown are pointwise 95% regions (for individual locations) and not simultaneous regions.

*** Figure 9 about here. ***

The credible regions show a separation of the CC and CV treatments. That the CC region contains both the OV and OC treatment location point estimates suggests that these treatment have a similar effect in maintaining bug group compositions. Similarly, the 95% credible region for CV contains the location point estimate for the SC and SV treatments. It suggests that increasing specialist predators does not mitigate the species composition shift caused by reduced vegetation.

Finally, we consider the effect of the vegetation removal separately for each of the predator manipulations. Recall that the "difference" in compositions, $\mathbf{z}_1$ and $\mathbf{z}_2$ (say), can be computed through the perturbation operator as follows (see section 2).

$$\mathbf{z}_1 \ominus \mathbf{z}_2 = \mathbf{z}_1 \oplus (\mathbf{z}_2)^{-1}$$

Using this we can evaluate the distance and direction of changes in group compositions associated with vegetation removal.

Figure 10 shows the change from background vegetation to 50% vegetation removal for each of the three predator manipulations. In addition, approximate 95% credible regions for each difference of locations is shown. If there was no effect attributable to vegetation removal, the differences would be centered at $\mathcal{I}_2$, the center of the simplex.

*** Figure 10 about here. ***

Figure 10 shows that the increased omnivory treatments respond differently to vegetation removal than do

the control or increased specialist predator treatments. Specifically, plots with increased omnivorous predators show increased proportion predators and decreased proportion specialist herbivores when vegetation is removed. Conversely, the increased specialist and control predator treatments show a decrease in the proportion of predators and an increase in specialist herbivores with vegetation removal. The large areas of the 95% credible regions, particularly for increased specialist predators, indicate that the magnitude of these changes is difficult to pin down; likely owing to the small number of plots per treatment (5) and to the small number of bugs observed per plot (as few as 7 bugs on some plots).

Diagnostics

A "leave-one-out" diagnostic procedure (Besag et al., 1995) was used to evaluate the adequacy of the statistical model. The (approximate) predictive distribution for a plot composition is obtained by setting its group counts equal to zero for all k groups, and collecting the MCMC realizations for the plot composition. The zero count for all categories is equivalent to having a missing observation for that plot. It should be noted that the prediction region is constructed for the (unobservable) plot species composition ($\mathbf{z}_{tj}$) and not the discrete observation vector ($\mathbf{y}_{tj}$). The prediction region for the discrete observation can be constructed (under the model) by sampling from a Multinomial distribution with parameter vector equal to the realized values of $\mathbf{z}_{tj}$, and sample size equal to $\sum_{i=1}^{k} [\mathbf{y}_{tj}]_i$.

The leave-one out procedure results in 95% prediction regions that contain the omitted (observed) plot compositions. This suggests that the statistical model is adequate in capturing the observed variability in the data.

Conclusion

These results indicate that increased omnivory helps to maintain a stable species composition in the presence of 50% vegetation removal. Further, background predator levels or increased specialist predators do not facilitate this stability when vegetation is removed. The omnivores' broad diets allow them to feed on a diversity of species that would otherwise increase in abundance in response to the vegetation thinning; effectively buffering the community from compositional shifts induced by disturbance.

Because experimental plots had small total counts, explicit inclusion of a discrete observation model better reflects the true variability of the observed compositions than does the method of simply computing the composition of the observed group counts (possibly adjusted for zeros). Aitchison's model using the observed compositions as data (ignoring their discrete nature) underestimates the actual variability of the observations. This results in confidence regions (or credible regions) that are too small, and tests that do not maintain the nominal level. Alternatively, by including an observation distribution, these problems can be avoided. Incorporation

of the Multinomial observation model extends Aitchison's approach for independent, continuous compositions to observation vectors of discrete counts.

6 Biological Monitoring in the Delaware Bay – Conditional Autoregressive Spatial Model

We further illustrate analysis methods for discrete, compositional data in a biological monitoring problem. In this application, we evaluate the natural variability of the benthic invertebrate population in the 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. (1997). Here we illustrate the methods for a problem where covariates and spatial dependence are important characteristics affecting the biological response.

In 1990, 25 locations in Delaware Bay were sampled to evaluate the benthic community, as well as physical and chemical characteristics at each sample site. The locations of the sites are shown in Figure 11. The stations are identified by their station ID number, and the area of the hexagon at each site is proportional to the number of benthic organisms observed. The background shading denotes the depth.

*** Figure 11 about here. ***

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 palp worms (see Billheimer, et al., 1997 for details). The relative abundance of organisms in the three groups (based on the combined counts of the three subsamples) is shown in figure 12. The observed compositions are identified by their station ID numbers. The plotting symbols indicate the salinity (in parts per thousand) measured at each site.

*** Figure 12 about here. ***

Preliminary analysis indicates the three subsamples at each location exhibit more variability than expected for

multinomial observations with a single proportion vector parameter. As is typical of many biological problems, repeated samples from a single site exhibit super-Multinomial variability.

Statistical Model Description

To incorporate spatial dependence into the statistical modeling structure, we couple a conditional autoregressive (CAR) model (Besag, 1974; Mardia, 1988) with the multinomial observation model. 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.

Suppose the group composition for site j and subsample t, $\mathbf{z}_{jt},\ j = 1, 2, ..., s,\quad t = 1, 2, ..., T_j$, depends on a mean zero (scalar) covariate x_j through the regression relationship described in section 2, and a conditional autoregressive spatial process. That is,

$$\mathbf{z}_{jt} \sim L_{k-1}(\boldsymbol{\theta}_j + \boldsymbol{\beta}\, x_j,\ \Psi)$$

where $\boldsymbol{\beta} \in \Re^{k-1}$ is a regression parameter vector, and $\boldsymbol{\theta}_j \in \Re^{k-1}$ results from a CAR spatial process "adjusted" for the effect of the covariate. The matrix $\Psi_{(k-1)\times(k-1)}$ describes the within site variance-covariance structure. We may interpret $\boldsymbol{\theta}_j$ and $\boldsymbol{\beta}$ as compositions via the alr^{-1} transformation.

To account for spatial structure, let the prior distribution for $\boldsymbol{\theta}_j$ follow a multivariate normal CAR model. Then,

$$E\left[\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{-j}\right] = \boldsymbol{\mu} + \sum_{r \in \delta j} \Lambda_{jr}\ [\boldsymbol{\theta}_r - \boldsymbol{\mu}]$$

where $\boldsymbol{\eta} = \text{alr}^{-1}(\boldsymbol{\mu})$ is the (compositional) location parameter vector of the spatial process, and Λ_{jr} is a $(k-1)\times(k-1)$ matrix of spatial dependence parameters. Let δj denote the set of neighbors of site j. In addition, the conditional variance for the alr-transformed composition at site j is specified

$$\text{Var}\left[\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{-j}\right] = \Gamma_j$$

where Γ_j is symmetric and positive definite. Assuming $\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{-j}$ is $(k-1)$-multivariate normal, Mardia's (1988) derivation yields the following result $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', ..., \boldsymbol{\theta}_s')$ is $s(k-1)$ multivariate normal with density given by

$$\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_s \mid \boldsymbol{\mu}, \Sigma) = \left(\frac{1}{2\pi}\right)^{\frac{s(k-1)}{2}} \mid \Sigma \mid^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{j=1}^{s}\sum_{r=1}^{s}(\boldsymbol{\theta}_j - \boldsymbol{\mu})' \Gamma_j^{-1} \Lambda_{jr} (\boldsymbol{\theta}_r - \boldsymbol{\mu})\right]$$

The variance-covariance matrix is $\Sigma = \left\{\text{Block}(-\Gamma_j^{-1}\Lambda_{jr})\right\}^{-1}$, where $\Lambda_{jj} = -I_{k-1}$. The Λ_{jr} matrices are restricted only to the extent that Σ is symmetric and positive definite.

In this application we make several simplifying assumptions regarding the prior spatial dependence structure. These are

1. The conditional variance at each site is inversely related to the number of neighbors of the site.

2. The influence of the neighbors of a site is the same for all neighbors.

3. The spatial dependence is the same for all groups of organisms.

We write these more formally as follows. (To ease notation let $\boldsymbol{\xi}_j = \text{alr}^{-1}(\boldsymbol{\theta}_j + \boldsymbol{\beta}\, x_j)$ denote the compositional location parameter vector for site j.)

1) Because sites may have from 1 to 6 "first-order" neighbors, we assume the conditional variance at site j depends on its number of neighbors as

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

where n_j is the number of neighbors of site j. The prior distribution specifies that the site composition, $\boldsymbol{\xi}$, (via $\boldsymbol{\theta}_j$) is predicted with greater precision as the number of neighbors increases. This assumption provides a mechanism for allowing increased variability at "edge" sites.

2) Influence of the neighbors of site j is the same for all neighbors. Hence,

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

3) Finally, for the spatial dependence to act identically for all groups of organisms (actually, for all logits $\log(\xi_{ji}/\xi_{jk})$), $\Lambda_j = \lambda/n_j \;\; I_{k-1}$. Note this also implies $\log(\xi_{ji}/\xi_{jk})$ and $\log(\xi_{rm}/\xi_{rk})$ are conditionally independent, given all other logits.

This 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.

Together these assumptions result in the following form for the matrix Block$(-\Lambda_{jr})$.

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

Here, I_{k-1} denotes the $(k-1)$ identity matrix, and $\mathbf{1}_{(r\in\delta j)}$ denotes the indicator function for site r being a neighbor of site j. 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 subsample composition, $\mathbf{z}_{jt}$.

$$p(\mathbf{y}_{jt} \mid \mathbf{z}_{jt}, m_{jt}) = \frac{m_{jt}!}{\prod_{i=1}^{k} [y_{jt}]_i!} \prod_{i=1}^{k} [\mathbf{z}_{jt}]_i^{[y_{jt}]_i}$$

where $m_{jt} = \sum_{i=1}^{k} [y_{jt}]_i$, $[y_{jt}]_i$ is the observed number in group i for site j, subsample t, and $[\cdot]_i$ is the i^{th} component of the k-vector. Further, $\mathbf{z}_{jt}$ is assumed to have conditional distribution $L_{k-1}(\boldsymbol{\theta}_j + \boldsymbol{\beta} x_j, \ \Psi)$.

The prior distributions for λ, $\boldsymbol{\beta}$, $Q = \Gamma^{-1}$, $R = \Psi^{-1}$ and $\boldsymbol{\mu}$ are specified as follows. (Recall that Q is the between–site precision matrix, while R is the within site precision matrix.)

The prior distribution for λ is a scaled Beta distribution; scaled to have support on $(-1, 1)$.

$$\begin{aligned} \pi(\lambda) &\sim \text{Scaled Beta}(\alpha_1, \alpha_2) \\ &\propto \left(\frac{\lambda+1}{2}\right)^{\alpha_1} \left(1 - \frac{\lambda+1}{2}\right)^{\alpha_2} \end{aligned}$$

Vectors $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are $(k-1)$-multivariate normal, each with mean 0_{k-1}, and covariance matrices $d\mathcal{N}$ and $c\mathcal{N}$, respectively. Matrices R and Q are Wishart distributed, with parameters ρ_1 and $a\mathcal{N}^{-1}$ for R, and ρ_2 and $b\mathcal{N}^{-1}$ for Q.

Recall that $\mathcal{N} = I_{k-1} + \mathbf{j}_{k-1}\mathbf{j}_{k-1}^{'}$. Typical choices for a, b, c, and d are $c = d = 0.5$, and $a = b = 1$. Choices for $\alpha_1 = \alpha_2 = 1$ specify a symmetric, unimodal distribution for λ. The hyperparameters ρ_1 and ρ_2 must each be at least $(k-1)$ to make $\pi(Q)$ and $\pi(R)$ proper distributions.

Combining the prior distributions with the likelihood, we obtain the expression for the posterior distribution (up to a constant of proportionality). From this, the full conditional distributions are easily derived and are available for MCMC implementation.

MCMC Implementation

MCMC is used to obtain a Markov chain realization from the joint posterior distribution. The algorithm updates $\mathbf{z}$'s, $\boldsymbol{\theta}$'s, $\boldsymbol{\mu}$, λ, $\boldsymbol{\beta}$, Q, and R each conditional on all other parameters (and on the data, $\mathbf{y}$). Hastings' algorithm (1970) is used to update the $\mathbf{z}$'s. The spatial dependence parameter, λ, is updated via a symmetric, uniform proposal density and Metropolis algorithm acceptance probability (Metropolis, et al., 1953). Gibbs updating (Geman and Geman, 1984) is used for all other model parameters. Details of the MCMC implementation are described in Billheimer and Guttorp (1995).

Inference about the site compositions, the spatial dependence parameter (λ), and the regression parameter vector ($\boldsymbol{\beta}$) result from a MCMC run with a burn-in of 200 cycles, and a collection phase of 20,000 cycles. Graphical inspection of realizations and diagnostics evaluating MCMC performance (Raftery and Lewis, 1992, 1995) indicate that 20,000 cycles are adequate to evaluate the posterior distribution.

Statistical Modeling Results

The CAR 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. Salinity and water depth are each evaluated as potential covariates.

The 95% credible region and point estimate for the salinity regression effect is shown in Figure 13. The point estimate for this composition is (0.33, 0.38, 0.29). The shift of the point estimate and credible region from the center of the simplex indicates an association between salinity and benthic composition. The regression effect 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). The point estimate indicates that as salinity increases, the proportion of palp worms decreases. Palp worms are replaced by pollution intolerant organisms. The proportion of tolerant organisms is not much affected by salinity.

*** Figure 13 about here. ***

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). Further, the 95% credible region for this composition is quite small relative to that for salinity, covers the identity element, $\mathcal{I}_2 = (1/3,\ 1/3,\ 1/3)$. This result indicates little association between water depth and benthic composition.

The realized values of the spatial dependence parameter (λ) are shown in Figure 14. (This estimated distribution corresponds to the model that includes salinity as a covariate.) The figure suggests that there is spatial similarity between neighboring sites (i.e., $\lambda > 0$). The median value for this realization is 0.60, while the observed mean is 0.63. The observed mode is about 0.80. Nearly 93% of the realized values are greater than zero, and 70% are greater than 0.5.

*** Figure 14 about here. ***

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 0.26 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 $0.5/0.26 = 1.9$. 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 integrating over the effect of salinity. Thus, λ denotes spatial dependence above that explained by the salinity gradient. The MCMC realizations suggest partial confounding of the salinity effect with the spatial structure of the observations. Such confounding makes separation of the covariate and spatial effects difficult. We take up this point again in the discussion.

We evaluate the adequacy of the statistical model using prediction for hold-out samples and residual diagnostics for the salinity covariate (on the log-odds scale). The results are summarized in Billheimer, et al. (1997). In short, our evaluation suggests the statistical model captures observed variability, both within and between sites, and that "linear" adjustment of compositions for the effect of salinity is adequate over the range of values observed.

7 Discussion

In this paper we present methods for statistical modeling and interpretation of compositional data. We begin with an algebra for compositions that includes addition, scalar multiplication and a norm. Our framework provides intuitive definitions for additive error, evaluation and interpretation of covariates, and distances between compositions. We also extend Aitchison's (1982, 1986) methods for compositional data analysis to discrete observations via a hierarchical model. Our approach uses a conditional multinomial observation model, but admits other discrete or continuous observation models. The flexibility afforded by this approach relies on MCMC for inference. This inference tool allows us to construct detailed stochastic descriptions of the problems under study.

The primary benefit of our approach is the interpretability of location, covariate, and interaction parameter estimates and credible regions in terms of compositions. Because proportions are a natural scale of measurement for many probelms, interpretation here allows scientists greater insights from statistical modeling results. Specifically, we contrast our approach with results on the multivariate logit scale. Not only are the inferences reported for the logarithm of a ratio of proportions, the estimators are not invariant to permutations of category order. Our approach overcomes both problems.

Interpretability relies fundamentally on the algebra for compositions. In the food web application of section 5 we demonstrate estimation of treatment location parameters and their interactions. The definition of addition of compositions (and consequently subtraction) makes estimation of the treatment interactions straightforward. Similarly, interpretation of the effect of a spatially varying covariate (in the benthic invertebrate example of section 6) is simple once the notion of scalar multiplication is established. Also note that the $\mathcal{L}^2$-type norm defined in section 2 suggests a path toward robust statistical methods. For example, one might define an $\mathcal{L}^\alpha$ norm ($1 \leq \alpha < 2$), and construct methods more resistant to outlying observations. Development of the algebra for compositions allows analogy with standard linear statistical models. In turn, this provides insight into both interpretation of results and extension to new methods.

We use the logistic normal distribution to model latent compositions. This distribution provides a flexible, powerful approach for such quantities. Two strengths of this model are

1. the powerful statistical methods developed for the multivariate normal distribution (for alr$(\cdot)$ transformed data), and

2. its ability to describe complicated covariance structure between components of general compositions.

Specifically, we find the rich covariance structure to be of tremendous benefit in biological applications.

However, there are a number of weaknesses with the logistic normal model. This distribution does not have "nice" mathematical properties of closure when combining elements of a composition (amalgamation), nor when marginalizing over a component. These do not appear to be serious limitations in the applications of the model.

A more serious shortcoming is that the logistic normal distribution requires that all components be positive. A zero proportion actually results in a $(k-1)$ component assemblage (and a $(k-2)$-dimensional logistic normal). Further, the alr$(\cdot)$ transformation is not defined when one or more components are zero. This restriction that all components be present in all samples may be restrictive in applications where one or more components are known to be absent, or where inference of absence is important.

Among other potential approaches for modeling compositional data, the Dirichlet distribution (ref ?) is best known. This distribution exhibits many convenient mathematical properties including closure under amalgamations of categories and marginalization over a category. Its primary limitation in modeling scientific data is its rigid variance-covariance structure prescribed by the summation constraint. As a consequence, the Dirichlet distribution is inadequate for describing complicated covariance relationshipe between elements of a composition.

Two other interesting approaches to compositional data are given by Stephens (1982) and Barndorff-Nielsen and Jørgensen (1991). Stephens suggests treating the square roots of proportions as directional data and using the von Mises spherical distribution to model the compositions. This model appears to be infrequently used in applications. The lack of use may be related to the relative complexity of the von Mises distribution.

Barndorff-Nielsen and Jørgensen (1991) define a class of distributions on the simplex by considering independent generalized inverse Gaussian random variables conditional on their sum. A special sub-class, termed the S^- distribution are constructed from the conditional distribution of identical inverse Gaussian random variables, given their sum. The S^- distribution exhibits several elegant mathematical properties including closure under amalgamations and marginalization over components. In addition, inference in the S^- model leads, in certain cases, to tests based on independent χ^2 variables. Conversely, the S^- model does not allow dependence between components except that induced by the summation constraint. For this reason the distribution is limited for analysis of general compositional data.

References

[1] Aitchison, J. (1982). "The statistical analysis of compositional data (with discussion)." *J. R. Statist. Soc. B.,* **44,** 139–177.

[2] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* Chapman & Hall, New York.

[3] Aitchison, J. and Shen, S. M. (1980). "Logistic-normal distributions: some properties and uses." *Biometrika,* **67,** 261–272.

[4] Albert, J. H. and Chib, S. (1993). "Bayesian Analysis of Binary and Polychotomous Response Data." *J. Amer. Statist. Assoc.,* **88,** 669–679.

[5] Allenby, G. and Lenk, P. (1994). "Modeling household purchase behavior with logistic normal regression". *J. Amer. Statist. Assoc.,* **89,** 1218–1231.

[6] Barndorff-Nielsen, O. E., and Jørgensen, B. (1991). "Some parametric models on the simplex." *J. Multiv. Anal.,* **39,** 106-116.

[7] Besag, J. E. (1974). "Spatial interaction and the statistical analysis of lattice systems" (with Discussion). *J. R. Statist. Soc. B,* **36,** 192–236.

[8] Besag, J. E., Green, P. J., Higdon, D. M. and Mengersen K. (1995). "Bayesian computation and spatial systems" (with Discussion). *Statist. Sci.,* **10,** 3–66.

[9] Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H., and Silkey, M. (1996) "Natural Variability of Benthic Species Composition in the Delaware Bay". *J. Environ. and Ecol. Statist.* to appear.

[10] Billheimer, D. D. and Guttorp, P. (1995). "Spatial statistical models for discrete compositional data". Technical Report, Dept. of Statistics, University of Washington, Seattle.

[11] Cerioli, A. (1992). "On the analysis of spatial categorical data." in *Statistical Modelling.* (ed. van der Heijden, P. G. M., Jansen, W, Francis, B., and Seeber, G. U. H.), Elsevier Science Publishers B.V., Amsterdam, The Netherlands. pp. 45-54.

[12] Fagan, W.F. (1996). "Population dynamics, movement patterns, and community impacts of omnivorous arthropods." Ph.D. Dissertation, University of Washington, Seattle, WA.

[13] Fagan, W.F. (1997). "Omnivory as a stabilizing feature of natural communities." *American Naturalist.* **150,** 554-568.

[14] Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). "Illustration of Bayesian inference in normal data models using Gibbs sampling." *J. Am. Statist. Assn.,* **85,** 972-985.

[15] Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika* **57:**97–109.

[16] Kass, R. E. and Raftery, A. E. (1995). "Bayes factors." *J. Amer. Statist. Assoc.,* **90,** 773–795.

[17] Mardia, K. V. (1988). "Multidimensional multivariate Gaussian Markov random fields with applications to image processing." *J. Multivariate Anal.* **24,** 265–284.

[18] Menge, B. and Sutherland, J. (1987). "Community regulation: variation in disturbance, competition, and predation in relation to environmental stress and recruitment." *American Naturalist* **130,** 563-576.

[19] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller E. (1953). "Equations of state calculations by fast computing machines." *J. Chemical Physics* **21,**108–1091.

[20] Overton, W. S., White, D., and Stevens, D. K. (1990). "Design Report for Emap: Environmental Monitoring and Assesment Program." EPA/600/3-91/053, U.S. Environmental Protection Agency, Washington, D.C.

[21] Pimm, S.L. (1982). *Food webs.* Chapman and Hall. New York.

[22] Pimm, S.L. and Lawton, J.H. (1978). "On feeding on more than one trophic level." *Nature* **275,** 542-544.

[23] Pawlowski, V. and Burger, H. (1992). "Spatial structure analysis of regionalized compositions." *Mathematical Geology,* **24,** 675-691.

[24] Polis, G.A. (1994). "Food webs, trophic cascades and community structure." *Australian Journal of Ecology* **19,** 121-136.

[25] Raftery, A. E. and Lewis, S. M. (1992). "How many iterations in the Gibbs sampler?" in *Bayesian Statistics 4.* (ed. Bernardo, J., Berger, J., Dawid, A. P. and Smith, A. F. M.). Oxford University Press, pp. 765–776.

[26] Raftery, A. E. and Lewis, S. M. (1995). "The number of iterations, convergence diagnostics and generic Metropolis algorithms." In *Practical Markov Chain Monte Carlo* (ed. Gilks, W. R., Spiegelhalter, D. J. and Richardson, S.). Chapman & Hall, London.

[27] Rayens, W. S. and Srinivasan, C. (1994). "Dependence Properties of Generalized Liouville Distributions on the Simplex." *J. Amer. Statist. Assoc.,* **89,** 1465–1470

[28] Strong, D.R. (1992). "Are trophic cascades all wet? Differentiation and donor control in speciose ecosystems." *Ecology* **73,** 747-754.

[29] Statistical Sciences, Inc. (1995). *S-PLUS User's Manual, Version 3.4 for Unix,* Seattle, Statistical Sciences, Inc.

[30] Stephens, M. A. (1982). "Use of the von Mises distribution to analyze continuous proportions." *Biometrika,* **69,** 197-203.

[31] Swayne, D. F., Cook, D., and Buja, A. (1991). "XGobi: Interactive dynamic graphics in the X window system with a link to S." *ASA Proc. Statist. Graphics,* 1-8.

[32] Upton, J. G. and Fingleton, B. (1989). *Spatial Data Analysis by Example, Vol. 2,* Wiley, New York.

8 Appendix I

Property 8.1 *For $\mathcal{I}_{k-1} = (\frac{1}{k}, \frac{1}{k}, ..., \frac{1}{k})$, the operation $\oplus\, \mathcal{I}_{k-1}$ is the identity operator for any $\mathbf{u} \in \nabla^{k-1}$, i.e., $\mathbf{u} \oplus \mathcal{I}_{k-1} = \mathbf{u}$*

Proof.

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

Property 8.2 *The operation $\oplus$ is commutative. For $\mathbf{u}$ and $\mathbf{a}$ in ∇^{k-1},*

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

Proof.

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

Property 8.3 *The operation $\oplus$ is associative. For $\mathbf{u}$, $\mathbf{a}$, and $\mathbf{z}$ in ∇^{k-1},*

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

Proof.

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

Theorem 8.1 *∇^{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. $(\mathbf{u} \oplus \mathbf{z})^a = \mathbf{u}^a \oplus \mathbf{z}^a$

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

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

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

Theorem 8.2 *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. $\langle \mathbf{u}, \mathbf{z} \rangle = \langle \mathbf{z}, \mathbf{u} \rangle$

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

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

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

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 8.3 *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 $\langle \mathbf{u}, \mathbf{u} \rangle$.

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

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 8.4 *∇^{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 $\|\mathbf{u}_m \ominus \mathbf{u}_n\| < \epsilon$.

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

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

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.

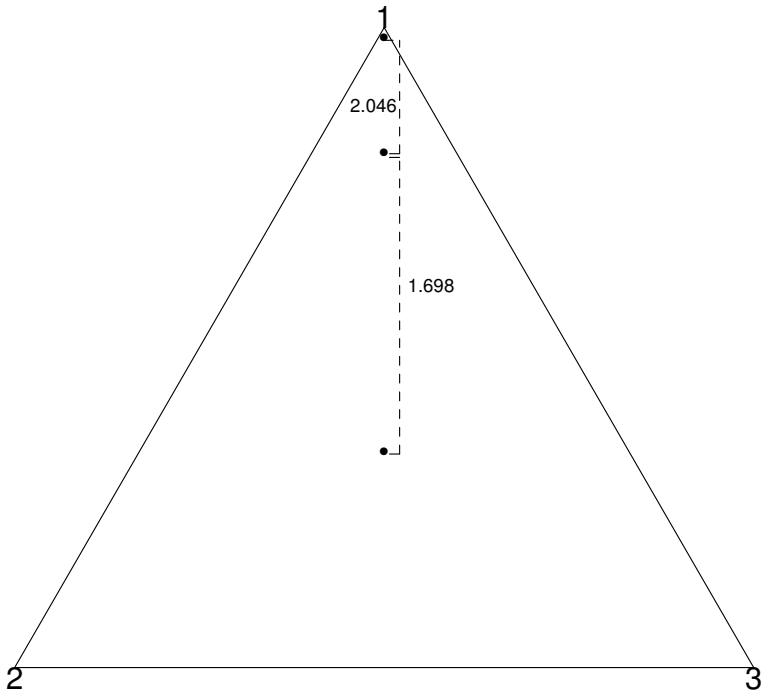

Figure 1: Graphical display of 3 three-part compositions in a ternary diagram. The points shown correspond to compositions of $(1/3,\ 1/3,\ 1/3), (0.80,\ 0.15,\ 0.05)$, and $(0.96,\ 0.03,\ 0.01)$. The numbers on the figure denote the distances between the points.

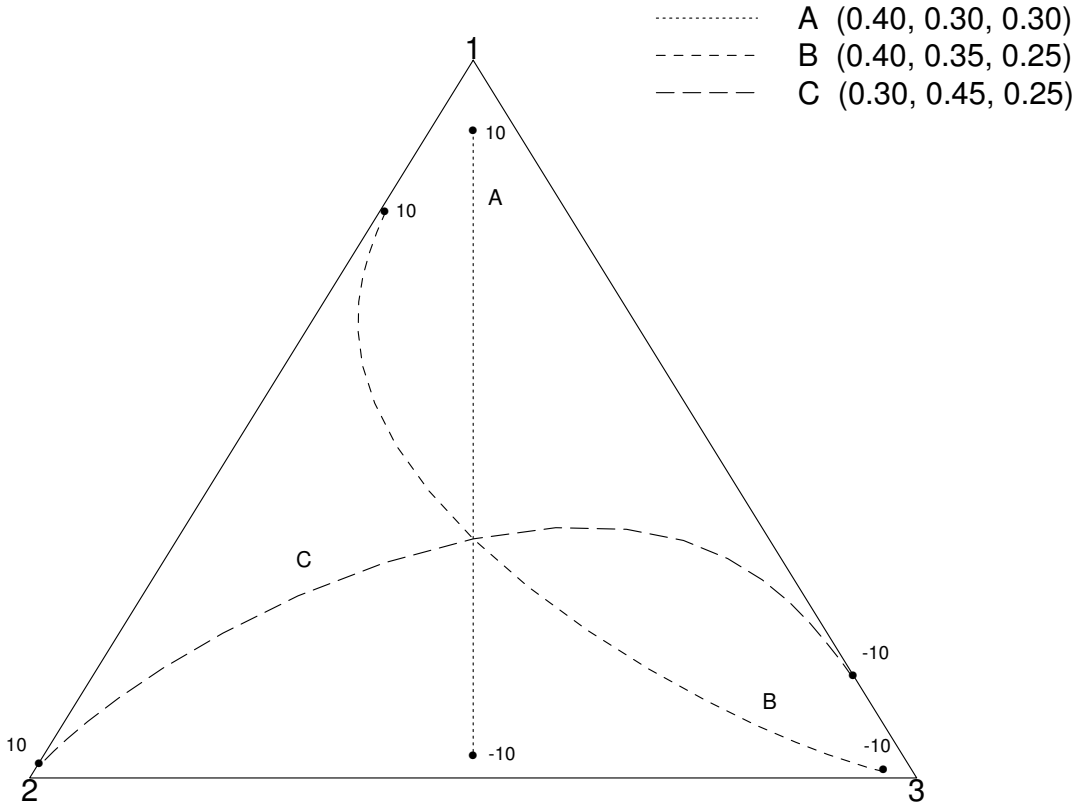

Figure 2: "Regression lines" for different parameter vectors.

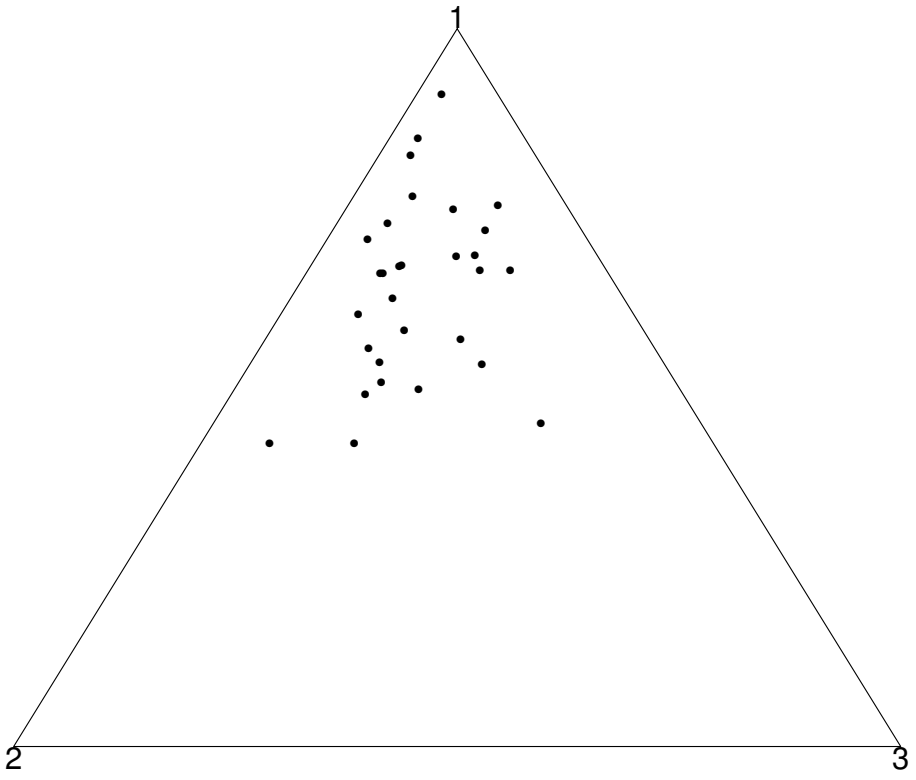

Figure 3: 30 realizations from a logistic normal model with $\boldsymbol{\xi} = (0.7,\ 0.2,\ 0.1)$ and $\Sigma = 0.2\mathcal{N}$

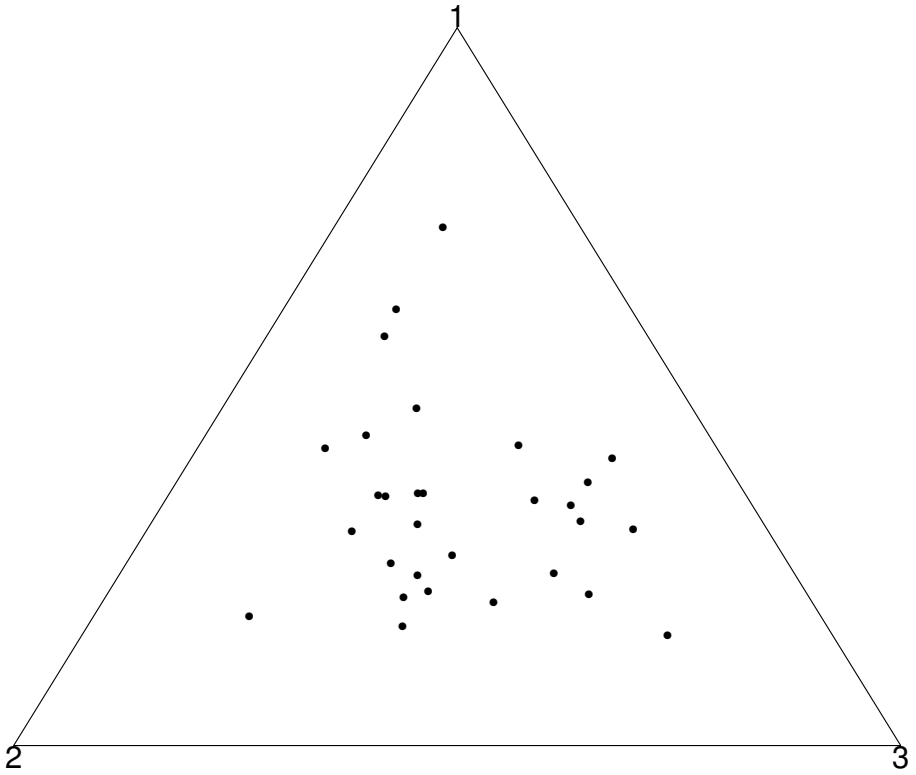

Figure 4: 30 residuals from a logistic normal model. Estimated location parameter vector $\hat{\boldsymbol{\xi}} = (0.659,\ 0.216,\ 0.125)$.

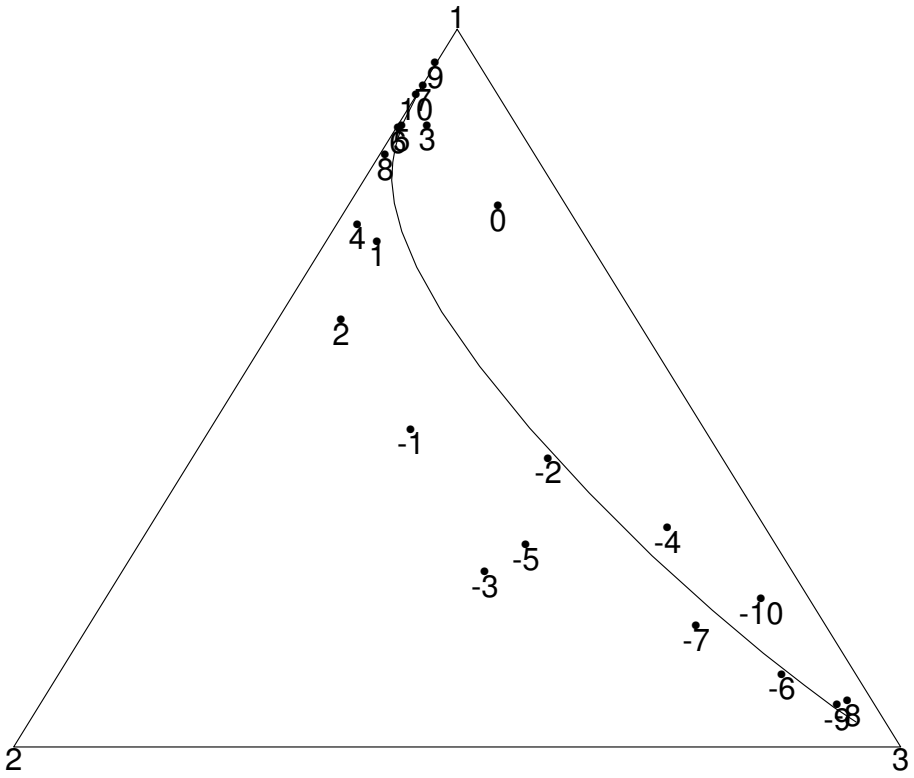

Figure 5: Realizations from a linear regression model for compositions. Regression parameter vector (0.40, 0.35, 0.25)

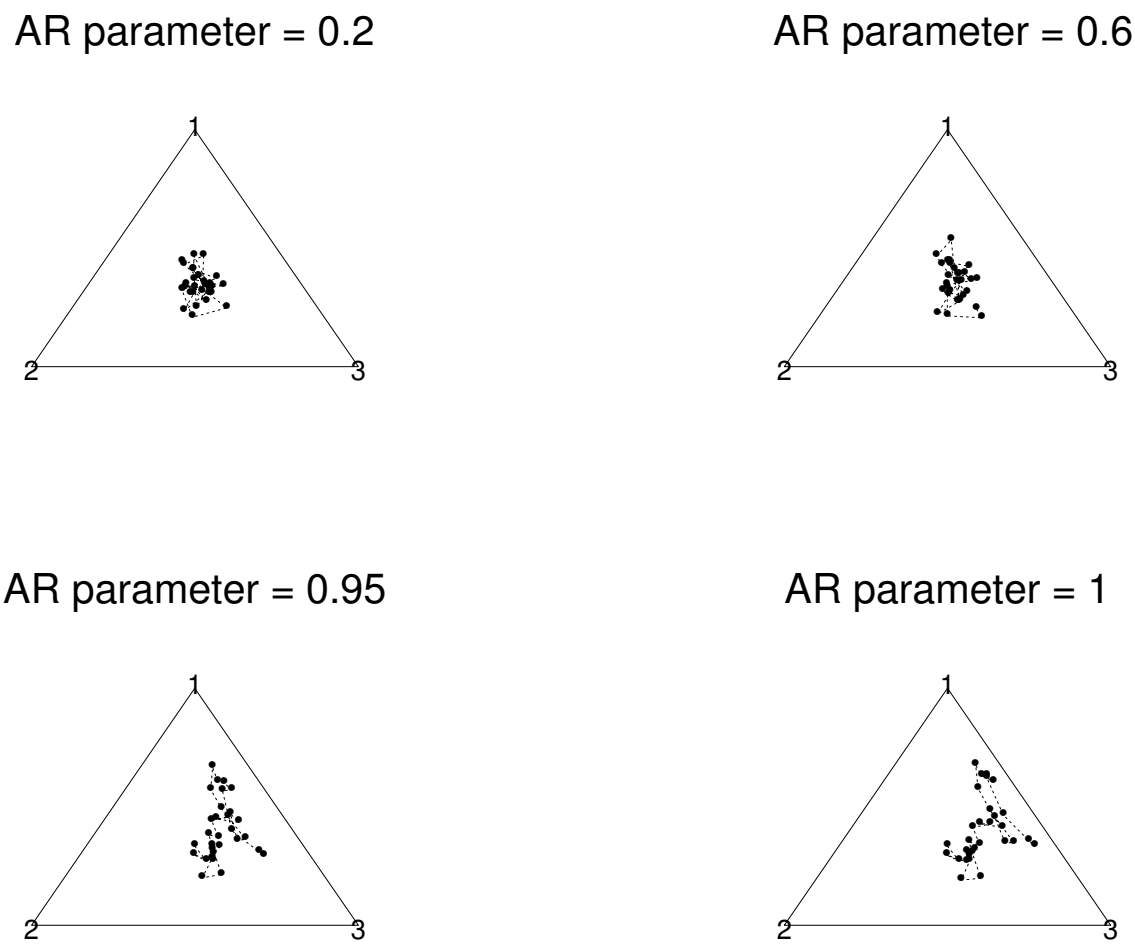

Figure 6: Comparison of first-order autoregressive processes. AR parameters are 0.2, 0.6, 0.95, and 1, respectively.

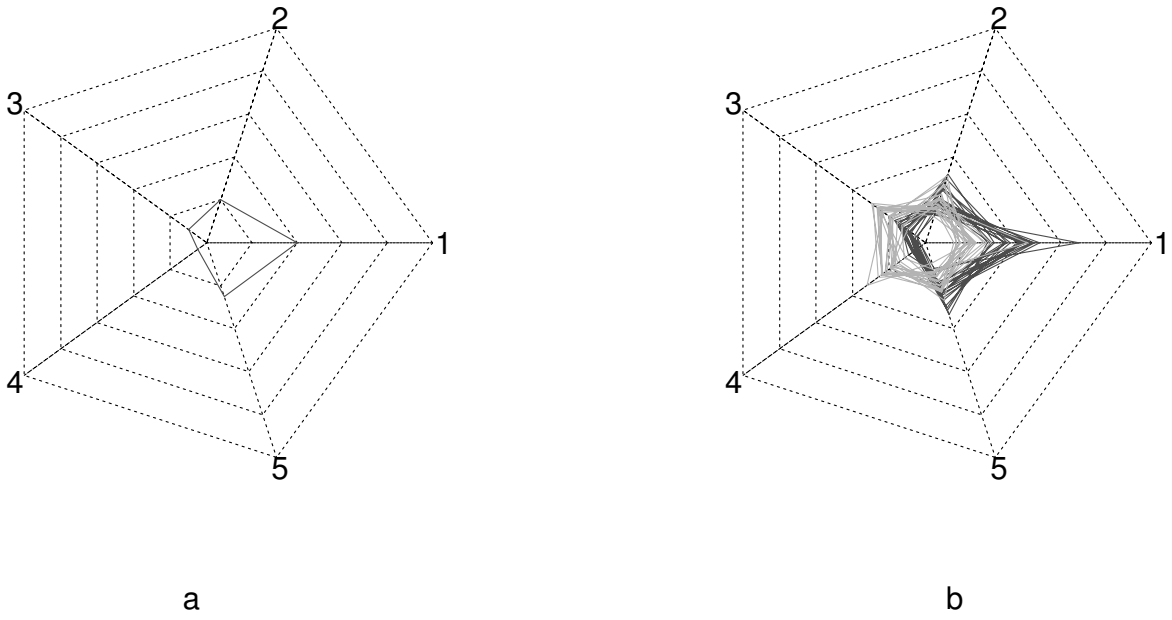

Figure 7: Figure (a), on the left, shows a single composition of (0.40, 0.20, 0.10, 0.05, 0.25). Figure (b) shows realizations from two distributions. Note that the dotted lines joining constant radii represent the coordinate axes. These denote "tick marks" at 0.2, 0.4, 0.6, 0.8, and 1.0, respectively.

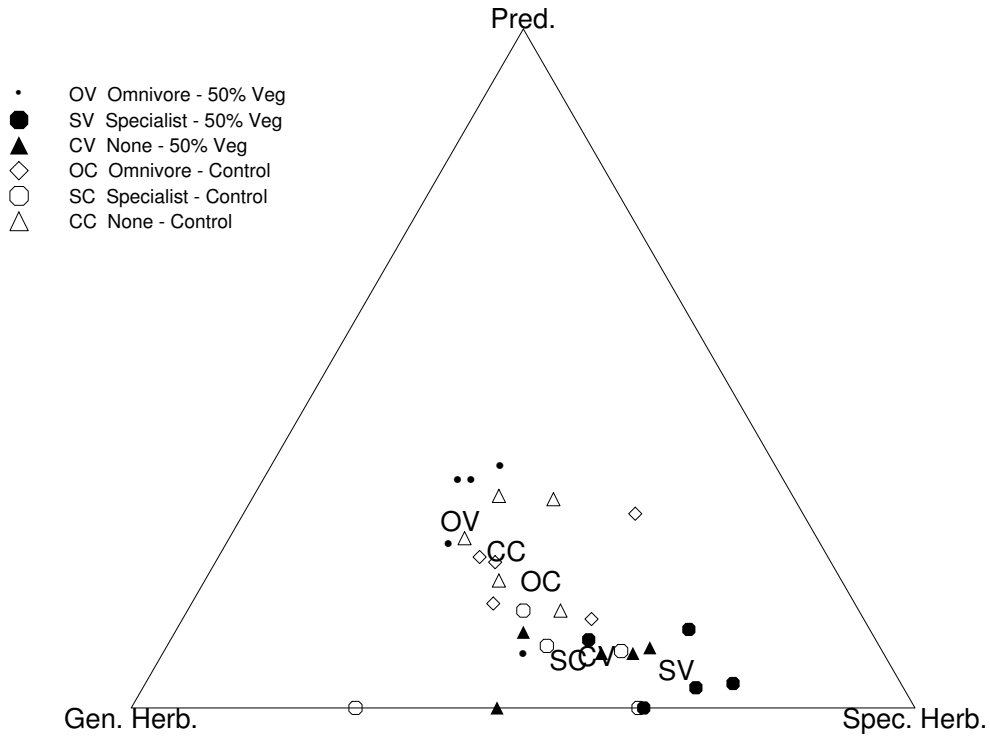

Figure 8: Location parameter estimates for the six experimental treatments.

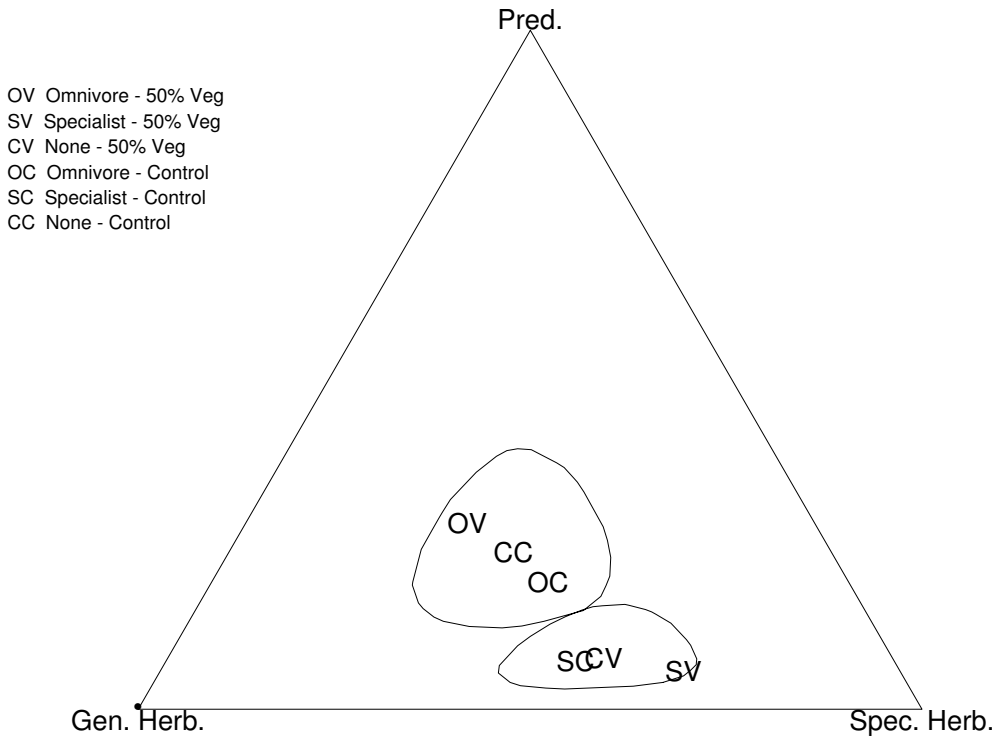

Figure 9: 95% credible regions for the location parameter estimates.

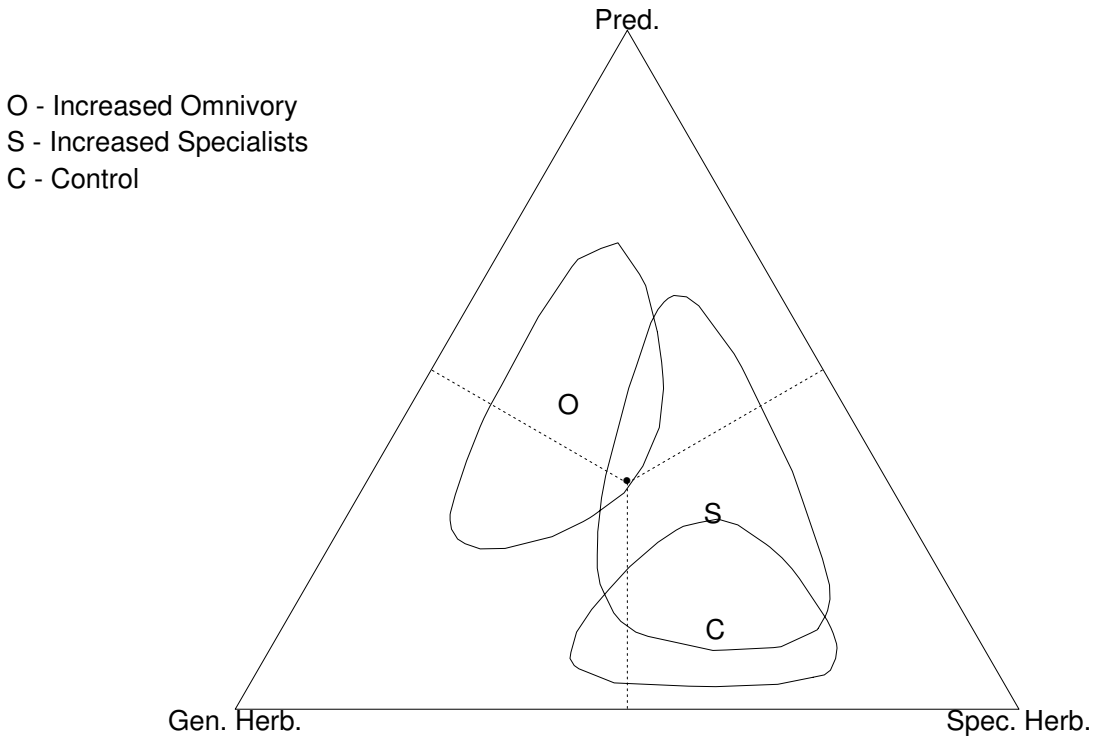

Figure 10: Point estimates and 95% credible regions for omnivory by vegetation interaction.

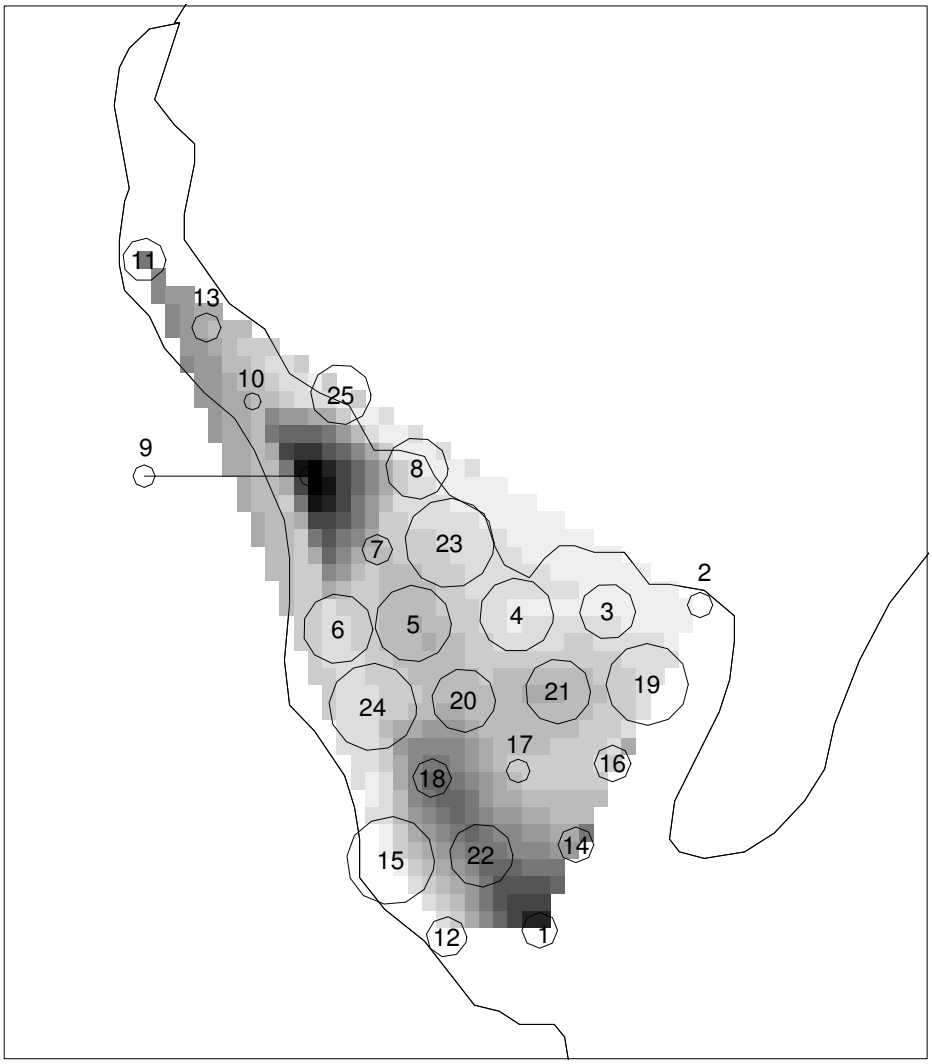

Figure 11: Benthic sample locations in the Delaware Bay. The sample locations are identified by their station ID number. The area of the hexagon at each site is proportional to the number of benthic organisms observed. The background shading denotes the depth.

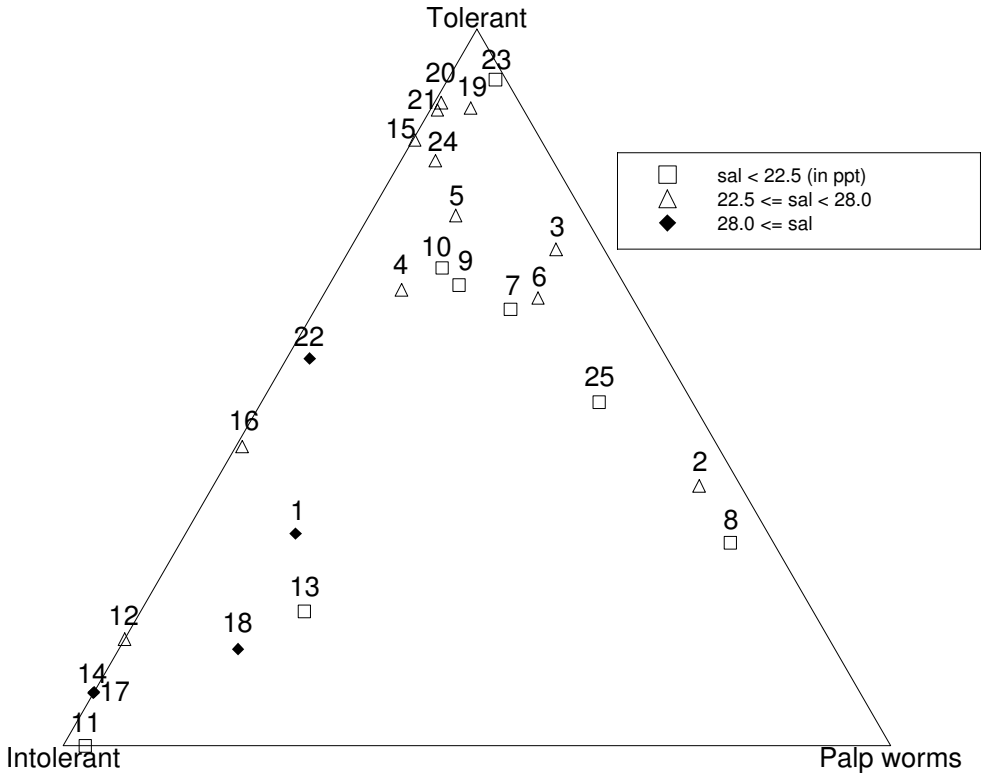

Figure 12: Observed benthic invertebrate compositions from Delaware Bay. The numbers denote the sample location, while the plotting symbol indicates the measured salinity at the sample site.

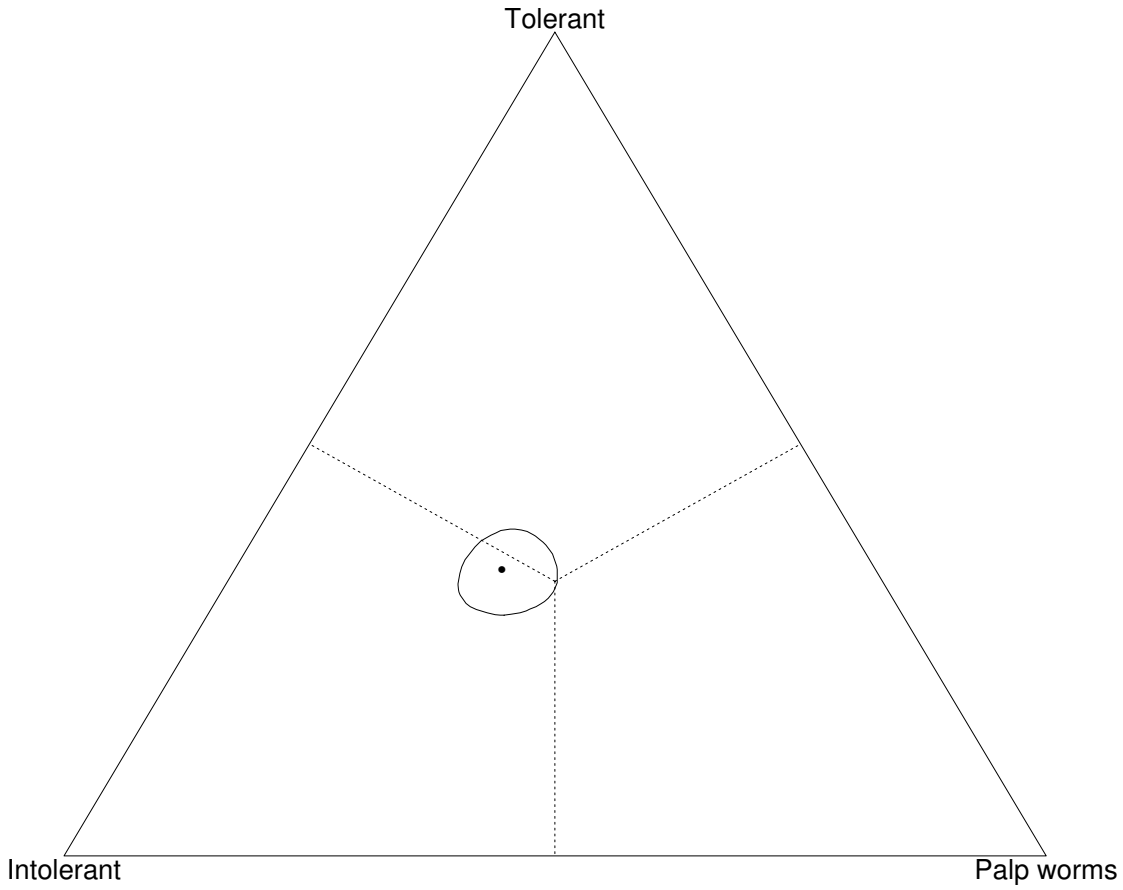

Figure 13: Point estimate and 95%credible region for salinity regression parameter composition. The point estimate is (0.33, 0.38, 0.29) indicating increasing proportions of intolerant organisms and decreasing palp worms.

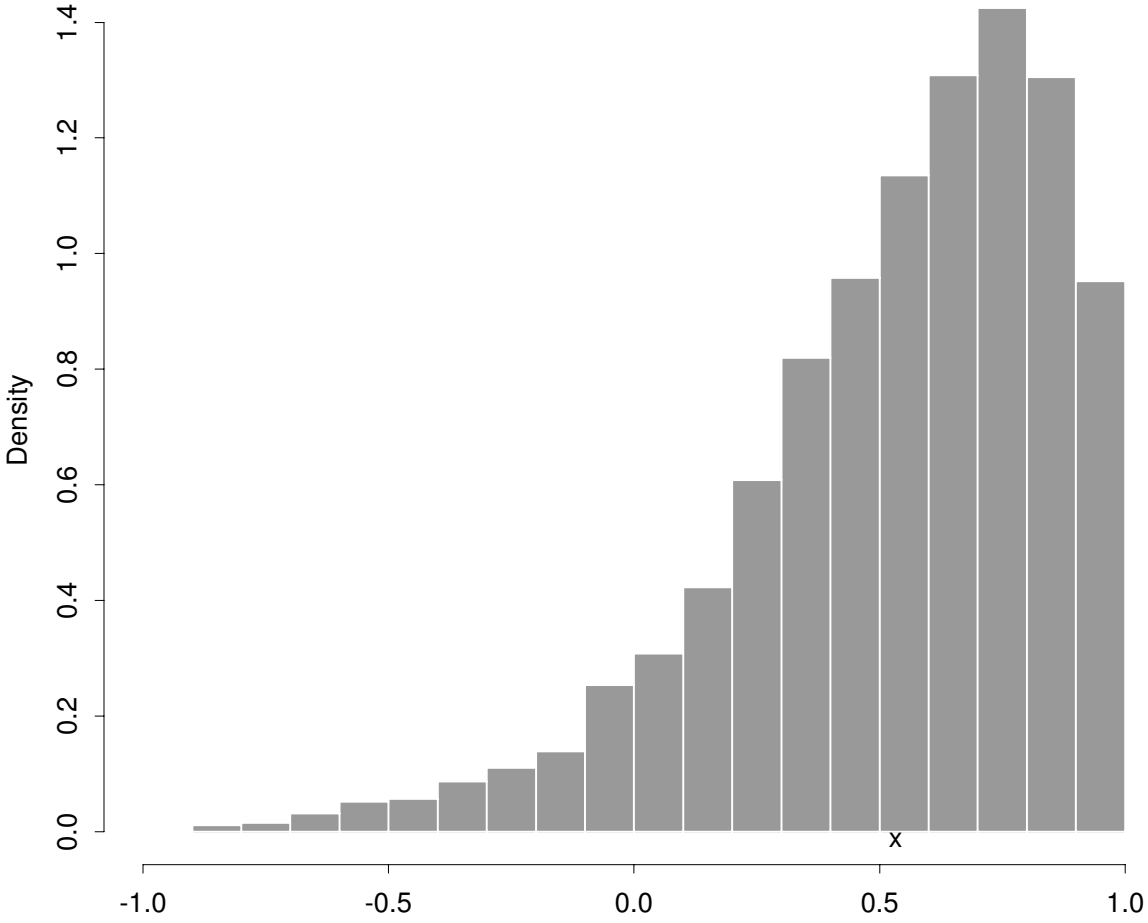

Figure 14: MCMC realizations for the spatial dependence parameter, λ. The median for λ is 0.60, and the mode is about 0.80. The $\times$ indicates the mean of the realizations of 0.63.