Supplementary materials for this article are available at 10.1007/s13253-011-0058-6.

A Baseline Category Logit Model for Assessing Competing Strains of *Rhizobium* Bacteria

C. BROPHY, J. CONNOLLY, I. L. FAGERLI, S. DUODU, and M. M. SVENNING

In this paper we describe novel methodology for evaluating competition among strains of *Rhizobium* bacteria which can be found naturally occurring in or can be introduced into soil. Rhizobia can occupy nodules on the roots of legume plants allowing the plant to 'fix' atmospheric nitrogen. Our model defines competitive outcomes for a community (the multinomial count of nodules occupied by each strain at the end of a time period) relative to the past state of the community (the proportion of each strain present at the beginning of the time period) and incorporates this prior information in the analysis. Our approach for assessing competition provides an analogy to multivariate methods for continuous responses in competition studies and an alternative to univariate methods for discrete responses that respects the multivariate nature of the data. It can also handle zero values in the multinomial response providing an alternative to compositional data analysis methods, which traditionally have not been able to facilitate zero values. The proposed experimental design is based on the simplex design and the model is an extension of multinomial baseline category logit models that includes multiple offsets and random terms to allow for correlation among clustered responses. Supplemental materials for this article are available from the journal website.

Key Words: Competition with discrete response; Compositional data analysis; Discrete multivariate analysis; Random effects; Simplex design; Zero values.

1. INTRODUCTION

Competition occurs among species when a required resource is limited and the species 'compete' to each obtain the resource. Competition has been widely studied experimentally across many organisms (Nicol and Thornton 1941; Connell 1983; Schoener 1983;

C. Brophy (✉) is a Lecturer in Statistics, Department of Mathematics & Statistics, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland (E-mail: *caroline.brophy@nuim.ie*). J. Connolly is Associate Professor in Statistics, UCD School of Mathematical Sciences, Environmental & Ecological Modelling Group, University College Dublin, Belfield, Dublin 4, Ireland. I. L. Fagerli was a Research Student and M. M. Svenning is Professor and Head, Department of Arctic and Marine Biology, University of Tromsø, 9037 Tromsø, Norway. S. Duodu is a Researcher, National Veterinary Institute, P.O. Box 750, Sentrum, 0106 Oslo, Norway.

© 2011 International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics, Volume 16, Number 3, Pages 409–421
DOI: 10.1007/s13253-011-0058-6

Firbank and Watkinson 1985; Goldberg and Barton 1992; Iwasa, Nakamaru, and Levin 1998). The analytical approaches for assessing effects range from multivariate models for continuous responses (Connolly and Wayne 2005) to univariate approaches for discrete responses (May 2001) to compositional methods (Aitchison 1986; Aitchison and Ng 2005). Here we develop a modeling approach for discrete multinomial response data that extends the current competition literature in three ways: (1) it is analogous to a competition model derived for continuous responses by Connolly and Wayne (2005) that defines competitive outcomes relative to the past state of the community and incorporates this prior information in the analysis, (2) it allows for the multivariate nature of the response data, (3) it will handle zero response values. Our model is a baseline category logit model extended to include random effects (Hartzel, Agresti, and Caffo 2001) to allow for correlated responses and multiple offset terms to allow for initial starting values of species. Offsets have previously been used with models for discrete responses (logistic regression in Agresti 2002) but multiple offsets have not been used with multinomial models or for the purpose of assessing competition among species.

The models developed in this paper are motivated by a study of competition among strains of rhizobia bacteria, which are found naturally occurring in soil or can be introduced deliberately into soil. Rhizobia can occupy nodules on the root of legume plant species resulting in atmospheric nitrogen fixation and thereby supply the host plant with N and provide additional N in the legume environment. This natural source of N can be beneficial to the productivity of grassland systems and can reduce the cost of running the system. It is possible that some strains of rhizobia are superior at occupying nodules and at fixing N. Does the proportion of strains of rhizobia present in the soil at a given point in time affect the proportion of nodules that the strain will occupy at a later time? To answer this question we applied three strains of rhizobia to the roots of a legume species in a range of initial proportions and after a period of time counted the number of nodules each strain occupied. There were a limited number of available sites for nodulation and the strains competed to occupy them. For each community (root section) we have a vector of initial proportions and a final multinomial response vector. We modeled the change from initial proportion applied to final proportion of nodules occupied for each strain.

In a community, a good competitor is one that gains proportionately more over time than other species (Connolly, Wayne, and Bazzaz 2001). Connolly and Wayne (2005) and Ramseier, Connolly, and Bazzaz (2005) developed a multivariate modeling approach to assessing the effects of the species identity, environment and species initial relative abundance on the outcome of competition. The continuous and multivariate response measured was the relative growth rate of each species in a community over a period of time. The variable(s) modeled were the differences in relative growth rates between pairs of species in a community, giving the name RGRD (relative growth rate difference) to the models. The RGRD model does not currently facilitate discrete responses.

When the response for each species in a community is a discrete whole number each experimental community provides a multinomial response vector. There is a long history of modeling approaches to community dynamics for such discrete responses (May 2001) and these can been related to a discrete version of the Lotka–Volterra model

(Leslie 1958). However, these approaches rarely deal with the multivariate nature of these types of data. Other approaches have been to use compositional data analysis methods for changing compositions (Aitchison 1986; Aitchison and Ng 2005), but these methods break down when species with zero compositions occur in the response. Some approaches to facilitate zero methods have been developed (e.g. Aitchison and Kay 2003; Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn 2003; Butler and Glasbey 2008) but these rely on assumptions about the type of zero or are suited only to analysis for specific hypotheses e.g. to compare compositions of different groups.

In a simplex design (Scheffe 1963; Cornell 2002), the initial relative abundances of competing species are manipulated so that not all experimental communities have all species equally present to begin with. This design has been used in a range of multispecies competition studies (e.g. Ramseier, Connolly, and Bazzaz 2005; Kirwan et al. 2007; Suter et al. 2007) as it allows a broad coverage of the design space and facilitates the simultaneous assessment of species identity, the effect of species on each other and, if required, environmental effects (Connolly, Wayne, and Bazzaz 2001). Ideally, in competition studies, the simplex design would comprise a wide range of compositions in the simplex space at a number of overall densities (Ramseier, Connolly, and Bazzaz 2005; Kirwan et al. 2007).

In this paper we propose an experimental and analytical framework for assessing competition among species where the outcome is discrete. The experimental response in our motivating example is the number of nodules acquired by each strain of rhizobia in each community and this is a multinomial vector. We describe a multinomial modeling framework for discrete responses from this multi-strain competition experiment and the experimental design needed to estimate model parameters, and we detail how to predict and test predictions from the models. The novel features are the marrying of simplex designs with multinomial responses in a discrete modeling framework that defines competitive outcomes for a community of species relative to a previous state of the community and incorporates this prior information in the analysis.

2. METHODS

We propose a multinomial baseline category logit model (Agresti 2002) to measure the competition between J species (categories) that will allow the assessment of competitive relationships among species and consequences for community structure. The categorical response vector is $(y_{i1}, \ldots, y_{iJ})$ for $i = 1, \ldots, c$ (the number of communities) and $j = 1, \ldots, J$ (the number of species) and represents the number of 'success counts' for each species at time t with $\sum_{j=1}^{J} y_{ij} = n_i$ being the total number of success counts for community i. A multinomial baseline category logit model is a series of $J - 1$ models relating the jth to the Jth species where the Jth species is called the baseline category. The ordering of the $j = 1$ to J species and the use of a particular species as the 'baseline' is arbitrary and independent of interpretation. We can model the vector of parameters $(\pi_{i1}, \ldots, \pi_{iJ})$, the proportion of success counts for each species in the ith community at

time t, with $\sum_{j=1}^{J} \pi_{ij} = 1$, as

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{x}_i'\boldsymbol{\beta}_j \quad \text{for } j = 1, \ldots, J-1 \tag{2.1}$$

where $\mathbf{x}_i$ denotes the vector of K explanatory variables for the ith community, $\boldsymbol{\beta}_j$ is the parameter vector of coefficients for the jth model and could include abiotic effects such as an environmental treatment. If $\boldsymbol{\beta}_j = 0$, then $\frac{\pi_{ij}}{\pi_{iJ}} = 1$ and we conclude that species j and J have the same proportion of success counts at time t. While model (2.1) can assess proportion of success counts by species relative to the baseline species at a given point in time (t), it can not address questions of competitive relations or consequences for community dynamics without incorporating information on the proportions of each species in the community at time 0 (or some other reference time) (Connolly, Wayne, and Bazzaz 2001). If the proportion of each species initially present in the ith community at time 0 is given by the vector $(p_{i1}, \ldots, p_{iJ})$, then we propose the model:

$$\log\left(\frac{\pi_{ij}/p_{ij}}{\pi_{iJ}/p_{iJ}}\right) = \mathbf{x}_i'\boldsymbol{\beta}_j \quad \text{for } j = 1, \ldots, J-1 \tag{2.2}$$

which can be rewritten as

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{x}_i'\boldsymbol{\beta}_j + \log\left(\frac{p_{ij}}{p_{iJ}}\right) \quad \text{for } j = 1, \ldots, J-1 \tag{2.3}$$

where $\log(\frac{p_{ij}}{p_{iJ}})$ is an offset term, i.e. a regression term with known coefficient equal to 1. If $\boldsymbol{\beta}_j = 0$, it indicates no change in relative abundance from time 0 to time t between the two competing species j and J and implies that the two species are equally competitive. This model is analogous to the specification of the RGRD model in Connolly and Wayne (2005, Equation (4)).

We extend this model to include a community specific random effect to allow for variation from community-to-community (Hartzel, Agresti, and Caffo 2001). The model comparing the jth to the Jth species is

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{x}_i'\boldsymbol{\beta}_j + \log\left(\frac{p_{ij}}{p_{iJ}}\right) + \mathbf{z}_i'\mathbf{u}_{ij} \quad \text{for } j = 1, \ldots, J-1 \tag{2.4}$$

where $\mathbf{z}_i$ denotes the design vector for the random effect for the ith community and $\mathbf{u}_{ij}$ is assumed multivariate normal with an unstructured covariance matrix ($\boldsymbol{\Sigma}$) to keep independence of the choice of baseline category (Hartzel, Agresti, and Caffo 2001).

We can fit model (2.4) using maximum likelihood. Denoting the linear predictor, $\text{lp}_{ij} = \mathbf{x}_i'\boldsymbol{\beta}_j + \log(\frac{p_{ij}}{p_{iJ}}) + \mathbf{z}_i'\mathbf{u}_{ij}$, the likelihood function for the ith response vector is, integrating out the random effects and omitting a fixed constant:

$$\prod_i \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_j^{J-1} \left[\frac{\exp(\text{lp}_{ij})}{1 + \sum_{j=1}^{J-1} \exp(\text{lp}_{ij})} \right]^{y_{ij}} \left[\frac{1}{1 + \sum_{j=1}^{J-1} \exp(\text{lp}_{ij})} \right]^{y_{iJ}} \times f(\mathbf{u}_{ij}; \boldsymbol{\Sigma})\, \mathrm{d}\mathbf{u}_{ij} \right). \tag{2.5}$$

We predict (denoted by the ˆ symbol, which is also used to denote the maximum likelihood estimate of model parameters) the proportion of success counts for the jth species from the model at the median value of the random effect using the equations:

$$\begin{aligned}\hat{\pi}_{ij} &= \frac{\exp\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}_j + \log\left(\frac{p_{ij}}{p_{iJ}}\right)\right)}{1+\sum_{j=1}^{J-1}\exp\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}_j + \log\left(\frac{p_{ij}}{p_{iJ}}\right)\right)} \quad \text{for } j = 1,\ldots,J-1,\\ \hat{\pi}_{iJ} &= 1 - \sum_{j=1}^{J-1}\hat{\pi}_{ij} \quad \text{for } J.\end{aligned} \tag{2.6}$$

While this model may be applied to a wide range of count data it is particularly relevant to data from experiments based on a simplex design (Scheffe 1963; Cornell 2002) in which the initial p_{ij} values and overall initial density of species are deliberately manipulated. The relative abundance of each species at time 0, $(p_{ij},\ldots,p_{iJ})$, may be important determinants of species relative competitiveness and hence of the final composition $(\pi_{ij},\ldots,\pi_{iJ})$ of the ith community. At its simplest, the $\mathbf{x}$ matrix in model (2.4) would include the relative abundances $p_{ij},\ldots,p_{iJ}$ giving:

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \sum_{k=1}^{J}\beta_{jk}p_{ik} + \beta_{jD}D_i + \log\left(\frac{p_{ij}}{p_{iJ}}\right) + u_{ij} \quad \text{for } j = 1,\ldots,J-1 \tag{2.7}$$

where p_{ik} is the initial proportion of the kth species for $k = 1,\ldots,J$, D_i is the total density of the ith community and u_{ij} is a random effect with variance σ_j^2 and may be correlated with the other $J-2$ random effects. Interactions among the p_{ik}'s and between the p_{ik}'s and other independent variables, such as a treatment factor or community density (D) may also be included in the model specification.

For model (2.7), if $\beta_{jk} = 0$ for all $k = 1,\ldots,J$ and $\beta_{jD} = 0$, then the relative proportions of the jth and Jth species are the same at times 0 and t, and species j and J are equally competitive i.e. $(\frac{\pi_{ij}}{\pi_{iJ}}) = (\frac{p_{ij}}{p_{iJ}})$. When these parameters are not zero and interaction effects are present, the number of competition coefficients may mean it is difficult to see their combined impact on community relative composition. To interpret the model the final proportions of success counts for each species can be predicted for a range of initial communities and these predictions used to determine the outcome of competition. Predictions can be displayed graphically using ternary diagrams (where there are three competing species), and we distinguish between two numerical comparisons. Compositional change measure (1): $\hat{\pi}_{ij}/p_{ij}$ compares the predicted proportion of success counts relative to initial proportion present for an individual species. This measure determines how a species performs relative to its own expectation (p_{ij}) but even a species that performs better than expected may not be the most competitive species. Compositional change measure (2): $\frac{\hat{\pi}_{ij}/p_{ij}}{\hat{\pi}_{ij'}/p_{ij'}}$ for $j \neq j'$, compares two species and determines which is the more competitive of the two.

3. EXAMPLE ASSESSING COMPETITION BETWEEN COMPETING STRAINS OF RHIZOBIUM FOR NODULE OCCUPANCY

3.1. EXPERIMENTAL DESIGN

When a *Rhizobium* strain has occupied a nodule on the root of a legume, it normally has the ability to 'fix' nitrogen (N) from the atmosphere and supply the host plant with N and provide additional N in the legume environment. Competition was investigated among *M. loti* strains *Ml*8, *Ml*19 and *Ml*16; named A, B and C, respectively, from here on. For inoculation, the three strains were mixed in proportions based on a simplex design both for high (1×10^4 cells ml^{-1}) and low (1×10^2 cells ml^{-1}) cell densities (Table 2: Density and Initial proportions columns). Four seedlings of *L. corniculatus* cultivar Leo (A/S Strand Brænderi, Norway) were transferred to each jar, and 5 jars (communities) were used as replicates for each of the seven inoculum mixtures at each density, giving a total of 70 communities. To each jar, 1 ml inoculum was added. The plants were grown for 4 weeks and rhizobia were re-isolated from nodules as described by Fagerli and Svenning (2005). Nodule occupancy of the rhizobial strains was determined by ERIC-PCR fingerprinting (de Bruijn 1992). The number of nodules occupied by each strain in each community was recorded. Several of the responses for individual species were zero. The raw data is available in online Appendix A. Further experiment details are in Fagerli (2006).

3.2. MODEL FITTING

We fitted a series of multinomial baseline category logit random effects models to the multinomial data. We used strain C as the baseline category and let $j = \mathrm{A}, \mathrm{B}$ and $J = \mathrm{C}$. We modeled the probabilities $(\pi_{iA}, \pi_{iB}, \pi_{iC})$ relative to the initial proportion of each strain present in the inoculum for the ith community (p_{iA}, p_{iB}, p_{iC}). We maximized the log of the likelihood function given in (2.5) using the NLMIXED procedure in SAS software (the code is available in online Appendix B). Explanatory variables included in the linear predictor were the proportion of each strain present initially in the inoculum and initial cell density giving the model:

$$\log\left(\frac{\pi_{ij}}{\pi_{iC}}\right) = \beta_{jA} p_{iA} + \beta_{jB} p_{iB} + \beta_{jC} p_{iC} + \beta_{jD} D_i + \log\left(\frac{p_{ij}}{p_{iC}}\right) + u_{ij} \quad \text{for } j = \mathrm{A}, \mathrm{B}. \tag{3.1}$$

where D_i is coded 0 and 1 for low and high initial cell density respectively; u_{iA} and u_{iB} are random effects assumed normally distributed with mean 0 and variance σ_A^2 and σ_B^2, respectively, and the covariance between them is γ_{AB}. We tested for the random components and for interactions among the fixed effect terms. We predicted from the fitted model for a range of initial compositions using Equation (2.6). We tested for a difference between the initial proportion and predicted final probability for each strain: $\hat{\pi}_j - p_j$ and also tested for a difference between $\hat{\pi}_j / p_j - \hat{\pi}_{j'} / p_{j'}$ for each pair-wise comparison (tests of compositional change). The first test assesses each individual strain's competitive outcome in

Table 1. Parameter estimates and standard errors for the fixed terms in the multinomial baseline category multinomial models describing the competition between the three strains of *Mesorhizobium loti*. The first model ($j = \text{A}$) compares strains A and C and the second ($j = \text{B}$) compares B and C.

	$j = \text{A}$		$j = \text{B}$	
Model Term	Coefficient estimate	SE	Coefficient estimate	SE
p_{iA}	−0.7	0.4	−4.4	1.2
p_{iB}	−2.8	1.1	−5.4	1.1
p_{iC}	−1.0	0.6	−2.9	1.3
D_i	−1.3	0.3	−2.2	0.3
$p_{iA} * p_{iB}$	2.1	2.6	16.7	5.0
$p_{iB} * p_{iC}$	2.1	2.6	16.7	5.0

comparison to its expectation while the second test compares the competitive outcomes for each pair-wise set of strains. Approximate standard errors were generated for these tests using the Delta method (Billingsley 1986). A Bonferonni correction was incorporated into the tests to allow for multiple comparisons. Although these tests are carried out on the difference scale, they are presented on the more intuitive ratio scale.

3.3. RESULTS

The final model, after extensive model selection using AIC (online Appendix Table C.1), included the initial proportions of each strain and density in the linear predictor. Two interaction terms $p_{iA}p_{iB}$ and $p_{iB}p_{iC}$ were also included and these interactions were found to be of similar strength and so were constrained to be equal. While the inclusion of the random effects to account for variation from community-to-community was not significant (online Appendix Table C.1), this component was included in the models to respect the structure in the experimental design. The variance components estimates were $\hat{\sigma}_A^2 = 0.021$, $\hat{\sigma}_B^2 = 0.072$ and $\hat{\gamma}_{AB} = -0.039$. The parameter estimates of fixed terms from the fitted multinomial baseline category logit models are given in Table 1. The estimates of the density effect (coefficient of D_i) are negative for $j = \text{A}$ and $j = \text{B}$ indicating increased competitiveness for strain C over both strain A and B at high density when compared with low density conditions. Predicted proportions of nodules occupied by each strain for a range of initial strain proportions at each density are in Table 2 and Figure 1. Compositional change measures (1) and (2) are also presented for these communities (Table 2).

The ratio of predicted proportion to initial proportion for strain C was always significantly higher than one at high density and some of the time at low density showing that strain C performed better than would be expected by its initial proportion in general and particularly at high density (Table 2 compositional change measure 1; Figure 1). Strain C always out-competed strain A and B at high density and generally out-competed both strains at low density (Table 2 Compositional change measure 2; Figure 1). Based on these two compositional change measures, strain C was the most competitive strain, particularly at high density, while there was no out-right winner between strains A and B.

Table 2. The initial proportion of each strain present in the inoculum at each density with predicted proportion of occupied sites for strains A, B and C. Two measures of compositional change in communities are also presented: (1) $\hat{\pi}_j/p_j$ and (2) $\hat{\pi}_j/p_j/\hat{\pi}_{j'}/p_{j'}$ for $j, j' =$ A, B and C and $j \neq j'$. Measure (1) assesses strain competitive outcome in comparison to its expectation and a value equal to 1 indicates it matches expectation. Measure (2) compares one strain's competitive outcome to another strain's and a value equal to 1 suggests the two strains are equally competitive. Deviations from 1 are highlighted in bold (at $\alpha = 0.05$) and are tested as indicated in text.

	Initial proportion			Predicted proportion of occupied sites			Compositional change measure (1)			Compositional change measure (2)		
Density	p_A	p_B	p_C	$\hat{\pi}_A$ (s.e.)	$\hat{\pi}_B$ (s.e.)	$\hat{\pi}_C$ (s.e.)	$\hat{\pi}_A/p_A$	$\hat{\pi}_B/p_B$	$\hat{\pi}_C/p_C$	$\frac{\hat{\pi}_A/p_A}{\hat{\pi}_B/p_B}$	$\frac{\hat{\pi}_C/p_C}{\hat{\pi}_A/p_A}$	$\frac{\hat{\pi}_C/p_C}{\hat{\pi}_B/p_B}$
Low	0.85	0.07	0.08	0.83 (0.040)	0.01 (0.005)	0.17 (0.039)	0.97	**0.08**	2.02	**12.5**	2.1	**26.0**
	0.15	0.74	0.11	0.09 (0.036)	0.48 (0.073)	0.43 (0.067)	0.59	0.65	**3.75**	0.9	**6.3**	**5.7**
	0.13	0.08	0.79	0.06 (0.021)	0.01 (0.010)	0.93 (0.024)	0.45	**0.15**	**1.17**	3.1	**2.6**	**7.9**
	0.41	0.26	0.33	0.28 (0.039)	0.17 (0.035)	0.55 (0.043)	0.67	0.66	**1.69**	1.0	**2.5**	**2.5**
	0.55	0.35	0.10	0.45 (0.063)	0.34 (0.067)	0.21 (0.038)	0.81	0.98	2.14	0.8	2.6	2.2
	0.52	0.07	0.41	0.35 (0.050)	0.01 (0.006)	0.64 (0.050)	0.68	**0.10**	**1.56**	**6.6**	**2.3**	**15.1**
	0.14	0.39	0.48	0.05 (0.014)	0.40 (0.071)	0.55 (0.067)	**0.37**	1.03	1.15	0.4	**3.1**	1.1
High	0.85	0.07	0.08	0.56 (0.062)	0.00 (0.001)	0.44 (0.061)	**0.66**	**0.02**	**5.26**	**28.9**	**8.0**	**231.4**
	0.15	0.74	0.11	0.04 (0.020)	0.11 (0.036)	0.85 (0.040)	**0.30**	**0.14**	**7.38**	2.1	**24.3**	**51.1**
	0.13	0.08	0.79	0.02 (0.007)	0.00 (0.001)	0.98 (0.007)	**0.12**	**0.02**	**1.24**	7.1	**9.9**	**70.5**
	0.41	0.26	0.33	0.11 (0.023)	0.03 (0.011)	0.86 (0.025)	**0.27**	**0.12**	**2.63**	2.3	**9.7**	**22.6**
	0.55	0.35	0.10	0.32 (0.053)	0.11 (0.035)	0.57 (0.054)	**0.58**	**0.30**	**5.89**	1.9	**10.1**	**19.4**
	0.52	0.07	0.41	0.13 (0.027)	0.00 (0.001)	0.87 (0.027)	**0.24**	**0.02**	**2.14**	**15.2**	**8.8**	**134.5**
	0.14	0.39	0.48	0.02 (0.007)	0.07 (0.026)	0.91 (0.027)	**0.16**	**0.19**	**1.89**	0.8	**11.9**	**9.9**

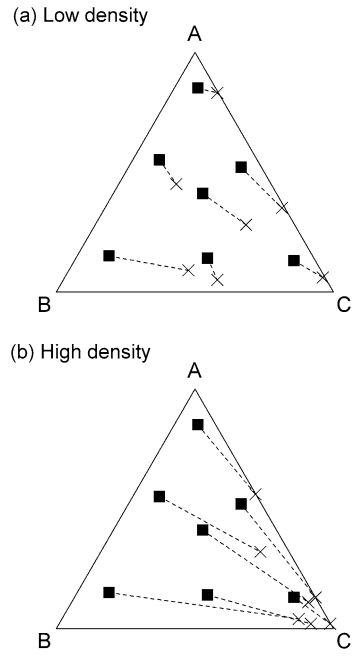

Figure 1. Initial compositions of strains (■) and predicted compositions of nodules infected by each strain (×) for (a) low density and (b) high density. Each symbol represents a proportion of strains A, B and C (either initial or predicted as given in Table 2). Associated tests of significance of community shifts, from initial to final, are in Table 2.

4. DISCUSSION

In this paper we present an experimental design and modeling framework for assessing multinomial responses from multiple species competition studies. For experiments we propose a simplex design which manipulates both species initial relative abundance and density. The modeling approach developed combines multinomial baseline category logit random effects models with multiple offsets that avoids the confounding of initial composition and competitive effects. It provides an analogue to RGRD methods for assessing continuous response competition (Connolly and Wayne 2005; Ramseier, Connolly, and Bazzaz 2005). It also provides a multivariate alternative to the univariate methods used for discrete responses based on Lotka–Volterra models (Leslie 1958; May 2001) that allows for correlation among responses within a community. The method can handle zero values within the multinomial responses, a problem which has plagued other approaches such as compositional data analysis methods previously considered for this type of problem (Aitchison and Bacon-Shone 1984; Aitchison 1986; Billheimer 2001; Billheimer, Guttorp, and Fagan 2001; Aitchison and Kay 2003; Aitchison and Ng 2005). The simplex design is an instrumental part of the framework as it allows the estimation of the parameters in model (2.7) which can identify how competitors react to changing initial presence of other species in addition to its own changing initial presence and to environmental change (Ramseier, Connolly, and Bazzaz 2005; Kirwan et al. 2007).

We applied the methodology to an experiment where three strains of *M. loti* were competing for nodule occupancy on the legume host *L. corniculatus.* The methodology was successful in identifying strain C as the most competitive strain for nodulation among the three rhizobial strains. Strain C occupied a large number of the nodules even when it was least represented in the inoculum particularly at high inoculum density, which in general is an indication of a highly competitive strain (Thies, Benbohlool, and Singleton 1992). This method has also successfully described competition effects in a similar experiment (Duodu et al. 2009). Our experimental design and model would be of interest in a wide range of other competition studies where the response multinomial proportions relative to proportions at an earlier point in time is of interest.

Success in a competitive environment can be highly dependent on the species' ability to compete. If there is no reference point prior in time for comparison, the assessment at a given time can be grossly biased, confounding competitive effects with differences in initial proportional composition of communities (Connolly, Wayne, and Bazzaz 2001). Suppose two species are assessed at a given time and both have occupied the same number of sites. The interpretation might be that the species are equally competitive; however this interpretation changes dramatically in favor of the second species if at time 0 there was 90% of the first and only 10% of second species present. Having at least one reference point prior in time validates conclusions about species competitiveness. The inclusion of the offsets in the model (2.3) incorporates a reference point into modeling. The focus from the model changes from whether π_{ij} differs from π_{iJ} to whether $\frac{\pi_{ij}}{\pi_{iJ}}$ differs from $\frac{p_{ij}}{p_{iJ}}$. If there is a difference between $\frac{p_{ij}}{p_{iJ}}$ and $\frac{\pi_{ij}}{\pi_{iJ}}$, the fixed effects parameter estimates allow us to determine the cause of the shift in community composition and test if individual species initial abundance affects it or if any interactions are occurring between competitors.

Many competition experiments use initial starting points of equal representation of all species in all communities (e.g. Firbank and Watkinson 1985; Hector et al. 1999; Bell et al. 2005). While this design can answer several interesting questions, including some on the effects of changing environment, interpretations are somewhat limited. The β_{jk} coefficients in model (2.7) can not be estimated in a design containing only communities with each species equally present (Ramseier, Connolly, and Bazzaz 2005). We have shown that using an appropriate simplex design allows the fitting of model (2.7) through which we can assess the relative competitiveness of species, whether species interfere with or interact with each other, and the outcome of these interspecific relationships on community composition.

The experimental and analytical framework developed here has been described in the context of competition. For competition to occur a desired resource must be limited and thus species compete to acquire it. In the example illustrated here, we assumed a finite number of sites were available and that when a site was invaded it was no longer available to the other competitors. In the event that the desired resource is not limited, the framework described here would still be appropriate but species performance rather than competition would be assessed, i.e. the 'winning' species is not out-competing other species but rather is quicker at acquiring the desired resource in the available timeframe and, given time, the other species would be able to acquire any resources they desired. The 'compositional

change measures' described above are still appropriate measures to compute, however their interpretations should change from a competitive context to a performance context. In the example here, the total number of sites occupied in each community was similar across all communities and was not affected by cell density ($p = 0.65$) providing evidence that our 'competitive' interpretations are well founded.

There are many parallels between the analysis carried out here and several other analytical approaches, in particular (i) Compositional data analysis (CDA), (ii) Relative growth rate difference analysis (RGRD) (as already discussed) and (iii) Lotka–Volterra differential equations for competing species. Our analysis can be considered to be a compositional data analysis problem (Aitchison 1986) as it is essentially an extension of the 'paired comparison lattice' example presented in Aitchison and Ng (2005). They test whether the shift (the perturbation) in community compositions differs between two treatments. We extend their model to determine the effect of the initial compositions, initial density and any interactions on the community compositional shift and to include the random effect to allow for clustering within each community. An advantage to our approach is that is can handle zero values in the multinomial response which cannot be done using the 'paired comparison lattice' approach. Advances have been made on how to handle zero values in CDA methods; some of these methods rely on the zero values being rounded zeros rather than true zeros (Aitchison and Kay 2003; Martín-Fernández, Barceló-Vidal, and Pawlowsky-Glahn 2003), while others are either work in progress (Aitchison and Kay 2003) or have been developed in the context of group comparisons (Butler and Glasbey 2008) rather than examining competitive changes over time. We have chosen the alternative approach of modeling the underlying discrete multinomial count process and using the models to generate compositions as has also been done with count data with an underlying Poisson process (Aitchison and Ho 1989; Bacon-Shone 2008). Our model is also closely related to the Lotka–Volterra differential equations for competing species. An analogue to the Lotka–Volterra equations to deal with discrete rather than continuous measures of species performance was developed by Leslie (1958). It is easy to show that the approach developed above is also a discrete analogue to the Lotka–Volterra equations. However, previous approaches that respect the discrete nature of responses rarely deal with the multivariate nature of these types of data (e.g. Law and Watkinson 1987). The approach outlined here respects the multivariate nature of the data and also includes the random community term to allow for variability from community-to-community.

SUPPLEMENTARY MATERIALS

Appendix A: The raw experimental data (excel file)
Appendix B: The SAS code to fit the final model (pdf file)
Appendix C: Model fitting with AIC (pdf file)

ACKNOWLEDGEMENTS

CB was funded by the Irish Environmental Protection Agency. Travel support for SD to Ireland and for CB to Norway was funded by COST 852 as a short-term scientific mission. The authors thank the editor, the associate editor and an anonymous referee for their useful comments on earlier drafts.

[Accepted February 2011. Published Online March 2011.]

REFERENCES

Agresti, A. (2002), *Categorical Data Analysis* (2nd ed.), New York: Wiley.

Aitchison, J. (1986), *The Statistical Analysis of Compositional Data*, London: Chapman and Hall. Reprinted in 2003 by The Blackburn Press.

Aitchison, J., and Bacon-Shone, J. (1984), "Log Contrast Models for Experiments With Mixtures," *Biometrika*, 71, 323–330.

Aitchison, J., and Ho, C. H. (1989), "The Multivariate Poisson-log Normal Distribution," *Biometrika*, 76, 643–653.

Aitchison, J., and Kay, J. W. (2003), "Possible Solutions of Some Essential Zero Problems in Compositional Data Analysis," in *Proceedings of Compositional Data Analysis Workshop—CoDaWork '03* (Girona), Electronic publication available at *http://ima.udg.es/Activitats/CoDaWork03*.

Aitchison, J., and Ng, K. W. (2005), "The Role of Perturbation in Compositional Data Analysis," *Statistical Modelling*, 5, 173–185.

Bacon-Shone, J. (2008), "Discrete and Continuous Compositions," in *CODAWORK 08*, Girona, Spain.

Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L., and Lilley, A. K. (2005), "The Contribution of Species Richness and Composition to Bacterial Services," *Nature*, 436, 1157–1160.

Billheimer, D. (2001), "Compositional Receptor Modeling," *Environmetrics*, 12, 451–467.

Billheimer, D., Guttorp, P., and Fagan, W. F. (2001), "Statistical Interpretation of Species Composition," *Journal of the American Statistical Association*, 96, 1205–1214.

Billingsley, P. (1986), *Probability and Measure*, New York: Wiley.

Butler, A., and Glasbey, C. (2008), "A Latent Gaussian Model for Compositional Data With Zeros," *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 57, 505–520.

Connell, J. H. (1983), "On the Prevalence and Relative Importance of Interspecific Competition—Evidence From Field Experiments," *The American Naturalist*, 122, 661–696.

Connolly, J., and Wayne, P. (2005), "Assessing Determinants of Community Biomass Composition in Two-Species Plant Competition Studies," *Oecologia*, 142, 450–457.

Connolly, J., Wayne, P., and Bazzaz, F. A. (2001), "Interspecific Competition in Plants: How Well Do Current Methods Answer Fundamental Questions?" *The American Naturalist*, 157, 107–125.

Cornell, J. A. (2002), *Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data* (3rd ed.), Chichester: Wiley.

de Bruijn, F. J. (1992), "Use of Repetitive (Repetitive Extragenic Palindromic and Enterobacterial Repetitive Intergeneric Consensus) Sequences and the Polymerase Chain-Reaction to Fingerprint the Genomes of *Rhizobium meliloti* Isolates and Other Soil Bacteria," *Applied and Environmental Microbiology*, 58, 2180–2187.

Duodu, S., Brophy, C., Connolly, J., and Svenning, M. M. (2009), "Competitiveness of a Native *Rhizobium Leguminosarum* Biovar *Trifolii* Strain for Nodule Occupancy is Manifested During Infection," *Plant and Soil*, 318, 117–126.

Fagerli, I. L. (2006), "Distribution and Diversity of Rhizobial Soil Populations From Subarctic and Arctic Regions in Norway," Dr. Scient Thesis, University of Tromsø, Norway. ISBN 10 82-90256-019.

Fagerli, I. L., and Svenning, M. M. (2005), "Arctic and Subarctic Soil Populations of *Rhizobium leguminosarum* Biovar *trifolii* Nodulating Three Different Clover Species: Characterisation by Diversity at Chromosomal and Symbiosis Loci," *Plant and Soil*, 275, 371–381.

Firbank, L. G., and Watkinson, A. R. (1985), "On the Analysis of Competition Within Two-Species Mixtures of Plants," *Journal of Applied Ecology*, 22, 503–517.

Goldberg, D. E., and Barton, A. M. (1992), "Patterns and Consequences of Interspecific Competition in Natural Communities—A Review of Field Experiments With Plants," *The American Naturalist*, 139, 771–801.

Hartzel, J., Agresti, A., and Caffo, B. (2001), "Multinomial Logit Random Effects Models," *Statistical Modelling*, 1, 81–102.

Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M. C., Diemer, M., Dimitrakopoulos, P. G., Finn, J. A., Freitas, H., Giller, P. S., Good, J., Harris, R., Hogberg, P., Huss-Danell, K., Joshi, J., Jumpponen, A., Korner, C., Leadley, P. W., Loreau, M., Minns, A., Mulder, C. P. H., O'Donovan, G., Otway, S. J., Pereira, J. S., Prinz, A., Read, D. J., Scherer-Lorenzen, M., Schulze, E. D., Siamantziouras, A. S. D., Spehn, E. M., Terry, A. C., Troumbis, A. Y., Woodward, F. I., Yachi, S., and Lawton, J. H. (1999), "Plant Diversity and Productivity Experiments in European Grasslands," *Science*, 286, 1123–1127.

Iwasa, Y., Nakamaru, M., and Levin, S. A. (1998), "Allelopathy of Bacteria in a Lattice Population: Competition Between Colicin-Sensitive and Colicin-Producing Strains," *Evolutionary Ecology*, 12, 785–802.

Kirwan, L., Lüscher, A., Sebastià, M. T., Finn, J. A., Collins, R. P., Porqueddu, C., Helgadottir, A., Baadshaug, O. H., Brophy, C., Coran, C., Dalmannsdóttir, S., Delgado, I., Elgersma, A., Fothergill, M., Frankow-Lindberg, B. E., Golinski, P., Grieu, P., Gustavsson, A. M., Höglind, M., Huguenin-Elie, O., Iliadis, C., Jørgensen, M., Kadziuliene, Z., Karyotis, T., Lunnan, T., Malengier, M., Maltoni, S., Meyer, V., Nyfeler, D., Nykanen-Kurki, P., Parente, J., Smit, H. J., Thumm, U., and Connolly, J. (2007), "Evenness Drives Consistent Diversity Effects in Intensive Grassland Systems Across 28 European Sites," *Journal of Ecology*, 95, 530–539.

Law, R., and Watkinson, A. R. (1987), "Response-Surface Analysis of Two-Species Competition—an Experiment on *Phleum-arenarium* and *Vulpia-fasciculata*," *Journal of Ecology*, 75, 871–886.

Leslie, P. H. (1958), "A Stochastic Model for Studying the Properties of Certain Biological Systems by Numerical Methods," *Biometrika*, 45, 16–31.

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

May, R. M. (2001), *Stability and Complexity in Model Ecosystems*, Princeton: Princeton University Press.

Nicol, H., and Thornton, H. G. (1941), "Competition Between Related Strains of Nodule Bacteria and Its Influence on Infection of the Legume Host," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 130, 32–59.

Ramseier, D., Connolly, J., and Bazzaz, F. A. (2005), "Carbon Dioxide Regime, Species Identity and Influence of Species Initial Abundance as Determinants of Change in Stand Biomass Composition in Five-Species Communities: An Investigation Using a Simplex Design and RGRD Analysis," *Journal of Ecology*, 93, 502–511.

SAS software, Version 9.1 of the SAS System for Windows. Copyright ©2002–2003 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Scheffe, H. (1963), "Simplex-Centroid Design for Experiments With Mixtures," *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 25, 235–263.

Schoener, T. W. (1983), "Field Experiments on Interspecific Competition," *The American Naturalist*, 122, 240–285.

Suter, M., Ramseier, D., Guesewell, S., and Connolly, J. (2007), "Convergence Patterns and Multiple Species Interactions in a Designed Plant Mixture of Five Species," *Oecologia*, 151, 499–511.

Thies, J. E., Benbohlool, B., and Singleton, P. W. (1992), "Environmental-Effects on Competition for Nodule Occupancy Between Introduced and Indigenous *Rhizobia* and Among Introduced Strains," *Canadian Journal of Microbiology*, 38, 493–500.