Ecological Monographs, 81(4), 2011, pp. 635–663

Random-effects ordination: describing and predicting multivariate correlations and co-occurrences

Steven C. Walker[1] and Donald A. Jackson

Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3G5 Canada

Abstract. Ecology is inherently multivariate, but high-dimensional data are difficult to understand. Dimension reduction with ordination analysis helps with both data exploration and clarification of the meaning of inferences (e.g., randomization tests, variation partitioning) about a statistical population. Most such inferences are asymmetric, in that variables are classified as either response or explanatory (e.g., factors, predictors). But this asymmetric approach has limitations (e.g., abiotic variables may not entirely explain correlations between interacting species). We study symmetric population-level inferences by modeling correlations and co-occurrences, using these models for out-of-sample prediction. Such modeling requires a novel treatment of ordination axes as random effects, because fixed effects only allow within-sample predictions.

We advocate an iterative methodology for random-effects ordination: (1) fit a set of candidate models differing in complexity (e.g., number of axes); (2) use information criteria to choose among models; (3) compare model predictions with data; (4) explore dimension-reduced graphs (e.g., biplots); (5) repeat 1–4 if model performance is poor. We describe and illustrate random-effects ordination models (with software) for two types of data: multivariate-normal (e.g., log morphometric data) and presence–absence community data. A large simulation experiment with multivariate-normal data demonstrates good performance of (1) a small-sample-corrected information criterion and (2) factor analysis relative to principal component analysis. Predictive comparisons of multiple alternative models is a powerful form of scientific reasoning: we have shown that unconstrained ordination can be based on such reasoning.

Key words: *AIC; community ecology; cross-validation; ecoinformatics; information criteria; multivariate; ordination; random effects; statistical ecology.*

Introduction

Biologists are often challenged to understand the interrelationships between large numbers of variables and observations. Some biological subfields are well known to be inherently multivariate, meaning their questions necessarily involve numerous variables (e.g., genomics, community ecology, morphometrics). Additionally, biological data sets are becoming increasingly multivariate with the growth of bio- and ecoinformatics (Michener et al. 1997, Jones et al. 2006). And as Houle (2007) has recently reminded us, biological entities—from macromolecules to the biosphere—cannot be fully described with only one or two variables; thus, uni- and bivariate biological explanations will always be incomplete.

Still, it is difficult to make sense of the interrelationships between large numbers of variables. For example, the well-studied tropical tree data set from Barro Colorado Island, Panama (Condit et al. 2002) contains abundances of 225 species (or variables); to understand the relationships between species in this data set one must examine 25 200 pairwise relationships and almost two million three-way relationships. It is difficult with data sets like this for biologists to develop and convey an intuitive understanding of the information in multivariate data, and, as a result, it is challenging to learn from such data. Put differently, visually representing all of these relationships in a graph is impossible.

These multivariate challenges can be overcome with ordination analysis (e.g., Legendre and Legendre 1998), which summarize sets of observed variables with a smaller, more manageable, number of uncorrelated variables called axes. The axes are obtained by a numerical transformation of the original variables. Ordinations are most useful when they consist of a small number of axes that summarize most of the information in the data (i.e., dimension reduction). It is hoped that these axes will clarify which observational units (e.g., sites, lakes, quadrats, individuals, etc.) are most similar to each other. In contrast, when looking at all of the variables at the same time it is essentially impossible to be able to comprehend patterns of similarity, particularly for very large numbers of variables.

Although dimension reduction is necessary for exploring complex multivariate data, it is not always

Manuscript received 17 May 2011; accepted 20 May 2011. Corresponding Editor: A. M. Ellison.
[1] E-mail: steve.walker@utoronto.ca

sufficient. Effective ordination analyses often couple dimension reduction with inferences about the statistical population from which the observational units were sampled. For example, Clarke (1993) was unsatisfied with the inferential capabilities of multidimensional scaling (MDS) ordination (e.g., Minchin 1987), and so he developed statistical tests for detecting patterns summarized by ordination axes (e.g., nonparametric analysis of dissimilarities). These tests provided tools for addressing questions such as: does species composition (as summarized by ordination axes) differ between sites near a pollution source vs. control sites? Another important development in dimension-reduced inference was constrained ordination (e.g., redundancy analysis [RDA, Rao 1964]; canonical correspondence analysis [CCA, ter Braak 1986]), which allows ecologists to explain multivariate variation using a set of explanatory variables (e.g., Legendre and Legendre 1998). These methods provided tools for addressing questions such as: do spatial or abiotic variables explain more variation in community composition (Borcard et al. 1992)? When ordinations are used in conjunction with population-level inferences, the summaries they provide become more meaningful.

Our focus here is on methods for dimension-reduced prediction, which is less common than either dimension-reduced variation explanation or hypothesis testing. Variation explanation is concerned with estimating measures of our ability to predict (e.g., R^2), and hence is related to prediction. But we stress the importance of going beyond measures of predictive ability to actually making predictions. In particular, we explore ordination methods for predicting the values of the variables at observational units that were not used to construct the ordination (e.g., validation vs. training data). Prediction and forecasting are becoming increasingly important as ecologists are being called upon to predict effects of anthropogenic global change (e.g., Clark 2007). Prediction is also useful for graphically checking the assumptions of models we use to interpret data. For example, few would take seriously data summaries provided by a linear regression (i.e., slope and intercept), when predicting nonlinear data. Such reasoning is useful because it provides a mechanism for the empirical criticism of model-based statistical summaries. However, this important type of model checking is much less common in evaluations of the suitability of summaries provided by ordination axes.

An available method for dimension-reduced prediction in ecology is redundancy analysis (RDA), which involves regressing multivariate response data against a set of predictors, resulting in predicted values for each response variable. The matrix of predicted values is then summarized with principal component analysis (PCA), providing a summary of the variation explained by the predictors. Although RDA is typically used for explaining variation, the fitted regression models can be used to make predictions. Many similar methods have been developed (e.g., CCA [ter Braak 1986], distance-based RDA [Legendre and Anderson 1999], generalized dissimilarity models [Ferrier 2002]); however, all of these methods are asymmetric in the sense that before conducting the analysis each variable is classified as either a predictor or a response.

We focus on symmetric multivariate prediction and dimension reduction, in which variables are not classified a priori as either responses and predictors. For example, such symmetry is sensible in exploratory phases of research where directions of causation are not yet clear. For another example, when the variables are species occurrences or abundances in a community, we usually do not think of some species as responses and others as predictors. Instead we think of associations (e.g., correlations, co-occurrences) between the variables (e.g., species). Our approach is to develop models of such associations and then use these models for both dimension reduction and prediction. Once symmetric models are fitted, each variable may be used as either a predictor or a response in subsequent predictions. For example, the probability of occurrence of species A could be predicted given the presence or absence of species B, but B may also be predicted given A using the same model; symmetric models predict in both directions. Predicting species occurrences using the other species in the community may be more effective than an asymmetric approach using environmental predictors, as species co-occurrences naturally integrate information on a broad range of biotic and abiotic processes (e.g., Austin 2002). Furthermore, it is useful to analyze patterns of co-occurrences before trying to explain them with abiotic predictors, because the predictors may fail to identify important patterns (e.g., Whittaker 1967, Gauch et al. 1974). For example, species interactions can result in co-occurrence patterns that cannot be explained by abiotic variables (e.g., Leathwick and Austin 2001). More generally, accounting for co-occurrence patterns that cannot be explained by measured environmental predictor variables is a major challenge in predictive ecological modeling (Elith and Leathwick 2009).

We base our approach on the latent variable theory of ordination (e.g., Gauch et al. 1974, ter Braak 1985, ter Braak and Prentice 1988, Yee 2004). In this theory, ordination axes are interpreted as estimates of unobserved latent variables, on which the observed variables depend. The process of ordination then consists of estimating the values of these latent variables for each of the observational units. To make these estimates, a parametric model is assumed that specifies an explicit form for the relationships between the axes and the observed variables; for example, a linear relationship is assumed in PCA and a unimodal relationship in correspondence analysis. The classic synthesis by ter Braak and Prentice (1988) remains an excellent reference on how to use these ideas in practice; Yee (2004) updates this theory with fewer approximations.

Most symmetric latent variable ordination models in ecology are treated as fixed-effects models. These types of models treat the axes (i.e., latent variables) as a fixed effect, and therefore make no assumptions about their distribution among the observational units in the sampled statistical population. For this reason, the approach can only be used to make inferences about the specific observational units that happened to be sampled. To make inferences about all observational units within the statistical population, it is necessary to model the variation in the latent variables. Such an approach would treat the values of the axes as random effects. While some of the numerical methods of classical ordination analysis could presumably be modified to treat axes as random-effects, we have never seen this done explicitly, although ter Braak et al. (2003) have explored random-effects ecological cluster analysis. Our purpose here is to present a case for symmetric ordination analysis with random effects, and describe methods for conducting such analyses in practice. Although our random-effects approach can also be used for hypothesis testing and variation explanation, as well as with asymmetric ordination, we focus on symmetric prediction because this is where its advantages will be most apparent.

In this monograph, we provide an extensive first systematic study of the use of random-effects ordination models in ecology. We begin by illustrating the basic idea of random-effects ordination. Drawing on previous statistical work outside of ecology, we describe a linear random-effects ordination model for understanding data with an approximately multivariate-normal distribution and illustrate its use with limnological data. Because PCA is the most commonly used numerical procedure for ordination analysis of linear data, we show how it is related to the linear random-effects ordination model. We report on an extensive simulation experiment comparing two methods for selecting the number of axes for these linear models, while providing the first test of a conjecture of Burnham and Anderson (2002) about model selection in small multivariate samples. Then we describe how to conduct random-effects ordination of presence–absence data and illustrate its use with fish community data. We finish with practical recommendations for using random-effects ordination.

Random Effects Ordination

The primary goal of ordination in general is to summarize a matrix (i.e., table) of data, $\mathbf{Y}$, with a smaller matrix of ordination scores, $\hat{\mathbf{X}}$, that is easier to interpret. The ordination, $\hat{\mathbf{X}}$, is a numerical function—or transformation—of the original data, $\mathbf{Y}$:

$$\mathbf{Y} \rightarrow \hat{\mathbf{X}} \tag{1}$$

where the arrow represents this transformation. The "hat" over the ordination matrix indicates that it is a statistic calculated from the data, $\mathbf{Y}$. Each of these two matrices, $\hat{\mathbf{X}}$ and $\mathbf{Y}$, has n rows; one row for each observational unit (e.g., individuals, lakes, quadrats, sites). Each of the p columns of $\mathbf{Y}$ correspond to a variable, indicating that p variables were measured on each of the n observational units. $\hat{\mathbf{X}}$ has fewer columns than $\mathbf{Y}$; one column for each of the d ordination axes. Each observational unit is characterized by d ordination scores, one score along each ordination axis. As $d < p$, the ordination, $\hat{\mathbf{X}}$, is a dimension-reduced model of the original data, $\mathbf{Y}$. In much the same way that ordinary least-squares regression summarizes the relationship between a predictor variable and response variable with a slope and intercept, ordination summarizes $\mathbf{Y}$ with $\hat{\mathbf{X}}$. Many existing analyses can be used for ordination, as defined above, but we specifically consider a random-effects approach.

The idea of random-effects ordination is similar to the idea of random effects in ANOVA models. For most species, an ANOVA factor such as sex is best treated as a fixed effect because it usually has only two levels, and so it is very unlikely that we would want to make inferences about levels other than male and female. However, a factor such as parent is typically best treated as a random effect because we are usually interested in making inferences about the statistical population of parent (e.g., maternal) effects and not only the effects of those parents that happened to be sampled. In ecological ordination analysis, the observational units (e.g., lakes in a watershed; quadrats in a field) are typically a random sample from a statistical population and therefore the ordination axes will often be best treated as random effects. To describe how random-effects ordination works we consider a series of simple statistical analyses, without going too deeply into the details, which come later and in Appendix A.

As is often the case when describing multivariate analyses, it is useful to begin with a univariate analogue of the type of model we have in mind. Suppose we are interested in inferring the distribution of a single response variable. We would typically measure this variable for a sample of observational units from a statistical population (Fig. 1A). These data are approximately normally distributed and so we fit a normal distribution to them (dashed curve). This distribution is an estimate of the population distribution, and we can therefore use it to make predictive probability statements about unobserved observational units. For example, the 95% prediction interval in Fig. 1A asserts a probability of 0.95 that any particular observation will be no greater than 3.0 and no less than −3.5.

As ecological systems are usually inherently multivariate, we will often need to study multiple response variables simultaneously. To keep our illustrations simple yet multivariate, we consider measuring two variables in a sample of observational units from a statistical population (Fig. 1A and B). These data are approximately bivariate normal and so we fit such a distribution to them. Again, this distribution is an estimate of the population distribution, and can be used

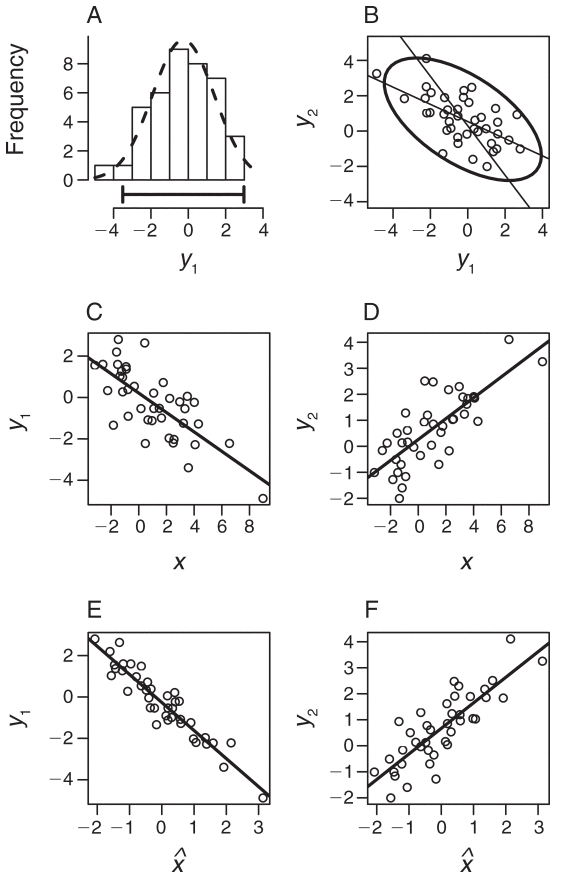

FIG. 1. From univariate to dimension-reduced multivariate predictive inference, with illustrative simulations. The two y's denote bivariate normally distributed simulated response variables, which are described by (A) a histogram and (B) a scatterplot. Some variation in both of the responses can be explained by a simulated explanatory variable, x (C, D). Covariation in the y's can also be explained by an ordination axis, $\hat{x}$ (E, F). See *Random effects ordination* for a detailed explanation.

to make predictive probability statements. For example, this estimate predicts with probability 0.95 that any particular observation will be within the 95% prediction ellipse in Fig. 1B.

The main complication introduced by multivariate inference is that we must model any relationships between response variables. The bivariate normal distribution models such relationships linearly by tilting prediction ellipses, resulting in a model of the correlation between variables. In our example there is a negative correlation which means that the model predicts a low probability density of observing relatively large (or small) values for both variables at the same observational unit. This type of prediction is symmetric because neither variable is singled out as a predictor or response. Symmetric prediction models also have asymmetric forms, based on the expected value of one variable given that the other is known (lines in Fig. 1B; see also ter Braak et al. [2003]).

The need to account for relationships between response variables is what makes multivariate inference more difficult to visualize, and therefore why we often need to reduce the dimensionality of the problem. One simple asymmetric way to reduce dimensionality is to identify and measure an explanatory variable (x-axes in Fig. 1C and D) that is related to both response variables. We fit one linear regression model to each response variable separately (Fig. 1C and D). The residuals of these regressions have a very low correlation (–0.06), despite the strong correlation between the responses (Fig. 1B). Hence controlling for the effect of the explanatory variable eliminates the correlation.

This simple model demonstrates the idea that a correlation may be modeled by identifying an explanatory variable that explains variation in both of the correlated response variables (e.g., ter Braak and Prentice 1988). Such an explanatory variable is a good candidate for an ordination axis—called a direct ordination axis (Whittaker 1967)—because it provides a good one-variable summary of more than one (in this case two) response variables, which is the goal of ordination. Such an asymmetric approach to ordination does not work however if the explanatory variable is not a good predictor of the response variables or if we are curious about whether there may be other ordination axes that better summarize them. Hence we often need a less restrictive approach to deriving ordination models. The approach that we take is based on the concept of latent independent variables, pioneered by ter Braak (1985).

The idea of ter Braak (1985) was to treat the explanatory variable as a latent variable that is not observed; rather, the value of the latent explanatory variable at each observational unit is estimated based on the information in the observable response data. Such estimates (x-axes in Fig. 1E and F) can be made using ter Braak's (1987) reciprocal summation algorithm, which he showed to be equivalent to PCA. Using this approach, even if we do not measure any explanatory variables, we can still estimate them. Such latent variables are called ordination axes and can be used to explain variation and covariation in multivariate data. However, weighted summation does not provide an estimate of the bivariate distribution of the two response variables (Fig. 1B); the reason for this is that the latent variables are treated as fixed effects, which work for asymmetric (Fig. 1C and D) but not for symmetric (Fig. 1B) prediction.

There is another interpretation of PCA, called probabilistic principal component analysis (PPCA; Tipping and Bishop 1999), that uses the latent variables to define random effects. Under this interpretation, the estimated slope parameters that define the relationships between observed and latent variables (Fig. 1E and F) can also be used to infer a multivariate distribution of the observed variables in the statistical population (Fig. 1B). These inferences are made using a model of a two-

step hierarchical process. First, the values of the latent ordination axes at a particular observational unit are randomly generated. Second, the values of the observable variables are randomly generated from a regression model that uses the axes as its independent variables. Such hierarchical processes specify a multivariate probability distribution (Fig. 1B, called a marginal distribution; see *Linear random-effects ordination by factor analysis: Marginal distribution of the observable data*) of the observable variables. By fitting such a model to a sample of data, we can (1) use the fitted multivariate distribution to make predictions about the statistical population from which the sample came and (2) summarize the sample by estimating the values of the random ordination axes. In the next section, we begin to describe random-effects ordination models more precisely. Appendix A provides relevant statistical modeling background for these descriptions.

Linear Random-Effects Ordination by Factor Analysis

We now describe a linear random-effects ordination model that is suitable for multivariate normal data. Such data arise in a wide variety of ecological contexts (e.g., log-transformed morphometric data or species densities). This linear model, called the exploratory factor analysis model (Lawley and Maxwell 1962), is the most important special case of the family of random-effects ordination models that we consider and forms the foundation for nonlinear models (e.g., see *Logistic random-effects ordination by latent traits*). Appendix A contains a general abstract definition of this family.

Assumptions

We use subscripts $i = 1, \ldots, n$, $j = 1, \ldots, p$, and $k = 1, \ldots, d$ to index observational units, variables, and axes. In these models, the fitted mean, $\hat{\mathbf{y}}_i = [\hat{y}_{ij}]$, of the observed response, $\mathbf{y}_i = [y_{ij}]$, at the ith observational unit in the data set, $\mathbf{Y}$, is linearly related to the axis scores, $\mathbf{x}_i = [x_{ik}]$:

$$\hat{y}_{ij} = a_j + \sum_{k=1}^{d} b_{jk} x_{ik} \text{(scalar form)}$$

$$\hat{\mathbf{y}}_i = \mathbf{a} + \mathbf{B}\mathbf{x}_i \text{(matrix form)} \qquad (2)$$

where $\mathbf{a} = [a_j]$ is a column vector of intercepts (one for each variable) and $\mathbf{B} = [b_{jk}]$ is a p-by-d matrix of coefficients relating the p variables to the d axes. There is variation around $\hat{\mathbf{y}}_i$ such that the residuals, $\mathbf{y}_i - \hat{\mathbf{y}}_i$, are normally distributed. This normality assumption means that the jth variable, y_{ij}, has a normal distribution with mean, $\hat{y}_{ij}$, and variance, ψ_j. In psychometrics, these residual variances, ψ_j, and standardized versions of them, are often called *uniquenesses*. This name is meant to remind us that ψ_j quantifies variation that is unique to variable j and therefore not correlated with any of the other $p - 1$ variables. We denote the p residual variances by a column vector, $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_p]'$, where the prime means matrix and vector transpose. Each axis score, x_{ik}, is also assumed to be distributed normally with mean zero and variance one—it is this assumption that makes this model a random effects model. All residuals and axis scores are independent of each other. The $b_{jk}x_{ik}$ terms are random effects (sensu Pinheiro and Bates 2000:8); they are random because the axes are treated as random and they are effects because they represent a deviation from the overall mean, a_j, of each variable, j.

Marginal distribution of the observable data

The normal distributions of the observed response variables are said to be conditional on the latent variables (i.e., axes); this means that the distribution of the observed response variables depends on the values of the latent variables, just as a dependent variable in regression is assumed to depend on independent variables. But in latent variable ordination, the independent variables are not observed. The fixed-effects paradigm of ordination modeling handles this problem by making point estimates of the latent variables (e.g., ter Braak 1987). Here we treat the latent variables as random and therefore to make inferences we need to know the marginal distribution—implied by the model assumptions—of the observed data. The concept of a marginal distribution is important for all models with random effects (Pinheiro and Bates 2000); we provide a brief tutorial in Appendix A. For our purposes, the marginal distribution can be understood as the assumed distribution of the observed data, without explicit reference to the latent variables (e.g., Fig. 1B).

For the linear model, the marginal distribution of $\mathbf{y}$ is multivariate-normal (e.g., Johnson and Wichern 1992). The multivariate-normal distribution is a generalization of the familiar normal distribution, which has two parameters: a mean and a variance. In the multivariate case, the number of parameters increases from 2 to $2p + (1/2)p(p - 1)$. The $2p$ part accounts for the mean and variance of each of the p variables. The $(1/2)p(p - 1)$ part is the number of covariances between each pair of variables. It is customary to collect these parameters into a mean vector, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_p]'$, and a p-by-p covariance matrix, $\mathbf{C}$ (e.g., Johnson and Wichern 1992). The entry in the ith row and jth column of $\mathbf{C}$ contains either the covariance between variables i and j (if $i \neq j$) or the variance of variable i (if $i = j$). The marginal mean and covariance matrix can be written as explicit functions of the intercept, $\mathbf{a}$, coefficients, $\mathbf{B}$, and the residual variances, $\boldsymbol{\psi}$ (Lawley and Maxwell 1962):

$$\boldsymbol{\mu} = \mathbf{a} \qquad (3)$$

$$\mathbf{C} = \mathbf{B}\mathbf{B}' + \Psi \qquad (4)$$

where Ψ is a p-by-p diagonal matrix with the residual variances on the diagonal; such a marginal result is not possible from the fixed-effects perspective, where the

distribution of the observed data is only modeled conditionally on the unobserved latent variables.

Estimation

In real data analysis situations, we do not know the values of the **a**, **B**, and Ψ parameters and so we must estimate them using data. Many procedures exist for making such estimates, but we use maximum likelihood (ML; Lawley and Maxwell 1962; Appendix A). As the marginal distribution of the data in this model is multivariate normal, we need a log-likelihood function that is based on this distribution. The log-likelihood function for all multivariate-normal models is as follows (e.g., Tipping and Bishop 1999):

$$\mathcal{L} = -\frac{n}{2}[p\log(2\pi) + \log|\mathbf{C}| + \operatorname{tr}(\mathbf{C}^{-1}\mathbf{S})] \quad (5)$$

where

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})' \quad (6)$$

and the vertical bars denote the determinant of the matrix between the bars and tr is the matrix trace function, which is the sum of the diagonal elements of its argument. For our ordination model, the mean vector and covariance matrix depend on the p intercepts, **a**, pd coefficients, **B**, and p residual variances, $\boldsymbol{\psi}$. Hence we write this log-likelihood as a function of these parameters: $\mathcal{L}(\mathbf{a}, \mathbf{B}, \boldsymbol{\psi})$.

The maximum likelihood estimate of the intercept **a** does not depend on **B** or $\boldsymbol{\psi}$ and can be given in closed form as the sample mean of the observed dependent variables (Tipping and Bishop 1999):

$$\hat{\mathbf{a}} = \bar{y} \equiv \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i. \quad (7)$$

Therefore, combining Eqs. 3, 6, and 7 shows that **S** at the maximum likelihood estimate is the sample covariance matrix:

$$\hat{\mathbf{S}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'. \quad (8)$$

The intercept is often of limited interest in many nonpredictive multivariate analyses because data are typically centered to have a mean of zero (see Legendre and Legendre 1998). However, an intercept is required for making predictive inferences on the original uncentered measurement scale.

Ideally, we would like to estimate the remaining parameters, **B** and $\boldsymbol{\psi}$, by maximizing

$$\mathcal{L}(\hat{\mathbf{a}}, \mathbf{B}, \boldsymbol{\psi}) = -\frac{n}{2}[p\log(2\pi) + \log|\mathbf{C}| + \operatorname{tr}(\mathbf{C}^{-1}\hat{\mathbf{S}})]. \quad (9)$$

To get an intuitive understanding of this function (technically called a profile log-likelihood function [Royall 1997]), note that it increases as the model covariance matrix, **C**, becomes more similar to the sample covariance matrix, $\hat{\mathbf{S}}$; this means that fitted models are well supported when their covariance matrix resembles the covariance matrix implied by the data. Unfortunately ML estimates for **B** and Ψ are not available in closed form. However efficient iterative algorithms exist for finding ML estimates, for example, we use the factanal function in the R stats package (version 2.13.0; R Development Core Team 2011). For technical reasons, it is not possible to fit models with more than $p + 1/2 - \sqrt{2p + (1/4)}$ axes in this way.

The maximum likelihood estimates, $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Psi}}$, can be substituted into Eq. 4 to calculate the estimated model covariance matrix, $\hat{\mathbf{C}} = \hat{\mathbf{B}}\hat{\mathbf{B}}' + \hat{\boldsymbol{\Psi}}$. This covariance matrix specifies the predictions that the fitted model makes about the interrelationships between variables in the statistical population (e.g., Fig. 1B). In *Probabilistic principal component analysis* we consider an alternative to maximum likelihood for estimating the parameters of this model that is based on PCA, which is more familiar to ecologists. However, we recommend maximum likelihood and explain why in *A simulation experiment*.

Selecting the number of axes

We use information criteria to help make this choice. Information criteria are suitable for use with our predictive approach to ordination, because they are designed to select models with good out-of-sample predictive ability; in particular, the information criteria we consider are estimates of a measure of out-of-sample prediction error called expected Kullback-Leibler information (Appendix A). To select the number of axes, we recommend calculating an information criterion for each of a number of factor analysis models with different numbers of axes, and selecting the model with the lowest criterion.

For factor analysis, the most common information criterion (Akaike 1973; Appendix A) is given by

$$\text{AIC} = -2\mathcal{L}(\hat{\mathbf{a}}, \hat{\mathbf{B}}, \hat{\Psi}) + 2\mathcal{P}$$

where $\mathcal{P}$ is the number of free parameters. In multivariate-normal models it is convenient to consider the number, ν, of parameters that determines the covariance matrix, **C**, separately from the number, $\mathcal{P} - \nu$, of parameters that determines the unconditional mean, $\boldsymbol{\mu}$. Then we can write $\mathcal{P} = \nu + p$ because the unconditional mean, $\boldsymbol{\mu}$, equals the intercept vector, **a**, which contains one intercept for each of the p variables; the problem of finding $\mathcal{P}$ reduces to finding ν. For factor analysis, a naïve guess at ν might be $p + pd = p(1 + d)$; p parameters for $\boldsymbol{\psi}$ and one for each of the pd elements of **B**. However, as is often the case with random latent variable models, ν is less than this naïve guess (Grace 2006): rather $\nu = p(1 + d) - 0.5d(d - 1)$. The reason for the reduction by $0.5d(d - 1)$ is because this ordination model is overparameterized. In statistical jargon we would say that $0.5d(d - 1)$ of the parameters in **B** are not identifiable and so constraints on allowable parameter combinations are made that effectively reduces the number of free

parameters to $v = p + pd - 0.5d(d-1)$ (Appendix A). See Grace (2006) for a discussion on identifiability in random latent variable models. Such a treatment is beyond the scope of this monograph and so we only point out that the number of free parameters in random-effects ordination models is not always so easily determined. For all of the models that we discuss however, we will simply give the number of truly free parameters, $\mathcal{P}$.

AIC in general performs better in large samples. In univariate statistics a popular modification of AIC, called AIC_c, provides better estimates in small samples (Burnham and Anderson 2002). A difficulty is that AIC_c is not appropriate for correcting AIC in small multivariate samples. Burnham and Anderson (2002) suggested that

$$MAIC_c = AIC + 2\frac{\mathcal{P}(\mathcal{P}+\nu)}{np-\mathcal{P}-\nu} \tag{11}$$

be used for models with a multivariate normal distribution of the data such as the factor analysis model. We use M (for multivariate) in front of AIC_c to distinguish it from the univariate version. Note that $MAIC_c \rightarrow AIC$ as $n \rightarrow \infty$ because n is in the denominator of the correction term but not the numerator, and so the two criteria agree in large samples. In effect, $MAIC_c$ penalizes complex models more harshly when sample size is small.

For the purpose of evaluating whether or not any of the candidate ordination models are appropriate, we also compare the candidates with two other models: a full model and a null model. In the full model, all pairs of variables are assumed to potentially have a non-zero correlation, whereas in the null model all variables are assumed to be uncorrelated with each other. If the full model is selected by the information criterion, this will indicate that the data set contains much information on a complex correlation structure and that the fitted ordination models are not capable of summarizing these patterns to the same extent as the full model. If the null model is selected, this will indicate that there is little information in the data about the correlations between variables and therefore it is not appropriate to exploit estimates of these correlations to summarize the data with ordination axes. Further details about these models are in Appendix A.

Estimating the ordination axes and biplots

One benefit of our probabilistic framework is that estimation of the axis scores is conceptually straightforward. This simplicity arises because the laws of probability tell us how to convert what we know (i.e., the data, $\mathbf{y}_i$) into an estimate of what we do not know (i.e., the axis scores, $\mathbf{x}_i$). In particular, for any fitted random-effects latent variable model, Bayes' theorem can be used to specify the conditional mean of $\mathbf{x}_i$, given the observed data, $\mathbf{y}_i$. In the linear model, this conditional mean is

$$\hat{\mathbf{x}}_i = \hat{\mathbf{B}}'\hat{\mathbf{C}}^{-1}(\mathbf{y}_i - \hat{\mathbf{a}}) \tag{12}$$

and so we take this mean as our point estimate of the axes, which is standard in factor analysis. This procedure is analogous to estimating random effects in an ANOVA model. The resulting estimates can be used for the same purposes as classical ordination axes.

It is standard to overlay onto such plots information about how the observed variables relate to the axes. Such plots are called biplots, because they convey information about both the observational units and the variables. In linear models, the variables can be visualized as arrows radiating from the origin of the ordination space, as is common in PCA. The component of the arrow for a particular variable along a particular axis is given by the coefficient in $\hat{\mathbf{B}}$ relating that variable and axis, divided by the observed standard deviation of that variable. The length of the component of an arrow along a particular axis indicates how much variation in that variable is explained by the axis. The direction of an arrow indicates whether that variable increases or decreases along each plotted axis. The angle between arrows tends to be smaller when variables are more correlated.

Fitted values and predictions

The main advantage of treating axes as random effects is that we can use our model to go beyond the biplot, and make predictions and inferences about the variables in the statistical population. The linear random-effects ordination model makes three basic types of predictions: (1) the mean and (2) variance of each variable and (3) the covariance (and correlation) between each pair of variables. These predictions can be checked by plotting data against their associated 95% prediction ellipses (e.g., Fig. 1B).

Using the established properties of the multivariate normal distribution, we can also convert these predicted means, variances, and correlations into regression equations by computing the modeled probability distribution of one set of variables, $\mathbf{y}_1$, given another set, $\mathbf{y}_2$ (Lawley and Maxwell 1973); these two vectors each represent a number of variables measured at the same observational unit. The $\mathbf{y}_1$ and $\mathbf{y}_2$ vectors play the roles of response and predictor variables, but the symmetry of the model allows any particular variable to play either role. Conditional on $\mathbf{y}_2$, the model predicts that $\mathbf{y}_1$ is distributed multivariate normally with mean vector, $\hat{\mathbf{y}}_1$, and residual covariance, $\hat{\mathbf{C}}_{\mathbf{1|2}}$, given by

$$\hat{\mathbf{y}}_1 = \hat{\mathbf{a}}_1 + \hat{\mathbf{B}}_1\hat{\mathbf{B}}_2'\hat{\mathbf{C}}_2^{-1}(\mathbf{y}_2 - \hat{\mathbf{a}}_2) \tag{13}$$

$$\hat{\mathbf{C}}_{\mathbf{1|2}} = \hat{\mathbf{C}}_1 - \hat{\mathbf{B}}_1\hat{\mathbf{B}}_2'\hat{\mathbf{C}}_2^{-1}\hat{\mathbf{B}}_2\hat{\mathbf{B}}_1' \tag{14}$$

where $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{a}}_2$ are the elements of $\hat{\mathbf{a}}$ corresponding to $\mathbf{y}_1$ and $\mathbf{y}_2$; and $\hat{\mathbf{C}}_2$ is the part of $\hat{\mathbf{C}}$ containing variances of and covariances between the variables in $\mathbf{y}_2$. When only one response variable is considered, Eq. 14 specifies its

predicted residual variance, $\hat{c}_{1|2}$, which can be used to construct a 95% prediction interval, $\hat{y}_1 \pm 1.96\sqrt{\hat{c}_{1|2}}$; with two or more responses, we have 95% prediction ellipses (e.g., Fig. 1B) or hyper-ellipses. Appendix A discusses technical advantages of Eq. 13 over ordinary least squares regression.

Limnology Example

To illustrate the use of factor analysis as a random-effects ordination model, we analyzed a limnological data set (Jackson 1988) consisting of four morphological (log lake area, log maximum depth, log volume, and log shoreline length), one topographic (log elevation) and three chemical (pH, calcium, and conductivity) measurements ($p = 8$ variables) on 52 lakes ($n = 52$ observational units) in the Black-Hollow watershed in central Ontario, Canada (Supplement 1). We used our R package, reo (for random effects ordination; Appendix B, Supplement 2), to conduct these analyses. To demonstrate out-of-sample prediction, we randomly split the 52 lakes into training ($n = 34$) and validation ($n = 18$) sets ($\approx$ 2:1 ratio).

We used the pcout function in the R mvoutlier package (*available online*)[2] to detect and remove scatter outliers from the training data (Appendix B). Scatter outliers are observational units that were estimated to have been generated by processes with more variation than the process that generated the majority of the data. Note that more challenging data sets may also contain location outliers, which have a different mean (or median); such data will appear skewed or even slightly bimodal, but it was only necessary to remove scatter outliers in our case. We removed two outliers, which reduced our training sample size to $n = 32$ lakes. Therefore, we expect that the inferences we make will apply to approximately 94% of the lakes in the watershed.

The three-axis model had the lowest $MAIC_c$ (Fig. 2A), suggesting that the four axis and full models are relatively overfitted and should not be used to make inferences. Fig. 2B shows that axes I and II reflect morphology and chemistry respectively. Axis III largely explains correlations between morphology and chemistry variables (Fig. 2C and D), and explains a relatively small fraction of variation: 13.8% compared with 40.1% and 27.7% for the other axes (total, 81.6%; Fig. 2A inset). Elevation only loads strongly onto axis III, suggesting that it is implicated in correlations between morphology and chemistry. Note that the interpretation of inferred relationships between the variables in factor analysis biplots essentially follows the interpretation of PCA biplots.

To further demonstrate that factor analyses can be interpreted in much the same way as other ordinations, we briefly show how to use estimates of the ordination axes (Eq. 12) to explore patterns of similarity among lakes. For example, to explore potential spatial patterns, we plotted the sizes of the symbols representing lakes in proportion to their latitude (Supplement 1) with circles and squares for training and validation lakes. In general, northern lakes tend to be more elevated, larger, less acidic, and calcium rich relative to southern lakes. This pattern holds true in both the training and validation sets. From this exploration, we might consider updating our model to explicitly include latitude; such iterative model building can be effective but is beyond our scope. The biplots also suggest that the validation lakes are representative of the total sample of lakes; predictions about the validation data are thus better described as interpolative rather than extrapolative.

Before considering out-of-sample predictions, we look at the fit of the model to the training data. The simplest such fits are the pair-wise regressions of one observed variable on another (Fig. 3). Note that these regressions are a direct prediction of our symmetric model, and not asymmetric least-squares fits. Therefore, because of symmetry, each variable plays the role of both predictor (x-axes) and response (y-axes). The two thin lines in each subplot define 95% prediction intervals. If the model assumptions are being met we would expect that between only one or two points will fall outside of each plot on average, which is approximately true. Such a good fit indicates that the biplots provide a good summary of the training data.

The most important benefit of random-effects ordinations is their ability to make out-of-sample predictions. To demonstrate, we predicted each variable given all other variables (Eqs. 13 and 14) in the validation sample (Fig. 4). With a validation set of 18 lakes, we expect approximately one lake to fall outside of the 95% prediction intervals. Given that we removed approximately 5% of the training data as outliers, we might expect approximately two lakes to fall outside of the intervals assuming that the proportion of outliers is similar in both the training and validation sets. Although most variables fit the model very well, pH is an exception as six validation lakes fall outside of the prediction intervals for pH. One lake has much higher elevation than predicted by the model, but the two outlying training lakes prepared us for the possibility of outlying validation lakes. The letters in Fig. 4 show the predictions for these training set outliers. Not surprisingly the model does not fit these data well. Such graphical assessments of out-of-sample predictions clarify the extent to which inferences (e.g., correlations between pH and the other variables) can be extended to the statistical population.

Probabilistic Principal Component Analysis

We used the method of maximum likelihood factor analysis to fit the linear random-effects ordination model. We used this method because it has become a standard, outside of ecology, for estimating such models. However, in ecology principal component

[2] ⟨http://CRAN.R-project.org/package=mvoutlier⟩

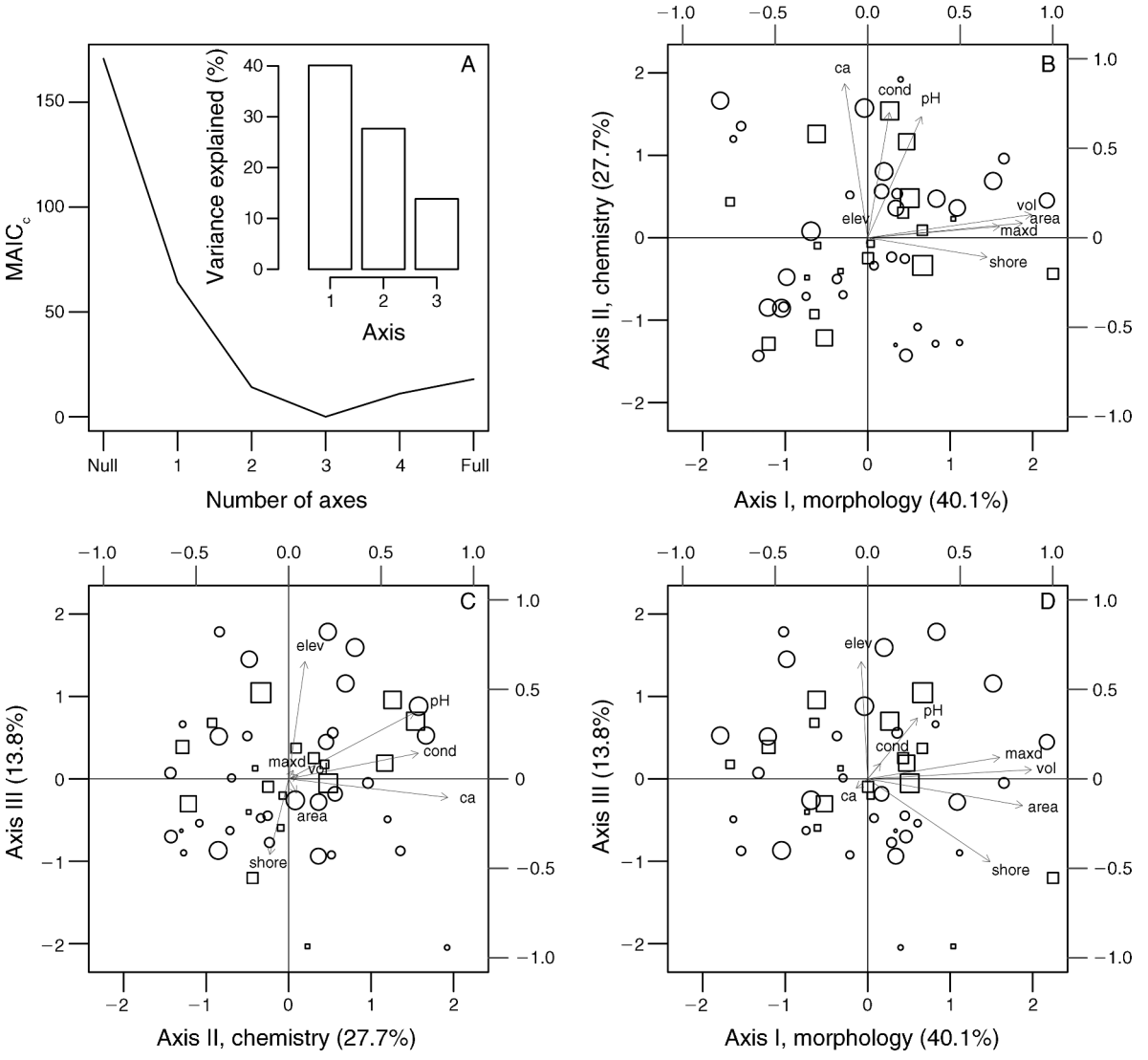

FIG. 2. (A) Model selection and B–D biplots for the limnology data. The inset in panel (A) shows the percentage of variance explained by the axes. Circles and squares in panels (B)–(D) are axis scores for training and validation lakes; shape size is proportional to latitude. Gray arrows (upper horizontal and right-hand vertical axes) give standardized coefficients relating each observed variable to the ordination axes. $MAIC_c$ is the multivariate Akaike information criterion adjusted for small sample sizes.

analysis (PCA) is currently the standard approach for the ordination of linear data within a fixed-effects perspective. Therefore, we evaluate the ability of PCA at estimating the parameters of the linear random-effects ordination model; we conclude that factor analysis is more appropriate from a random-effects perspective.

Although PCA was originally devised as a purely geometric procedure (Pearson 1901), it has subsequently been given various interpretations as a model estimation technique (e.g., ter Braak 1987). Working in the field of machine learning, Tipping and Bishop (1999) showed that it can be used to estimate the parameters of the linear random-effects model described above. They referred to their approach as probabilistic principal component analysis (PPCA).

The main assumption of PPCA is that all of the residual variances are identical:

$$\psi_1 = \ldots = \psi_p = \psi. \tag{15}$$

Under this simplifying assumption, it turns out that the log-likelihood function (Eq. 5) is maximized at values of $\mathbf{B}$ and $\boldsymbol{\psi}$ that depend only on an eigen-analysis of the sample covariance matrix, $\hat{\mathbf{S}}$ (Tipping and Bishop 1999):

$$\hat{\psi} = \frac{1}{p-d}\sum_{j=d+1}^{p}\lambda_j$$

$$\hat{\mathbf{B}} = \mathbf{U}_d(\Lambda_d - \hat{\psi}\mathbf{I})^{1/2} \tag{16}$$

where λ_j is the jth eigenvalue of $\hat{\mathbf{S}}$, $\mathbf{U}_d$ is a matrix whose

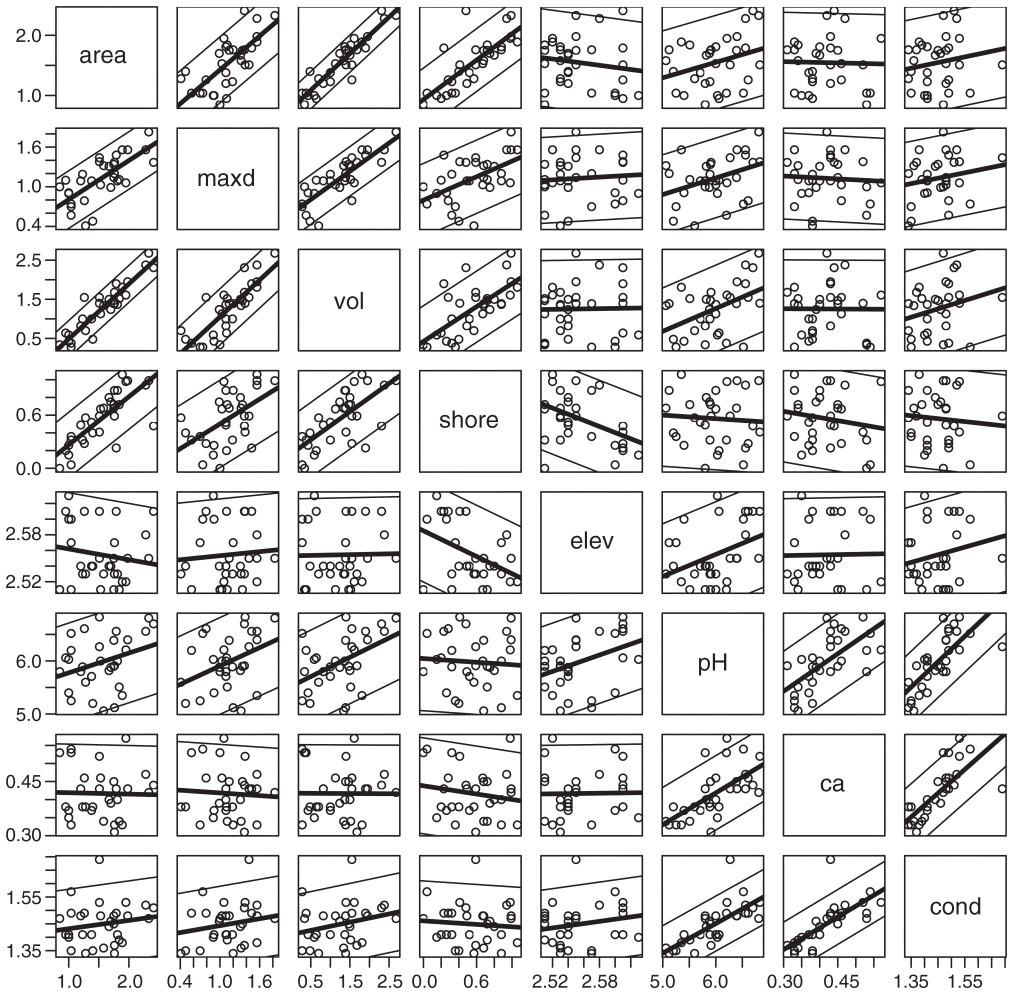

FIG. 3. Pairwise fit of the three-axis model to the training limnology data. Thick lines are model-fitted conditional means of the y-axis variable given the x-axis variable. Thin lines give 95% prediction intervals. Axis labels are listed on the diagonal: area, log lake area (ha); maxd, log maximum depth (m); vol, log lake volume (10^4 m^3); shore, log shoreline length (km); elev, log elevation (m); pH; ca, log calcium concentration (mg/L); cond, log conductivity (μmhos/cm).

columns are the first d eigenvectors of $\hat{\mathbf{S}}$ and $\Lambda_d = \text{diag}(\lambda_1, \ldots, \lambda_d)$. These estimates are based on the same eigen-analysis as used in regular PCA. The difference is that in probabilistic PCA, these estimates lead to a random-effects model that can be used to make inferences beyond the specific sample of observational units (Tipping and Bishop 1999).

We do not expect PPCA to be applicable to ecological and environmental data. It is well-known that eigen-analyses of covariance matrices are unlikely to lead to good ordinations, especially when variables are on different measurement scales as with the Black-Hollow environmental variables (Legendre and Legendre 1998). PPCA is based on the covariance matrix (Eq. 16), suggesting that PPCA will typically not be appropriate for environmental data. Our random-effects approach sheds light on why covariance matrix PCA is not appropriate. The problem stems from the critical assumption that all variables have an identical residual variance (Eq. 15). This is an extremely poor assumption given that many of the variables were measured on completely different measurement scales. Hence the observed variances of the variables differed widely and as a result it is unreasonable to assume equal residual variances, as PPCA does.

Correlation matrix PPCA

While factor analysis is able to avoid the issue of different measurement scales, ecologists have tended to prefer another solution. The usual approach to this issue has been to standardize the data prior to analysis by subtracting the mean and dividing by the standard deviation of each variable, resulting in z scores (Legendre and Legendre 1998). This standardized analysis is known as correlation matrix PCA. Following the standardization, all variables are measured in units of standard deviations. Correlation matrix PCA solves

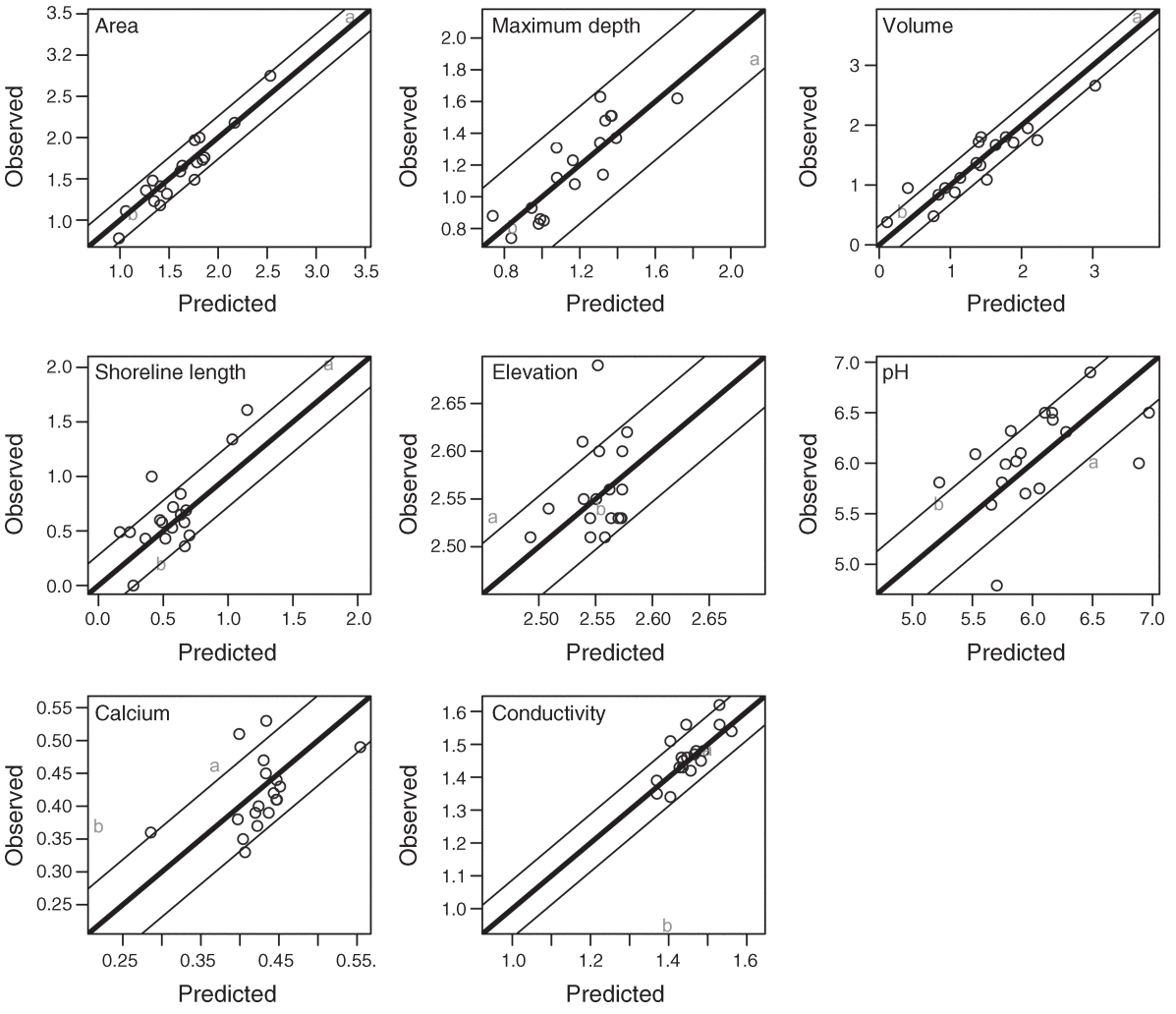

FIG. 4. Out-of-sample predictions of the validation limnology data (open circles) and outlying lakes (gray letters) removed from the training data, using the three-axis ordination model. Predictions for each variable are conditional on all other variables. Thick lines give conditional mean, and thin lines give 95% prediction intervals. Variables are as in Fig. 3.

the problem of differences in the measurement scale by putting all variables on the same scale.

There are potentially numerous ways that correlation matrix PCA can be used to estimate the $\mathbf{B}$ and $\boldsymbol{\psi}$ parameters of the linear random-effects ordination model. To bound the scope of this monograph we will consider only one such estimate that is based on a modification of PPCA, which we refer to as correlation matrix PPCA (PPCAcor). Hence the conclusions that we draw about PPCAcor do not necessarily apply generally to all uses of correlation matrix PCA; however, in *A simulation experiment: PPCAcor in large samples* we develop theory to explore how general these conclusions might be.

The PPCAcor estimates of $\mathbf{B}$ and $\boldsymbol{\psi}$ are

$$\hat{\boldsymbol{\psi}} = \left(\frac{1}{p-d}\sum_{j=d+1}^{p}\lambda_j\right)\mathrm{diag}(\mathbf{DD}') \tag{17}$$

$$\hat{\mathbf{B}} = \mathbf{D}\mathbf{U}_d(\Lambda_d - \hat{\psi}\mathbf{I})^{1/2}$$

where λ_j is the jth eigenvalue of the sample correlation matrix of $\mathbf{Y}$, $\mathbf{U}_d$ is a matrix whose columns are the first d eigenvectors of the sample correlation matrix of $\mathbf{Y}$, $\boldsymbol{\Lambda}_d = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ and $\mathbf{D}$ is a p-by-p diagonal matrix with the sample standard deviations of the p variables on the diagonal. Note that Eq. 17 does not give maximum likelihood estimates and therefore AIC or $\mathrm{MAIC_c}$ should not be used to estimate expected Kullback-Leibler information; instead, we therefore use a cross-validation information criterion (CVIC), which is described in Appendix A.

The critical assumption behind PPCAcor is that all variables leave the same proportion of variance unexplained. So while PPCAcor relaxes the assumption of equal residual variances, as in PPCA, it still assumes proportional equivalence, which is not reasonable given the estimates of proportional residual variances obtained

by factor analysis: area, 0.088; maximum depth, 0.316; volume, 0.005; shoreline length, 0.161; elevation, 0.418; pH, 0.124; calcium, 0.045; conductivity, 0.317. As a result, PPCAcor only partially solves the problem of different measurement scales as we will see in the next section.

A Simulation Experiment

We replicate a simulation experiment of Peres-Neto et al. (2005) to compare factor analysis with PPCAcor, in terms of the number of axes selected by information criteria. We used Peres-Neto et al.'s (2005) 36 multivariate-normal simulation models. Half of these models contained $p = 9$ variables and half contained $p = 18$. All variables have mean zero and variance one. The models differ in their correlation matrices (Figs. 5); the 18-variable matrices are exactly the same as the nine-variable matrices, but with two copies of each variable. For each nine- and 18-variable matrix, we simulated 1000 data sets with $n = 50$ or $n = 100$ observations each, respectively. We fitted PPCAcor ($d = 1, \ldots, p - 2$), factor analysis ($d = 1, \ldots, 5$ when $p = 9$ and $d = 1, \ldots, 12$ when $p = 18$), null and full models to each simulated data set. CVIC was calculated for each fitted PPCAcor model whereas both AIC and $MAIC_c$ were calculated for each factor analysis (see Appendix A for a discussion of the differences between these criteria). All three information criteria were calculated for each null and full model. In total 1 242 000 information criteria were calculated.

This experiment has been used several times (Jackson 1993, Peres-Neto et al. 2005, Dray 2008) to evaluate procedures that select the number of axes, and so it provides a useful benchmark. However, previous interpretations of the results of these experiments have been somewhat incongruent with recent thinking in statistical ecology from the perspective of information criteria (e.g., Burnham and Anderson 2002). For this reason, we will now briefly address some philosophical issues relating to these kinds of experiments.

Two targets for axis selection methods

A reoccurring theme in recent statistical thought is that studies in model selection should explicitly acknowledge the fact that all models are wrong (e.g., Burnham and Anderson 2002); although models can be useful, the concept of a "true" model is unrealistic because nature is more complex than any model. These arguments question the relevance of using simulation studies to evaluate model selection procedures based on their propensity for identifying the known simulation model, because such simulation models are at best an imperfect approximation to nature. As an alternative, model selection procedures may be evaluated based on their out-of-sample prediction error, as measured by expected Kullback-Leibler information and its estimators (e.g., AIC, CVIC; see Appendix A). It has been argued that this alternative is more relevant to the analysis of real data, because making good predictions is possible while identifying the true model is not.

A corresponding counter-theme is that, although it is true that all models are wrong, sometimes nature may be very simple, or at least well approximated by a very simple model, and in these circumstances we would like model selection procedures to lead us to such simple models (Taper 2004). Furthermore, by using probabilistic models, we implicitly account for the complex details of nature using stochastic fluctuations (Taper 2004). Indeed, most published assessments of ordination selection methods seem to use the concept of a true ordination (Jackson 1993, Peres-Neto et al. 2005, Dray 2008). These studies assess methods for choosing the number of principal components, also called the true dimensionality, by analyzing simulated samples from multivariate distributions with simple structure. The criteria are judged based on the likelihood that the selected number of components is equal to the true dimensionality.

Because these philosophical issues are far from resolved (e.g., Taper 2004), we assess the ability of our information criterion approach to both identify the model with (1) the true number of axes (i.e., the "true" model) and (2) the lowest Kullback-Leibler prediction error (i.e., the KL-best model). Still, it is important to keep in mind that the ability to select the 'true' number of axes is not the primary goal of the information criterion approach. Expected Kullback-Leibler information (i.e., prediction error) was calculated for each model as the simulation average Kullback-Leibler information, using the formula of Bedrick and Tsai (1994) for multivariate-normal models, over 500 simulated data sets. Determining the true number of axes requires a consideration of the conspicuous block structure of Peres-Neto et al.'s (2005) matrices. We use the number, $g = g' + \upsilon$, of variable groups to distinguish different types of group structure; the two components, g' and υ, represent two different types of groups. The number of groups composed of more than one variable, such that all within-group correlations are greater than zero, is denoted by g'. The number of variables that are uncorrelated with all other variables is denoted by υ. For example, the nine-variable matrix 1 has $g' = 3$ and $\upsilon = 0$ whereas the nine-variable matrix 11 has $g' = 2$ and $\upsilon = 3$ (Fig. 5). Each of the g' groups exhibiting non-zero within-group correlations should require one single axis to summarize its covariance, because all variables in the group are correlated in exactly the same way to all other variables in the group. The υ uncorrelated variables should not be summarized by any axis, because axes should summarize covariation only. Hence the true number of axes is g', not g.

PPCAcor in large samples

We present some theory on the behavior of PPCAcor in large samples because it provides insight into when PPCAcor might be appropriate and it generates predictions for the simulation experiment. Matrices 1, 4, 6, 7, and 9–18 can all be considered special cases of a

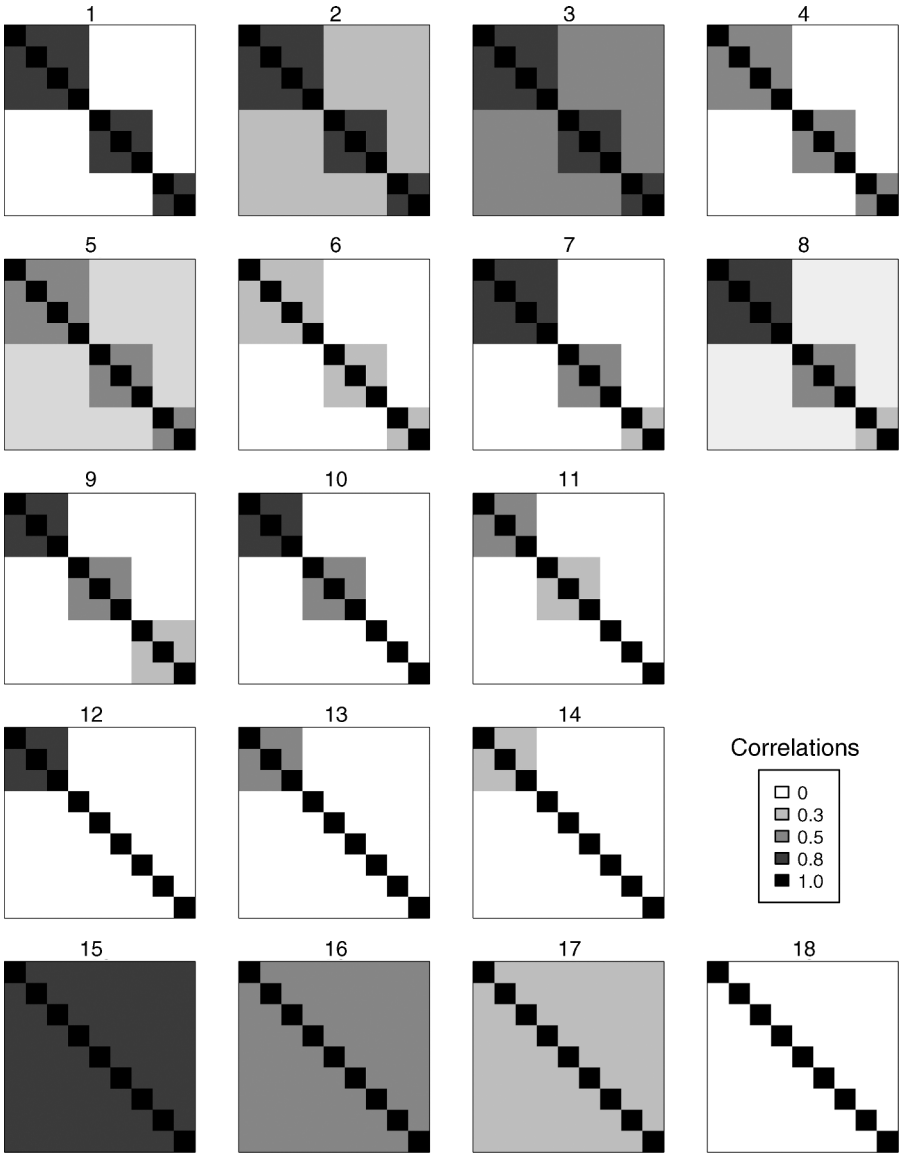

FIG. 5. Peres-Neto et al.'s (2005) nine-variable correlation matrices. The small squares in each matrix represent the correlation (indicated by shading) between two variables. The 18-variable matrices are identical to the nine-variable matrices, but with two copies of each variable.

general type of correlation matrix. In such matrices, all pairs of variables between groups have correlation zero. Let the within-group correlation in group m be r_m and the number of variables in group m be p_m, so by definition,

$$p = \sum_{m=1}^{g} p_m.$$

Substantial insight into how PPCAcor works on such matrices can be obtained by letting sample size approach infinity: $n \to \infty$. In such circumstances the sample correlation matrix will equal the true correlation matrix. Carrying out a PPCAcor on the true correlation matrix will therefore characterize the behavior of the method when sample sizes are large.

We now describe a simple equation (Appendix A) for establishing when a g'-axis PPCAcor model will separate covariance from residual variance in large samples. In the infinite sample limit, the g'-axis PPCAcor estimate of the within-group covariance is

$$\hat{\sigma}_m = r_m\left(1 - \frac{1}{p_m}\right) + \bar{r}\left(\frac{1}{p_m}\right) \tag{18}$$

where

$$\bar{r} = \frac{1}{p - g}\sum_{m=1}^{g}(p_m - 1)r_m \tag{19}$$

is a weighted average of the within-group correlations. Because all of the variables have variance one, the true covariances are identical to the true correlations. Therefore if g' axes summarize all of the covariance, then the estimated covariance, $\hat{\sigma}_m$, should equal the true correlation, r_m, in every group. Setting $\hat{\sigma}_m = r_m$ in Eq. 18 we see that this condition is met whenever $r_m = \bar{r}$ for every group, which can only be true if the within-group correlations are identical in all groups

$$r_1 = \ldots = r_g. \tag{20}$$

Based on this result, we conclude that PPCAcor is appropriate for matrices 1–6 and 15–18, because these matrices have identical within-group correlations for all g' groups. On the other hand, matrices 7–14 are expected to require more than g' axes to summarize all of the covariance.

Eq. 18 provides further insight. If the number of variables in each group is large, then $\hat{\sigma}_m \approx r_m$. Hence we expect the 18-variable matrices to be less influenced by differences in within-group correlations compared with the nine-variable matrices. However, as sample size gets truly large, there may be enough information to detect small differences between $\hat{\sigma}_m$ and r_m. Therefore, large numbers of variables are expected to mitigate, but not eliminate, the problems with PPCAcor.

These results have implications for the analysis of real ecological study systems. Even if the correlations between real variables are approximately clustered into groups, these groups will differ somewhat in their within-group correlations. Therefore, these results suggest that PPCAcor may not be appropriate for the ordination of real ecological data because it will almost always summarize both covariation and residual variation on at least some axes.

Finally, these conclusions are not restricted to the particular use of the correlation matrix eigen-analysis that characterizes PPCAcor. For example, Johnson and Wichern (1992) describe another use of this eigen-analysis for which the analogue of Eq. 18 is (Appendix A),

$$\hat{\sigma}_m = r_m\left(1 - \frac{1}{p_m}\right) + \left(\frac{1}{p_m}\right). \tag{21}$$

Following the same logic as with Eq. 18, the conclusion here is that g' axes will adequately summarize covariance if the within-group correlations are perfect for every group (i.e., $r_1 = \ldots = r_g = 1$); this is an unreasonable condition for real ecological data. Therefore, despite the problems with PPCAcor, it will likely be more appropriate than the method described by Johnson and Wichern (1992).

Simulation study expectations and results

Bias for expected Kullback-Leibler information.—Information criteria are constructed to estimate expected Kullback-Leibler information (i.e., prediction error) with minimal bias. We explored bias by plotting the simulation average information criteria against the expected Kullback-Leibler information for various models. Bias is approximately zero when there is no difference between the average information criterion and expected Kullback-Leibler information. CVIC will almost certainly be a nearly unbiased estimate of expected Kullback-Leibler information (see Appendix A for a justification). We therefore checked four simulation models to confirm that the bias is indeed negligible. It is possible that AIC will be badly biased with factor analysis estimation. We know of no theory for predicting whether the sample sizes used in this study are sufficiently large to reduce this bias. There is also no theory underlying Burnham and Anderson's (2002) conjecture that MAIC_c (Eq. 11) will reduce bias relative to the AIC case. To our knowledge, the present study provides the first evaluation of this conjecture; to address this research gap we evaluated the bias of AIC and MAICc for all factor analysis models.

As expected, CVIC is essentially an unbiased estimate of expected Kullback-Leibler information (Fig. 6 gives four examples of this lack of bias). However, AIC was noticeably biased for the factor analysis models considered (Fig. 7). As hypothesized by Burnham and Anderson (2002), MAIC_c reduced this bias in every case but did not eliminate it (Fig. 7). For some simulation models (such as nine-variable matrix 18) MAIC_c actually overcompensated for the bias in AIC such that the absolute value of the bias was relatively similar between the two criteria, but just in different directions. However MAIC_c very rarely resulted in a worse absolute bias, suggesting that MAIC_c should be recommended over AIC for factor-analysis model selection. Accordingly, we will not present any more AIC results.

Comparing the true and KL-best models.—In the general literature on information criteria, the KL-best model is typically less complex than the true model (e.g., Burnham and Anderson 2002). This tendency results because less complex models often produce better predictions when the sample size is too small to reliably estimate the parameters of more complex models. We apply this logic to ordination selection, but add a few caveats.

The number of axes of the KL-best factor analysis model should tend to be less than or equal to the true number, g', of axes, especially for simulations with low within-group correlations; groups with low correlations have little covariance to be explained by ordination axes. By the same logic, we also expect this tendency for PPCAcor with matrices 1–6 and 15–18. However, as explained in *PPCAcor in large samples*, matrices 7–14 are expected to require more than the true number of axes. In these cases, it will therefore not be surprising if the KL-best PPCAcor model has more than g' axes.

There were very few cases of disagreement between the KL-best and true models (Tables 1, 2, 3, 4), and so we only highlight these few exceptions. As predicted,

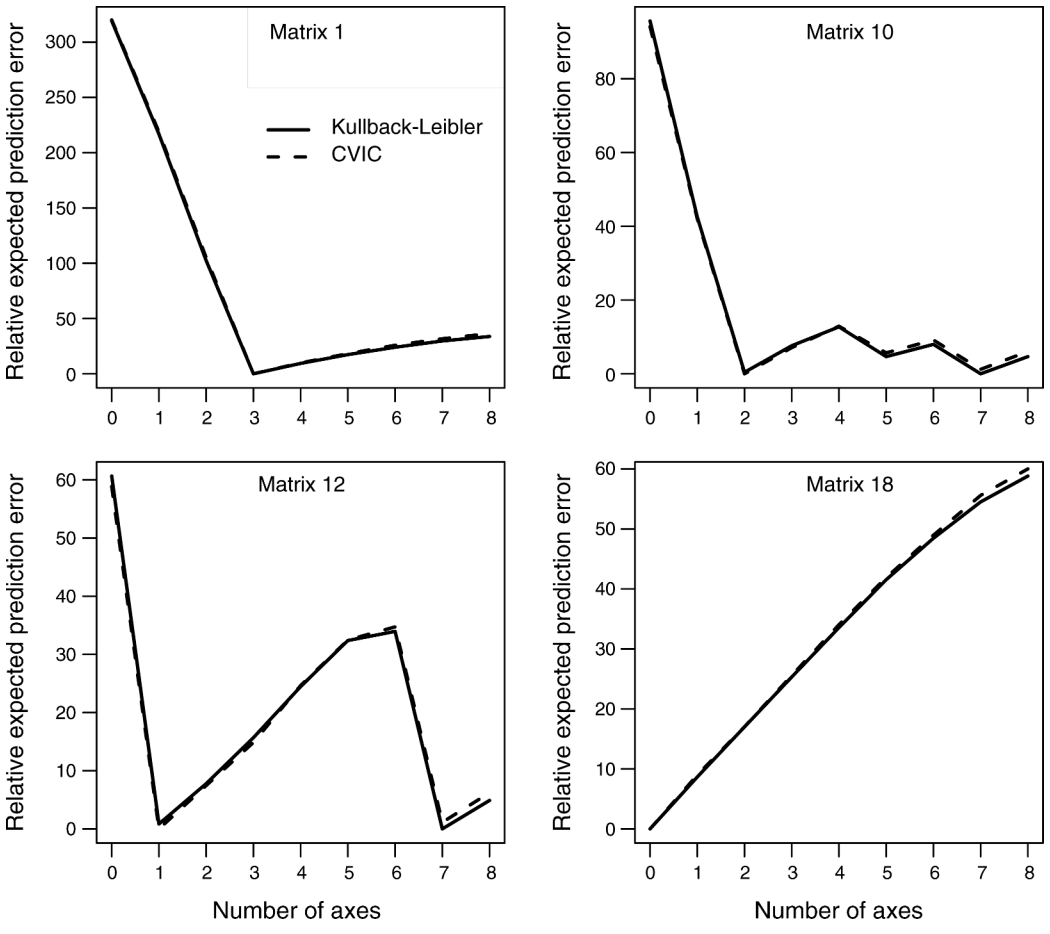

FIG. 6. Correlation matrix probabilistic principal component analysis: expected Kullback-Leibler information and average cross-validation information criterion (CVIC) for four of the matrices of Peres-Neto et al. (2005). Numbers of axes for each ordination model are on the *x*-axes.

the three exceptions for factor analysis (Table 1, matrices 6–8) had low within-group correlations and the KL-best had fewer axes than the true model. A similar pattern was observed for PPCAcor with the nine-variable matrix 6 (Table 2). The only other PPCAcor disagreement involved a KL-best model that was more complex than the true model (Table 2, matrix 8); in agreement with theory (*PPCAcor in large samples*), the within-group correlations of this matrix differed between groups.

Model-selection frequencies.—Because information criteria are meant to estimate expected Kullback-Leibler information, the most frequently selected model will typically be the KL-best model; but this need not be the case in all scenarios. For example, if the information criterion is badly biased, the most frequently selected model will differ from the KL-best model. Such differences are more likely to occur with factor analysis estimation because AIC and $MAIC_c$ are much more likely than CVIC to be biased. Bias is more likely to cause disagreement between the KL-best and most frequently selected models when all models have similar expected Kullback-Leibler information.

The simple structure of the Peres-Neto simulation models ensures that samples from the same model will tend to show very similar patterns. Hence a good model-selection procedure should choose the same (or similar) model(s) in repeated samples. Otherwise, the procedure would be capable of representing similar structure in a variety of ways, which makes for interpretation difficulties. Ideally a model-selection procedure will choose the same model with high probability. However it is also acceptable if a procedure selects one of two adjacent models with high probability. Two models are defined as adjacent if they have d and $d+1$ axes. In contrast, model selection variability over models with very different numbers of axes indicates that very different candidate ordinations have similar capabilities for summarizing the patterns generated by the simulation model. Such variation suggests that the list of candidates is too long. We predicted that PPCAcor model selection would be more variable for matrices 7–14 (see *PPCAcor in large samples*). Factor analysis estimation should lead to

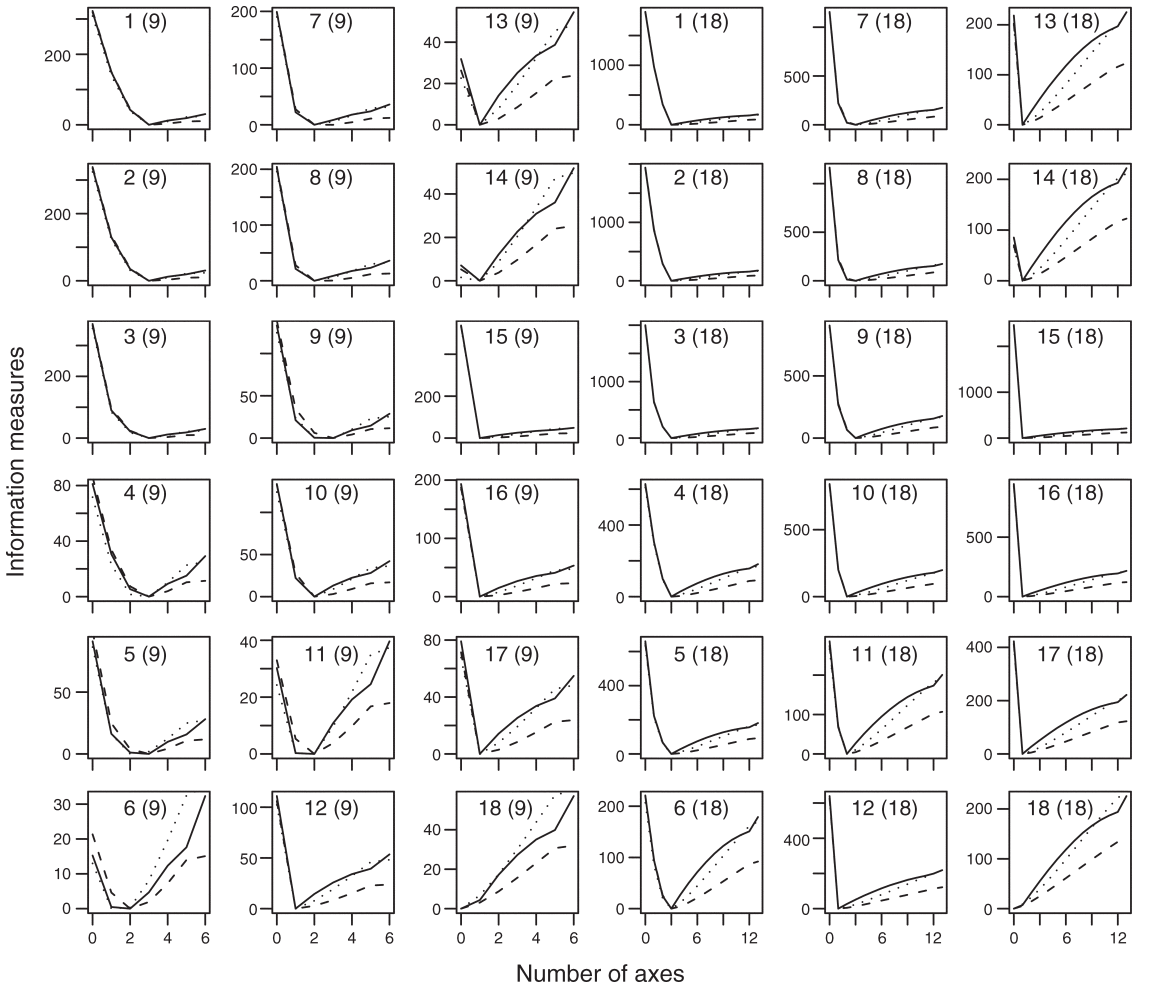

FIG. 7. Factor analysis of expected Kullback-Leibler information (solid), average AIC (dashed), and average $MAIC_c$ (dotted) for all of the matrices of Peres-Neto et al. (2005) (number of variables in parentheses). Numbers of axes for each ordination model are on the *x*-axes.

much less of this variation because it is more efficiently able to summarize residual variation without axes.

As predicted, the most frequently selected model was almost always the KL-best model, with only four exceptions (Tables 1, 2, 3, 4). Also as predicted, all of those exceptions were for factor analysis (Table 1; matrices 5, 6, 9, and 11). The reason for these four exceptions may relate to the fact that the expected Kullback-Leibler information for these matrices was very similar for more than one model (Fig. 7), thereby exaggerating the effects of a small bias in $MAIC_c$. For every 18-variable simulation matrix, the most frequently selected model was the KL-best model for both estimation procedures. The factor analysis procedure selected one single (or two adjacent) model(s) with greater than 80% frequency for all matrices (Tables 1 and 3). PPCAcor had similarly small variability to factor analysis for matrices 1–6 and 15–18, but was highly variable for the remaining matrices as predicted. Overall, factor analysis with $MAIC_c$ provided more reliable model selection than PPCAcor.

Logistic Random-Effects Ordination by Latent Traits

A very common type of data in ecology for which transformation to normality is not possible are presence–absence data, where the presence or absence of *p* species (i.e., variables) is recorded as either a one or zero respectively at each of *n* sites (i.e., observational units). In univariate contexts, such presence-absence data are commonly handled by logistic regression. Here we show that a logistic random-effects ordination model is also available for presence-absence data. Our logistic ordination model would be called a latent trait model (e.g., Rizopoulos 2006) in the psychometric field of item response theory (e.g., Lord 1986). Our latent trait model is closely related to ter Braak's Gaussian-logit model (see also Yee 2004), the major difference being whether

TABLE 1. Factor analysis model-selection percentages using multivariate Akaike information criterion adjusted for small sample sizes ($MAIC_c$) for the nine variable simulations.

Matrix	KL	g'	Number of axes						
			Null	1	2	3	4	5	Full
1	3	3	0.0	0.0	0.0	**98.0**	1.9	0.0	0.1
2	3	3	0.0	0.0	0.3	**97.8**	1.8	0.0	0.1
3	3	3	0.0	0.0	3.4	**95.1**	1.5	0.0	0.0
4	3	3	0.0	1.3	44.9	**52.9**	0.9	0.0	0.0
5	3	3	0.0	6.5	**57.0**	36.1	0.4	0.0	0.0
6	2	3	7.3	**48.9**	38.8	5.0	0.0	0.0	0.0
7	2	3	0.0	2.5	**84.5**	13.0	0.0	0.0	0.0
8	2	3	0.0	2.5	**86.0**	11.5	0.0	0.0	0.0
9	3	3	0.0	1.3	**51.8**	46.6	0.3	0.0	0.0
10	2	2	0.0	1.9	**95.6**	2.5	0.0	0.0	0.0
11	2	2	0.6	**51.7**	46.5	1.2	0.0	0.0	0.0
12	1	1	0.0	**94.6**	5.3	0.1	0.0	0.0	0.0
13	1	1	2.2	**93.4**	4.4	0.0	0.0	0.0	0.0
14	1	1	47.5	**50.3**	2.2	0.0	0.0	0.0	0.0
15	1	1	0.0	**95.2**	4.8	0.0	0.0	0.0	0.0
16	1	1	0.0	**94.6**	5.3	0.1	0.0	0.0	0.0
17	1	1	0.0	**96.1**	3.8	0.1	0.0	0.0	0.0
18	0	0	**92.6**	7.3	0.1	0.0	0.0	0.0	0.0

Notes: KL identifies the number of axes associated with the Kullback-Leibler-best model; g' is the true number of axes. Boldface numbers indicate the model with the highest selection percentage.

axes are treated as random or fixed (ter Braak 1985:866–867).

Assumptions

The major difference between linear and logistic random-effects ordination is that the linear model (Eq. 2) relating axes and variables is replaced by a logistic model:

$$\text{logit}(\hat{y}_{ij}) = a_j + \sum_{k=1}^{d} b_{jk}x_{ik}. \quad (22)$$

The probabilities of occurrence, $\hat{y}_{ij}$, in this model are determined by a monotonic S-shaped function of the latent variables. However, often with real ecological data we would like our ordination axes to reflect variation in environmental gradients; for example, physiological stress models often predict that species will take their maximum probability of occurrence at intermediate sites along such gradients (e.g., ter Braak and Prentice 1988). Yet the above logistic ordination model predicts maximum probabilities of occurrence at extreme, not intermediate, sites. ter Braak (1985) introduced the Gaussian-logit model, which showed that it is possible to relax this assumption of monotonicity by modeling the probabilities of occurrence as

TABLE 2. Correlation matrix probabilistic principal component analysis (PPCAcor) cross-validation information criterion (CVIC) model-selection percentages for the nine variable simulations.

Matrix	KL	g'	Number of axes								
			Null	1	2	3	4	5	6	7	Full
1	3	3	0.0	0.0	0.0	**91.9**	7.1	0.9	0.1	0.0	0.0
2	3	3	0.0	0.0	0.0	**92.1**	6.5	1.3	0.1	0.0	0.0
3	3	3	0.0	0.0	0.1	**92.9**	6.4	0.4	0.1	0.0	0.1
4	3	3	0.0	0.2	7.5	**87.5**	4.4	0.3	0.1	0.0	0.0
5	3	3	0.0	4.0	26.7	**64.2**	4.3	0.7	0.1	0.0	0.0
6	2	3	4.5	22.4	**38.5**	31.6	2.6	0.4	0.0	0.0	0.0
7	3	3	0.0	0.0	10.5	**33.1**	15.5	7.6	29.5	3.2	0.6
8	6	3	0.0	0.0	12.2	27.1	14.5	10.5	**33.1**	2.4	0.2
9	3	3	0.0	0.0	3.6	**57.5**	8.6	6.4	4.5	17.6	1.8
10	2	2	0.0	0.1	**37.8**	9.5	4.7	11.8	7.8	25.7	2.6
11	2	2	2.5	27.0	**54.2**	12.9	2.9	0.4	0.0	0.1	0.0
12	1	1	0.0	**43.9**	11.6	2.2	0.6	0.3	1.1	37.5	2.8
13	1	1	3.7	**78.2**	14.7	2.8	0.4	0.2	0.0	0.0	0.0
14	1	1	40.9	**48.7**	8.4	1.9	0.1	0.0	0.0	0.0	0.0
15	1	1	0.0	**94.2**	5.4	0.4	0.0	0.0	0.0	0.0	0.0
16	1	1	0.0	**95.0**	4.3	0.5	0.2	0.0	0.0	0.0	0.0
17	1	1	0.0	**94.1**	4.8	0.9	0.2	0.0	0.0	0.0	0.0
18	0	0	**87.1**	10.4	2.1	0.3	0.0	0.1	0.0	0.0	0.0

Note: Boldface numbers indicate the model with the highest selection percentage.

TABLE 3. Factor analysis model-selection percentages using $MAIC_c$ for the 18-variable simulations.

Matrix	KL	g'	Number of axes							
			Null	1	2	3	4	5	>6	Full
1	3	3	0.0	0.0	0.0	**96.0**	3.9	0.1	0.0	0.0
2	3	3	0.0	0.0	0.0	**96.4**	3.4	0.2	0.0	0.0
3	3	3	0.0	0.0	0.0	**95.5**	4.5	0.0	0.0	0.0
4	3	3	0.0	0.0	0.0	**95.7**	4.2	0.1	0.0	0.0
5	3	3	0.0	0.0	0.1	**96.3**	3.5	0.1	0.0	0.0
6	3	3	0.0	0.0	7.0	**89.3**	3.5	0.2	0.0	0.0
7	3	3	0.0	0.0	6.0	**90.8**	3.0	0.2	0.0	0.0
8	3	3	0.0	0.0	9.8	**86.4**	3.8	0.0	0.0	0.0
9	3	3	0.0	0.0	0.0	**96.5**	3.4	0.1	0.0	0.0
10	2	2	0.0	0.0	**95.9**	4.0	0.1	0.0	0.0	0.0
11	2	2	0.0	0.0	**94.9**	5.1	0.0	0.0	0.0	0.0
12	1	1	0.0	**94.1**	5.5	0.4	0.0	0.0	0.0	0.0
13	1	1	0.0	**95.6**	4.4	0.0	0.0	0.0	0.0	0.0
14	1	1	0.1	**94.3**	5.4	0.2	0.0	0.0	0.0	0.0
15	1	1	0.0	**94.2**	5.5	0.3	0.0	0.0	0.0	0.0
16	1	1	0.0	**94.4**	5.6	0.0	0.0	0.0	0.0	0.0
17	1	1	0.0	**95.3**	4.5	0.2	0.0	0.0	0.0	0.0
18	0	0	**92.8**	6.8	0.4	0.0	0.0	0.0	0.0	0.0

Note: Boldface numbers indicate the model with the highest selection percentage.

unimodal functions of the latent independent variables

$$\text{logit}(\hat{y}_{ij}) = a_j + \sum_{k=1}^{d} b_{jk}x_{ik} + c_{jk}x_{ik}^2 \qquad (23)$$

where the c coefficients are constrained to be negative, which ensures unimodal relationships rather than valley-shaped relationships.

As in logistic regression, each y_{ij} is assumed to have a Bernoulli distribution with probability of occurrence, $\hat{y}_{ij}$. There is no need to specify a parameter for residual variation, as was required for normal ordination models, because $\hat{y}_{ij}$ determines both the mean and variance of the Bernoulli distribution. Because the axes are treated as random effects, we need to specify a distribution for them as well. As in our multivariate-normal model, we assume that each axis score has a standard normal distribution.

Estimation and model selection

These logistic random-effects ordination models can be fitted by maximum likelihood using the ltm function in the R ltm package (Rizopoulos 2006), provided that the number of axes does not exceed $d = 2$. This function uses an iterative procedure that improves parameter estimates at each iteration. However, for ecologically relevant unimodal models in particular, we found that the algorithm was not very stable (Appendix B). It frequently converges to coefficient estimates of positive and negative infinity, implying unrealistically strong associations between species that cross-validate very

TABLE 4. PPCAcor CVIC model-selection percentages for the 18-variable simulations.

Matrix	KL	g'	Number of axes																	
			Null	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	15	Full
1	3	3	0.0	0.0	0.0	**96.8**	2.9	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	3	3	0.0	0.0	0.0	**97.5**	2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	3	3	0.0	0.0	0.0	**97.4**	2.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	3	3	0.0	0.0	0.0	**97.9**	1.9	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	3	3	0.0	0.0	0.0	**97.9**	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	3	3	0.0	0.0	1.4	**94.4**	3.9	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	3	3	0.0	0.0	0.3	**69.2**	3.8	2.5	1.6	0.4	0.1	0.5	1.6	17.8	2.0	0.2	0.0	0.0	0.0	0.0
8	3	3	0.0	0.0	0.6	**69.2**	4.0	1.4	1.2	0.6	0.3	0.6	1.5	18.6	1.8	0.2	0.0	0.0	0.0	0.0
9	3	3	0.0	0.0	0.0	**91.5**	3.8	0.5	0.1	0.0	0.2	0.1	0.0	0.0	0.2	3.2	0.4	0.0	0.0	0.0
10	2	2	0.0	0.0	**74.7**	6.8	0.9	0.2	0.1	0.6	1.5	0.1	0.2	0.5	1.8	11.5	0.6	0.5	0.0	0.0
11	2	2	0.0	0.0	**95.1**	4.7	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	1	1	0.0	**47.7**	6.2	1.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	41.3	2.9	0.0	0.1	0.0
13	1	1	0.0	**93.9**	5.4	0.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	1	1	0.1	**94.5**	4.8	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	1	1	0.0	**96.9**	3.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	1	1	0.0	**98.1**	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	1	1	0.0	**97.7**	2.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	0	0	**94.4**	5.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Note: Boldface numbers indicate the model with the highest selection percentage.

poorly; infinite coefficients lead to fitted probabilities of co-occurrence of zero or one, which implies far too much certainty to be a good model of ecological data. Furthermore, runs from different random initial values would often lead to vastly different estimates. These numerical difficulties are not entirely surprising, as similar experiences with ter Braak's Gaussian-logit model motivated statistical ecologists to recommend using eigen-analyses such as correspondence analysis to approximate model-based fits (Gauch 1982, ter Braak 1985). Another difficulty with the ltm function is that it does not constrain quadratic coefficients to be negative, allowing valley-shapes instead of unimodal relationships between species and the axes. In an effort to address these issues, we developed a procedure that is more stable and ensures unimodal or S-shaped relationships (Appendix A). We developed an R function that implements our fitting procedure for $d = 2$ axes, which is based on the same iterative principles as ltm but is tailored to better suit ecological data.

The increased stability of our algorithm arises from two additions to the algorithm in ltm. First, we choose initial values for the parameters using the first two axes of correspondence analysis (Appendix A for details). Other geometric ordination procedures can be used to provide initial values, but we chose correspondence analysis because ter Braak (1985) showed that it provides an approximate maximum likelihood solution to a fixed-effects version of this model. Second, we shrank the absolute values of coefficients towards zero at each iteration and set near-zero coefficients to exactly zero, thereby preventing parameter estimates from becoming infinite. We used the LASSO (least absolute shrinkage and selection operator) method of coefficient shrinkage for this purpose (Tibshirani 1996), which has been successfully used to prevent unrealistically large coefficients in a wide variety of applications (e.g., Dahlgren 2010). The LASSO method requires the specification of an additional parameter, λ, called a regularization parameter, which specifies the degree to which coefficients are shrunk towards zero. Larger values of λ cause more shrinkage, which leads to more coefficients being set to exactly zero and therefore to less complex models. We used CVIC to select a value of lambda that balances goodness-of-fit and parsimony.

Our procedure has several advantages over existing methods for interpreting ordinations of presence–absence data. In classical ecological fixed-effects ordination, there are generally two types of models: linear (e.g., PCA) and unimodal (e.g., correspondence analysis; e.g., ter Braak and Prentice 1988). However, natural functional diversity makes it likely that some species will respond linearly (or S-shaped) while others will be unimodal. Therefore this dichotomy between models in which all species are S-shaped or all are unimodal, will often fail to reflect natural among-species variation. With our procedure, some species can be inferred to be S-shaped, some as unimodal, and others as a combination of both types; as the LASSO sets some coefficients to exactly zero, some species will have no quadratic terms and therefore be S-shaped. Similarly, the LASSO may set all axis I coefficients to zero but retain non-zero axis II coefficients. Therefore, some species may be related to only one axis or the other, which is an advantage over classical procedures that will include all species in all axes even if the variation being summarized is largely noise.

Our LASSO-based procedure also has advantages over more recent fixed-effects ordination models (e.g., Yee 2004). As with our procedure, coefficients can be set to zero in these models. However, the decision to set certain parameters to zero must be made by the analyst using prior information. In contrast, our procedure allows the data to select which coefficients will be set to zero. Furthermore, if relevant prior information exists, our procedure can be modified to allow researchers to directly set certain coefficients to zero; this possibility is not explored here as we focus on exploratory analysis.

Estimating the axes and biplots

The estimate, $\hat{\mathbf{x}}_i$, of the axes at the ith observational unit is given by the conditional mean of $\mathbf{x}_i$, given the observed data, $\mathbf{y}_i$. This approach is identical in principle to the estimates used for the linear model (Eq. 12, *Linear random-effects ordination by factor analysis*). However, unlike the linear case, there is no simple expression for these means (it involves integrals)—our R function computes them numerically (Appendices A and B, Supplement 2). We used arrows to represent variables in linear model biplots, as is common in PCA (e.g., ter Braak and Prentice 1988). In classical unimodal models, the variables are represented as points (e.g., correspondence analysis), such that observational units with ordination scores that are near to particular variables can be interpreted as tending to have higher values of these variables. As our procedure estimates some species to be S-shaped and others to be unimodal, we need a different way to plot species information on biplots that will allow us to visualize this additional structure.

Many different possibilities are conceivable for this purpose, but we suggest using contour lines that connect points in the ordination space for which the probability of occurrence is 1/2. We illustrate four broad types of species responses that can be estimated using our framework (Fig. 8A); the points represent fictitious observational units. Species 1 is monotonic along both axes; on one side of the line the probability of occurrence is greater than 1/2 and on the other it is less than 1/2. The arrow points in the direction of increasing probabilities of occurrence. Species two is unimodal along both axes; inside (outside) the ellipse probabilities of occurrence are greater (less) than 1/2. Species 3 is monotonic along axis one but unimodal along axis two; on the inside (outside) of the parabola probabilities of occurrence are greater (less) than 1/2. Finally species 4 is unimodal along axis one but

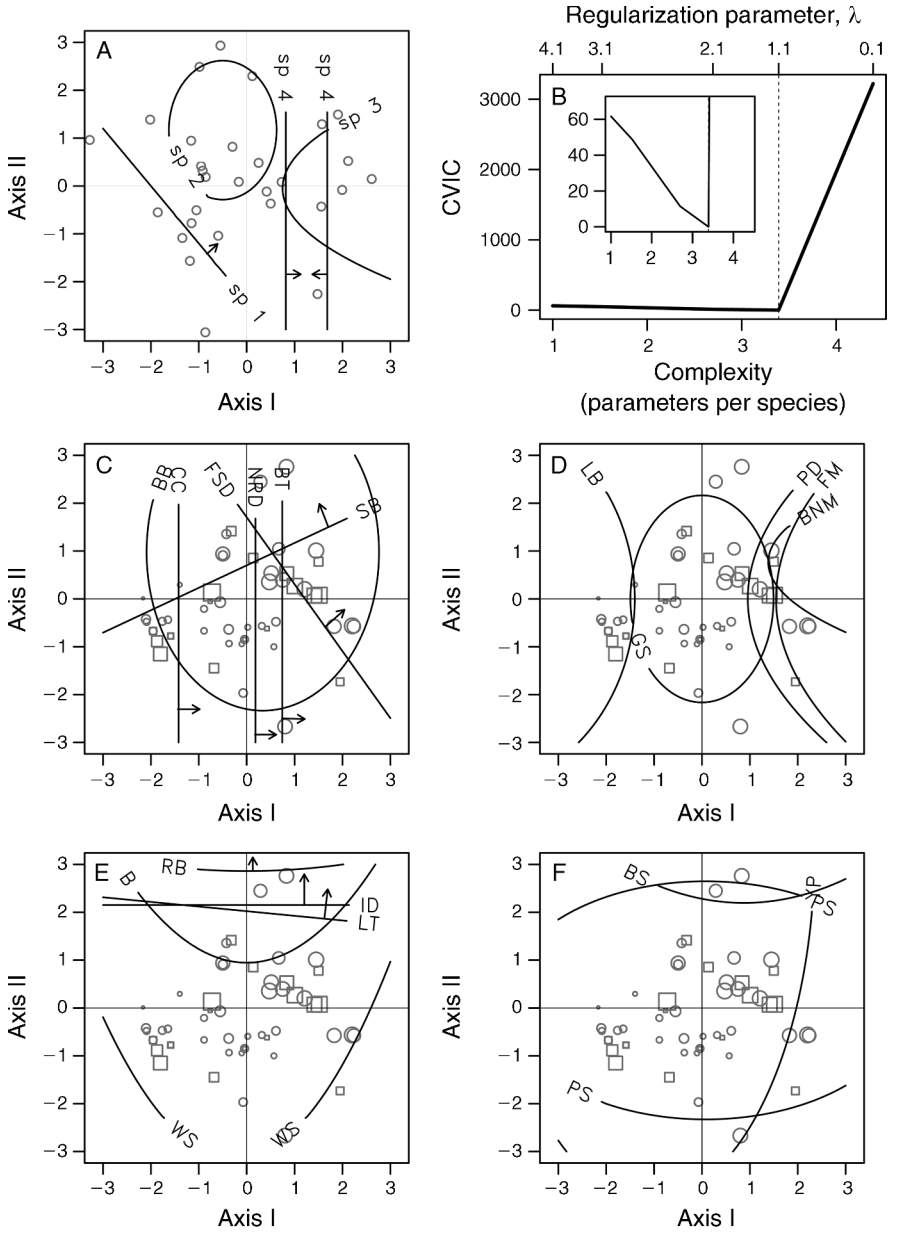

FIG. 8. (A) Cartoon example biplot, (B) CVIC model selection (dotted vertical line at best model) with zoomed-in inset to show details, and (C–F) biplots from the selected latent trait model fitted to the fish community data. Circles and squares in panels C–F are axis scores for training and validation lakes; shape size is proportional to latitude. Thick lines are contours for 50% probabilities of occurrence for each species. Missing species did not have probabilities greater than 50% anywhere on the ordination space. Arrows indicate direction of increasing probability.

completely unrelated to axis two, requiring two lines to represent it; the arrows associated with it indicate that probabilities of occurrence between (outside of) the lines are greater (less) than 1/2.

Fitted probabilities of occurrence and predictions

In the linear case, we explored such predictions through regression equations describing relationships between observed variables that are predicted by the ordination model. With presence–absence data, we cannot construct regression equations per se. Instead, we calculate probabilities of occurrence for one species that are conditional on the presence or absence of other species, and then check these conditional probabilities against observed data (see Appendix A for details). For example, our fitted latent trait models can be used to predict the probability of observing species A at a particular observational unit, given that species B, C, and

D were observed but E, F, and G were not. If the ordination model is working well, we should expect to see higher probabilities associated with observational units at which species A was actually present than at units at which it was absent. We can also compute the probability that species A and B will co-occur, given that species C and D were observed but not E and F. In general, we can compute the probability of observing any particular presence-absence pattern for one set of species (i.e., the response variables), given a particular presence-absence pattern for another mutually exclusive set (i.e., the predictors). We can also compute these conditional probabilities at observational units that were not used to fit the model, because our random-effects approach allows us to make inferences about the statistical population. Such out-of-sample predicted probabilities will be less accurate than within-sample fitted probabilities, but they allow us to explore the reliability of population-level inferences made by the fitted model.

Fish Community Example

Using our R package, reo (Appendix B, Supplement 2), we applied this modeling procedure to the data on the presences and absences of 30 fish species (variables) in the 52 lakes (observational units) of the Black-Hollow watershed (Supplement 1). We used acronyms to refer to the species (Supplement 1). We fit five candidate logistic ordination models to the same 34 lakes that were used as a training set for the limnology data example (*Limnology example*). Only 23 of the 30 species were present at two or more of these 34 lakes, which led us to remove the other seven species. The candidate models were each associated with a value for the regularization parameter, $\lambda = 0.1$, 1.1, 2,1, 3.1, 4.1 (larger values denote simpler models). CVIC selected a fitted model with 3.4 parameters per species ($\lambda =$ 1.1) in the training sample (Fig. 8B), and therefore we used this model to make inferences and predictions. The most complex model has a particularly large CVIC, indicating that regularization vastly improves prediction; in fact, when λ is allowed to go all the way to zero (i.e., maximum complexity), CVIC $= \infty$.

Fig. 8C–F gives four ordination biplots of this model, each with different species represented. Species were divided among panels to reduce visual clutter. The 1/2 probability contours for four species are absent because their probability of occurrence is less then 1/2 over the entire ordination space. The gray circles and squares represent the training and validation lakes; sizes of these shapes are proportional to latitude. The plots suggest that more northern lakes tend to be positioned at positive values of both axes, although this trend is not particularly strong. Such trends could be assessed with significance tests on the estimated axis scores but this is beyond our scope as many such methods already exist (e.g., Clarke 1993). As with the limnological example the validation lakes are well interspersed amongst the training lakes in the ordination space, meaning that predictions of the validation lakes will be best described as interpolative rather than extrapolative.

The variety of contour shapes highlights the utility of our approach for capturing complex patterns relative to classical latent variable ordination (e.g., ter Braak and Prentice 1988), which requires that species be classified a priori as either all S-shaped or all unimodal. The biplots tell us that several species are extremely widespread, having probabilities of occurrence greater than 1/2 over most of the ordination space (PS, WS, YP, BB, CC). The species contours also provide a great deal of information about how species are associated in the training lakes. For example, BT and NRD have parallel contours with arrows pointing in the same direction (Figs. 8C), suggesting that they tend to co-occur. The raw data support this interpretation, as NRD was 1.8 times more likely to be present at lakes with BT than without BT. Non-monotonic relationships are also illustrated; for example, LB, GS, and PD are all more likely to be present when the other two are absent, as their optimum probabilities of occurrence occur at different places along axis I (Fig. 8D). We also see relationships between monotonic and unimodal species (e.g., a negative association between LB and NRD). The biplots visually illustrate many such co-occurrence patterns simultaneously; as always with ordination analysis, implied trends should be checked by examining their consistency with the raw data.

But how appropriate are the biplots as summaries of the data? To answer this, we can look at the model on which the biplots are based in more detail. Fig. 9 shows the fitted probabilities of occurrence of the twenty most widespread species in the training lakes, given the presence-absence pattern of all other species; the probabilities are shown as separate boxplots for lakes in which the focal species was present and lakes in which it was absent. In general, the model fits well, in that it tends to predict higher probabilities of occurrence when the response species is actually present. Therefore, we can be more confident that the trends displayed by the contour biplots give an honest representation of the patterns in the data.

But how well does the model perform on new data? Our random-effects approach allows us to test the model in the 18 validation lakes (Fig. 10). For the majority of species, the good fits to the training data correspond to good out-of-sample predictions. However, there were exceptions. For example, the most (i.e., YP) and least (i.e., BS,BD) common species were well fitted to the training data (Fig. 9) but were poorly predicted (Fig. 10). Such a finding is not surprising because there is little information in the data about the lakes at which common (rare) species are absent (present). However, even some moderately common species were poorly predicted in the validation lakes (e.g., GS). Importantly, classical fixed-effects approaches are unable to detect such poor out-of-sample predictive performance. In contrast, our random-effects approach helps to highlight

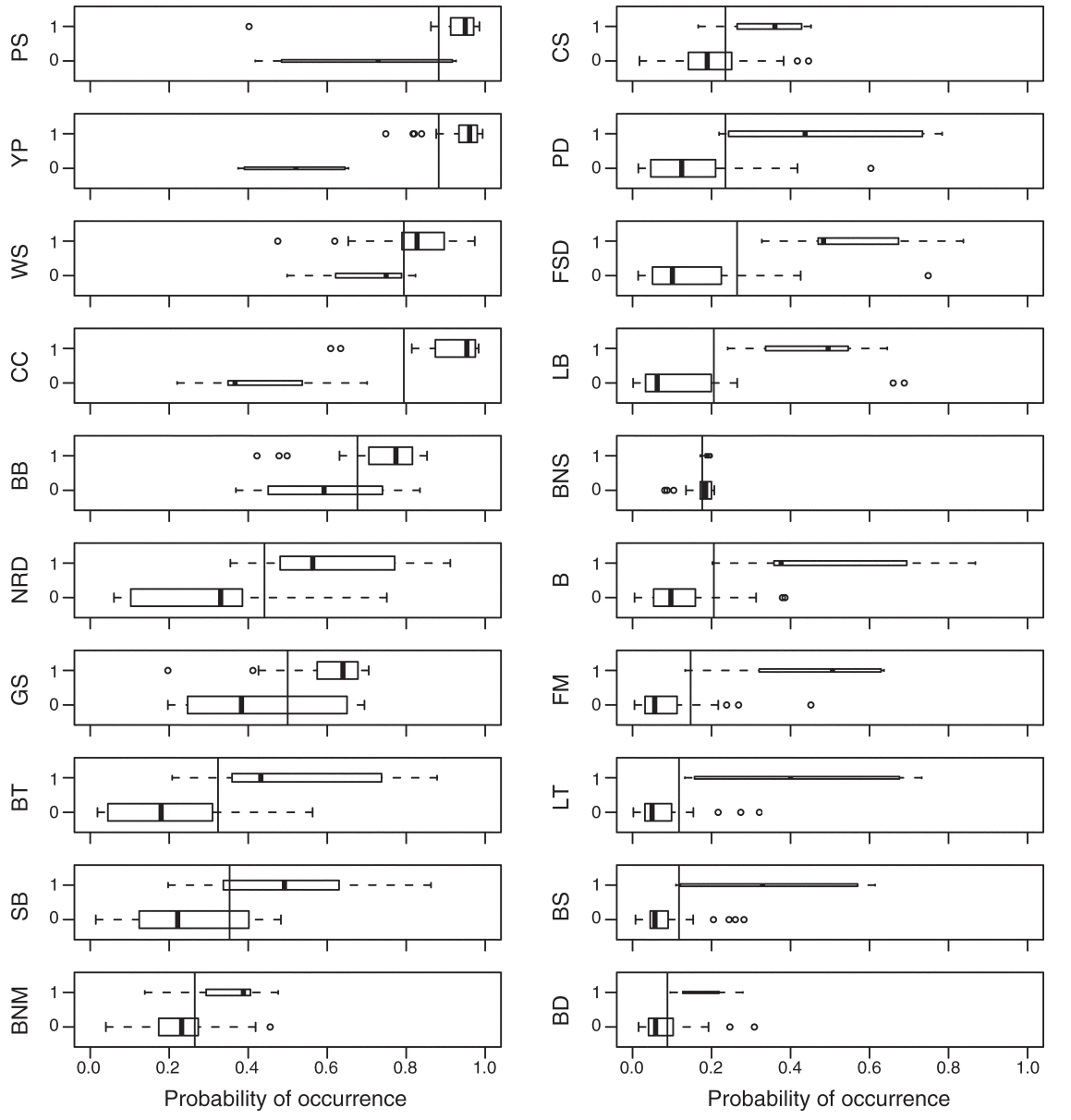

FIG. 9. Fits of each variable in the fish community training data to the selected latent trait model. The plots show the fitted probability of occurrence (x-axis) against observed presence or absence (y-axis) with boxplots giving variation in the fits; points are outliers, whiskers give extremes, and hinges give the first and third quartiles. Vertical lines give the proportion of lakes at which the species was present. Labels on y-axes are species abbreviations (see Supplement 1).

the predictive limitations of multivariate information that is visualized in biplots and therefore to avoid over-interpreting our data.

To demonstrate that our symmetric predictions based on species occurrences can outperform standard asymmetric approaches that use environmental predictor variables, we compared our model predictions with those of logistic regressions of each species on lake pH (Fig. 11). Acidity has been identified as an important factor in this system (Jackson 1988). Each regression was fitted to the training lakes and tested on the validation lakes. Although the logistic regressions outperformed our symmetric model for some species (most notably the rare ones, BS, BD), the opposite was true in the majority of cases. In one species (GS), logistic regression predicted higher probabilities of occurrence when GS was absent then when it was present. One interesting reason for the relatively good prediction of our community model over logistic regression is that species naturally integrate information on a wide variety of environmental variables, whereas pH only integrates information on variables to which it is correlated. Of course this is only one example, but it suggests that predictions based solely on co-occurrence patterns can

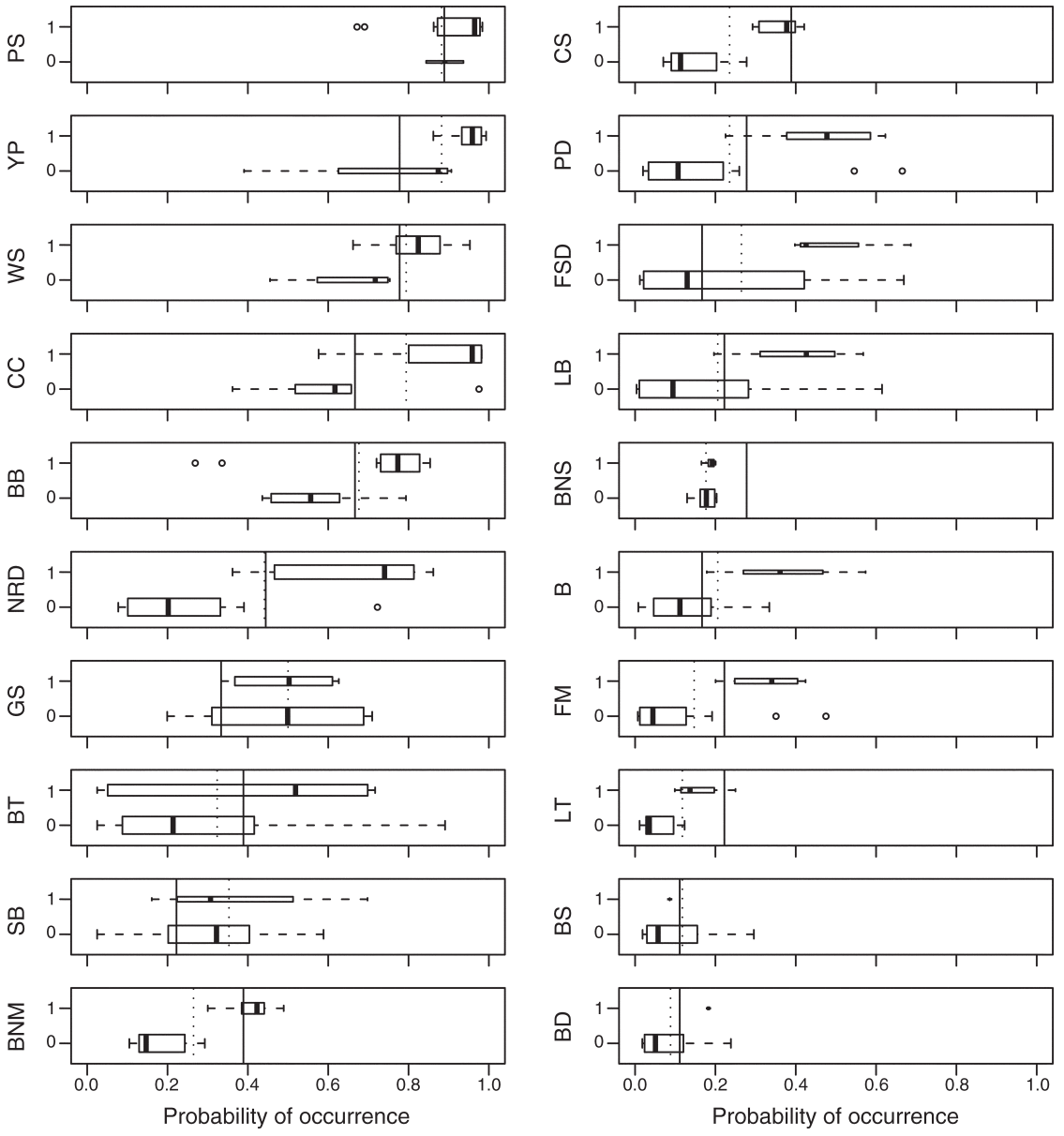

FIG. 10. Predictions of each variable in the fish community validation data using the selected latent trait model. The plots show the predicted probability of occurrence (x-axis) against observed presence or absence (y-axis) with boxplots giving variation in the predictions. Solid (dotted) vertical lines give the proportion of validation (training) lakes at which the species was present.

be useful. We predict that models that include both external environmental predictors and latent random effects may prove to be particularly useful, because the latent variables could "fill in" any patterns that are unrelated to the environmental predictors.

PRACTICAL RECOMMENDATIONS

Although random-effects ordination is new to ecology, the idea of using random latent variables to explore, describe, and predict multivariate relationships has been extensively developed in other fields. Maximum likelihood factor analysis and latent trait models were originally developed to address questions in psychology (e.g., Lawley and Maxwell 1962, Lord 1986). More recently, the field of machine learning has seen an explosion of methodological work on random latent variables (e.g., Lawrence 2005). There are therefore tremendous opportunities for developing more flexible random-effects ordination methodologies then those outlined here.

Still, via transformation and outlier removal, the multivariate-normal and binary methods developed here will provide appropriate ordinations of many ecological data sets. Fig. 12 gives a key for deciding whether a data set is suitable for the methods covered here or if it will be

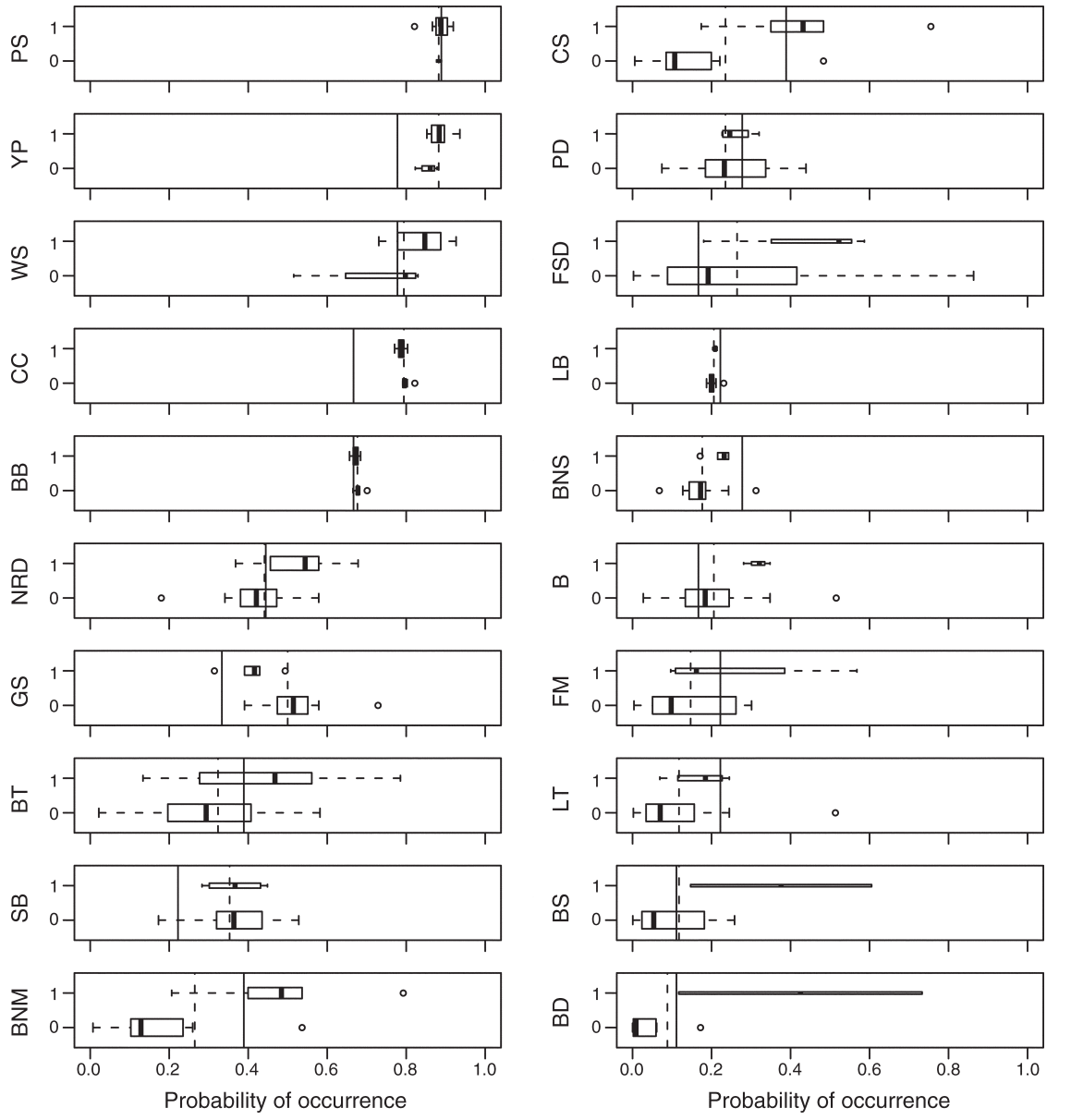

FIG. 11. Predictions of each variable in the fish community validation data using logistic regressions with a pH predictor variable. The plots show the predicted probability of occurrence (x-axis) against observed presence or absence (y-axis) with boxplots giving variation in the predictions. Solid (dotted) vertical lines give the proportion of validation (training) lakes at which the species was present.

necessary to explore the methodological literature outside of ecology: we are less enthusiastic about the third option of treating axes as fixed effects if they are indeed best modeled as random. Furthermore, as more and more latent variable methods are made available through R, these concerns will quickly vanish.

The first decision in the key is whether or not the data are approximately multivariate-normal. There are several ways to check for multivariate-normality. We have generally found informal inspection of pairwise scatterplots to be adequate for such checks; if the pairwise relationships are approximately linear with homogeneous residuals then it is reasonable to tentatively assume multivariate normality. If non-normality is suspected, a transformation may normalize the data. For example, most of the variables in our limnological data were log transformed before analysis. For community data, a Hellinger transformation can often help (Legendre and Gallagher 2001), but we have found it to be less effective for rare species. A decision could be made to remove rare species but this will obviously result in a loss of information.

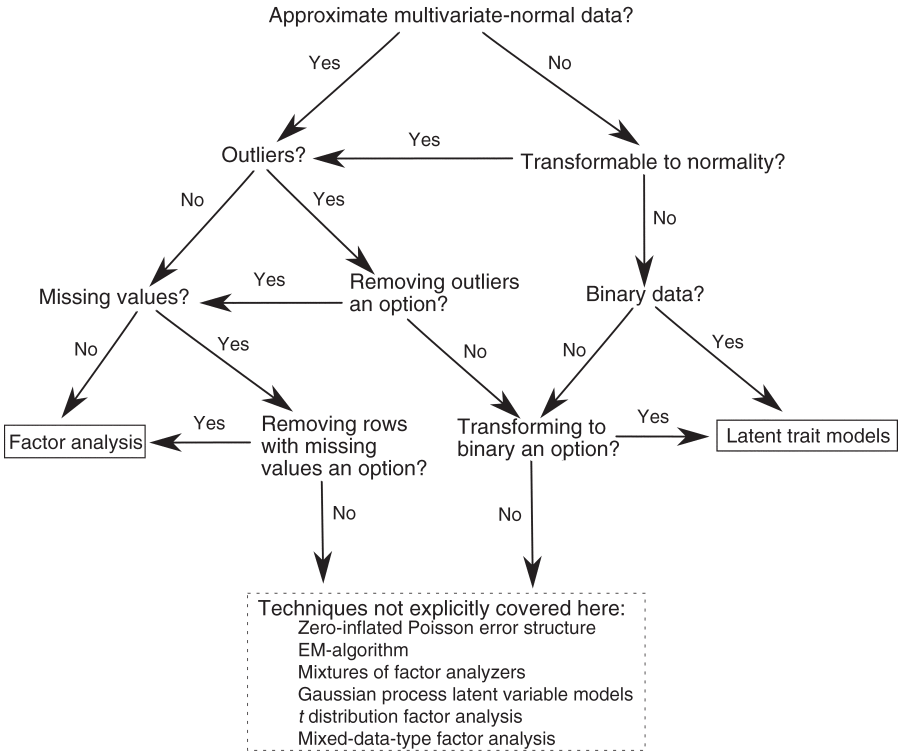

FIG. 12. A decision key for assembling a candidate set of random-effects ordination models. There are three possible endpoints: (1) factor analysis (see *Linear random-effects ordination by factor analysis* and *Limnology example*); (2) latent trait models (see *Logistic random-effects ordination by latent traits* and *Fish community example*; or (3) the dotted box with techniques that are not explicitly covered in this manuscript. See *Practical recommendations* for a more detailed explanation of the key.

If the data are reasonably assumed to be multivariate-normal, the next step is to check for outliers and missing variables. The R mvoutlier package is useful for detecting multivariate outliers (see footnote 2). If observational units with outlying or missing variables are present, one must decide whether to remove them or not. Removing them results in a loss of information, possibly leading to the selection of uninteresting simple models. However, removal has the benefit of leading to data that can be appropriately analyzed by factor analysis: the simplest procedure for random-effects ordination; often the number of observational units is not appreciably impacted if problematic data are simply removed, as in our example. When deciding the fate of outliers in particular, remember that if $x\%$ of the data are identified as outliers, inferences based on the remaining data will apply to approximately $(100 - x)\%$ of the observational units in the statistical population. If this percentage is deemed too low for a particular application, then multivariate normal methods are probably not appropriate. If the data are transformable to approximate normality and outliers and missing values can be removed, then we recommend factor analysis for random-effects ordination.

A common type of data that cannot be transformed to normality are binary presence–absence data. We recommend latent trait models for such data sets. Latent trait models may also be suitable for data sets that are transformable to binary. For example, abundance data can always be transformed to binary by setting all abundances greater than zero to one. Of course there is a loss of information with such transformations, but often this lost information is related to uninteresting processes and is best removed anyway. For example, some fish species are more likely to be detected by certain sampling methods making interspecific comparisons of abundances difficult; transformation to binary data is often more appropriate in these cases (Jackson and Harvey 1997). Such problems often arise in studies of animal communities because animal species often differentially avoid capture. Furthermore, transforming to binary has the benefit of down-weighting the influence of outliers.

Although transformations and outlier removal simplifies analysis, this is not an ideal strategy because it tends to de-emphasize what might be the most interesting features of the data! The dotted box in Fig. 12 lists some techniques that can model such anomalous features rather than remove them. Although these methods are not explicitly covered in our study, the methods we present can be modified to account for more complex situations. For example, our latent trait model

could be modified to handle abundance data by changing the logit-link function to the log-link and changing the Bernoulli distribution to the Poisson. For abundance data with many zeros, a zero-inflated Poisson distribution (e.g., Hall 2000) might be more appropriate. Statisticians have also developed versions of factor analysis that are more robust to outliers. Instead of multivariate-normality these methods assume a multivariate t distribution, which is more likely to predict extreme observations.

One important extension of our methodology would be to allow for observational units with missing values. Information in such units can be utilized by an iterative model-fitting technique called the EM-algorithm (expectation-maximization) that is designed specifically to handle missing values (Dempster et al. 1977). The algorithm that we used to estimate latent trait models is based on EM and therefore could be modified to accept data with missing values. In fact, the EM algorithm is a powerful tool in random latent variable modeling, as the latent variables themselves can be thought of as "missing." For example, the EM-algorithm can also be used to fit models to data that are neither normal nor binary, without having to resort to transformation or outlier removal (McLachlan and Peel 2000: chapter 8). This approach, called mixture modeling, uses normal (or other simple) distributions as building blocks to build up more interesting distributions; an additional latent variable is used to assign each observational unit to one of the building block distributions. The FLXMCfactanal function in the R flexmix package can be used to fit such models (called mixtures of factor analyzers). Machine learning researchers have developed software for such latent variable EM algorithms in languages such as C and MAT LAB. Several books on mixture modeling are useful for help with constructing EM algorithms (e.g., McLachlan and Peel 2000). As free statistical software continues to be developed for environments such as R, it will not be long before these more sophisticated methods become common and user friendly.

The steps of random-effects ordination analysis

Assemble a set of candidate models.—It is important to ensure wide variation in complexity among the candidates, so that information criteria can be used to select an appropriate level of complexity. For factor analysis, the set of candidates consists of the null and full models as well as models with $d = 1, \ldots, \lfloor p + (1/2) - \sqrt{2p + (1/4)} \rfloor$ axes ($\lfloor \cdot \rfloor$ is the floor function, p the number of variables); complexity is then measured by the number of axes. If our latent trait models are used, each candidate corresponds to a different value of the regularization parameter, λ. We compared models with $\lambda = 0.1, 1.1, 2.1, 3.1, 4.1$. It may be necessary to consider $\lambda > 4.1$ for some data sets to ensure that a null model is included with all non-intercept coefficients set to zero. We do not recommend using $\lambda < 0.1$, because in our experience this results in models with large coefficients that are over-fitted and poorly predict new data. If neither factor analysis nor latent trait modeling are used, the principle is the same: assemble a set of related models that differ in complexity.

Identify/develop software.—If factor analysis or latent trait modeling is used, our R package, reo, is available for model fitting (Appendix B, Supplement 2). If a different set of models is deemed necessary, alternative software is required. The R packages sem, ltm, flexmix, and MCMCpack contain functions for fitting many such models. We emphasize that coding your own algorithms leads to a much better understanding of the methods and how to interpret them; for this purpose, we recommend a good book on statistical modeling in ecology (e.g., Hilborn and Mangel 1997, Burnham and Anderson 2002, Clark 2007, Bolker 2008) and a good book on the EM algorithm (e.g., McLachlan and Peel 2000).

Model selection.—We recommend selecting the model with the lowest value of an information criterion, while a simpler model can be used if it is adequate for the question at hand (e.g., Cudeck and Browne 1983). However, it is never appropriate to use a model that is more complex than the selected model. With factor analysis we recommend MAIC_c, as it performed well in our simulations. When maximum likelihood is not used, such as with our latent trait models, CVIC will usually be necessary unless it can be shown that AIC-based criteria give approximately unbiased estimates of Kullback-Leibler information. Although CVIC is more widely applicable, it is computationally slower, because each model must be fitted n times. Our R function took approximately three minutes to compute the CVIC values for our fish example. For larger data sets, this time will obviously increase, perhaps beyond acceptable amounts of time. Recent statistical work on developing modifications of AIC for LASSO-based regression models (Zou et al. 2007) and regularized latent trait models (Houseman et al. 2007) are promising and approximately n times more computationally efficient than CVIC, but are currently untested on multivariate ecological data. Until these approaches become better developed, we recommend using $\lambda = 1$ in situations where cross-validation is not computationally possible, as this λ value cross-validates well in our experience.

Compare fitted values and predictions with data.—Although information criteria select models that make good predictions relative to the other candidates, it is important to check the absolute quality of selected models. A first step is to consider the percentage of variation explained by the ordination, as in classical analyses (e.g., PCA). But such simple measures do not address the details of model successes and failures. We recommend making predictions that can be graphically checked (e.g., Figs. 3, 4, 9, and 10), just as we would check the predictions of regression models against data. We also recommend making out-of-sample predictions

on data that were not used to fit the model, if such data are available; out-of-sample predictions test the strength of the inferences that the model makes about the statistical population. It will not always be feasible to include all possible predictions in a research paper. However, we strongly recommend including as much material as possible, in electronic appendices perhaps, so that people can make up their own minds about the quality of the models.

Make biplots.—The production of a biplot (e.g., Figs. 2 and 8) is the classical endpoint of ordination analysis. Within our random-effects approach such summaries should be based on the coefficients that relate the variables to the axes, because these coefficients determine the population-level inferences and predictions made by the fitted model. Therefore, the utility of the summaries can be assessed by how well the model predictions succeed (previous step). In factor analysis we use arrows to represent the variables, in the familiar way that they are used in PCA. In latent trait models, we used contour lines connecting points in the ordination space of equal probabilities of occurrence. Our R package includes functions for computing these biplots, but we look forward to new inventive visual representations of random-effects ordination models.

Repeat.—We view random-effects ordination as an iterative process requiring human ingenuity: information criteria are only guides. At every stage of analysis, the models should be scrutinized. If model inadequacies are identified, it might be a good idea to start again from the first step with a new set of candidate models. For example, we first fitted latent trait models to the fish community data using the R ltm package. But we found that the fitted models were predicting probabilities of occurrence that were too extreme to be ecologically reasonable (*Estimation and model selection*; Appendix B). We therefore modified our fitting procedure with the LASSO to obtain better models. In general, we have an increased chance of finding problems with random-effects ordinations relative to classical procedures, because of the additional population-level inferences that random-effects models make. This might seem like a disadvantage of the random-effects approach, but we see it as beneficial for the goal of developing believable quantitative models. Still, it is important to eventually stop looking for model inadequacies, which will always be present.

Conclusion

We have argued that ordination analysis can and should be done from a random-effects perspective, and provided practical guidelines and software for doing so. This perspective has the benefit of allowing us to tackle multivariate problems using many of the tools from univariate statistics, including the checking of model assumptions, model validation on independent data, and model selection using information criteria. We conclude with some implications of our approach for ecology.

Our random-effects approach addresses the interplay between exploratory and confirmatory multivariate ecological analysis. The spirit of ordination is to analyze multivariate relationships without making assumptions about causality. The models that we use here are consistent with this spirit; all variables are treated symmetrically, in the sense that there are no a priori variable-specific assumptions. This symmetry gives random-effects ordination models the exploratory emphasis they should have. At the same time, our approach is closely related to structural equation modeling (e.g., Grace 2006), which uses random latent variables to assess the evidence in multivariate data for a priori causal hypotheses. Random-effects ordination models may be modified to include information from causal hypotheses. One exciting prospect here is the potential for using information criteria to select between exploratory random-effects ordinations and confirmatory structural equation models; if this approach is done appropriately then selecting an ordination might suggest that our scientific ideas are less predictively successful than data exploration, indicating that our ideas require refinement. Work outside of ecology has been done on the relationship between exploratory and confirmatory structural equation modeling (e.g., Bollen 1989), providing a rich literature to draw on.

While we see great potential in random-effects ordination, it may not always be appropriate. For example, if our observational units consist of nature preserves, then the axes may truly be fixed effects if we sampled all nature preserves of interest. In general, fixed-effects ordination will be more appropriate whenever it is possible to exhaustively sample the target statistical population of observational units.

Although we studied ordination models in which all effects are random, we see the development of mixed-effects ordination as an important next step. These models would include both fixed-effects predictors as well as the random-effects we consider, and are beginning to be explored outside of ecology (Houseman et al. 2007). For example, it is possible to produce mixed-effects versions of classical asymmetric ordination models (e.g., redundancy analysis; canonical correspondence analysis), by adding random axis terms. We demonstrated the possibility of using random axes to model between-species associations that are left unexplained by environmental predictors (e.g., Section 8). Such associations can arise from species interactions or missing predictors, and are considered an important challenge in predictive ecological modeling (Elith and Leathwick 2009).

Advocates of multidimensional scaling may eschew our emphasis on explicit model assumptions, arguing instead for a robust approach to ordination that is able to handle a wide variety of underlying patterns (e.g., Minchin 1987). But no method is assumption free. Explicit probabilistic modeling lays bare the assumptions behind the methods. The probabilistic approach to

ordination encourages ecologists to construct explicit working models of their study systems. The strategy of checking and comparing the assumptions and predictions of alternative models is central to science: we have formalized ordination analysis from this perspective.

Acknowledgments

We thank Laura Timms, Marie-Josee Fortin, Ben Bolker, Keith Somers, Cajo ter Braak, Nick Collins, Angela Strecker, Stéphane Dray, Tahira Jamil, and two anonymous reviewers for their substantial comments on various earlier drafts and Simon Prince, Pierre Legendre, and Peter Minchin for discussions about our ideas in this monograph. We gratefully acknowledge funding from the Natural Sciences and Engineering Research Council of Canada, an Ontario Graduate Scholarship, and support to S. C. Walker from P. Legendre.

Literature Cited

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267–281 *in* B. Petrov and F. Csaki, editors. Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101–118.

Bedrick, E. J., and C.-L. Tsai. 1994. Model selection for multivariate regression in small samples. Biometrics 50:226–231.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey, USA.

Bollen, K. A. 1989. Structural equations with latent variables. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, New York, USA.

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. Ecology 73:1045–1055.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Second edition. Springer, New York, New York, USA.

Clark, J. S. 2007. Models for ecological data. Princeton University Press, Princeton, New Jersey, USA.

Clarke, K. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117–143.

Condit, R., et al. 2002. Beta-diversity in tropical forest trees. Science 295:666–669.

Cudeck, R., and M. W. Browne. 1983. Cross-validation of covariance structures. Multivariate Behavioral Research 18:147–167.

Dahlgren, J. P. 2010. Alternative regression methods are not considered in Murtaugh (2009) or by ecologists in general. Ecology Letters 13:E7–E9.

Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B (Methodological) 39:1–38.

Dray, S. 2008. On the number of principal components: a test of dimensionality based on measurements of similarity between matrices. Computational Statistics and Data Analysis 52:2228–2237.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology and Systematics 40:677–697.

Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? Systems Biology 51:331–363.

Gauch, H. G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge, UK.

Gauch, H. G., G. B. Chase, and R. H. Whittaker. 1974. Ordination of vegetation samples by Gaussian species distributions. Ecology 55:1382–1390.

Grace, J. B. 2006. Structural equation modeling and natural systems. Cambridge University Press, Cambridge, UK.

Hall, D. B. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56:1030–1039.

Hilborn, R., and M. Mangel. 1997. The ecological detective. Princeton University Press, Princeton, New Jersey, USA.

Houle, D. 2007. A dispatch from the multivariate frontier. Journal of Evolutionary Biology 20:22–23.

Houseman, E. A., C. Marsit, M. Karagas, and L. M. Ryan. 2007. Penalized item response theory models: application to epigenetic alterations in bladder cancer. Biometrics 63:1269–1277.

Jackson, D. A. 1988. Fish communities in lakes of the Black and Hollow River watersheds, Ontario. Thesis. University of Toronto, Toronto, Ontario, Canada.

Jackson, D. A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology 74:2204–2214.

Jackson, D. A., and H. H. Harvey. 1997. Qualitative and quantitative sampling of lake fish communities. Canadian Journal of Fisheries and Aquatic Science 54:2807–2813.

Johnson, R. A., and D. W. Wichern. 1992. Applied multivariate statistical analysis. Third edition. Prentice Hall, Englewood Cliffs, New Jersey, USA.

Jones, M. B., M. P. Schildhauer, O. Reichman, and S. Bowers. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annual Review of Ecology and Systematics 37:519–544.

Lawley, D., and A. Maxwell. 1962. Factor analysis as a statistical method. Journal of the Royal Statistical Society Series D (The Statistician) 12:209–229.

Lawley, D., and A. Maxwell. 1973. Regression and factor analysis. Biometrika 60:331–338.

Lawrence, N. 2005. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. Journal of Machine Learning Research 6:1783–1816.

Leathwick, J. R., and M. Austin. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. Ecology 82:2560–2573.

Legendre, P., and M. J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecological Monographs 69:1–24.

Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Number 20 in developments in environmental modelling. Second edition. Elsevier, Amsterdam, The Netherlands.

Lord, F. M. 1986. Maximum likelihood and Bayesian parameter estimation in item response theory. Journal of Educational Measurement 23:157–162.

McLachlan, G. J., and D. Peel. 2000. Finite mixture models. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, New York, USA.

Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications 7:330–342.

Minchin, P. R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69:89–107.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2:559–572.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers. 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics and Data Analysis 49:974–997.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in S and S-plus. Statistics and computing. Springer, New York, New York, USA.

R Development Core Team. 2011. R version 2.13.0. R Project for Statistical Computing, Vienna, Austria. ⟨www.r-project.org⟩

Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. Sankhyā: The Indian Journal of Statistics, Series A 26:329–358.

Rizopoulos, D. 2006. ltm: An R package for latent variable modeling and item response theory analyses. Journal of Statistical Software 17:1–25.

Royall, R. 1997. Statistical evidence: a likelihood paradigm. Chapman and Hall, London, UK.

Taper, M. L. 2004. Model identification from many candidates. Pages 488–524 *in* M. L. Taper. The nature of scientific evidence: statistical, philosophical, and empirical considerations. University of Chicago Press, Chicago, Illinois. USA.

ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. Biometrics 41:859–873.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179.

ter Braak, C. 1987. Ordination. Pages 91–169 *in* Data analysis in community and landscape ecology. Cambridge University Press, Cambridge, UK.

ter Braak, C. J. F., H. Hoijtink, W. Akkermans, and P. F. Verdonschot. 2003. Bayesian model-based cluster analysis for predicting macrofaunal communities. Ecological Modelling 160:235–248.

ter Braak, C. J. F., and I. C. Prentice. 1988. A theory of gradient analysis. Advances in Ecological Research 18:271–317.

Tibshirani, R. 1996. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society Series B (Methodological) 51:267–288.

Tipping, M. E., and C. M. Bishop. 1999. Probabilistic principal component analysis. Journal of the Royal Statistical Society Series B (Methodological) 61:611–622.

Whittaker, R. 1967. Gradient analysis of vegetation. Biological Review 42:207–264.

Yee, T. W. 2004. A new technique for maximum-likelihood canonical Gaussian ordination. Ecological Monographs 74:685–701.

Zou, H., T. Hastie, and R. Tibshirani. 2007. On the "degrees of freedom" of the LASSO. Annals of Statistics 35:2173–2192.

APPENDIX A

Mathematical details of random-effects ordination (*Ecological Archives* M081-023-A1).

APPENDIX B

Tutorial on using the reo package in R (*Ecological Archives* M081-023-A2).

SUPPLEMENT 1

Data from the Black-Hollow watershed (*Ecological Archives* M081-023-S1).

SUPPLEMENT 2

R source code and manual for the reo package (*Ecological Archives* M081-023-S2).