# CANONICAL ANALYSIS OF PRINCIPAL COORDINATES: A USEFUL METHOD OF CONSTRAINED ORDINATION FOR ECOLOGY

Marti J. Anderson[1,3] and Trevor J. Willis[1,2]

[1]*Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand*
[2]*Leigh Marine Laboratory, University of Auckland, P.O. Box 349, Warkworth, New Zealand*

*Abstract.* A flexible method is needed for constrained ordination on the basis of any distance or dissimilarity measure, which will display a cloud of multivariate points by reference to a specific a priori hypothesis. We suggest the use of principal coordinate analysis (PCO, metric MDS), followed by either a canonical discriminant analysis (CDA, when the hypothesis concerns groups) or a canonical correlation analysis (CCorA, when the hypothesis concerns relationships with environmental or other variables), to provide a flexible and meaningful constrained ordination of ecological species abundance data. Called ''CAP'' for ''Canonical Analysis of Principal coordinates,'' this method will allow a constrained ordination to be done on the basis of any distance or dissimilarity measure. We describe CAP in detail, including how it can uncover patterns that are masked in an unconstrained MDS ordination. Canonical tests using permutations are also given, and we show how the method can be used (1) to place a new observation into the canonical space using only interpoint dissimilarities, (2) to classify observations and obtain misclassification or residual errors, and (3) to correlate the original variables with patterns on canonical plots. Misclassification error or residual error is used to obtain a non-arbitrary decision concerning the appropriate dimensionality of the response data cloud (number of PCO axes) for the ensuing canonical analysis. We suggest that a CAP ordination and an unconstrained ordination, such as MDS, together will provide important information for meaningful multivariate analyses of ecological data by reference to explicit a priori hypotheses.

*Key words: canonical ordination; classification; community structure; distance matrix; MDS; multivariate analysis; principal coordinate analysis; species abundances; statistical methods.*

## Introduction

Many ecological studies involve the investigation of the simultaneous response of many species by reference to some specified hypothesis. For example, it may be hypothesized that a whole set of species (an assemblage or community) will change in response to some experimental treatments, through time or space, or in response to some changes in environmental or other predictor variables. In these situations, multivariate ordination methods are required to reduce dimensionality and to visualize patterns in multivariate data.

Ordination procedures can be classified as either constrained or unconstrained. Unconstrained methods include principal component analysis (PCA), correspondence analysis (CA), metric multidimensional scaling (also called principal coordinate analysis or PCO; Torgerson 1958, Gower 1966*a*), and nonmetric multidimensional scaling (MDS; Shepard 1962, Kruskal 1964, Kruskal and Wish 1978). An unconstrained ordination procedure does not use a priori hypotheses in any way, but reduces dimensions on the basis of some general criterion, such as minimizing residual variance (as in PCA; Rao 1964) or minimizing a stress function (as in

nonmetric MDS; Meulman 1986, 1992). Table 1 summarizes the most commonly used unconstrained procedures.

Unconstrained ordinations are generally extremely useful for visualizing broad patterns across the entire data cloud. Also, for cases in which data are classified into a priori groups, one can visualize potential patterns of differences in the location or relative dispersion among groups. Metric and nonmetric MDS can be based on any distance or dissimilarity measure of choice, including ecologically meaningful measures such as the Bray-Curtis measure (Bray and Curtis 1957), the Kulczynski measure (Faith et al. 1987), or the CY measure (Cao et al. 1997), making them very flexible. Nonmetric MDS has been demonstrated as a particularly robust and useful unconstrained ordination procedure for ecology (Field et al. 1982, Kenkel and Orlóci 1986, Minchin 1987). Importantly, however, certain patterns of overall dispersion can sometimes mask real patterns of differences in multivariate location among groups in an unconstrained ordination.

Constrained ordinations, on the other hand, use an a priori hypothesis in some manner to produce the plot. That is, they can be used to relate a matrix of response variables, **Y** (species abundance variables) with some predictor variables, **X** (such as quantitative environmental variables or dummy variables that identify AN-

[3] E-mail: mja@stat.auckland.ac.nz

TABLE 1. Methods of unconstrained ordination of a multivariate set of data, **Y**.

| Name of method (acronyms, synonyms) | Distance measure preserved | Relationship of ordination axes with original variables | Criterion for drawing ordination axes |
|---|---|---|---|
| Principal Component Analysis (PCA) | Euclidean distance | linear | finds axis that maximizes the total variance (or, equivalently, that minimizes the total residual variation) |
| Correspondence Analysis (CA, reciprocal averaging, dual scaling) | chi-square distance | unimodal (approximately Gaussian) | finds axis that maximizes dispersion of species scores (which are themselves weighted averages of site scores) |
| Principal Coordinate Analysis (PCO, PCoA, metric multidimensional scaling, classical scaling, Torgerson scaling) | any chosen distance or dissimilarity measure | unknown; depends on distance measure chosen | Euclidean distances in new full-dimensional space are equal to original distances (or dissimilarities). |
| Nonmetric Multidimensional Scaling (MDS, NMDS) | any chosen distance or dissimilarity measure | unknown, depends on distance measure chosen | The number of dimensions for the new space is chosen a priori (reduced). Euclidean distances in new space are monotonically related to original distances. |

OVA factors or groups). The most commonly used constrained ordination methods are summarized in Table 2. Constrained ordinations include canonical discriminant analysis (CDA), in which ordination axes are drawn so as to maximize differences among groups (e.g., James and Wilkinson 1971, Mardia et al. 1979), and canonical correlation analysis (CCorA), in which axes are drawn so as to maximize their correlation with linear combinations of some quantitative predictor variables (e.g., Gittins 1985). Ter Braak (1995) refers to constrained and unconstrained ordination methods as ''direct'' and ''indirect'' gradient analysis, respectively.

We note that CDA has also been described by Legendre and Legendre (1998:617) as a multiple regression of the group structure (as coded in matrix **X,** say) on the species data (**Y**), reversing the roles of ''predictor'' and ''response.'' However, CDA is equivalent to CCorA when the **X** matrix codes for group structure, rather than containing another set of quantitative variables (Table 2). Thus, for CDA, as for CCorA, **X** and **Y** are treated symmetrically, so in fact one may view either matrix as the ''response.'' Here, we shall continue to refer to **Y** (species data) as the response matrix and to **X** (design matrix or environmental data) as the predictor matrix, respectively.

Existing constrained ordination techniques are limited to being based implicitly on some particular (metric) distance measure. For example, RDA preserves Euclidean distances, CDA preserves Mahalanobis distances (Gower 1966b, Meulman 1992), and canonical correspondence analysis (CCA) preserves chi-square distances (ter Braak 1986a, b). Although chi-square distance is generally more meaningful than Euclidean distance for many ecological applications (e.g., Palmer 1993), experimenters may well wish to have the flexibility to choose some other distance measure. There-

fore, what is needed is a constrained ordination procedure that can be used to investigate hypotheses visually in reduced dimension, but with the flexibility to allow any distance or dissimilarity measure to be used as the basis for analysis.

Some multidimensional scaling techniques that involve external constraints (not listed in Table 2) have been developed (Borg and Lingoes 1980, Digby and Gower 1981, Heiser and Meulman 1983, Meulman 1986, 1988, ter Braak 1992, Borg and Groenen 1997, Legendre and Anderson 1999, McArdle and Anderson 2001). However, these methods do not take into account the correlation structure among the variables in the response (species) data cloud. This contrasts with the traditional methods of CDA and CCorA that do. We note that Legendre and Legendre (1998) refer to all of the methods listed in Table 2 as ''canonical.'' In this paper, we shall refer only to the methods of CDA and CCorA as ''canonical,'' as these are the only methods that explicitly account for correlation structure among response variables, whereas we refer to the general class of methods as ''constrained'' ordination techniques.

Legendre and Legendre (1998:634–635) hinted that CDA could be done on the axes obtained from a principal coordinate analysis, but they did not provide any examples of this approach, nor did they discuss any possible advantages compared to existing methods. Here, we propose that this general approach, which we shall refer to as the canonical analysis of principal coordinates (CAP), is a flexible and particularly useful constrained ordination procedure for ecology. It has the advantage of allowing any distance or dissimilarity measure to be used, but also takes into account correlation structure among variables in the response data cloud. Thus, like the traditional canonical methods, it

can uncover important patterns in the multivariate data by reference to relevant hypotheses.

CAP may be used where the a priori hypothesis concerns differences among groups, as in ANOVA; i.e., the **X** matrix contains orthogonal variables coding for the classification criteria for the factor concerned (e.g., Neter et al. 1996, Legendre and Anderson 1999). In this case, CAP is a canonical discriminant analysis of principal coordinates, called ''generalized discriminant analysis based on distances'' by Anderson and Robinson (*in press*). The analysis may also be done for hypotheses involving one or more quantitative explanatory variables, in which case CAP is a canonical correlation analysis of principal coordinates.

First, we demonstrate how ecologically important patterns can be masked in unconstrained ordination procedures, such as PCA or MDS. We then provide a description of CAP for ordination, including extensions such as the placement of a new observation into the canonical ordination and the calculation of misclassification error (and residual error) using the ''leave-one-out'' procedure (e.g., Lachenbruch and Mickey 1968). Tests of hypotheses are achieved using permutation procedures on canonical test statistics. Further use of the method includes calculating a posteriori correlations to assess relationships of original species variables with canonical axes to characterize multivariate effects. We also give an ecological example using CAP and compare it with alternative approaches. The asymptotic properties of test statistics obtained using CAP, including simulations comparing them to other permutation tests designed to detect group differences (e.g., McArdle and Anderson 2001), are described elsewhere (Anderson and Robinson, *in press*).

## Constrained vs. Unconstrained Criteria for Ordination

We start with a simple example, in Euclidean space, to demonstrate how the constrained and unconstrained methods differ in terms of the criterion used for drawing ordination axes. Consider a simple data set with two response variables measured for each of 100 observation points in each of two groups, as shown in Fig. 1. Imagine that we are unable to see two dimensions and can, unfortunately, only see one. Thus, we need to obtain a one-dimensional ordination of the two-dimensional cloud of points in order to visualize patterns.

The first principal component axis (PC1) for an unconstrained ordination (using PCA) is a linear combination of the original variables that minimizes the sum of squared deviations of points to the new axis (i.e., PC1 is drawn in the direction of maximum variation through the cloud; Fig. 1a). There is no indication, however, of any difference between the two groups at all along PC1. In contrast, the first canonical variable (CV1) for a constrained ordination (using CDA) is an axis drawn through the cloud of points that

best separates the groups (Fig. 1b). This is equivalent to finding a linear combination of the original variables that maximizes the between- to within-group variation. CV1 clearly shows separation of the groups, a real pattern that was evident in the higher two-dimensional plot (Fig. 1b).

Thus, different criteria used to obtain a lower dimensional ordination can radically affect our view of multivariate patterns. Note that the cloud of points is exactly the same in either case, but our ''view'' of the data has been changed by our choice of ordination procedure. In this case, the direction of greatest total variation was completely different from the direction of group differences in multivariate space. An unconstrained ordination may be useful to visualize overall patterns of dispersion, but this simple example also demonstrates how real differences in location, which were masked in the PCA, were uncovered by the canonical approach.

## Description of Canonical Analysis of Principal Coordinates (CAP)

The previous example was done in Euclidean space with traditional methods to introduce and clarify concepts. However, we wish to obtain a constrained ordination based on any distance or dissimilarity measure. Full details of the analysis, with a mathematical description, are given in the Appendix (see also Anderson and Robinson, *in press*).

A short summary of the CAP approach for a generalized constrained ordination on the basis of any dissimilarity measure is as follows:

1) Do a principal coordinate analysis (PCO, metric MDS, or classical scaling) on the data matrix **Y** ($N$ observations $\times$ $p$ variables), using a dissimilarity measure of choice, yielding orthonormal axes **Q**. No scaling of these axes is necessary.

2) Choose an appropriate number of axes $m$ as a subset of **Q** (matrix $\mathbf{Q}_m$) to use for the ensuing canonical analysis. The choice can be made on a non-arbitrary basis as the number of axes resulting in minimum misclassification error or minimum residual sum of squares, as we will describe.

3) Do a traditional canonical analysis (CDA if **X** contains dummy variables corresponding to group specifications, or CCorA if **X** contains quantitative variables) on the first $m$ axes of **Q**.

These three steps are sufficient to produce an ordination. As a further option, one may test the hypothesis of either (1) no significant differences in multivariate location among groups (CDA), or (2) no significant relationship with quantitative environmental or other variables (CCorA). This is done by using the trace statistic (sum of canonical eigenvalues = sum of squared canonical correlations; see the Appendix) and obtaining a $P$ value by permutation (e.g., Anderson 2001*b*).

TABLE 2. Methods of constrained ordination relating response variables, **Y** (species abundance variables) with predictor variables, **X** (such as quantitative environmental variables or qualitative variables that identify factors or groups as in ANOVA).

| Name of method (acronyms, synonyms) | Distance measure preserved | Relationship of ordination axes with original variables | Takes into account correlation structure† |
|---|---|---|---|
| Redundancy Analysis (RDA) | Euclidean distance | linear with **X**, linear with fitted values, $\hat{Y} = X(X'X)^{-1} X'Y$ | ⋯ among variables in **X**, but *not* among variables in **Y** |
| Canonical Correspondence Analysis (CCA)§ | chi-square distance | linear with **X**, approximately unimodal with **Y**, linear with fitted values, $\hat{Y}$* | ⋯ among variables in **X**, but *not* among variables in **Y** |
| Canonical Correlation Analysis (CCorA, COR) | Mahalanobis distance | linear with **X**, linear with **Y** | ⋯ among variables in **X**, and among variables in **Y** |
| Canonical Discriminant Analysis (CDA; Canonical Variate Analysis, CVA; Discriminant Function Analysis, DFA) | Mahalanobis distance | linear with **X**, linear with **Y** | ⋯ among variables in **X**, and among variables in **Y** |
| Canonical Analysis of Principal Coordinates (CAP; Generalized Discriminant Analysis) | any chosen distance or dissimilarity | linear with **X**, linear with $Q_m$; unknown with **Y** (depends on distance measure) | ⋯ among variables in **X**, and among principal coordinates $Q_m$ |

† Note that CCorA (and CDA) use standardized $Y_s$ rather than unstandardized **Y** or $Y_c$, thus taking into account correlation structure among variables in the response data cloud. CAP also does this as the vectors of $Q_m$ are orthonormal variables.

‡ $C^{-1/2}$ is a $p \times p$ diagonal matrix containing the square root of column totals. **R** is a diagonal matrix of row totals. **Y** standardized by its variance–covariance matrix is $Y_s = Y_c(Y_c'Y_c)^{-1/2}$, and the **Y** variables centered on their means comprise $Y_c = Y - j\bar{y}'$, with **j** being an $n \times 1$ column of 1's and $\bar{y}$ being a $p \times 1$ column of means, with a similar definition for $X_c$. Also, $(Y_c'Y_c)$ is the sum of squares and cross-product matrix for **Y** and $(Y_c'Y_c)^{-1/2}$ is the Cholesky decomposition of the inverse of this matrix.

§ There is a trivial solution to the eigenvalue decomposition of the matrix given for CCA above ($\lambda = 1$), which can be avoided by first subtracting from each variable in **X** its weighted mean, i.e., $\bar{x}_j = \Sigma_{i=1}^{n} R_i x_{ij}/T$, with $R_i$ the row total for the $i$th observation in **Y**, and $T$ the total sum of values in **Y** (for further details see ter Braak [1995:163]).

‖ For further details, see ter Braak (1995:160–165), Legendre and Legendre (1998:594–597), and Legendre and Gallagher (2001).

¶ See Legendre and Legendre (1998:626) and Gittins (1985).

### Classifying a new observation

In traditional canonical discriminant analysis, one may use Fisher's discriminant rule (Fisher 1936, Mardia et al. 1979, Seber 1984) to classify a new observation into one of the groups. This works because, for the traditional analysis in which the result is obtained in Euclidean space, the canonical axes are linear combinations of the original variables. However, when the analysis is based on a dissimilarity matrix, as in CAP, the canonical axes do not have this nice linear relationship with the original variables. Nevertheless, the canonical axes are linear combinations of the orthonormal principal coordinate axes. Thus, we can place a new observation into the canonical space, provided that we can place a new observation into the space of principal coordinates, which can be done using only the interpoint distances of the new observation with all previous observations. A full description of this is given in the Appendix, following Gower (1968), Cuadras et al. (1997), and Anderson and Robinson (*in press*). Once the new observation has been placed in the canonical space, we can then classify it into one of the groups by observing which group centroid is the closest to it in the canonical space (e.g., Seber 1984).

### Goodness of fit

The relative sizes of the canonical correlations provide a measure of the relationship between the response data represented by the PCO axes and the **X** matrix. If the **X** matrix contains group identifiers, we may wish to know how distinct the groups are in multivariate space. A method that will provide a statistical estimate of misclassification error is the "leave-one-out" approach of Lachenbruch and Mickey (1968). Here, we take out a single observation and do the CAP analysis on the remaining observations without it. We then use the method just described to classify the "left out" observation in the canonical space determined by the rest of the observations. We know a priori which group the left-out observation really came from. Was the classification successful? We can do this for every single observation in the data set and we can then calculate the proportion of the observations that were incorrectly classified. This is the misclassification error, which equals one minus the proportion of correct classifications. This approach is sometimes erroneously referred to as "jackknife classification," but is more appropriately called the "leave-one-out" method (e.g., Seber 1984). Although it will provide an estimate of the misclassification error with greater variance than some alternatives, such as the bootstrap (Efron 1979), it does, nevertheless, give us a reasonable and nearly unbiased measure of how distinct the groups are in multivariate space (for further details, see Seber [1984]:288–293).

The misclassification error can provide a measure of goodness of fit when **X** contains dummy variables cor-

Table 2. Extended.

| Criterion for drawing ordination axes | Matrix for eigenvalue decomposition‡ |
|---|---|
| finds axis of maximum correlation between **Y** and some linear combination of variables in **X** (i.e., multivariate multiple regression of **Y** on **X**, followed by PCA on fitted values, **Ŷ**) | $\mathbf{Y}_c' \mathbf{X}_c \, (\mathbf{X}_c' \mathbf{X}_c)^{-1} \, \mathbf{X}_c' \mathbf{Y}_c$ |
| same as RDA, but **Y** are transformed to **Y**\* and weights (square roots of row sums) are used in multiple regression‖ | $\mathbf{C}^{-1/2} \, \mathbf{Y}' \mathbf{X}_c \, (\mathbf{X}_c' \mathbf{R} \mathbf{X}_c)^{-1} \, \mathbf{X}_c' \mathbf{Y} \mathbf{C}^{-1/2}$ |
| finds linear combination of variables in **Y** and **X** that are maximally correlated with one another | $\mathbf{Y}_s' \mathbf{X}_c \, (\mathbf{X}_c' \mathbf{X}_c)^{-1} \, \mathbf{X}_c' \mathbf{Y}_s$ |
| finds axis that maximizes differences among group locations. Same as CCorA when **X** contains group identifiers. Equivalent analysis is regression of **X** on **Y**, provided **X** contains orthogonal contrast vectors | as for CCorA, but CDA eigenvalues are $\delta_i^2/(1 - \delta_i^2)$, and corresponding eigenvectors are $(1 - \delta_i^2)^{-1/2}$ times the CCorA eigenvectors¶ |
| finds linear combination of axes in $\mathbf{Q}_m$ and in **X** that are maximally correlated, or (if **X** contains group identifiers) finds axis in PCO space that maximizes differences among group locations | $\mathbf{Q}_m' \mathbf{X}_c \, (\mathbf{X}_c' \mathbf{X}_c)^{-1} \, \mathbf{X}_c' \mathbf{Q}_m$ |

responding to group structure. More generally, the goodness of fit of the relationship between the response data cloud and any set of variables (such as environmental variables) contained in **X** may be provided by the sum of squared distances from each individual point in the multivariate cloud to their position predicted on the fitted canonical axis (or axes) obtained using the rest of the observations, as a ''leave-one-out'' residual sum of squares. Details of the calculation of this residual error are given in the Appendix.

### How should the number of PCO axes (m) be chosen?

An extremely important point is to determine how many PCO axes should be retained; i.e., what should be the choice for $m$? One might be concerned that the method could suffer from a poor choice for $m$. If $m$ is too small, then there may be some ecologically important information in the data that will not be included in the canonical analysis. On the other hand, if $m$ is too large relative to $N$, then a misleading canonical plot could result. For example, we cannot set $m = N - 1$, for then the squared canonical correlations would all be equal to 1.0, because we have $N$ points! That is, we can of course achieve a perfect fit of our hypothesis to the $N$ points if we use all of the nonzero $N - 1$ coordinate axes. Although this was suggested as a constrained ordination technique by ter Braak (1992) and McArdle and Anderson (2001), in the one-way ANOVA case it yields points in a canonical space that have a within-group variability of zero. That is, all of the observation points within a single group are superimposed on one another at a single point, so the canonical plot of, say, three groups, shows three points. This is simply replicating the hypothesis itself, rather than showing the potential relationship of the multivariate data to it. In fact, $m$ cannot approach the value

of $N$ or the patterns on the canonical plot will suggest strong separation of groups that may not actually be present in multivariate space.

If $N > p$ and all eigenvalues from the PCO are positive, i.e., a metric distance measure has been chosen that does not produce any problems of ''non-Euclideanarity'' (e.g., Legendre and Legendre 1998:432–433), then the PCO will produce a maximum of $p$ positive eigenvalues and all corresponding axes may be used for the canonical analysis, (i.e., one may choose to use $m = p$). Also, in the event that $N > p$ and one wishes to base the analysis on certain distance measures (including chi-square distance, chi-square metric, Orlóci's chord distance, distance between species profiles, or Hellinger distance), then one can simply transform the data in the particular manner described for each of these distance measures by Legendre and Gallagher (2001) and proceed with the CDA (or the CCorA) on the transformed data. This approach has the advantage of cutting out the need for the PCO altogether. Furthermore, the new axes in the constrained ordination would be linear combinations of the transformed variables, so determining the relative importance of the original transformed response variables and the production of biplots would also be straightforward. If $p$ approaches $N$, however, some method for the reduction of dimensions will be needed to avoid the production of a misleading canonical plot.

In many situations, the dissimilarity measure chosen is non-Euclidean (such as Bray-Curtis, CY dissimilarity, etc.) or $p \geq N$, in which case a subset of the PCO axes will have to be chosen. In choosing $m$, we will wish to include all of the most salient patterns and ecologically important information in the original data set, without including extraneous random variation. This can be done by sequentially increasing the value of $m$ for the CAP analysis and, each time, calculating
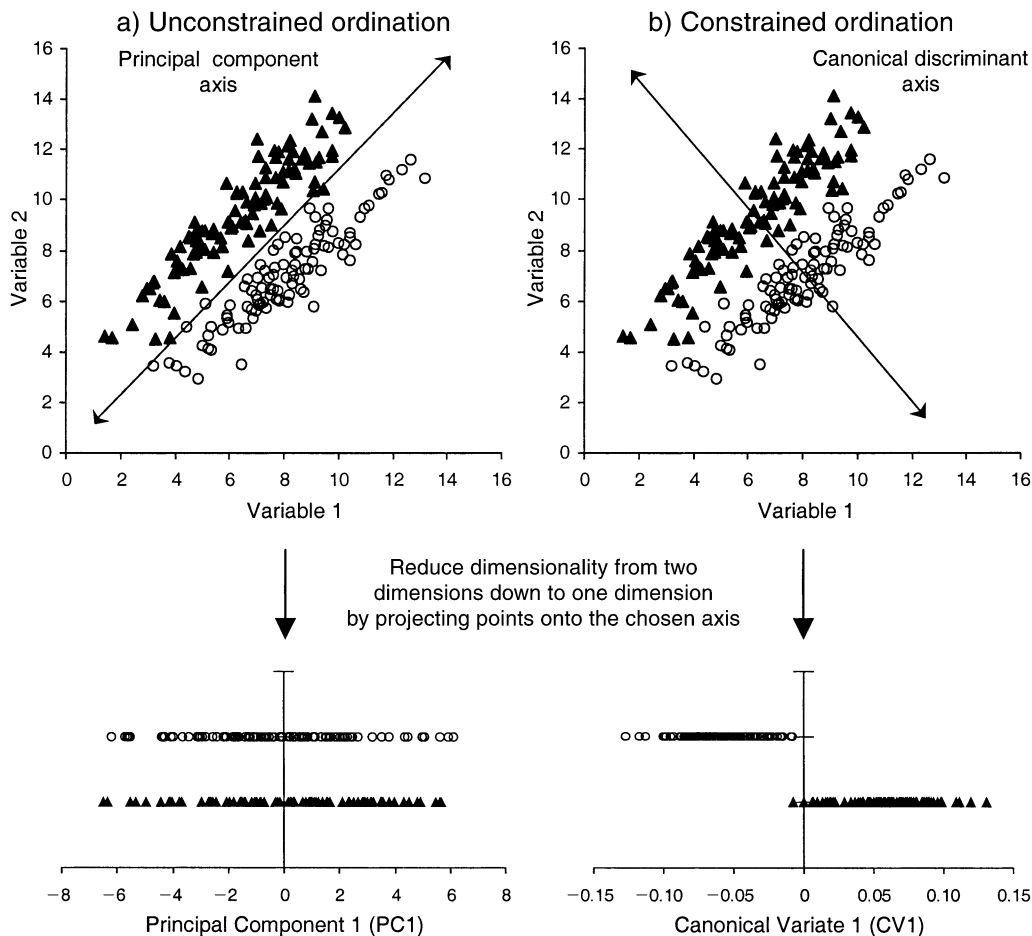
FIG. 1. Visual comparison of the method used to reduce dimensions in (a) an unconstrained and (b) a constrained ordination procedure. Data were simulated from a multivariate normal distribution with the two groups having different centroids (6, 9) and (9, 7), but both variables had a standard deviation of 2, and the correlation between the two variables was 0.9. Note the difference in scale between the first canonical axis (CV1) and the first principal component (PC1).

the residual error and (in the case in which **X** specifies groups) the misclassification error. We shall choose $m$ where the residual error (or the misclassification error) is at its minimum.

Even in cases where $p < N$ and the chosen measure has Euclidean properties, an important strength of the CAP approach is that the PCO can be used to reduce the number of dimensions of the problem. The problem of choosing an appropriate subset of variables for traditional discriminant analysis when $p$ is too large is well known (e.g., Williams 1983, Seber 1984, Williams and Titus 1988, ter Braak 1995, Legendre and Legendre 1998). As pointed out by Legendre and Legendre (1998), Williams and Titus (1988) suggested for CDA that the number of observations per group should be at least three times the number of variables. The essential problem is that if the number of variables ($p$ or $m$) even begins to approach $N$, then any patterns seen in the canonical analysis may simply be "fortuitous," and some check against the arbitrariness of the canonical results is necessary (e.g., Williams 1983). Mini-

mization of the misclassification error or of the residual sum of squares using the leave-one-out procedure, as we suggest here, provides such a necessary check. In traditional CDA, the problem of variable selection has all of the problems associated with choosing appropriate subsets of variables in multiple regression (i.e., issues of multicollinearity in the search for the "best" possible subset). In contrast, the CAP procedure uses PCO as a first step, which yields a series of ordered, orthogonal axes that explain smaller and smaller proportions of the original variance in the species data, and thus can simply be added sequentially until a minimum error is achieved.

As an added note, we consider that there is also a logical upper limit for the choice of $m$: it should not include PCO axes such that their total sum of squares exceeds the total sum of squares in the original dissimilarity matrix. This may occur in the event that the PCO produces negative eigenvalues.

Thus, the PCO, coupled with the calculation of misclassification error (or residual error) for increasing

values of $m$, as suggested here for the CAP procedure, will not only provide us with a non-arbitrary method of choosing an appropriate value for $m$, but also will provide us with a check on the arbitrariness of the canonical analysis itself for larger numbers of dimensions. We suggest that this (or a similar) check should be used even if $p < N$ and the method of transforming the data in the manner suggested by Legendre and Gallagher (2001) is possible.

As a general rule of thumb, we also suggest that the choice of $m$ should be made such that the percentage of the total variance in the dissimilarity matrix explained by the first $m$ PCO axes should exceed 60%, but should not exceed 100%. This prevents $m$ from being so small as to not encompass the majority of the information in the dissimilarity matrix. Within these bounds, $m$ should be chosen to minimize either the misclassification error (in the case of CDA) or the residual sum of squares (in the case of CCorA).

### Which species are responsible for multivariate patterns?

A further use of the proposed canonical analysis is its potential for identifying the species (or original variables) responsible for multivariate patterns. In particular, one can simply calculate the product-moment correlation coefficient for each individual species with each of the canonical axes. For the case in which $\mathbf{X}$ specifies group identifiers, species with high absolute correlations (positive or negative) will probably have an important story to tell with regard to the nature of group differences. If $\mathbf{X}$ contains quantitative variables, then these correlations will indicate which species may show important relationships with those variables. In either case, correlations of species with canonical axes will provide a good indication of which species should be investigated in more detail with univariate analysis. Note that the original transformation or standardization used on the response variables before calculating dissimilarities should also be done before calculating such correlations with the canonical axes. Note also that, in keeping with the philosophy that species may have nonlinear relationships with canonical axes, a rank correlation or some other modeling measure of choice may be used here, rather than the simple Pearson's correlation, if desired.

Clearly, this use of correlations with canonical axes is an indirect ''post hoc'' way of identifying possible contributions of individual species to differences among groups. Nevertheless, correlations with canonical axes will provide a quantitative and objective approach to usefully direct subsequent univariate analyses, allowing multivariate effects to be meaningfully characterized in terms of the original variables. Other approaches that could be used to advantage here include the IndVal method (Dufrêne and Legendre 1997) or the SIMPER method (Clarke 1993).

### ECOLOGICAL EXAMPLE

#### Fish assemblages from the Poor Knights Islands at three different times

These data come from a study by Willis and Denny (2000) on the effects of a marine reserve on assemblages of fish at the Poor Knights Islands, New Zealand. Only a small part of a much larger study is discussed here. The data consisted of counts of abundances of species from underwater visual censuses taken along transects measuring $25 \times 5$ m at a depth of 8–20 m. Counts from nine such transects done at a single site were pooled to provide a single observation row vector. The censuses were done at each of $g = 3$ different times, September 1998 ($n_1 = 15$), March 1999 ($n_2 = 21$), and September 1999 ($n_3 = 20$), with the first census occurring before the Poor Knights Marine Reserve was established on 1 October 1998. There were 47 species of fish recorded (i.e., 47 variables) and $N = 56$. Data from the three census times were considered independent of one another, because fish are highly mobile and the exact transect locations were not repeated in the different surveys. Thus, under the null hypothesis of no differences in fish assemblages among the three times, the observation row vectors were exchangeable.

For multivariate analysis, the data were transformed to $y' = \ln(y + 1)$ to remove large differences in scale among the original variables. Then Bray-Curtis dissimilarities were calculated between every pair of observations, and an unconstrained ordination was done using PCO (metric MDS) on the dissimilarity matrix. The first two PCO axes explained 20.72% and 12.37% of the variability in the original dissimilarity matrix. The scatter of points for the three groups in this ordination suggested that there were no important differences among the groups (Fig. 2a). A similar pattern was obtained using nonmetric MDS (ordination not shown). However, a nonparametric multivariate analysis of variance (NPMANOVA; Anderson 2001$a$) indicated that there were significant differences among the groups ($F_{2,53} = 2.81$, $P = 0.0004$, with 4999 permutations). This result seemed perplexing. How could the groups be significantly different when the scatter of points in the MDS plot suggested a great deal of overlap among the groups?

A constrained ordination provided important insight. The first step was to choose the number of PCO axes ($m$) to include in the CAP analysis. This was done by plotting the proportion of correct allocations obtained with increases in the number of axes included in the analysis (Fig. 3a). We chose to use $m = 7$, which achieved the maximum proportion of correct allocations (71.43%) of any choice of $m$ (i.e., minimum misclassification error = 28.57%). Any value of $m$ greater than this would increase the within-group variability more than the between-group variability, and thus would be of no use for discrimination among groups.
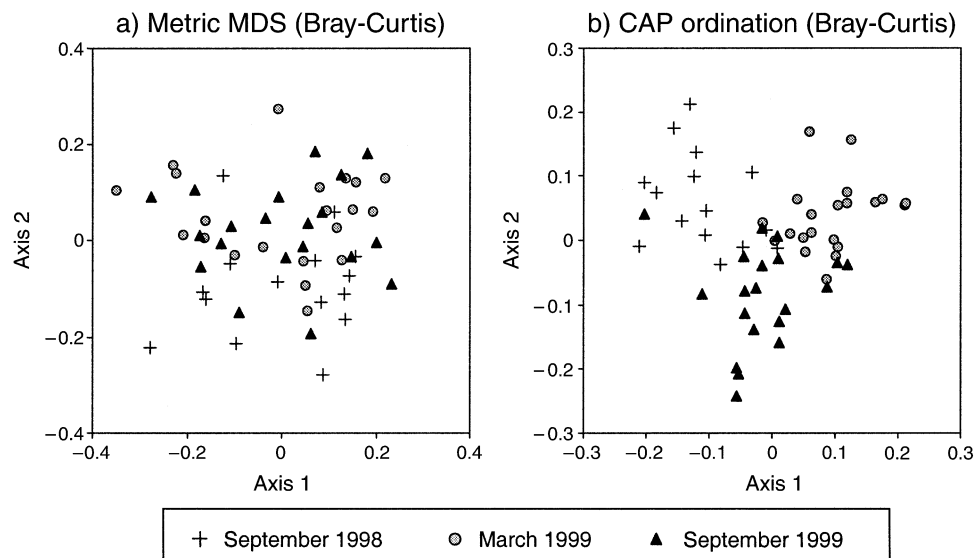
FIG. 2. (a) Unconstrained and (b) constrained ordinations of fish assemblage data from the Poor Knights Islands, New Zealand, at three different times of sampling.
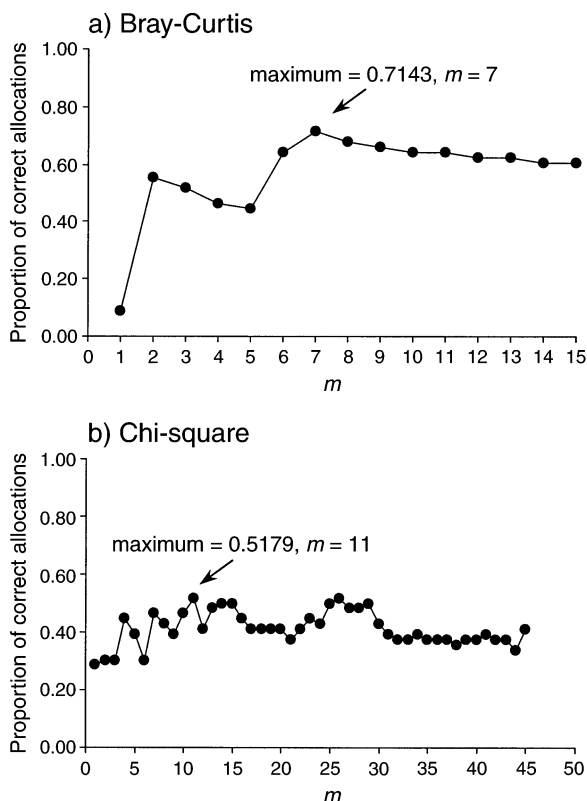


FIG. 3. Plots of the proportion of correct allocations of observations to groups (= 1 minus the misclassification error) with increases in the number of principal coordinate axes ($m$) used for the CAP procedure on data from the Poor Knights Islands at three different times on the basis of (a) the Bray-Curtis dissimilarity measure on data transformed to $y' = \ln(y + 1)$, and (b) the chi-square distance measure.

The first seven PCO axes explained 71.38% of the variability in the original dissimilarity matrix.

The canonical analysis (CAP) yielded two canonical axes (as $\min(g - 1, m) = 2$ in this case), with squared canonical correlations of $\delta_1^2 = 0.610$ and $\delta_2^2 = 0.478$. The two canonical test statistics were highly significant ($P = 0.0001$ for both tests, using 9999 permutations), which is consistent with the result obtained using NPMANOVA.

Importantly, a plot of the first two canonical axes shows a pattern of differences among the groups; i.e., observations in the same group are clustered together on the plot (Fig. 2b). This is a good example of a situation in which real group differences in multivariate space were not apparent in the unconstrained ordination, but were uncovered and displayed by the CAP analysis. The reduction of dimensions from $p = 47$ down to $m = 7$, chosen by checks on the misclassification error (Fig. 3a), ensured that patterns seen on the plot would not be overparameterized and misleading. Thus, it is clear that there were significant changes in the multivariate fish assemblages through time after the establishment of the marine reserve at the Poor Knights.

We next considered the correlations of individual species (after transformation) with canonical axis 1 to characterize the multivariate effect (Table 3). Also, correlations of species with both CAP axes are shown graphically in Fig. 4, which, when superimposed on Fig. 2b, produces a biplot. We did not include in Table 3 any species with an absolute correlation of <0.20 or that occurred in fewer than six observations. Note that a positive correlation for a species with axis 1 indicates increasing numbers after establishment of the marine reserve in October of 1998, whereas a negative cor-

TABLE 3.　Individual fish species showing absolute correlations of $|r| \geq 0.20$ with canonical axis 1 that separated the three different times of sampling (see Fig. 2b), along with their average abundance at each time.

| Fish species | Correlation | $\bar{y}_{Sep98}$ | $\bar{y}_{Mar99}$ | $\bar{y}_{Sep99}$ |
|---|---|---|---|---|
| *Pagrus auratus* | 0.7451 | 0.20 | 7.71 | 2.05 |
| *Pseudolabrus luculentus* | 0.5656 | 1.47 | 7.62 | 6.20 |
| *Dasyatis brevicaudata* | 0.4596 | 0.53 | 1.52 | 0.05 |
| *Coris sandageri* | 0.4184 | 14.07 | 30.43 | 20.85 |
| *Girella cyanea* | 0.3842 | 0.13 | 2.76 | 1.05 |
| *Seriola lalandi* | 0.3698 | 0.07 | 1.48 | 1.45 |
| *Nemadactylus douglasii* | 0.3219 | 2.47 | 3.76 | 3.25 |
| *Amphichaetodon howensis* | 0.3138 | 0.13 | 0.90 | 0.25 |
| *Aplodactylus etheridgii* | 0.3118 | 0.07 | 0.24 | 0.05 |
| *Dasyatis thetidis* | 0.3106 | 0.00 | 0.86 | 0.00 |
| *Notolabrus fucicola* | −0.2959 | 3.53 | 1.71 | 2.65 |
| *Odax pullus* | −0.2777 | 5.13 | 2.14 | 0.75 |
| *Parika scaber* | −0.2252 | 10.73 | 7.48 | 6.40 |
| *Atypichthys latus* | 0.2105 | 0.27 | 0.67 | 0.50 |

*Note:* Species occurring in fewer than six observations were not included.

relation with axis 1 indicates decreasing numbers after this time (Fig. 2b).

Although no univariate statistical analyses are given here, we noted several patterns by reference to the magnitude of the correlation and the average abundances of fish in each of the three sampling times. First, it is clear that cessation of fishing inside the reserve was correlated with an increase in the numbers of commercially targeted fish species, such as snapper (*Pagrus auratus*) and kingfish (*Seriola lalandi*). In addition, other fish that may be prone to increased mortality due to fishing have high correlations with the canonical axis, e.g., bluefish (*Girella cyanea*) and porae (*Nemadactylus douglasii*). Similarly, orange wrasse (*Pseudolabrus luculentus*) and Sandager's wrasse (*Coris sandageri*) also increased in density after the cessation of fishing, although this was more likely due to strong recruitment events prior to the March survey rather
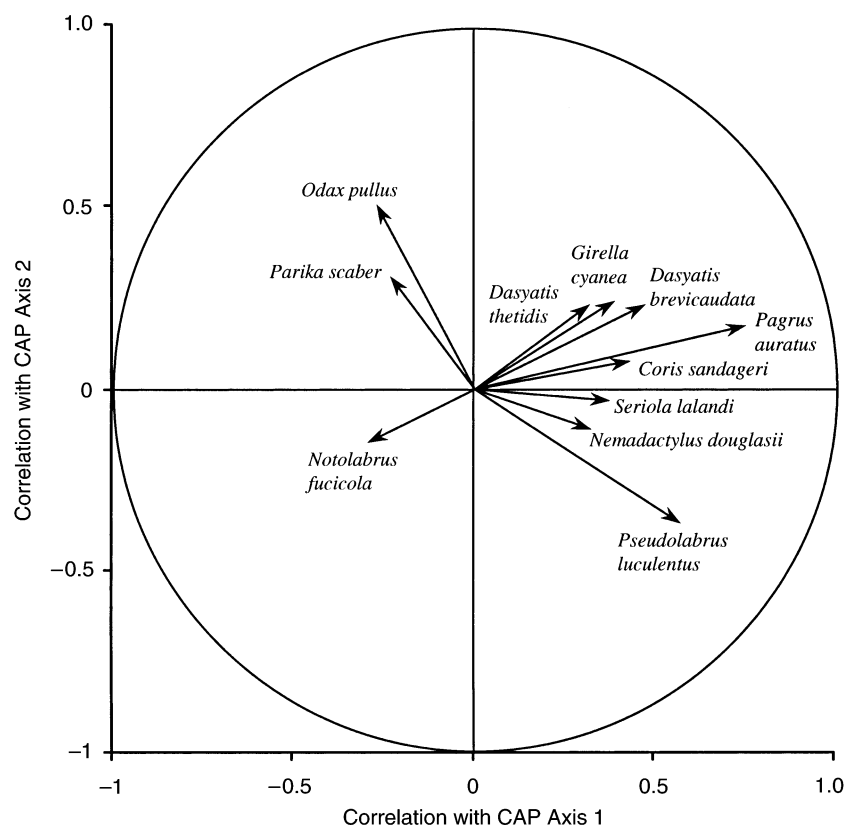


FIG. 4.　Correlations of several original species variables with the two CAP axes of Fig. 2b.

than to any change in fishing mortality (T. J. Willis, *personal observation*). Some of the species that showed correlation with the canonical axis may have changed in abundance through time because of seasonal differences between September (relatively cold) and March (warm). This pattern was shown by the two species of stingray (*Dasyatis thetidis* and *D. brevicaudata*), which tend to move to shallower waters in warmer seasons, and by the Lord Howe coralfish (*Amphichaetodon howensis*), which may be observed in southern temperate areas in warm seasons.

There were more species that showed positive correlations than negative correlations with the canonical axis, suggesting that the overall abundance and diversity of fish increased through time at the Poor Knights Islands after the establishment of the marine reserve. Some fish species, however, did decrease in abundance: banded wrasse (*Notolabrus fucicola*), butterfish (*Odax pullus*), and leatherjacket (*Parika scaber*). This may have been due to competitive pressure from increases in other species or large-scale changes in oceanic climate. These issues and more specific univariate analyses will be addressed in a future paper.

## COMPARISONS WITH ALTERNATIVES

There are essentially two important strengths of the CAP approach over existing methods. First, the CAP approach allows any dissimilarity or distance measure of choice to be used as the basis of the analysis. This flexibility is important. Different dissimilarity measures emphasize different aspects of the multivariate data set. Thus, one may investigate how well the data fit an a priori hypothesis when different aspects of the multivariate data are considered through the use of different distance measures. For example, chi-square distances emphasize differences in composition of assemblages, whereas Bray-Curtis dissimilarities tend to emphasize differences in relative abundance (which is one reason that a transformation is usually recommended before applying Bray-Curtis if variables are on different scales; Clarke 1993). Furthermore, in the context of modeling, some dissimilarity measures may be better than others for providing discrimination among a priori groups.

Second, the use of a principal coordinate analysis as part of the CAP procedure allows the dimensionality of the system to be reduced to $m$ orthonormal axes prior to the canonical analysis. This reduction is achieved with an important check on the arbitrariness of the canonical results and, as a consequence, produces a very reliable procedure for canonical ordination.
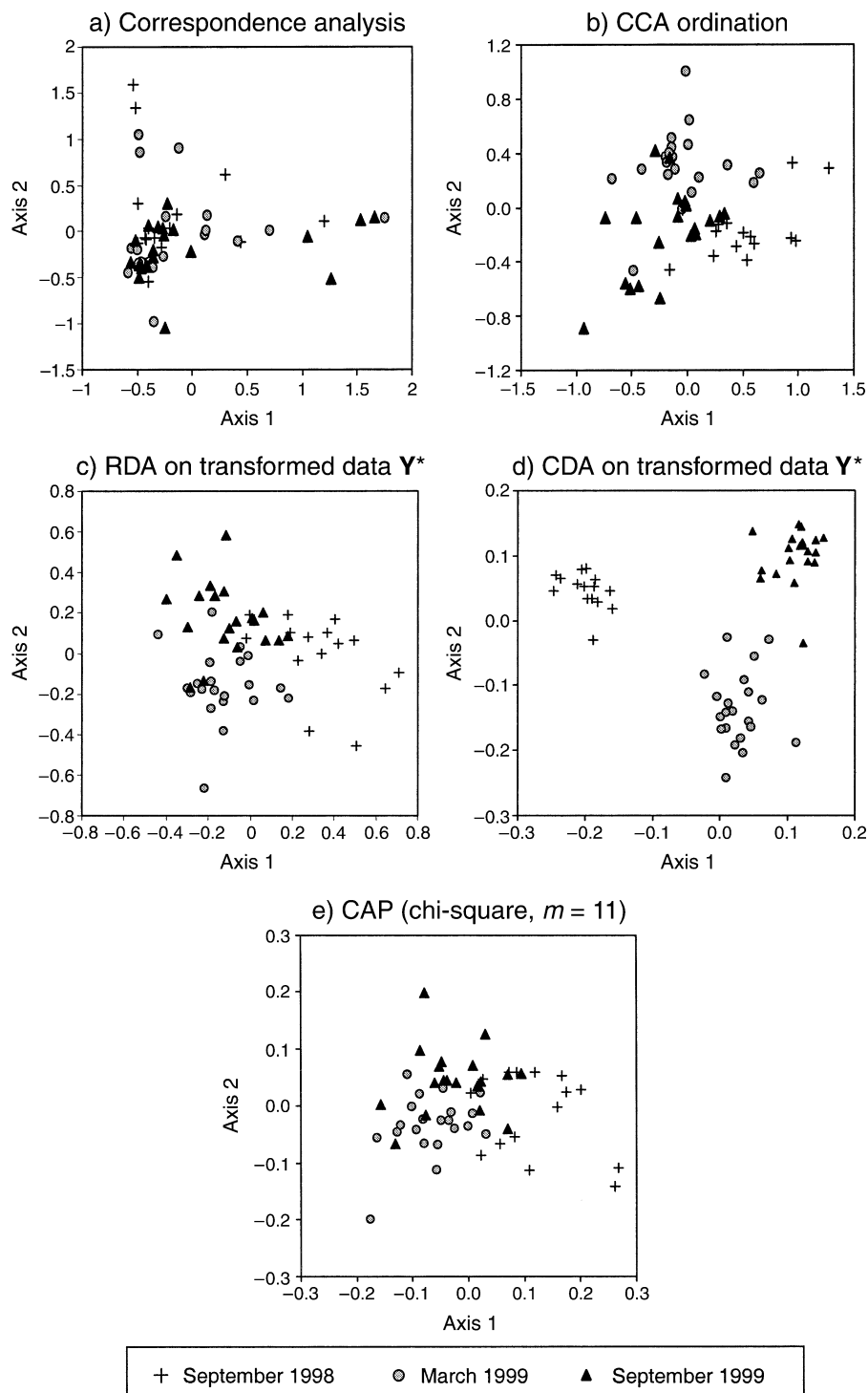
To illustrate these points, we considered several other ordinations of the fish data (Fig. 5). First, we considered a correspondence analysis (Fig. 5a). The first two CA axes together explained 32.5% of the variability in the original species data and, like the unconstrained MDS plot (Fig. 2a), suggested no apparent differences in the fish assemblages among the three

times of sampling. Two traditional constrained ordinations were then done: CCA (Fig. 5b) and RDA on data transformed, as suggested by Legendre and Gallagher (2001), to $\mathbf{Y}^* = (y_{ij}^*)$:

$$y_{ij}^* = y_{ij} \frac{\sqrt{T}}{R_i \sqrt{C_j}}$$

where $T$ is the sum of all values in the data table; $R_i$ is the sum of values in row $i$; and $C_j$ is the sum of values in column $j$ (Fig. 5c). Euclidean distances among observation rows in $\mathbf{Y}^*$ are equivalent to chi-square distances. The first two CCA axes accounted for 6.6% of the original species data, with canonical correlations of 0.666 and 0.590. The first two RDA axes on transformed data accounted for 6.2% of the original species data, with canonical correlations of 0.775 and 0.630. The tests by permutation indicated, in each case, that there were significant differences in fish assemblages among the three times of sampling ($F_{3,53}^{\text{CCA}} = 1.875$, $P = 0.0074$; and $F_{3,53}^{\text{RDA}} = 1.749$, $P = 0.0060$, each with 4999 permutations). The patterns in these two plots are similar, and only differ because of the weights used in CCA, as discussed in detail by Legendre and Gallagher (2001).

A traditional canonical analysis (CDA) was also done on the transformed variables, $\mathbf{Y}^*$ (Fig. 5d). This is equivalent to a CAP analysis based on chi-square distances, but using all of the principal coordinate axes corresponding to nonzero eigenvalues. For these data, 100% of the variability in the distance matrix was explained by $m = 46$ PCO axes. The canonical correlations were very high ($\delta_1 = 0.980$, $\delta_2 = 0.925$) and the pattern on the plot suggests that the groups are very distinct. It is tempting to jump to the conclusion that the method has produced an ordination much better than that of the other approaches. However, this is dangerous, for we must consider how much of this apparently ''clear'' pattern is ''fortuitous,'' i.e., due to the simple fact that the number of axes used is quite large and is approaching the total number of observations. That is, although the groups appear distinct on the plot, this ordination is misleading, caused by overparameterization ($m$ is too close to $N$) and does not reflect true patterns, given the within-group variability in the data. A check on the arbitrariness of the dimensionality is provided by the calculation of leave-one-out misclassification error for increasing values of $m$. This indicates that the best choice for the CAP analysis on the basis of the chi-square distance is $m = 11$, which gives 51.79% correct allocations (Fig. 3b). Contrast this with the analysis using $m = 46$, which resulted in only 32.14% correct allocations (a much higher misclassification error of 67.86%), in spite of the apparent distinctness of groups on the plot (Fig. 5d). The CAP ordination using $m = 11$ shows a similar pattern to that of the RDA plot, with $\delta_1 = 0.750$, and $\delta_2 = 0.514$ ($P = 0.0525$, with 9999 permutations). These results of

FIG. 5. Comparison of patterns obtained using five different ordination procedures on the fish assemblage data from the Poor Knights Islands at three different times of sampling.

CAP on the basis of chi-square distances may be compared with the results of the CAP ordination on the basis of Bray-Curtis dissimilarities (Fig. 2b), with $\delta_1$ = 0.784, and $\delta_2$ = 0.714 ($P$ = 0.0001, with 9999 permutations).

Thus, the CAP method will not necessarily draw a "better" picture of the data compared to other constrained ordination techniques; it will not necessarily show the clarity of the hypothesis in the data where other methods will or will not fail. Rather, its advantage

TABLE 4. CAP analysis of the multivariate fish species data at three different times from the Poor Knights Islands, New Zealand, done using several different distance or dissimilarity measures.

| Measure | $m$ | Proportion of $ss_T$ | $\delta_1^2$ | $\delta_2^2$ | Correct allocations (%) |
|---|---|---|---|---|---|
| Bray-Curtis† | 7 | 0.7138 | 0.6100 | 0.4781 | 71.43 |
| Canberra† | 13 | 0.9706 | 0.5919 | 0.5190 | 48.21 |
| Chi-square | 11 | 0.7561 | 0.5620 | 0.2638 | 51.79 |
| Euclidean† | 27 | 0.9756 | 0.8516 | 0.6569 | 66.07 |
| CY dissimilarity | 11 | 0.8511 | 0.6941 | 0.5442 | 75.00 |
| Gower‡ | 17 | 0.8427 | 0.6290 | 0.4900 | 55.36 |
| Hellinger | 12 | 0.8511 | 0.7193 | 0.4349 | 71.43 |
| Kulczynski† | 8 | 0.7701 | 0.6186 | 0.5108 | 64.29 |
| Orloci's chord | 19 | 0.9859 | 0.6913 | 0.4550 | 64.29 |

*Notes:* The value of $m$ in each case was obtained using criteria given in the text. ''Proportion of $ss_T$'' is the proportion of the total sum of squared interpoint dissimilarities divided by the number of points (total variance) explained by the first $m$ PCO axes. The last column indicates the percentage of correct allocations obtained using the ''leave-one-out'' procedure. Details of the measures are given in Legendre and Legendre (1998:Chapter 7), except the CY dissimilarity measure, which is described in Cao et al. (1997).

† The data were transformed to $y' = \ln(y + 1)$ in all of these cases; the other measures have intrinsic standardizations for differences in the scale of variables.

‡ Gower's measure is a flexible one. The form used here was described in Legendre and Legendre (1998:266–267).

is in providing the experimenter with the flexibility to choose any measure of dissimilarity as the basis of the analysis and in providing a non-arbitrary way of reducing the dimensionality of the problem and, concomitantly, any potential arbitrariness in the canonical results.

In addition to the Bray-Curtis and chi-square distance measures, the CAP analysis was done on the fish data set using a suite of many different dissimilarity measures, with results shown in Table 4. For these data, three measures seemed to provide the best fit to the hypothesis of three separate groups at the three separate times: Bray-Curtis, Hellinger, and the CY measure. The CY measure was developed to provide a sensitive measure that would give equal weighting to different types of multivariate variation in species abundances between samples (Cao et al. 1997). It is, essentially, an average of the proportional (multiplicative) change in each species (i.e., changes on the log scale); see Cao et al. (1997) for details. The Hellinger distance uses differences in the square root of row-proportioned abundance values. Although Bray-Curtis is known to be overly sensitive to differences in absolute abundances, with the transformation, in this case to $y' = \ln(y + 1)$, this effect is largely reduced (Clarke 1993, Cao et al. 1997) and also places observations on the logarithmic (multiplicative) scale. What these three techniques have in common suggests that multiplicative changes or changes in the proportional abundances of species present at the Poor Knights may have contributed more to changes through time than did differences in absolute abundances or composition of the assemblages, although obviously more study would be needed to investigate this idea.

We do not suggest from the study of this data set that one or the other of these distance measures is ''best'' for any given situation. Rather, it is clear that different measures will show different pictures of (i.e., will emphasize different aspects of) the multivariate data. An advantage of the CAP approach is that it will allow any measure to be chosen for the analysis.

By having the flexibility to choose any measure, the experimenter is also free to develop any model of choice regarding the relationship between the original variables and the canonical axes. There is no constraint that this should necessarily be unimodal or linear, as is the case with CCA or RDA, respectively (see Table 2). Although the metric PCO stipulates a linear relationship between the original dissimilarities among observations and their respective interpoint distances in the Euclidean PCO space, this does not in any way constrain the relationship of the PCO axes (or the canonical axes) with the original variables to be linear.

Furthermore, the unscaled orthonormal PCO axes are already standardized to unit variance, ready for the ensuing canonical analysis. Note that, in general, if matrix $\mathbf{Y}$ is standardized by its variance–covariance matrix (i.e., to $\mathbf{Y}_s = \mathbf{Y}_c(\mathbf{Y}_c'\mathbf{Y}_c)^{-1/2}$, where $\mathbf{Y}_c = \mathbf{Y} - \mathbf{j}\bar{\mathbf{y}}'$, with $\mathbf{j}$ being an $n \times 1$ column of 1's and $\bar{\mathbf{y}}$ being a $p \times 1$ column of means, Table 2), then RDA of $\mathbf{Y}_s$ vs. $\mathbf{X}$ is equivalent to CCorA of $\mathbf{Y}$ with $\mathbf{X}$ (e.g., ter Braak 1995, Mielke and Berry 1999). Thus, an important difference between RDA and CCorA is that the former does not (generally) take into account variance–covariance (or correlation) structure among variables in the response data cloud (Table 2). Note that if the PCO axes ($\mathbf{Q}$) were scaled according to their eigenvalues (as is generally done to draw the PCO ordination; Gower 1966$a$), then the first step for the canonical analysis would be to rescale these axes back to unit variance. Anderson and Robinson (*in press*) found that CAP is more powerful than RDA on the full set of PCO axes (scaled to

their respective eigenvalues, as described by Legendre and Anderson [1999] and McArdle and Anderson [2001]), when there are several highly correlated variables with relatively large variances that are unrelated to group structure. This is consistent with the results of Mielke and Berry (1999), who found that the relative power of multi-response permutation tests (e.g., Mielke et al. 1976), which do not take into account correlation structure, vs. traditional canonical tests, which do, depended on the direction of group differences by reference to the correlation structure among variables.

If desired, as a variation on the CAP approach, one could consider using the axes from a nonmetric MDS as input into a traditional canonical analysis. There are three potential drawbacks to this approach. First, the nonmetric MDS algorithm requires the dimensionality to be chosen a priori. It will usually not be clear ahead of time what an appropriate dimensionality should be for any particular problem at hand. Second, sequential axes in a nonmetric MDS solution are not necessarily orthogonal to one another, as they are in the case of PCO. Thus, even if an appropriate dimensionality could somehow be determined by some sequential addition of axes to a nonmetric MDS, an extra step would be required to standardize the axes according to their variance–covariance structure for the canonical analysis. Third, it is not clear how one would place a new observation point into an existing nonmetric MDS plot; the entire MDS algorithm must begin again from scratch when a new observation is introduced. This would make it impossible to place new observations into the canonical plot, impossible to calculate leave-one-out misclassification or residual errors, and thus impossible to assess the potential arbitrariness of the resulting canonical ordination. Given that the metric and nonmetric solutions are often very similar for a reasonable number of dimensions, the problems introduced by attempting to use a nonmetric MDS do not seem to provide any advantages in this context.

After considering the CAP approach, one might be tempted to ask the question: why do the unconstrained ordination at all? Shouldn't the CAP analysis be used preferentially to nonmetric or metric MDS when one has an explicit multivariate hypothesis a priori? It is important not to fall into the trap of believing that there is one single ordination that will ''do it all!'' A constrained ordination such as CAP is good for observing the potential patterns in our data with respect to an a priori hypothesis (e.g., examining location differences among groups), but it will provide us with no information at all concerning the overall pattern of dispersion of points in the multivariate cloud, or potential differences in multivariate variability, or dispersion among groups. In other words, the canonical analysis may add something to our understanding of patterns in multivariate space over and above what can be seen in an MDS plot, but it does not in any way ''replace'' the

MDS in terms of some of the important information that is useful, in general, for multivariate analysis.

## DISCUSSION

We suggest that a scientist would ideally wish to see four items in the analysis of multivariate ecological data: (1) a robust unconstrained ordination (such as a nonmetric MDS plot); (2) an appropriate constrained analysis (such as a CAP plot) by reference to a specific hypothesis; (3) a rigorous statistical test of the hypothesis; and (4) characterization of species responsible for multivariate patterns or effects. Both the constrained and unconstrained ordinations should be done on the basis of the same distance measure and the same transformation or standardization prior to analysis, in order for their joint information to be interpretable.

Unconstrained and constrained ordination procedures are both useful to obtain a more complete understanding of patterns in multivariate data. They each have a specific purpose, are drawn on the basis of different criteria, and thus demonstrate two very different, but important, aspects of the multivariate data cloud. An unconstrained ordination will allow overall patterns across the entire data cloud to be examined, as well as any differences in within-group variability or spread. In the words of Clarke (1993) and Clarke and Ainsworth (1993), the unconstrained MDS plot ''lets the data speak for themselves.'' Although group differences may be seen in an unconstrained ordination, they can also be masked by high variability and high correlation structure among variables unrelated to group differences. A constrained ordination, on the other hand, does not allow any assessment of either total or relative within-group variability, but it does allow location differences among groups to be seen, if indeed any at all are discernable in multivariate space. If the direction(s) of difference(s) among groups is (are) also generally the direction(s) of greatest total variability in the data cloud, then it is likely that group differences will be evident in the unconstrained ordination.

Previously proposed methods of constrained ordination may be problematic because (1) they have unrealistic assumptions about the distribution of the response variables (e.g., multivariate normality for discriminant analysis); (2) they require a fixed relationship between response variables and explanatory variables (e.g., unimodal for CCA, linear for CDA, CCorA, or RDA; but see Makarenkov and Legendre [2002]); or (3) they do not take into account correlation structure in the response data cloud (RDA, CCA, or MDS with constraints). It is possible to take into account correlation structure in the beginning, by standardizing the original variables by their variance–covariance matrix (e.g., see Mielke and Berry 1999). However, such a standardization is not possible if the number of variables exceeds the number of observations, which is quite common in ecological data sets. It is also not logical to do this prior to the calculation of non-Eu-

clidean dissimilarities (such as Bray-Curtis dissimilarities) among observations.

What we suggest here is a canonical analysis of principal coordinates (CAP) as a useful procedure for constrained ordination. It allows the experimenter to produce a meaningful constrained ordination on the basis of any dissimilarity matrix. This approach was first hinted at by Legendre and Legendre (1998), more fully developed by Anderson and Robinson (*in press*), and is really only new wine in old wineskins. It is a two-step process combining two existing multivariate techniques: classical multidimensional scaling (PCO) followed by a classical canonical analysis (either CDA or CCorA) on unscaled orthonormal PCO axes. An additional strength of this approach is derived from its use in placing new observations into canonical ordinations and, furthermore, from the use of misclassification errors (or residual error) to obtain a non-arbitrary decision concerning the appropriate number of dimensions to include in the canonical analysis.

The CAP approach does not replace existing robust unconstrained ordination methods, but it does add something extremely useful to these, we feel, for the analysis of multivariate data in ecology by reference to a priori hypotheses.

### Literature Cited

Anderson, M. J. 2001*a*. A new method for non-parametric multivariate analysis of variance. Austral Ecology **26**:32–46.

Anderson, M. J. 2001*b*. Permutation tests for univariate or multivariate analysis of variance and regression. Canadian Journal of Fisheries and Aquatic Sciences **58**:626–639.

Anderson, M. J., and J. Robinson. *In press.* Generalised discriminant analysis based on distances. Australian and New Zealand Journal of Statistics.

Borg, I., and P. Groenen. 1997. Modern multidimensional scaling: theory and applications. Springer-Verlag, Berlin, Germany.

Borg, I., and J. C. Lingoes. 1980. A model and algorithm for multidimensional scaling with external constraints on the distances. Psychometrika **45**:25–38.

Bray, J. R., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs **27**:325–349.

Cao, Y., P. Williams, and A. W. Bark. 1997. Similarity measure bias in river benthic Aufwuchs community analysis. Water Environment Research **69**:95–106.

Clarke, K. R. 1993. Nonparametric multivariate analyses of changes in community structure. Australian Journal of Ecology **18**:117–143.

Clarke, K. R., and M. Ainsworth. 1993. A method of linking multivariate community structure to environmental variables. Marine Ecology Progress Series **92**:205–219.

Cuadras, C. M., J. Fortiana, and F. Oliva. 1997. The proximity of an individual to a population with applications in discriminant analysis. Journal of Classification **14**:117–136.

Digby, P. G. N., and J. C. Gower. 1981. Ordination between- and within-groups applied to soil classification. Pages 63–75 *in* D. F. Merriam, editor. Down to earth statistics: solutions looking for geological problems. Syracuse University Geology Contributions, Syracuse, New York, USA.

Dufrêne, M., and P. Legendre. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monographs **67**:345–366.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics **7**:1–26.

Faith, D. P., P. R. Minchin, and L. Belbin. 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio **69**:57–68.

Field, J. G., K. R. Clarke, and R. M. Warwick. 1982. A practical strategy for analysing multispecies distribution patterns. Marine Ecology Progress Series **8**:37–52.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics **7**:179–188.

Gittins, R. 1985. Canonical analysis: a review with applications in ecology. Springer-Verlag, Berlin, Germany.

Gower, J. C. 1966*a*. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika **53**:325–338.

Gower, J. C. 1966*b*. A Q-technique for the calculation of canonical variates. Biometrika **53**:588–590.

Gower, J. C. 1968. Adding a point to vector diagrams in multivariate analysis. Biometrika **55**:582–585.

Heiser, W. J., and J. J. Meulman. 1983. Analyzing rectangular tables by joint and constrained multidimensional scaling. Journal of Econometrics **22**:139–167.

James, A. T., and G. N. Wilkinson. 1971. Factorization of the residual operator and canonical decomposition of non-orthogonal factors in the analysis of variance. Biometrika **58**:279–294.

Kenkel, N. C., and L. Orlóci. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology **67**:919–928.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing a goodness of fit to a nonmetric hypothesis. Psychometrika **29**:1–28.

Kruskal, J. B., and M. Wish. 1978. Multidimensional scaling. Sage Publications, Beverly Hills, California, USA.

Lachenbruch, P. A., and M. R. Mickey. 1968. Estimation of error rates in discriminant analysis. Technometrics **10**:1–11.

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. Second English edition. Elsevier Science, Amsterdam, The Netherlands.

Makarenkov, V., and P. Legendre. 2002. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. Ecology **83**:1146–1161.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. Multivariate analysis. Academic Press, New York, New York, USA.

McArdle, B. H., and M. J. Anderson. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology **82**:290–297.

Meulman, J. J. 1986. A distance approach to nonlinear multivariate analysis. DSWO Press, Leiden, Germany.

Meulman, J. J. 1988. Nonlinear redundancy analysis via distances. Pages 515–522 *in* H. H. Bock, editor. Classification

and related methods. Elsevier Science BV, Amsterdam, The Netherlands.

Meulman, J. J. 1992. The integration of multidimensional scaling and multivariate analysis with optimal transformations. Psychometrika **57**:539–565.

Mielke, P. W., and K. J. Berry. 1999. Multivariate tests for correlated data in completely randomized designs. Journal of Educational and Behavioral Statistics **24**:109–131.

Mielke, P. W., K. J. Berry, and E. S. Johnson. 1976. Multi-response permutation procedures for a priori classifications. Communications in Statistics—Theory and Methods **A5**(14):1409–1424.

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

Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. 1996. Applied linear statistical models. Fourth edition. Irwin, Chicago, Illinois, USA.

Palmer, M. W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. Ecology **74**:2215–2230.

Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. Sankhya A **26**: 329–358.

Seber, G. A. F. 1984. Multivariate observations. John Wiley, New York, New York, USA.

Shepard, R. N. 1962. The analysis of proximities: multidimensional scaling with an unknown distance function. Parts I and II. Psychometrika **27**:125–140,219–246.

ter Braak, C. J. F. 1986*a*. The analysis of vegetation–environment relationships by canonical correspondence analysis. Vegetatio **69**:69–77.

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

ter Braak, C. J. F. 1992. Multidimensional scaling and regression. Statistica Applicata **4**:577–586.

ter Braak, C. J. F. 1995. Ordination. Pages 91–173 *in* R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, editors. Data analysis in community and landscape ecology. Second edition. Cambridge University Press, Cambridge, UK.

Torgerson, W. S. 1958. Theory and methods of scaling. John Wiley, New York, New York, USA.

Williams, B. K. 1983. Some observations of the use of discriminant analysis in ecology. Ecology **64**:1283–1291.

Williams, B. K., and K. Titus. 1988. Assessment of sampling stability in ecological applications of discriminant analysis. Ecology **69**:1275–1285.

Willis, T. J., and C. M. Denny. 2000. Effects of Poor Knights Islands Marine Reserve on demersal fish populations. Report to the Department of Conservation, Research Grant No. 2519, Leigh Marine Laboratory, University of Auckland, Auckland, New Zealand.

## APPENDIX

A detailed statistical description of canonical analysis of principal coordinates (CAP) is available online in ESA's Electronic Data Archive: *Ecological Archives* E084-011-A1.

## SUPPLEMENT

A computer program for performing the CAP method, including all necessary diagnostics for choosing an appropriate value for *m*, is available online in ESA's Electronic Data Archive: *Ecological Archives* E084-011-S1.