A new method for non-parametric multivariate analysis of variance

MARTI J. ANDERSON
Centre for Research on Ecological Impacts of Coastal Cities, Marine Ecology Laboratories A11, University of Sydney, New South Wales 2006, Australia

Abstract   Hypothesis-testing methods for multivariate data are needed to make rigorous probability statements about the effects of factors and their interactions in experiments. Analysis of variance is particularly powerful for the analysis of univariate data. The traditional multivariate analogues, however, are too stringent in their assumptions for most ecological multivariate data sets. Non-parametric methods, based on permutation tests, are preferable. This paper describes a new non-parametric method for multivariate analysis of variance, after McArdle and Anderson (in press). It is given here, with several applications in ecology, to provide an alternative and perhaps more intuitive formulation for ANOVA (based on sums of squared distances) to complement the description provided by McArdle and Anderson (in press) for the analysis of any linear model. It is an improvement on previous non-parametric methods because it allows a direct additive partitioning of variation for complex models. It does this while maintaining the flexibility and lack of formal assumptions of other non-parametric methods. The test-statistic is a multivariate analogue to Fisher's F-ratio and is calculated directly from any symmetric distance or dissimilarity matrix. P-values are then obtained using permutations. Some examples of the method are given for tests involving several factors, including factorial and hierarchical (nested) designs and tests of interactions.

Key words: ANOVA, distance measure, experimental design, linear model, multifactorial, multivariate dissimilarity, partitioning, permutation tests, statistics.

INTRODUCTION

The analysis of multivariate data in ecology is becoming increasingly important. Ecologists often need to test hypotheses concerning the effects of experimental factors on whole assemblages of species at once. This is important for core ecological research and in studies of biodiversity or environmental impacts in many habitats, including marine subtidal environments (Warwick *et al.* 1988; Gray *et al.* 1990; Chapman *et al.* 1995; Glasby 1997), mangroves (Skilleter 1996; Kelaher *et al.* 1998), freshwater systems (Faith *et al.* 1995; Quinn *et al.* 1996) and terrestrial systems (Oliver & Beattie 1996; Anderson & Clements, in press).

Univariate analysis of variance (ANOVA) provides an extremely powerful and useful tool for statistical tests of factors and their interactions in experiments (Underwood 1981, 1997). Partitioning variation, as in multifactorial ANOVA, is particularly important for testing hypotheses in complex ecological systems with natural temporal and spatial variability. This partitioning is also needed to test multivariate hypotheses in ecology for experimental designs involving several factors.

*Present address: Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand (Email: mja@stat.auckland.ac.nz).

This paper describes a new non-parametric test of the general multivariate hypothesis of differences in the composition and/or relative abundances of organisms of different species (variables) in samples from different groups or treatments. This test is a significant advance on previous methods because it can be based on any measure of dissimilarity and can partition variation directly among individual terms in a multifactorial ANOVA model. The test is applicable to any situation where the simultaneous responses of many potentially non-independent variables (usually abundances of species in an assemblage) have been measured in samples from a one-factor or multifactorial ANOVA design.

Powerful multivariate statistical methods, such as the traditional multivariate analysis of variance (MANOVA), have existed for decades (Hotelling 1931; Wilks 1932; Fisher 1936; Bartlett 1939; Lawley 1939; Pillai 1955), but tests using these statistics rely on assumptions that are not generally met by ecological data. The assumption that the data conform to a multivariate normal distribution is particularly unrealistic for most ecological data sets. This is because the distributions of abundances of individual species are usually highly aggregated or skewed (e.g. Gaston & McArdle 1994). Also, abundances take discrete values rather than being continuous, species with small means often have asymmetric distributions because they are necessarily

truncated at zero, and rare species contribute lots of zeros to the data set. MANOVA test statistics are not particularly robust to departures from multivariate normality (Mardia 1971; Olson 1974; Johnson & Field 1993). Finally, many of these test statistics are simply impossible to calculate when there are more variables than sampling units, which often occurs in ecological applications.

Many non-parametric methods for tests of differences among *a priori* groups of observations (as in MANOVA) have been developed (Mantel 1967; Mantel & Valand 1970; Hubert & Schultz 1976; Mielke *et al.* 1976; Clarke 1988, 1993; Smith *et al.* 1990; Excoffier *et al.* 1992; Edgington 1995; Pillar & Orlóci 1996; Legendre & Anderson 1999). These methods generally have two things in common. First, they are based on measures of distance or dissimilarity between pairs of individual multivariate observations (which I will refer to generally as distances) or their ranks. A statistic is constructed to compare these distances among observations in the same group versus those in different groups, following the conceptual framework of ANOVA. Second, they use permutations of the observations to obtain a probability associated with the null hypothesis of no differences among groups.

These non-parametric methods generally fall into two categories. First, there are those that can be based on any chosen distance measure. There are many such measures and these have different properties, which make them appropriate for different kinds of data (Legendre & Legendre 1998). For example, to express differences in community structure, the semimetric Bray–Curtis measure of ecological distance (Bray & Curtis 1957) or Kulczynski's (1928) semimetric measure are generally preferred over metric measures, like Euclidean distance (Odum 1950; Hajdu 1981; Faith *et al.* 1987; Clarke 1993). The methods that are flexible enough to be used with any such distance measure (e.g. Mantel 1967; Hubert & Schultz 1976; Smith *et al.* 1990; Clarke 1993) have much to recommend them for this reason.

The drawback to using these methods is that they are not able to cope with multifactorial ANOVA. That is, they are not able to partition variation across the many factors that form part of the experimental design. Consequently, for most complex designs, one must analyse data as one-way analyses in multiple subsets within particular levels of factors. These multiple one-way analyses and qualitative interpretations of ordination plots are then used to infer something about interactions or variability at different spatial scales (e.g. Anderson & Underwood 1994; Kelaher *et al.* 1998).

Some of the proposed non-parametric methods do allow partitioning for a complex design (e.g. Excoffier *et al.* 1992; Edgington 1995; Pillar & Orlóci 1996), but these are restricted for use with metric distance measures, which are not ideal for ecological applications.

Furthermore, even if these statistics were to be used, there has been disagreement concerning appropriate permutational strategies for complex ANOVA, particularly for tests of interactions (e.g. Edgington 1995; Manly 1997). There have been some recent examples of direct statistical analyses of Bray–Curtis distances (Faith *et al.* 1995; Underwood & Chapman 1998). These are restricted, however, to very specific experimental designs or hypotheses and cannot be used for any multifactorial ANOVA design.

Ecologists need a non-parametric multivariate method that can partition variation based on any distance measure in any ANOVA design. The method needs to be robust, interpretable by reference to the experimental design, and should lack formal assumptions concerning distributions of variables. The purpose of this paper is to outline just such a method and to give some ecological examples of its use. The more general mathematical theory underlying this method, along with simulations and a comparison with the related approach of Legendre and Anderson (1999), is described elsewhere (McArdle & Anderson, in press).

STRATEGY FOR NON-PARAMETRIC MULTIVARIATE ANALYSIS

An outline for a general approach to the analysis of multivariate data in ecology was given by Clarke and Green (1988) and Clarke (1993). For experimental designs used to test hypotheses defined *a priori*, there are essentially four steps: (i) a choice is made concerning an appropriate transformation and/or standardization (if any) to apply to the data, given the hypothesis and the scales and nature of the species variables; (ii) a choice is made concerning the distance measure to be used as the basis of the analysis (e.g. Bray–Curtis, Euclidean, χ^2 or other measure); (iii) ordination (and/or clustering) is performed in order to visualize patterns of resemblance among the observations based on their community composition; and (iv) a non-parametric multivariate test for differences among groups is done to obtain a rigorous probabilistic statement concerning multivariate effects of *a priori* groups. Note that (iii) is not essential in terms of the statistical test; ordination simply gives a visual representation by reducing the dimensionality of the data. In this paper, I focus on step (iv) of this procedure, which currently poses a problem for multifactorial designs.

DESCRIPTION OF THE TEST: ONE-WAY DESIGN

The two essential considerations for the test are: (i) the construction of the test-statistic, and (ii) the calculation

of a P-value using some method of permutation. I will describe the method, which I shall simply call non-parametric MANOVA, first for the one-way design and then for more complex designs, followed by some ecological examples. I deal here only with the case of balanced ANOVA designs, but analogous statistics for any linear model, including multiple regression and/or unbalanced data, can be constructed, as described by McArdle and Anderson (in press).

The test statistic: an *F*-ratio

The essence of analysis of variance is to compare variability within groups versus variability among different groups, using the ratio of the F-statistic. The larger the value of F, the more likely it is that the null hypothesis (H_0) of no differences among the group means (i.e. locations) is false. For univariate ANOVA, partitioning of the total sum of squares, SS_T, is achieved by calculating sums of squared differences (i) between individual replicates and their group mean (SS_W, the within-group sum of squares; Table 1a), and (ii) between group means and the overall sample mean (SS_A, the among-group sum of squares). Next, consider the multivariate case where p variables are measured simultaneously for each of n replicates in each of a groups, yielding a matrix of data where rows are observations and columns are variables. A natural multivariate analogue may be obtained by simply adding up the sums of squares across all variables (Table 1b). An F-ratio can then be constructed, as in the univariate case.

This multivariate analogue can also be thought of geometrically (e.g. Caliński & Harabasz 1974; Mielke *et al.* 1976; Edgington 1995; Pillar & Orlóci 1996), as shown in Fig. 1 for the case of two groups and two variables (dimensions). Here, SS_W is the sum of the squared Euclidean distances between each individual replicate and its group centroid (the point corresponding to the averages for each variable, Fig. 1 and
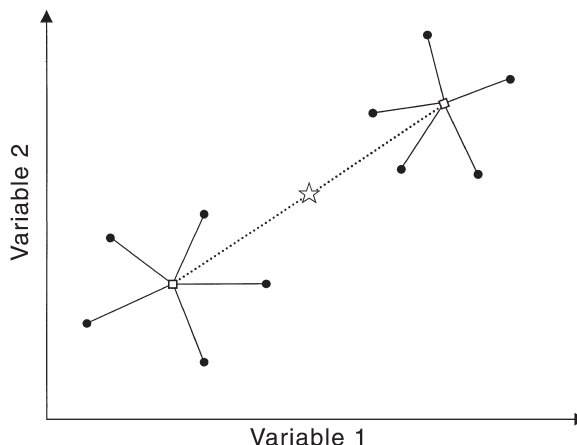

Fig. 1. A geometric representation of MANOVA for two groups in two dimensions where the groups differ in location. The within-group sum of squares is the sum of squared distances from individual replicates to their group centroid. The among-group sum of squares is the sum of squared distances from group centroids to the overall centroid. (——) Distances from points to group centroids; (·······) distances from group centroids to overall centroid; (☆), overall centroid; (□), group centroid; (●), individual observation.

Table 1. Calculations of within-group sums of squares for partitioning in (a) univariate ANOVA, (b) a multivariate analogue obtained by summing across variables, (c) a multivariate analogue equivalent to (b) obtained using sums of squared Euclidean distances, (d) the traditional MANOVA approach, which yields an entire matrix (W) of within-group sums of squares and cross products, and (e) the partitioning using inter-point distances advocated here, equivalent to (b) and (c) if Euclidean distances are used

Univariate	
(a) One variable	$SS_W = \Sigma_{i=1}^{a} \Sigma_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2$
Multivariate	
(b) Summed across variables	$SS_W = \Sigma_{i=1}^{a} \Sigma_{j=1}^{n} \Sigma_{k=1}^{p} (y_{ijk} - \bar{y}_{i.k})^2$
(c) Geometric approach (inner product, a scalar, based on Euclidean distances, correlations between variables ignored)	$SS_W = \Sigma_{i=1}^{a} \Sigma_{j=1}^{n} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^{\mathrm{T}} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})$
(d) Traditional MANOVA (outer product, a matrix, based on Euclidean distances, correlations between variables matter)	$\mathbf{W} = \Sigma_{i=1}^{a} \Sigma_{j=1}^{n} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.}) (\mathbf{y}_{ij} - \bar{\mathbf{y}}_{i.})^{\mathrm{T}}$
(e) Inter-point geometric approach (a scalar, based on any distance measure, correlations between variables ignored)	$SS_I = \frac{1}{n} \Sigma_{i=1}^{N-1} \Sigma_{j=i+1}^{N} d_{ij}^2 \epsilon_{ij}$

y_{ij}, univariate observation of the jth replicate ($j = 1,…, n$) in the ith group ($i = 1,…, a$); y_{ijk}, observation of y_{ij} for the kth variable ($k = 1,…, p$); $\mathbf{y}_{ij}$, vector of length p, indicating a point in multivariate space according to p variables (dimensions) for observation j in group i. A superscript 'T' indicates the transpose of the vector, bars over letters indicate averages and a dot subscript indicates averaging was done over that subscripted variable.

Table 1c). Note that this additive partitioning using a geometric approach yields one value for each of SS_W, SS_A and SS_T as sums of squared Euclidean distances. This geometric approach gives sums of squares equivalent to the sum of the univariate sums of squares (added across all variables) described in the previous paragraph. This differs from the traditional MANOVA approach, where partitioning is done for an entire matrix of sums of squares and cross-products (e.g. Mardia *et al.* 1979; Table 1d).

The key to the non-parametric method described here is that the sum of squared distances between points and their centroid is equal to (and can be calculated directly from) the sum of squared interpoint distances divided by the number of points. This important relationship is illustrated in Fig. 2 for points in two dimensions. The relationship between distances to centroids and interpoint distances for the Euclidean measure has been known for a long time (e.g. Kendall & Stuart 1963; Gower 1966; Caliński & Harabasz 1974; Seber 1984; Pillar & Orlóci 1996; Legendre & Legendre 1998; see also equation B.1 in Appendix B of Legendre & Anderson 1999). What is important is the implication this has for analyses based on non-Euclidean distances. Namely, an additive partitioning of sums of squares can be obtained for any distance measure directly from the distance matrix, without calculating the central locations of groups.

Why is this important? In the case of an analysis based on Euclidean distances, the average for each variable across the observations within a group constitutes the measure of central location for the group in Euclidean space, called a centroid. For many distance measures, however, the calculation of a central location may be problematic. For example, in the case of the semimetric Bray–Curtis measure, a simple average across replicates does not correspond to the 'central location' in multivariate Bray–Curtis space. An appropriate measure of central location on the basis of Bray–Curtis distances cannot be calculated easily directly from the data. This is why additive
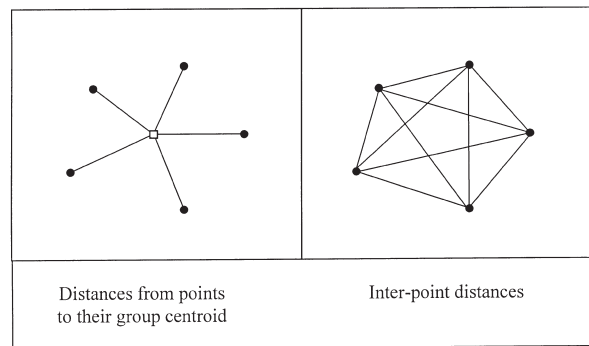

Fig. 2. The sum of squared distances from individual points to their centroid is equal to the sum of squared interpoint distances divided by the number of points.
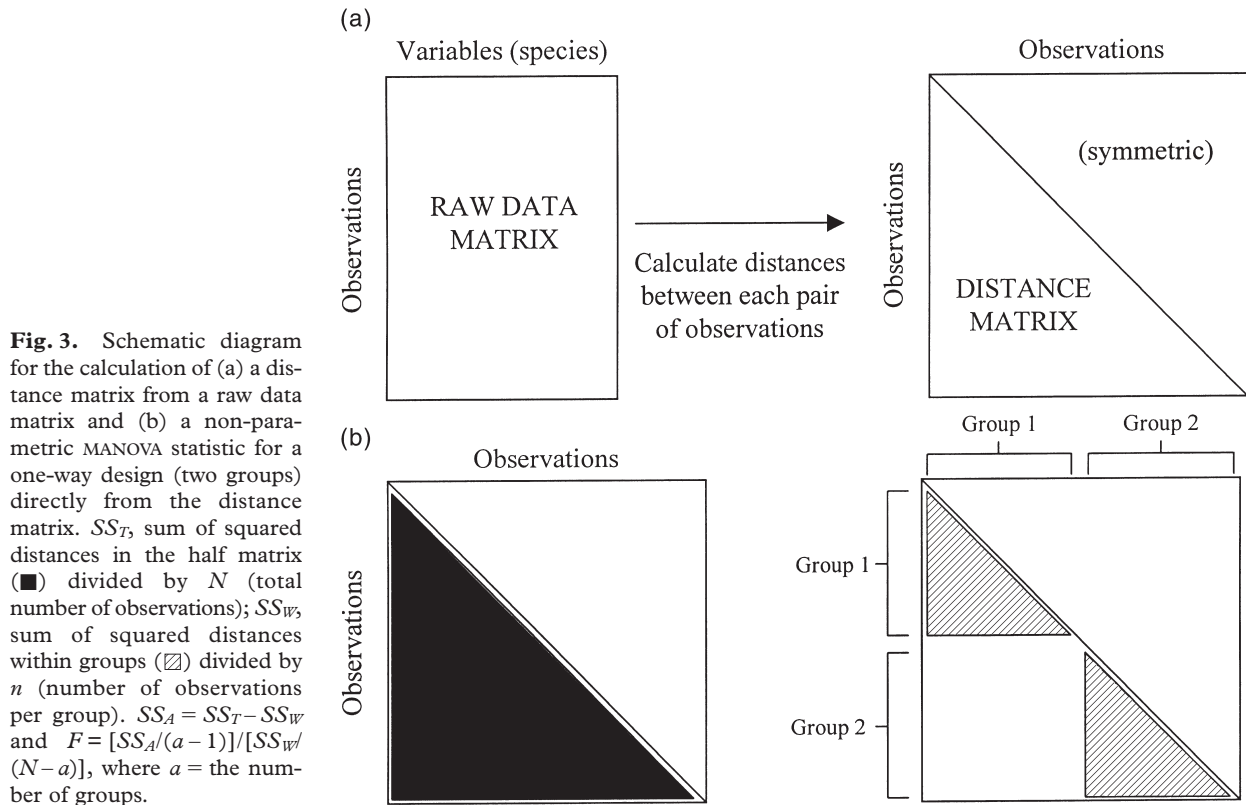

Fig. 3. Schematic diagram for the calculation of (a) a distance matrix from a raw data matrix and (b) a non-parametric MANOVA statistic for a one-way design (two groups) directly from the distance matrix. SS_T, sum of squared distances in the half matrix (■) divided by N (total number of observations); SS_W, sum of squared distances within groups (▨) divided by n (number of observations per group). $SS_A = SS_T - SS_W$ and $F = [SS_A/(a-1)]/[SS_W/(N-a)]$, where a = the number of groups.

partitioning (in terms of 'average' differences among groups) has not been previously achieved using Bray–Curtis (or other semimetric) distances. However, the relationship shown in Fig. 2 can be applied to achieve the partitioning directly from interpoint distances.

Thus, consider a matrix of distances between every pair of observations (Fig. 3a). If we let $N = an$, the total number of observations (points), and let d_{ij} be the distance between observation $i = 1,…, N$ and observation $j = 1,…, N$, the total sum of squares is

$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \qquad (1)$$

That is, add up the squares of all of the distances in the subdiagonal (or upper-diagonal) half of the distance matrix (not including the diagonal) and divide by N (Fig. 3b). In a similar fashion, the within-group or residual sum of squares is

$$SS_W = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \, \epsilon_{ij} \qquad (2)$$

where ϵ_{ij} takes the value 1 if observation i and observation j are in the same group, otherwise it takes the value of zero. That is, add up the squares of all of the distances between observations that occur in the same group and divide by n, the number of observations per group (Fig. 3b). Then $SS_A = SS_T - SS_W$ and a pseudo F-ratio to test the multivariate hypothesis is

$$F = \frac{SS_A/(a-1)}{SS_W/(N-a)} \qquad (3)$$

If the points from different groups have different central locations (centroids in the case of Euclidean distances) in multivariate space, then the among-group distances will be relatively large compared to the within-group distances, and the resulting pseudo F-ratio will be relatively large.

One can calculate the sums of squares in equations (1) and (2) and the statistic in equation (3) from a distance matrix obtained using *any* distance measure. The statistic in equation (3) corresponds exactly to the statistic in equation (4) of McArdle and Anderson (in press), who have shown more generally how partitioning for any linear model can be done directly from the distance matrix, regardless of the distance measure used. Another important aspect of the statistic described above is that, in the case of a Euclidean distance matrix calculated from only one variable, equation (3) gives the same value as the traditional parametric univariate F-statistic.

This is proposed as a new non-parametric MANOVA statistic that is intuitively appealing, due to its analogy with univariate ANOVA, and that is extremely relevant for ecological applications. The results (in terms of sums of squares, mean squares and pseudo F-ratios) obtained for individual terms in a multivariate analysis

can be interpreted in the same way as they usually are for univariate ANOVA. The difference is that the hypothesis being tested for any particular term is a multivariate hypothesis.

OBTAINING A *P*-VALUE USING PERMUTATIONS

The multivariate version of the F-statistic described here is not distributed like Fisher's F-ratio under the null hypothesis. This is so because (i) we do not expect the individual variables to be normally distributed, and (ii) we do not expect that the Euclidean distance will necessarily be used for the analysis. Even if each of the variables were normally distributed and the Euclidean distance used, the mean squares calculated for the multivariate data would not each consist of sums of independent χ^2 variables, because, although individual observations are expected to be independent, individual species variables are not independent of one another. Thus, traditional tabled P-values cannot be used. A distribution of the statistic under the null hypothesis can be created, however, using permutations of the observations (e.g. Edgington 1995; Manly 1997). The only situation in which one could use the traditional tabled probabilities would be if one had a single variable that could reasonably be assumed to be normally distributed and one used Euclidean distances for the analysis.

Suppose the null hypothesis is true and the groups are not really different (in terms of their composition and/or their relative abundances of species, as measured by the Bray–Curtis distances). If this were the case, then the multivariate observations (rows) would be exchangeable among the different groups. Thus, the labels on the rows that identify them as belonging to a particular group could be randomly shuffled (permuted) and a new value of F obtained (called, say, F^π). This random shuffling and re-calculation of F^π is then repeated for all possible re-orderings of the rows relative to the labels. This gives the entire distribution of the pseudo F-statistic under a true null hypothesis for our particular data. Comparing the value of F obtained with the original ordering of the rows to the distribution created for a true null by permuting the labels, a P-value is calculated as

$$P = \frac{(\text{No. of } F^\pi \geq F)}{(\text{Total no. of } F^\pi)} \qquad (4)$$

Note that we consider the original observed value of F to be a member of the distribution of F^π under permutation (i.e. it is one of the possible orderings of the labels on the rows). The usual scientific convention of an *a priori* significance level of $\alpha = 0.05$ is generally used for interpreting the significance of the result, as in other statistical tests. It is also possible to view the

P-value as a measure of confidence concerning the null hypothesis (Fisher 1955; Freedman & Lane 1983).

With *a* groups and *n* replicates per group, the number of distinct possible outcomes for the *F*-statistic in a one-way test is $(an)!/(a!(n!)^a)$ (Clarke 1993). As it is usually not practical to calculate all possible permutations, because of the time involved, *P* can be calculated using a large random subset of all possible permutations (Hope 1968). However, the precision of the *P*-value will increase with increasing numbers of permutations. Generally, at least 1000 permutations should be done for tests with an α-level of 0.05 and at least 5000 permutations should be done for tests with an α-level of 0.01 (Manly 1997).

ASSUMPTIONS

The only assumption of the test is that the observations (rows of the original data matrix) are exchangeable under a true null hypothesis. To assume exchangeability under the null hypothesis is generally to assume that the observations are independent and that they have similar distributions (e.g. Boik 1987; Hayes 1996). By 'similar distributions', I mean similar multivariate dispersions of points, not that the points are necessarily multivariate normal. The test described here is a test for differences in location (means or centroids) among groups of multivariate observations based on the chosen distance measure. Like its univariate counterpart, which is sensitive to heterogeneity of variances, this test and its predecessors that use permutations, like ANOSIM (Clarke 1993), will also be sensitive to differences in the dispersions of points, even if the locations do not differ.

The sensitivity of ANOSIM to differences in dispersion has been suggested as an advantage by Clarke (1993). This is because it was introduced in the context of detecting environmental impacts, for which detection of differences of any kind between control and impacted locations is very important for environmental reasons. Here, I simply suggest that caution be exercised in interpreting the results of tests of significance. Determining if significant differences among groups may be due to differences in dispersion versus differences in location (or some combination of the two) is an important statistical and ecological issue. The use of permutation tests to obtain *P*-values does not avoid this issue.

A useful comparative index of multivariate dispersion has been given by Warwick and Clarke (1993). Also, a separate permutation test for significant differences in multivariate dispersions (after removing effects of differences in location), as an accompaniment to the non-parametric MANOVA approach given here, will be described elsewhere (Anderson, Dutilleul, Lapointe & Legendre, unpubl. data).

DISTINCTION FROM TRADITIONAL TEST STATISTICS

Although the statistic described here is sensitive to differences in the relative dispersion of points among groups, it takes no account of the correlations among variables. In traditional MANOVA, the test-statistics (such as Wilks' Lambda) use information contained in the between-group and/or within-group sample variance–covariance matrices (e.g. Table 1d, see Olson 1974; Johnson & Field 1993). The traditional MANOVA tests assume not only that the variance for each variable remains constant across different groups (i.e. the points in different groups have similar scatter), they also assume that the relationships among the variables (their covariances or correlations) do not differ across groups.

These differences in the sensitivities of different multivariate test statistics are shown diagrammatically in Fig. 4 for two variables (two dimensions). Figure 4(a) shows two groups that differ in their correlation structure, but not in their variances or location. Figure 4(b) shows two groups that differ only in their dispersions, but not in their correlation or location. Although all MANOVA statistics are designed to test for differences in location, the traditional statistics will also be sensitive to differences in correlations (Fig. 4a) as well as differences in dispersion (Fig. 4b). The method of non-parametric MANOVA described here will only be sensitive to differences in dispersion (Fig. 4b). The correlations among variables play no role in the analysis. If differences in the relationships amongst variables form a hypothesis of interest, then some other non-parametric techniques may be relevant (e.g. Biondini *et al.* 1991; Krzanowski 1993).

ONE-WAY EXAMPLE: EFFECTS OF GRAZERS

Consider the following example, taken from an ecological study by Anderson and Underwood (1997). The study was designed to test the hypothesis that grazing by gastropods affects intertidal estuarine assemblages. The experiment was done at an intertidal oyster farm from January to July 1994 in Quibray Bay, south of Sydney, New South Wales, Australia. Experimental surfaces (10 cm × 10 cm) were enclosed in cages to exclude gastropod grazers, while other surfaces were left open to grazing. A third treatment consisted of caged areas where natural densities of grazers were included. This was a control for the effect of the cage itself on assemblages. There were *n* = 20 surfaces in each of the three treatments. The numbers of individuals of each of 21 taxa (invertebrates and algae) colonizing each surface were recorded.

The rationale for increasing the severity of the transformation to increase the relative contribution of rare

versus abundant species in the analysis, given by Clarke and Green (1988), is followed here. Note that the transformation is not done in an effort to make data conform to any assumptions of the analysis. In this example, the data contained some species that occurred on a very large relative scale of abundance (e.g. Spirorbid worms occurred in the thousands), so the data were transformed by taking double-square roots before the analysis. To visualize the multivariate patterns among observations, non-metric multi-dimensional scaling (MDS) was performed on the Bray–Curtis distances (Kruskal & Wish 1978), using the PRIMER computer program. Non-parametric MANOVA was then done on Bray–Curtis distances, as described in the previous section, using the computer program NPMANOVA, written by the author in FORTRAN.
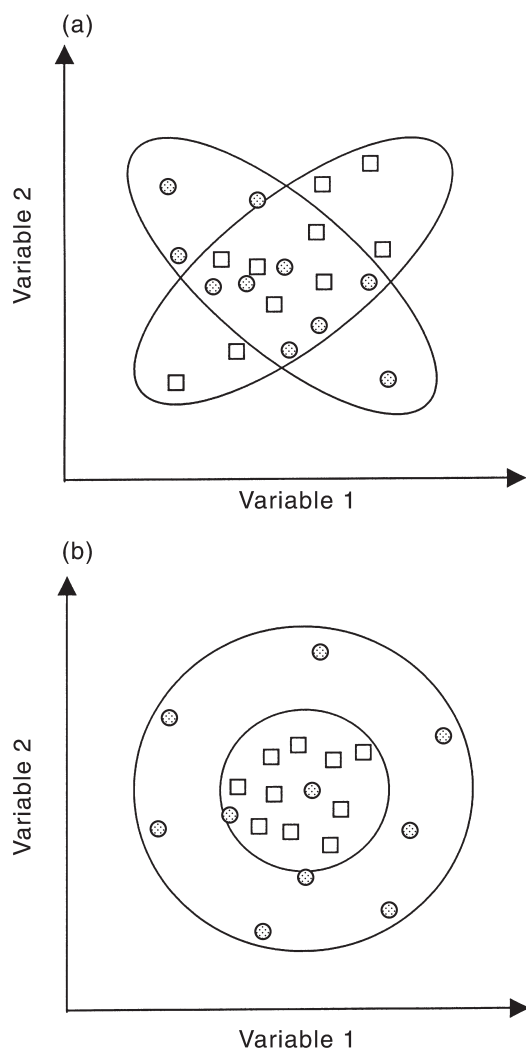
The number of possible permutations for the one-way test in the case of the grazing experiment is 9.6×10^{25}. A random subset of 4999 permutations was used (Fig. 5). In this case, the null hypothesis of no differences among groups was rejected, as the observed value was much larger than any of the values obtained under permutation (Fig. 5, Table 2).

A POSTERIORI TESTS

As in univariate ANOVA where there is a significant result in a comparison of 3 or more treatments, we may wish to ask for the multivariate case: wherein does the significant difference lie? This can be done by using the same test, given above for the one-way comparison of groups, but where individual pair-wise comparisons between particular groups are done. To continue with the logic of the analogous univariate situation, we can use a t-statistic (which is simply the square root of the value of the F-statistic described above) for these

(a)

(b)

Fig. 4.   Two variables in each of two groups of observations where (a) the groups differ in correlation between variables, but not in location or dispersion and (b) the groups differ in dispersion, but not in location or correlation between variables.

Table 2.   Non-parametric MANOVA on Bray–Curtis distances for assemblages of organisms colonizing intertidal surfaces in estuaries in three grazing treatments (grazers excluded, grazers inside cages, and surfaces open to grazers)

Source	d.f.	SS	MS	F	P
Grazers	2	18 657.65	9328.83	36.61	0.0002
Residual	57	14 520.89	254.75		
Total	59	33 178.54			

Comparison*		t		P
Open versus caged		8.071		0.0002
Open versus cage control		3.268		0.0002
Caged versus cage control		6.110		0.0002

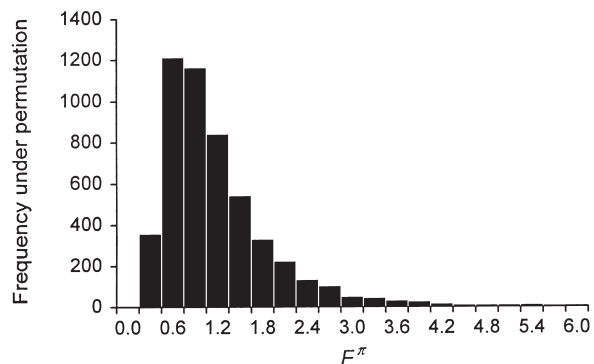*Pair-wise *a posteriori* tests among grazing treatments.

Fig. 5.   Distribution of the non-parametric MANOVA F-statistic for 4999 permutations of the data on assemblages in different grazing treatments. The real value of F for these data is very extreme by reference to this distribution ($F = 36.62$): thus there are strong differences among the assemblages in different grazing treatments.

pairwise comparisons. These have the same interpretation as univariate t-tests, but they test the general multivariate hypothesis of no difference between the groups on the basis of the Bray–Curtis (or other chosen) distances. This is Student's univariate t-statistic if Euclidean distances are chosen for the analysis of only one variable. P-values for each test are obtained using separate sets of permutations that are only done across the pair of groups being compared. In this example, a random subset of 4999 permutations was used (out of a possible 6.8×10^9) for each pairwise comparison.

For the analysis of the experimental removal of grazers, there was a significant difference among all pairs of treatments: assemblages colonizing surfaces in cages differed from those in the open or in cage controls ($t = 8.07$, $P = 0.0002$ and $t = 6.11$, $P = 0.0002$, respectively). Grazers had a significant effect on the assemblages, which is consistent with the pattern of separation of points corresponding to different treatments in the non-metric MDS plot (Fig. 6). The fact that assemblages on open surfaces also differed significantly from those in cage controls ($t = 3.27$, $P = 0.0002$) suggested that there was some additional artifact due to the presence of a cage in the experiment.

An important point here is that the *a posteriori* comparisons just described did not make any correction for experiment-wise error rate (Day & Quinn 1989). Similarly, the multivariate pair-wise tests available in the computer program NPMANOVA are not corrected for experiment-wise error rate. This means that with an *a priori* significance level of $\alpha = 0.05$, one should expect to obtain a significant result in one out of every 20
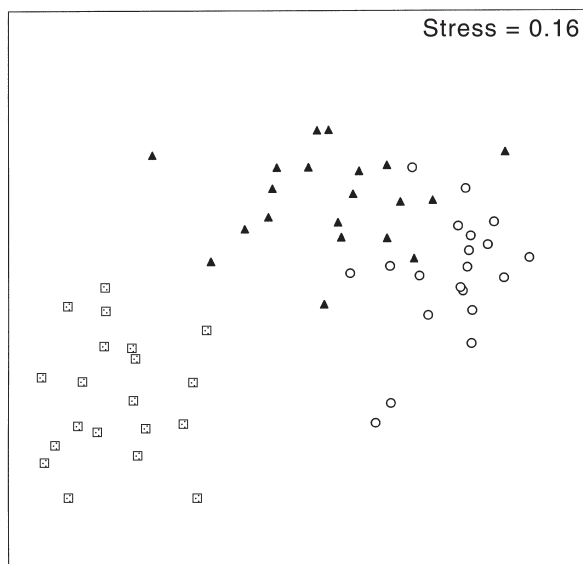

Fig. 6. Non-metric MDS plot of assemblages colonizing intertidal surfaces in Quibray Bay in each of three different grazing treatments: (▲), cage control; (○), open; (⊡) caged.

independent tests by chance alone. Nevertheless, the P-value obtained under permutation for any individual pair-wise test is exact. Many of the methods used for correcting error rates for multiple comparisons, such as the Bonferroni method, are very conservative but may be applied.

MORE COMPLEX DESIGNS

For more complex designs, we can start by considering the situation with two factors in a factorial (or orthogonal) design. The principles used to partition the variation directly from the distance matrix and to obtain the statistics and permutation methods for individual terms in the model can be readily generalized to other multifactorial cases. The logic applied to multifactorial ANOVA of univariate data (e.g. see Underwood 1981, 1997) also applies to the analysis of multivariate data using this non-parametric procedure. For example, tests of main effects should be examined after tests for interactions.

Calculating the statistic

Let A designate factor 1 with a levels (treatments or groups) and B designate factor 2 with b levels, with n replicates in each of the ab combinations of the two factors. The total number of observations is $N = abn$. The total sum of squares in the analysis is calculated as for the one-way case according to equation (1). To partition the variation, the within-group sum of squares for factor A, ignoring any influence of B, is calculated as

$$SS_{W(A)} = \frac{1}{bn} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \epsilon_{ij}^{(A)} \qquad (5)$$

where $\epsilon_{ij}^{(A)}$ takes the value 1 if observation i and observation j are in the same group of factor A, otherwise it takes the value of zero. Similarly, the within-group sum of squares for factor B, ignoring any influence of A, is

$$SS_{W(B)} = \frac{1}{an} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \epsilon_{ij}^{(B)} \qquad (6)$$

Then, the corresponding sums of squares for each of the main effects in the analysis are $SS_A = SS_T - SS_{W(A)}$ and $SS_B = SS_T - SS_{W(B)}$.

The residual sum of squares is calculated by considering the interpoint distances within each of the ab combinations of factor A and B, thus:

$$SS_R = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_{ij}^2 \epsilon_{ij}^{(AB)} \qquad (7)$$

where $\epsilon_{ij}^{(AB)}$ takes the value 1 if observation i and observation j are in the same combination of factors A and B, otherwise it takes the value of zero. We then can

easily obtain the sum of squares corresponding to the interaction term: $SS_{AB} = SS_T - SS_A - SS_B - SS_R$. It may be easier to consider the squared distances being summed in equations (5) through (7) by reference to their physical location in the distance matrix itself, as illustrated in Fig. 7.

In the case of a two-factor design where one factor is nested in the other, the same general approach is used. In this case, however, there is no interaction term in the analysis and we have instead $SS_{B(A)} = SS_T - SS_A - SS_R$, where $B(A)$ denotes that factor B is nested in factor A.

Having obtained appropriate sums of squares, the construction of the pseudo F-statistic for each term in the analysis for non-parametric MANOVA then follows the same rules and formulae as it would for the usual univariate ANOVA. The construction of the F-ratio depends on the experimental design, that is, whether factors are nested or factorial and whether they are fixed or random, exactly as for univariate ANOVA (e.g. Underwood 1981, 1997; Winer *et al.* 1991).

Doing the permutations

The method of permutation required to obtain an exact test is not so simple if there is more than one factor in the design. The choice of an appropriate permutation method is not trivial and should be considered carefully for each term in the model. Indeed, the lack of exact tests or knowledge of how the available approximate permutation tests might behave for complex models has been a sticking point in the development of multivariate non-parametric methods (e.g. Crowley 1992; Clarke 1993). To construct exact tests, two important issues must be considered (Anderson & ter Braak, unpublished data). First, which units should be permuted (i.e. what are exchangeable under the null hypothesis) and second, should any restrictions be imposed on the permutations to account for other factors in the design?

In many important situations, such as tests of interactions, no exact permutation test can be done. Also, there are times when the exact test imposes so many
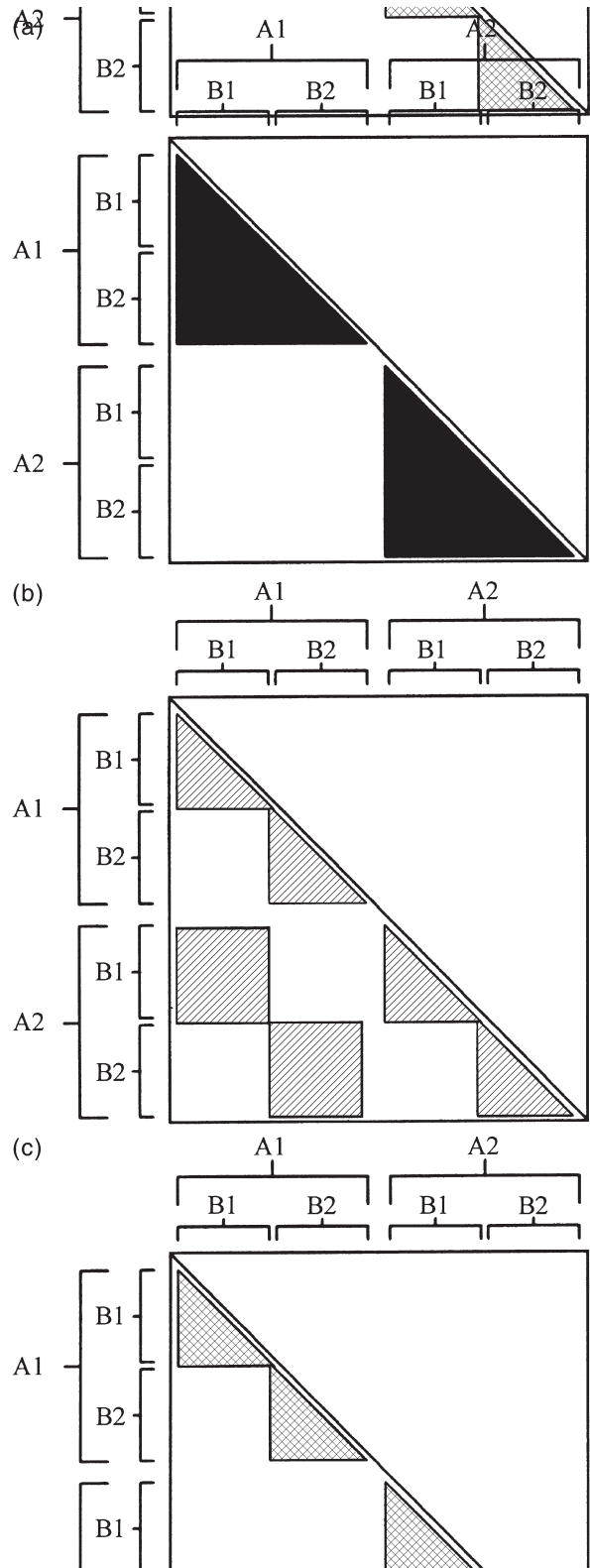

Fig. 7. Schematic diagram of the interpoint distances used to partition the variability in the multivariate data set and to calculate the sum of squares for each term in a two-factor orthogonal design (each factor has two groups or levels). (a) $SS_{W(A)}$ = sum of squared distances within groups of A (■), divided by (bn). (b) $SS_{W(B)}$ = sum of squared distances within groups of B (▨), divided by (an). (c) SS_R = sum of squared distances within combinations of AB (▨), divided by (n) (residual sum of squares). SS_T = sum of squared distances in the total half matrix, divided by (abn), $SS_A = SS_T - SS_{W(A)}$, $SS_B = SS_T - SS_{W(B)}$, $SS_{AB} = SS_T - SS_A - SS_B - SS_R$.

restrictions as to render the test meaningless, due to there being too few possible permutations left. In these cases, approximate permutation tests should be used, of which there are several alternatives, including permutation of residuals and permutation of raw data across all terms in the analysis (e.g. Freedman & Lane 1983; ter Braak 1992; Manly 1997). Some empirical comparisons of these methods are provided by Gonzalez and Manly (1998) and Anderson and Legendre (1999).

ECOLOGICAL EXAMPLES

Two-way factorial design

The first example is from an experiment in Middle Harbour (part of Sydney Harbour) to test for the effect of shade and proximity to the seafloor on assemblages of invertebrates and algae colonizing subtidal hard surfaces near marinas (Glasby 1999). The experiment was a two-way crossed (orthogonal) design with $n = 4$ replicate settlement plates (15 cm × 15 cm sandstone) either far from or near to the seafloor (the factor of 'position'; all plates were at a similar depth of approximately 2 m below low water) and in one of three shading treatments: (i) shade (an opaque plexiglass roof), (ii) a procedural control (a clear plexiglass roof), and (iii) no shade. Organisms colonizing the plates after 33 weeks were counted and a total of 46 taxa were included in the analyses. Organisms that occurred only once across the entire data set were not included. Non-parametric MANOVA was done on Bray–Curtis distances calculated from double-root transformed data using the FORTRAN program NPMANOVA. The sample size was reasonably small for this study ($n < 5$), so the

test was done using unrestricted permutation of raw data (e.g. Manly 1997; Gonzalez & Manly 1998) with 4999 random permutations. Similar results were obtained using permutation of residuals under a reduced model (not shown).

There was no significant interaction of shade and position, but both main effects were significant (Table 3, Fig. 8). It was then of interest to compare the groups corresponding to different shading treatments using *a posteriori* tests (Table 3). It was not necessary to do this for the effect of position, because this factor only had two groups. Assemblages of organisms on settlement plates near the bottom were extremely different from those far away from the bottom (Fig. 8). Also, assemblages on shaded plates were significantly different from those on either the procedural control or on unshaded plates, which themselves did not differ (Table 3, Fig. 8). This analysis also shows how the effect of position relative to the bottom was much greater than the effect of shading on assemblages in this experiment (compare the values of their mean squares in Table 3). The non-parametric approach advocated here allows tests of significance, but it also allows relative sizes of effects to be compared directly through the partitioning of the variation and examination of mean squares.

Three-way design, including nesting

The second example comes from an experiment to test the hypothesis that the size of a patch available for

Table 3. Non-parametric MANOVA on Bray–Curtis distances for assemblages of organisms colonizing subtidal sandstone settlement panels after 33 weeks in an estuary at different distances from the seafloor (positions either near or far) and in three different shading treatments

Source	d.f.	SS	MS	F	P
Position	1	5595.40	5595.40	13.536	0.0002
Shade	2	3566.44	1783.22	4.314	0.0006
Position × shade	2	1238.94	619.47	1.499	0.1394
Residual	18	7440.66	413.37		
Total	23	17 841.43			

Comparison*		t	P
Shade versus control		1.783	0.0154
Shade versus no shade		1.987	0.0018
Control versus no shade		0.866	0.5560

*Pair-wise *a posteriori* tests among shading treatments.
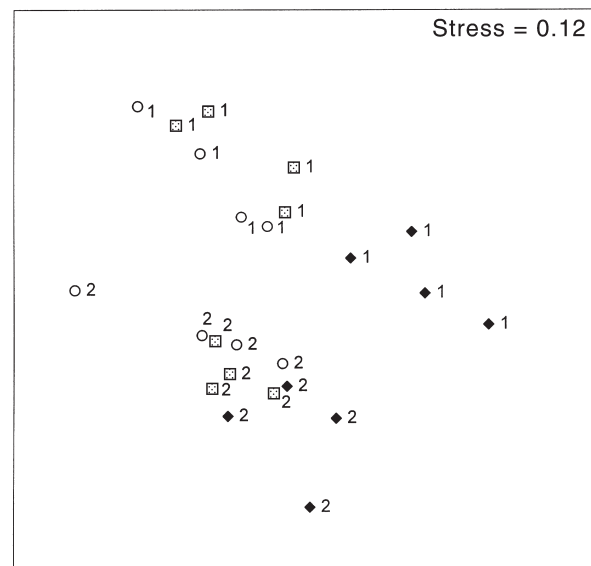

Fig. 8. Two-factor non-metric MDS plot of subtidal assemblages colonizing sandstone settlement plates after 33 weeks in Middle Harbour that were either near to or far from the seafloor and in one of three shading treatments. (◆), Shaded; (⊡), control; (○), no shade; 1, far from the sea floor; 2, close to the sea floor.

colonization would affect the succession of assemblages in equal areas on those patches (Anderson 1998). This hypothesis was tested using wooden panels of three different sizes (10 cm × 10 cm, 20 cm × 20 cm and 40 cm × 40 cm). Two panels of a given size were attached to sticks that were then strapped to the structure of an intertidal oyster farm in the Port Stephens estuary in New South Wales, Australia in January of 1995 (see Anderson 1998 for details). Six sticks (two for each of the three patch sizes) were then collected independently after periods of 3, 6, 9, 12 and 18 months of exposure to colonization. The experimental design thus consisted of three factors: time (5 periods of submersion), patch (3 sizes) and sticks (2 sticks per time × patch combination, a random nested factor), with $n = 2$ panels per stick. Organisms colonizing panels were quantified in a 10 cm × 10 cm area from each panel (chosen randomly from the larger-sized panels). A total of 33 taxa were included in multivariate analyses.

The analysis was done using NPMANOVA on Bray–Curtis distances calculated on double-root transformed data, as for the previous examples. In this case, however, an exact permutation test for the nested factor (sticks) was carried out by permuting the observations randomly across sticks, but only within the 5 × 3 combinations of levels of time × patch. Then, for the test of the upper-level terms (the main effects of time, patch size and their interaction), individual replicates on a stick were permuted together as a unit (i.e. whole sticks were permuted). This is done so that the upper-level effects can be tested against the variability across sticks, not across individual replicates, as is necessary under the null hypothesis for a nested hierarchy (e.g. Clarke 1993). For all tests, a subset of 4999 permutations was used.

The nested factor of sticks was not significant in the multivariate analysis, but the time × patch size interaction was significant (Table 4). Individual pair-wise comparisons of effects of patch size for each time were nearly identical to the one-way tests given in Anderson (1998) using analysis of similarities (ANOSIM, Clarke 1993). Assemblages were significantly different on the smallest patches compared to the other sized patches after 3, 6, 9 or 12 months. Assemblages on the two larger sized patches did not differ significantly from one another except after 12 months. After 18 months, assemblages were similar on all patch sizes.

Non-metric MDS plots helped to interpret these results. Two separate ordinations were done on these data, as the stress value for the non-metric MDS plot that included all observations was too high for a reasonable interpretation. The effect of different patch sizes appeared to be fairly consistent (in terms of its magnitude and direction) after 3, 6 or 9 months (Fig. 9a). After 12 or 18 months, the observations become more scattered within and across the groups

and the effects of patch size become less clear (Fig. 9b). Increased dispersion (variability in assemblages) after these longer periods of time, compared to earlier periods, is seen clearly in the two-factor plot of stick centroids (Fig. 9c, which includes all data). As noted earlier, the tests are sensitive to such differences in dispersion. Overall, although the two factors did interact, the effect of time (i.e. succession) was relatively more important in distinguishing assemblages than the size of the patch (compare their mean squares in Table 4), and effects of patch size decreased through time for these assemblages (see Anderson 1998 for further details).

DISCUSSION

Natural temporal and spatial variability is intrinsic to ecological systems. Indeed, variability might be considered the currency of ecological scientific work. It is for this reason that statistical analysis plays such an important role in the development of ecology as a science. In *Design of Experiments*, R. A. Fisher (1935, p. 4) wrote:

> We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of uncertainty may itself be capable of rigorous expression.

Quantitative statistical inference is indeed what is needed for the rigorous interpretation of mensurative or manipulative ecological experiments. Although our conclusions may be uncertain, they are still rigorous in the sense that the degree of uncertainty can be expressed in terms of mathematical probability. In univariate analysis, W. S. Gosset ('Student' 1908) made this possible for comparisons of two treatments, while R. A. Fisher made this possible for many treatments and experimental factors. In a complex and intrinsically variable world, ANOVA allows us to identify simultaneous effects and interactions of more than one factor, and to identify the uncertainty of our inferences with rigour (e.g. Underwood 1981, 1997).

An important advance in the analysis of multivariate data in ecology was the development of non-parametric methods for testing hypotheses concerning whole communities (e.g. Clarke 1988, 1993; Smith *et al.* 1990; Biondini *et al.* 1991). Some parallel advances were made in the context of tests for significant clusters in cluster analysis (e.g. Good 1982; Gordon 1994). Before these applications, particularly ANOSIM (Clarke 1993), became widely available, most multivariate analyses in ecology focused on the reduction of dimensionality to produce and interpret patterns (ordination
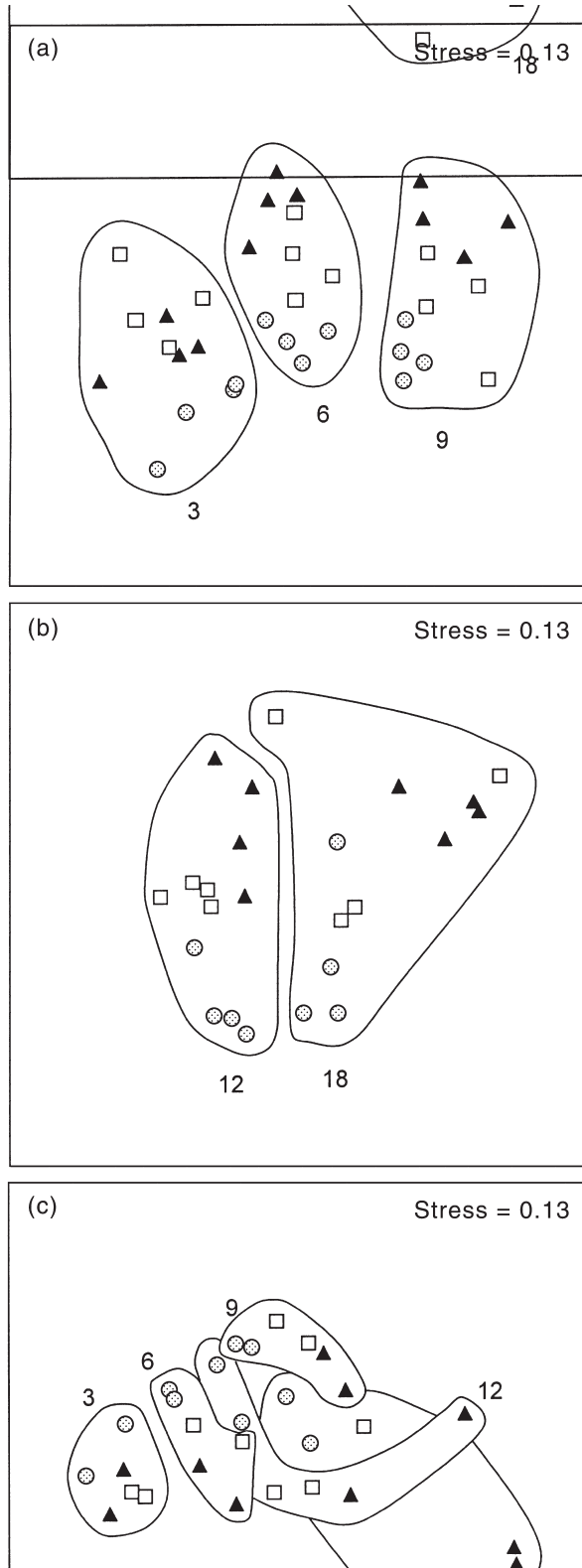
methods) and the use of numerical strategies for placing observations into natural groups (clustering). These methods, although extremely useful towards their purpose, do not rigorously express the nature and degree of uncertainty concerning *a priori* hypotheses. Methods like Mantel's test (Mantel 1967), ANOSIM (Clarke 1993) and multiresponse permutation procedures (Mielke *et al.* 1976) allow such rigorous probabilistic statements to be made for multivariate ecological data.

The drawback to such non-parametric tests is that they cannot easily be extended to the multifactorial designs so common now in ecological studies. Two sticking points prevented this: (i) the lack of a generalized statistic for partitioning variation, and (ii) the lack of appropriate permutation methods (e.g. Clarke 1993; Legendre & Anderson 1999). Although traditional test-statistics used for MANOVA allow partitioning, their restrictive assumptions have prevented their effective use in ecology. The method of distance-based redundancy analysis (Legendre & Anderson

Table 4. Non-parametric MANOVA on Bray–Curtis distances for assemblages of organisms colonizing wooden settlement panels of three different sizes after 3, 6, 9, 12 or 18 months on an intertidal oyster farm

Source	d.f.	*SS*	*MS*	*F*	*P*
Time	4	30 305.71	7576.43	20.50	0.0002
Patch Size	2	6414.99	3207.49	8.68	0.0002
Time × patch	8	6224.03	778.004	2.10	0.0062
Sticks (time × patch)	15	5544.66	369.64	1.28	0.3384
Residual	30	8697.09	289.90		
Total	59	57 186.48			

Comparison	3 months	6 months	9 months	12 months	18 months
Small versus medium	2.24★	1.86★	1.48★	2.00★	1.38
Small versus large	2.30★	2.87★	2.38★	2.54★	3.02★
Medium versus large	1.49	1.47	1.69	1.75★	1.43

★$P < 0.05$; pair-wise *a posteriori* tests among patch sizes within each time using the *t*-statistic. Sticks were ignored in the pair-wise tests. There were 35 possible permutations for each.

Fig. 9. Two-factor non-metric MDS plots of assemblages colonizing intertidal wooden settlement panels of three different sizes for (a) periods of 3, 6 or 9 months and (b) periods of 12 or 18 months for raw data and (c) where centroids were plotted corresponding to each stick across all combinations of time × patch size. The points corresponding to assemblages after particular times of submersion (numbers indicate the period in months) have been outlined for clarity. (●), Small; (□), medium; (▲), large.

1999) largely solved these issues, but it has many rather complicated steps and involves the use of a correction constant to distances. Although this correction does not adversely affect the test (generally making it more conservative, if anything), accurate *P*-values are not given by this method in the case of anything other than a one-factor design (McArdle & Anderson, in press).

The method presented here has, in some sense, combined the best of both worlds. Like the traditional test-statistics, it can partition variation according to any ANOVA design. Like the most flexible non-parametric methods, it can be based on any symmetric dissimilarity or distance measure (or their ranks) and provides a *P*-value using appropriate permutation methods. That is, one can still choose a relevant transformation and an appropriate distance measure (or use ranks of distances), consistent with the method of ordination used to visualize patterns. By using permutations, the test requires no specific assumption concerning the number of variables or the nature of their individual distributions or correlations. The statistic used is analogous to Fisher's *F*-ratio and is constructed from sums of squared distances (or dissimilarities) within and between groups. Another feature of this statistic is that it is equal to to Fisher's original *F*-ratio in the case of one variable and when Euclidean distances are used.

It is perhaps important to point out that the Bray–Curtis measure of dissimilarity may or may not be the most appropriate measure to use in any given situation. A point still commonly ignored is that Bray–Curtis and related measures, such as Kulczynski's coefficient, will tend to under-estimate true ecological distances when distances become large (i.e. when observations have very few species in common), as determined by simulations (Faith *et al.* 1987; Belbin 1991). The Bray–Curtis measure may therefore only be useful insofar as it will produce reasonable ecological ordinations, through the ranks used for MDS. In light of this issue, Faith (1990) proposed doing MANOVA on ordination scores obtained from MDS.

An analysis based only on ordination scores will omit some portion of the ecological information contained in the original distance matrix and will depend on the number of dimensions chosen for the MDS (Clarke 1993). The severity of this potential problem will obviously increase with increases in the stress value associated with an MDS plot. The actual amount and the kind of information lost in reducing dimensions using MDS are generally unknown and will depend on the particular data set. Also, any subsequent statistical inferences on ordination scores (using traditional MANOVA, as described in Faith 1990; or based on permutation tests) would be limited to points in the ordination plots, rather than being applicable to the original observations.

It is not possible to identify a single 'best' distance measure for ecological data. The method described here may be used with any distance measure chosen (or on ranks of distances). It is useful to have the flexibility to choose a distance measure appropriate for the data and hypothesis being tested. Although the Bray–Curtis measure has now become commonplace in ecological studies (perhaps due to its availability in the PRIMER computer program, or due to its intuitive interpretation as 'percentage difference', or due to the results presented in Faith *et al.* 1987), there are still many rivals. Over 60 measures of similarity or dissimilarity have been described, with very few actual comparisons of their performance with different kinds of ecological data (see Lamont & Grant 1979; Legendre & Legendre 1998). This is undoubtedly an area needing further research.

The approach advocated here is that multifactorial analysis of variance, as successfully applied to univariate data in ecology (e.g. Underwood 1981, 1997), can and should also be applied to multivariate data for testing hypotheses in a logical and rigorous way. It stands to reason that the developments in experimental design for ecology that require multifactorial ANOVA, in order to, for example, avoid pseudo-replication (Hurlbert 1984), test for generality (Beck 1997) or test for environmental impact (e.g. Green 1979, 1993; Underwood 1993; Glasby 1997), should be incorporated into multivariate analysis. The method described here allows that to happen, but within a framework that is general enough to suit our need for few assumptions and flexibility in the multivariate analysis of ecological data.

ACKNOWLEDGEMENTS

REFERENCES

Anderson M. J. (1998) Effects of patch size on colonisation in estuaries: revisiting the species-area relationship. *Oecologia* **118,** 87–98.

Anderson M. J. & Clements A. (in press) Resolving environmental disputes: a statistical method for choosing among competing cluster models. *Ecol. Applic.*

Anderson M. J. & Legendre P. (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J. Stat. Comput. Simul.* **62,** 271–303.

Anderson M. J. & Underwood A. J. (1994) Effects of substratum on the recruitment and development of an intertidal estuarine fouling assemblage. *J. Exp. Mar. Biol. Ecol.* **184,** 217–36.

Anderson M. J. & Underwood A. J. (1997) Effects of gastropod grazers on recruitment and succession of an estuarine assemblage: a multivariate and univariate approach. *Oecologia* **109,** 442–53.

Bartlett M. S. (1939) A note on tests of significance in multivariate analysis. *Proc. Camb. Philos. Soc.* **35,** 180–5.

Beck M. W. (1997) Inference and generality in ecology: current problems and an experimental solution. *Oikos* **78,** 265–73.

Belbin L. (1991) Semi-strong hybrid scaling, a new ordination algorithm. *J. Veg. Sci.* **2,** 491–6.

Biondini M. E., Mielke P. W. & Redente E. F. (1991) Permutation techniques based on Euclidean analysis spaces: a new and powerful statistical method for ecological research. In: *Computer Assisted Vegetation Analysis* (eds E. Feoli & L. Orlóci) pp. 221–40. Kluwer Academic Publishers, Dordrecht.

Boik R. J. (1987) The Fisher–Pitman permutation test: a non-robust alternative to the normal theory F test when variances are heterogeneous. *Br. J. Math. Stat. Psychol.* **40,** 26–42.

Bray J. R. & Curtis J. T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27,** 325–49.

Caliñski T. & Harabasz J. (1974) A dendrite method for cluster analysis. *Commun. Stat.* **3,** 1–27.

Chapman M. G., Underwood A. J. & Skilleter G. A. (1995) Variability at different spatial scales between a subtidal assemblage exposed to discharge of sewage and two control sites. *J. Exp. Mar. Biol. Ecol.* **189,** 103–22.

Clarke K. R. (1988) Detecting change in benthic community structure. In: *Proceedings XIVth International Biometric Conference, Namur: Invited Papers*, pp. 131–42. Société Adolphe Quetelet, Gembloux.

Clarke K. R. (1993) Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18,** 117–43.

Clarke K. R. & Green R. H. (1988) Statistical design and analysis for a 'biological effects' study. *Mar. Ecol. Prog. Ser.* **46,** 213–26.

Crowley P. H. (1992) Resampling methods for computation-intensive data analysis in ecology and evolution. *Ann. Rev. Ecol. Syst.* **23,** 405–47.

Day R. W. & Quinn G. P. (1989) Comparison of treatments after an analysis of variance. *Ecol. Monogr.* **59,** 433–63.

Edgington E. S. (1995) *Randomization Tests,* 3rd edn. Marcel Dekker, New York.

Excoffier L., Smouse P. E. & Quattro J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131,** 479–91.

Faith D. P. (1990) Multivariate methods for biological monitoring based on community structure. In: *The Australian Society of Limnology 29th Congress*, p. 17 (Abstract). Alligator Rivers Region Research Institute.

Faith D. P., Dostine P. L. & Humphrey C. L. (1995) Detection of mining impacts on aquatic macroinvertebrate communities: results of a disturbance experiment and the design of a multivariate BACIP monitoring program at Coronation Hill, N. T. *Aust. J. Ecol.* **20,** 167–80.

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

Fisher R. A. (1935) *Design of Experiments.* Oliver & Boyd, Edinburgh.

Fisher R. A. (1936) The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7,** 179–88.

Fisher R. A. (1955) Statistical methods and scientific induction. *J. Roy. Stat. Soc.* **17,** 69–78.

Freedman D. & Lane D. (1983) A nonstochastic interpretation of reported significance levels. *J. Bus. Econ. Stat.* **1,** 292–8.

Gaston K. J. & McArdle B. H. (1994) The temporal variability of animal abundances: measures, methods and patterns. *Phil. Trans. Roy. Soc. Lond.* **345,** 335–58.

Glasby T. M. (1997) Analysing data from post-impact studies using asymmetrical analyses of variance: a case study of epibiota on marinas. *Aust. J. Ecol.* **22,** 448–59.

Glasby T. M. (1999) Interactive effects of shading and proximity to the seafloor on the development of subtidal epibiotic assemblages. *Mar. Ecol. Prog. Ser.* **190,** 113–24.

Gonzalez L. & Manly B. F. J. (1998) Analysis of variance by randomization with small data sets. *Environmetrics* **9,** 53–65.

Good I. J. (1982) An index of separateness of clusters and a permutation test for its significance. *J. Stat. Comput. Simul.* **15,** 81–4.

Gordon A. D. (1994) Identifying genuine clusters in a classification. *Comput. Stat. Data Anal.* **18,** 561–81.

Gower J. C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53,** 325–38.

Gray J. S., Clarke K. R., Warwick R. M. & Hobbs G. (1990) Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. *Mar. Ecol. Prog. Ser.* **66,** 285–99.

Green R. H. (1979) *Sampling Design and Statistical Methods for Environmental Biologists*. Wiley, New York.

Green R. H. (1993) Application of repeated measures designs in environmental impact and monitoring studies. *Aust. J. Ecol.* **18,** 81–98.

Hajdu L. J. (1981) Graphical comparison of resemblance measures in phytosociology. *Vegetatio* **48,** 47–59.

Hayes A. F. (1996) Permutation test is not distribution free. *Psychol. Methods* **1,** 184–98.

Hope A. C. A. (1968) A simplified Monte Carlo significance test procedure. *J. Roy. Stat. Soc.* **30,** 582–98.

Hotelling H. (1931) The generalization of Student's ratio. *Ann. Math. Stat.* **2,** 360–78.

Hubert L. & Schultz J. (1976) Quadratic assignment as a general data analysis strategy. *Br. J. Math. Stat. Psychol.* **29,** 190–241.

Hurlbert S. H. (1984) Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* **54,** 187–211.

Johnson C. R. & Field C. A. (1993) Using fixed-effects model multivariate analysis of variance in marine biology and ecology. *Oceanogr. Mar. Biol. Ann. Rev.* **31,** 177–221.

Kelaher B. P., Chapman M. G. & Underwood A. J. (1998) Changes in benthic assemblages near boardwalks in

temperate urban mangrove forests. *J. Exp. Mar. Biol. Ecol.* **228,** 291–307.

Kendall M. G. & Stuart A. (1963) *The Advanced Theory of Statistics*, Vol. 1, 2nd edn. Charles Griffith, London.

Kruskal J. B. & Wish M. (1978) *Multidimensional Scaling*. Sage Publications, Beverly Hills.

Krzanowski W. J. (1993) Permutational tests for correlation matrices. *Statistics and Computing* **3,** 37–44.

Kulczynski S. (1928) Die Pflanzenassoziationen der Pieninen. *Bull. Int. Acad. Pol. Sci. Lett. Cl. Sci. Math. Nat. Ser.* B, (Suppl. II) **1927,** 57–203.

Lamont B. B. & Grant K. J. (1979) A comparison of twenty-one measures of site dissimilarity. In: *Multivariate Methods in Ecological Work* (eds L. Orloci, C. R. Rao & W. M. Stiteler) pp. 101–26. International Co-operative Publishing House, Fairland.

Lawley D. N. (1939) A generalization of Fisher's *IX* test. *Biometrika* **30,** 180–7 (Corrections in *Biometrika* **30,** 467–9).

Legendre P. & Anderson M. J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* **69,** 1–24.

Legendre P. & Legendre L. (1998) *Numerical Ecology*, 2nd English edn. Elsevier Science, Amsterdam.

McArdle B. H. & Anderson M. J. (in press) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*.

Manly B. F. J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology,* 2nd edn. Chapman & Hall, London.

Mantel N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27,** 209–20.

Mantel N. & Valand R. S. (1970) A technique of nonparametric multivariate analysis. *Biometrics* **26,** 547–58.

Mardia K. V. (1971) The effect of non-normality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika* **58,** 105–21.

Mardia K. V., Kent J. T. & Bibby J. M. (1979) *Multivariate Analysis*. Academic Press, London.

Mielke P. W., Berry K. J. & Johnson E. S. (1976) Multi-response permutation procedures for a priori classifications. *Commun. Stat. Theory Methods* **5** (14)**,** 1409–24.

Odum E. P. (1950) Bird populations of the Highlands (North Carolina) Plateau in relation to plant succession and avian invasion. *Ecology* **31,** 587–605.

Oliver I. & Beattie A. J. (1996) Designing a cost-effective invertebrate survey: a test of methods for rapid assessment of biodiversity. *Ecol. App.* **6,** 594–607.

Olson C. L. (1974) Comparative robustness of six tests in multivariate analysis of variance. *J. Am. Stat. Assoc.* **69,** 894–908.

Pillai K. C. S. (1955) Some new test criteria in multivariate analysis. *Ann. Math. Stat.* **26,** 117–21.

Pillar V. D. P. & Orlóci L. (1996) On randomization testing in vegetation science: multifactor comparisons of relevé groups. *J. Veg. Sci.* **7,** 585–92.

Quinn G. P., Lake P. S. & Schreiber S. G. (1996) Littoral benthos of a Victorian lake and its outlet stream: spatial and temporal variation. *Aust. J. Ecol.* **21,** 292–301.

Seber G. A. F. (1984) *Multivariate Observations*. John Wiley and Sons, New York.

Skilleter G. A. (1996) Validation of rapid assessment of damage in urban mangrove forests and relationships with Molluscan assemblages. *J. Mar. Biol. Ass. UK* **76,** 701–16.

Smith E. P., Pontasch K. W. & Cairns J. (1990) Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.* **24,** 507–14.

Student. (1908) The probable error of a mean. *Biometrika* **6,** 1–25.

ter Braak C. J. F. (1992) Permutation versus bootstrap significance tests in multiple regression and ANOVA. In: *Bootstrapping and Related Techniques* (eds K. H. Jöckel, G. Rothe & W. Sendler) pp. 79–86. Springer-Verlag, Berlin.

Underwood A. J. (1981) Techniques of analysis of variance in experimental marine biology and ecology. *Oceanogr. Mar. Biol. Ann. Rev.* **19,** 513–605.

Underwood A. J. (1993) The mechanics of spatially replicated sampling programmes to detect environmental impacts in a variable world. *Aust. J. Ecol.* **18,** 99–116.

Underwood A. J. (1997) *Experiments in Ecology: Their Logical Design and Interpretation Using Analysis of Variance*. Cambridge University Press, Cambridge.

Underwood A. J. & Chapman M. G. (1998) A method for analysing spatial scales of variation in composition of assemblages. *Oecologia* **107,** 570–8.

Warwick R. M., Carr M. R., Clarke K. R., Gee J. M. & Green R. H. (1988) A mesocosm experiment on the effects of hydrocarbon and copper pollution on a sublittoral soft-sediment meiobenthic community. *Mar. Ecol. Prog. Ser.* **46,** 181–91.

Warwick R. M. & Clarke K. R. (1993) Increased variability as a symptom of stress in marine communities. *J. Exp. Mar. Biol. Ecol.* **172,** 215–26.

Wilks S. S. (1932) Certain generalizations in the analysis of variance. *Biometrika* **24,** 471–94.

Winer B. J., Broan D. R. & Michels K. M. (1991) *Statistical Principles in Experimental Design*, 3rd edn. McGraw-Hill, Sydney.