CHANGE IN MARINE COMMUNITIES:

An Approach to Statistical Analysis and Interpretation

2nd edition

K R Clarke & R M Warwick

Plymouth Marine Laboratory, UK

Published 2001
by
PRIMER-E Ltd

Registered office: Plymouth Marine Laboratory
Prospect Place
West Hoe
Plymouth PL1 3DH
United Kingdom

Business Office: 3 Meadow View
Lutton
Ivybridge
Devon PL21 9RH
United Kingdom

Directors: Prof K Robert Clarke MSc PhD Raymond N Gorley MA

First edition 1994
Second edition 2001

Printed by Bretonside Copy, Plymouth, UK (www.bretonsidecopy.com)

Clarke, K.R., Warwick, R.M., 2001
Change in marine communities: an approach to statistical analysis and interpretation, 2nd edition. PRIMER-E: Plymouth

© Copyright 2001 PRIMER-E Ltd, all rights reserved

CONTENTS

INTRODUCTION

CHAPTER 1 A framework for studying changes in community structure

CHAPTER 2 Measures of similarity of species abundance/biomass between samples

CHAPTER 3 Hierarchical clustering

CHAPTER 4 Ordination of samples by Principal Components Analysis (PCA)

CHAPTER 5 Ordination of samples by Multi-Dimensional Scaling (MDS)

CHAPTER 6 Testing for differences between groups of samples

CHAPTER 7 Species analyses

CHAPTER 8 Diversity measures, dominance curves and other graphical analyses

CHAPTER 9 Transformations

CHAPTER 10 Species removal and aggregation

CHAPTER 11 Linking community analyses to environmental variables

CHAPTER 12 Causality: community experiments in the field and laboratory

CHAPTER 13 Data requirements for biological effects studies: which components and attributes of the marine biota to examine?

CHAPTER 14 Relative sensitivities and merits of univariate, graphical/distributional and multivariate techniques

CHAPTER 15 Multivariate measures of community stress

CHAPTER 16 Further comparison of multivariate patterns

CHAPTER 17 Biodiversity and dissimilarity measures based on relatedness of species

APPENDIX 1 Index of example data

APPENDIX 2 Principal literature sources and further reading

APPENDIX 3 References cited

CHAPTER 6: TESTING FOR DIFFERENCES BETWEEN GROUPS OF SAMPLES

Many community data sets possess some *a priori* defined structure within the set of samples, for example there may be replicates from a number of different sites (and/or times). A pre-requisite to interpreting community differences between sites should be a demonstration that there *are* statistically significant differences to interpret.

UNIVARIATE TESTS

When the species abundance (or biomass) information in a sample is reduced to a single index, such as Shannon diversity (see Chapter 8), the existence of replicate samples from each of the groups (sites/times etc.) allows formal statistical treatment by analysis of variance (ANOVA). This requires the assumption that the univariate index is normally distributed and has constant variance across the groups, conditions which are normally not difficult to justify (perhaps after transformation, see Chapter 9). A so-called ***global test*** of the ***null hypothesis*** (H_o), that there are no differences between groups, involves computing a particular ratio of variability in the group means to variability among replicates within each group. The resulting ***F statistic*** takes values near 1 if the null hypothesis is true, larger values indicating that H_o is false; standard tables of the *F* distribution yield a significance level *(p)* for the observed *F* statistic. Roughly speaking, *p* is interpreted as the probability that the group means we have observed (or a set of means which appear to differ from each other to an even greater extent) could have occurred if the null hypothesis H_o is actually true.

Fig.6.1 and Table 6.1 provide an illustration, for the 6 sites and 4 replicates per site of the Frierfjord macrofauna samples. The mean Shannon diversity for the 6 sites is seen in Fig.6.1, and Table 6.1 shows that the *F* ratio is sufficiently high that the probability of observing means as disparate as this by chance is $p<0.001$ (or $p<0.1\%$), if the true mean diversity at all sites is the same. This is deemed to be a sufficiently unlikely chance event that the null hypothesis can safely be rejected. Convention dictates that values of $p<5\%$ are sufficiently small, *in a single test*, to discount the possibility that H_o is true, but there is nothing sacrosanct about this figure: clearly, values of $p = 4\%$ and 6% should elicit the same inference. It is also clear

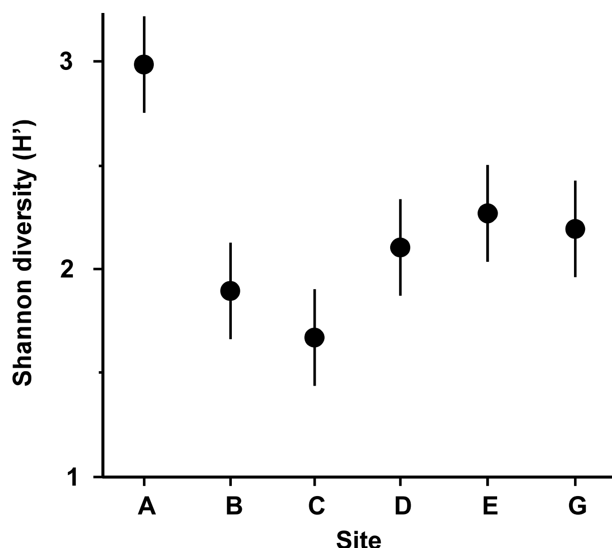

Fig. 6.1. Frierfjord macrofauna {F}. *Means and 95% confidence intervals of Shannon diversity (H') at the 6 field sites (A-E, G) shown in Fig. 1.1.*

that repeated significance tests, each of which has (say) a 5% possibility of describing a chance event as a real difference, will cumulatively run a much greater risk of drawing at least one false inference. This is one of the (many) reasons why it is not usually appropriate to handle a multi-species matrix by performing an ANOVA on each species in turn. (More decisive reasons are the complexities of dependence between species and the general inappropriateness of normality assumptions).

Fig. 6.1 shows the main difference to be a higher diversity at the outer site, A. The intervals displayed are ***95% confidence intervals*** for the true mean diversity at each site; note that these are of equal width because they are based on the assumption of constant variance, that is, they use a pooled estimate of replication variability from the residual mean square in the ANOVA table.

Table 6.1. Frierfjord macrofauna {F}. *ANOVA table showing rejection (at a significance level of 0.1%) of the global hypothesis of "no site-to-site differences" in Shannon diversity (H').*

	Sum of squares	Deg. of freedom	Mean Square	F ratio	Sig. level
Sites	*3.938*	*5*	*0.788*	*15.1*	*<0.1%*
Residual	*0.937*	*18*	*0.052*		
Total	*4.874*	*23*			

Further details of how confidence intervals are determined, why the ANOVA F ratio and F tables are defined in the way they are, how one can allow to some extent for the repeated significance tests in pairwise comparisons of site means etc, are not pursued here. This is the ground of basic statistics, covered by many standard texts, for example Sokal and Rohlf (1981), and such computations are available in all general-purpose statistics packages. This is not to imply that these concepts are elementary; in fact it is ironic that a proper understanding of *why* the univariate F test works requires a level of mathematical sophistication that is not needed for the simple permutation approach to the analogous global test for differences in *multivariate* structure between groups, outlined below.

MULTIVARIATE TESTS

One important feature of the multivariate analyses described in earlier chapters is that they in no way utilise any known structure among the samples, e.g. their division into replicates within groups. (This is in contrast with Canonical Variate Analysis, for example, which deliberately seeks out ordination axes that, in a certain well-defined sense, best separate out the known groups; e.g. Mardia *et al*, 1979). Thus, the ordination and dendrogram of Fig 6.2, for the Frierfjord macrofauna data, are constructed only from the pairwise similarities among the 24 samples, treated simply as numbers 1 to 24. By superimposing the group (site) labels A to G on the respective replicates it becomes immediately apparent that, for example, the 4 replicates from the outer site (A) are quite different in community composition from both the mid-fjord sites B, C and D and the inner sites E and G. A statistical test of the hypothesis that there are no site-to-site differences overall is clearly unnecessary, though it is less clear whether sufficient evidence exists to assert that B, C and D differ.

This simple structure of groups, and replicates within groups, is referred to as a ***1-way layout***, and it was seen above that 1-way ANOVA would provide the appropriate testing framework *if* the data were univariate (e.g. diversity or total abundance across all species). There *is* an analogous multivariate analysis of variance (MANOVA, e.g. Mardia *et al*, 1979), in which the F test is replaced by a test known as Wilks' Λ, but its assumptions will never be satisfied for typical multi-species abundance (or biomass) data. This is the problem referred to in the earlier chapters on choosing similarities and ordination methods; there are typically many more species (variables) than samples and the probability distribution of counts could never be reduced to approximate (multivariate) normality, by any transformation, because of the dominance of zero values. For example, for the Frierfjord data, as many as 50% of the entries in the species/samples matrix are zero, even after reducing the matrix to only the 30 most abundant species!

A valid test can instead be built on a simple non-parametric permutation procedure, applied to the (rank) similarity matrix underlying the ordination or classification of samples, and therefore termed an

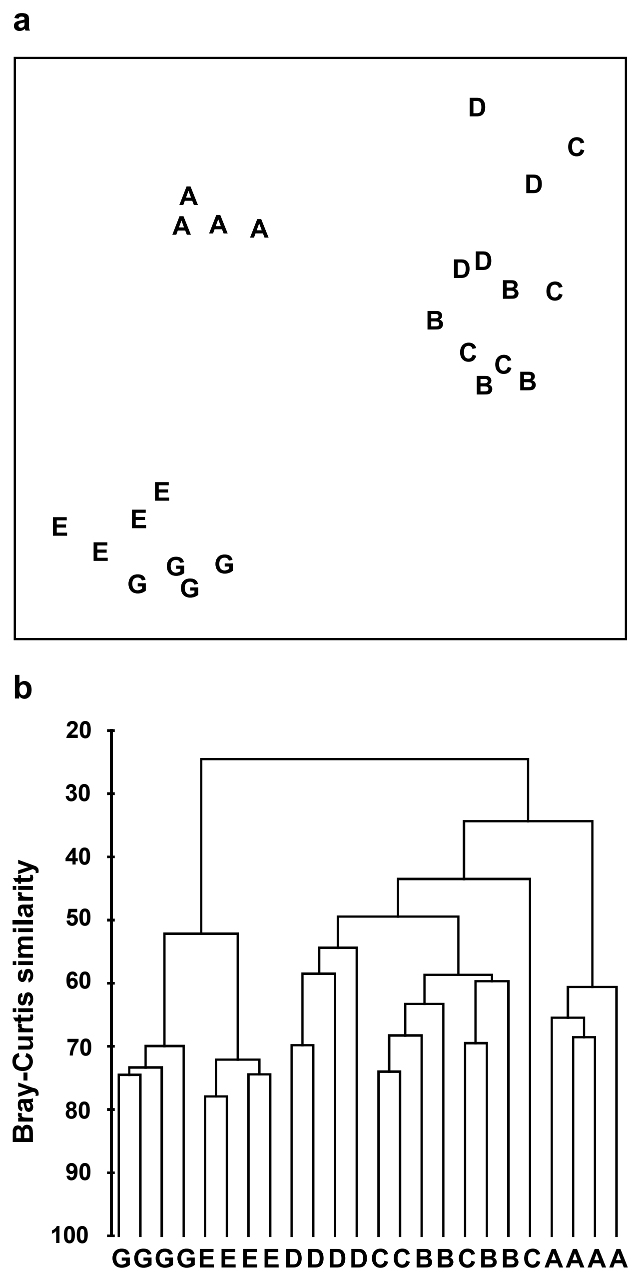

Fig. 6.2 Frierfjord macrofauna {F}*. a) MDS plot, b) dendrogram, for 4 replicates from each of the 6 sites (A-E and G), from Bray Curtis similarities computed for √√-transformed species abundances (MDS stress = 0.05).*

ANOSIM test (***analysis of similarities***)[¶], by analogy with the acronym ANOVA (analysis of variance). The history of such permutation tests dates back to the epidemiological work of Mantel (1967), and this is combined with a general randomization approach to the generation of significance levels (***Monte Carlo tests***, Hope 1968). In the context below, it was described by Clarke and Green (1988).

'ANOSIM' FOR THE 1-WAY LAYOUT

Fig.6.3 displays the MDS based only on the 12 samples (4 replicates per site) from the B, C and D sites of the Frierfjord macrofauna data. The null hypothesis (H_o) is that there are no differences in community composition at these 3 sites. In order to examine H_o, there are 3 main steps:

1) ***Compute a test statistic*** reflecting the observed differences *between* sites, contrasted with differences among replicates *within* sites. Using the MDS plot of Fig. 6.3, a natural choice might be to calculate the average distance between every pair of replicates within a site, and contrast this with the average distance apart of all pairs of samples corresponding to replicates from different sites. A test could certainly be constructed from these distances but it would have a number of drawbacks.

a) Such a statistic could only apply to a situation in which the method of display was an MDS rather than, say, a cluster analysis.

b) The result would depend on whether the MDS was constructed in two, three or higher dimensions. There is often no "correct" dimensionality and one may end up viewing the picture in several different dimensions – it would be unsatisfactory to generate different test statistics in this way.

c) The configuration of B, C and D replicates in Fig. 6.3 also differs slightly from that in Fig. 6.2a, which includes the full set of sites A-E, G. It is again undesirable that a test statistic for comparing *only* B, C and D should depend on which other sites are included in the picture.

These three difficulties disappear if the test is based not on distances between samples in an MDS but on the corresponding rank similarities between samples

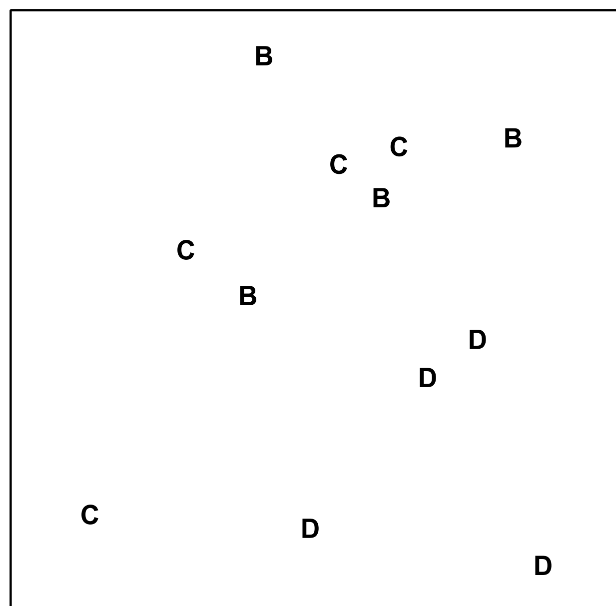

Fig. 6.3. Frierfjord macrofauna {F}*. MDS ordination as for Fig. 6.2 but computed only from the similarities involving sites B, C and D (stress = 0.11).*

in the underlying triangular similarity matrix. If $\bar{r}_W$ is defined as the average of all rank similarities among replicates *within* sites, and $\bar{r}_B$ is the average of rank similarities arising from all pairs of replicates *between* different sites, then a suitable test statistic is

$$R = \frac{(\bar{r}_B - \bar{r}_W)}{\frac{1}{2}M} \qquad (6.1)$$

where $M = n(n-1)/2$ and n is the total number of samples under consideration. Note that the highest similarity corresponds to a rank of 1 (the lowest value), following the usual mathematical convention for assigning ranks.

The denominator constant in equation (6.1) has been chosen so that:

a) R can never technically lie outside the range (-1,1);

b) R = 1 only if *all* replicates within sites are more similar to each other than *any* replicates from different sites;

c) R is approximately zero if the null hypothesis is true, so that similarities between and within sites will be the same on average.

R will usually fall between 0 and 1, indicating some degree of discrimination between the sites. R substantially less than zero is unlikely since it would correspond to similarities across different sites being *higher* than those within sites; such an occurrence is more likely

[¶] *The PRIMER ANOSIM routine covers tests for replicates from 1-way and 2-way (nested or crossed) layouts; the ANOSIM2 routine tackles the special case of a 2-way layout with no replication, which needs a modified style of test described at the end of this chapter.*

to indicate an incorrect labelling of samples.[†] The R statistic itself is a useful comparative measure of the degree of separation of sites, and its value is at least as important as its statistical significance (arguably more so). As with standard univariate tests, it is perfectly possible for R to be significantly different from zero yet inconsequentially small, if there are many replicates at each site.

2) ***Recompute the statistic under permutations*** of the sample labels. Under the null hypothesis H_o: "no difference between sites", there will be little effect on average to the value of R if the labels identifying which replicates belong to which sites are arbitrarily rearranged; the 12 samples of Fig. 6.3 are just replicates from a single site if H_o is true. This is the rationale for a ***permutation test*** of H_o; all possible allocations of four B, four C and four D labels to the 12 samples are examined and the R statistic recalculated for each. In general there are

$$(kn)!/[(n!)^k k!] \qquad (6.2)$$

distinct ways of permuting the labels for n replicates at each of k sites, giving 5775 permutations here. It is computationally possible to examine this number of re-labellings but the scale of calculation can quickly get out of hand with modest increases in replication, so the full set of permutations is randomly sampled (usually with replacement) to give the null distribution of R. In other words, the labels in Fig. 6.3 are randomly reshuffled, R recalculated and the process repeated a large number of times (T).

3) ***Calculate the significance level*** by referring the observed value of R to its permutation distribution. *If* H_o is true, the likely spread of values of R is given by the random rearrangements, so that if the *true* value of R looks unlikely to have come from this distribution there is evidence to reject the null hypothesis. Formally, if only t of the T simulated values of R are as large (or larger than) the *observed* R then H_o can be rejected at a ***significance level*** of $(t+1)/(T+1)$, or in percentage terms, $100(t+1)/(T+1)\%$.

[†] *Chapman and Underwood (1999) point out some situations in which negative R values (though not necessarily significantly negative) do occur in practice, when the community is species-poor and individuals have a heavily clustered spatial distribution, so that variability within a group is extreme. It usually also requires a design failure, e.g. a major stratifying factor (a differing substrate, say) is encompassed within each group but its effect is ignored in the analysis.*

EXAMPLE: Frierfjord macrofauna

The rank similarities underlying Fig. 6.3 are shown in Table 6.2 (note that these are the similarities involving only sites B, C and D, extracted from the matrix for *all* sites and re-ranked). Averaging across the 3 diagonal sub-matrices (within groups B, C and D) gives $\bar{r}_W = 22.7$, and across the remaining (off-diagonal) entries gives $\bar{r}_B = 37.5$. Also $n = 12$ and $M = 66$, so that $R = 0.45$. In contrast, the spread of R values possible from random re-labelling of the 12 samples can be seen in the histogram of Fig. 6.4: the largest of $T = 999$ simulations is less than 0.45 ($t = 0$). An observed value of $R = 0.45$ is seen to be a most unlikely event, with a probability of less than 1 in a 1000 if H_o is true, and we can therefore reject H_o at a significance level of $p<0.1\%$ (at least, because $R = 0.45$ may still have been the most extreme outcome observed had we chosen an even larger number of simulations).

Table 6.2. ***Frierfjord macrofauna {F}***. *Rank similarity matrix for the 4 replicates from each of B, C and D, i.e. C3 and C4 are the most, and B1 and C1 the least, similar samples.*

	B1	B2	B3	B4	C1	C2	C3	C4	D1	D2	D3	D4
B1	–											
B2	*33*	–										
B3	*8*	*7*	–									
B4	*22*	*11*	*19*	–								
C1	*66*	*30*	*58*	*65*	–							
C2	*44*	*3*	*15*	*28*	*29*	–						
C3	*23*	*16*	*5*	*38*	*57*	*6*	–					
C4	*9*	*34*	*4*	*32*	*61*	*10*	*1*	–				
D1	*48*	*17*	*42*	*56*	*37*	*55*	*51*	*62*	–			
D2	*14*	*20*	*24*	*39*	*52*	*46*	*35*	*36*	*21*	–		
D3	*59*	*49*	*50*	*64*	*54*	*53*	*63*	*60*	*43*	*41*	–	
D4	*40*	*12*	*18*	*45*	*47*	*27*	*26*	*31*	*25*	*2*	*13*	–

Pairwise tests

The above is a ***global*** test, indicating that there are site differences *somewhere* that may be worth examining further. Specific pairs of sites can then be compared: for example, the similarities involving only sites B and C are extracted, re-ranked and the test procedure repeated, giving an R value of 0.23. This time there are only 35 distinct relabellings so, under the null hypothesis H_o that sites B and C do not differ, the full permutation distribution of possible values of R can be computed; 12% of these values are equal to or larger than 0.23 so H_o cannot be rejected. By contrast, $R = 0.54$ for the comparison of B against D, which is the most extreme value possible under the 35 permutations.

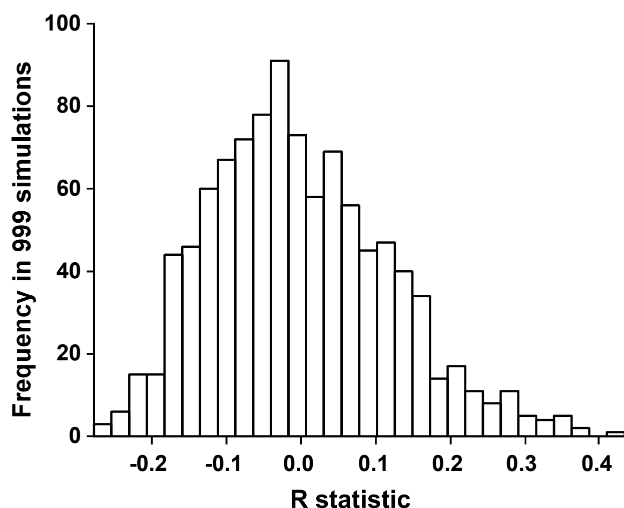

Fig. 6.4. Frierfjord macrofauna {F}. *Simulated distribution of the test statistic R (equation 6.1) under the null hypothesis of 'no site differences'; this contrasts with an observed value for R of 0.45.*

B and D are therefore inferred to differ significantly at the $p< 3\%$ level. For C against D, R = 0.57 similarly leads to rejection of the null hypothesis ($p<3\%$).

There is a danger in such repeated significance tests which should be noted (although rather little can be done to ameliorate it here). To reject the null hypothesis at a significance level of 3% implies that a 3% risk is being run of drawing an incorrect conclusion (a ***Type I error*** in statistical terminology). If many such tests are performed this risk will cumulate. For example, all pairwise comparisons between 10 sites, each with 4 replicates (allowing 3% level tests at best), would involve 45 tests, and the overall risk of drawing at least one false conclusion is high. For the analogous pairwise comparisons following the global F test in a univariate ANOVA, there exist ***multiple comparison*** tests which attempt to adjust for this repetition of risk. One straightforward possibility, which could be carried over to the present multivariate test, is a ***Bonferroni correction***. In its simplest form, this demands that, if there are *n* pairwise comparisons in total, each test uses a significance level of $0.05/n$. The so-called ***experiment-wise*** Type I error, the overall probability of rejecting the null hypothesis at least once in the series of pairwise tests, when there are no genuine differences, is then kept to 0.05.

However, the difficulty with such a Bonferroni correction is clear from the above example: with only 4 replicates in each group, and thus only 35 possible permutations, a significance level of 0.05/3 (=1.7%) can never be achieved! It may be possible to plan for a modest improvement in the number of replicates: 5 replicates from each site would allow a 1% level test for a pairwise comparison, equation (6.2) showing that there are then 126 permutations, and two groups of 6 replicates would give close to a 0.2% level test. However, this may not be realistic in some practical contexts, or it may be inefficient to concentrate effort on too many replicates at one site, rather than (say) increase the spatial coverage of sites. Also, for a fixed number of replicates, a too demandingly low Type I error (significance level) will be at the expense of a greater risk of ***Type II error***, the probability of *not* detecting a difference when one genuinely exists.

Strategy for interpretation

The solution, as with all significance tests, is to treat them in a more pragmatic way, exercising due caution in interpretation certainly, but not allowing the formality of a test procedure for pairwise comparisons to interfere with the natural explanation of the group differences. Herein lies the real strength of defining a test statistic, such as *R*, which has an absolute interpretation of its value. This is in contrast to a standard ***Z-type statistic***, which typically divides an appropriate measure (taking the value zero under the null hypothesis) by its standard deviation, so that interpretation is limited purely to statistical significance of the departure from zero.

The recommended course of action, for a case such as the above Frierfjord data, is therefore always to carry out, and take totally seriously, the global ANOSIM test for overall differences between groups. Usually the total number of replicates, and thus possible permutations, is relatively large, and the test will be reliable and informative. If it is not significant, then generally no further interpretation is permissible. If it *is* significant, it is legitimate to ask where the main between-group differences have arisen. The best tool for this is an examination of the *R* value for each pairwise comparison: large values (close to unity) are indicative of complete separation of the groups, small values (close to zero) imply little or no segregation. If the MDS is of sufficiently low stress to give a reliable picture, then the relative group separations will also be evident from this.[¶] The *R* value itself is not unduly affected by the number of replicates in the two groups being compared; this is in stark contrast to its statistical significance, which is dominated by the group sizes (for large numbers of replicates, *R* values near zero

[¶] *But the comparison of ANOSIM R values is the more generally valid approach, e.g. when the two descriptions do not appear to be showing quite the same thing. Calculation of R is in no way dependent on whether the 2-dimensional approximation implicit in an MDS is satisfactory or not, since R is computed from the underlying, full-dimensional similarity matrix.*

could still be deemed "significant", and conversely, few replicates could lead to R values close to unity being classed as "non-significant").

The analogue of this approach in the univariate case (say in the comparison of species richness between sites) would be firstly to compute the global F test for the ANOVA. If this establishes that there are significant overall differences between sites, the size of the effects would be ascertained by examining the differences in mean values between each pair of sites, or equivalently, by simply looking at a plot of how the mean richness varies across sites (perhaps with the replicates also shown). It is then immediately apparent where the main differences lie, and the interpretation is a natural one, emphasising the important biological features (e.g. absolute loss in richness is 5, 10, 20 species, or relative loss is 5%, 10%, 20% of the species pool, or whatever), rather than putting the emphasis solely on significance levels in pairwise comparisons of means, which runs the risk of missing the main message altogether.

So, returning to the multivariate data of the above Frierfjord example, interpretation of the ANOSIM tests is seen to be straightforward: a significant level (p<0.1%) and a mid-range value of R (= 0.45) for the global test of sites B, C and D establishes that there are statistically significant differences between these sites. Similarly mid-range values of R (slightly higher, at 0.54 and 0.57) for the B v D and C v D comparisons, contrasted with a much lower value (of 0.27) for B v C, imply that the explanation for the global test result is that D differs from both B and C, but the latter sites are not distinguishable.

The above discussion has raised the issue of Type II error for an ANOSIM permutation test, and the complementary concept, that of the ***power*** of the test, namely the probability of detecting a difference between groups when one genuinely exists. Ideas of power are not easily examined for non-parametric procedures of this type, which make no distributional assumptions and for which it is difficult to specify a precise non-null hypothesis. All that can be obviously said in general is that power will improve with increasing replication, and some low levels of replication should be avoided altogether. For example, if comparing only two groups with a 1-way ANOSIM test, based on only 3 replicates for each group, then there are only 10 distinct permutations and a significance level better than 10% could never be attained. A test demanding a significance level of 5% would then have *no* power to detect a difference between the groups, however large that difference is!

Generality of application

It is evident that few, if any, assumptions are made about the data in constructing the 1-way ANOSIM test, and it is therefore very generally applicable. It is not restricted to Bray-Curtis similarities or even to similarities computed from species abundance data: it could provide a non-parametric alternative to Wilks' Λ test for data which *are* more nearly multivariate-normally distributed, e.g. for testing whether groups (sites or times) can be distinguished on the basis of their environmental data (see Chapter 11). The latter would involve computing a Euclidean distance matrix between samples (after suitable transformation of the environmental variables) and entering this as a dissimilarity matrix to the ANOSIM procedure. Clearly, if multivariate normality assumptions are genuinely justified then the ANOSIM test must lack sensitivity in comparison with standard MANOVA, but this would seem to be more than compensated for by its greater generality.

Note also that there is no restriction to a balanced number of replicates. Some groups could even have only one replicate provided enough replication exists in other groups to generate sufficient permutations for the global test (though there will be a sense in which the power of the test is compromised by a markedly unbalanced design, here as elsewhere). More usefully, note that no assumptions have been made about the variability of within-group replication needing to be similar for all groups. This is seen in the following example, for which the groups in the 1-way layout are not sites but samples from different years at a single site.

EXAMPLE: Indonesian reef-corals

Warwick *et al* (1990b) examined data from 10 replicate transects across a single coral-reef site in S. Tikus Island, Thousand Islands, Indonesia, for each of the six years 1981, 1983, 1984, 1985, 1987 and 1988. The community data are in the form of % cover of a transect by each of the 58 coral species identified, and the analysis used Bray-Curtis similarities on untransformed data to obtain the MDS of Fig. 6.5. There appears to be a strong change in community pattern between 1981 and 1983 (putatively linked to the 1982/3 El Niño) and this is confirmed by a 1-way ANOSIM test for these two years alone: R = 0.43 (p< 0.1%). Note that, though not really designed for this situation, the test is perfectly valid in the face of greater "variability" in 1983 than 1981; in fact it is mainly a change in variability rather than location in

the MDS plot that distinguishes the 1981 and 1983 groups (a point returned to in Chapter 15).¶ This is in contrast with the standard univariate ANOVA (or multivariate MANOVA) test, which will have no power to detect a variability change; indeed it is invalid without an assumption of approximately equal variances (or variance-covariance matrices) across the groups.

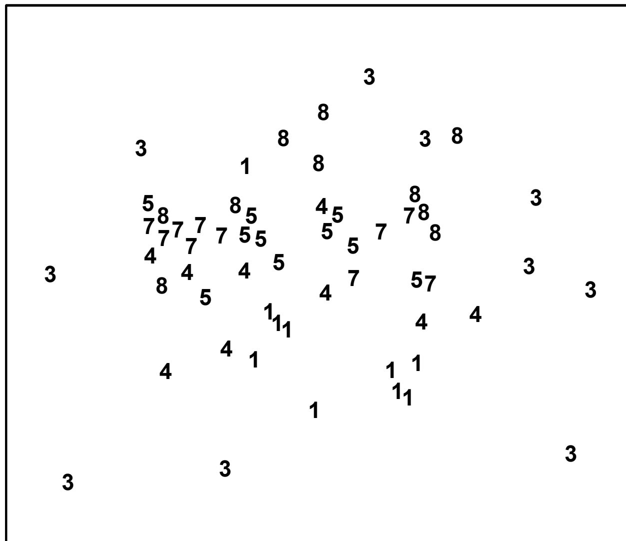

Fig. 6.5. Indonesian reef corals, S. Tikus Island {I}. *MDS of % species cover from 10 replicate transects in each of 6 years: 1 = 1981, 3 = 1983 etc (stress = 0.19).*

The basic 1-way ANOSIM test can also be extended to cater (to some degree) for more complex sample designs, as follows.

ANOSIM FOR TWO-WAY LAYOUTS

Three types of field and laboratory designs are considered here:

a) the ***2-way nested*** case can arise where two levels of spatial replication are involved, e.g. sites are grouped *a priori* to be representative of two "treatment" categories (control and polluted) but there are also replicate samples taken within sites;

b) the ***2-way crossed*** case can arise from studying a fixed set of sites at several times (with replicates at each site/time combination), or from an experimental study in which the same set of "treatments" (e.g. control and impact) are applied at a number of locations ("blocks"), for example in the different mesocosm basins of a laboratory experiment;

c) a 2-way crossed case ***with no replication*** of each treatment/block combination can also be catered for, to a limited extent, by a different style of permutation test.

The following examples of cases a) and b) are drawn from Clarke (1993) and the two examples of case c) are from Clarke and Warwick (1994).

EXAMPLE: Clyde nematodes (2-way nested case)

Lambshead (1986) analysed meiobenthic communities from three putatively polluted (P) areas of the Firth of Clyde and three control (C) sites, taking three replicate samples at each site (with one exception). The resulting MDS, based on fourth-root transformed abundances of the 113 species in the 16 samples, is given in Fig. 6.6a. The sites are numbered 1 to 3 for both conditions but the numbering is arbitrary – there is nothing in common between P1 and C1 (say). This is what is meant by sites being "nested" within conditions. Two hypotheses are then appropriate:

H1: there are no differences among sites *within* each "treatment" (control or polluted conditions);

H2: there are no differences *between* control and polluted conditions.

The approach to H2 might depend on the outcome of testing H1.

H1 can be examined by extending the 1-way ANOSIM test to a *constrained* randomisation procedure. The presumption under H1 is that there may be a difference between general location of C and P samples in the MDS plots but within each condition there cannot be any pattern in allocation of replicates to the three sites. Treating the two conditions entirely separately, one therefore has two separate 1-way permutation analyses of exactly the same type as for the Frierfjord macrofauna data (Fig. 6.3). These generate test statistics R_C and R_P, computed from equation (6.1), which can be combined to produce an average statistic $\overline{R}$. This can be tested by comparing it with $\overline{R}$ values from all possible permutations of sample labels permitted under the null hypothesis. This does not mean that all 16 sample labels may be arbitrarily permuted; the randomisation is constrained to take place only within the

¶ *Of course it could equally be argued that, as with any* ***portmanteau*** *test, this is a drawback rather than an advantage of ANOSIM. The price for being able to detect changes of different types is arguably a loss of specificity in interpretation, in cases where it is important to ascribe differences solely to a shift in the "mean community" rather than variation changes.*

separate conditions: P and C labels may not be switched. Even so, the number of possible permutations is large (around 20,000).

Notice again that the test is *not* restricted to ***balanced designs***, i.e. those with equal numbers of replicate samples within sites and/or equal numbers of sites within treatments (although lack of balance causes a minor complication in the efficient averaging of R_C and R_P, see Clarke, 1988, 1993). Fig. 6.6b displays the results of 999 simulations (constrained relabellings) from the permutation distribution for $\overline{R}$ under the null hypothesis H1. Possible values range from –0.3 to 0.6, though 95% of the values are seen to be <0.27 and 99% are <0.46. The observed $\overline{R}$ of 0.75 therefore provides a strongly significant rejection of hypothesis H1.

H2, which will usually be the more interesting of the two hypotheses, can now be examined. The test of H1 demonstrated that there are, in effect, only three genuine "replicates" (the sites 1-3) at each of the two conditions (C and P).

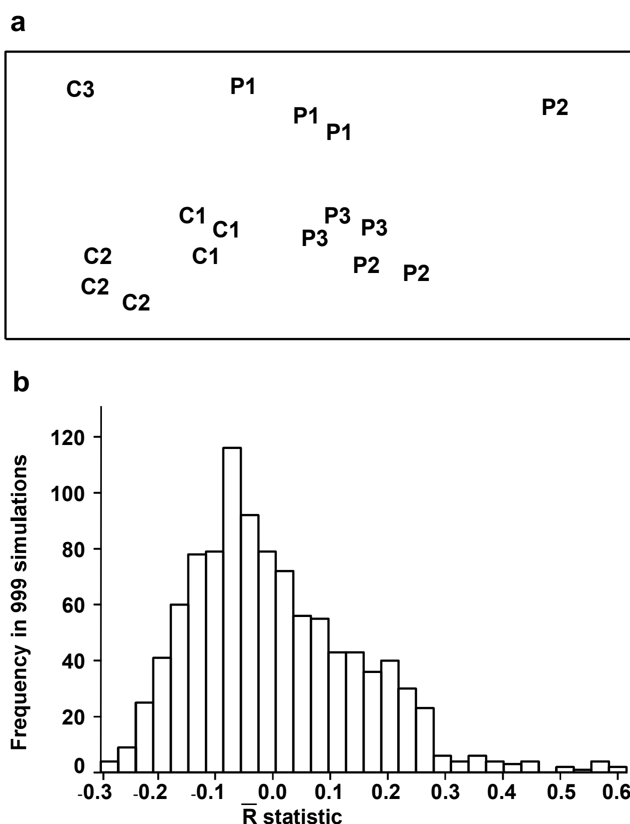

Fig. 6.6. Clyde nematodes {Y}*. a) MDS of species abundances from three "polluted" (P1-P3) and three "control" sites (C1–C3), with three replicate samples at most sites (stress = 0.09). b) Simulated distribution of the test statistic* $\overline{R}$ *under the hypothesis H1 of "no site differences" within each condition; the observed* $\overline{R}$ *is 0.75.*

This is a 1-way layout, and H2 can be tested by 1-way ANOSIM but one first needs to combine the information from the three original replicates at each site, to define a similarity matrix for the 6 "new" replicates. Consistent with the overall strategy that tests should only be dependent on the rank similarities in the original triangular matrix, one first averages over the appropriate ranks to obtain a reduced matrix. For example, the similarity between the three P1 and three P2 replicates is defined as the average of the nine inter-group *rank* similarities; this is placed into the new similarity matrix along with the 14 other averages (C1 with C2, P1 with C1 etc) and all 15 values are then *re-ranked*; the 1-way ANOSIM then gives R = 0.74. There are only 10 distinct permutations so that, although this is actually the most extreme R value possible, H2 is only able to be rejected at a p<10% significance level.

The other scenario to consider is that the first test fails to reject H1. There are then two possibilities for examining H2:

a) Proceed with the average ranking and re-ranking exactly as above, on the assumption that even if it cannot be *proved* that there are no differences between sites it would be unwise to *assume* that this is so; the test may have had rather little power to detect such a difference.

b) Infer from the test of H1 that there *are* no differences between sites, and treat all replicates as if they were separate sites, e.g. there would be 7 replicates for control and 9 replicates for polluted conditions in a 1-way ANOSIM test applied to the 16 samples in Fig. 6.6a.

Which of these two courses to take is a matter for debate, and the argument here is exactly that of whether "to pool" or "not to pool" in forming the residual for the analogous univariate 2-way ANOVA. Option b) will certainly have greater power but runs a real risk of being invalid; option a) is the conservative test and it is certainly unwise to design a study with anything other than option a) in mind.¶

¶ *The ANOSIM program in the PRIMER package always takes this first option.*

EXAMPLE: Eaglehawk Neck meiofauna (2-way crossed case)

An example of a two-way crossed design is given in Warwick *et al* (1990a) and is introduced more fully here in Chapter 12. This is a so-called ***natural experiment***, studying disturbance effects on meiobenthic communities by the continual reworking of sediment by soldier crabs. Two replicate samples were taken from each of four disturbed patches of sediment, and from adjacent undisturbed areas, on a sand flat at Eaglehawk Neck, Tasmania; Fig. 6.7a is a schematic representation of the 16 sample locations. There are two factors: the presence or absence of disturbance by the crabs and the "block effect" of the four different disturbance patches. It might be anticipated that the community will change naturally across the sand flat, from block to block, and it is important to be able to separate this effect from any changes associated with the disturbance itself. There are parallels here with impact studies in which pollutants affect sections of several bays, so that matched control and polluted conditions can be compared against a background of changing community pattern across a wide spatial scale. There are presumed to be replicate samples from each treatment/block combination (the meaning of the term ***crossed***), though balanced numbers are not essential.

For the Eaglehawk Neck data, Fig. 6.7b displays the MDS for the 16 samples (2 treatments × 4 blocks × 2 replicates), based on Bray-Curtis similarities from root-transformed abundances of 59 meiofaunal species. The pattern is remarkably clear and a classic analogue of what, in univariate two-way ANOVA, would be called an ***additive*** model. The meiobenthic community is seen to change from area to area across the sand flat but also appears to differ consistently between disturbed and undisturbed conditions. A test for the latter sets up a null hypothesis that there are no disturbance effects, *allowing* for the fact that there may be block effects, and the procedure is then exactly that of the 2-way ANOSIM test for hypothesis H1 of the nested case. For each *separate* block an R statistic is calculated from equation (6.1), as if for a simple one-way test for a disturbance effect, and the resulting values averaged to give $\overline{R}$. Its permutation distribution under the null hypothesis is generated by examining all simultaneous re-orderings of the four labels (two disturbed, two undisturbed) *within* each block. There are only three distinct permutations in each block, giving a total of 3^4 (= 81) combinations overall and

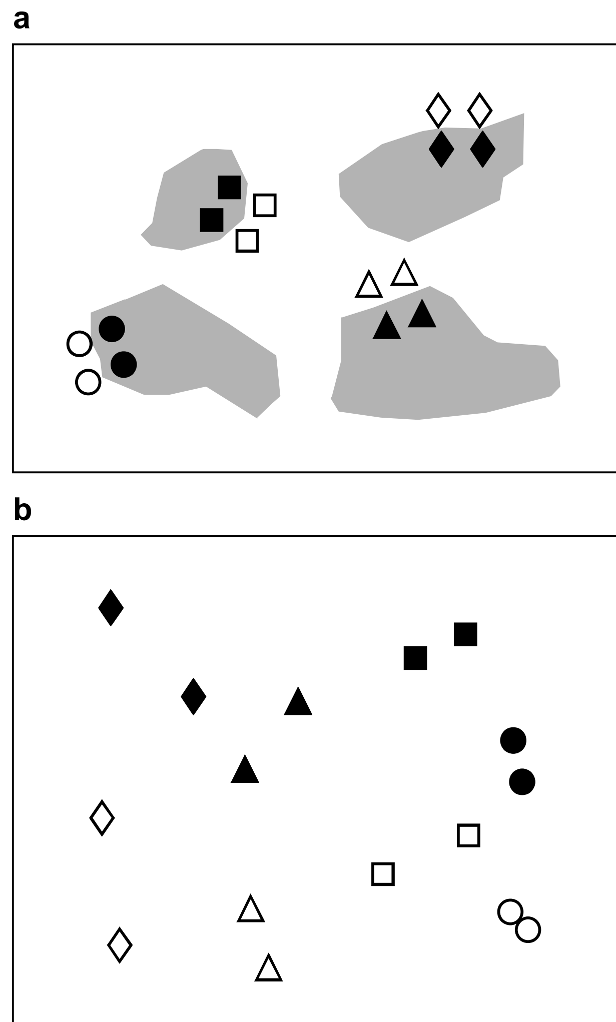

***Fig. 6.7. Tasmania, Eaglehawk Neck {T}**. a) Schematic of the '2-way crossed' sampling design for 16 meiofaunal cores with two disturbed and two undisturbed replicates from each of four patches of burrowing activity by soldier crabs (shaded). b) MDS of species abundances for the 16 samples, showing separation of the blocks on the x-axis and discrimination of disturbed from undisturbed communities on the y-axis (stress = 0.11).*

the observed value of $\overline{R}$ (= 0.94) is the highest value attained in the 81 permutations. The null hypothesis is therefore rejected at a significance level of just over 1%.

The procedure departs from the nested case because of the symmetry in the crossed design. One can now test the null hypothesis that there are no block effects, allowing for the fact that there *are* treatment (disturbance) differences, by simply reversing the roles of treatments and blocks. $\overline{R}$ is now an average of two R statistics, separately calculated for disturbed and undisturbed samples, and there are $8!/[(2!)^4 4!]$ = 105 permutations of the 8 labels for each treatment. A random selection from the 105^2 = 11,025 possible combinations must therefore be made. In 1000 trials

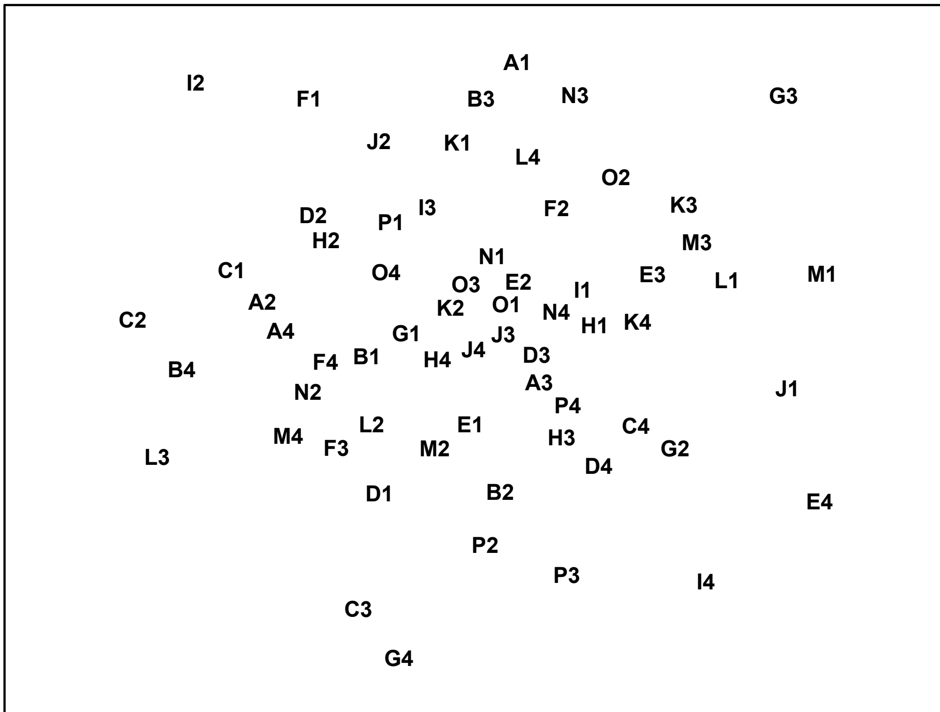

Fig. 6.8. Westerschelde nematodes experiment {W}*. MDS of species abundances from 16 different nutrient-enrichment treatments, A to P, applied to sediment cores in each of four mesocosm basins, 1 to 4 (stress = 0.28).*

the true value of $\bar{R}$ (=0.85) is again the most extreme and is almost certainly the largest in the full set; the null hypothesis is decisively rejected. In this case the test is inherently uninteresting but in other situations (e.g. a sites × times study) tests for both factors could be of practical importance.

EXAMPLE: Mesocosm experiment (2-way crossed case with no replication)

Although the above test may still function if a few random cells in the 2-way layout have only a single replicate, its success depends on reasonable levels of replication overall to generate sufficient permutations. A commonly arising situation in practice, however, is where the 2-way design includes no replication at all.¶ Typically this could be a sites × times field study (see next section) but it may also occur in experimental work: an example is given by Austen and Warwick (1995) of a laboratory mesocosm study in which a complex array of treatments was applied to soft-sediment cores taken from a single, intertidal location in the Westerschelde estuary, Netherlands. A total of 64 cores were randomly divided between 4 mesocosm basins, 16 to a basin.

The experiment involved 15 different nutrient enrichment conditions and one control, the treatments being applied to the surface of the undisturbed sediment cores. After 16 weeks controlled exposure in the mesocosm environment, the meiofaunal communities in the 64 cores were identified, and Bray-Curtis similarities on root-transformed abundances gave the MDS of Fig. 6.8. The full set of 16 treatments was repeated in each of the 4 basins (blocks), so the structure is a 2-way treatments × blocks layout with only one replicate per cell. Little, if any, of this structure is apparent from Fig. 6.8 and a formal test of the null hypothesis

H_o: there are no treatment differences (but allowing the possibility of basin effects)

is clearly necessary before any sort of interpretation is attempted.

In the absence of replication, a test is still possible in the *univariate* case, under the assumption that interaction effects are small in relation to the main treatment or block differences (Scheffé, 1959). In a similar spirit, a *global* test of H_o is possible here, relying on the observation that *if* certain treatments are responsible for community changes, in a more-or-less consistent way across blocks, *separate* MDS analyses for each block should show a repeated treatment pattern. This is illustrated schematically in the top half of Fig. 6.9: the fact that treatment A is consistently close to B (and C to D) can only arise is H_o is false. The analogy

¶ *As noted earlier, this case is not covered by the PRIMER ANOSIM routine. It uses a separate routine, ANOSIM2.*

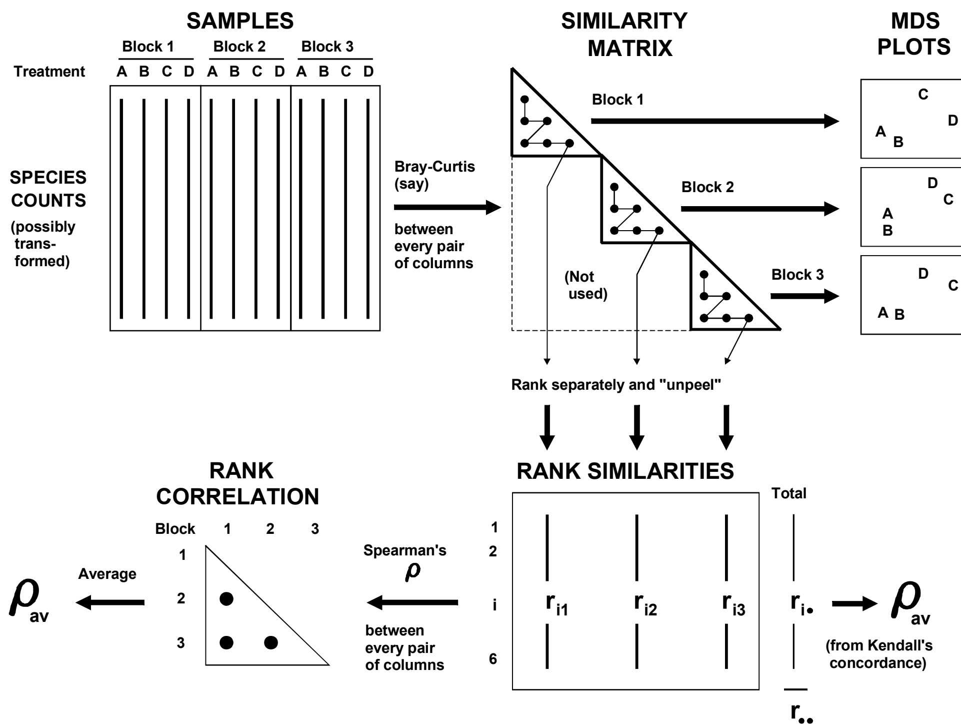

Fig. 6.9. Schematic diagram illustrating the stages in defining concordance of treatment patterns across the blocks, and the two computational routes for ρ_{av}

with the univariate test is clear: large interaction effects imply that the treatment pattern differs from block to block and there is little chance of identifying a treatment effect; on the other hand, for a treatment × block design such as the current mesocosm experiment there is no reason to expect treatments to behave very differently in the different basins.

What is therefore required is a measure of how well the treatment patterns in the ordinations for the different blocks match; this statistic can then be recomputed under all possible (or a random subset of) permutations of the treatment labels within each block. As previously, if the observed statistic does not fall within the body of this (simulated) distribution there is significant evidence to reject H_o. Note that, as required by the statement of H_o, the test makes no assumption about the absence of block effects; between-block similarities are irrelevant to a statistic based only on agreement in within-block patterns.

In fact, for the same reasons advanced for the previous ANOSIM tests (e.g. arbitrariness in choice of MDS dimensionality), it is more satisfactory to define agreement between treatment patterns by reference to the underlying similarity matrix and not the MDS locations. Fig. 6.9 indicates two routes, which lead to equivalent formulations. If there are n treatments and thus $N = n(n-1)/2$ similarities within a block, a natural choice for agreement of two blocks, j and k, is the Spearman correlation coefficient

$$\rho_{jk} = 1 - \frac{6}{N(N^2-1)} \sum_{i=1}^{N} (r_{ij} - r_{ik})^2 \qquad (6.3)$$

between the matching elements of the two *rank* similarity matrices $\{r_{ij}, r_{ik};\ i=1,\dots,N\}$, since these ranks are the only information used in successful MDS plots. The coefficients can be averaged across all $b(b-1)/2$ pairs from the b blocks, to obtain an overall measure of agreement ρ_{av} on which to base the test. A short cut is to define, from the row totals $\{r_{i\cdot}\}$ and grand total $r_{\cdot\cdot}$ shown in Fig. 6.9, Kendall's (1970) ***coefficient of concordance*** between the full set of ranks:

$$W = \frac{12}{b^2 N(N^2-1)} \sum_{i=1}^{N} \left(r_{i\cdot} - \frac{r_{\cdot\cdot}}{N}\right)^2 \qquad (6.4)$$

and then exploit the known relationship between this and ρ_{av}:

$$\rho_{av} = (bW - 1)/(b - 1) \qquad (6.5)$$

As a correlation coefficient, ρ_{av} takes values in the range (–1, 1), with $\rho_{av} = 1$ implying perfect agreement and $\rho_{av} \approx 0$ if the null hypothesis H_o is true.

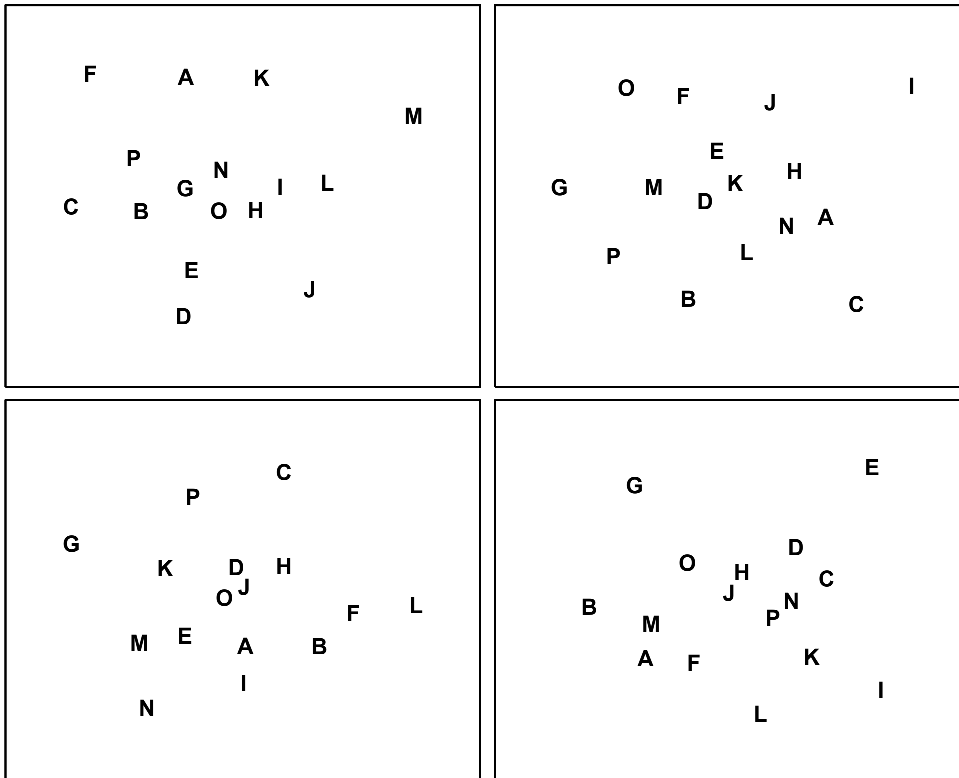

Fig. 6.10. Westerschelde nematodes experiment {W}. MDS for the 16 treatments (A to P), performed separately for each of the four basins; no shared treatment pattern is apparent (stress ranges from 0.16 to 0.20).

Note that standard significance tests and confidence intervals for ρ or W (e.g. as given in basic statistical tables) are totally invalid, since they rely on the ranks $\{r_{ij}; i=1,...,N\}$ being from independent variables; this is obviously not true of similarity coefficients from all possible pairs of a set of (independent) samples. This does not make ρ_{av} any the less appropriate as a measure of agreement whose departure from zero (rejection of H_o) is testable by permutation.

For the nutrient enrichment experiment, Fig. 6.10 shows the separate MDS plots for the 4 mesocosm basins. Although the stress values are rather high (and the plots therefore slightly unreliable as a summary of the among treatment relationships), there appears to be no commonality of pattern, and this is borne out by a near zero value for ρ_{av} of –0.03. This is central to the range of simulated values for ρ_{av} under H_o (obtained by permuting treatment labels separately for each block and recomputing ρ_{av}), so the test provides no evidence of any treatment differences. Note that the symmetry of the 2-way layout also allows a test of the (less interesting) hypothesis that there are no block effects, by looking for any consistency in the among-basin relationships across separate analyses for each of the 16 treatments. The test is again non-significant, with ρ_{av} = –0.02. The overall negative conclusion to the tests should bar any further attempts at interpretation of these data.

EXAMPLE: Exe nematodes (no replication and missing data)

A final example demonstrates a positive outcome to such a test, in a common case of a 2-way layout of sites and times with the additional feature that samples are *missing* altogether from a small number of cells. Fig. 6.11 shows again the MDS, from Chapter 5, of nematode communities at 19 sites in the Exe estuary.

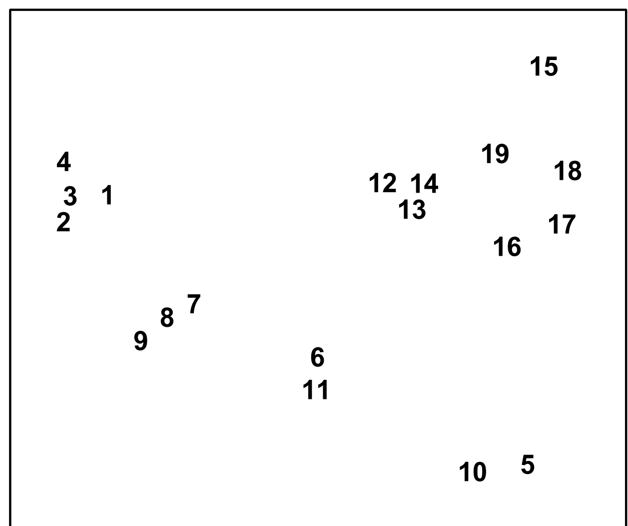

Fig. 6.11. Exe estuary nematodes {X}. MDS, for 19 inter-tidal sites, of species abundances averaged over 6 bi-monthly sampling occasions; see also Fig.5.1 (stress = 0.05).

In fact, this is based on an average of data over six successive bi-monthly sampling occasions. For the individual times, the samples remain strongly clustered into the 4 or 5 main groups apparent from Fig. 6.11. Less clear, however, is whether any structure exists within the largest group (sites 12 to 19) or whether the scatter in Fig. 6.11 is simply the consequence of sampling variation.

Rejection of the null hypothesis of "no site differences" would be suggested by a common site pattern in the separate MDS plots for the 6 times (Fig. 6.12). At some of the times, however, one of the site samples is missing (site 19 at times 1 and 2, site 15 at time 4 and site 18 at time 6). Instead of removing these sites from *all* plots, in order to achieve matching sets of similarities, one can remove for each *pair* of times only those sites missing for either of that pair, and compute the Spearman correlation ρ between the remaining rank similarities. The ρ values for all pairs of times are then averaged to give ρ_{av}, i.e. the left-hand route is taken in the lower half of Fig. 6.9. This is usually referred to as ***pairwise removal*** of missing data, in contrast to the ***listwise removal*** that would be needed for the right-hand route. Though increasing the computation time, pairwise removal clearly utilises more of the available information.

Fig. 6.12 shows evidence of a consistent site pattern, for example in the proximity of sites 12 to 14 and the tendency of site 15 to be placed on its own; the fact that site 15 is missing on one occasion does not undermine this perceived structure. Pairwise computation gives ρ_{av} = 0.36 and its significance can be determined by a Monte Carlo test, as before. The (non-missing) site labels are permuted amongst the available samples, separately for each time, and these designations fixed whilst all the paired ρ values are computed (using pairwise removal) and averaged. Here the, largest such ρ_{av} value in 999 simulations was 0.30, so the null hypothesis is rejected at the p<0.1% level.

In the same way, one can also carry out a test of the hypothesis that there are no differences across *time* for sites 12 to 19. The component plots, of the 4 to 6 times for each site, display no obvious features and ρ_{av} = 0.08 (p<18%). The failure to reject this null hypothesis justifies, to some extent, the use of averaged data across the 6 times, in the earlier analyses.

Tests of this form, searching for agreement between two or more similarity matrices, occur also in Chapter 11 (in the context of matching species to environmental data) and Chapter 15 (where they link biotic patterns to some model structure). The discussion there includes use of measures other than a simple Spearman coefficient, for example a weighted Spearman coefficient ρ_w (suggested for reasons explained in Chapter 11), and these adjustments could certainly be implemented here also if desired, using the left-hand route in the lower half of Fig.6.9. In the present context, this type of "matching" test is clearly an inferior one to that possible where genuine replication exists within the 2-way layout. It cannot cope with follow-up tests for differences between specific pairs of treatments, and it can have little sensitivity if the numbers of treatments and blocks are both small. A test for two treatments is impossible note, since the treatment pattern in all blocks would be identical.

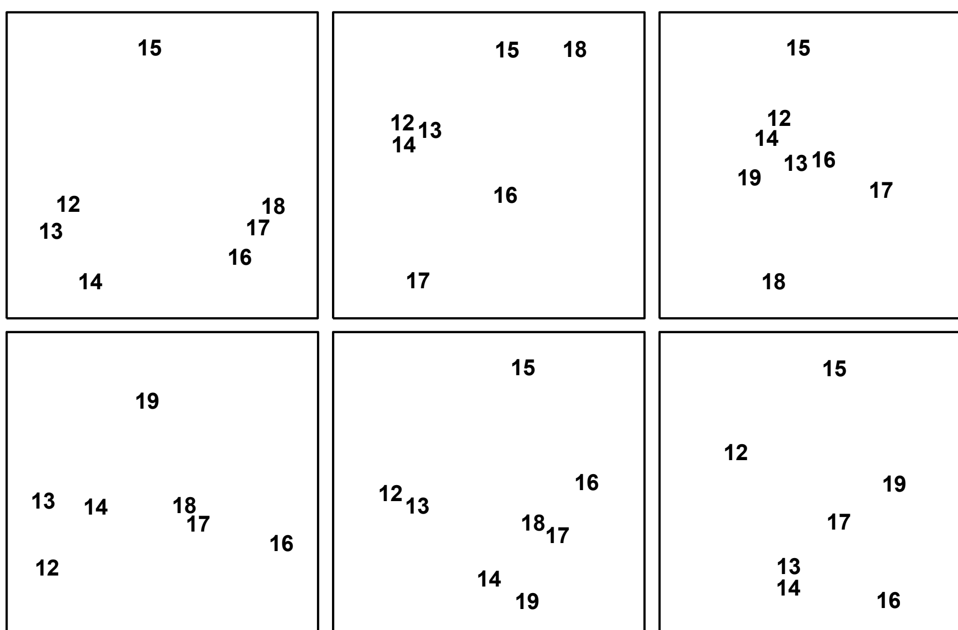

Fig. 6.12. Exe estuary nematodes {X}. MDS for sites 12 to 19 only, performed separately for the 6 sampling times (read across rows for time order); in spite of the occasional missing sample some commonality of site pattern is apparent (stress ranges from 0.01 to 0.08).

Higher-way layouts

The ANOSIM tests described above cover only some of the simpler design layouts: one-way and two-way nested and crossed layouts, the latter restricted to a test for an overall effect of (either) one of the factors. There is no direct test for an ***interaction*** between the two factors. Indeed, without further model assumptions of some form, it is not even clear that interaction is a precisely-defined concept. (This is very much tied up with the issue of linearity of models. In a univariate context note that an interaction term in a two-way layout can sometimes be made to disappear simply by a monotonic transformation of the variable – non-parametrically speaking, no change at all!)

RECOMMENDATIONS

1) For typical species abundance matrices, it is much better to use an ANOSIM-type permutation procedure rather than a classical MANOVA test; the latter will almost always be totally invalid.

2) Choice of the level and type of replication should be carefully considered. Though it is difficult to define power for any of the ANOSIM tests, it is clearly important to take sufficient replicates to generate a large enough set of permutations for meaningful significance levels. Equally important is that replicates should *genuinely* represent the condition being sampled: ***pseudo-replication*** (see Hurlbert, 1984) is commonplace, e.g. analyses of sub-cores of single cores, or sets of spatially contiguous samples which are unrepresentative of the extent of a site. For pseudo-replicates in a 2-way layout, the only valid course is to average them and carry out the above global test for the case of “no replication”.

3) A point that cannot be over-stressed is that ANOSIM tests only apply to groups of samples specified *prior* to seeing (or collecting) the data. A dangerous misconception is that one can use a cluster analysis of the species abundance data to define sample groupings, whose statistical validity can be established by performing an ANOSIM test for differences between these groups. *This is entirely erroneous*, the argument being completely circular. The only safe course here is to use this first set of data to define potential groups of interest, i.e. to erect the hypothesis, and then to collect a further set of data to test that hypothesis.

APPENDIX 2: PRINCIPAL LITERATURE SOURCES AND FURTHER READING

This manual chiefly reflects an approach to multivariate and other community analyses that has been adopted and developed at the Plymouth Marine Laboratory (PML) for well over a decade, and has benefited from experience at numerous IOC/UNEP and commercial training courses. Methods papers from work at PML covered in this manual include: Field *et al* (1982), Warwick (1986), Clarke and Green (1988), Clarke (1990), Warwick and Clarke (1993a & b, 1995a), Clarke and Ainsworth (1993), Clarke *et al* (1993), Clarke and Warwick (1994, 1998a & b, 1999, 2001) and Somerfield and Clarke (1995, 1997). Clarke (1993, 1999), Warwick (1993) and Warwick and Clarke (2001) give general reviews, and a large number of papers from PML and authors worldwide exemplify their use via the PRIMER package (some are listed in Appendix 3 but there are currently over a thousand in the SCI).

Of course, the exposition here draws on a wider body of statistical techniques, and there follows a brief list of the main sources that can be consulted for more detail on the methods and analyses of each Chapter.

Chapter 1: Framework. The categorisation here is an extension of that given by Warwick (1988a). The Frierfjord macrofauna data and analyses (Tables 1.2 & 1.6 and Figs. 1.1, 1.2 & 1.7) are extracted and re-drawn from Bayne *et al* (1988), Gray *et al* (1988) and Clarke and Green (1988), the Loch Linnhe macrofauna data (Table 1.4 and Fig. 1.3) from Pearson (1975), and the ABC curves (Fig. 1.4) from Warwick (1986). The species abundance distribution for Garroch Head macrofauna (Fig. 1.6) is first found in Pearson *et al* (1983), and the multivariate linking to environmental variables (Fig. 1.9) in Clarke and Ainsworth (1993). The mesocosm data and analysis (Table 1.7 and Fig. 1.10) are extracted and redrawn from Gee *et al* (1985).

Chapters 2 and 3: Similarity and Clustering. These methods originated in the 1950's and 60's (e.g. Florek *et al,* 1951; Sneath, 1957; Lance and Williams, 1967). The description here widens that of Field *et al* (1982), with some points taken from the general texts of Everitt (1980) and Cormack (1971). The dendrogram of Frierfjord macrofaunal samples (Fig.3.1) is redrawn from Gray *et al* (1988), and the zooplankton example (Figs. 3.2 & 3.3) from Collins and Williams (1982).

Chapter 4: Ordination by PCA. This is a founding technique of multivariate statistics, see for example Chatfield and Collins (1980) and Everitt (1978). The final example (Fig. 4.2) is from Warwick *et al* (1988).

Chapter 5: Ordination by MDS. Non-metric MDS was introduced by Shepard (1962) and Kruskal (1964); two standard texts are Kruskal and Wish (1978) and Schiffman *et al* (1981). Here, the exposition parallels that in Field *et al* (1982) and Clarke (1993); the Exe nematode graphs (Figs. 5.1–5.4) are redrawn from the former. The dosing experiment (Fig. 5.5) is discussed in Warwick *et al* (1988).

Chapter 6: Testing. The basic permutation test and simulation of significance levels can be traced to Mantel (1967) and Hope (1968), respectively. In this context (e.g. Figs. 6.2 & 6.3 and eqt. 6.1) it is described by Clarke and Green (1988). A fuller discussion of the extension to 2-way nested and crossed ANOSIM tests (including Figs. 6.4 & 6.6) is in Clarke (1993) (with some asymptotic results in Clarke, 1988); the coral analysis (Fig. 6.5) is in Warwick *et al* (1990b), and the Tasmanian meiofaunal MDS (Fig. 6.7) in Warwick *et al* (1990a). The 2-way design without replication (Figs. 6.8–6.12) is tackled in Clarke and Warwick (1994); see also Austen and Warwick (1995).

Chapter 7: Species analyses. Clustering and ordination of species similarities is as illustrated in Field *et al* (1982), for the Exe nematode data (Figs 7.1 & 7.2, redrawn); see also Clifford and Stephenson (1975). The SIMPER ("similarity percentages") procedure is described in Clarke (1993).

Chapter 8: Univariate/graphical analyses. Pielou (1975), Heip *et al* (1988) and Magurran (1991) are useful texts, summarising a large literature on a variety of diversity indices and ranked species abundance plots. The diversity examples here (Figs. 8.1 & 8.2) are discussed by Warwick *et al* (1990c, 1990b respectively) and the Caswell *V* computations (Table 8.1) are from Warwick *et al* (1990c). The Garroch Head species abundance distributions (Fig. 8.4) are first found in Pearson *et al* (1983); Fig. 8.3 is redrawn from Pearson and Blackstock (1984). Warwick (1986) introduced Abundance–Biomass Comparison curves, and the Loch Linnhe and Garroch Head illustrations (Figs. 8.7 & 8.8) are redrawn from Warwick (1986) and Warwick *et al* (1987). The transformed scale and partial dominance curves of Figs. 8.9–8.11 were suggested by Clarke (1990), which paper also tackles issues of summary statistics (Fig. 8.12, equation 8.7, and as employed in Fig. 8.13) and significance tests for dominance curves.

Chapter 9: Transformations. This chapter is an expansion of the discussion in Clarke and Green (1988);

Fig. 9.1 is recomputed from Warwick *et al* (1988).

Chapter 10: Aggregation. This description of the effects of changing taxonomic level is based on Warwick (1988b), from which Figs. 10.2–10.4 and 10.7 are redrawn. Fig. 10.1 is discussed in Gray *et al* (1988), Fig. 10.5 and 10.8 in Warwick *et al* (1990b) and Fig. 10.6 in Gray *et al* (1990) (or Warwick and Clarke, 1993a, in this categorisation). More recent work on the effects on the analysis of choice of taxonomic level (and transform) can be found in Olsgard *et al* (1997, 1998) and Olsgard and Somerfield (2000).

Chapter 11: Linking to environment. For wider reading on this type of "canonical" problem, see Chapter 5 of Jongman *et al* (1987), including ter Braak's (1986) method of canonical correspondence analysis. The approach here of performing environmental and biotic analyses separately, and then comparing them, combines that advocated by Field *et al* (1982: superimposing variables on the biotic MDS) and by Clarke and Ainsworth (1993: the BIO-ENV program). The data in Table 11.1 is from Pearson and Blackstock (1984). Fig 11.3 is redrawn from Collins and Williams (1982) and Fig. 11.6 from Field *et al* (1982); Figs. 11.7–11.8, 11.10 and Table 11.2 are from Clarke and Ainsworth (1993).

Chapter 12: Community experiments. Influential papers and books on field experiments, and causal interpretation from observational studies in general, include Connell (1974), Hurlbert (1984), Green (1979) and many papers by A J Underwood, M G Chapman and collaborators, in particular the Underwood (1997) book. Underwood and Peterson (1988) give some thoughts specifically on mesocosm experiments. Lab-based "microcosm" experiments on community structure, using this analysis approach, are typified by Austen and Somerfield (1997) and Schratzberger and Warwick (1998b). Figs. 12.2 and 12.3 are redrawn from Warwick *et al* (1990a) and Figs. 12.5, 12.6 from Gee *et al* (1985).

Chapter 13: Data requirements. The exposition parallels that in Warwick (1993) but with additional examples. Figs. 13.1–13.3 and 13.8 are redrawn from Warwick (1993), and earlier found in Colebrook (1986), Dawson-Shepherd *et al* (1992), Warwick (1988b) and Gray *et al* (1988) respectively. Fig. 13.4 is redrawn from Warwick *et al* (1990a), Fig. 13.5 from Warwick *et al* (1990c), Fig. 13.6 from Warwick *et al* (1990b) and Fig. 13.7 from Warwick and Clarke (1991).

Chapter 14: Relative sensitivities. This parallels the earlier sections of Warwick and Clarke (1991), from which all these figures (except Figs. 14.11 & 14.14) have been redrawn. Primary source versions of the figures can be found as follows: Figs. 14.1–14.3, Gray *et al* (1988); Figs. 14.5–14.7, Warwick *et al* (1990b); Figs 14.9–14.10, Dawson-Shepherd *et al* (1992); Figs. 14.11–14.12, Gee and Warwick (1994); Figs. 14.14–14.16, Austen and Warwick (1989).

Chapter 15: Multivariate measures of disturbance. This follows the format of Warwick and Clarke (1995) and is an amalgamation of ideas from three primary papers: Warwick and Clarke (1993a) on "meta-analysis" of NE Atlantic macrobenthic studies, Warwick and Clarke (1993b) on the increase in multivariate dispersion under disturbance, and Clarke *et al* (1993) on the breakdown of multivariate seriation patterns. Figs. 15.1–15.3 and Table 15.1 are redrawn and extracted from the first, Fig. 15.4 and Table 15.2 from the second and Figs. 15.5 & 15.6 and Table 15.5 from the third. The analysis in Table 15.4 is from Warwick *et al* (in press).

Chapter 16: Comparing multivariate patterns. The general extension of the BIO-ENV approach of Chapter 11, to combinations other than selecting environmental variables to match biotic patterns, is described in Clarke and Warwick (1998a). This details the forward/backward stepping search algorithm BVSTEP, and uses it to select subsets of "influential species" from a biotic matrix. Second-stage MDS is defined in Somerfield and Clarke (1995) and early examples of its use can be found in Olsgard *et al* (1997, 1998). Figs. 16-1 to 16-3, and Tables 16-1, 16-2, are extracted from Clarke and Warwick (1998a), and Fig. 16-5 from Somerfield and Clarke (1995).

Chapter 17: Taxonomic distinctness measures. Warwick and Clarke (1995a) first defined taxonomic diversity/distinctness. Earlier work, from a conservation perspective, and using different species relatedness properties (such as PD), can be found in, e.g. Faith (1992, 1994), Vane-Wright *et al* (1991) and Williams *et al* (1991). The superior sampling properties of average taxonomic distinctness (Δ^+), and its testing structure in the case of simple species lists, are given in Clarke and Warwick (1998b), and applied to UK nematodes by Warwick and Clarke (1998) and Clarke and Warwick (1999). Variation in taxonomic distinctness (Λ^+) was introduced, and its sampling properties examined, in Clarke and Warwick (2001), and a review of the area can be found in Warwick and Clarke (2001), from which Figs. 17.1, 17.2, 17.5, 17.11, 17.12 are redrawn. Fig. 17.3 is discussed in Warwick and Clarke (1995a), Fig. 17.4 in Warwick *et al* (in press), Figs. 17.6, 17.8, 17.9, 17.14, 17.17 in Clarke and Warwick (2001), Fig. 17.7 in Clarke and Warwick (1998b) and Figs. 17.10, 17.13 in Rogers *et al* (1999).

APPENDIX 3: REFERENCES CITED

Addison, R.F., Clarke, K.R. (eds.) (1990). *Biological effects of pollutants in a subtropical environment. J. exp. mar. Biol. Ecol.* 138

Agard, J.B.R., Gobin, J., Warwick, R.M. (1993). Analysis of marine macrobenthic community structure in relation to pollution, natural oil seepage and seasonal disturbance in a tropical environment (Trinidad, West Indies). *Mar. Ecol. Prog. Ser.* 92: 233–243

Anderson, M.J. (2001a). Permutation tests for univariate or multivariate analysis of variance and regression. *Can. J. Fish. Aquat. Sci.* 58: 626-639

Anderson, M.J. (2001b). A new method for non-parametric multivariate analysis of variance. *Austral. Ecol*. 26: 32-46

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: 442-453

Austen, M.C., McEvoy, A.J. (1997). The use of offshore meiobenthic communities in laboratory microcosm experiments: response to heavy metal contamination. *J. exp. Mar. Biol. Ecol.* 211: 247-261

Austen, M.C., Somerfield, P.J. (1997). A community level sediment bioassay applied to an estuarine heavy metal gradient. *Mar. Envir. Res.* 43: 315-328

Austen, M.C., Thrush, S.F. (2001). Experimental evidence suggesting slow or weak response of nematode community structure to a large suspension-feeder. *J. Sea Res.* 46: 69-84

Austen, M.C., Warwick, R.M. (1989). Comparison of univariate and multivariate aspects of estuarine meiobenthic community structure. *Est. cstl. shelf Sci.* 29: 23–42

Austen, M.C., Warwick, R.M. (1995). Effects of manipulation of food supply on estuarine meiobenthos. *Hydrobiologia* 311: 175–184

Austen, M.C., Widdicombe, S., Villano-Pitacco, N. (1998). Effects of biological disturbance on diversity and structure of meiobenthic nematode communities. *Mar. Ecol. Prog. Ser.* 174: 233-246

Bayne, B.L., Clarke, K.R., Gray, J.S. (eds) (1988). *Biological effects of pollutants: results of a practical workshop. Mar. Ecol. Prog. Ser.* 46

Beukema, J.J. (1988). An evaluation of the ABC–method (abundance /biomass comparison) as applied to macrozoobenthic communities living on tidal flats in the Dutch Wadden Sea. *Mar. Biol.* 99: 425–433

Box, G.E.P., Cox, D.R. (1964). An analysis of transformations. *J. R. Statist. Soc. Ser. B* 26: 211–243

Bray, J.R., Curtis, J.T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27: 325–349

Brown, B.E., Clarke, K.R., Warwick, R.M. (in press). Serial patterns of biodiversity change in corals across shallow reef flats in Ko Phuket, Thailand, due to the effects of local (sedimentation) and regional (climatic) perturbations. *Mar. Biol.*

Buchanan, J.B. (1993). Evidence of benthic pelagic coupling at a station off the Northumberland coast. *J. exp. Mar. Biol. Ecol.* 172: 1–10

Caswell, H. (1976). Community structure: a neutral model analysis. *Ecol. Monogr.* 46: 327–354

Chapman, M.G., Underwood, A.J. (1999). Ecological patterns in multivariate assemblages: information and interpretation of negative values in ANOSIM tests. *Mar. Ecol. Prog. Ser.* 180: 257–265

Chatfield, C., Collins, A.J. (1980). *Introduction to multivariate analysis*. Chapman and Hall, London

Clarke, K.R. (1988). Detecting change in benthic community structure. Proceedings XIVth International Biometric Conference, Namur: Invited Papers. Societe Adolphe Quetelet, Gembloux, Belgium

Clarke, K.R. (1990). Comparisons of dominance curves. *J. exp. mar. Biol. Ecol.* 138: 143–157

Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18: 117–143

Clarke, K.R. (1999). Non-metric multivariate analysis in community-level ecotoxicology. *Environ. Toxicol. Chem.* 18: 118–127

Clarke, K.R., Ainsworth, M. (1993). A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser*. 92: 205–219

Clarke, K.R., Green, R.H. (1988). Statistical design and analysis for a ‘biological effects’ study. *Mar. Ecol. Prog. Ser.* 46: 213–226

Clarke, K.R., Warwick, R.M. (1994). Similarity-based testing for community pattern: the 2-way layout with no replication. *Mar. Biol*. 118: 167–176

Clarke, K.R., Warwick, R.M. (1998a). Quantifying structural redundancy in ecological communities. *Oecologia* 113: 278–289

Clarke, K.R., Warwick, R.M. (1998b). A taxonomic distinctness index and its statistical properties. *J. appl. Ecol.* 35: 523–531

Clarke, K.R., Warwick, R.M. (1999). The taxonomic distinctness measure of biodiversity: weighting of step lengths between hierarchical levels. *Mar. Ecol. Prog. Ser.* 184: 21–29

Clarke, K.R, Warwick, R.M. (2001). A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Mar. Ecol. Prog. Ser.* 216: 265–278

Clarke, K.R., Warwick, R.M., Brown, B.E. (1993). An index showing breakdown of seriation, related to disturbance, in a coral-reef assemblage. *Mar. Ecol. Prog. Ser*. 102: 153–160

Clifford, D.H.T., Stephenson, W. (1975). *An introduction to numerical classification*. Academic Press, New York

Colebrook, J.M. (1986). Environmental influences on long-term variability in marine plankton. *Hydrobiologia* 142: 309–325

Collins, N.R., Williams, R. (1982). Zooplankton communities in the Bristol Channel and Severn Estuary. *Mar. Ecol. Prog. Ser.* 9: 1–11

Connell, J.H. (1974). Field experiments in marine ecology. In: Mariscal, R. (ed.) *Experimental marine biology*. Academic Press, New York

Connell, J.H. (1978). Diversity in tropical rain forests and coral reefs. *Science N.Y.* 199: 1302–1310

Cormack, R.M. (1971). A review of classification. *J. R. Statist. Soc. Ser.* A 134: 321–367

Dauvin, J-C. (1984). Dynamique d'ecosystemes macrobenthiques des fonds sedimentaires de la Baie de Morlaix et leur perturbation par les hydrocarbures de l'Amoco-Cadiz. Doctoral thesis, Univ. Pierre et Marie-Curie, Paris

Dawson-Shepherd, A., Warwick, R.M., Clarke, K.R., Brown, B.E. (1992). An analysis of fish community responses to coral mining in the Maldives. *Environ. Biol. Fish*. 33: 367–380

Everitt, B. (1978). *Graphical techniques for multivariate data.* Heinemann, London

Everitt, B. (1980). *Cluster analysis, 2nd edn*. Heinemann, London

Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61: 1-10

Faith, D.P. (1994). Phylogenetic pattern and the quantification of organismal biodiversity. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 345: 45-58

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

Field, J.G., Clarke, K.R., Warwick, R.M. (1982). A practical strategy for analysing multispecies distribution patterns. *Mar. Ecol. Prog. Ser.* 8: 37–52

Fisher, R.A., Corbet, A.S., Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. anim. Ecol.* 12: 42-58

Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, H., Zubrzycki, S. (1951). Sur la liason et la division des points d'un ensemble fini. *Colloquium Math*. 2: 282–285

Gage, J.D., Coghill, G.G. (1977). Studies on the dispersion patterns of Scottish sea-loch benthos from contiguous core transects. In: Coull, B. (ed) *Ecology of marine benthos*. University of South Carolina Press, Columbia.

Gee, J.M., Somerfield, P.J. (1997). Do mangrove diversity and leaf litter decay promote meiofaunal diversity? *J. exp. mar. Ecol. Biol.* 218: 13-33

Gee, J.M., Warwick, R.M. (1994a). Metazoan community structure in relation to the fractal dimensions of marine microalgae. *Mar. Ecol. Prog. Ser.* 103: 141–150

Gee, J.M., Warwick, R.M. (1994b). Body-size distribution in a marine metazoan community and the fractal dimensions of macroalgae. *J. exp. mar. Biol. Ecol.* 178: 247–259

Gee, J.M., Warwick, R., Schaanning, M., Berge, J.A., Ambrose Jr, W.G. (1985). Effects of organic enrichment on meiofaunal abundance and community structure in sublittoral soft sediments. *J. exp. mar. Biol. Ecol*. 91: 247–262

Goldman, N., Lambshead, P.J.D. (1989). Optimization of the Ewens /Caswell neutral model program for community diversity analysis. *Mar. Ecol. Prog. Ser.* 50: 255–261

Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–328

Gower, J.C. (1971). Statistical methods of comparing different multivariate analyses of the same data. pp. 138–149 in F.R. Hodson, D.G. Kendall and P. Tautu (eds.), *Mathematics in the archaeological and historical sciences*. Edinburgh University Press, Edinburgh

Gower, J.C., Ross, G.J.S. (1969). Minimum spanning trees and single linkage cluster analysis. *Appl. Statist.* 18: 54–64

Gray, J.S., Aschan, M., Carr, M.R., Clarke, K.R., Green, R.H., Pearson, T.H., Rosenberg, R., Warwick, R.M. (1988). Analysis of community attributes of the benthic macrofauna of Frierfjord /Langesundfjord and in a mesocosm experiment. *Mar. Ecol. Prog. Ser.* 46: 151–165

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: 285–299

Gray, J.S., Pearson, T.H. (1982). Objective selection of sensitive species indicative of pollution-induced change in benthic communities. I. Comparative methodology. *Mar. Ecol. Prog. Ser.* 9: 111–119

Green, R.H. (1979). *Sampling design and statistical methods for environmental biologists.* Wiley, New York

Greenacre, M.J. (1984). *Theory and applications of correspondence analysis.* Academic Press, London

Hall, S.J., Greenstreet, S.P. (1998). Taxonomic distinctness and diversity measures: responses in marine fish communities. *Mar. Ecol. Prog. Ser.* 166: 227-229

Harper, J.L., Hawksworth, D.L. (1994). Biodiversity: measurement and estimation. Preface. *Phil. Trans. Roy. Soc. Lond. Ser. B* 345: 5-12.

Heip, C., Herman, P.M.J., Soetaert, K. (1988). Data processing, evaluation, and analysis. pp. 197–231 in R.P. Higgins and H. Thiel (eds.), *Introduction to the study of meiofauna.* Smithsonian Institution, Washington DC

Hill, M.O. (1973a). Reciprocal averaging: an eigenvector method of ordination. *J. Ecol*. 61: 237–249

Hill, M.O. (1973b). Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427–432

Hill, M.O. (1979a). *DECORANA – A FORTRAN program for detrended correspondence analysis and reciprocal averaging*. Cornell University, Ithaca, New York

Hill, M.O. (1979b). *TWINSPAN – A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of individuals and attributes.* Cornell University, Ithaca, New York

Hill, M.O., Gauch, H.G. (1980). Detrended correspondence analysis, an improved ordination technique. *Vegetatio* 42: 47–48

Hope, A.C.A. (1968). A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc. Ser. B* 30: 582–598

Howson, C.M. (ed.) (1987). *Directory of the British marine fauna and flora*. Marine Conservation Society, Ross-on-Wye, Hertfordshire

Humphries, C.J., Williams, P.H., Vane-Wright, R.I. (1995). Measuring biodiversity value for conservation. *Ann. Rev. Ecol. Syst.* 26: 93-111

Hurlbert, S.H. (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52: 577-586

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

Huston, M. (1979). A general hypothesis of species diversity. *Am. Nat.* 113: 81–101

Ibanez, F., Dauvin, J.-C. (1988). Long-term changes (1977–1987) in a muddy fine sand *Abra alba – Melinna palmata* community from the Western English Channel: multivariate time-series analysis. *Mar. Ecol. Prog. Ser.* 49: 65–81

Izsak, C., Price, A.R.G. (2001). Measuring β-diversity using a taxonomic similarity index, and its relation to spatial scale. *Mar. Ecol. Prog. Ser.* 215: 69-77

Jayasree, K. (1976). Systematics and ecology of free-living marine nematodes from polluted intertidal sand in Scotland. Ph.D. thesis, University of Aberdeen

Jongman, R.H.G., ter Braak, C.F.J., van Tongeren, O.F.R. (1987). *Data analysis in community and landscape ecology*. Pudoc, Wageningen

Kendall, M.A., Widdicombe, S. (1999). Small scale patterns in the structure of macrofaunal assemblages of shallow soft sediments. *J. exp. mar. Biol. Ecol.* 237: 127-140

Kendall, M.G. (1970). *Rank correlation methods*. Griffin, London

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

Krzanowski, W.J. (in press). Multifactorial analysis of distance in studies of ecological community structure. *J. Agric. Biol. Environ. Sci.*

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1–27

Kruskal, J.B., Wish, M. (1978). *Multidimensional scaling*. Sage Publications, Beverley Hills, California

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

Lambshead, P.J.D. (1986). Sub-catastrophic sewage and industrial waste contamination as revealed by marine nematode faunal analysis. *Mar. Ecol. Prog. Ser.* 29: 247–260

Lambshead, P.J.D., Platt, H.M., Shaw, K.M. (1983). The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. *J. nat. Hist.* 17: 859–874

Lance, G.N., Williams, W.T. (1967). A general theory of classificatory sorting strategies: 1 Hierarchical Systems. *Comp. J.* 9: 373–380

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 Engl. edn*. Elsevier, Amsterdam

Lorenzen, S. (1994). *The phylogenetic systematics of free-living nematodes*. Ray Society, London

Magurran, A.E. (1991). *Ecological diversity and its measurement.* Chapman and Hall, London

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res*. 27: 209–220

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

May, R.M. (1990). Taxonomy as destiny. *Nature* 347: 129-130

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

McEachran, J.D., Miyake, T. (1990). Phylogenetic interrelationships of skates: a working hypothesis (Chondrichthyes, Rajoidea). In: *Elasmobranchs as living resources: advances in the biology, ecology, systematics, and the status of the fisheries*, H.L. Pratt *et al* (eds). NOAA Technical Report NMFS 90: 285-304

Morrisey, D.J., Underwood, A.J., Howitt, L. (1996). Effects of copper on the faunas of marine soft-sediments: an experimental field study. *Mar. Biol.* 125: 199-213

Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Ann. Rev. Genet.* 30: 371-403

Nelson, J.S. (1994). *Fishes of the world, 3rd edn*. Wiley, New York

Olafsson, E., Moore, C.G. (1992). Effects of macroepifauna on developing nematode and harpacticoid assemblages in a subtidal muddy habitat. *Mar. Ecol. Prog. Ser.* 84: 161–171

Olsgard, F., Somerfield, P.J. (2000). Surrogates in marine benthic investigations – which taxonomic unit to target? *J. Aquat. Ecosyst. Stress Recov.* 7: 25–42

Olsgard, F., Somerfield, P.J., Carr, M.R. (1997). Relationships between taxonomic resolution and data transformations in analyses of a macrobenthic community along an established pollution gradient. *Mar. Ecol. Prog. Ser.* 149: 173–181

Olsgard, F., Somerfield, P.J., Carr, M.R. (1998). Relationships between taxonomic resolution, macrobenthic community patterns and disturbance. *Mar. Ecol. Prog. Ser.* 172: 25–36

Pearson, T.H. (1975). The benthic ecology of Loch Linnhe and Loch Eil, a sea-loch system on the west coast of Scotland. IV. Changes in the benthic fauna attributable to organic enrichment. *J. exp. mar. Biol. Ecol.* 20: 1–41

Pearson, T.H., Blackstock, J. (1984). Garroch Head sludge dumping ground survey, final report. Dunstaffnage Marine Research Laboratory (unpublished)

Pearson, T.H., Gray, J.S., Johannessen, P.J. (1983). Objective selection of sensitive species indicative of pollution-induced change in benthic communities. 2. Data analyses. *Mar. Ecol. Prog. Ser.* 12: 237–255

Pielou, E.C. (1975). *Ecological diversity*. Wiley, New York

Pielou, E.C. (1984). *The interpretation of ecological data. A primer on classification and ordination*. Wiley, New York

Piepenburg, D., Voss, J., Gutt, J. (1997). Assemblages of sea stars (Echinodermata: Asteroidea) and brittle stars (Echinodermata: Ophiuroidea) in the Weddell Sea (Antarctica) and off Northeast Greenland (Arctic): a comparison of diversity and abundance. *Polar Biol.* 17: 305-322

Platt, H.M., Warwick, R.M. (1983). *Freeliving marine nematodes. Part I. British Enoplids.* Synopses of the British Fauna no 28. Cambridge University Press, Cambridge

Platt, H.M., Warwick, R.M. (1988). *Freeliving marine nematodes. Part II. British chromadorids.* E.J. Brill, Leiden

Platell, M.E., Potter, I.C., Clarke, K.R. (1998). Resource partitioning by four species of elasmobranchs (Batoidea: Urolophidae) in coastal waters of temperate Australia. *Mar. Biol.* 131: 719–734

Potter, I.C., Claridge, P.N., Hyndes, G.A., Clarke, K.R. (1997). Seasonal, annual and regional variations in ichthyofaunal composition in the inner Severn Estuary and inner Bristol Channel. *J. mar. biol. Ass. U.K.* 77: 507–525

Price, A.R.G., Keeling, M.J., O'Callaghan, C.J. (1999). Ocean-scale patterns of 'biodiversity' of Atlantic asteroids determined from taxonomic distinctness and other measures. *Biol. J. Linn. Soc*. 66: 187-203

Raffaelli, D., Mason, C.F. (1981). Pollution monitoring with meiofauna using the ratio of nematodes to copepods. *Mar. Poll. Bull.* 12: 158–163

Reid, P.C., Barges, M.dF., Svendsen, E. (in press). A regime shift in the North Sea circa 1988 linked to changes in the North Sea fishery. *Fish. Res*.

Rogers, S.I., Clarke, K.R., Reynolds, J.D. (1999). The taxonomic distinctness of coastal bottom-dwelling fish communities of the North-east Atlantic. *J. Anim. Ecol.* 68: 769–782

Sanders, H.L. (1968). Marine benthic diversity: a comparative study. *Am. Nat*. 102: 243–282

Scheffe, H. (1959). *The analysis of variance*. Wiley, New York

Schiffman, S.S., Reynolds, M.L., Young, F.W. (1981). *Introduction to multi-dimensional scaling. Theory, methods and applications.* Academic Press, London

Schratzberger, M., Warwick, R.M. (1998a). Effects of the intensity and frequency of organic enrichment on two estuarine nematode communities. *Mar. Ecol. Prog. Ser.* 164: 83-94

Schratzberger, M., Warwick, R.M. (1998b). Effects of physical disturbance on nematode communities in sand and mud: a microcosm experiment. *Mar. Biol.* 130: 643-650

Schratzberger, M., Warwick, R.M. (1999). Differential effects of various types of disturbances on the structure of nematode assemblages: an experimental approach. *Mar. Ecol. Prog. Ser.* 181: 227-236

Schwinghamer, P. (1981). Characteristic size distributions of integral benthic communities. *Can. J. Fish. aquat. Sci.* 38: 1255–1263

Seber, G.A.F. (1984). *Multivariate observations*. Wiley, New York

Shimatani, K. (2001). On the measurement of species diversity incorporating species differences. *Oikos* 93: 135-147

Shephard, R.N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. *Psychometrika* 27: 125–140

Simpson, E.H. (1949). Measurement of diversity. *Nature* 163: 688

Sneath, P.H.A. (1957). The application of computers to taxonomy. *J. gen. Microbiol.* 17: 201–226

Sneath, P.H.A., Sokal, R.R. (1973). *Numerical taxonomy.* Freeman, San Francisco

Sokal, R.R., Rohlf, F.J. (1981). *Biometry.* Freeman, San Francisco

Somerfield, P.J., Gage, J.D. (2000). Community structure of the benthos in Scottish sea-lochs. IV. Multivariate spatial pattern. *Mar. Biol.* 136: 1133-1145

Somerfield, P.J., Gee, J.M., Warwick, R.M. (1994a). Benthic community structure in relation to an instantaneous discharge of waste water from a tin mine. *Mar. Pollut. Bull.* 28: 363–369

Somerfield, P.J., Gee, J.M., Warwick, R.M. (1994b). Soft sediment meiofaunal community structure in relation to a long-term heavy metal gradient in the Fal estuary system. *Mar. Ecol. Prog. Ser.* 105: 79–88

Somerfield, P.J., Gee, J.M., Widdicombe, S. (1993). *The use of meiobenthos in marine pollution monitoring programmes.* Plymouth Marine Laboratory Miscellaneous Publication LIB-33A,B, Plymouth

Somerfield, P.J., Clarke, K.R. (1995). Taxonomic levels, in marine community studies, revisited. *Mar. Ecol. Prog. Ser.* 127: 113–119

Somerfield, P.J., Clarke, K.R. (1997). A comparison of some methods commonly used for the collection of sublittoral sediments and their associated fauna. *Mar. Environ. Res.* 43: 143–156

Somerfield, P.J., Olsgard, F., Carr, M.R. (1997). A further examination of two new taxonomic distinctness measures. *Mar. Ecol. Prog. Ser.* 154: 303-306

Somerfield, P.J., Rees, H.L., Warwick, R.M. (1995). Interrelationships in community structure between shallow-water marine meiofauna and macrofauna in relation to dredgings disposal. *Mar. Ecol. Prog. Ser.* 127: 103–112

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

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. (1992). Beyond BACI: the detection of environmental impact on populations in the real, but variable, world. *J. exp. Mar. Biol. Ecol*. 161: 145–178

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* 117: 570-578

Underwood, A.J., Peterson, C.H. (1988). Towards an ecological framework for investigating pollution. *Mar. Ecol. Prog. Ser.* 46: 227–234

Vane-Wright, R.I., Humphries, C.J., Williams, P.H. (1991). What to protect? Systematics and the agony of choice. *Biol. Conserv.* 55: 235-254

Ward, P., Woodd-Walker, R., Clarke, A. (submitted). Seasonality rather than temperature regulates zooplankton diversity. *Nature*

Warwick, R.M. (1971). Nematode associations in the Exe estuary. *J. mar. Biol. Ass. U.K.* 51: 439–454

Warwick, R.M. (1984). Species size distributions in marine benthic communities. *Oecologia (Berlin)* 61: 32–41

Warwick, R.M. (1986). A new method for detecting pollution effects on marine macrobenthic communities. *Mar. Biol.* 92: 557–562

Warwick, R.M. (1988a). Effects on community structure of a pollutant gradient – summary. *Mar. Ecol. Prog. Ser.* 46: 207–211

Warwick, R.M. (1988b). The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. *Mar. Pollut. Bull.* 19: 259–268

Warwick, R.M. (1993). Environmental impact studies on marine communities: pragmatical considerations. *Aust. J. Ecol.* 18: 63–80

Warwick, R.M., Ashman, C.M., Brown, A.R., Clarke, K.R., Dowell, B., Hart, B., Lewis, R.E., Shillabeer, N., Somerfield, P.J., Tapp, J.F. (in press). Inter-annual changes in the biodiversity and community stucture of the macrobenthos in Tees Bay and the Tees estuary, UK, associated with local and regional environmental events. *Mar. Ecol. Prog. Ser.*

Warwick, R.M., Buchanan, J.B. (1970). The meiofauna off the coast of Northumberland. I: The structure of the nematode population. *J. mar. biol. Assoc. UK* 50: 129-146

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: 181–191

Warwick, R.M., Clarke, K.R. (1991). A comparison of methods for analysing changes in benthic community structure. *J. mar. Biol. Ass. U.K.* 71: 225–244

Warwick, R.M., Clarke, K.R. (1993a). Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data. *Mar. Ecol. Prog. Ser.* 92: 221–231

Warwick, R.M., Clarke, K.R. (1993b). Increased variability as a symptom of stress in marine communities. *J. exp. mar. Biol. Ecol.* 172: 215–226

Warwick, R.M., Clarke, K.R. (1994). Relearning the ABC: taxonomic changes and abundance/biomass relationships in disturbed benthic communities. *Mar. Biol.* 118: 739–744

Warwick, R.M., Clarke, K.R. (1995a). New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Mar. Ecol. Prog. Ser.* 129: 301–305

Warwick, R.M., Clarke, K.R. (1995b). Multivariate measures of community stress and their application to marine pollution studies in the East Asian region. *Phuket mar. biol. Cent. Res. Bull.* 60: 99–113

Warwick, R.M., Clarke, K.R. (1998). Taxonomic distinctness and environmental assessment. *J. appl. Ecol.* 35: 532–543

Warwick, R.M., Clarke, K.R. (2001). Practical measures of marine biodiversity based on relatedness of species. *Oceanogr. Mar. Biol. Ann. Rev.* 39: 207–231

Warwick, R.M., Clarke, K.R., Gee, J.M. (1990a). The effects of disturbance by soldier crabs, *Mictyris platycheles* H. Milne Edwards, on meiobenthic community structure. *J. exp. Mar. Biol. Ecol.* 135: 19–33

Warwick, R.M., Clarke, K.R., Suharsono. (1990b). A statistical analysis of coral community responses to the 1982–3 El Nino in the Thousand Islands, Indonesia. *Coral Reefs* 8: 171–179

Warwick, R.M., Coles, J.W. (1977). The marine flora and fauna of the Isles of Scilly. Free-living Nematoda. *J. Nat. Hist.* 11: 393–407

Warwick, R.M., Light, J. (in press). Death assemblages of molluscs on St. Martin's Flats, Isles of Scilly: a surrogate for regional biodiversity? *Biodivers. Conserv.*

Warwick, R.M., Pearson, T.H., Ruswahyuni (1987). Detection of pollution effects on marine macrobenthos: further evaluation of the species abundance/biomass method. *Mar. Biol.* 95: 193–200

Warwick, R.M., Platt, H.M., Clarke, K.R., Agard, J., Gobin, J. (1990c). Analysis of macrobenthic and meiobenthic community structure in relation to pollution and disturbance in Hamilton Harbour, Bermuda. *J. exp. Mar. Biol. Ecol.* 138: 119–142

Warwick, R.M., Platt, H.M., Somerfield, P.J. (1998). *Freeliving marine nematodes. Part III. British Monhysterida.* Synopses of the British Fauna no 53. Field Studies Council, Shrewsbury

Widdicombe, S., Austen, M.C. (1998). Experimental evidence for the role of *Brissopsis lyrifera* (Forbes, 1841) as a critical species in the maintenance of benthic diversity and the modification of sediment chemistry. *J. exp. Mar. Biol. Ecol.* 228: 241–255

Widdicombe, S., Austen, M.C. (2001). The interaction between physical disturbance and organic enrichment: an important element in structuring benthic communities. *Limnol. Oceanog.* 46: 1720-1733

Wilkinson, D.M. (1999). The disturbing history of intermediate disturbance. *Oikos* 84: 145-147

Williams, P.H., Humphries, C.J., Vane-Wright, R.I. (1991). Measuring biodiversity: taxonomic relatedness for conservation priorities. *Aust. Syst. Bot.* 4: 665-679

ACKNOWLEDGEMENTS

We are grateful to a large number of individuals for helpful collaboration and discussions: Paul Somerfield, Martin Carr, Ray Gorley, Paul Dowland, Ben Knight, Marios Mitella, Jane Addy, John Hall, Martin Budge, Mike Ainsworth, Robert Pritchard, Charlie Green, John Bramley, Simon Frith, Roger Carter, Mike Gee, Mel Austen, John Gray, Frode Olsgard and many others who in various ways contributed to the analyses featured in this manual.

We also gratefully acknowledge institutional support from a diversity of sources, primarily the Plymouth Marine Laboratory of the Natural Environment Research Council, UK (which employed both of us for our entire research careers, until 2001), IOC, UNEP and FAO, who supported the earlier development work, and the UK Ministry of Agriculture, Fisheries and Food (now the Department of Environment, Food & Rural Affairs) who jointly funded, with the NERC, some of the more recent methods developments.

K R Clarke
R M Warwick
2001