Nonparametric Multivariate Analyses of Changes in Community Structure

Article *in* Austral Ecology · March 1993

DOI: 10.1111/j.1442-9993.1993.tb00438.x

1 author:

K. Robert Clarke
Plymouth Marine Laboratory
161 PUBLICATIONS **74,157** CITATIONS

Australian Journal of Ecology (1993) **18**, 117–143

Non-parametric multivariate analyses of changes in community structure

K. R. CLARKE
Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth, PL1 3DH, United Kingdom

Abstract In the early 1980s, a strategy for graphical representation of multivariate (multi-species) abundance data was introduced into marine ecology by, among others, Field, *et al.* (1982). A decade on, it is instructive to: (i) identify which elements of this often-quoted strategy have proved most useful in practical assessment of community change resulting from pollution impact; and (ii) ask to what extent evolution of techniques in the intervening years has added self-consistency and comprehensiveness to the approach. The pivotal concept has proved to be that of a biologically-relevant definition of similarity of two samples, and its utilization mainly in simple rank form, for example 'sample A is more similar to sample B than it is to sample C'. Statistical assumptions about the data are thus minimized and the resulting non-parametric techniques will be of very general applicability. From such a starting point, a unified framework needs to encompass: (i) the display of community patterns through clustering and ordination of samples; (ii) identification of species principally responsible for determining sample groupings; (iii) statistical tests for differences in space and time (multivariate analogues of analysis of variance, based on rank similarities); and (iv) the linking of community differences to patterns in the physical and chemical environment (the latter also dictated by rank similarities between samples). Techniques are described that bring such a framework into place, and areas in which problems remain are identified. Accumulated practical experience with these methods is discussed, in particular applications to marine benthos, and it is concluded that they have much to offer practitioners of environmental impact studies on communities.

INTRODUCTION AND RATIONALE

Strategy of Field *et al.* (1982)

Field *et al.* (1982) outlined a strategy for the analysis of data on community structure, that is, an abundance (or biomass) array whose columns represent separate samples and whose rows are the full set of species present in those samples. (Note that all such assemblage information is referred to, rather loosely, throughout the current paper as 'community' data.) Their approach has the following components.

(1) The biotic relationship between any two samples is distilled into a coefficient measuring similarity (or dissimilarity) in species composition.

(2) The resulting triangular matrix of similarities between every pair of samples is used to classify the samples into groups, by hierarchical agglomerative clustering with group-average linking (e.g. Clifford & Stephenson 1975), or to map the sample inter-relationships in an ordination, by non-metric multi-dimensional scaling (MDS, e.g. Kruskal & Wish 1978).

(3) Relationships between the species are examined by, in effect, transposing the data matrix and repeating the classification and ordination on a similarity matrix computed between every pair of species. Species which are indicative of particular groups of samples are determined by so-called information statistic (I-) tests (Field 1969).

(4) Having allowed the community data to 'tell its own story', its relationship to matching environmental data is examined by superimposing the values of each abiotic variable separately onto the biotic ordination.

The above strategy has been adopted in a sizeable number of published studies, for example papers in Bayne *et al.* (1988), Addison and Clarke (1990), Warwick *et al.* (1991), Agard *et al.* (1993)

Accepted for publication October 1992.

and many others. (There are about 80 non-self citations to Field *et al.* 1982 in the Science Citation Index.) Many of these studies are concerned with the effects of pollutants. For example, samples might consist of a set of replicate sediment cores at different sites or times, chosen with the intention of displaying pollution-induced change in benthic community structure.

Fundamental role of rank similarities among samples

The effectiveness of this strategy has depended on the flexibility inherent in the use of non-metric MDS as an ordination technique. This is based only on the similarity matrix between samples, as defined by the biologist to reflect the particular aspects of community structure that are biologically meaningful for that study. The ordination technique should not force a specific definition of similarity (either explicitly or implicitly) onto the practitioner. He or she should control the following initial stages, by answering the questions posed.

(1) Selection of community attribute, for example 'is the biological hypothesis best examined by data on species abundance or biomass, or by some combination of these reflecting production (as in Warwick & Clarke 1993a; Warwick 1993)?'

(2) Standardization of data to relative rather than absolute values, for example 'are differences in total abundance between samples of any biological significance?'

(3) Transformation of the data matrix (after standardization, if appropriate), for example 'which end of the spectrum, of common to rare species, should the analysis chiefly reflect?' Untransformed data will typically lead to a shallow interpretation in which only the pattern of a few, very common species is represented (although ordinations will then fit more readily into two dimensions). The transformation sequence of $y^{0.5}$, $y^{0.25}$, $\log y$ and ultimately simple presence/absence, allows progressively greater contribution from the rarer species (e.g. Clarke & Green 1988).

(4) Choice of similarity coefficient, for example 'should the similarity between two samples depend in any way on those species that are absent from both (but present in other samples in the data set)?' Ecologists generally seem to feel that so-called 'joint absences' are not germane, ruling out some common coefficients. Other desirable features include invariance (even under power transformation) to a scale change, for example of biomass measurements, and some form of standardization ensuring that the extreme values of 100 and 0 correspond, respectively, to a complete match of the measurements and to a complete lack of species in common. These properties are all satisfied by a number of coefficients, including the Bray–Curtis similarity (Bray & Curtis 1957), widely used in ecology. Faith *et al.* (1987) discuss the robustness of several such coefficients, in a simulation study of a range of (non-linear) ecological response models; the Bray–Curtis coefficient is seen to be one of the most reliable performers.

Some care will therefore have been taken in defining similarity to reflect biological reality, but it is still true that no particular meaning can be attached to an isolated value, say a similarity of 64.3 between samples A and B. The absolute levels will depend markedly on the chosen coefficient and transformation (typically Bray–Curtis similarities will tend to increase with increasing severity of transformation). It is *relative* levels which have a natural interpretation, in particular the ranks of the similarity matrix, which summarize the data through statements such as 'sample A is more similar to sample B than it is to sample C'. This is an intuitively appealing and very generally applicable base from which to build a graphical representation of the sample patterns and, in effect, this is the only information used by a successful non-metric MDS ordination. The rank similarity matrix thus plays a fundamental role in defining and visualizing the community pattern. It is then natural to demand consistency in answering subsequent questions concerning that pattern, by utilizing only these same rank similarities.

Weaknesses in the earlier approach and new requirements

Some of the components of the Field *et al.* (1982) approach fail this consistency test.

(1) The I-statistic method for identifying indicator species is a function only of the presence or absence of species and thus has a rather tenuous link with the Bray–Curtis similarity matrix on

quantitative data, used elsewhere in the strategy. Its provenance as a hypothesis testing procedure does not stand up to scrutiny in this case, as Field *et al.* (1982) pointed out. At a more pragmatic level, it has an inevitable tendency to select indicator species that are rare, particularly for small clusters where chance occurrence of single individuals in one group can make a rare species look like a 'perfect' indicator. It shares this property with correspondence analysis and two-way indicator species analysis (Hill 1973; 1979; Greenacre 1984), as these methods are deliberately designed to identify clusters with mutually exclusive species sets.

(2) The species analysis, based on a matrix of similarities between species in their disposition across samples, has also been found to have limited practical use. This is principally because of the high degree of noise in individual species patterns, causing difficulty in representing the among-species relationships in a low-dimensional ordination. The approach tends to be informative only in strongly clustered situations in which most species divide into mutually exclusive groups. (For further discussion of why ordination of species might be expected to cause more problems than ordination of samples see Faith 1991.) In practice, questions concerning species identities often turn into queries about what effect individual species have in determining the among-*sample* relationships seen in a cluster dendrogram or MDS plot. There is thus a requirement to identify individual species contributions to the underlying sample similarity matrix.

(3) Hierarchical cluster analysis of samples, using group-average linking, is not a function only of the rank similarities, though single-linkage clustering would be. Use of single-linkage is not advised, however, because it has a tendency to give rise to unhelpful dendrogram plots, with chain linking of samples rather than clearly defined groups (Everitt 1980), even in situations where the latter are to be expected. The requirement here is to reconcile the minor inconsistency between the information used in a non-metric MDS and that exploited by group-average clustering, particularly where the point of doing a cluster analysis is to check the adequacy of an MDS ordination, by superimposition of the clusters (see later discussion). This is not a requirement of any great theoretical or practical import but can be simply achieved.

(4) A cluster analysis or MDS ordination of the full set of samples deliberately makes no use of the way the samples are structured (e.g. replicates within different sites, times, etc.). The overall sample pattern is displayed and one can then judge visually whether, for example, different sites appear to have differing community composition, based on the variation among replicates within a site. This visual comparison may be rather inadequate in pollution impact studies for which there may often be a priori hypotheses about (lack of) differences before and after an impact, or between control and impacted sites. There is also increasing interest in designed studies of pollution or disturbance effects on communities, involving field manipulations or experimental mesocosms (e.g. Gee *et al.* 1985). These designs are handled in classical statistics by multivariate analysis of variance (MANOVA, e.g. Mardia *et al.* 1979) but the required assumptions of multivariate normality are impossible to meet for many benthic community data sets. There is a clear requirement, not addressed by Field *et al.* (1982), to extend the ad hoc multivariate methods to encompass formal hypothesis-testing situations, without sacrificing the strongly non-parametric (distribution-free) nature of analyses based on rank similarities.

(5) Finally, in many impact studies, the biotic samples will be supplemented by matched environmental data, both on levels of pollutants and underlying physical variables that could be structuring the community (such as sediment grain size or depth of the water column). Field *et al.* (1982) illustrated the linking of environmental information to the biological analysis, by superimposing the values of the abiotic variables, one at a time, on to the respective sample positions in the biotic MDS. While this can be a powerful visual tool in simple situations, it gives no basis for answering such questions as 'how well does the full set of recorded environmental data explain the observed community pattern?' and 'is there a subset of the environmental variables that explains the pattern equally well, or better?' These questions are answered in classical multivariate statistics by techniques such as canonical correlation (e.g. Mardia *et al.* 1979) but, as remarked above, there will be few impact studies on complete communities for which classic multivariate assumptions are valid. What is needed here is an analogue of canonical analysis based only on rank similarities.

Outline of proposed strategy

This paper therefore attempts to erect a coherent framework for non-parametric multivariate analysis of community data, that is, a combination of techniques that acknowledge the primacy of the among-sample similarity matrix and whose inferences are drawn only from its ranks (or at least consistently with that starting point). The consequent lack of model assumptions will confer a general validity of application which it would be hard to improve upon, though it should be recognized that there will be a price to pay: some features of classical multivariate analysis have no obvious analogue in this similarity-based setting.

The main sections of this paper deal with the following components of this strategy.

(1) *Display of community pattern* through ordination and clustering. The rationale and utility of MDS are illustrated with several examples, including accumulated practical experiences of the algorithm's behaviour. The possibility of applying group-average clustering to the rank similarities, rather than the absolute values, is discussed in passing.

(2) *Determining the species responsible for sample groupings* observed in a cluster analysis. Naturally, because information identifying individual species has been entirely lost from the among-sample similarity matrix, some return to an earlier phase of the analysis is essential for this objective. In line with the underlying principle, however, the only information exploited is the contribution each species makes to the chosen similarity coefficient (after the chosen standardization, transformation etc.). The mean contribution of each species to the dissimilarity of two clusters is defined as an average over all cross-group pairs of samples. This yields an assessment of which species are good discriminators of these two groups. A subtly different question is to ask which species are typical of specific groups, in the sense of making a large contribution to the average similarity between every pair of samples *within* a group. (Species can be typical of more than one group and thus poor discriminators between groups.)

(3) *Testing for spatial and temporal differences* in community structure when samples are adequately replicated and hypotheses defined a priori. For example, with suitable replicate samples from each of a number of sites, the hypothesis of 'no site-to-site differences' can be tested by permutations of the rank similarity matrix. This 'analysis of similarities' (ANOSIM) test is a distribution-free analogue of one-way ANOVA. Higher-way analogues also occur in practice. For example, adding another level of structure, in which sites are selected to represent impacted or control locations, produces a two-way nested design. Both this and a two-way crossed layout can be handled, at least in part, by a further permutation or randomization procedure.

(4) *Linking community patterns to environmental variables*, where a suite of abiotic data has been collected to match each biotic sample. The questions posed earlier, about the extent to which the abiotic data 'explains' the community structure, are answered by an indirect analysis, again only involving rank similarities. It is based on the premise that pairs of samples which are rather similar in their values for a set of environmental data would be expected to have rather similar species composition, provided the relevant variables determining community structure have been identified correctly. If this is so, separate ordinations of biotic and abiotic data would be expected to show a close match. This gives rise to a simple optimization routine (the BIO-ENV procedure), which selects that subset of environmental variables maximizing a (modified) rank correlation between the biotic and abiotic similarity matrices. Clarke and Ainsworth (1993) describe the procedure in detail so the technique is only outlined here and illustrated by a new example.

Throughout the paper, there is a degree of promiscuity in the use of illustrations, which are all re-analyses of published data. While several do not refer specifically to biological effects of pollutants, they are all chosen to exemplify analyses with obvious parallels in environmental impact studies.

DISPLAY OF COMMUNITY PATTERN

Non-metric multi-dimensional scaling

As implied earlier, non-metric MDS is often the method of choice for graphical representation of community relationships (e.g. Everitt 1978; Kenkel & Orloci 1986), principally because of the flexibility and generality bestowed by:

(1) its dependence only on a biologically mean-

ingful view of the data, that is, choice of standardization, transformation and similarity coefficient appropriate to the hypotheses under investigation;

(2) its distance-preserving properties, that is, preservation of the rank order of among-sample dissimilarities in the rank order of distances.

The computational algorithm, an iterative optimization procedure (Kruskal & Wish 1978), is fairly complex but the principle is very simple — indeed its conceptual clarity is perhaps MDS's most important asset when communicating results of impact studies to environmental managers.

Analogy of reconstructing a map of the world

The main features of MDS are well illustrated by the following question. Starting from the triangular matrix of (great-circle) distances between every pair of major cities in the world, can one reconstruct the map of the globe, that is, place the cities in their correct location? This is what non-metric MDS sets out to do except that, in effect, it attempts to solve a harder problem, that of reconstructing the world map from only the rank order of inter-city distances (e.g. statements of the form 'Sydney is closer to Canberra than London is to Paris'). Somewhat surprisingly, it succeeds; the result is a near-perfect map of the relative locations of the cities in three dimensions. The algorithm actually works in a brute force manner by initially placing the cities in three-dimensional space at entirely arbitrary locations, and then gradually refining their relative positions in an iterative cycle (involving a combination of the numerical techniques of monotonic regression and steepest descent). The intention, though not necessarily the detail, is clear throughout: to move cities into positions in which the rank order of the inter-city distances becomes ever closer to the rank order in the original triangular matrix. The extent to which the two disagree is reflected in a stress coefficient, where stress tends to zero when the rank orders reach perfect agreement.

Perfect agreement will not always be possible. For example, starting with the same triangular distance matrix, suppose we now wish to construct a map of the cities in two dimensions; some distortion of the true relative positions is inevitable. Figure 1 shows the result of two-dimensional MDS applied to all great-circle distances between 39 cities (*Reader's Digest Great World Atlas* 1968). The reconstruction is still remarkably good, with a seemingly accurate placement of cities in local regions. The distortion is most evident for San Francisco; the MDS is attempting to reconcile the

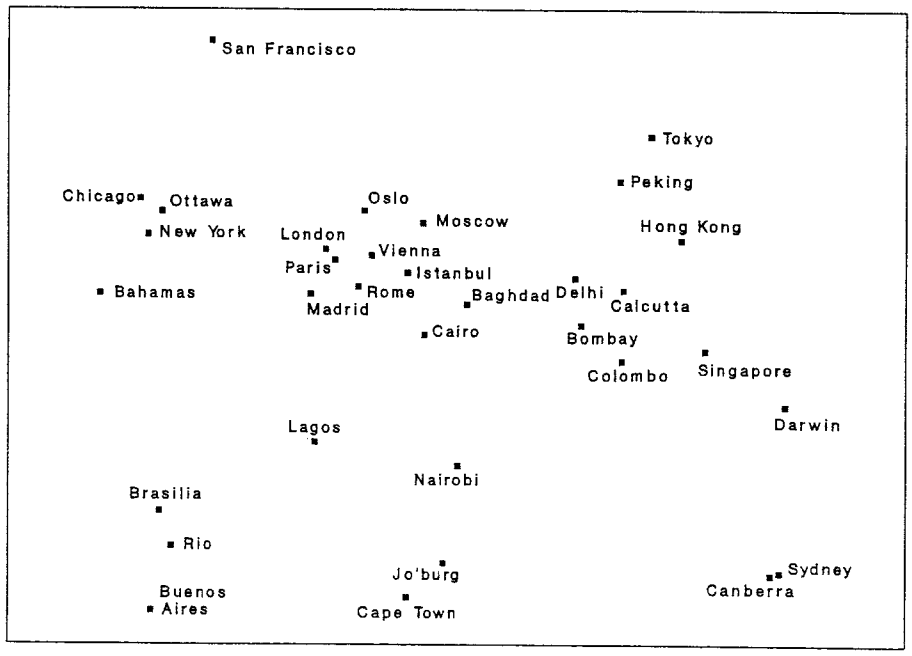

Fig. 1. Two-dimensional ordination of 39 world cities from non-metric multi-dimensional scaling (MDS), applied to a triangular matrix of 'great-circle' distances between every pair of cities. Stress=0.13.

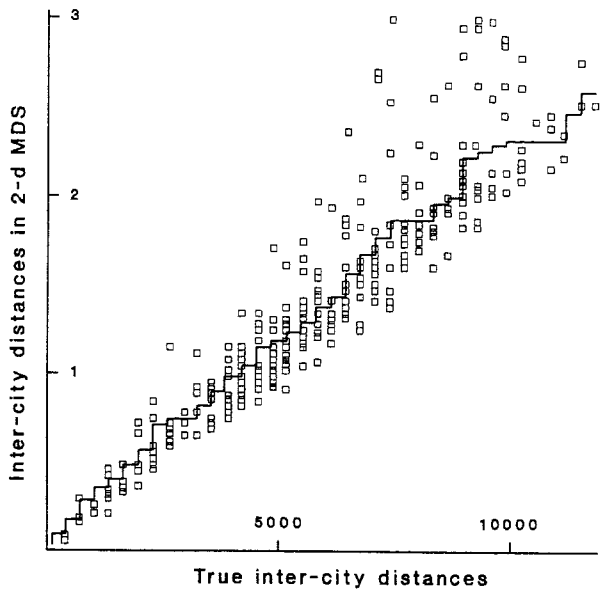

Fig. 2. Scatter plot ('Shepard diagram') of all inter-city distances from the MDS of Fig. 1 (y axis) against the corresponding true great-circle distances (x axis). The line denotes the best-fit monotonic (increasing) regression of y on x; scatter about this line defines the MDS 'stress'.

shorter (cross-Pacific) distance from San Francisco to Tokyo with the cumulatively longer distance expected from the intermediate steps of, say, San Francisco to New York to London to Delhi to Tokyo.

The resulting stress is usefully displayed in a Shepard diagram (Fig. 2), a simple scatter plot of the distances in the original triangular matrix against the corresponding distances between cities in the final MDS. There is little scatter at low distances in Fig. 2, bearing out the observed accuracy of the map for local regions. The continuous line denotes the fitted (monotonic increasing) regression, MDS's best estimate of the relationship between original and final distance. By definition, if rank order distances agree then there is no stress and no scatter around this line. Stress is therefore defined in terms of total scatter, here taking the value 0.13 (Kruskal's stress formula 1). This stress is by no means small in relation to other quoted values in this paper, yet the final plot gives a rather accurate representation in two dimensions. The overall picture can certainly be interpreted more easily than the original distance matrix, and is unlikely to mislead. This is reassuring for interpretation of the later ecological examples.

Community changes following the *Amoco Cadiz* oil spill (Morlaix)

The above example may have little to do with community data but it is a helpful analogy. The triangular matrix of similarity (or dissimilarity) in species composition between every pair of samples provides rank-order statements such as 'sample A is more similar to sample B than it is to sample C'; these are entirely equivalent to the previous example's 'city A is closer to city B than to city C'. One can thus use non-metric MDS to reconstruct a map of the samples in two or more dimensions, in which relative distance apart of the samples reflects relative similarity in species composition. Just as for the world cities, there is no guarantee that the rank similarities can be accurately preserved in (say) a two-dimensional layout of the samples, and it is important to note the level of stress involved.

Figure 3 gives one example of the use of MDS in a temporal study of environmental impact. Samples of subtidal macrobenthic communities were taken by Dauvin (1984) at a single station in the Bay of Morlaix, Brittany, France, at roughly 3-monthly intervals between April 1977 and February 1982, covering the period of the *Amoco Cadiz* oil-tanker disaster. This took place in March 1978 at a distance of some 50 km from the Morlaix site, but the resultant oil slick was dispersed quite widely along the Brittany coast. Abundances for a

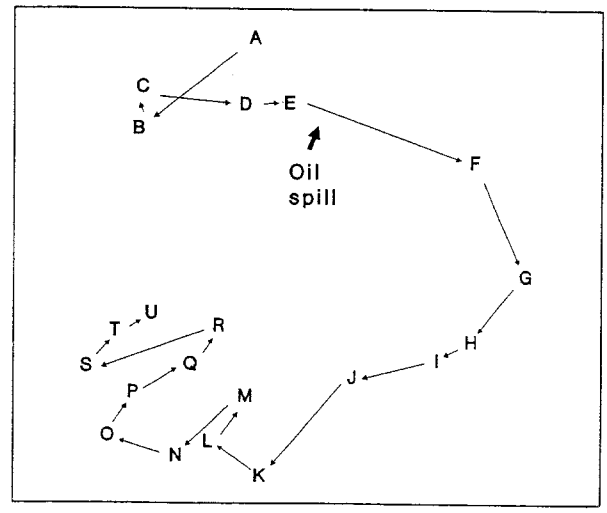

Fig. 3. MDS of approximately quarterly samples (A,B,C, . . .) of macrobenthic communities in the Bay of Morlaix, Brittany, covering the period of the *Amoco Cadiz* oil spill. Stress=0.09.

total of 257 species were recorded in the 21 samples; after double-root transformation, Bray-Curtis similarities gave rise to the MDS ordination in Fig. 3. The 3-monthly samples are coded A, B, C, . . ., with the oil spill taking place between samples E and F. The stress coefficient is low, and the resulting pattern invites ready interpretation: seasonal changes in community structure (e.g. A to D) were rather small in relation to the large perturbation evident following the oil spill. After 2 years the community had begun to settle down to a more stable mix, with comparable seasonal variation, but with a somewhat different species composition than originally found.

Sampling design

Several caveats to these conclusions are necessary, from the perspective of experimental design. The absence of replicates on each sampling occasion makes it largely impossible to infer the presence of a seasonal signal at all — the variation in quarterly samples in the pre-impact phase could simply be a reflection of sampling variability from replicate cores on a single occasion. More importantly for the main conclusions of the study, there is no spatial control (Green 1979), that is, a comparable site sampled over the same period but which remains unimpacted (or preferably several such sites, Underwood 1992). It is at least possible that all such sites along the Brittany coast, whether subject to the impact or not, exhibited a similar pattern in which marked change in community structure was forced at the same time (e.g. by local climatic factors). Observational studies are prone to such caveats, of course, and even the addition of good spatial as well as temporal controls does not guarantee immunity from criticism. It is inevitable in a purely observational study that there could be unrecorded environmental factors which co-vary closely with the contaminant signal and are themselves the cause of any observed community change (e.g. Clarke & Green 1988). Such factors can only be fully controlled by 'randomizing them out', to use the jargon of statistical experimental design. This would involve experimental protocols in which treatments (e.g. pollution impacts) are randomly allocated to experimental units (e.g. different sites), rarely a credible option for environmental impact studies in general! Nonetheless, careful design can do much to reduce the likelihood of drawing misleading inferences from biological effects studies (e.g. Underwood & Peterson 1988).

Community pattern around a North Sea oil field (Ekofisk)

An example of a spatial, rather than temporal, study is that of macrobenthic communities sampled in July 1987, at 39 locations around the Ekofisk oil field in the Norwegian sector of the

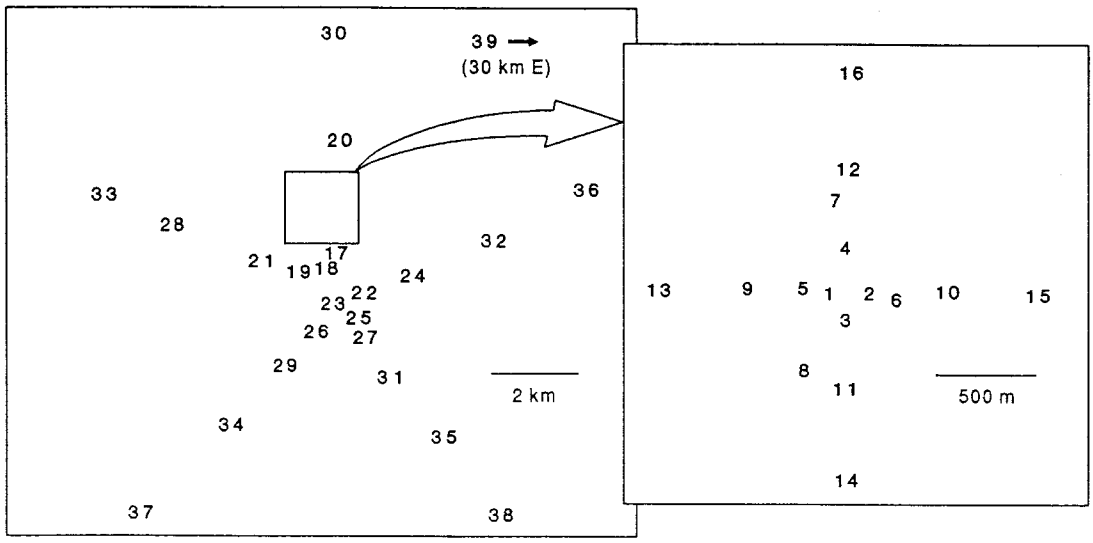

Fig. 4. Sampling design for a macrobenthic community study around the Ekofisk oil field, North Sea. Note that sites have been renumbered from Gray *et al.* (1990), in order of distances from the current centre of drilling activity.

North Sea (Gray *et al.* 1990). The design consisted of several radial transects stretching up to 4 km away from the field, with more concentrated sampling within 1 km of the main drilling activity in preceding years. A single site 30 km distant was also sampled. Figure 4 shows the sampling locations; note that to aid the later displays the sites have been renumbered from Gray *et al.* (1990), the new site numbers being in order of distance from the centre of the oil-field (here defined as the active 2/4B&K rig complex).

The data consist of abundance for 209 species for three replicate grabs from each of the 39 locations. Totalling across the replicates, root-transforming counts and computing Bray–Curtis similarities leads to the MDS ordination of Fig. 5. The MDS stress is acceptably low (see the later discussion) and the interpretation is very clear. There is a gradation of community change from the distant to central sites with, for example, all sites at a distance of around 3 km or more from the centre (including the 30 km reference) falling together at the left hand side of the plot, and clearly distinguishable from sites at intermediate distances (*ca* 1–3 km). This consistency of pattern is all the more remarkable because it is not at all evident from viewing the original data matrix and is not detectable in simple summary measures such as diversity indices. This is a good illustration of the sensitivity of multivariate analyses (e.g. Warwick & Clarke 1991).

The caveats in the above discussion on sampling design also apply here. There is no proof that the drilling activity is causal to the change in the benthic community but the good design allows a strong prima facie case to be made. To deny such causality, one has to invoke less credible alternative hypotheses to explain such 'concentric circles' of differing community pattern surrounding the oil field, with the most geographically dispersed sites showing no more community variation (and arguably less variation, Warwick & Clarke 1993b) than the closely located samples at the centre.

Such alternative hypotheses exist, of course, and are most satisfactorily refuted by repeating the 'experiment' in an independent setting, that is, repeating the treatment. This is an unlikely scenario for many environmental impact studies (Underwood 1992) but does pertain to oil field monitoring in the North Sea. Gray *et al.* (1990) described a parallel community analysis for the Eldfisk drilling complex, a distance of 10 km from Ekofisk. This was a smaller study involving 'cross-hair' transects supplemented by additional samples near to the centre of activity; the 20 sites showed a similar pattern of community change with decreasing distance from the rigs, albeit with contours which were slightly more elliptical. The same pattern is also apparent from recent analysis of other North Sea fields (J. S. Gray, pers. comm.), and this is clearly a case where a series of purely observational spatial studies is capable of building up decisive evidence of causality.

This conclusion needs to be circumscribed in two ways. First, there is no statement here about the causative mechanism. Several features of oil field activity could be responsible for a community impact, for example toxic effects from contami-

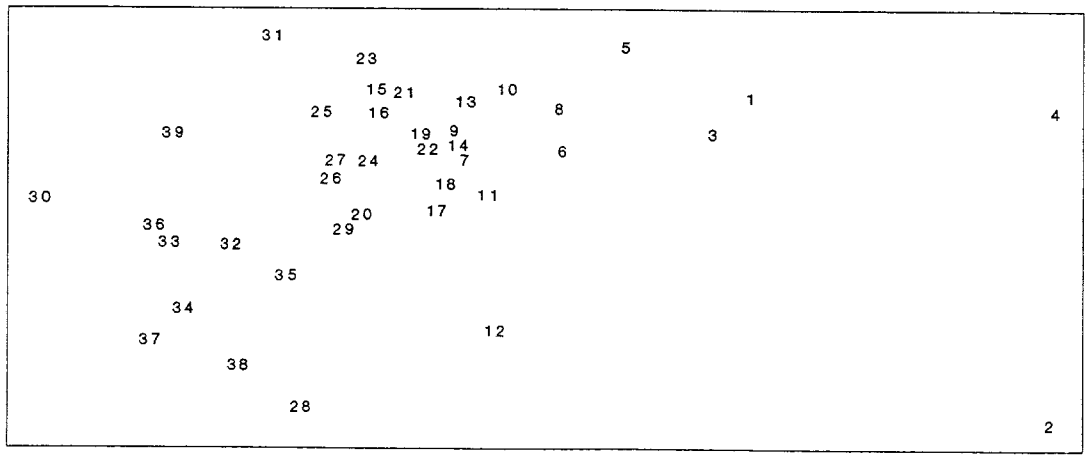

Fig. 5. MDS plot for macrobenthos at 39 sampling sites (Fig. 4) around the Ekofisk oil field. Note the strong gradation of community pattern with increasing distance (site number) from the centre of drilling activity. Stress = 0.12.

nants in the drilling muds, or simply their modification of the grain size structure of sediments. Such variables are certain to be confounded with each other, to some extent. (The question of whether more progress can be made in differentiating correlative links with particular contaminant variables is returned to later.) One should distinguish between use of the term 'causality' to imply that a particular event or activity is responsible for an observed change, and its use in defining the mechanism by which this change is engineered. Second, and perhaps tritely, while the sensitivity of the multivariate approach has here provided solid evidence of biological change, there can be no imputation from this analysis alone that the effect is deleterious. Such judgements are outside the scope of this paper, though the issues are touched on in the companion paper (Warwick 1993).

Practical experiences with the MDS algorithm

Experience in ecological application of non-metric MDS over the last decade has highlighted a number of technical issues that are sometimes overlooked.

(1) An MDS configuration can be arbitrarily rotated, reflected or expanded, explaining the absence of axis annotation in plots such as Figs 3 and 5. This is clear from the world map example; exploiting only knowledge of which two cities are closest, which two are next closest et cetera there can be no information on orientation or units of measurement. What is not arbitrary is the *relative* position of points in the final plot. The arbitrariness of orientation can be a practical nuisance when comparing different ordinations with the same label set, and it can be helpful to rotate an MDS configuration so that its direction of maximal variation always lies along the x axis. (This is simply achieved by applying principal component analysis [PCA] to the two-dimensional MDS configuration; this is not the same thing as applying PCA to the original data matrix of course.)

(2) The nature of the iterative algorithm, successively refining an initially arbitrary configuration of samples, will often generate solutions which are sub-optimal. It is imperative to repeat the iterative process a reasonable number of times, typically eight or nine would be advisable, and to check that the same (lowest) stress is achieved in several of the repeats. (Configurations with the same stress value, to three decimal places, are almost always identical.)

(3) Degenerate solutions can occur, in which groups of samples collapse into single points on the MDS plot. Sometimes this can be a genuine artefact and will not be found in repeat iterations; more often however it *is* repeatable and results from a total disjunction in the data. If the data divide into two groups, which have absolutely no species in common, then there is clearly no yardstick for determining how far apart the groups should be placed in the MDS plot. They are infinitely far apart, in effect, and it is not surprising to find that the samples in each group then collapse to a point. The solution is to split the data and carry out an ordination separately on each group. In fact, this sequential approach could be taken beneficially in less extreme situations. If an initial MDS shows the samples to be strongly clustered, and the individual groups are quite large, then separate ordinations for each group are likely to reveal the fine structure more accurately. Such an approach may be essential when the total number of samples is very large. MDS on much more than 100 samples is not only rather computation intensive — though ever-increasing computing power is rendering this a less important consideration than it once was — but more importantly it is unlikely to reveal a clear and reliable pattern. In general, the greater the number of samples the harder it will be to reflect the complexity of their inter-relationship in a two-dimensional plot, whatever the ordination technique employed. An initial cluster analysis might then form the basis for separate ordinations of two or three major groups.

(4) All ordination methods are a compromise; inherently high-dimensional data are being viewed in a lower-dimensional (often two-dimensional) plot. It can be claimed that non-metric MDS makes the best possible job of preserving among-sample relationships accurately in a low-dimensional picture; that is, after all, its *raison d'être*. Nonetheless, it is important to assess how well it succeeds in any specific case and modify interpretation accordingly. The simplest indicator is the stress value. Traditionally (Kruskal & Wish 1978), this is examined for MDS solutions in a range of dimensions; as the dimensionality increases, a sudden drop in the stress value indicates that a valid configuration has been found. In practice, a clear shoulder in the

graph of stress versus dimensionality is rarely seen for ecological data, and practical experience suggests the following rule of thumb for interpreting Kruskal's stress formula 1.

Stress <0.05 gives an excellent representation with no prospect of misinterpretation.

Stress <0.1 corresponds to a good ordination with no real risk of drawing false inferences. A higher-dimensional plot is unlikely to add to the overall picture, though in a strongly clustered situation the fine structure of individual groups might bear separate examination.

Stress <0.2 can still lead to a usable picture, although for values at the upper end of this range there is potential to mislead; too much reliance should not be placed on the details of the plot — a higher-dimensional solution could show a somewhat different picture.

Stress >0.2 is likely to yield plots which could be dangerous to interpret. Certainly by the time stress reaches 0.35–0.4 the samples are effectively randomly placed, bearing little relation to the original similarity ranks. (Such large values can also be generated by the user inputting similarities to a routine that expects dissimilarities, or vice versa, and the MDS plot will then tend to string the samples around the circumference of a circle.)

These guidelines are over-simplistic. For example, stress tends to increase with increasing numbers of samples. Also, it makes a difference to interpretation if contributions to the stress arise roughly evenly over all points or if most of the stress results from difficulty in placing a single point in the two-dimensional picture (as in Fig. 1). The latter problem can be identified by closer study of the Shepard diagram (in Fig. 2 many of the points furthest from the monotonic regression line are from distances involving San Francisco) and by noting that, in repeated runs of the algorithm, several of the near-optimal solutions are identical to the optimal configuration, except that one point has moved to a quite different position.

(5) A useful approach in cases of non-negligible stress is to check reliability of interpretation by superimposing underlying or complementary information onto the ordination. For example, one could connect all points whose corresponding similarities are ranked in the top 10 or 20%. A less-than-faithful preservation of these rank similarities in the MDS would be evidenced by unnatural connections. A more sophisticated variant of this is to superimpose the minimum spanning tree (Gower & Ross 1969). A third possibility, recommended by Field *et al.* (1982), is to indicate groupings from a cluster analysis in the MDS plot. Again, unnatural groupings suggest that among-sample relationships are more complex than can be accurately portrayed in a two-dimensional configuration. Warwick *et al.* (1988) gave a detailed example of this, contrasting the relative success of MDS and PCA in representing meiobenthic community response to contaminant dosing in a mesocosm experiment.

(6) If such techniques show that a two-dimensional picture is an inadequate summary, one may occasionally be able to divide the data into subsets that are capable of accurate portrayal; it is more likely though that a three- or higher-dimensional solution must be sought. Good software for three-dimensional plots is now much more widely available and could beneficially be used more often for ordinations, even where stress in two dimensions is not particularly large (Warwick & Clarke 1993a give a recent example).

Clustering on rank similarities: Estuarine nematode communities (Exe)

To maximize the effectiveness of superimposing clusters on an MDS, one requires the cluster analysis to exploit the same information as the ordination; inconsistencies between the two displays can then be attributed unambiguously to the inadequacy of a two-dimensional description. In keeping with the earlier rationale, this suggests that group-average clustering be performed on the ranks of the similarities. For n samples, the ranks are just the integers $1,2, \ldots, n(n-1)/2$, although these will usually need some adjustment for ties (by simple averaging). Depending on the clustering software used, they also need initial rescaling to lie in the interval 0–100%. There is no practical difficulty with all of this. (Although theoretically, as D. P. Faith, pers. comm. 1992, has pointed out, a closer parallel still to the advocated ordination technique would be a clustering algorithm which minimized a stress value computed between the original dissimilarities and the metric relations represented by the dendrogram, the iterative algorithm being constrained only by the rank ordering of the dissimilarities. The simpler idea of perform-

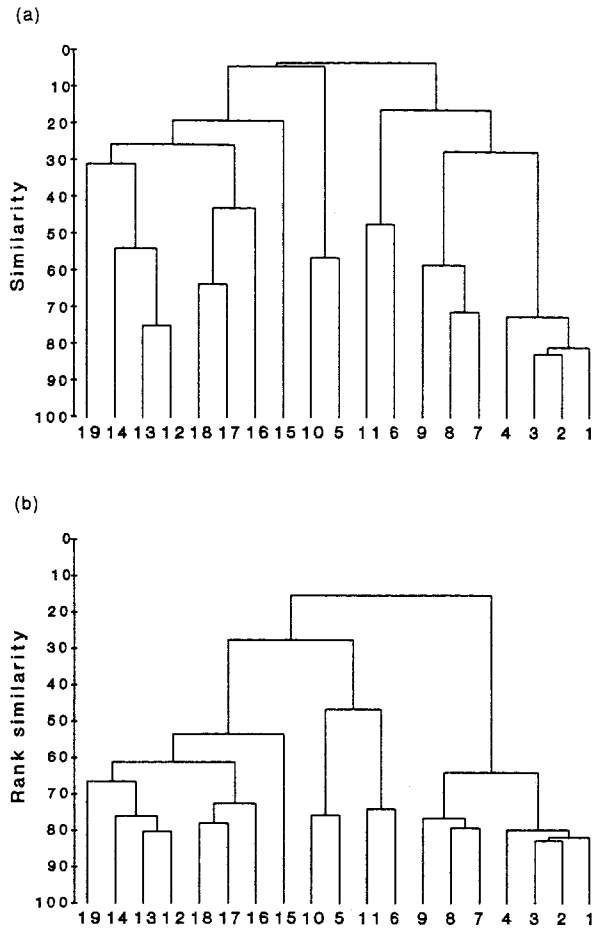

Fig. 6. Hierarchical agglomerative clustering of nematode communities from 19 intertidal sites in the Exe estuary, UK, using group-average linking on: (a) the Bray–Curtis similarities themselves and (b) only the ranks of the Bray-Curtis similarities (rescaled).

ing standard clustering on the rank similarities can be thought of as an approximation to such a procedure.)

The practical consequences of replacing similarities by their ranks, prior to group-average clustering, are not expected to be great. An example is shown in Fig. 6, for the data used as an illustration by Field *et al.* (1982). Warwick (1971) sampled nematode communities in intertidal sediments at 19 locations in the Exe estuary, UK; this was not a pollution study but a basic investigation of meiofaunal community pattern in relation to differing environmental conditions. Figure 6a is the original group-average cluster analysis, based on Bray–Curtis similarities from fourth-root transformed abundances of 182 species, and Fig. 6b is the modified dendrogram using only ranks of these similarities. Note that in order to concentrate attention on the significant feature of these two figures, viz the group structure, the rank similarities have been rescaled to the same range as the original Bray–Curtis values (cluster groupings are invariant to a scale or location change in the similarity matrix). As expected, the dendrograms are similar though not identical, the (5, 10) and (6, 11) groups now joining together at a late stage in the analysis, instead of their previous attachment to separate groups. Rather than signalling any significant change in interpretation, this is more a reflection of the arbitrariness of cluster analysis, and its necessity to form hierarchical clusters come what may. The MDS for this data (see the later Fig. 13a) confirms the structure of tight groups (1–4), (7–9), (5, 10), (6, 11) et cetera but shows these groups to be fairly evenly spaced. It is then something of a hair-line decision as to how these groups combine. This is precisely the reason why the continuum of an MDS ordination is preferred to the discreteness of a cluster analysis, provided the MDS plot has low enough stress to be reliable, as is the case here.

DETERMINING SPECIES RESPONSIBLE FOR SAMPLE GROUPINGS

Discriminating between two groups using Bray–Curtis dissimilarities

In keeping with the rationale of the opening section, it is consistent to assess the role of individual species only through their contributions to the among-sample similarity coefficients. When a cluster analysis divides a set of samples into (say) two clear-cut groups, it may be important to know which species are contributing principally to this division.

The Bray–Curtis dissimilarity δ_{jk} between any two samples j and k can be defined as

$$\delta_{jk} = \Sigma_{i=1}^{p} \delta_{jk}(i) \qquad (1)$$

where

$$\delta_{jk}(i) = 100|y_{ij} - y_{ik}| / \Sigma_{i=1}^{p}(y_{ij} + y_{ik}) \qquad (2)$$

y_{ij} is the (transformed) abundance of the ith species in the jth sample and p is the number of species. There is no unambiguous partition of δ_{jk} into contributions from each species, since the standardizing term in the denominator of equation (2) is a function of all species values, but one natural defi-

nition would be to take $\delta_{jk}(i)$ as the 'contribution of the ith species' to δ_{jk}. Averaging δ_{jk} over all sample pairs (j, k), with j in the first group and k in the second, gives the overall average dissimilarity, $\bar{\delta}$, between groups 1 and 2. The same averaging taken over each δ_{jk} (i) gives the average contribution, $\bar{\delta}_i$, from the ith species to this overall dissimilarity $\bar{\delta}$.

Typically there are many pairs of samples (j, k) making up the average $\bar{\delta}_i$, and a useful measure of how consistently a species contributes to $\bar{\delta}_i$ is the standard deviation, $SD(\delta_i)$, of the $\delta_{jk}(i)$ values. If $\bar{\delta}_i$ is large and $SD(\delta_i)$ small (and thus the ratio $\bar{\delta}_i/SD(\delta_i)$ is large), then the ith species not only contributes much to the dissimilarity between groups 1 and 2 but also does so consistently; it is thus a good discriminating species. (Note that, while $SD(\delta_i)$ is a permissible and convenient measure of variation here, the $\delta_{jk}(i)$ values are not independent and one cannot, for example, use a conventional mean-to-SD ratio to test if the average contribution of the ith species is effectively zero.)

For the Exe nematode data of Fig. 6, Table 1 shows which species contribute most to the dissimilarities between two of the groups, the clusters termed 1A (sites 1–4) and 1B (sites 7–9) by Field *et al.* (1982). The first two columns give the abundances for each species, averaged across the sites making up groups 1A and 1B (although note that this is an average of untransformed values and the dissimilarity computations are based on fourth-root transformed data). Table 1 is then ordered by the values in the third column, the decreasing contribution, $\bar{\delta}_i$, to the total dissimilarity $\bar{\delta}$ (=72.6).

It can be seen that many species play some part in determining the dissimilarity between the two groups, and this is typical of such analyses. In this tightly clustered situation, it is no surprise to find that the principal contributions come from species that are abundant in one group and largely absent (though not necessarily totally absent) from the other; the balance of contributions in this case is from species that are numerous in 1B but rare in 1A. Note the inconsistent contribution of certain species, such as *Oxystomina elongata* (high $SD(\delta_i)$ and low mean-to-SD ratio), implying that this would not be a very useful discriminating species for the two groups.

The final column in Table 1 cumulates the contributions, having rescaled the $\bar{\delta}_i$ to percentages of

Table 1. Average abundance $(\bar{y})$ of important nematode species in groups 1A (=1–4) and 1B (=7–9) of Exe estuary sites. Species are listed in order of their contribution $(\bar{\delta}_i)$ to the average dissimilarity $\bar{\delta}$(=72.6) between the two groups, with a cut-off when the cumulative per cent contribution $(\Sigma\bar{\delta}_i\%)$ to $\bar{\delta}$ reaches 70%

Species	$\bar{y}_{1A}$	$\bar{y}_{1B}$	$\bar{\delta}_i$	$SD(\delta_i)$	$\bar{\delta}_i/SD(\delta_i)$	$\Sigma\bar{\delta}_i\%$
Hypodontolaimus ponticus	140.0	0.5	3.9	1.2	3.2	5.3
Axonolaimus spinosus	0.0	60.8	3.7	0.8	5.0	10.4
Adoncholaimus fuscus	55.0	0.0	3.4	0.6	6.3	15.2
Viscosia viscosa	84.0	1.0	3.3	1.0	3.2	19.7
Sphaerolaimus balticus	38.0	0.0	3.2	0.9	3.7	24.2
Axonolaimus paraspinosus	58.0	0.3	3.0	1.0	2.9	28.3
Oxystomina elongata	145.0	0.0	2.8	3.2	0.9	32.2
Hypodontolaimus geophila	0.0	38.8	2.7	1.1	2.6	36.0
Tripyloides gracilis	58.7	0.0	2.5	1.9	1.3	39.4
Daptonema oxycerca	96.7	3.5	2.5	0.5	5.2	42.8
Monhystera sp	8.0	0.0	2.2	0.2	9.4	45.8
Praeacanthonchus punctatus	15.0	0.0	2.1	0.9	2.5	48.7
Adoncholaimus thalassophygas	0.0	7.3	2.0	0.6	3.3	51.5
Anoplostoma viviparum	11.3	122.3	2.0	1.0	1.9	54.2
Microlaimus robustidens	3.0	0.0	1.8	0.3	5.6	56.7
Enoploides spiculohamatus	2.3	0.0	1.7	0.4	4.0	59.0
Ascolaimus elongatus	12.0	0.0	1.6	1.3	1.2	61.1
Leptolaimus papilliger	0.7	7.5	1.5	0.9	1.7	63.3
Daptonema flevensis	0.0	3.5	1.4	1.0	1.4	65.2
Sabatieria pulchra	128.7	47.5	1.4	1.1	1.3	67.1
Rhabditid sp a	0.0	3.0	1.3	0.9	1.5	68.9
Halichoanolaimus robustus	2.7	0.0	1.3	1.0	1.3	70.7

$\bar{\delta}$. It can be seen that only part of the analysis is tabled, up to the point where 70% of the total dissimilarity is accounted for. This has involved 22 species and there are a further 27 species which contribute the remaining 30% of the dissimilarity (the other 133 species are absent from both groups so contribute nothing).

Typicality of species within a group

In much the same way, though with less practical significance, one can examine the contribution each species makes to the average similarity *within* a group, $\bar{S}$. The average contribution of the ith species, $\bar{S}_i$, is defined by taking the average over all pairs of samples (j, k) within a group, of the ith term $S_{jk}(i)$ in the alternative definition of Bray–Curtis similarity:

$$S_{jk} = \Sigma_{i=1}^{p} S_{jk}(i) \quad (3)$$

where

$$S_{jk}(i) = 200 \min (y_{ij}, y_{ik}) / \Sigma_{i=1}^{p} (y_{ij} + y_{ik}) \quad (4)$$

(It may not be immediately apparent, but can be simply demonstrated, that $S_{jk} = 100 - \delta_{jk}$, that is similarity and dissimilarity add to 100, the common convention.)

The more abundant a species is within a group the more it will tend to contribute to the intra-group similarities. It typifies that group if it is found at a consistent abundance throughout, so the standard deviation SD(S_i) of its contribution is low, and the ratio $\bar{S}_i$/SD(S_i) high.

Table 2. Average abundance ($\bar{y}$) of important nematode species in group 1A of Exe estuary sites, and their contribution $\bar{S}_i$ to the average similarity $\bar{S}$ (=76.8) within the group

Species	$\bar{y}_{1A}$	$\bar{S}_i$	SD(S_i)	$\bar{S}_i$/SD(S_i)	$\Sigma\bar{S}_i$%
Anoplostoma viviparum	122.3	12.0	1.5	8.1	15.6
Daptonema setosa	60.8	11.5	1.4	8.3	30.6
Axonolaimus spinosus	60.8	10.8	2.5	4.3	44.6
Sabatieria pulchra	47.5	8.6	1.5	5.7	55.8
Desmolaimus fennicus	16.8	6.9	0.9	7.4	64.7
Hypodontolaimus geophila	38.8	6.0	1.6	3.9	72.6
Adoncholaimus thalassophygas	7.3	5.1	1.3	3.8	79.2
Daptonema oxycerca	3.5	5.1	0.5	10.4	85.8
Leptolaimus papilliger	7.5	5.0	0.7	7.4	92.3

Table 3. Species contributions to similarity within group 1B of Exe estuary sites

Species	$\bar{y}_{1A}$	$\bar{S}_i$	SD(S_i)	$\bar{S}_i$/SD(S_i)	$\Sigma\bar{S}_i$%
Daptonema oxycerca	96.7	5.9	0.3	21.4	9.4
Viscosia viscosa	84.0	5.6	0.4	13.9	18.4
Sabatieria pulchra	128.7	5.5	1.5	3.7	27.2
Hypodontolaimus ponticus	140.0	4.9	1.1	4.4	35.1
Adoncholaimus fuscus	55.0	4.2	0.8	5.5	41.8
Sphaerolaimus balticus	38.0	4.0	1.3	3.1	48.1
Axonolaimus paraspinosus	58.0	3.7	0.5	7.6	54.1
Daptonema setosa	20.7	2.8	0.6	4.9	58.6
Monhystera sp	8.0	2.8	0.2	15.0	63.1
Anoplostoma viviparum	11.3	2.8	0.6	4.7	67.6
Microlaimus robustidens	3.0	2.4	0.3	9.0	71.5
Desmolaimus fennicus	7.3	2.2	0.3	7.8	75.0
Praeacanthonchus punctatus	15.0	2.1	0.3	6.8	78.3
Enoploides spiculohamatus	2.3	2.1	0.3	6.8	81.7
Tripyloides gracilis	58.7	1.5	2.5	0.6	84.0
Metachromadora remanei	3.7	0.9	1.6	0.6	85.5
Halichoanolaimus robustus	2.7	0.9	1.5	0.6	86.9
Paracanthonchus multitubifer	2.7	0.8	1.4	0.6	88.2
Oxystomina elongata	145.0	0.8	1.4	0.6	89.5
Trefusia longicaudata	1.3	0.8	1.4	0.6	90.8

Table 2 gives the species breakdown of similarities for group 1A of the Exe estuary nematode study. The first column is again a simple mean abundance within the group and the table is ordered by the decreasing values of its second column, the contributions $\overline{S}_i$ to the total similarity $\overline{S} = 76.8$. The table is again incomplete, the first column showing that only nine species contributed 90% of this similarity. Table 3 is the equivalent information for group 1B, where there are now more species contributing to the total similarity $\overline{S} = 62.6$. It is worth noting again that the species with the highest average abundance, *Oxystomina elongata*, contributes little because it is not found consistently within the group. Additionally, note that typicality within a group does not guarantee that a species is a good discriminator between two groups. One example is *Sabatieria pulchra* which features quite high in both Tables 2 and 3 but less so in Table 1; it is abundant in both groups 1A and 1B and will therefore distinguish poorly between them.

These similarity/dissimilarity breakdowns (termed the 'similarity percentages' or SIMPER procedure) have been carried out for the other groups in the Exe nematode study, including comparisons made for these data in Field *et al.* (1982), by the 'I-test' procedure. (The latter does not lend itself to small numbers of samples, so the comparison between groups 1A and 1B was not performed by Field.) The conclusions bear out the remarks in the Introduction, that the I-test approach is biased towards rarer species and does not highlight the range of species responsible for defining the clustering pattern, as seen in the SIMPER analyses.

TESTING FOR SPATIAL AND TEMPORAL DIFFERENCES

Impact of coral mining on reef-fish communities (Maldives)

In the previous section, the emphasis was on a posteriori grouping of the samples in examining which species are principally responsible for an observed clustering. In other situations the grouping may be an a priori one and the need is for a formal statistical framework within which to test hypotheses about differences in community structure between groups. The simplest examples are of 'one-way layouts', to use the terminology of analysis of variance (ANOVA).

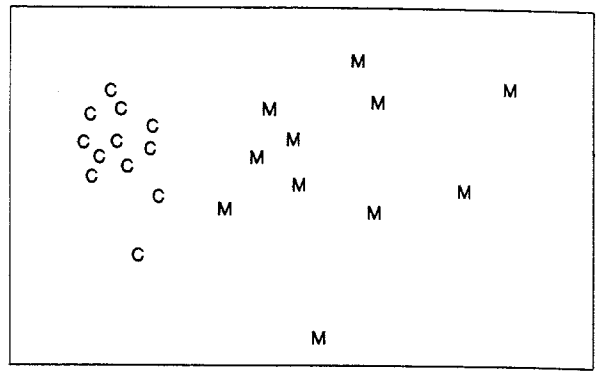

Fig. 7. MDS ordination showing a clear distinction between reef-flat fish communities from 11 mined (M) and 12 control (C) sites in the Maldive Islands. Stress = 0.08.

Figure 7 is an MDS plot based on abundance data for reef-flat fish communities (152 species) recorded at 23 sites in the Maldives (Dawson-Shepherd *et al.* 1992). The purpose of the study was to examine if there were detectable impacts on the fish communities from the widespread mining of reef corals, for building materials, that takes place around certain of the Maldive Islands. A number of reef sites were therefore selected from mined areas (denoted M in Fig. 7) and a roughly equal number of non-mined sites were chosen to act as controls (C in Fig. 7). The effect of coral mining is to reduce the complexity of the reef habitat yet, surprisingly, the mean Shannon diversity of the fish communities does not differ between mined and control sites (Dawson-Shepherd *et al.* 1992). By contrast, the multivariate analysis (Fig. 7) shows a clear-cut difference in community structure with a near-total separation of mined and control sites. A statistical test is obviously unnecessary in this case but there will be many analogous situations in which a formal test of the null hypothesis of 'no impact' would be desirable.

As discussed in the Rationale, such a test would be an analogue of the one-way ANOVA used for testing diversity indices, but the classical multivariate equivalent (MANOVA) applied to the full species matrix is inappropriate for many community data sets and is also not in keeping with the distribution-free stance taken by this paper. (An alternative which is sometimes suggested is inference on the ordination co-ordinates, e.g. a two-dimensional MANOVA, Faith 1990; this also has a number of

problems, such as the formal lack of independence of samples in the ordination space, the usual difficulties with making MANOVA assumptions of equal variance–covariance matrices and the arbitrariness of the choice of ordination dimensionality within which to carry out the test — see below.) A simple non-parametric procedure which avoids these problems is possible, however, and the principle is outlined for a 'biological effects' study by Clarke and Green (1988). The history of such permutation tests can be traced to epidemiological work by Mantel (1967) and the randomization principle employed to generate significance levels is due to Hope (1968). It is convenient to illustrate the details by a more borderline case than the Maldives study.

A permutation test for the one-way layout: Frierfjord macrobenthos

Gray *et al.* (1988) described a study of subtidal benthic communities at several sites in Frierfjord/Langesundfjord, in the context of the IOC Oslo Workshop on Biological Effects of Pollutants (Bayne *et al.* 1988). Figure 8a is from an extract of these data, and shows the MDS for a 12 sample × 110 species abundance matrix, in which samples have the structure of four replicates at each of three sites (denoted B, C, D). This example can be thought of as representative of many other situations in which it is desirable to establish site-to-site differences in community structure before going on to interpret those differences in terms of the biology or the environmental conditions. The null hypothesis is therefore that of 'no differences between sites'.

To examine this hypothesis, one first needs to construct a test statistic reflecting observed differences between sites, contrasted with differences among replicates within sites. From Fig. 8a, a natural choice might be to compute 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 has a number of drawbacks.

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

(2) The result would depend on whether the MDS was constructed in two, three or higher dimensions. The earlier discussion showed that 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.

(3) The configuration of B, C and D replicates in an MDS would also differ slightly if the MDS included additional samples, for example the full set of sites A–E, G in fig. 2a of Gray *et al.* (1988). It is again undesirable that a test statistic for comparing only B, C and D depends on what happens at other sites.

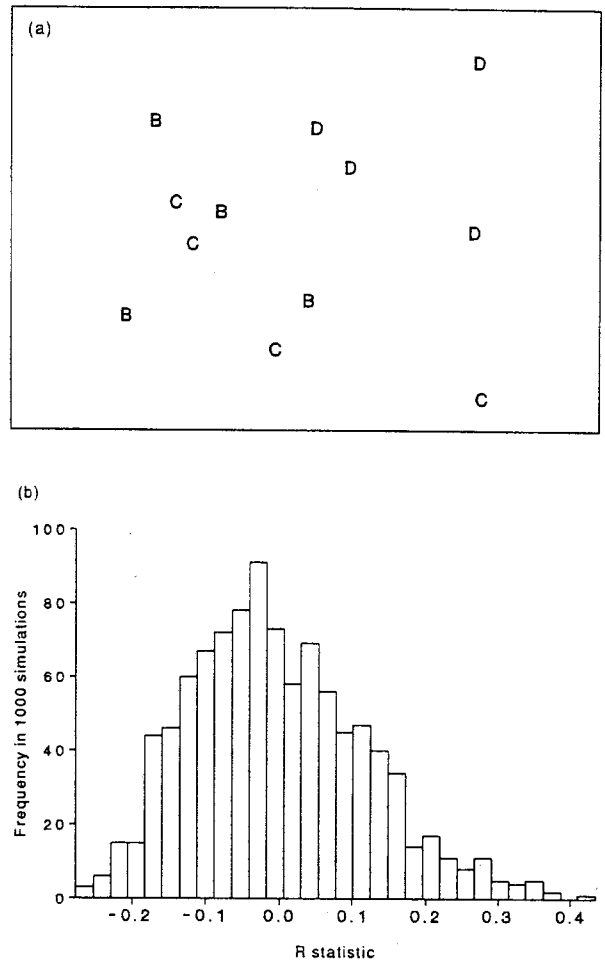

Fig. 8. Testing for benthic community differences between three sites (B,C,D) in Frierfjord, Norway, based on four replicate cores per site. (a) MDS of the 12 samples, stress=0.12 (b) Simulated distribution of the test statistic R, a difference of rank similarities between and within sites, under the null hypothesis of 'no site-to-site changes' in community structure. The true value of R, 0.45, falls above the simulated range.

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 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

$$R = (\bar{r}_B - \bar{r}_W)/(M/2) \qquad (5)$$

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 (5) has been chosen so that: (i) R can never technically lie outside the range $(-1, 1)$; (ii) $R = 1$ only if all replicates within sites are more similar to each other than any replicates from different sites; and (iii) 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. For the Oslo Workshop data of Fig. 8a, $n = 12$, $M = 6$, $\bar{r}_B = 37.5$, $\bar{r}_W = 22.7$, so that R takes the value 0.45. (This is based on the similarities only for the sites B, C, D, extracted from the matrix for all sites and re-ranked.)

R substantially less than zero is an unlikely contingency since it would correspond to similarities across different sites being higher than those within sites; such an occurrence is more likely to indicate an incorrect labelling of samples. The R statistic itself is a useful comparative measure of the degree of separation of sites, though one is often initially concerned with the simple question of whether it is significantly different from zero. (It should not be forgotten though that, 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.)

Under the null hypotheses H_0: 'no differences 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 are just replicates from a single site if H_0 is true. This is the rationale for a permutation test of H_0; 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

$$(kn)!/[(n!)^k k!] \qquad (6)$$

distinct ways of permuting the labels for n replicates at each of k sites, and the equation gives 5775 permutations in this case. It is computationally possible to examine such a 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. 8a are randomly reshuffled, R recalculated and the process repeated, say, 1000 times. In this case, the resulting spread of R values is shown in the histogram of Fig. 8b; they range from about -0.3 to just over 0.4, with a right-skewed frequency distribution. This is the range of likely values of R if H_0 is correct. The true value of R, at 0.45, is therefore seen to be an unlikely event, at least a one in 1000 chance, so the null hypothesis can be rejected with significance level $P < 0.1\%$. This is the *randomization test* principle of Hope (1968). (Note that this approach should not be confused with the more approximate technique of boot-strapping, Efron 1979, in which the *replicate data* would be resampled, with replacement. Here we are only sampling from the permutation distribution, and are doing this only to save unnecessary computation in evaluating the full set of permutations. By increasing the number of randomizations, significance levels can be determined as accurately as is necessary to demonstrate, or fail to demonstrate, rejection of H_0. In that sense it is an exact test.)

It is possible to derive some 'large sample' results (Mantel 1967) for the null frequency distribution of statistics such as equation (5), and for the approximate behaviour of their variance (Clarke 1990). In practice though, the simplest and safest approach is to test R by evaluating a large number of random rearrangements, as above. This was carried out here by a specially written FORTRAN program (ANOSIM) but would be relatively easy to implement in macro-languages of standard statistical packages.

Generality of application: Coral communities (South Tikus Island)

Though the above exposition has been in terms of testing for site differences, by referral to replicate samples within a site, it clearly is equally applicable to other one-way layouts. A more practically important case is represented by the earlier Maldives example (Fig. 7), where the replicates were the different sites, which divide into two treatment levels: mining impact and control conditions. A test of the null hypothesis of 'no effect of mining on fish communities' gives a large, positive value for R of 0.74 and, not surprisingly, the hypothesis is rejected at virtually any significance level one cares to nominate. In drawing conclusions, however, one should bear in mind that this is an observational study and not an experiment in which mining impact has been allocated to the sampling sites at random. Mining activity is concentrated in a more local region than the full geographic range covered by the control sites (Dawson-Shepherd *et al.* 1992); other factors could be responsible for the different communities in that region. This is mitigated to some extent by the observations that: (i) there are one or two control sites within the general area of mining damage, and these group with the other control sites; and (ii) the community variation among control sites appears to be smaller than among mined sites, in spite of the control sites' greater geographic spread. A simple model in which community differences were a function of spatial separation and not mining activity could generate fallacious discrimination of mined and non-mined sites but would reverse the pattern of variability observed in Fig. 7.

Temporal rather than spatial studies also generate one-way layouts, which are amenable to testing in the above fashion. A further example of coral-reef work is reported by Warwick *et al.* (1990b), this time a study of the coral communities themselves. Figure 9 displays the MDS of replicate transects across a single reef site in South Tikus Island, Thousand Islands, Indonesia, with 10 replicates from each of the two years 1981 and 1983. (The data were percentage cover of the transect by each of 58 coral species and it was not considered necessary to transform the observations prior to calculating Bray–Curtis similarities.) A change in the community pattern between the two years is evident, reflecting a coral-bleaching episode putatively linked to the 1982–83 El Nino. It is interesting to ask whether the permutation/randomization test described above has any validity or power to detect this change. This question is by no means an 'Aunt Sally'; the conventional univariate ANOVA or t-test does *not* detect a change which corresponds to a variance increase rather than a location shift. Furthermore, the standard test is invalid if the variance differs substantially between the two groups. Similar caveats apply to the 'classical' multivariate analysis of variance (MANOVA) tests, such as Wilks' lambda (Mardia *et al.* 1979), which make an assumption of equal variance–covariance matrices for the two groups, in addition to approximate normality of the species data.

By contrast, no such assumptions have been spelt out for the randomization test, and they are not required. The strength of this test is undoubtedly its simplicity and validity in almost any situation. It *will* have some power to detect the sort of change seen in Fig. 9; the R statistic is 0.43 and this is significant at the $P< 0.1\%$ level. Naturally, in cases where the strong assumptions of multivariate normality are justified, and the change is only a shift in location not variance, a classical test would be expected to have greater power and should be used — much richer inference is possible. In many practical situations, however, the distribution-free approach is likely to be more appropriate, provided sufficient attention is paid to the replication level to generate a reasonable number of possible permutations. If only two replicates are taken for each

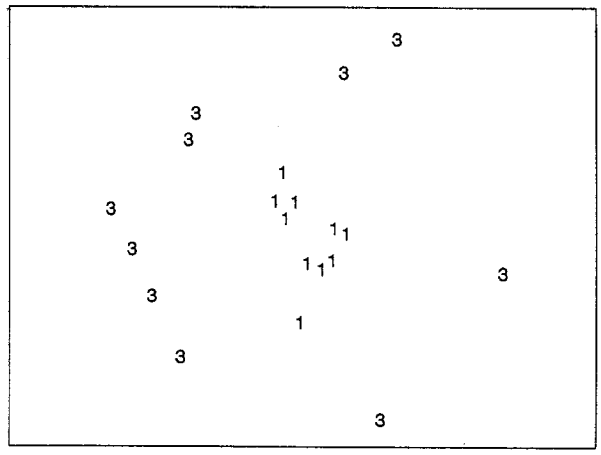

Fig. 9. MDS ordination indicating a difference in variability of coral community structure at a single site from South Tikus Island, Indonesia. There were 10 replicate transects in each of 1981 (1) and 1983 (3). Stress = 0.11.

of two groups then there are only three distinct permutations and a 5 or 10% significance level test could not be constructed. Four replicates from each of two groups (35 permutations) are needed for a 5% level test, though the number of possibilities rises steeply once there are more than two groups; for example there are 280 permutations of three groups of three replicates and thus a potential $P < 0.5\%$ level test.

Before leaving the South Tikus Island data, it is worth noting that Warwick and Clarke (1993b) discuss this example in the context of increased variability of community pattern as an indicator of pollution or disturbance impact. (Increased variability in population numbers through space or time has previously been adduced as a consequence of disturbance, e.g. by Underwood 1991.) Warwick and Clarke (1993b) propose a comparative index of multivariate dispersion which follows the rationale of this present paper, being only a function of the rank similarity matrix.

Many impact studies are not confined to one-way layouts; some recommended designs mix both spatial and temporal components and have hierarchies of spatial sampling (e.g. Green 1979, 1993; Underwood 1992). These designs have been evolved in the context of normality-based ANOVA, applied to abundance data for a single species or to a diversity measure. In the attempt to reduce the likelihood of observed changes not being causally related to the impact, some of these designs can become very complex (Underwood 1992). It is a major challenge to even begin to translate such structures into the distribution-free multivariate context of the present paper, and one that could only ever be partially successful. A first step is possible, namely some results for the two basic types of two-way layout.

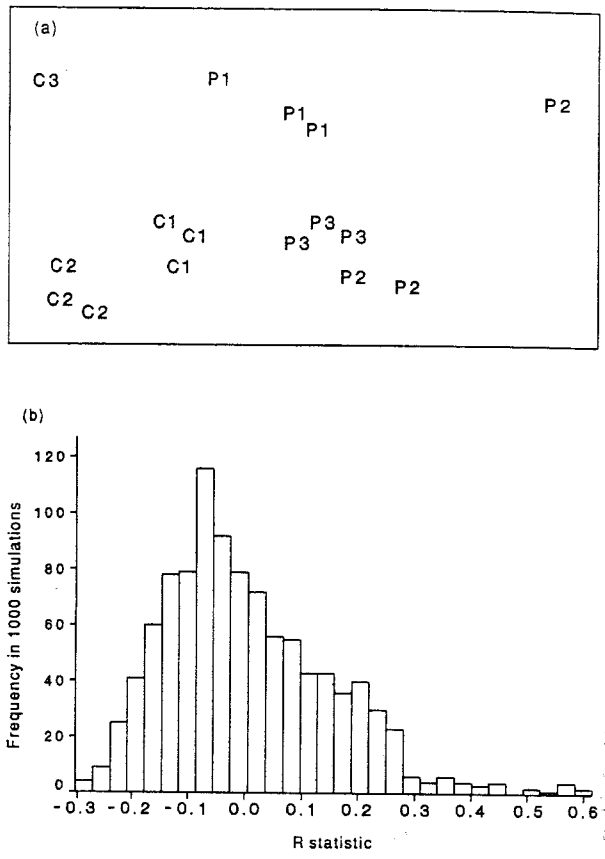

Fig. 10. Testing for community differences in a two-way nested layout of 16 nematode samples from the Firth of Clyde. (a) MDS for three 'polluted' (P1,P2,P3) and three 'control' sites (C1,C2,C3), with three replicate samples at most sites (stress = 0.09). (b) Simulated distribution of the $\overline{R}$ test statistic, under the null hypothesis of 'no difference between sites' within either condition (C or P).

Two-way nested layout: Impacts on nematode communities (Clyde)

Simple nested designs arise in spatial studies, where two levels of spatial replication are involved. For example, Lambshead (1986) analysed meiobenthic communities from three putatively impacted areas of the Firth of Clyde and three control sites, taking three replicate samples from most of the sites. The resulting MDS, based on fourth-root transformed abundances of the 113 species in the 16 samples, is given in Fig. 10a. Note that the third control site has only one replicate. The sites are numbered 1 to 3 for both conditions but the numbering is arbitrary; for example, there is nothing in common between P1 and C1. This is what is meant by sites being 'nested' within conditions. Two questions are then appropriate.

(Q1) Can we reject the null hypothesis of no difference among sites within each treatment (control or polluted conditions)?

(Q2) Can we reject the null hypothesis of no difference between control and polluted conditions? The approach to question 2 might depend on the outcome of question 1.

Question 1 can be answered by extending the one-way permutation test of the last section to a constrained randomization procedure. The presumption under question 1 is that there may be a difference between general location of C and P

samples in the MDS plot (Fig. 10a) but within each condition there cannot be any pattern in allocation of replicates to the three sites. This is patently false in Fig. 10a, so one would expect to reject the null hypothesis here. Treating the two conditions entirely separately, one therefore has two separate one-way permutation analyses of exactly the same type as for the Frierfjord macrobenthic samples (Fig. 8a). These generate test statistics R_C and R_P, say, computed from equation (5). These may be combined to produce an average statistic, $\overline{R}$, which 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 randomization is constrained to take place only within the separate conditions: P and C labels may not be switched. Even so, the number of possible permutations is large; for a balanced design it would be given by the square of equation (6). Here the design is slightly unbalanced by the loss of two replicates, so equation (6) needs some modification, but there are around 20 000 distinct permutations. Notice incidentally that there is nothing inherent in the randomization procedure, for either one-way or two-way cases, that restricts its use to balanced designs. It may even, as here, cope with some sites that are represented by a single sample, provided there are enough replicates elsewhere to generate sufficient permutations. There must be a sense in which the power of the test is weakened by the lack of balance but this is a difficult area to examine in any formal way — any definition of power requires a precise alternative hypothesis to be nominated and it is hard to see how this could be specified. The lack of balance also causes a minor complication in the definition of the average $\overline{R}$ of R_P and R_C; some minor efficiency gain will be possible if $\overline{R}$ is a weighted average in the unbalanced case. It is relatively straightforward to exploit the variance approximation given by Clarke (1990) to effect an optimal weighting, though this will not be pursued here. The present example uses a straight average of R_P and R_C.

Figure 10b displays the results of simulating the full permutation distribution for $\overline{R}$ under the null hypothesis. That is, $\overline{R}$ is computed for 1000 different (constrained) re-labellings of the points in Fig. 10a. Possible values range from -0.3 to 0.6, although 95% of the values are <0.27 and 99% are <0.46. The true value of $\overline{R}$ at 0.75 thus provides a strongly significant rejection of the null hypothesis that there are no site-to-site differences within a specific condition.

Question 2, which will usually be the more interesting of the two hypotheses, can now be examined. The test of question 1 demonstrated that there are, in effect, only three replicates (the sites 1–3) at each of the two conditions (C and P). This is a one-way layout, and the null hypothesis that there is no pollution impact can be tested by the one-way ANOSIM procedure of the previous subsection. (This is exactly homologous to the more familiar univariate nested analysis of variance, where treatments would now be tested by an F statistic on (1, 4) degrees of freedom, having discovered significant variations among sites within treatments.)

One initial decision still needs to be made, namely how best to combine the information from the three replicates at each site, so that a similarity matrix can be defined for the six new 'replicates' (sites C1–C3, P1–P3). One possibility is simply to pool the original data across the initial replicates and compute an entirely new similarity matrix. More satisfactory, however, certainly within the rationale of this paper, would be to retain dependence only on the rank similarities in the original triangular matrix, by averaging 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. Applying the one-way test to this re-ranked matrix gives $R = 0.74$. There are only 10 distinct permutations (6!/3!3!2!) in this case so that, although the true R of 0.74 is actually the most extreme value possible, the null hypothesis of no difference between control and polluted conditions is only able to be rejected at a $P \leqslant 10\%$ significance level.

These data perhaps provide a rather weak exemplar of the methodology (though to be fair to the original author, they were probably never intended to be exploited in this way). The number of sites within each of the conditions is too limited for strong inference and the spatial distribution of the sites themselves, and the rather ill-defined nature of the pollution status, are not really satisfactory for

the hypotheses here erected, so the main purpose of this example is to illustrate a methodological possibility. One scenario is not evidenced by these data: what if question 1 returns a 'no' verdict, that is we cannot reject the null hypothesis of 'no difference among sites within each condition'? There are then two possibilities for answering question 2.

(1) 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.

(2) Infer from the test results that there *are* no differences between sites, and treat all replicates as if they were separate sites, for example there would be seven replicates for control and nine replicates for polluted conditions in the example of Fig. 10a. A one-way ANOSIM procedure is then carried out on these 16 samples.

Which of these two possibilities should be pursued is to some extent an open question. Option (2) will certainly have greater power but runs a real risk of being invalid; option (1) is the conservative test and it is certainly unwise to design the study with anything other than option (1) in mind. There is little that is new, or specific to this multivariate approach, about these options; they are paralleled by the similar dilemma over whether 'to pool or not to pool' in construction of residuals in standard univariate ANOVA (e.g. Winer 1971).

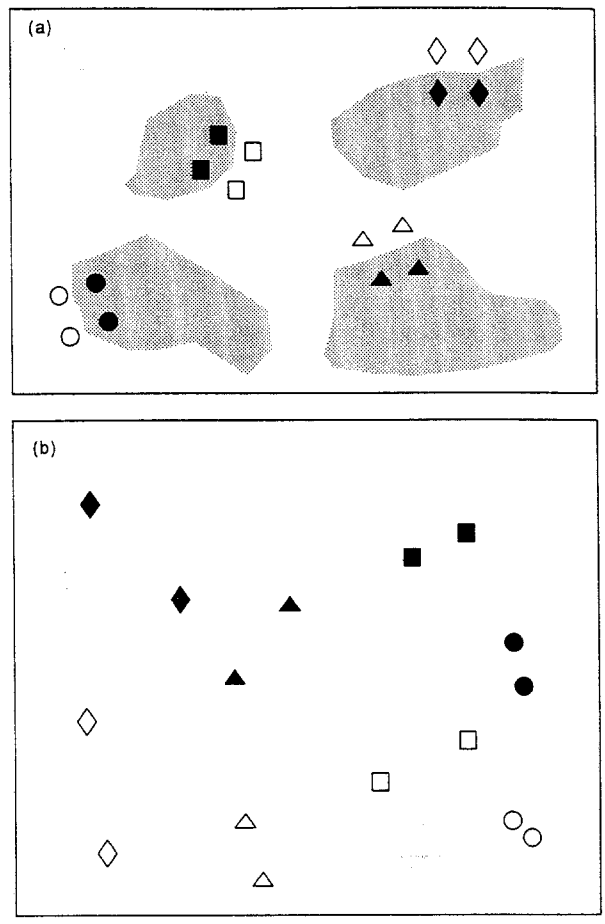

Fig. 11. A two-way crossed layout arising from disturbance of meiobenthic communities by soldier-crab burrowing, Eaglehawk Neck, Tasmania. (a) Schematic sampling design of four cores from each of four 'blocks' (disturbed patches, shaded). (b) MDS of the 16 meiofauna samples, with the *x* axis separating the blocks and the *y* axis discriminating disturbed from undisturbed communities. Stress = 0.11.

Two-way crossed layout: A natural disturbance 'experiment' (Eaglehawk Neck)

An example of a two-way crossed design is given in Fig. 11a. This is not a study of anthropogenic impact but of natural disturbance to meiobenthic communities by continual reworking of the sediment by soldier crabs (Warwick *et al.* 1990a). Two replicate samples were taken from each of four disturbed patches of sediment, and from adjacent undisturbed areas, stretching across a sand flat at Eaglehawk Neck, Tasmania; Fig. 11a is a schematic representation (rather than a detailed map) of the 16 sample locations. There are thus 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, that is 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 obvious parallels here with environmental impact studies in which (say) sewage pollution or human disturbance affects sections of several embayments, so that matched control and polluted conditions can be compared against a background of changing community structure across a wide spatial scale. In another scenario, the blocks would be sampling occasions in a time series, each point in time having replicate observations from control and polluted conditions; the objective would again be to separate the effect of natural fluctuations through time from differences associated with the pollution impact.

A feature of these designs is that there are replicate samples from all blocks for both conditions: blocks and treatments are said to be 'crossed' and the stages in this two-way layout are handled slightly differently from the previous nested case. Returning to the Eaglehawk Neck example, Figure 11b displays the MDS for the 16 core samples (2 treatments $\times$ 4 blocks $\times$ 2 replicates), based on Bray–Curtis similarities from fourth-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 (separation of symbol types on the x axis) but also differs between disturbed and undisturbed conditions (separation of closed and open symbols on the y axis). An immediate corollary is the importance of the block design in eliciting the disturbance effects; the spatial changes appear to be as great as community differences associated with disturbance.

Statistical tests to support the above conclusions are probably desirable, even in an apparently clear-cut case such as this, because of the minimal number of replicates for each block–treatment combination. A test of the null hypothesis that there are no disturbance effects, allowing for the fact that there may be block effects, can be carried out using precisely the same two-way ANOSIM procedure as initially employed in the two-way nested layout (the Question 1 test). For each separate block an R statistic is calculated from equation (5), as if for a simple one-way test for a disturbance effect. The resulting values R_1 to R_4 are then averaged to give the test statistic $\overline{R}$. Its permutation distribution under the null hypothesis is generated by examining all re-orderings of the four labels (two disturbed, two undisturbed) within each block — labels are not switched between blocks. There is no necessity to sample from the permutation distribution in this case; there are only three distinct permutations in each block, giving a total of $3^4 = 81$ combinations overall. The observed value of $\overline{R}$ for the similarity matrix underlying Fig. 11b is 0.94 and, not unexpectedly, this is the highest value attained in the 81 permutations. The null hypothesis is therefore rejected at close to a significance level of 1%.

The strict conclusion is not that soldier crab disturbance is necessarily causal to the meiobenthic community change, though this is a very plausible mechanistic explanation, but that there is statistical evidence of an association between the disturbance and community differences. So-called 'natural experiments' are still observational studies and not experiments at all, in the strict statistical sense of randomly allocating treatments to experimental units (see the previous discussion). Nonetheless, this example does demonstrate that better observational design can lead to more powerful inference.

There is a symmetry in the crossed design that is not present in the nested case. 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 in the above test. $\overline{R}$ 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 eight labels for each treatment. One must therefore randomly select from the 11 025 possible combinations. In 1000 simulations the true value of 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 one can envisage many situations where tests for both factors in a crossed design are of practical significance.

Interaction in a two-way layout: A microcosm experiment

Notice that the above two-way ANOSIM procedure is not the analogue of a test for treatment main effects in a univariate two-factor (treatments $\times$ blocks) ANOVA. Rather it is equivalent to pooling the sums of squares for main effects and interactions, and comparing this with the residual to give an *overall* test for presence of a treatment effect. In the current context, this is saying that the two-way ANOSIM test has the potential to reject the null hypothesis of 'no disturbance effect' either when (as above) there is a consistent treatment difference across all blocks or when this difference is strongly present in some blocks but not others. An example of the latter is seen in Fig. 12, for data on nematode community structure in a microcosm experiment performed by Austen (1989). Again the details and objectives of this work are of less importance here than the structure of the data and the close parallels

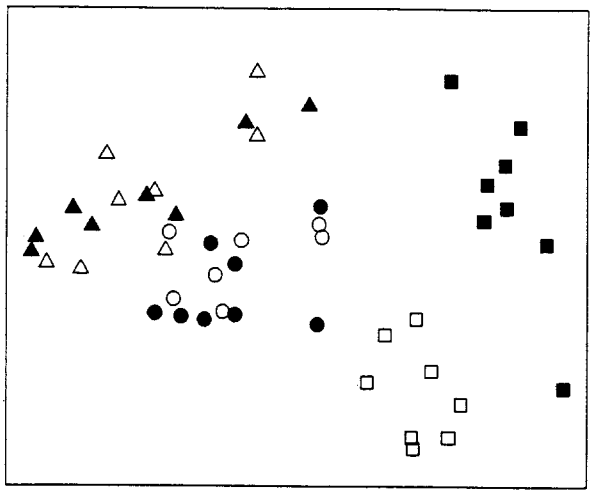

Fig. 12. A two-way crossed layout demonstrating 'interaction effects'; an MDS of nematode communities from a microcosm experiment with six independently replicated treatments, involving food limitation and osmotic stress. Stress=0.13. (△) 25‰ (control); (○) 15‰; (□) 5‰; closed symbols indicate food limited.

that can be drawn with designs for environmental impact studies.

As part of a larger series of experiments, Austen (1989) examined the effects on estuarine meiobenthic communities of prolonged food limitation (factor 1 at two levels: control and food-limited) and osmotic stress (factor 2 at three levels: control and two reduced salinity levels). Figure 12 displays the MDS of nematode communities at the termination of the experiment, based on square-root transformed abundances and Bray–Curtis similarities (Austen 1989). It is apparent that salinity had a strong effect on final community type; the two-way ANOSIM test of 'no salinity differences', allowing for possible differences from food limitation, gives $\overline{R} = 0.72$ ($P < 0.1\%$). There was also an overall effect of food limitation, allowing for the salinity differences, although the $\overline{R}$ is much smaller, at 0.32 ($P < 0.1\%$). In fact, it is clear from the plot that food limitation only had a marked effect on those communities which were subject to the greatest osmotic stress; differences were slight or negligible at the 25‰ and 15‰ salinity levels. This, of course, is an interaction effect between the two factors. Its presence can be formally established, indirectly, by noting that the one-way ANOSIM of control versus food-limited conditions, separately for each salinity level, gives: $R = -0.05$ for 25‰, $R = 0.20$ for 15‰ and $R = 0.81$ for 5‰. On an overall 5% significance level for the three tests, the first two R values are not significant but the last is — and highly so.

This indirect approach to examining interactions will be inadequate in some important practical scenarios for environmental impact studies. Green (1979, 1993), Underwood (1991, 1992) and others have discussed a class of spatial and temporal layouts known as BACI designs (Before/After, Control/Impact). A putatively impacted area and a matching control site (or sites) are monitored in a time series straddling the impact event. Pollution effects will then show up as interactions between the temporal and spatial factors. At present, such a design could be formally tested by the ANOSIM procedure only if there were no differences between control and (putatively) polluted sites prior to the impact. A general interaction effect could, of course, arise as a significant difference between control and impacted sites before the impact and a larger difference after the impact. The difficulty in extending the above methodology to cover this case is not so much in defining an interaction statistic based only on the rank similarity matrix (though that is not trivial), as in testing such a statistic using an appropriate permutation of sample labels. This would appear to defy development within the similarity-based framework of this paper and must be accepted as a limitation of the current methodology, though there is clearly scope for further study here.

LINKING COMMUNITY PATTERNS TO ENVIRONMENTAL VARIABLES

'Explaining' Exe nematode communities

The final major area of practical community studies is the attempted explanation of community patterns by linking the biotic analysis to physical or contaminant data from the same set of samples. The Introduction and Rationale outlined an approach which is designed to fit the minimalist assumptions of this paper: it uses only rank similarities between samples and attempts to avoid explicit assumptions about the form of biota–environment relationships. The basic premise appears straightforward: if the suite of physico-chemical data responsible for structuring the community were known, then samples having rather similar values for these variables would be expected to have rather similar species composition,

and an ordination based on this abiotic information would group sites in the same way as for the biotic plot. If key environmental variables are omitted, the match between the two plots will deteriorate. By the same token, the match will also worsen if abiotic data which are irrelevant to the community structure are *included*. An example is given in Fig. 13, based on the Exe nematode data used by Field *et al.* (1982).

Figure 13a shows the MDS for nematode communities at 19 intertidal sites in the Exe estuary, based on the similarity matrix employed in the earlier cluster analysis (Fig. 6), with which there is very close agreement. The remaining plots in Fig. 13 are of specific combinations of the six sediment variables recorded for each sample: the depth of the H_2S layer (H_2S), the interstitial salinity, the median particle diameter (MPD), % organics, depth of the water table and height up the shore. For consistency, these plots are MDS ordinations based on a Euclidean distance matrix from the normalized variables, though they will be virtually identical to configurations from principal component analysis (PCA, e.g. Everitt 1978) in this instance. (Fig. 13b is effectively just a scatter plot, since it involves only two variables.) In contrast to biotic data, quantitative environmental data is almost always well-handled by PCA, possibly after transformation.

The point to note here is the remarkable degree of concordance between biotic and abiotic plots, particularly Figs. 13a and c: both group the samples in very similar fashion. Leaving out MPD (Fig. 13b), the (7–9) group is less clearly distinguished from (6, 11) and one also loses some matching structure in the (12–19) group. Adding variables such as depth of the water table and height up the shore (Fig. 13d), the (1–4) group becomes more widely spaced than is in keeping with the biotic plot, sample 9 is separated from 7 and 8 et cetera and the fit again deteriorates. In fact, Fig. 13c represents the 'best fitting' environmental combination, in the sense defined below, and therefore best 'explains' the community pattern.

Quantifying the match between any two plots could be accomplished by a Procrustes analysis (Gower 1971), in which one plot is rotated, scaled

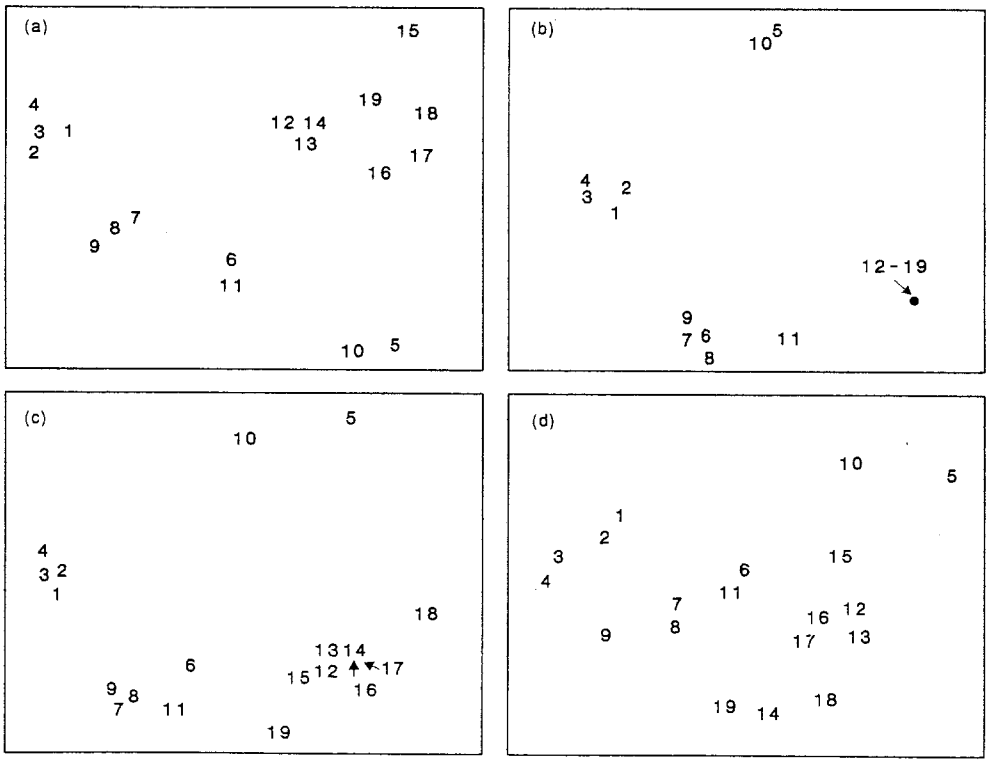

Fig. 13. MDS ordinations for the 19 Exe estuary sites, based on: (a) nematode counts; (b) sediment variables recording depth of the H_2S layer and interstitial salinity; (c) H_2S, salinity and MPD, the environmental combination 'best matching' Fig. 13a (i.e. maximizing the rank correlation of the respective similarity matrices); (d) all six recorded abiotic variables. Stress values are: (a) 0.05, (b) 0, (c) 0.04, (d) 0.06.

or reflected to fit the other in such a way as to minimize a sum of squared distances between the superimposed configurations. This is unsatisfactory, however, for exactly the same reasons as advanced earlier in deriving the ANOSIM statistic: the 'best match' should not be a function of the dimensionality in which one chooses to view the two patterns. The fundamental constructs are, as usual, the similarity matrices underlying both biotic and abiotic ordinations. These are chosen differently to match the respective form of the data (e.g. Bray–Curtis for biota, Euclidean distance for environmental variables) and will not be scaled in the same way. Their ranks, however, can be compared through a rank correlation coefficient and this is a very natural measure to adopt within the framework of this paper.

Clarke and Ainsworth (1993) describe this whole approach in detail, including the systematic search of all variable combinations to find the optimal match. (This uses a FORTRAN routine, BIO–ENV, though again it would not be difficult to implement the procedure in macro-languages of other systems.) They also contrast the use of simple Spearman rank correlation (Kendall 1970) of the two similarity matrices with a weighted Spearman coefficient, the latter placing more emphasis on matching the local rather than global structure of the biotic and abiotic patterns. For this latter coefficient (ρ_W), the combination of variables in Fig. 13c is optimal, with $\rho_W = 0.80$. It is perhaps unnecessary to repeat the warning that, as this is an observational study, one cannot infer that a combination of depth of H_2S layer, interstitial salinity and median particle diameter are directly causal in shaping the community pattern at these sites. They may, for example, be highly correlated with unrecorded variables which *are* causal, although in this instance they form a very plausible 'explanation'.

The Ekofisk oil field study

The above example was of a fundamental study in a largely unpolluted estuary, but these ideas clearly have potential for use in the monitoring of environmental impact. Clarke and Ainsworth (1993) discuss their application to macrobenthic samples from the Clyde sewage-sludge dumping ground (Pearson 1987), for which the BIO-ENV results are encouragingly informative.

A further example of an impact study, not discussed by Clarke and Ainsworth (1993), is the Ekofisk oil field data of Gray *et al.* (1990), analysed earlier (Figs 4,5). Figure 14 shows an MDS of the 39 sites based on just the three sediment variables quoted by Gray *et al.* (1990): 'total hydrocarbon concentration (THC)', 'barium concentration' and '% mud'. Barium is present in (and therefore a good tracer for) drilling muds, although it is not known to be toxic to the marine benthos; the percentage mud fraction might also be expected to reflect effects of the finer drilling muds. Barium and THC levels were initially log-transformed and all three variables normalized to improve the appropriateness of Euclidean distance as a measure of among-sample dissimilarity (Clarke & Ainsworth

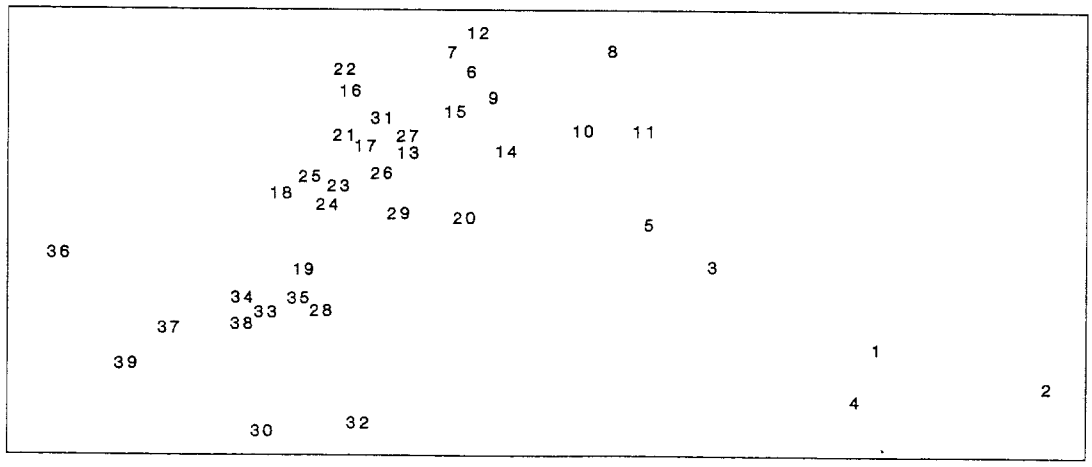

Fig. 14. MDS ordination of three environmental variables (% mud, log total hydrocarbons and log barium) for the 39 Ekofisk sites, Fig. 4. Note the fair match with the biotic ordination, Fig. 5. Stress = 0.05.

1993). The Euclidean distance matrix was then used to construct the MDS of Fig. 14 and also input to the BIO-ENV procedure. This showed that it was necessary to retain all three variables to optimize the correlation ($\rho_W = 0.59$) with the rank similarities underlying the biotic MDS (Fig. 5). The detailed patterns are fairly well matched and certainly the broad sweep of community change moving away from the oil field is well mirrored in the contaminant plot, including the marked division between intermediate and distant sites.

It must be emphasised that the BIO-ENV procedure has so far only been applied to a limited number of data sets, although with some apparent success, and needs more rigorous examination, both practically and theoretically. For example, no explicit assumptions appear to have been made about the form of the relationship between species abundance and environmental gradient, yet there could be functional forms for which this matching procedure is unsound. Simulation studies are likely to be informative here.

CONCLUDING REMARKS

The methods described here have demonstrated their utility in a variety of applications, particularly in pollution studies, where multivariate analyses have been shown to be both sensitive in their elicitation of community change and robust to a substantial degree of taxonomic aggregation (see the companion paper, Warwick 1993). The examples given here are exclusively marine, and predominantly involve soft-sediment benthos, but application of these techniques is by no means confined to marine communities. Certainly, the methods are well suited to some common features of soft-sediment data, such as a large species set (always in excess of the number of samples) and a sparse, highly-skewed abundance matrix, arising from marked spatial heterogeneities (clumping) in species distributions. These characteristics, however, are shared by many other fields of application.

The reader should be warned that not all of the methods of this paper represent received wisdom in the multivariate area. Earlier studies emphasised the role in ordination of PCA, with its assumption that sample dissimilarities are well-represented by Euclidean distances for both environmental and species data. Abundances are also assumed to be linearly related to environmental gradients. (The genesis for this is the classic statistical model of multivariate normality.) For a canonical analysis, that is the linking of community pattern to an environmental matrix, PCA leads naturally to canonical correlation, or a variant of it known as redundancy analysis (Rao 1964), but the assumptions underlying such techniques rarely seem to be satisfied for field community data. In contrast, ecologists now more often use a form of correspondence analysis for ordination, usually detrended correspondence analysis (DCA, Hill & Gauch 1980) with its implicit definition of 'χ^2 distance' as the measure of sample dissimilarity. (This has its genesis in a standard categorical data model, in which abundances are the multinomial frequencies that would be expected from sampling spatial distributions of organisms which are locally homogeneous — known technically as 'homogeneous planar Poisson processes', Diggle 1983). The matching canonical analysis, predicated on unimodal response models of species abundances across measured environmental variables, is provided by the more recent, important development of Ter Braak (1986): canonical correspondence analysis. A good description of these competing ordination and canonical methods can be found in Jongman *et al.* 1987.

Detrended correspondence analysis has attracted statistical criticism for some time, mainly for the somewhat arbitrary and 'overzealous' nature of its detrending process, but also because of the sometimes inappropriate imposition and non-robust behaviour of an underlying χ^2 distance measure (Pielou 1984; Faith *et al.* 1987; Gower 1992). This paper takes the stance that the choice of dissimilarity measure should be dictated by relevant ecological assumptions and not the mechanics of the ordination method. The fact that these measures (and even the ordination technique) may be very different for biotic and abiotic variables, and an unwillingness to constrain species–environment relationships to linear, monotonic, unimodal or even multimodal forms (in practice all four combinations may be present), has led to the proposed 'matching approach' to canonical analysis. This contrasts strongly with the direct gradient methods, in which species–environment relationships are embedded at an early stage of the analysis and will influence the observed biotic pattern. The matching procedure avoids this (arguably) undesir-

able feature, and possesses a seductive simplicity, but how well it stands up to further detailed scrutiny remains to be seen.

In conclusion, it should be stressed that the driving force for all four main sections of this paper *is* one of simplicity. The advocated non-parametric techniques may lack some of the sophistication of other multivariate methods but it is suggested that this is more than compensated for by their widespread validity and the comparative ease with which they can be understood. The latter is a major asset in communicating results from the large data sets typically arising in environmental impact studies.

ACKNOWLEDGEMENTS

The evolution of the components of this framework owes much to discussion and input from several colleagues, in particular R. Warwick and M. Carr of the Plymouth Marine Laboratory (PML). Three students, J. Hall, M. Ainsworth and C. Green, were responsible for coding FORTRAN versions of the ANOSIM and BIO-ENV procedures, to supplement the PRIMER programs (Plymouth Routines In Multivariate Ecological Research) developed at PML principally under the supervision of M. Carr. These latter programs were used for most of the analyses of this paper. Credit for the SIMPER procedure belongs to M. Carr and J. Hall. M. Austen (PML) also contributed significantly to initial discussions on several of the paper's topics. This work forms part of the Community Ecology project of the PML. My thanks are also due to A. Underwood (Institute of Marine Ecology, University of Sydney) for the opportunity to participate in the Workshop and contribute to this special issue. His suggestions, and those of D. Faith (CSIRO, Canberra), have significantly improved the paper but there remain a number of idosyncrasies that are entirely my own.

REFERENCES

Addison R. F. & Clarke K. R. ed. (1990) *Biological Effects of Pollutants in a Subtropical Environment. J. Exp. Mar. Biol. Ecol.* **138**, 1–166.

Agard J. B. R., Gobin J. & Warwick R. M. (1993) Analysis of marine macrobenthic community structure in relation to natural and man induced perturbations in a tropical environment. *Mar. Ecol. Prog. Ser.* (in press).

Austen M. C. (1989) Factors affecting estuarine meiobenthic assemblage structure: A multifactorial microcosm experiment. *J. Exp. Mar. Biol. Ecol.* **130**, 167–87.

Bayne B. L., Clarke K. R. & Gray J. S. ed. (1988) *Biological Effects of Pollutants: Results of a Practical Workshop. Mar. Ecol. Prog. Ser.* **46**, 1–278.

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

Clarke K. R. (1990) Detecting change in benthic community structure. In: *Proceedings XIVth International Biometric Conference, Namur: Invited Papers*. pp. 131–42. Société Adolphe Quetelet, Gembloux, Belgium.

Clarke K. R. & Ainsworth M. (1993) A method of linking multivariate community structure to environmental variables. *Mar. Ecol. Prog. Ser.* (in press).

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

Clifford D. H. T. & Stephenson W. (1975) *An Introduction to Numerical Classification*. Academic Press, New York.

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. PhD thesis, Université Pierre et Marie Curie.

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. *Envir. Biol. Fishes.* **33**, 367–80.

Diggle P. J. (1983) *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.

Efron B. (1979) Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **7**, 1–26.

Everitt B. (1978) *Graphical Techniques for Multivariate Data*. Heinemann, London.

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

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

Faith D. P. (1991) Effective pattern analysis methods for nature conservation. In: *Cost Effective Survey Methods for Nature Conservation* (ed. C. R. Margules and M. P. Austin) pp. 47–53. CSIRO and NSW NPWS, Canberra.

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. (1969) The use of the information statistic in the numerical classification of heterogeneous systems. *J. Ecol.* **57**, 565–9.

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.

Gee J. M., Warwick R. M., Schaanning M. *et al.* (1985) Effects of organic enrichment on meiofaunal abundance and community structure in sublittoral soft sediments. *J. Exp. Mar. Biol. Ecol.* **91**, 247–62.

Gower J. C. (1971) Statistical methods of comparing different multivariate analyses of the same data. In: *Mathematics in the Archaeological and Historical Sciences* (ed. F. R. Hodson, D. G. Kendall and P. Tautu) pp. 138–49. Edinburgh University Press, Edinburgh.

Gower J. C. (1992) Generalized biplots. *Biometrika* **79**, 475–93.

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

Gray J. S., Aschan M., Carr M. R. *et al.* (1988) Analysis of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment. *Mar. Ecol. Prog. Ser.* **46**, 151–65.

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

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

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

Greenacre M. J. (1984) *Theory and Applications of Correspondence Analysis*. Academic Press, London.

Hill M. O. (1973) Reciprocal averaging: An eigenvector method of ordination. *J. Ecol.* **61**, 237–49.

Hill M. O. (1979) *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**, 47–58.

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

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. 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**, 919–28.

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

Lambshead P. J. D. (1986) Sub-catastrophic sewage and industrial waste contamination as revealed by marine nematode faunal analysis. *Mar. Ecol. Prog. Ser.* **29**, 247–59.

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

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

Pearson T. H. (1987) The benthic biology of an accumulating sludge disposal ground. In: *Biological Processes and Wastes in the Ocean* (ed. J. Capuzzo and D. Kester) pp. 195–220. Krieger, Melbourne.

Pielou E. C. (1984) *The Interpretation of Ecological Data. A Primer on Classification and Ordination*. Wiley, New York.

Rao C. R. (1964) The use and interpretation of principal component analysis in applied research. *Sankhya A* **26**, 329–58.

Reader's Digest Great World Atlas, 2nd edn (1968) The Reader's Digest Association Ltd, London.

Ter Braak C. F. J. (1986) Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**, 1167–79.

Underwood A. J. (1991) Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Aust. J. Mar. Freshwat. Res.* **42**, 569–87.

Underwood, A. J. (1992) Beyond BACI: The detection of environmental impacts on populations in the real, but variable, world. *J. Exp. Mar. Biol. Ecol.* **161**, 145–78.

Underwood A. J. & Peterson C. H. (1988) Towards an ecological framework for investigating pollution. *Mar. Ecol. Prog. Ser.* **46**, 227–34.

Warwick R. M. (1971) Nematode associations in the Exe estuary. *J. Mar. Biol. Ass. UK* **51**, 439–54.

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

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

Warwick R. M. & Clarke K. R. (1991) A comparison of some methods for analysing changes in benthic community structure. *J. Mar. Biol. Ass. UK* **71**, 225–44.

Warwick R. M. & Clarke K. R. (1993a) Comparing the severity of disturbance: A meta-analysis of marine macrobenthic community data. *Mar. Ecol. Prog. Ser.* (in press).

Warwick R. M. & Clarke K. R. (1993b) Increased variability as a symptom of stress in marine communities. *J. Exp. Mar. Biol. Ecol.* (in press).

Warwick R. M., Clarke K. R. & Gee J. M. (1990a) The effect 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 1982–83 El Nino in the Thousand Islands, Indonesia. *Coral Reefs* **8**, 171–9.

Warwick R. M., Goss-Custard J. D., Kirby R., George C. L., Pope N. D. & Rowden A. A. (1991) Static and dynamic environmental factors determining the community structure of estuarine macrobenthos in SW Britain: Why is the Severn Estuary different? *J. Appl. Ecol.* **28**, 329–45.

Winer, B. J. (1971) *Statistical Principles in Experimental Design*, 2nd edn. McGraw-Hill, Kogakusha, Tokyo.