# A Practical Strategy for Analysing Multispecies Distribution Patterns

## J. G. Field[1], K. R. Clarke[2] and R. M. Warwick[2]

[1] Zoology Department and Institute of Oceanography, University of Cape Town, Rondebosch 7700, South Africa
[2] Institute for Marine Environmental Research, Prospect Place, Plymouth PL1 3DH, United Kingdom

ABSTRACT: A strategy is presented for analysing marine biological survey data and relating the biotic patterns to environmental data. To avoid circular argument, biotic and environmental data are kept separate. The strategy is illustrated by a worked example using data on the distribution of 182 nematode species in 107 samples in the River Exe estuary. Nineteen stations are grouped into 4 main clusters using complementary classification and multi-dimensional scaling (MDS) ordination techniques. These are both based on root-root transformed abundance data with the Bray-Curtis measure of similarity. Indicator species characterising each group are extracted using information statistics. Inverse analyses give clusters of co-occurring species which are strongly related to the station groups. Relationships of station groups to environmental variables are revealed by superimposing data for one variable at a time on the MDS plot, showing that some station groups differ in sediment granulometry and others in salinity, for example. Some of the other factors plotted show no difference between station groups. Similarly, physiognomic characteristics of the species are superimposed on the MDS plots of the inverse analysis of species groups, revealing differences in setal length and trophic status between the species groups. Finally, the 4 major station groups and species groups are related to one another in terms of morphological adaptation to the habitat.

## INTRODUCTION

Biological surveys whether of benthos, plankton or nekton, usually result in complex bodies of biotic and environmental data from which patterns and relationships need to be extracted. Although such multispecies data sets have much in common, a confusing variety of numerical techniques has been used in the marine ecological literature, often simply because a computer program happened to be handy and without consideration of its suitability for the data. Numerical techniques have been most commonly applied to benthic data (e.g. Sanders, 1958; Cassie and Michael, 1968; Lie and Kelley, 1970; Day et al., 1971; Hughes and Thomas, 1971; Popham and Ellis, 1971; Stephenson et al., 1972; Poore and Mobley, 1980; Shin, 1982; and several other recent papers). Plankton workers have also used numerical methods (e.g. Cassie, 1961; Fager and McGowan, 1963; McConnaughey, 1964; Thorrington-Smith, 1971; Angel and Fasham, 1975), and some similar analyses have been done on fish distribution (e.g. Fager and Longhurst, 1968; Peters, 1971; Haedrich et al., 1980).

In this paper we present an overall strategy for the analysis of multispecies data and the associated environmental variables which we believe has wide applicability in marine ecology. A set of robust and tested numerical techniques is presented stage by stage and illustrated by a simple example. We do not claim to review all the useful techniques available, but merely to outline numerical methods which we have successfully applied to a variety of ecological data. For a more complete review of many of the techniques see Clifford and Stephenson (1975).

Walker et al. (1979) have summarized the 3 alternative approaches to analysis of survey data:

(1) A search for patterns amongst the biological variables with an attempt to interpret these in terms of the environmental data.

(2) A search for patterns of relationship between the biotic and environmental data simultaneously, e.g. canonical analysis (Cassie, 1972).

(3) A search for patterns amongst the physical variables followed by a search for related patterns in the biotic data.

Walker et al. chose the third approach which may be

suitable in pollution surveys, or when one knows in advance which physical variables are likely to be dominant, but we prefer the first approach as a general procedure. In other words we analyse the biotic data first, 'letting the species tell their story' (Day et al., 1971) and once groups of biotically similar samples have been recognized we test the environmental variables for statistical differences, an approach also adopted by Green and Vascotto (1978) inter alia. This strategy keeps the analyses of biotic and environmental data completely separate, avoiding the influence of any previous assumptions about relationships between the biota and environment, and minimizing the danger of circular argument in seeking to deduce relationships.

## METHODS

### Normal Analysis

For simplicity we deal first with the commonest type of analysis, normal or q-type analysis, in which samples (or stations) are arranged into groups which each have a similar biotic composition. Fig. 1 summarizes the stages of analysis.

### Raw Data

The biotic data consist of a matrix in which $n$ samples (or stations) are described by $s$ species (or other taxa). Data may be categorized into 5 types:

(1) Presence-absence data, in which species are recorded as being present (1) or absent (0) in each sample.

(2) Coded abundances: semi-quantitative data coded on an arbitrary scale to indicate relative abundance (e.g. from absent, 0, present, 1, fairly common, 2, to dominant, 5; Field, 1970). Coding usually tends to have the effect of normalizing data and subsequent transformation is then unnecessary.

(3) Frequencies: The number of occurrences of each species is counted, e.g. the number of grabs or hauls in which the species was identified at a particular station. Frequency data are often obtained by reducing a large number of samples (hauls, cores, grabs) to a smaller number of stations at which they were taken.

(4) Densities: the commonest data, in which the number of individuals per sample is recorded. Densities often provide very skewed data, and transformation is normally advisable.

(5) Biomass: colonial animals and plants cannot be meaningfully counted, whereas all organisms can be expressed in units of weight. There is a danger of occasional, random inclusion of particularly large organisms which may swamp the other data, therefore transformation is usually advisable.

### Transformation

Clifford and Stephenson (1975) discuss a number of methods of transformation and standardization, the essential difference being that transformation alters the score for each species in each sample without reference to the range of scores in the rest of the data. A number of transformations are commonly used, the commonest being the logarithmic transformation:

$$Y_{ij} = \log (X_{ij} + 1)$$

where $X_{ij}$ = raw data score of the $i$th species in the $j$th
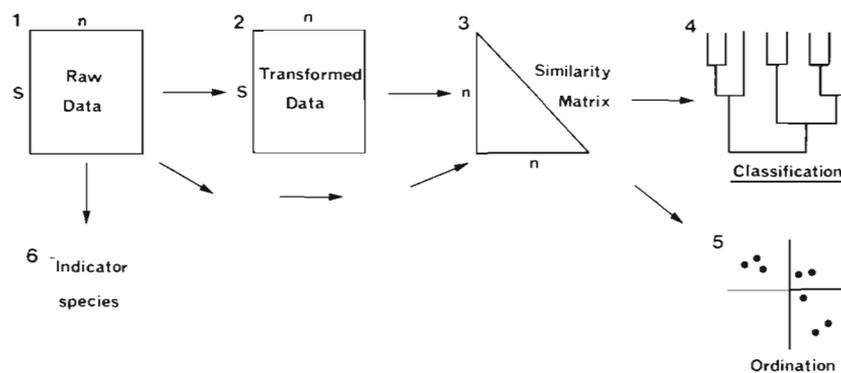


Fig. 1. Normal (q-type) analysis: diagrammatic summary of stages leading to classification and ordination of samples. Raw data (Stage 1) are represented in a matrix of $n$ samples by $s$ species. It may be necessary to transform data (Stage 2). Comparison of each sample with every other sample using a measure of similarity leads to a triangular similarity matrix (Stage 3). Classification (Stage 4) and ordination (Stage 5) are complementary pictorial summaries of the relationships between 8 samples in this diagram. 'Indicator species' are obtained directly from raw data. Stages are referred to by number in the text

sample; $Y_{ij}$ = corresponding transformed score. The log-transformation has the effect of scaling down the scores of abundant species so that they do not swamp the other data (Field and McFarlane, 1967; Clifford and Stephenson, 1975). This is usually desirable for density and biomass data.

We now prefer the 'root-root' transform:

$$Y_{ij} = \sqrt{\sqrt{X_{ij}}} = X_{ij}^{1/4}$$

The root-root transform has a similar effect in reducing the weighting of abundant species, but in addition has the advantage that, when similarity is assessed by the Bray-Curtis measure (see next section), the similarity coefficient is invariant to a scale change, e.g. it does not matter whether scores are expressed per cm² or per m². It has previously been used in this context by Stephenson and Burgess (1980).

## Measurement of Similarity

A variety of measures of distance, information, correlation, similarity and dissimilarity have been used to summarize the overall similarity between 2 samples, taking all the species into consideration. Many of these measures and their properties are summarized in Clifford and Stephenson (1975). A frequent feature of marine survey data is that many of the species are absent from a majority of the samples, so that usually more than half the data matrix entries are zeros. Transformation of the data does not alter this. Thus measures which take account of joint absences, including the product-moment correlation coefficient which is based on deviations from the mean score, are not robust enough to be generally applicable. Taking account of joint absences has the effect of saying 'estuarine and abyssal samples are similar because both lack outer-shelf species'.

We have adopted the measure used in plant ecology by Bray and Curtis (1957), after applying it to many types of data. It is not affected by joint absences (Field and McFarlane, 1968) and is therefore sufficiently robust for marine data, yet it gives more weight to abundant species (in comparing samples) than to rare ones, which is what most ecologists do intuitively. The Bray-Curtis measure has the form

$$\delta_{jk} = \frac{\sum\limits_{i=1}^{s} \left| Y_{ij} - Y_{ik} \right|}{\sum\limits_{i=1}^{s} (Y_{ij} + Y_{ik})}$$

where $Y_{ij}$ = score for the $i$th species in the $j$th sample; $Y_{ik}$ = score for the $i$th species in the $k$th sample; $\delta_{jk}$ = dissimilarity between the $j$th and $k$th samples summed over all $s$ species. $\delta_{jk}$ ranges from 0 (identical scores for all species) to 1 (no species in common) and is the complement of the similarity $S_{jk}$:

$$S_{jk} = 1 - \delta_{jk}$$

The Bray-Curtis measure is algebraically equivalent to the Czekanowski coefficient as used by Field and McFarlane (1968) and Day et al. (1971). It can be used on all types of data, on presence-absence data it reduces to the coefficient of Dice (Sneath and Sokal, 1973).

The Canberra metric (Lance and Williams, 1967) has properties similar to the Bray-Curtis measure, but equal weight is given to each species. This appears to be a desirable property when using biomass data in which there is danger of a chance occurrence of one large organism in a sample swamping all the other data. However, in practice we have never found the Bray-Curtis measure over-influenced by chance occurrences after log transformation of data, and Stephenson and his co-workers (pers. comm.) have now abandoned use of the Canberra metric in favour of Bray-Curtis. However, caution is required if the biota is very impoverished, as Stephenson et al. (1977) found after floods had swept away most animals and the samples were characterized by very few or even no species. Under these circumstances the Manhatten metric proved a suitable measure of dissimilarity, although it is normally not used because it is very sensitive to the number of species present.

Application of the measure of similarity results in a triangular matrix whose entries compare each of $n$ samples with every other sample (Fig. 1). This matrix could be arranged in a trellis diagram (Sanders, 1957) but is more conveniently summarised in diagrammatic form as a dendrogram or an ordination.

## Classification

The various hierarchical sorting strategies available to produce a dendrogram from the similarity matrix are described by Clifford and Stephensen (1975). The most successful method appears to be group-average sorting, which joins 2 groups of samples together at the average level of similarity between all members of one group and all members of the other.

Dendrograms have the advantage of simplicity: samples are clustered into distinct groups, although the cut-off levels are arbitrary and depend upon convenience. There are 4 main disadvantages to dendrograms:

(1) The hierarchy is irreversible—once a sample has been placed in a group its identity is lost.

(2) Dendrograms only show inter-group relationships; the level of similarity indicated is the average inter-group value.

(3) The sequence of individuals (here samples) in a

dendrogram is arbitrary and 2 adjacent samples are not necessarily the most similar. This can be illustrated by visualizing an upside-down dendrogram as a suspended mobile with the threads holding samples which are free to rotate in a horizontal plane.

(4) Dendrograms tend to over emphasize discontinuities and may force a graded series into discrete classes.

In view of these disadvantages, it is advisable to employ an additional method of presentation to show individual relationships. If the 2 complementary methods agree, then discontinuities can be accepted as real; if not, one has to temper the interpretation of results accordingly, but it may be helpful to use the distinct classes shown by the dendrogram to simplify description, e.g. in plotting on a map or by delineating dendrogram classes on the corresponding ordination.

## Ordination

Our preferred method of ordination is multi-dimensional scaling (MDS), a technique first derived for use in psychology and sociology (see Kruskal and Wish, 1978, for an introductory survey) but potentially of wide application to the biological sciences. In the context of sample analysis, MDS produces an ordination of the $n$ stations in a specified number of dimensions. Firstly, it interprets some function of the dissimilarity measure between each pair of stations as a distance in ordinary Euclidean space. It then seeks the best possible reconciliation (according to some optimality criterion) of these $n(n-1)/2$ inter-station 'distances' with the physical distances between $n$ points on a 2 (or higher) dimensional map. Metric MDS (essentially Principal Co-ordinates Analysis) stems from the work of Torgerson (1958); non-metric MDS, a much more flexible tool, dates to Shepard (1962), Kruskal (1964) and others. Kruskal (1977) reviews 2 of the more widely available computer programs for implementing non-metric MDS, namely M-D-SCAL and KYST. (The version M-D-SCAL 5MS has been used to derive the results for our example). Non-metric MDS is an iterative procedure with the following steps:

(1) A starting 'map' of the $n$ stations is constructed, in the required number of dimensions. This could be the result of an ordination by another method, e.g. principal co-ordinates analysis, but will often be simply an arbitrary configuration chosen at random.

(2) The interpoint distances $\{d_{jk} : k > j; j = 1, \ldots n\}$ of this configuration are then regressed on the corresponding dissimilarities $\{\delta_{jk}\}$. In the very simplest case, where there is reason to believe that the dissimilarities are directly proportional to distance, this could be by simple least squares linear regression. However, for

biological q-type analysis the relation is usually non-linear, as the scatter plot for the final configuration of the Exe estuary stations illustrates (Shepard diagram, Fig. 5). Non-metric MDS allows for this by fitting a general monotonic (increasing) regression of $\{d_{jk}\}$ on $\{\delta_{jk}\}$, the only information used from the dissimilarities being their rank order.

(3) The 'goodness-of-fit' of this regression is measured by some criterion, usually the stress formula:

$$\text{Stress 1} = \sum_{j}^{n} \sum_{k>j}^{n} (d_{jk} - \hat{d}_{jk})^2 \Big/ \sum_{j}^{n} \sum_{j>k}^{n} d_{jk}^2$$

where $\hat{d}_{jk}$ = distance estimated from the regression, corresponding to dissimilarity, $\delta_{jk}$. If stress is large, the current 'map' tallies poorly with the observed dissimilarities; conversely, low stress indicates that the sample relationships can be well represented by a station 'map' in the specified dimensionality. Stress may be thought of as the distortion involved in 'compressing' the data to a small number of dimensions.

(4) The current configuration is perturbed in a direction which decreases the stress (the method of steepest descent is usually employed here) and stages (2) to (4) repeated until no further reduction in stress is possible.

As with many multi-dimensional minimisation problems, convergence may often be to a local rather than global minimum of the stress function. Small perturbations of the present configuration may all lead to higher stress values but an entirely different configuration may have a lower stress minimum. It is therefore advisable to repeat the iteration process with several different starting configurations. If essentially the same configurations (with the same lowest stress) are achieved on a number of occasions, then the optimal solution has almost certainly been found, though this can never be guaranteed. It is important to note that the final map is only determined to within an arbitrary orientation and reflection, and arbitrary location and scale; this explains the omission of axes in Figs. 4, 6, and 8 to 10.

Step (2) highlights the advantages of MDS over techniques such as principal co-ordinates, reciprocal averaging and correspondence analysis for obtaining a simple ordination. The latter are all essentially based on the eigenvalue method of principal components (though they differ in the standardisations and transformations initially applied to the data) and are relatively inflexible, particularly with regard to the large number of zero counts generally present in a species-samples matrix, as discussed under 'Measurement of Similarity'. By contrast, great flexibility is bestowed on MDS by its less direct approach – first constructing a dissimilarity matrix to suit the particular form of the data and then allowing a general monotonic transformation to distance. In fact it is surprising that such a

precise 'map' of the relation of samples to each other can be constructed solely from information of the type 'species composition at Station A is more like that at B than at C'.

Further advantages of MDS are its ability to handle, with comparative ease, missing data, replication and data of non-uniform reliability for which it is desirable to give unequal weights to the dissimilarities in seeking the 'best' map. A definite disadvantage is the escalation of the computing problem for large numbers of stations; computer time increases proportionally to $n$ and tends to become prohibitive for 3 figure values of $n$.

## Indicator Species

Having summarized the analysis of distribution patterns in 2 complementary diagrams, classification and ordination, one has lost track of the species differences which cause the patterns. This information can be regained by returning to the raw data, or preferably to frequencies derived from the raw data.

It may be useful to know which species are characteristic of one group of samples but absent from another. Information statistic $(I-)$ tests (Field, 1969; Velimirov et al., 1977) provide a means of assessing which species differ most between one group of samples and another. Comparing Cluster 1 with Cluster 2 for any one taxon (species), $i$:

$$2\triangle I_i = 2 (I_{ti} - I_{1i} - I_{2i})$$

where $I_{ti}$ = total information content of both clusters combined, $I_{ti} = N_t \log N_t - A_{ti} \log A_{ti} - (N_t - A_{ti}) \log (N_t - A_{ti})$; $N_t$ = number of samples in both clusters together ('potential presences'); $A_{ti}$ = number of samples in which Species $i$ is actually present; $(N_t - A_{ti})$ = number of samples from which Species $i$ is absent. Similarly, the information content $I_{1i}$ and $I_{2i}$ are obtained for Clusters 1 and 2 respectively. Thus any clusters may be compared, pairwise, to see which species discriminate best between them.

Under the hypothesis that the probability of observing Species $i$ is the same for samples in both Clusters, $2\triangle I$, has an approximate chi-square distribution, and the scores of each species might be looked up in chi-square tables (1 d.f.) to see whether they differ significantly at the 5 % or 1 % probability levels. However, these are best regarded as arbitrary cut-off levels, a convenient rule of thumb for selecting indicator species but lacking statistical rigor. This is because several assumptions implicit in chi-square testing are not met (e.g. large sample conditions such as expected values exceeding 5, repeated significance tests, and the fact that classification categories have already partly been determined by the same data used in 'I-tests' to pick out discriminating species). Nevertheless, we have found it a very useful way to re-examine the data in practice and recommend its use, provided it is not regarded as a statistical test of significance.

Another technique that has been used to extract the species that differ most between sample-groups is the F-ratio (Stephenson et al., 1977; Shin, 1982) which can be used to find the species that differ most among all the sample groups simultaneously, as compared to the pairwise Information Statistic tests. The F-ratio should be applied to transformed data. Similar constraints preclude its use as a rigorous statistical test.

## Inverse Analysis

Grouping of species may be of greater interest than sample-groups in which case inverse- or r-analysis is appropriate as a complement or substitute for analysis of samples.

If presence-absence data are to be analysed either the Bray-Curtis measure or that of McConnaughey (1964) have proved satisfactory (Field, 1970), but to avoid spurious groups of rare species the number of species should be reduced. If semi-quantitative or quantitative data are used, the Bray-Curtis measure is appropriate only if the data are first standardized (W. Stephenson, pers. comm.).

A disadvantage of inverse analysis alone is that later analysis of the species groups in relation to the environmental data is more complex and simple tests for significant differences between groups are not appropriate. This can be partially remedied in a presence-absence analysis by contradicting our overall strategy and categorizing the environmental data, the classes of which can be treated as 'species' which are 'present' or 'absent' from samples and grouped into clusters along with the biotic species (Field, 1971). Steps 2–5 are followed as with normal analysis, but additional stages (7) and (8) are also required (Fig. 2).

## Data Reduction

In any survey, some species occur too seldom to form an analysable pattern. These add to computer time but they do not affect normal analysis of samples (Day et al., 1971). However, these problems become more serious when species are compared (inverse analysis) for 2 reasons. Firstly, computer time increases with the square of the number of individuals (now species) being compared; secondly, random co-occurrence of 2 rare species which only appear once in the analysis, can result in their being grouped together as having identical distributions.
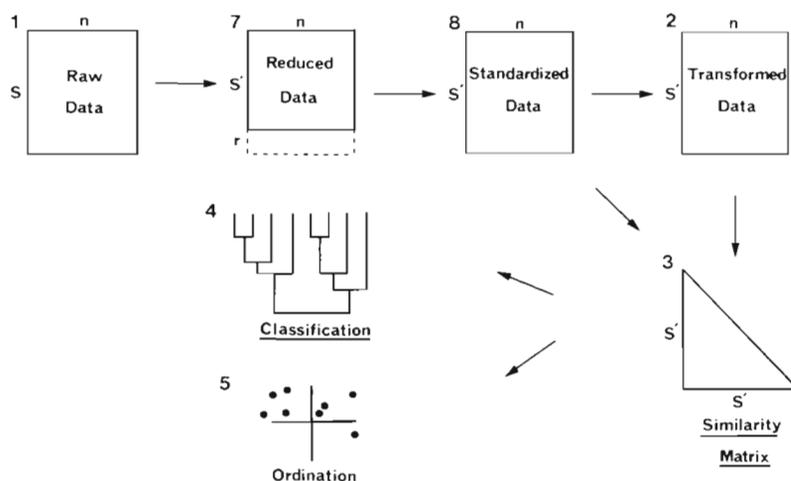
Fig. 2. Inverse (r-type) analysis: diagrammatic summary of stages leading to classification and ordination of species. Additional steps (7 and 8), reduction of the number of species and standardisation of abundances respectively, are usually required before comparing each of the s' species with every other species. The classification suggests that there are 2 main clusters of 4 species each; the ordination indicates which individual species are close or distant in distribution

Clifford and Stephenson (1975) discuss several possible ways of eliminating rare species, and Stephenson and Cook (1980) describe a method of eliminating species prior to analysis. In many analyses, as in the example below, some stations will tend to have low species diversity but high numerical abundance, and others a high diversity and low abundance. Thus, to select say the 50 overall most abundant species, biases the selection towards the species at the low diversity/high abundance stations whereas many species which may be absolutely characteristic of the high diversity/low abundance stations will be disregarded. To overcome this problem, we have selected on the basis of species having above an arbitrary percentage dominance at any one station.

Standardisation

Noy-Meir (1970, 1973) has discussed fully various types of standardization used in plant ecology, and a summary is given by Clifford and Stephenson (1975). For normal analysis, standardization is not required but it is necessary prior to inverse (species) comparisons when quantitative data are used. This is because perfectly correlated species which always occur together (e.g. host and parasite) but in different abundances might be separated from one another because their scores are not the same. The recommended standardisation for species comparison is:

$$Y_{ij} = 100\, X_{ij} / \sum_{j=1}^{n} X_{ij}$$

where $X_{ih}$ = abundance of the $i$th species in the $j$th sample; $\sum_{i=1}^{} X_{ij}$ = summed abundance of the $i$th species over all samples; $Y_{ij}$ = corresponding standardized score. This type of standardisation is also referred to as 'relativized data' (Whittaker and Gauch, 1973; Campbell, 1978).

**Relationship to Environmental Data**

Having obtained groups of samples by analysis of the biotic data, most ecologists are interested in seeking the environmental factors that are likely to be responsible for the patterns found. One approach is to do this in a separate statistical analysis of the environmental data.

A variety of techniques is available for this, ranging from simple t-tests or their non-parametric counterpart, Mann-Whitney U-tests (Siegel, 1956) to analysis of variance and multiple discriminant analysis (Polgar, 1975; Green and Vascotto, 1978).

In the simplest case, 2 sample-groups are compared. All observations of an environmental variable (e.g. temperature) pertaining to one sample group are tested against the corresponding observations of the other sample group. This is done using each suggested variable in turn and the ones that differ significantly are noted as being possible factors responsible for the biotic groups (Field, 1971). This approach has the usual drawbacks inherent in repeated significance tests and, because only one pair of sample-groups can be compared at a time, the overall effect of environmental parameters on the observed patterns is not clear. With ordination techniques it is often found that the samples or stations assume configurations which are orientated along particular dominating environmental gradients.

In the worked example below we have therefore ranked several environmental parameters and plotted them on the station configurations of the ordination to give a visual impression of possible correlations.

## A Worked Example

By way of illustration, we have subjected the data described by Warwick (1971) on the distribution of free-living nematodes in the Exe estuary to numerical analysis. In this paper, Warwick gave a subjective assessment of station and species groupings and of the main environmental factors controlling distribution: it is of some interest therefore to see whether numerical techniques bear out this subjective assessment.

Briefly, 4 more or less equally spaced transects on the east bank of the Exe estuary were sampled at 5 tide levels (except at Topsham where the sea-wall precluded sampling at MHWST), giving a total of 19 stations. Each station was sampled bimonthly between October 1966 and September 1967, at the time of the lowest monthly spring tides, by taking three 3.5 cm internal diameter cores of sediment to a maximum depth of 20 cm. The nematodes were subsequently extracted from the sediment, identified, and counted. On each sampling occasion several environmental parameters were measured: granulometry, organic content, interstitial salinity, depth of blackened H2S layer and depth of water table. An outline of the environmental conditions at each station is given in Table 1. Further details, together with tabulations of the nematode species composition at each station, are given in Warwick (1971). In general the muddy stations had a high population density but low species diversity and the sandy stations the reverse. The MHWST sandy stations at Lympstone and Shelly Bank had a low diversity/low density fauna, attributed by Warwick (1971) to seepage of low salinity coastal subsoil water.

## RESULTS

### Classification

Fig. 3 is a dendrogram showing station affinities, based on the mean root-root transformed abundance of all 182 species of nematodes found in the study, using the Bray-Curtis measure of similarity and group-average sorting. A broken line drawn at the arbitrary similarity level of 15 % delineates 4 major groups of stations: Group 1 divides into 2 homogeneous subgroups which have been designated 1A and 1B. Group 1A comprises all the Topsham stations, Group 1B the lower 3 tide-levels at Lympstone, Group 2 the MHWNT levels at Lympstone and Shelly Bank, Group 3 the MHWST levels at Lympstone and Shelly Bank, and Group 4 the remaining sand stations (the 3 lower levels at Shelly Bank and all the Orcombe Point stations).

### Ordination

Fig. 4 shows the results of multidimensional scaling using the same similarity matrix as above, delineating groups of stations from the dendrogram (Fig. 3). This

Table 1. Summary of environmental conditions at the 19 sampling stations

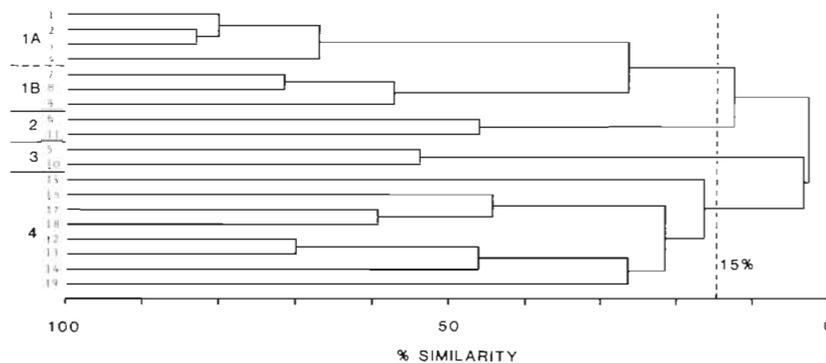| Locality | Station No. | Shore level | Sediment | Organic content (%) | Interstitial salinity | H$_2$S layer | Water table |
|---|---|---|---|---|---|---|---|
| Topsham | 1 | MHWNT | Mud | 6.43 | low | + | surface |
| (head of estuary) | 2 | MTL | Mud | 7.06 | low | + | surface |
| | 3 | MLWNT | Mud | 7.99 | low | + | surface |
| | 4 | MLWST | Mud | 7.15 | low | + | surface |
| Lympstone | 5 | MHWST | Coarse sand | 0.24 | v. low | − | absent |
| (mid-estuary) | 6 | MHWNT | Muddy-sand | 0.37 | moderate | + | variable |
| | 7 | MTL | Mud | 1.98 | moderate | + | surface |
| | 8 | MLWNT | Sandy-mud | 2.22 | moderate | + | surface |
| | 9 | MLWST | Mud | 5.88 | moderate | + | surface |
| Shelly Bank | 10 | MHWST | Sand | 0.09 | v. low | − | absent |
| (inside mouth | 11 | MHWNT | Muddy-sand | 0.39 | high | + | variable |
| of estuary) | 12 | MTL | Sand | 0.09 | high | − | absent |
| | 13 | MLWNT | Sand | 0.06 | high | − | absent |
| | 14 | MLWST | Sand | 0.09 | high | − | surface |
| Orcombe Point | 15 | MHWST | Sand | 0.06 | high | − | variable |
| (outside mouth | 16 | MHWNT | Sand | 0.04 | high | − | variable |
| of estuary) | 17 | MTL | Sand | 0.06 | high | − | variable |
| | 18 | MLWNT | Sand | 0.07 | high | − | variable |
| | 19 | MLWST | Fine sand | 0.09 | high | − | surface |

Fig. 3. Dendrogram showing classification of 19 stations in Exe estuary based on mean bimonthly abundances of nematodes over 1 y. Abundances were root-root transformed before comparing stations using the Bray-Curtis measure, and the dendrogram formed by group-average sorting. Four main station groups (1–4) are distinguished at an arbitrary similarity level of 15 % (X-axis)
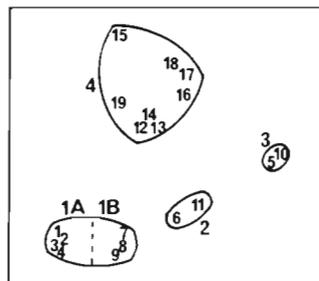


Fig. 4. Ordination of 19 stations in Exe estuary in 2-dimensions using multi-dimensional scaling on the same similarity matrix as Fig. 3. Clusters 1–4 and subclusters 1A and 1B were distinguished in the dendrogram and are superimposed here by encircling each cluster of stations. Axis scales are purely arbitrary and therefore not given

analysis gives essentially the same picture as the dendrogram: Stations 1, 2, 3, 4; Stations 7, 8, 9; Stations 6, 11, and Stations 5, 10 are closely clustered and conform to groupings defined from the dendrogram. Stations 14, 12 and 13 (the lowest 3 stations at Shelly Bank) form a tight cluster within the rather diffuse set of Group 4 stations.

MDS was first performed in 3 dimensions, with a resulting stress of 0.033. This configuration was then used as a starting 'map' for the 2-dimensional MDS, the latter then converging quickly to stress of 0.053 (Sufficient repetitions were carried out to ascertain with reasonable certainty that a global minimum had been obtained). Note that, unlike principal components based ordination methods, the 2-D configuration is not a projection of the higher dimensional solution onto some plane. Naturally, stress always increases as the dimensionality is reduced; the 1-D solution (not shown here) has stress 0.182.

The stress for the 2-D plot is rather low and the large reduction in passing from 1 to 2 dimensions and the comparatively slight decrease for the 3-D plot suggest

that a 2-D map adequately portrays the relationship between the stations. This is borne out by more formal reference to the simulation studies of Spence (1972) and Spence and Graef (1974), who give a detailed discussion of how to relate stress to the 'true' dimensionality of the data. However, just as it may be helpful to look at data plotted in the plane of the first 2 principal components when these do not account for most of the total variability, so a 2-D MDS plot can still be a useful tool when the stress indicates that the 'true' dimensionality is greater than 2. In the latter case though, it is strongly advisable to superimpose the results of a separate classification analysis as, for example, in Fig. 4.

The adequacy of the 2-dimensional representation is also apparent from the small residual variability about the regression line shown in the Shepard diagram (Fig. 5). The Shepard diagram is also a useful tool in detecting those stations least adequately represented by a 2-D configuration. Outliers in the plot could also indicate errors in particular dissimilarity values.

Both the classification and ordination discussed above are based on the mean abundances of species at each station from the seasonal series of samples. In order to justify the use of mean values (possible seasonal differences could invalidate this) we have run each individual sample on the MDS program to get a picture of the sample variability. To do this it has been necessary to divide the samples into a set of 55 summer and 52 winter (at present the program can cope with a maximum of 60 samples). Figs. 6a and 6b indicate that the replication is good: all replicate samples taken from the same station cluster in the same groups as Fig. 4, both in summer and winter. Note that in Figs. 4, 6a and 6b the relative positions of the clusters are similar but the exact configurations are arbitrary, and rotation and/or reflection of the axes would therefore be necessary in order to superimpose corresponding clusters.

Further discussion will only concern the station means data, since no seasonal differences are evident.

## Indicator Species

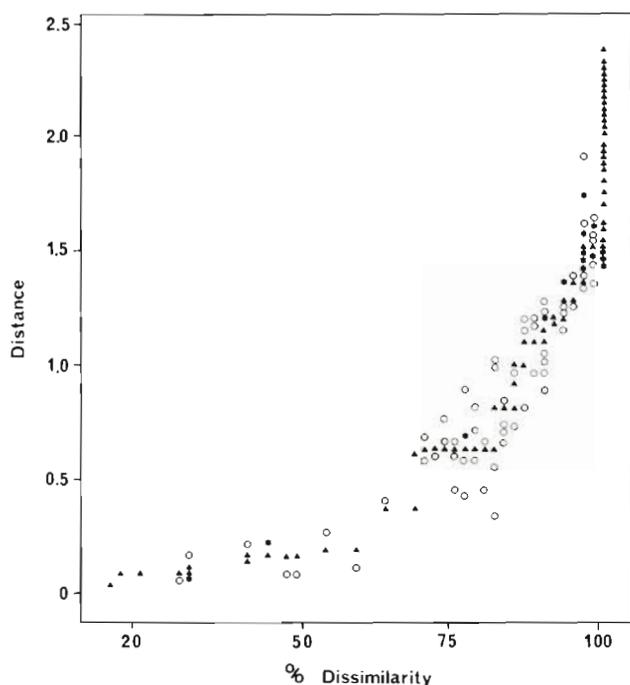Tables 2 and 3 list the species which are characteristic of Group 1 and distinguish it from Groups (2 + 3)



Fig. 5. Shepard diagram of MDS ordination shown in Fig. 4. Circles: scatter plot of inter-station distances $(d_{jk})$ for final 2-D configuration (Y-axis), against dissimilarities $(\delta_{jk})$ represented in dissimilarity matrix (X-axis). Triangles: monotonic, non-metric regression-estimated distances $(\hat{d}_{jk})$, plotted against dissimilarities $(\delta_{jk})$. Asterisks: 2 or more coincident points. Distances have arbitrary scale (Y-axis), while X-axis represents percent dissimilarity. Stress = 0.053 is a measure of scatter about regression-estimated values

Table 2. Frequencies of occurrence of indicator species, ranked according to information statistics, which distinguish stations of Group 1 from those of Groups (2 + 3). Rigorous statistical criteria are not met; species tabulated above the arbitrary horizontal line have $2 \triangle I > 6.63$, those below the line have $2 \triangle I > 3.84$. Number of occurrences in Group 1 are given, with those in Groups (2 + 3) in parentheses; maximum values are N 1 = 7 and N (2 + 3) = 4, respectively

| Species | Group 1 | (Groups 2+3) |
|---|---|---|
| *Mesotheristus setosus* | 7 | (1) |
| *Desmolaimus zeelandicus* | 7 | (1) |
| *Leptolaimus papilliger* | 5 | (0) |
| *Sabatieria pulchra* | 7 | (2) |
| *Cylindrotheristus oxycercus* | 7 | (2) |
| *Axonolaimus paraspinosus* | 4 | (0) |

* This species comprises less than 4 % of numbers at any one station and is excluded from the inverse species analysis (Table 6)

and Group 4 respectively. There are no 'perfect indicators', species which occur in all samples of Group 1 and in none of the compared groups, but *Mesotheristus setosus* and *Desmolaimus zeelandicus* are always present in Group 1 and occur very rarely (once in each comparison) in other groups. Information statistics reveal no species which distinguish Groups (2 + 3) combined, from Group 1; thus Group 1 is separated from Groups (2 + 3) by the presence of several additional species listed in Table 3. As another example, Tables 4 and 5 extract from the data the species characteristic of Group 4; *Sigmophora litoralis, Ditlevsenella danica* and *Theristus sp. G* occur in nearly all Group 4 stations and are absent from the other groups. The information tests thus provide a summary of which species differ most between the selected groups and are partly responsible for the separation of the groups in the classification and ordination analyses. A number



Fig. 6. MDS Ordination of replicate samples at 19 River Exe stations, based on root-root transformed abundances of 182 nematode species. Clusters delineated and numbered as in Figs. 3 and 4. Fig. 6a shows 55 summer replicate samples; Fig. 6b, 52 winter replicates
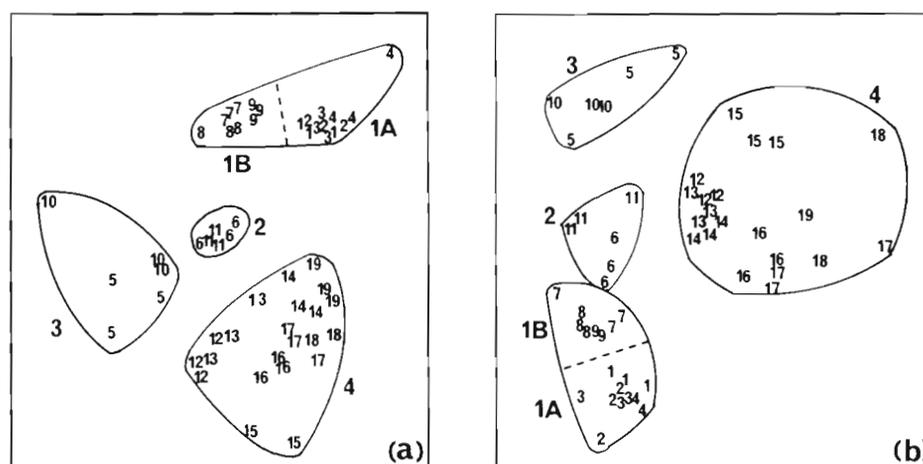
Table 3. Frequencies of indicator species, ranked according to information statistics, which distinguish stations of Group 1 from those in Group 4. Numbers of occurrences in Group 1 are given with those in Group 4 in brackets; maximum values are N 1 = 7 and (N 4 = 8), respectively

| Species | Group 1 | (Group 4) |
|---|---|---|
| Mesotheristus setosus | 7 | (1) |
| Desmolaimus zeelandicus | 7 | (1) |
| Sabatieria pulchra | 7 | (2) |
| Cylindrotheristus oxycercus | 7 | (2) |
| *Leptolaimus papilliger | 5 | (0) |
| Anoplostoma viviparum | 7 | (3) |
| Axonolaimus paraspinosus | 4 | (0) |
| Viscosia viscosa | 5 | (1) |
| Tripyloides gracilis | 3 | (1) |
| Sphaerolaimus hirsutus | 3 | (0) |
| Spilophorella paradoxa | 3 | (0) |
| Odontophora setosa | 3 | (0) |
| Oxystomatina elongata | 3 | (0) |
| Theristus procerus | 3 | (0) |

* This species comprises less than 4 % of numbers at any one station and is excluded from the inverse species analysis (Table 6)

of the species listed in Tables 2 to 5 may be useful indicators, but are not sufficiently dominant to be included in the inverse analysis which follows. It should be noted that the technique is most effective when the 2 groups being compared are both large, hence we combined Groups 2 and 3 in the comparisons.

Table 4. Frequencies of indicator species, ranked according to information statistics, which distinguish Group 4 from Group 1. There are no 'perfect indicators' present in all 8 stations of Group 4 and absent from all 7 stations of Group 1

| Species | N 4 = 8 | (N 1 = 7) |
|---|---|---|
| *Sigmophora litoralis | 7 | (0) |
| *Ditlevsenella danica | 6 | (0) |
| *Theristus sp. G | 6 | (0) |
| *Alaimella truncata | 5 | (0) |
| Epacanthion gorgonocephalum | 4 | (0) |
| *Chromaspirina parapontica | 4 | (0) |
| Oncholaimellus calvadoscus | 4 | (0) |
| *Thalassironus sp. | 4 | (0) |
| *Leptonemella sp. | 3 | (0) |
| *Linhomoeus sp. | 3 | (0) |
| *Theristus albigensis | 3 | (0) |
| *Trichenoplus sp. | 3 | (0) |
| *Linhomoeus sp. B | 3 | (0) |

* This species comprises less than 4 % of numbers at any one station and is excluded from the inverse species analysis (Table 6)

Table 5. Frequences of occurrence of indicator species, ranked according to information statistics, which distinguish Group 4 from Groups (2 + 3). There are no 'perfect indicators' present in all 8 stations of Group 4 and none of Groups (2 + 3)

| Species | N 4 = 8 | (N[2 + 3]) = 4 |
|---|---|---|
| *Sigmophora litoralis | 7 | (0) |
| *Ditlevsenella danica | 6 | (0) |
| *Theristus sp. G | 6 | (0) |
| *Alaimella truncata | 5 | (0) |
| Epacanthion gorgonocephalum | 4 | (0) |
| *Chromaspirina parapontica | 4 | (0) |
| Oncholaimellus calvadoscus | 4 | (0) |
| *Thalassironus sp. | 4 | (0) |

* This species comprises less than 4 % of numbers at any one station and is excluded from the inverse species analysis (Table 6)

## Inverse Analysis

For the species analysis we have reduced the total numbers from 182 to 55 by using only those species which have >4 % dominance at any one station. Species abundances have been standardised for each station as a percentage of the total abundance at all stations (i.e. if a species is found at only 1 station, its abundance there is 100 %). Fig. 7 shows the dendrogram for the inverse analysis. In order to define the same number of species groups as station groups we have drawn horizontal lines at 6 % similarity and at 8 % similarity giving 5 groups (if station groups 1A and 1B are considered separately). For convenience the species clusters on this dendrogram have been designated by the same notation as the station clusters, since they clearly correlate with them (see below). A list of the species in each cluster is given in Table 6. There is good general agreement between the species groups listed in Table 6 and the indicator species extracted by the information statistic technique. Five out of 6 species characterising Group 1 in Table 2 are in Group 1 of Table 6, the 6th being too rare for inclusion in the inverse analysis. Similarly, all but 1 of the included species characterising Group 1 in Table 3 are also clustered into Group 1 by the inverse analysis; Tripyloides gracilis is included in Group 2 (Table 6). However, only Epacanthion gorgonocephalum and Oncholaimellus calvadoscus are included in Group 4 by the inverse analysis (Table 6) and also distinguish Group 4 from Groups 1, 2 and 3 (Tables 4, 5). Thus it appears that a number of rarer species occur only in Group 4, and these are too rare to appear in the inverse analysis.

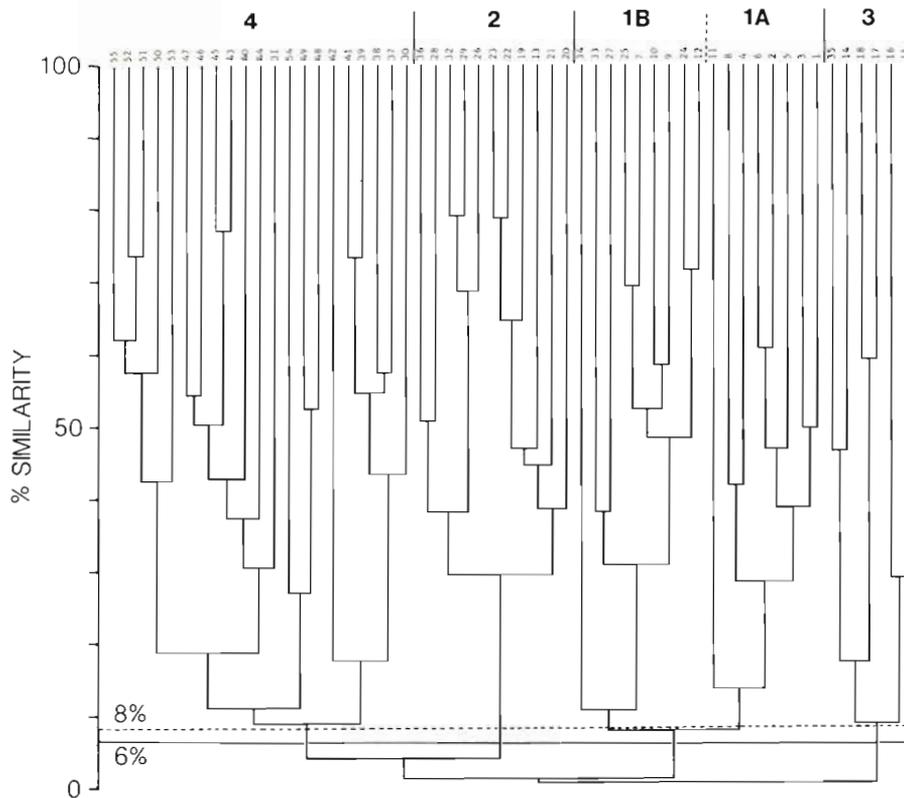Fig. 8a is the species analysis with multidimensional

Fig. 7. Dendrogram of inverse analysis comparing 55 nematode species occurring in more than 4% dominance at any of the 19 River Exe stations. Species abundances standardised and compared using the Bray-Curtis measure with group-average sorting. Species numbers listed and species named by cluster in Table 6

Table 6. Species groups distinguished by inverse (r-type) analysis. The groups are based on the dendrogram in Fig. 8; species numbers refer to numbers used in Fig. 7 and 8. Nomenclature follows that of Gerlach and Riemann (1973, 1974)

**Group 1A**
1 Mesotheristus setosus; 3 Sabatieria pulchra; 5 Hypodontolaimus geophilus; 2 Anoplostoma viviparum; 6 Desmolaimus zeelandicus; 4 Axonolaimus spinosus; 8 Adoncholaimus thalassophygas; 11 Theristus flevensis

**Group 1B**
12 Axonolaimus paraspinosus; 24 Paracanthonchus punctatus; 9 Viscosia viscosa; 10 Ptycholaimellus ponticus; 7 Cylindotheristus oxycercus; 25 Sphaerolaimus hirsutus; 22 Odontophora setosa; 33 Atrochromadora microlaima; 34 Terschellingia communis

**Group 2**
20 Paracanthonchus tyrrhenicus; 21 Tripiloides gracilis; 13 Oncholaimus brachycercus; 19 Ascolaimus elongatus; 22 Adoncholaimus fuscus; 23 Theristus acer; 26 Microlaimus honestus; 29 Microlaimus robustidens; 32 Trefusia longicaudata; 28 Theristus normandicus; 36 Bathylaimus assimilis

**Group 3**
15 Tripyla so.; 16 Rhabditid; 17 Dorylaimid; 18 Eurystomina terricola; 14 Bathylaimus stenolaimus; 35 Paracyatholaimus intermedius

**Group 4**
30 Enoplus brevis; 37 Mesacanthion africanthiforme; 38 Enoploides brunettii; 39 Enoplolaimus litoralis; 41 Trissonchulus benepapillosus; 42 Araeolaimus elegans; 48 Epacanthion gorgonocephalum; 49 Enoplolaimus denticulatus; 54 Axonolaimus hexapilus; 31 Chromadora nudicapitata; 44 Mesacanthion hirsutum; 40 Enoplolaimus propinquus; 43 Dichromadora hyalocheile; 45 Praeacanthonchus opheliae; 46 Viscosia cobbi; 47 Sigmophora rufum; 53 Bathylaimus paralongisetosus; 50 Pomonema reducta; 51 Oncholaimellus calvadoscus; 52 Axonolaimus orcombensis; 55 Chromaspirina inglisi

scaling; the species groups are delineated from the dendrogram. Fig. 8b shows the same configuration with the species numbers replaced by the station group(s) in which they are found to represent more than 4% of specimens in any one station. This indicates a high degree of agreement between the species and station clusters. Although the classification artificially forces a species to occur in only 1 group, it will be seen from Fig. 8b that, in this case, rather few species occur in significant numbers in more than 1
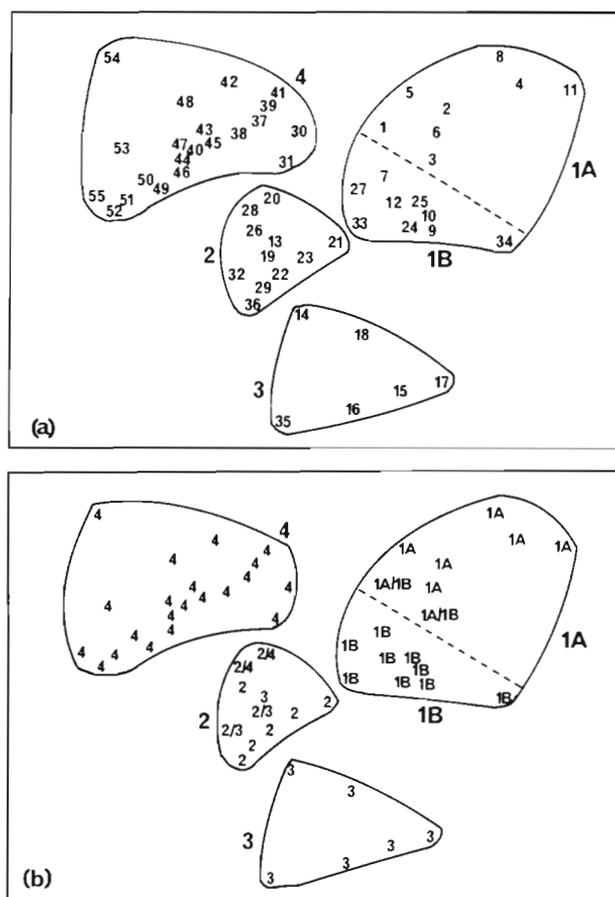
Fig. 8. MDS Inverse ordination comparing 55 nematode species of the River Exe, using standardised abundances and Bray-Curtis measure. Species groups delineated from dendrogram in Fig. 7. Fig. 8a shows relationships between species, whose numbers and names are listed in Table 6. In Fig. 8b the nematode species numbers have been replaced by station group(s) in which they represent more than 4% of specimens at any one station (see text). Most species are strongly related to particular station groups, although some Group 2 species are also seen to occur in Groups 3 or 4

group. The main exception to this is the central group of mean high water neap tide (MHWNT) muddy-sand stations (Group 2) which contain several species from Groups 3 and 4. This figure provides a good summary of the differences in species composition between the station groups.

## DISCUSSION AND CONCLUSIONS

The results presented above give an example of some analyses that can be applied to the biological data. In accordance with our strategy (p. 38), no use has been made of the environmental data. We now relate the groups of stations distinguished on the basis of species abundances to environmental data.

## Relation of Station Groups to the Environment

Figs. 9a to 9e represent the same station configurations as Fig. 4 with a variety of physico-chemical properties of the stations superimposed. There are important correlations between the clusters and the granulometry, the organic content, and depth of $H_2S$, and also with the interstitial salinity. Similar plots showing % silt-clay, not reproduced here, give the same results. These plots show quite convincingly, for example, that Groups 1A and 1B, and Groups 3 and 4 are distinguished from each other on salinity characteristics but not on sediment granulometry, whereas with Groups 1B and 2 the reverse is the case. On the other hand there appears to be no overall effect of certain other characters on the clustering, for example the height of the station above chart datum (Fig. 9e) and the depth of the water table (not shown).

## Characterisation of Species Groups

The species associations defined in the inverse analysis will clearly be confined to their respective station groups because of certain biochemical, physiological, morphological or behavioural adaptations.

In an earlier paper Warwick (1971) compared the distribution of several physiognomic characters of the nematodes in different habitats: (1) The feeding type (as deduced from the structure of the buccal cavity: 1A – selective deposit feeders, 1B – non-selective deposit feeders, 2A – epigrowth feeders, 2B – predators/omnivores. (2) Length of cephalic or body setae. (3) Body length. (4) Cuticular pattern: smooth, transversely striated or punctated. (5) The presence or absence of visual pigments.

In the same way that physico-chemical characters were plotted on the stations MDS output, so it is possible to plot these physiognomic characters on the species MDS configuration to investigate possible correlations of these characters with the clusters. Figs. 10 a–d show these plots for 4 of the characters (very few species had visual pigments, and this plot has been omitted).

The main features of these plots are: (1) Feeding types (Fig. 10a): The high proportion of predators/omnivores in Group 4 (14 out of a total of 21 species), and the high proportion of non-selective deposit feeders (6 out of 8) in Group 1A. (2) Setal length (Fig. 10b): A general tendency for increase in setal length from bottom right to top left of the configuration, culminating in a high proportion of species with long setae in Group 4. (3) Body length (Fig. 10c): Groups 1, 2 and 3 with a fairly equitable distribution of small and large
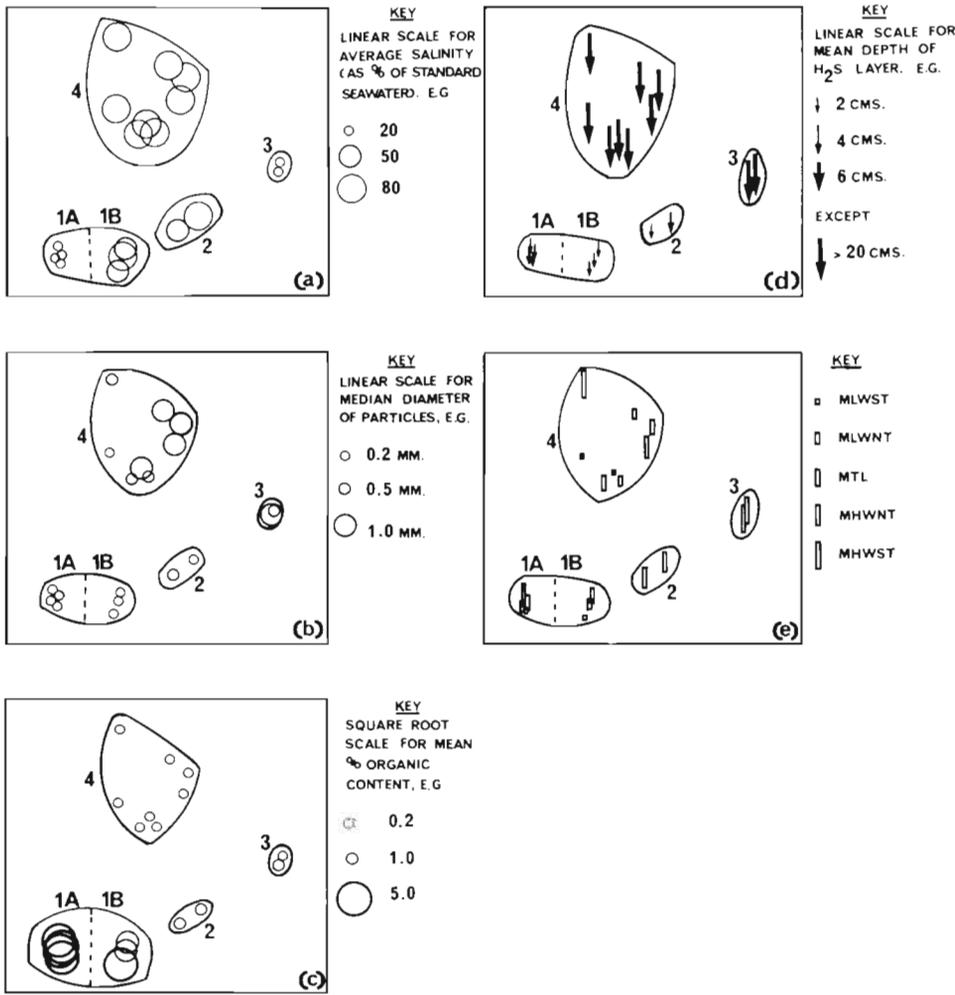
Fig. 9. Relation of station groups to environmental factors. MDS plot of 19 River Exe stations based on nematode fauna with clusters delineated as in Fig. 4. (a) At each station circles proportional in diameter to mean salinity of interstitial water are superimposed. (b) Relation of station groups to particle size, with circles proportional in diameter to average median sediment particle size at each station. (c) Relation of station groups to sediment organic content: circle diameters represent mean % organic content at each station on a square root scale. (d) Relation of station groups to depth of $H_2S$ layer in sediment: arrow lengths proportional to mean depth of $H_2S$ layer at each station. (e) Relation of station groups to tidal level: columns indicate height of each station above chart datum
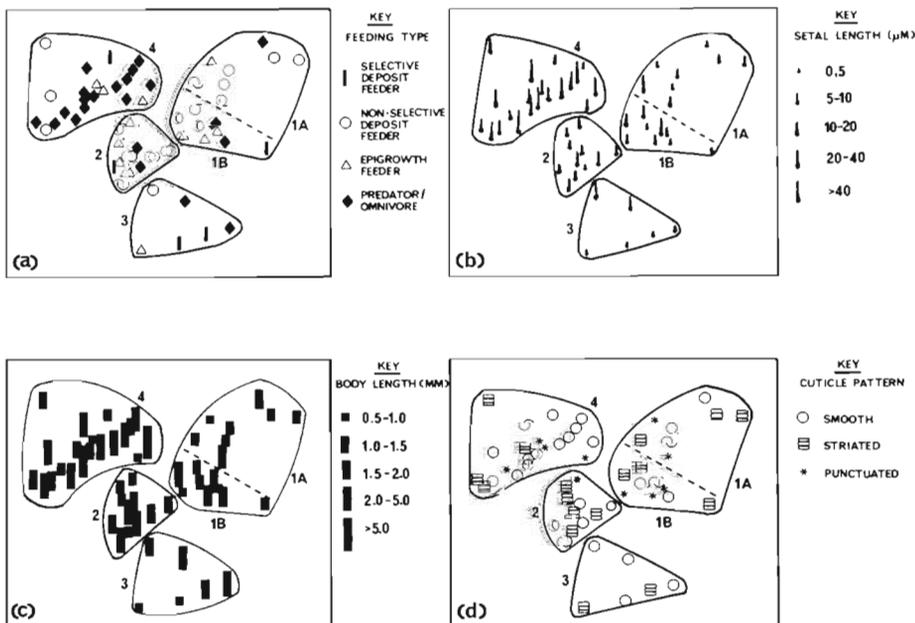


Fig. 10. Relation of species groups to morphological characteristics. MDS Plot of 55 nematode species in the River Exe with clusters delineated as in Fig. 8. (a) Superimposed symbols: feeding categories. (b) Relation of nematode species groups to setal length, with superimposed symbols scaled to setal length categories. (c) Relation of nematode species groups to body length, with superimposed symbols proportional to body length categories. (d) Relation of nematode species groups to cuticle pattern, with symbols representing cuticle patterns

bodied species, but there is a suggestion that Group 4 has proportionally fewer small (or more large) species. (4) Cuticle pattern (Fig. 10d): No marked difference in the distribution of cuticle patterns between groups.

The statistical significance of the differences noted above can be examined by a chi-square test. For a 2-way categorisation into 'feeding type' and 'species cluster', a test of the null hypothesis that there is no association between the classifications is just rejected at the 5 % level, whereas independence of 'cuticle pattern' and 'species cluster' cannot be rejected. The continuous variables can be categorised and a chi-square test applied similarly; 'setal length' is seen to be very strongly associated with species group ($x^2 r =$ 29.6 on 9 degrees of freedom), but the suggestion that 'body length' differs between the clusters is not strong enough to attain statistical significance.

As an alternative to categorisation, the observation that setal length tends to increase along in axis from bottom right to top left of the MDS plot can formally be verified by multiple linear regression. For the regression of log setal length of the two MDS co-ordinates, the slope of the fitted plane is in the anticipated direction and is very significantly non-zero (F = 12.5 on 2,52 degrees of freedom). For body length regressed in the same way, the slope does not achieve significance, in agreement with the earlier $x^2$ test.

## Main Conclusions from Worked Example

These analyses show that, in the Exe estuary, there are 4 major groups of stations representing, in physico-chemical terms, 4 major habitats each characterised by a set of nematode species which have morphological adaptations to suit the habitat.

(1) Muddy stations with a high organic content and a blackened anoxic layer a few cm below the surface, characterised typically by small deposit-feeding nematodes with short setae. Salinity either low (Group 1A, Stations 1–4) or moderate (Group 1B, Stations 7–9).

(2) Muddy-sand stations at MHWNT with moderate salinity and organic content with blackened $H_2S$ layer at variable depth. Physiognomic characters of the nematodes intermediate between mud (Group 1) and sand (Group 4) stations. Stations 6 and 11.

(3) Well-drained sands at MHWST with moderate organic content, very low interstitial salinity and no $H_2S$. Rather few species so that the distribution of physiognomic characters cannot be considered, but 3 of the 6 species in this group belong to fresh-water genera. Stations 5 and 10.

(4) Clean sands, high salinity, no $H_2S$, low organic content. A high proportion of large predatory or omnivorous nematode species with long setae which enable them to maintain their position in this highly dynamic environment. Stations 12–19.

Although these conclusions broadly agree with the subjective assessment made earlier (Warwick, 1971), they differ somewhat in detail. Warwick did not note the separation of Group 1 stations into 2 coherent sub-groups on the basis of salinity, and included Station 8 from Group 1B with Group 2 stations because of its slightly coarser sediment. The present study shows quite clearly that, faunistically, Station 8 belongs with Group 1B. Further, Warwick divided the Group 4 stations into 3 groups: Stations 12 and 13 – coarse sands which dry out at low tide, Stations 14 to 18 coarse sands with a more or less permanent water table, and Station 19 fine stable sand retaining a permanent water table. The present study suggests that there was no basis for this separation and that the water table or the precise grade of sand have little effect on the clustering of stations within Group 4. Warwick's observations on the distribution of physiognomic characters, bearing in mind some regrouping of stations, are largely borne out by the present study.

## General Conclusion

Application of our strategy has given as a first result a dendrogram which divides the stations into 4 groups, one of which is subdivided. The simplicity of the classification is useful for presentation but may force the data into artificially distinct classes when continua exist, hence a complementary method of analysis is advisable. Multi-dimensional scaling (MDS) confirmed the existence of clear groups which were emphasized by delineating the dendrogram groups on the MDS plot. Inverse MDS analysis (after excluding the rarer species to reduce computer storage and time) produced clusters of species which correspond well with the station groups. Information tests were used to find 'indicator' species characteristic of one group of stations and absent from another group.

The second stage of analysis, according to our strategy, is to relate to environmental data the groups formed by analysing the biotic data. This was done by superimposing a scaled symbol, representing one environmental variable, onto the MDS plot of the stations obtained previously. This demonstrated clearly whether a factor, such as salinity, differs markedly between the station groups. We could have used statistical tests (e.g. ANOVA or Kruskal-Wallis tests) to test the significance of differences in each environmental variable among the station groups; this was unnecessary in our worked example because of the very marked differences.

Finally, the same graphical approach of superimpos-

ing symbols on the MDS plots can be used to show relationships between attributes of the species (here feeding type and morphometry) on the one hand, and the species groups on the other. This enables one to synthesize results from complex biotic and environmental data in terms of adaptations to the habitat, as has been demonstrated by the example of estuarine nematodes.

## LITERATURE CITED

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

Campbell, B. M. (1978). Similarity coefficients for classifying releves. Vegetatio 37: 101–109

Cassie, R. M. (1961). The correlation coefficient as an index of ecological affinities in plankton populations. Memorie Ist. ital. Idrobiol. 13: 151–177

Cassie, R. M. (1972). Fauna and sediments of an intertidal mud-flat: an alternative multivariate analysis. J. exp. mar Biol. Ecol. 9: 55–64

Cassie, R. M., Michael, A. D. (1968). Fauna and sediments of an intertidal mud flat: a multivariate analysis. J. exp. mar Biol. Ecol. 2: 1–33

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

Day, J. H., Field, J. G., Montgomery, M. P. (1971). The use of numerical methods to determine the distribution of the benthic fauna across the continental shelf of North Carolina. J. Anim. Ecol. 40: 93–126

Fager, E. W., Longhurst, A. R. (1968). Recurrent group analysis of species assemblages of demersal fish in the Gulf of Guinea. J. Fish. Res. Bd Can. 25: 1045–1421

Fager, E. W., McGowan, J. (1963). Zooplankton species groups in the North Pacific. Science, N.Y. 140: 453–460

Field, J. G. (1969). The use of the information statistic in the numerical classification of heterogeneous systems. J. Ecol. 57: 565–569

Field, J. G. (1970). The use of numerical methods to determine benthic distribution patterns from dredgings in False Bay. Trans. R. Soc. S. Afr. 39: 183–200

Field, J. G. (1971). A numerical analysis of changes in the soft-bottom fauna along a transect across False Bay, South Africa. J. exp. mar Biol. Ecol. 7: 215–253

Field, J. G., McFarlane, G. (1968). Numerical methods in marine ecology. I. A quantitative similarity' analysis of rocky shore samples in False Bay, South Africa. Zool. afr. 3: 119–138

Gerlach, S. A., Riemann, F. (1973). The Bremerhaven checklist of aquatic nematodes (Part I). Veröff. Inst. Meeresforsch. Bremerh. 1 (Suppl. 4): 1–404

Gerlach, S. A., Riemann, F. (1974). The Bremerhaven checklist of aquatic nematodes (Part 2). Veröff. Inst. Meeresforsch. Bremerh. 2 (Suppl. 4): 405–734

Green, R. H., Vascotto, G. L. (1978). A method for the analysis of environmental factors controlling patterns of species composition in aquatic communities. Wat. Res. 12: 583–590

Haedrich, R. L., Rowe, G. T., Polloni, P. T (1980). The megabenthic fauna in the deep sea south of New England, USA. Mar. Biol. 57: 165–179

Hughes, R. N., Thomas, M. L. H. (1971). The classification and ordination of shallow-water benthic samples from Prince Edward Island, Canada. J. exp. mar Biol. Ecol. 7: 1–39

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

Kruskal, J. B. (1964). Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis. Psychometrika 29: 1–27

Kruskal, J. B. (1977). Multidimensional scaling and other methods for discovering structure. In: Enslein, Ralston, Wilf (eds.) Statistical methods for digital computers. John Wiley, New York

Lance, G. N., Williams, W. T (1967). A general theory of classificatory sorting strategies. I. Hierarchical systems. Computer J. 9: 373–380

Lie, U., Kelley, J. C. (1970). Benthic infauna communities off the coast of Washington and in Puget Sound: identification and distribution of the communities. J. Fish. Res. Bd Can. 27: 621–651

McConnaughey, B. H. (1964). The determination and analysis of plankton communities. Marine Research in Indonesia (Spec. No.): 1–25

Noy-Meir, I. (1970). Component analysis of semi-arid vegetation in southeastern Australia. Ph. D. thesis, Australian National University, Canberra

Noy-Meir, I. (1973). Data transformations in ecological ordinations. I. Some advantages of non-centering. J. Ecol. 61: 329–342

Peters, J. A. (1971). A new approach in the analysis of biogeographic data. Smithson. Contrib. Zool. 107: 1–28

Polgar, T. T. (1975). Characterization of benthic community responses to environmental variations by multiple discriminant analysis. In: Saila, S. B. (ed.) Fisheries and energy production. Lexington Books, D. C. Heath Co., Mass., pp. 267–293

Poore, G. C. B., Mobley, M. C. (1980). Canonical correlation analysis of marine macrobenthos survey data. J. exp. mar. Biol. Ecol. 45: 37–50

Popham, J. D., Ellis, D. V. (1971). A comparison of traditional cluster and Zurich Montpelier analyses of infaunal pelecypod associations from the adjacent sediment beds. Mar. Biol. 8: 260–266

Sanders, H. L. (1958). Benthic studies in Buzzards Bay. I. Animal-sediment relationships. Limnol. Oceanogr 3: 245–258

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

Shin, P. K. S. (1982). Multiple discriminant analysis of macrobenthic infaunal assemblages. J. exp. mar. Biol. Ecol. (in press)

Siegel, S. (1956). Non-parametric statistics for the behavioral sciences, McGraw Hill, New York

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

Spence, I. (1972). An aid to the estimation of dimensionality in non-metric multidimensional scaling. Univ. of Western Ontario Res. Bull. No. 229

Spence, I., Graef, J. (1974). The determination of the underlying dimensionality of an empirically obtained matrix of proximities. Multivar. Behav. Res. 9: 331–342

Stephenson, W. T., Burgess, D. (1980). Skewness of data in the analysis of species-in-sites-in-times. Proc. R. Soc. Queensland 91: 37–52

Stephenson, W., Cook, S. D. (1980). Elimination of species prior to cluster analysis. Austr. J. Ecol. 5: 263–273

Stephenson, W., Cook, S. D., Raphael, Y. I. (1977). The effect of a major flood on the macrobenthos of Bramble Bay Queensland. Mem. Qd Mus. 18: 95–119

Stephenson, W., Williams, W. T., Cook, S. D. (1972). Computer analyses of Petersen's original data on bottom communities. Ecol. Monogr. 42: 387–415

Thorrington-Smith, M. (1971). West Indian Ocean phytoplankton: a numerical investigation of phytohydrographic regions and their characteristic phytoplankton associations. Mar. Biol. 9: 115–137

Torgerson, W. S. (1958). Theory and methods of scaling, Wiley, New York

Velimirov, B., Field, J. G., Griffiths, C. L., Zoutenkyk, P. (1979). The ecology of kelp bed communities in the Benguela upwelling system. Analysis of biomass and spatial distribution. Helgoländer wiss. Meeresunters. 30: 495–518

Walker, H. A., Saila, S. B., Anderson, E. L. (1979). Exploring data structure of New York Bight benthic data using post-collection stratification of samples, and linear discriminant analysis of species composition comparisons. Estuar. coast. mar. Sci. 9: 101–120

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

Whittaker, R. H., Gauch, H. G. (1973). The Braun-Blanquet approach. In: Whittaker, R. H. (ed.) Ordination and classification of communities, Handbook of vegetation science, Part 5. Junk, The Hague, pp. 289–321